Designing AI Systems that Preserve and Promote Human Creative Agency

Eran Barak-Medina[1]

[1] *Holon Institute of Technology, Instructional Technologies Faculty*

Abstract

Artificial Intelligence (AI) systems are increasingly integrated into cognitively demanding and creative workflows—yet their influence on human agency remains under-theorized. This paper posits that beyond usability and accuracy, AI tools must be evaluated by their effects on *human creative agency*—the motivation, ownership, and accountability individuals feel when facing complex challenges. Drawing from current literature and professional practice, a conceptual framework to analyze how perceptions of AI functionality shape creative agency is proposed. Specifically, the framework identifies four factors: whether AI is perceived as *complementing* or *competing* with one's skills; its *perceived effectiveness*; the *stakes* of the task; and the user's level of *AI literacy*. The article explores how these factors can either sustain or diminish creative agency, with implications for the design of agency-aware AI systems. It concludes by outlining a preliminary method for operationalizing these factors into actionable principles for AI system design.

Keywords

Human-Centered AI, Creative Agency, Human-AI Collaboration, AI Literacy

1. Introduction

The goal of this article is to propose a practical approach to designing Artificial Intelligence (AI) systems and the interactions people can have with them, that not only avoid hindering but enhance human creative agency. Human *creative agency* is a person's overall motivation and sense of ownership and accountability in engaging with challenges that require a high level of cognitive ability, such as creative tasks, trying to acquire a new and challenging skill, complex decision-making or problem-solving [1]. Without creative agency related to different activities and challenges, people might refrain from those activities and challenges, and those would no longer be perceived by them as part of their identity, or as activities or challenges they acknowledge themselves as accountable for. In certain contexts, AI, its functionality and the way it is perceived, might lead to decreased creative agency. On the other hand, in different contexts and Human-AI dynamics, AI's perceived functionality can enhance creative agency, and help people perceive themselves as able to achieve more, while still identifying the achievement as their own and as an expression of their abilities, goals, ownership and accountability. This article suggests ways that developers of AI can design their systems in alignment with principles that promote human creative agency.

AI technology is developing today in an extremely fast and continuous way. AI, and especially Generative AI (GenAI) demonstrates capabilities of performing functions that until now belonged exclusively to the human realm, like reasoning, problem-solving, decision-making, producing creative outputs, etc., sometimes at an even greater quality and efficiency than most or all people. The reality in which we, people, share the world with a technology or an agent capable of high-

[1]Proceedings of the first SYNERGY Workshop 2025, June 9, 2025, Pisa, Italy

Eranba@hit.ac.il (E. Barak-Medina)

https://orcid.org/0009-0002-8542-0305 (E. Barak-Medina)

cognitive performance, brings forth many questions about human agency, identity, and uniqueness.

In this context, we must question the ways that people and AI can interact and collaborate in a manners that enhances people (and humanity in general), giving us a stronger sense of agency and efficacy, stronger ability to solve problems and promote desired outcomes, while leaving us with a coherent perception of identity and worth. This issue is the central objective of the field of Human-centered AI (HCAI). HCAI is defined as the design and development of AI systems that prioritize human needs, values and agency in the way they operate, rather than aiming to replace human [2]. Shneiderman [3] claims that it is possible to enable a high level of automation, and a high level of human control at the same time.

However, this is far from being an easy or already figured-out challenge. Evidence shows that using AI in various domains and for various tasks, can take a toll in diluting human skills [4], decreasing human learning effectiveness [5], and hindering a sense of agency [6]. On top of that, in certain use-cases there are even findings that shows that human involvement together with AI, might decrease the quality and efficiency of what can be achieved by AI with minimal human intervention [7], which could potentially, perhaps even rightfully so, trigger the question of the human added-value in those case-studies, challenges and domains entirely. Many AI systems are being developed with technologically and functionality-driven visions, while not fully considering the way people will interact with the system and the way it might enhance or decrease their agency.

Within the field of HCAI, several approaches and methodologies have been offered to guide the design of AI systems and the shaping of Human-AI collaboration loops and processes, that are meant to protect human control and enable an effective human-AI collaboration, among which consideration for human agency [3, 8, 9, 10]. However, these methodologies focus often place an emphasis on designing an effective collaborative work process, or on defining principles for this design, and offer less consideration to the practical ways that the interaction with AI can promote or decrease human creative agency in the context.

This article suggests that considering human creative agency and how it could be impacted within human-AI interaction, is an essential component to any design process of an AI system that strives to develop Human-centered AI. It proposes that for human–AI synergy to emerge, AI systems must be evaluated not only by their accuracy or fluency, but by their capacity to sustain and expand human creative agency. The article will also offer workflow analysis principles that could help developers who wish to design Human-centered AI systems that enhance creative agency.

2. The AI-Creative Agency Framework

Creative agency, or the willingness to invest in meeting challenges that require creative and high-level cognitive effort, is a crucial element in many of humanity's cultural achievements. Without it, people would not be involved in artistic expression, scientific exploration, economic growth and entrepreneurship, and many other challenges and activities that help society progress and overcome obstacles. It is also important for the growth of many professionals, experts and innovators, as they need to go through a series of challenges that require this kind of willingness in order to become fully mature professionals. Any creative outcome, any innovative solution, even any day-to-day achievements that are needed for us to survive and prosper (medical decision, public administration, teaching, and many more), are dependent on people who have the motivation to try to accomplish something which' outcome demands their highest skills and is not guaranteed. Therefore, all creativity (in art, business, science, etc.) is dependent on creative agency.

The relationship between AI and human creativity is complex. GenAI is displaying constantly increasing capabilities to produce outputs that would require a person to employ his creative and high-level cognitive skills. Pictures, paintings, stories, videos, data-based insights, articles, strategies, medical diagnosis and treatment plan, research ideas and more – these are all outputs

that GenAI can now produce at ease, fast, and at a quality that surpasses what most people can do. These are all outputs that in the past, and today as well, would require a person a lot of time, dedication and skill. Once AI can produce these outputs at a certain level, it poses a question for the person on how to relate to AI, and how to define his own merit and the challenges he would take upon himself. For our discussion's purpose, AI has the potential to directly impact people's creative agency.

One of the familiar principles in HCAI is the idea of "Human-in-the-loop", suggesting that while designing AI systems, it is advised to reserve control for the person within the collaborative workflow. However, even a "Human-in-the-loop" approach, while preserving human control and regulation, does not necessarily consider the specific elements in the workflow that could impact human creative agency, nor does it indicate a commitment to assigning human users with a role that emphasizes their added-value so the collaborative process would benefit from it.

The *AI-Creative Agency Framework* [1] tries to map the factors that might bridge between AI and human creative agency. The framework revolves around two assumptions: (a) That the way AI impacts human creative agency goes through how people perceive AI's functionality. (b) That this perception of functionality is task-specific. In other words, a person's creative agency would connect to a specific task, challenge or function, and would be affected not by the AI itself, but by how the person perceives what AI can do in that task.

Upon these two assumptions, there are three interacting factors in a person's perception of AI's functionality that would impact his creative agency (whether the AI is competing or complementing, AI's perceived effectiveness, and whether the function is high-stakes or low-stakes), and one factor that relates to the person's own attributes (his AI literacy). Following is a review of the four factors and their potential impact on human creative agency.

2.1. Competing or Complimenting AI

One of the first questions a person might ask regarding the AI's functionality, is – *"does it do what I do?"* If the answer is *"yes"* that would mean that the AI system is competing with the person's skills. If the answer is either *"no"*, *"yes – but partly"*, or *"no – but what it does could be beneficial for me"*, the system could be considered as complementing the person's skills. In general, a competing system would more likely hurt the person's creative agency (although this could be influenced by the other factors described later), and a complementing system would either be perceived as irrelevant or could enhance creative agency. A complementing system might help a person realize he could use the AI to handle parts of a workflow in a way that makes the workflow more effective and empowers the person to utilize his skills better.

2.2. Perceived Effectiveness of the AI system

Another question that a person would ask regarding the AI, is *"is it better than me?"*. Another variant of this comparison might be not personal but comparing the AI to other people in the person's profession or in this challenge. This question can also be asked in a social context by people who interact with the person, for example: clients who seek graphic design would compare between a human graphic designer and an AI, and their conclusion would have implications on the graphic designer as well. At face level, an AI system that is perceived as more effective than people, would have a negative effect on people's creative agency regarding that task or skill, and an AI system that is perceived as less effective, would have a neutral or slight positive effect on creative agency. But this is also interacting with the first factor – because if an AI system is perceived as complementing and highly effective – it could greatly boost human creative agency. For example: If an AI system upgrades my spelling and phrasing, which is part of my article writing process, then its strong effectiveness would encourage me to take more and tougher writing challenges (meaning my creative agency has increased).

2.3. High Stakes or Low Stakes

Another factor that can mediate how AI can influence human creative agency is whether the task, skill or challenge that is evaluated is high stakes or low stakes. The meaning of "high stakes" is the cost of mistake. Medical diagnosis, air traffic, law enforcement, pardon requests decisions, these are examples for decision-making that can vastly impact people's lives, health or well-being, and would be considered high stakes. High stakes decisions or challenges would create pressure (personal and social) and a tendency to prefer the more accurate decision or problem-solving process possible. Most patients would prefer being diagnosed accurately over preserving the physician's creative agency, as most physicians would, when given certainty and evidence, rather delegate the task to the superior mechanism over making a mistake. In this sense, the "stakes" factor, acts as an amplifier of preferring the AI system, once it's perceived as more effective and competing, thus decreasing the human creative agency to perform that task. On the other hand, the low stakes context might defuse the potential negative effect of a competing and effective AI system. An example for this is in games – where AI could perhaps enhance the gamer's abilities over non-AI-using gamers, yet that gamer will feel that using AI decreases his enjoyment of the game and therefore will not use it. As much as gaming is important to gamers' identity and well-being, winning the game is not a high stakes challenge (and certainly inferior to enjoying the game and expressing oneself), hence when it comes to using AI, it would be considered as a low stakes challenge.

2.4. AI Literacy

AI literacy is not an attribute of the AI system and how it is perceived, but rather a quality within the person in the human-AI interaction. AI literacy is a multi-facet concept that is defined in many ways [11, 12, 13, 14]. For this article and the AI-creative agency framework, it is not required to define the entire concept but rather relate to two of its main attributes: (a) The ability to learn how to use AI effectively, assess its functionality and incorporate it into the person's workflow and decision-making. Overall, people whose ability to learn new AI tools and use them effectively, are expected to have their creative agency strengthened, because they would realize they harness AI to achieve more. However, they could also be the first to recognize that AI makes their own skills redundant. (b) A less discussed facet of AI literacy is people's "self-management mindset" with AI. This idea refers to the way people approach using AI and their expectations regarding it. Watkins et al. [15] have described four such mindsets:

- Empowerment: This is a mindset that starts with the question – "how can me together with AI could achieve better than me alone and AI alone?". A person holding this mindset will seek ways to enhance his workflow and benefit from where AI offers an advantage, and where he has the added value to the process and maintains control and strategic intent. A person holding this mindset would also cope better with AI becoming competing and better than himself, because he might look for ways to take advantage of this growth to try and accomplish better results and create a new workflow and redefine his own contribution. This mindset would likely enhance a person's creative agency.
- Delegation: This mindset perceives AI as an opportunity to offload tasks to AI. The person operating from this mindset expects the AI to replace his own contribution. This mindset, for the most part, would potentially lead to decreasing creative agency and form a dependency on AI. When delegating is part of a more elaborate workflow that combines AI's and human contributions, in a deliberate way, while stressing the human unique contribution, then it will not count as "delegation" mindset, but rather an" empowerment" mindset.
- Avoidance: This mindset perceives AI as either a threat, a too-big challenge, or belittles its capabilities, to the point of not using it. In the short-term, this mindset might not have an impact on the person's creative agency, but in the long-term this person might realize the effectiveness by which he meets his challenges is topped by AI or people who use AI, which would also lower his creative agency.
- Suspension: This mindset and "self-management of using AI" is a deliberate decision to not use AI, even when knowing the advantages of AI. Suspension is a decision most likely to be

taken mainly in two contexts: (1) When the activity is enjoyable, and using AI would take the fun out of it (i.e., the gamer, and any hobby or cherished activity). (2) When one realizes that for the development of one's own skills, it is preferable to suspend using AI until new skills are honed. Self-managing the interaction with AI based on the suspension mindset would most likely enhance creative agency.

AI literacy is not, however, a factor that developers of AI systems can usually control, meaning that the main factors they should consider are the functionality elements of the system they try to develop.

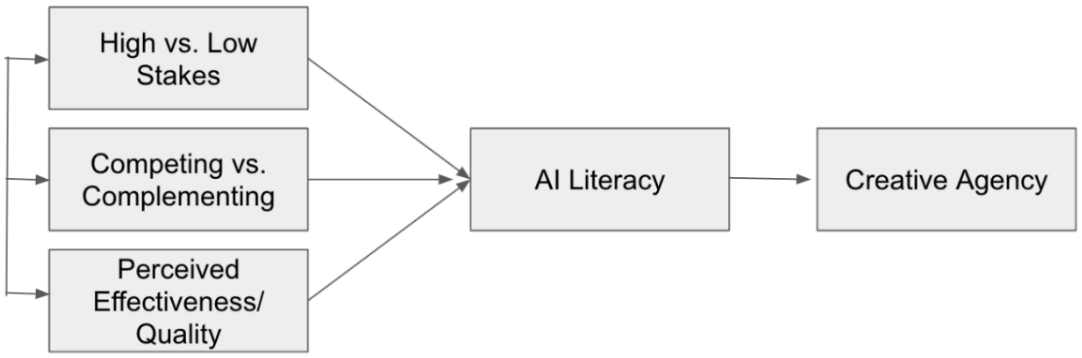

Figure 1: Proposed factors for AI's influence on human creative agency [1].

2.5. The AI-Creative Agency Framework's Usages

The AI-Creative Agency framework can serve two major purposes:

i. It can explain and predict the effects of AI, currently and in the future, on people's creative agency, across domains, activities and skills. Since AI keeps being developed and improved, even domains and activities that current AI is either complementing or ineffective and the people involved it them have their creative agency intact or enhanced, might change when AI gets better, and its functionality evolves.

ii. The framework can serve as a tool for designing Human-centered AI systems, when preserving or enhancing human creative agency is a desired value.

3. How to Design AI systems that Enhance Human Creative Agency?

There is an assumption that underlies this article, which is that designing AI systems that maintain and enhance human creative agency is desirable. This assumption is aligned with the values of the Human-Centered AI field. It could be argued that prioritizing the maintenance and enhancement of human creative agency might come with a cost of overall performance, that it would require restricting the system's capabilities and functions to allow people to keep certain functions. While this argument could occasionally be true, the larger vision is not necessarily to design systems but rather design collaborative workflows that blend people and AI in a way that achieves more than what people or AI could do on their own. AI systems that enhance human creative agency enable people to achieve better results and outcomes, while maintaining their ownership, strategical intent and sense of identity intact.

Based on the AI-Creative Agency framework, the best-case scenario for an AI system to enhance human creative agency, is to develop **complementing and effective systems** (in either low stakes or high stakes situations). Complementing and effective AI systems don't take over the entire workflow in the Human-AI collaboration loop, and when properly designed with the entire loop in mind, provide contributions that solve problems for people in the workflow or help magnify certain skills and performance.

3.1. Defining the Complementing Functionalities

When aiming to develop a complementing and effective AI system, the starting point needs to be in analyzing the current workflow people use without AI. Workflow analysis is figuring out the step-by-step process a person (or a profession) does when encountering a challenge that requires his creativity and high-level cognitive skills. Each step would usually target a small part of the overall challenge and employ unique skills. As an example, here is the possible rough (and unprecise) workflow for a screenwriter writing a screenplay for a movie:

1. Creating the story premise
2. Developing the story elements (Characters, conflicts, theme, etc.)
3. Creating a rough unifying story arc and structure
4. Developing the outline (plot)
5. Writing a draft
6. Evaluating and revising the draft

Mind that each step could be further analyzed to identify the step's inner workflow. The broader the challenge, the more layers of steps (overall workflow, steps workflows) could be explored.

Analyzing the workflow is a skill that can usually be performed with the help of cognitive psychologists, industrial and management engineers, instructional designers or experienced professionals and subject matter experts [16, 17, 18].

Once there is an established workflow, broken down into sufficient specific functions and skills (i.e., "elements"), the elements can be evaluated by the following dimensions:

- Human effectiveness and added value
- AI effectiveness and added value
- Human creative experience – core or marginal (or even "nagging")? This dimension reflects the level by which people perceive this skill as the core of their creative expression or unique skill

Table 1 demonstrates how this evaluation could apply to the screenwriting workflow we outlined.

Table 1

Screenwriting Workflow – AI-Creative Agency Evaluation

Head 1	Human & AI effectiveness	Human Creative Experience
Creating a story premise	Human: challenging but ultimately high AI: Limited, lacks nuance and originality	Core creative expression
Developing story elements	Human: High at professional levels AI: Moderate or more, can offer brainstorm support	High to moderate (not as high as premise)
Rough story structure	Human: High AI: moderate, can offer brainstorm support	High to moderate
Detailed outline	Human: moderate to high, requires effort and cognitive control AI: low to moderate. Can be effective in organizing plot	Moderate to high – varies between writers
First draft	Human: high, although a time-	Core creative expression

	consuming process AI: moderate to high. Can create a mediocre draft fast	
Rewriting	Human: moderate to high. Considered a tough skill, with difficulty in evaluating once own work. AI: low to moderate, with potential to evaluate drafts effectively and provide revision goals	Moderate – considered as a less desirable activity

Once the workflow is analyzed and its elements evaluated, the developer of an AI system who wishes to create a system that enhances human creative agency, will ideally look for the workflow elements for which AI could offer high effectiveness or an advantage compared to human ability, and for which the human creative experience is low. The values can, of course, be relative in various ways, for instance identifying only parts of the elements for which this evaluation combination exists or choosing a less ideal combination (relatively low creative expression, with relatively high AI effectiveness).

Based on the table, it can be spotted that the entire rewriting phase is relatively low in its creative expression value, and while AI isn't estimated as having the potential to do all of it at a high effectiveness level, it has the potential to provide the screenwriter with an evaluation of the screenplay draft at a relatively fast, accurate and instructive way, while a human screenwriter might be biased in favoring the already existing draft, and would find it difficult to assess how the screenplay can be improved. Following this logic, an AI system that helps a screenwriter quickly and effectively evaluate the draft and derive clear revision goals, would not contest the screenwriter's sense of creative expression and probably even enhance his creative agency, by allowing it to better handle a more challenging and less desirable aspect of his workflow. Such a system would be complementing and effective and therefore lead to enhanced creative agency.

4. Conclusions

This article has argued that creative agency is not merely a psychological by-product of human–AI interaction, but a critical design objective that should guide how AI systems and the interactions with it are envisioned, developed, and evaluated.

By applying the AI–Creative Agency Framework to workflow analysis, developers can identify which elements of a human workflow are most suitable for AI integration—specifically, those where AI is highly effective and where human creative investment is low. This approach enables the development of AI systems that are not only powerful but also complementary, fostering rather than diminishing human creative agency.

This paper contends that the ultimate aim is not to preserve human labor at all costs, nor to surrender it to automation, but to design collaborative workflows in which AI extends human capability—while preserving, and ideally enhancing, the sense of ownership, accountability, and personal growth that defines creative agency. When done right, AI's contribution should propel people to feel more competent, more confident, and more motivated to engage in challenging and meaningful tasks.

Creative agency should be treated as a primary axis of evaluation in AI system design—alongside interpretability, safety, fairness, and user satisfaction. We call on Human-Centered AI practitioners to incorporate creative agency audits into their design and assessment processes, especially in domains where identity, learning, and professional growth are at stake.

As AI capabilities continue to improve, the challenge of sustaining human creative agency will become increasingly complex. Many current AI systems are still perceived as complementing

human skills and thus may contribute positively to creative agency. But as AI becomes more capable across a broader range of tasks, the risk of displacing human agency rather than supporting it will grow. Embedding creative agency as a design consideration from the outset offers a path to ensure that future AI systems support not only high performance, but also foster human engagement, development, and a stronger sense of creative agency—encouraging people to take on tougher challenges with greater confidence and ownership.

Acknowledgements

The author would like to thank Prof. Ryan Watkins from George Washington University for his partnership in the development of this framework.

References

[1] Watkins, Ryan, and Eran Barak-Medina. "AI's Influence on Human Creative Agency." Creativity Research Journal 36.1 (2024): 1–13.
[2] Wei Xu and Zaifeng Gao. "Enabling Human-Centered AI: A Methodological Perspective." IEEE (2024).
[3] Ben Shneiderman. "Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy." arXiv:2002.04087 (2020).
[4] Brooke N. Macnamara, Ibrahim Berber, M. Cenk Çavuşoğlu, Elizabeth A. Krupinski, Naren Nallapareddy, Nadine Nelson, Philip J. Smith, Amy L. Wilson-Delfosse, and Soumya Ray. "Does using artificial intelligence assistance accelerate skill decay and hinder skill development without performers' awareness?" Cognitive Research: Principles and Implications 9.1 (2024).
[5] Hamsa Bastani, Osbert Bastani, Alp Sungu, Haosen Ge, Özge Kabakcı, and Rei Mariman. "Generative AI Can Harm Learning." SSRN 4895486 (2024).
[6] Adriana Salatino, Arthur Prével, Émilie A. Caspar, and Salvatore Lo Bue. "Influence of AI Behavior on Human Moral Decisions, Agency, and Responsibility." Research Square (2024).
[7] Ethan Goh, Robert Gallo, Jason Hom, Elizabeth Strong, Yuchuan Weng, Harlan Kerman, and Jason H. Chen. "Large Language Model Influence on Diagnostic Reasoning: A Randomized Clinical Trial." JAMA Network Open 7.10 (2024): e2440969.
[8] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, and Eric Horvitz. "Guidelines for Human-AI Interaction." Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (2019): 1–13.
[9] Aung Pyae. "The Human-AI Handshake Framework: A Bidirectional Approach to Human-AI Collaboration." arXiv:2502.01493 (2025).
[10] Jeba Rezwana and Mary Lou Maher. "Designing Creative AI Partners with COFI: A Framework for Modeling Interaction in Human-AI Co-Creative Systems." ACM Transactions on Computer-Human Interaction 30.5 (2023): 1–28.
[11] Andi Asrifan, Uswah Mujahidah Rasuna Said, Juvrianto Chrissunday Jakob, and Risman Wanci. "AI Literacy." Advances in Educational Technologies and Instructional Design (IGI Global, 2024): 17–48.
[12] Davy Tsz Kit Ng, Jac Ka Lok Leung, Samuel Kai Wah Chu, and Maggie Shen Qiao. "Conceptualizing AI Literacy: An Exploratory Review." Computers and Education: Artificial Intelligence 2 (2021): 100041.
[13] Thomas KF Chiu, Zubair Ahmad, Murodulla Kahramonovich Ismailov, and Ismaila Temitayo Sanusi. "What Are Artificial Intelligence Literacy and Competency? A Comprehensive Framework to Support Them." Computers and Education Open (2024): 100171.
[14] Miharaini Md Ghani, Wan Azani Mustafa, Durratul Laquesha Shaiful Bakhtiar, and Moh. Khairudin. "A Comprehensive Study: AI Literacy as a Component of Media Literacy." Journal of Advanced Research in Applied Sciences and Engineering Technology 53.2 (2024): 112–121.

[15] Ryan Watkins, Eran Barak-Medina, and Katrina Pugh. "Is AI Hijacking Our Agency?" Center for Open Science, Working Paper No. 8bpaq_v1 (2024).
[16] Colby Tofel-Grehl and David F. Feldon. "Cognitive Task Analysis–Based Training: A Meta-Analysis of Studies." Journal of Cognitive Engineering and Decision Making 7.3 (2013): 293–304.
[17] Richard E. Clark and Fred Estes. "Cognitive Task Analysis." International Journal of Educational Research 25.5 (1996): 403–417.
[18] Kenneth A. Yates and Richard E. Clark. "Cognitive Task Analysis." In *International Handbook of Student Achievement*, Routledge (2012).