The Five Pillars of Enaction as a Theoretical Framework for Co-Creative Artificial Intelligence

Nicholas Davis
Co-Creative AI Consulting
Elryia, OH 44035 USA
nicholas.davis10@.gmail.com

Manoj Deshpande
Expressive Machinery Lab
Georgia Institute of Technology
Atlanta, GA 30308 USA
manojdeshpande@gatech.edu

Jeba Rezwana
Computer & Information Systems
Towson University
Towson, MD 21252 USA
jrezwana@towson.edu

Brian Magerko
Expressive Machinery Lab
Georgia Institute of Technology
Atlanta, GA 30308 USA
magerko@gatech.edu

Abstract

This paper proposes the cognitive science paradigm of enaction as a theoretical framework for co-creative artificial intelligence (AI). Enaction describes how meaning emerges through the interaction of an agent with the environment in a process of sense-making. Enaction is different from information processing (IP) theories of cognition as it does not employ plans, but rather improvised and situated meaning construction processes. This paper argues that enaction can be used as the theoretical basis to design, evaluate, and describe co-creative AI systems. The five pillars of enaction are described: autonomy, sense-making, embodiment, emergence, and experience. Each category is applied to co-creative AI to create a descriptive framework to categorize and systematically describe co-creative AI systems. An analysis is conducted of 20 co-creative AI systems from the literature, including ChatGPT, Stable Diffusion, and Google's Gemini. Enactive design recommendations are provided for each of the enactive categories.

Introduction

Co-creative artificial intelligence (AI) systems that collaborate with users on a shared creative product or performance are becoming more prevalent with the advent of generative AI. The future is clear that co-creative systems will remain an important piece of the AI landscape. Co-creative AI is a subfield of computational creativity (Colton, Charnley, and Pease 2011; Toivonen and Gross 2015) and artificial intelligence. The concept of mixed-initiative co-creative AI, which originated in the game design domain, typically involves a turn-taking approach between humans and AI (Yannakakis, Liapis, and Alexopoulos 2014; Deterding et al. 2017) and share similarities with co-creative AI systems. Human-AI co-creativity researchers design, evaluate, and study systems that collaborate with users on a shared creative product or performance (Davis 2013). Co-creation can be an improvisational activity with contributions influencing both collaborators through time (Davis 2013). Co-creation features many of the same mechanisms as sense-making and participatory sense-making described by the enactive literature (Davis et al. 2016b), such as interaction dynamics (e.g. turn-taking, communication strategies, coordination) and interaction couplings (e.g. mutually influential turns).

The lenses that we have to look at co-creativity and co-creative systems for theorizing and designing are insufficient. There is a theoretical and practical need for a cognitive theory guiding co-creative AI. The dominant paradigm in cognitive science is the information processing (IP) theory of cognition, which views cognitive agents as information processing systems that sense and act based on goal-based planning (Schank and Abelson 1975; Laird 2019). However, co-creation is inherently improvisational, dynamic, and open-ended with meaning emerging from the interactions of the collaborators. The IP theory of cognition needs to be supplemented to describe the dynamics of co-creation. The cognitive science theory of enaction, on the other hand, is particularly well suited to study co-creative AI.

The cognitive science paradigm of enaction views cognition as emerging from the interaction of an agent with the environment and other agents in the environment (De Jaegher and Di Paolo 2007; Velmans 2007; Varela, Thompson, and Rosch 2017). The agent regulates its interaction with the environment to increase its autonomy in a process referred to as *sense-making* (De Jaegher and Di Paolo 2007; Thompson and Stapleton 2009). The agent's autonomy is the self-organizing and self-sustaining principle of living cognitive systems that enables them to survive and thrive in their environment (Varela, Thompson, and Rosch 2017). When multiple agents are engaged in sense-making together, a new domain of relational dynamics emerges through the coordination of interactions in a process referred to as participatory sense-making (Fuchs and De Jaegher 2009). Enactivists use dynamical systems theory to understand these relational dynamics, such as interaction coupling, i.e. the mutually influential turn taking process wherein some of the characteristics of the turn correspond to their partner's previous turn (De Jaegher and Di Paolo 2007). The dynamics of interaction are particularly relevant to the study and design of co-creative systems.

This paper begins by reviewing the co-creative AI literature and the enactive cognitive science literature. The paper strives to ignite a synergy between enaction and co-creative AI similar to the reciprocally beneficial relationship between traditional cognitive science and AI. Next, we develop and describe a framework to interpret and analyze co-creative AI using the five pillars of enaction. Applying the five pillars of enaction to co-creative AI systems can structure the design and study of such systems by mapping conceptual relations between enactive cognitive theory and AI. Then, an analysis is conducted of 20 co-creative systems using the enactive classification framework. A discussion follows that considers how to combine the techniques identified here to evaluate co-creative systems, and a set of enactive design recommendations for co-creative systems is provided. In summary, this paper makes the following contributions:

- This paper introduces a framework for interpreting and analyzing co-creative AI using the five pillars of the cognitive science theory of enaction.
- An analysis of 20 existing co-creative AI systems is presented using the framework.

Background

Co-Creative AI Systems

In co-creative AI systems, humans and AI contribute as collaborators in the creative process (Davis 2013), distinguished from autonomous creative AI systems, which generate creative products independently, and creativity support tools that support human creativity (Kantosalo and Takala 2020; Davis et al. 2015). Kantosalo & Takala provide a descriptive definition of human-computer co-creativity in their Five C framework for co-creation (Kantosalo and Takala 2020):

> "The creative human–computer collective consists of at least one human and one computational collaborator. The collaboration of the collective consists of individual and collaborative creative processes and interactions that support them. The collaboration results in an artefact or a product that represents the contributions of the collective. These contributions are communicated to and shared with a wider community of peers, audiences, and other social influences. The co-creative collaboration takes place in a context representing the environment of the creative act, including e.g. cultural artefacts and conventions, and more immediate factors such as material affordances and shared mental resources, such as the creative task."(Kantosalo and Takala 2020)

This definition takes into account multiple collaborators working together on a shared product in a creative process that is situated in a broader cultural context. The Five C framework is useful in a distributed cognitive analysis of co-creation by analyzing the artifacts and social processes onto which cognition is distributed during the co-creative process. However, the interaction dynamics of co-creation (e.g. turn-taking, communication strategies, feedback, coordination, and interaction modes) are absent from the definition of Kantosalo and Takala (2020) despite being an integral part of co-creation. In a co-creative system, interaction between humans and AI makes the creative process complex and emergent (Rezwana and Maher 2023)) and the creativity that emerges from the co-creation cannot be credited either to the human or to the AI alone (Yannakakis, Liapis, and Alexopoulos 2014). Fantasia et al. propose an embodied approach to collaboration, considering it as an intrinsic part of interaction processes, emphasizing the importance of investigating interaction dynamics, context, environment, and sense-making to enhance knowledge and understanding of collaboration. (Fantasia, De Jaegher, and Fasulo 2014).

Co-creative AI has been segmented into alternating co-creativity (turn-taking on a single task) and task-divided co-creativity (user and agent work on separate creative tasks) (Kantosalo and Toivonen 2016). The user and agent can both take initiatives in a creative process in co-creative AI (Yannakakis, Liapis, and Alexopoulos 2014).

The role a co-creative AI agent can play has been analyzed. Lubart (2005) proposes four modes a co-creative agent can take: nanny, pen-pal, coach, and colleague. A nanny watches over the work to check for mistakes. A pen-pal facilitates communication between collaborators. A coach tries to improve the creativity of the user. A colleague collaborates with the user on a shared product. On the other hand, Kantosalo et al. describe the roles of co-creative AI as generator, evaluator and concept definer (Kantosalo and Toivonen 2016).

Enactivism

Enactivism articulates a cognitive science perspective emphasizing that cognition is not a passive process but arises from active engagement with the environment (Shapiro and Spaulding 2021; Ward, Silverman, and Villalobos 2017; Gallagher and Bower 2013; Di Paolo, Rohde, and De Jaegher 2010; De Jaegher and Di Paolo 2007). This framework counters traditional views that cognition involves internal computations detached from real-world interactions. Instead, enactivism posits that cognitive processes emerge from an organism's sensorimotor activities, interactions, and broader environmental engagements, emphasizing the intertwined nature of perception, action, and cognition. There are broadly three varieties of enactivism: autopoietic enactivism, sensorimotor enactivism, and radical enactivism (Ward, Silverman, and Villalobos 2017; Gallagher and Bower 2013).

Autopoietic enactivism focuses on the self-organizing nature of living systems. Thompson (Thompson 2010) characterizes autopoietic enactivism as the idea that an organism's cognition is fundamentally tied to its ongoing pursuit of maintaining its own operational identity. This means that cognitive processes are continuous with the organism's life processes – there is no clear demarcation between what is mental and what is a life-sustaining activity. From this perspective, cognition is seen as the organism's active regulation of its exchanges with the environment, ensuring its continuity as a viable system.

Sensorimotor enactivism, another strand within the enactive approach, views cognition as arising from an organism's embodied action in the world. Accordingly, sensorimotor enactivism posits that cognitive processes develop through the organism's exploration of the environment, leading to a network of dependencies between sensory inputs, motor outputs, and the ecological context (Shapiro and Spaulding 2021). This exploration is not random but is structured by the organism's sensorimotor contingencies – the rules that govern the coordination between sensory experiences and motor actions. This strand of enactivism captures the essence of "thinking by doing," where cognition results from skillful interaction with the world.

Radical enactivism extends the enactive argument further by rejecting the traditional cognitivist reliance on mental representations. Ward et al. (2017) describe radical enactivism as focusing on the dynamic, adaptive patterns of interaction an organism has with its environment. In this view, cognition does not require the construction of internal models of the world; instead, it is directly embodied in the organism's actions and interactions. Radical enactivism, therefore, seeks to explain cognition without invoking internal mental states, emphasizing the importance of real-world engagements over internal computations.

Participatory Sense-Making (PSM) is a theoretical approach developed by Di Paolo and De Jaegher (2007) to provide a deeper insight into how we understand social cognition. Within the broader context of enactivism, particularly autopoietic enactivism, PSM extends the concept of cognition from an individual-centric to a social-centric framework. It underscores the role of interpersonal interactions in forging shared meanings and collective understandings, positing that the cognition observed during social engagements emerges through the dynamic interplay of these interactions rather than being a simple aggregation of individual cognitive processes. While PSM is primarily rooted in the principles of autopoietic enactivism, it also resonates with sensorimotor enactivism to some extent, particularly in its acknowledgment of cognition as an emergent and interactive phenomenon. The approach accentuates that social understanding is collaboratively constructed, evolving from the mutual and interconnected activities of individuals actively engaged in their social worlds.

While enactivism encompasses a broad spectrum of ideas, PSM specifically hones in on the interactional and social dimensions of cognition, positioning itself as a valuable framework for examining enaction in co-creative systems. This approach synthesizes elements from both autopoietic and sensorimotor enactivism, focusing on dynamic social engagements. Within this context, enactivism is founded upon five critical concepts: *autonomy*, *sense-making*, *embodiment*, *emergence*, and *experience* (De Jaegher and Di Paolo 2007; Di Paolo, Rohde, and De Jaegher 2010; Thompson and Stapleton 2009; Vernon 2010; Varela, Thompson, and Rosch 2017). Autonomy, in the five pillar framework, is defined as the self-organizing principle that enables an agent to maintain its identity through interactions with its environment. Sense-making describes the process by which agents derive meaning and understanding through their engagement with environmental stimuli, mediated by sensorimotor feedback. Embodiment emphasizes the role of the physical body in shaping cognitive processes and how cognitive experiences are inherently grounded in bodily interactions with the world. Emergence in enactivism pertains to how complex meanings and cognitive phenomena spontaneously arise from simpler interactive dynamics. Lastly, experience in this context refers to the cumulative history of an agent's interactions, shaping its ongoing responses to its environment and influencing its future interactions.

Enactive Co-Creative AI

The five pillar framework applies the five core concepts of enaction: *autonomy*, *sense-making*, *embodiment*, *emergence*, and *experience*, to analyze and articulate co-creative AI. Each category offers a conceptual framework with which to describe co-creative AI. Together, they offer a vision of what co-creative AI systems can potentially achieve. Using the five pillars of enaction, a fully enactive co-creative AI system can be defined as follows:

> At least one human and one agent collaborating on a shared creative product where the autonomy of the user and agent is maintained and meaning is built through interaction, coordination, communication, and feedback. The agent and user engage in sense-making (regulating interaction with the environment) and participatory sense-making (regulating a social sense-making process) to understand each other's creative intentions and enact or bring forth meaning in the environment. Both the user and agent are embodied, with interactions constrained and afforded by their bodies. The agent engages in improvised interaction to yield emergent interaction dynamics. The agent remembers its experience, storing the interaction history and utilizing that to inform the creative trajectory of the interaction.

There are degrees of enaction a system exhibits, from highly enactive to minimally enactive. Highly enactive systems would include features from most of the pillars, while minimally enactive systems would have one or a few of the enactive features. In the next sections, each of the enactive pillars is considered with *full*, *partial*, and *none* classifications for describing the presence of the feature in co-creative AI systems.

Autonomy

In the context of co-creative AI, autonomy refers to the ability of the agent to choose its creative action and determine when to contribute in a co-creative session. The ability to make creative choices and decide when to take actions are the defining characteristics of autonomy in the five pillars framework.

Fully autonomous Agent takes initiatives to act and chooses its creative action. Similar to mixed-initiative systems where both collaborators take turns "constraining, suggesting, producing, evaluating, modifying, or selecting creative outputs in response to the other, such that creative agency and initiative cannot be easily ascribed to one side alone." (Deterding et al. 2017). An example of this would be a drawing system that chooses when it wants to contribute as well as what it wants to contribute.

Semi-autonomous Agent does not take initiatives to act, but chooses its creative action. For example, the system may wait for the user to take a turn before taking a turn of its own. This would be a user-initiated semi-autonomous system. Some of the casual creativity literature (Compton and

Mateas 2015) describes systems that would fall into this category. The user can define some parameters and seed values to a system that then generates a creative product based on that specification.

User directed Agent is directed by the user on what to do. In this situation, the agency lies with the user. The user's creative process drives the interaction, and the system behaves in well-defined ways. Creativity support tools (Shneiderman 2007) are an example of this type of interaction where the user directs the system to do.

Embodiment

Embodiment refers to the bodily constraints and affordances offered by the unique capabilities of the agent's body and how the human's cognition and conceptual structures are inherently rooted in the body (Varela, Thompson, and Rosch 2017; Clark 1999). For co-creative AI, this would mean the an AI agent that has bodily constraints, affordances, and an embodied presence of a character to visualize the AI and with which the agent takes actions in the environment (Guckelsberger et al. 2021). Having an embodied presence can enhance the user experiences in co-creative systems, by making the system more relatable and helping to establish rapport between the user and system (Rezwana, Maher, and Davis 2021).

Fully embodied Agent has a body and actions constrained and afforded by that body. For example, in LuminAI (Long et al. 2017), the AI has a body it uses to dance, and the dance moves are constrained by the physical characteristics of the agent's body.

Partially embodied Agent has a (real/virtual) body, but actions are not constrained or afforded by the body. This example would include static icons and avatars representing the AI agent. The actions the AI takes are not animated by the agent.

Non-embodied In this scenario, there is no virtual or physical character representing the AI. The actions the agent takes are disembodied and just appear in the working creative space of the system. The Sentient Sketchbook is a MICI that supports users in generating game level designs. Similar to other MICIs, the Sentient Sketchbook does not have a virtual AI character and is therefore, disembodied.

Emergence

Emergent systems arise from the interaction of simpler systems and processes in a complex interplay. There are two forms of emergence: *strong emergence* and *weak emergence*. Strong emergence occurs when the resulting behavior of the phenomena cannot be deduced from the low level domains that make up the system (Chalmers 2006). Weak emergence, on the other hand, produces unexpected behavior, but it can be deduced from the initial conditions and rule-set of the system (Chalmers 2006). Cellular automata are an example of weak emergence, where they achieve complex and unexpected behavior with the interaction of a few simple rules.

Fully emergent Agent produces contributions that are strongly emergent, highly variable, spontaneous, and unpredictable and cannot be deduced from low-level features of the system (Chalmers 2006). Fully emergent systems are unpredictable, even with knowing how the algorithms of the system work. ChatGPT is an example of a fully emergent system as its responses are varied and unpredictable. It will give a different response to different users given the context of the conversation. The search space of the ChatGPT model is large enough to provide varied responses.

Partially emergent Agent produces weakly emergent contributions that are unexpected, but deductible from the initial conditions and rule-set of the agent. Here, the co-creative agent produces contributions that may seem unpredictable without knowledge of how the system works. However, given the rule-set of the agent, the responses are predictable. For example, in the Drawing Apprentice (Davis et al. 2015), the agent transforms user input (e.g. translating, rotating, scaling) and draws it on the canvas. Knowledge of this rule-set makes the contributions of the agent more predictable.

Non-emergent Agent's interaction is based on scripted knowledge. The agent does not improvise in this circumstance, it recognizes cases that it responds to with associated knowledge. The results of such algorithms will be more predictable than partially emergent and fully emergent approaches.

Sense-Making

Sense-making is the process of exploring the environment through interaction (De Jaegher and Di Paolo 2007; Thompson and Stapleton 2009). Agents engage in sense-making both individually with systems in the world, and together, as social coordination. Social coordination requires feedback and interaction and can lead to interaction couplings, where the structure of one turn is related to the structure of another turn (De Jaegher and Di Paolo 2007).

Full sense-making Agent regulates its interaction with the environment through coordination and feedback. This type of agent would be capable of achieving participatory sense-making or joint sense-making with the user as it would use feedback to coordinate the interaction. The use of feedback (e.g. positive/negative feedback) is a basic requirement for a full sense-making system. This feedback can be bidirectional—either from the user to the agent or vice versa. More nuanced feedback would enrich the communication channel and enable greater coordination.

Partial sense-making The AI agent orients itself to the user's input, but there is no communication channel for feedback and coordination. This type of agent would engage in more reactive sense-making where the agent is reacting to what is happening in the environment (Deshpande et al. 2023).

Non sense-making Agent does not regulate actions based on coordination and feedback from the partner. This would

Table 1: Enactive coding scheme for co-creative systems.

	Full (code = 3)	Partial (code = 2)	None (code = 1)
Autonomy	Agent can take initiative to act and choose their creative action.	Agent does not take initiative to act, but chooses its own creative action.	Agent is directed by the user what to do.
Sense-Making	Agent coordinates interaction with feedback. Agent achieves PSM.	Agent responds to its environment to achieve reactive sense-making.	Agent utilizes no feedback and engages in individual sense-making.
Embodiment	Agent has a (real/virtual) body with actions constrained and afforded by that body.	Agent has a (real/virtual) body (including static icons), but actions are not constrained or afforded by the body.	Agent has no body or representation of a body.
Emergence	Agent produces strongly emergent contributions that are highly variable, spontaneous, and unpredictable that cannot be deduced from low-level features of the system.	Agent produces weakly emergent contributions that are unexpected, but deductible from the initial conditions and ruleset of the agent.	Agent uses scripted interactions and predefined actions to interact.
Experience	Agent learns from interaction history and guides the trajectory of co-creative experiences.	Agent learns from experience to inform creative decision making.	Agent does not use any interaction history to inform its creative decision making.

equate to a system trying to follow a script in the interaction. One difficulty of using a script-based approach in improvisational contexts is the uncertainty, ambiguity, and flexibility of responses that can occur. The script may not be adequate to account for all the different variations a co-creative experience could take. In this type of system, meaning would not emerge through interaction, but rather could be pre-programmed in a large knowledge base.

Experience

The experience category represents the interaction history (and its use in decision making) between the cognitive agent and the environment, including other agents within that environment (Vernon 2010). In human interaction, the interaction history informs the trajectory of the interaction and serves as the groundwork upon which further interactions are based (De Jaegher and Di Paolo 2007). Co-creative systems can record the interaction history, analyze it, model it in some manner, and utilize it to inform creative interactions in the moment.

Fully experiential Agent learns from all aspects of previous interactions and applies that knowledge in the moment to inform its creative decision making. A fully experiential agent could learn from its experiences to inform its interactions in the moment. This type of agent would recognize the trajectory or arc of the creative experience and interact according to this trajectory.

Partially experiential Agent retains some knowledge from previous experience and applies that knowledge in the moment to inform creative decision making. In this scenario, the agent retains a subset of experiential data and stores it to serve as knowledge for the future. However, the agent is not aware of the creative trajectory or arc of the creative experience.

Non-experiential Agent does not use any interaction history knowledge to inform its creative decision making. This type of co-creative agent would rely on the interaction dynamics in the moment to inform its creative decisions. The agent would not record actions or use them to determine future actions. The system could respond to the last action alone.

Enactive Co-Creative AI Analysis

Convenience sampling was used to select systems from the literature seeded by systems the authors were familiar with. To perform the analysis, 3 researchers (the authors) independently coded the 20 co-creative systems based on the systems' descriptions in their respective publications, then convened with a fourth researcher to discuss discrepancies using the consensus assessment technique (CAT) (Hennessey, Amabile, and Mueller 1999). The coding rubric was analyzed to see where discrepancies were occurring and updated, and the systems were re-coded. Then, disagreements were discussed and updated to reflect the group consensus.

Highly Enactive Systems

Systems are considered highly enactive if their total score, calculated as the sum of the scores across all five pillars, is at least 12 out of a maximum possible score of 15. Highly enactive systems are either semi-autonomous or fully autonomous. They feature full sense-making processes with feedback and coordination in the moment. These systems have full or partial AI embodiment that is used during the co-creation. The actions that these AI agents perform are typically fully emergent, and they cannot be reduced to the

Table 2: Analysis of 20 co-creative systems. Dark blue = Fully present = 3; Blue = Partially present = 2; Light blue = Not present = 1.

Co-Creative System	**Autonomy**	**Sense-Making**	**Embodiment**	**Emergence**	**Experience**	**#**
Shimon (Hoffman and Weinberg 2010)						14
Github Copilot (Github 2024)						12
Google Gemini (Reid et al. 2024)						12
Cobbie (Lin et al. 2020)						12
ChatGPT (Achiam et al. 2023)						12
Drawing Apprentice (Davis et al. 2015)						11
LuminAI (Long et al. 2017)						11
Creative PenPal (Rezwana, Maher, and Davis 2021)						11
Story Drawer (Zhang et al. 2021)						11
CreativeConnect (Choi et al. 2023)						10
Reframer (Lawton et al. 2023)						9
Drawcto (Deshpande and Magerko 2021)						9
DuetDraw (Oh et al. 2018)						9
Stable Diffusion (Podell et al. 2023)						8
Creative Sketching Partner (Karimi et al. 2020)						8
CharacterChat (Schmitt and Buschek 2021)						7
ALYSIA (Cheatley et al. 2020)						7
Poetry Machine (Kantosalo and Riihiaho 2019)						7
Stable Walks (Rost and Andreasson 2023)						7
FashionQ (Jeon et al. 2021)						7
Average Code	1.75	2.35	1.65	2.4	1.55	
Code Count	35	47	33	48	31	

initial conditions and rule-set of the agent. Highly enactive agents include the experience of the interaction as a part of their knowledge set, adding a layer of context to their contributions.

Our analysis identifies Shimon, Google Gemini, ChatGPT, Github Copilot, and Cobbie, as highly enactive systems within the dataset. All these generative AI systems are semi-autonomous, with the exception of Shimon, which is fully autonomous and can define its own creative objectives. All these AI systems can coordinate their interaction with feedback, such as Shimon using embodied gestures as visual cues to communicate turn-taking and musical beats. Among these systems, only Shimon has a physical body, while others use icons for minimal virtual representation of the AI. All of these systems generate emergent contributions, such as ChatGPT produces highly emergent and variable contributions when using it as a creative storyteller. These systems keep interaction histories that shape creative decision making.

Moderately Enactive Systems

Moderately enactive systems, with a total score ranging from 11 to 9, typically exhibit semi-autonomous characteristics. They feature a mix of full sense-making (with feedback, coordination, and participatory sense-making) and partial sense-making processes, where the system reacts and adapts to the user's contribution. AI embodiment in moderately enactive systems ranges from fully embodied to no embodiment. The contributions of moderately enactive co-creative agents are a mix of strong emergence and weak emergence. Moderately enactive systems are typically partially experiential or non-experiential.

Within our dataset, we found 8 co-creative systems as moderately enactive. Most of these systems, such as Cobbie, are semi-autonomous, which can not take initiative but can define its creative objective. These systems have either full or partial sensemaking; for instance, Drawing Apprentice can coordinate contributions based on feedback, whereas LuminAI only reacts and adapts to user input. We see a lot of variability in terms of AI embodiment, with Cobbie boasting a physical body, Drawing Apprentice having a virtual one, and ALYSIA lacking embodiment. We see both strong emergence, like Reframer generating highly emergent drawings, and weak emergence, like Creative Penpal generating somewhat emergent inspirational sketches depending on the visual and structural similarity of users' contributions. LuminAI learns from user contribution without guiding the creative trajectory while ALYSIA does not learn from previous interactions.

Minimally Enactive Systems

Systems that have a total score below 9 reclassified as minimally enactive systems. Minimally enactive systems are typically either semi-autonomous or user-directed. They feature partial sense-making processes, where the system is reacting to the user's contribution but does not explicitly use feedback to coordinate its interaction. Minimally enactive systems typically do not use AI embodiment. The contributions of minimally enactive systems typically feature weak emergence, where they are unpredictable without knowledge of how the system works. Minimally enactive systems have a mix of non-experience and partial experience-based algorithms.

In our dataset, we identified 7 minimally enactive systems. For instance, Drawcto exhibits partial autonomy by defining creative objectives but lacks initiative, while systems like Stable Diffusion are entirely user-directed. Many of these systems, such as Creative Sketching Partner, employ partial sense-making by reacting and adapting to user input and do not use AI embodiment. Contributions from systems like DuetDraw exhibit some level of emergence, although they remain deducible by initial conditions and rules. Notably, systems like Poetry Machine do not learn from user interactions.

Discussion

Enaction offers a theoretical framework relevant for co-creative AI. Enaction describes cognition as a sense-making process of interacting with the environment to maintain autonomy. "Meaning" is an emergent property of that interaction, and it is dynamic and improvisational. The five pillars of enactive cognition provide a way to systematically dissect co-creative AI systems to analyze and compare their interaction and cognition. Each pillar provides a lens through which to view co-creative AI. When viewed in tandem, it is possible to envision a new type of enactive co-creative agent that has full autonomy, sense-making, embodiment, emergence, and experience.

Our analysis revealed that Shimon was a prime example of an enactive system with its autonomous improvisation and internally generated goals. Shimon also has a sense-making process of coordinating musical interaction with the user through head bob and gaze. The agent is fully embodied in a robot with a perceptual process rooted in that body. The system could become more enactive if it remembered previous parts of the musical interaction and incorporated them into the current context.

The analysis demonstrates that many of the co-creative systems are user-directed (7 systems), without any form of autonomy. This shows that the co-creative systems surveyed largely do not take the initiative in a creative process and contribute the content the user determines. Many of the co-creative systems analyzed were also disembodied (8/15 systems). They had no virtual character or presence in the interface other than their creative output. Additionally, animating the co-creative agent's response, such as done in ChatGPT and Gemini, can improve the interaction design and make the system more engaging.

From the enactive analysis, the average rating for *experience* is the lowest at 1.55. This demonstrates a significant opportunity for co-creative AI research to investigate systems that utilize their experience to inform their creative decision making. The average rating for *autonomy* and *embodiment* are similarly low at 1.75 and 1.65, respectively, which demonstrates the majority of the AI systems surveyed have limited embodied presence and are semi-autonomous or user directed. *Sense-making* and *emergence* were higher at 2.35 and 2.4, respectively. The generative AI systems surveyed generally have emergent properties, especially LLMs.

The *sense-making* category has all systems at least at partial sense-making except for one (Stable Walks), meaning they are analyzing the user's input and making sense of it in some way so as to act upon it. The *emergence* category is all partially emergent. All of the systems are demonstrating some unpredictability, with some of them achieving a coded value of 3, indicating that the user could not deduce the contribution based on the initial rule set of the agent.

Enactive Design Recommendations

The five pillar framework was synthesized with findings from the enactive co-creative AI analysis to yield a set of design recommendations. These recommendations emphasize how to include enactive features in a co-creative system.

Practitioners can choose to apply all or some of the principles to their work in co-creative AI depending on the context of the application and use case.

Autonomy Based on the purpose of the co-creative AI, autonomy could be included to make the co-creation more spontaneous. Long et al. suggest that AI agents should have creative autonomy (Long, Jacob, and Magerko 2019) and found that the level of AI autonomy that is appropriate is heavily dependent on context. They suggest designing the AI so that the degree of creative autonomy can be modulated according to the context of the co-creation. However, control between the AI and the humans should be considered when deciding on the degree of autonomy of a co-creative AI. Existing literature shows that human collaborators tend to want more control over the behavior and actions of co-creative AI (Rubidge 2002; Davis et al. 2016a). Research shows that users want to lead in co-creation while AI follows their lead (Oh et al. 2018; Winston and Magerko 2017). Future research is necessary to understand the balance between AI autonomy and control between humans and AI.

Sense-making Include explicit feedback mechanisms to help the co-creative system coordinate its interactions with the user. Several systems used binary voting mechanisms (e.g. Gemini, Drawing Apprentice, Cobbie, Creative PenPal) to inform the system whether the user liked its contribution. Shimon uses embodied feedback mechanisms, such as bobbing its head to the rhythm and directing its gaze at the person who is supposed to play next. The feedback mechanism could be more fine-grained than binary, such as informing the system what it was about the contribution they did not like (e.g. content, location, timing). Other types of communication could be present as well, such as both the user and system explaining their actions, providing instructions, and ideating content. This human-to-AI and AI-to-human communication (Rezwana and Maher 2023) relates to the recent push for explainable computational creativity (Llano et al. 2022).

Embodiment The creation of a virtual or physical character that embodies the AI and animates its actions may help the user relate to the agent and improve the perception of the AI system. Several co-creative AI systems had an AI virtual character, and some of these systems animated the actions of the system. This type of embodied creativity is more engaging in the interaction than static creative contributions that just appear. For example, ChatGPT has a black circle that animates the text it is typing as it types it, whereas Google's Gemini animates the response line by line, making it less embodied. Shimon and Cobbie have robot bodies within which their perception is rooted, making them fully embodied agents. The inclusion of a virtual/physical character or avatar improves the presence of the agent as a collaborative partner (Rezwana, Maher, and Davis 2021), improves user trust in the agent (Rezwana, Maher, and Davis 2021), and improves user confidence of an agent performing a task (Kim et al. 2018a). The literature asserts that embodied communication aids synchronization and coordination in improvisational human-computer co-creativity (Hoffman and Weinberg 2011). Users' confidence in an AI agent's ability to perform tasks is improved when imbuing the agent with embodiment compared to the agent solely depending on conversation (Kim et al. 2018b).

Emergence Consider the size of the co-creative agent's search space and the complexity of the rules governing its behavior. More emergent systems like ChatGPT and Google's Gemini are difficult to predict and cannot be reduced down to a rule set. More emergent systems may be able to sustain the user's creative engagement for longer periods due to the unpredictability and spontaneity of the system.

Experience Record and utilize the interaction history of the co-creative session to inform the agent's creative decision making. The interaction history can provide context and inform the trajectory of the interaction. Including interaction history as part of the design of a co-creative agent can make it more contextually aware and engaging. Several of the LLM dialog systems include interaction history, making them fully experiential. Other co-creative systems record the user's actions and feedback and utilize them in creative decision making (e.g. Drawing Apprentice, Cobbie, Creative PenPal). Consider adding a Creative Trajectory Monitor (Davis et al. 2014) to the co-creative system that models user interaction dynamics and feedback to determine how to respond in the moment.

Conclusions

This paper presented the five pillars of enaction as a potential theoretical framework for co-creative AI, which can be used in the design, articulation, and evaluation of co-creative AI systems. Enaction was described as a cognitive theory that emphasizes the role that interaction and coordination play in perception and meaning formation. The core pillars of enaction were described and applied to the context of co-creative AI by segmenting the categories into *fully present*, *partially present*, and *not present* in a co-creative system. 20 co-creative systems were analyzed using this enactive classification framework. Several highly enactive systems feature feedback (e.g. binary voting) to enhance the communication channel between the user and agent. All the systems were at least partially emergent, signifying they were not scripted interactions. Several of the systems using LLMs received a rating of fully emergent due to the size of the agent's search space and unpredictability of its responses. The robotic marimba player Shimon was found to be the most enactive of the systems reviewed given its full embodiment and autonomous nature.

References

Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.

Chalmers, D. J. 2006. Strong and weak emergence. *The re-emergence of emergence* 675:244–256.

Cheatley, L.; Ackerman, M.; Pease, A.; and Moncur, W. 2020. Co-creative songwriting for bereavement support. In *Eleventh International Conference on Computational Creativity: ICCC'20*, 33–41. Association for Computational Creativity.

Choi, D.; Hong, S.; Park, J.; Chung, J. J. Y.; and Kim, J. 2023. CreativeConnect: Supporting Reference Recombination for Graphic Design Ideation with Generative AI. arXiv:2312.11949 [cs].

Clark, A. 1999. An embodied cognitive science? *Trends in cognitive sciences* 3(9):345–351. Publisher: Elsevier.

Colton, S.; Charnley, J. W.; and Pease, A. 2011. Computational Creativity Theory: The FACE and IDEA Descriptive Models. In *ICCC*, 90–95. Mexico City.

Compton, K., and Mateas, M. 2015. Casual Creators. In *ICCC*, 228–235.

Davis, N. M.; Popova, Y.; Sysoev, I.; Hsiao, C.-P.; Zhang, D.; and Magerko, B. 2014. Building Artistic Computer Colleagues with an Enactive Model of Creativity. In *ICCC*, 38–45.

Davis, N.; Hsiao, C.-P.; Singh, K. Y.; Li, L.; Moningi, S.; and Magerko, B. 2015. Drawing Apprentice: An Enactive Co-Creative Agent for Artistic Collaboration. In *Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, 185–186. Glasgow United Kingdom: ACM.

Davis, N.; Hsiao, C.-P.; Singh, K. Y.; and Magerko, B. 2016a. Co-creative drawing agent with object recognition. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 12, 9–15.

Davis, N.; Hsiao, C.-P.; Yashraj Singh, K.; Li, L.; and Magerko, B. 2016b. Empirically Studying Participatory Sense-Making in Abstract Drawing with a Co-Creative Cognitive Agent. In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, 196–207. Sonoma California USA: ACM.

Davis, N. 2013. Human-computer co-creativity: Blending human and computational creativity. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 9, 9–12. Issue: 6.

De Jaegher, H., and Di Paolo, E. 2007. Participatory sense-making: An enactive approach to social cognition. *Phenomenology and the cognitive sciences* 6:485–507. Publisher: Springer.

Deshpande, M., and Magerko, B. 2021. Drawcto: A Multi-Agent Co-Creative AI for Collaborative Non-Representational Art. In Harrison, B., ed., *Joint Proceedings of the AIIDE 2021 Workshops*, volume 3217 of *CEUR Workshop Proceedings*. Virtual Event, Lexington: CEUR. ISSN: 1613-0073.

Deshpande, M.; Trajkova, M.; Knowlton, A.; and Magerko, B. 2023. Observable Creative Sense-Making (OCSM): A Method For Quantifying Improvisational Co-Creative Interaction. In *Creativity and Cognition*, 103–115. Virtual Event USA: ACM.

Deterding, S.; Hook, J.; Fiebrink, R.; Gillies, M.; Gow, J.; Akten, M.; Smith, G.; Liapis, A.; and Compton, K. 2017. Mixed-Initiative Creative Interfaces. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, 628–635. Denver Colorado USA: ACM.

Di Paolo, E.; Rohde, M.; and De Jaegher, H. 2010. Horizons for the enactive mind: Values, social interaction, and play. In *Enaction: Towards a new paradigm for cognitive science*.

Fantasia, V.; De Jaegher, H.; and Fasulo, A. 2014. We can work it out: an enactive look at cooperation. *Frontiers in psychology* 5:874.

Fuchs, T., and De Jaegher, H. 2009. Enactive intersubjectivity: Participatory sense-making and mutual incorporation. *Phenomenology and the Cognitive Sciences* 8(4):465–486.

Gallagher, S., and Bower, M. 2013. Making enactivism even more embodied.

Github. 2024. GitHub Copilot · Your AI pair programmer.

Guckelsberger, C.; Kantosalo, A.; Negrete-Yankelevich, S.; and Takala, T. 2021. Embodiment and computational creativity. *arXiv preprint arXiv:2107.00949*.

Hennessey, B. A.; Amabile, T. M.; and Mueller, J. S. 1999. Consensual assessment. *Encyclopedia of creativity* 1:347–359.

Hoffman, G., and Weinberg, G. 2010. Shimon: an interactive improvisational robotic marimba player. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems*, 3097–3102. Atlanta Georgia USA: ACM.

Hoffman, G., and Weinberg, G. 2011. Interactive improvisation with a robotic marimba player. *Autonomous Robots* 31(2-3):133–153.

Jeon, Y.; Jin, S.; Shih, P. C.; and Han, K. 2021. FashionQ: An AI-Driven Creativity Support Tool for Facilitating Ideation in Fashion Design. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–18. Yokohama Japan: ACM.

Kantosalo, A., and Riihiaho, S. 2019. Quantifying co-creative writing experiences. *Digital Creativity* 30(1):23–38.

Kantosalo, A., and Takala, T. 2020. Five C's for Human-Computer Co-Creativity-An Update on Classical Creativity Perspectives. In *ICCC*, 17–24.

Kantosalo, A., and Toivonen, H. 2016. Modes for creative human-computer collaboration: Alternating and task-divided co-creativity. In *Proceedings of the seventh international conference on computational creativity*, 77–84.

Karimi, P.; Rezwana, J.; Siddiqui, S.; Maher, M. L.; and Dehbozorgi, N. 2020. Creative sketching partner: an analysis of human-AI co-creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 221–230. Cagliari Italy: ACM.

Kim, K.; Boelling, L.; Haesler, S.; Bailenson, J.; Bruder, G.; and Welch, G. F. 2018a. Does a digital assistant need a body? The influence of visual embodiment and social behavior on the perception of intelligent virtual agents in AR.

In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 105–114. IEEE.

Kim, K.; Boelling, L.; Haesler, S.; Bailenson, J.; Bruder, G.; and Welch, G. F. 2018b. Does a digital assistant need a body? the influence of visual embodiment and social behavior on the perception of intelligent virtual agents in ar. In *2018 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 105–114. IEEE.

Laird, J. E. 2019. *The Soar cognitive architecture*. MIT press.

Lawton, T.; Ibarrola, F. J.; Ventura, D.; and Grace, K. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*, 264–277. Sydney NSW Australia: ACM.

Lin, Y.; Guo, J.; Chen, Y.; Yao, C.; and Ying, F. 2020. It Is Your Turn: Collaborative Ideation With a Co-Creative Robot through Sketch. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–14. Honolulu HI USA: ACM.

Llano, M. T.; d'Inverno, M.; Yee-King, M.; McCormack, J.; Ilsar, A.; Pease, A.; and Colton, S. 2022. Explainable Computational Creativity. arXiv:2205.05682 [cs].

Long, D.; Jacob, M.; Davis, N.; and Magerko, B. 2017. Designing for Socially Interactive Systems. In *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, 39–50. Singapore Singapore: ACM.

Long, D.; Jacob, M.; and Magerko, B. 2019. Designing co-creative ai for public spaces. In *Proceedings of the 2019 on Creativity and Cognition*. 271–284.

Lubart, T. 2005. How can computers be partners in the creative process: classification and commentary on the special issue. *International journal of human-computer studies* 63(4-5):365–369. Publisher: Elsevier.

Oh, C.; Song, J.; Choi, J.; Kim, S.; Lee, S.; and Suh, B. 2018. I Lead, You Help but Only with Enough Details: Understanding User Experience of Co-Creation with Artificial Intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–13. Montreal QC Canada: ACM.

Podell, D.; English, Z.; Lacey, K.; Blattmann, A.; Dockhorn, T.; Müller, J.; Penna, J.; and Rombach, R. 2023. Sdxl: Improving latent diffusion models for high-resolution image synthesis. *arXiv preprint arXiv:2307.01952*.

Reid, M.; Savinov, N.; Teplyashin, D.; Lepikhin, D.; et al. 2024. Gemini 1.5: Unlocking multimodal understanding across millions of tokens of context. *arXiv preprint arXiv:2403.05530*.

Rezwana, J., and Maher, M. L. 2023. 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):1–28.

Rezwana, J.; Maher, M. L.; and Davis, N. 2021. Creative PenPal: A Virtual Embodied Conversational AI Agent to Improve User Engagement and Collaborative Experience in Human-AI Co-Creative Design Ideation. In *IUI Workshops*.

Rost, M., and Andreasson, S. 2023. Stable Walk: An interactive environment for exploring Stable Diffusion outputs.

Rubidge, S. 2002. Digital technology in choreography: issues and implications. In *Published in the Proceedings of the 17th Annual Symposium of the Dance Society of Korea, Seoul, Korea*.

Schank, R. C., and Abelson, R. P. 1975. Scripts, plans, and knowledge. In *IJCAI*, volume 75, 151–157.

Schmitt, O., and Buschek, D. 2021. CharacterChat: Supporting the Creation of Fictional Characters through Conversation and Progressive Manifestation with a Chatbot. In *Creativity and Cognition*, 1–10. Virtual Event Italy: ACM.

Shapiro, L., and Spaulding, S. 2021. Embodied Cognition. In Zalta, E. N., ed., *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2021 edition.

Shneiderman, B. 2007. Creativity support tools: accelerating discovery and innovation. *Communications of the ACM* 50(12):20–32.

Thompson, E., and Stapleton, M. 2009. Making Sense of Sense-Making: Reflections on Enactive and Extended Mind Theories. *Topoi* 28(1):23–30.

Thompson, E. 2010. *Mind in life: Biology, phenomenology, and the sciences of mind*. Harvard University Press.

Toivonen, H., and Gross, O. 2015. Data mining and machine learning in computational creativity. *WIREs Data Mining and Knowledge Discovery* 5(6):265–275.

Varela, F. J.; Thompson, E.; and Rosch, E. 2017. *The embodied mind, revised edition: Cognitive science and human experience*. MIT press.

Velmans, M. 2007. Where experiences are: Dualist, physicalist, enactive and reflexive accounts of phenomenal consciousness. *Phenomenology and the Cognitive Sciences* 6(4):547–563.

Vernon, D. 2010. Enaction as a Conceptual Framework for Developmental Cognitive Robotics. *Paladyn, Journal of Behavioral Robotics* 1(2):89–98.

Ward, D.; Silverman, D.; and Villalobos, M. 2017. Introduction: The varieties of enactivism. *Topoi* 36(3):365–375. Publisher: Springer.

Winston, L., and Magerko, B. 2017. Turn-taking with improvisational co-creative agents. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 13, 129–135.

Yannakakis, G. N.; Liapis, A.; and Alexopoulos, C. 2014. Mixed-initiative co-creativity. Publisher: Foundations of Digital Games.

Zhang, C.; Yao, C.; Liu, J.; Zhou, Z.; Zhang, W.; Liu, L.; Ying, F.; Zhao, Y.; and Wang, G. 2021. StoryDrawer: A Co-Creative Agent Supporting Children's Storytelling through Collaborative Drawing. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–6. Yokohama Japan: ACM.