A Research Plan for Integrating Generative and Cognitive AI for Human Centered, Explainable Co-Creative AI

MARY LOU MAHER, University of North Carolina at Charlotte, USA
BRIAN MAGERKO, Georgia Tech, USA
DAN VENTURA, Brigham Young University, USA
DOUGLAS FISHER, Vanderbilt University, USA
ROGELIO E. CARDONA-RIVERA, University of Utah, USA
NANCY FULDA, Brigham Young University, USA
JOHN GERO, University of North Carolina at Charlotte, USA
MINWOO LEE, University of North Carolina at Charlotte, USA
DAVID WILSON, University of North Carolina at Charlotte, USA
JAMES C. KAUFMAN, University of Connecticut, USA

While both cognitive and computational models of creativity continue to receive a great deal of attention, we propose a synergy between the two approaches to creativity and a resulting co-creativity model that leverages strengths of both. Additionally, we propose three foundational use-cases for driving research in co-creativity: improvisation, design, and narrative.

CCS Concepts: • **Human-centered computing → Human computer interaction (HCI)**; **Collaborative and social computing**; • **Computing methodologies → Artificial intelligence**.

Additional Key Words and Phrases: cognitive models, computational creativity, co-creativity, improvisation, design, narrative

ACM Reference Format:
Mary Lou Maher, Brian Magerko, Dan Ventura, Douglas Fisher, Rogelio E. Cardona-Rivera, Nancy Fulda, John Gero, Minwoo Lee, David Wilson, and James C. Kaufman. 2022. A Research Plan for Integrating Generative and Cognitive AI for Human Centered, Explainable Co-Creative AI . In *CHI Workshop on Generative AI and HCI, May 10, 2022.* ACM, New York, NY, USA, 4 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Early approaches to computational creativity relied on symbolic models of creative cognition that automated the production of novel solutions in limited domains. More recent approaches rely on data-driven statistical and deep learning approaches that produce novel outputs, yet have little connection to a human-centered world. These two approaches can be mapped to System 1 and System 2 thinking [21]. Deep learning approaches to the generation of creative artifacts can be considered as System 1 thinking that is intuitive and not easily explained. Cognitive models of creativity can be considered as System 2 thinking that is deliberate and can be described and explained. We propose a path to human centered, explainable co-creative AI through the synthesis of neural/data-driven/generative approaches with cognitive models, with the potential to achieve foundational tools for human-centered co-creative AI systems [22].

Both socio-cognitive processes, such as sensemaking and shared mental model construction, as well as individual cognitive processes, such as analogical reasoning and emergence, can act as the foundations for a co-creative AI that will be able to make sense of the world with us and help us be more innovative. We suggest that a deeper understanding of both neural and cognitive models of creativity [10] will advance AI models of creativity to support synergistic situational human-machine partnerships, eliciting new solutions and experiences that are novel, surprising, and valuable.

1.1 Limitations of current approaches

The development of neurocognitive approaches to explainable co-creative AI addresses the following limitations in existing computational models of creativity:

- Research in cognitive models of creativity, such as idea generation and divergent-convergent thinking, are typically studied as components of human innovation [30] rather than as foundational models in human cognition. While there is a large literature on the human use of AI-based entities, that research has addressed co-execution of known tasks rather than co-creativity for open-ended tasks.
- Neuro-cognitive studies of human creativity have been limited to reductive problems (e.g., AUT, RAT) rather than rich problems embedded in natural settings [11].
- Current deep learning models require massive amounts of data and computation, are statistical or probabilistic in nature, do not admit divergent solutions, and lack explainability [15].
- Computational creativity research until recently has focused on developing individual AI abilities rather than augmenting human creativity [6].
- Embodied agents pose a special challenge for co-creative AI where actions taken in the environment have immediate functional and expressive consequences; moreover, these systems have not scaled in capacity the way computers (and super-computers) have [23].

We propose the study of situated, explainable, co-creative AI with interactive models of intuitive (System 1) and deliberative (System 2) thinking. For example, given a “black box" neural model of creativity, is it possible to construct a cognitive process that explains the creative outcome produced by the neural model [5]? This process can be understood as a communicative act–or a series of communicative acts with successive refinements [9]—and thereby the nature of each entity (human and AI) is important. Recognition of these dynamics will enable either party to ask “why" questions and to re-direct or even re-frame the partner’s creative direction [24]. Thus, modeling a creative process entails modeling the creative partner within that process. Cooperative AI systems that learn causal explanations for the behavior of learned AI systems [27] can provide new means for studying these creative architectures. Active learning methods for querying human judgments of similarity [8] or for eliciting their mental models or hypotheses about a problem (Markant, 2018) can be used to better map creative thought in individuals, providing new means for AI systems to become effective partners to humans. Implicitly elicited human guidance [36] can further explain the creative interactions. We expect such research will lead to a Mutual Theory of Mind between humans and AI [35].

2 A WAY FORWARD

We describe 3 use-inspired research areas to guide the development and evaluation of explainable co-creative AI: improvisation, design, and narrative.

2.1 Improvisation as in-the-moment creativity

Improvisation is a fundamental technique we use to express ourselves, experiment, connect, and make sense of the world around us [4]. This concept is well-understood when we work with other humans [19], but it has also been theorized when humans interact with non-human entities. When working with media or artifacts, designers are said to conduct a "reflective conversation with the materials of a design situation" [33], and a part of this reflective conversation is improvisation as those materials reveal their qualities [17]. This idea has been taken up in studies of human centered data science, where data work [26] involves intimate interactions with data as a kind of "material" [2] that can require a certain type of learned or prepared way-of-knowing or way-of-seeing [29]. However, improvising in our interactions with technology is not integrated into their design. Human-computer interactions tend to be scripted, predictable, and driven by the user or driven by an inflexible scripted workflow or set of policy-driven requirements on the user [31]. AI systems have been developed to partner with humans during improvisation in creative domains such as dance, music, play, conversation, and dialogic performance [14]. We will focus on improvisation contexts to develop and evaluate our advances in foundational AI models of co-creativity.

2.2 Design as structured creativity

Design is the foundation of all human-made physical and virtual objects and processes, generating new products and processes, in part, through mechanisms of abductive reasoning [18]. Design involves the generation of new spaces of possibilities. In this it uses processes that run parallel to those used in creativity (Gero, 2000). Creative design leads to innovation and increased competitiveness [32]. AI systems have been developed to support design through cognitive and neural models of analogy, transformation, and emergence - but not, as yet, with models of abduction or with models based on neurocognition findings. Recent findings have shown that not only neuronal activations but also inhibitions are involved in creativity (Carmada, et. al., 2019). We will move the neurocognition research from reductionist creative tasks to open creative tasks [11]. We will focus on design as providing ecologically valid contexts as we develop empirically-based neurocognitive models of creativity that expand the space of possibilities as one basis for AI and AI co-creativity [34].

2.3 Narrative intelligence as creative composition

Narrative intelligence is the ability to generate natural language descriptions of knowledge and scenarios with causal and temporal relevance. Narrative generation and storytelling are creative processes that humans use to communicate feelings, information, and intention [20]. They can augment human understanding of complex ideas and situations, provide explanation for action, and encourage trust and cooperation [16]. The importance of computational narrative intelligence has been acknowledged for decades [13], and it has been repeatedly demonstrated that narrative intelligence in AI systems improves human adoption [25], trust [28], and cooperation [12]. We propose to study models of narrative intelligence to extend the usefulness of natural language and word embedding models in human co-creative and computational creativity settings. Examples of use-cases include dialogic and other types of co-creative systems [7], extending the interestingness of data-driven storytelling [1], and developing AI intermediaries to promote productive discourse across community divides [3].

REFERENCES

[1] A. Al-Doulat, N. Nur, A. Karduni, A. Benedict, E. Al-Hossami, M.L. Maher, W. Dou, M. Dorodchi, and X. Niu. 2020. Making Sense of Student Success and Risk through Unsupervised Machine Learning and Interactive Storytelling. In *Proc. AI in Education*. 3–15.

[2] C. Aragon, S. Guha, M. Kogan, M. Muller, and G. Neff. 2022. *Human-Centered Data Science: An Introduction*. MIT Press, Cambridge, MA.

[3] L.P. Argyle, E.C. Busby, N. Fulda, J. Gubler, C. Rytting, and D. Wingate. 2022. Out of One, Many: Using Language Models to Simulate Human Samples. *Political Analysis* (2022), under review.

[4] B., W. Manzoul, M. Riedl, A. Baumer, D. Fuller, K. Luther, and C. Pearce. 2009. An Empirical Study of Cognition and Theatrical Improvisation. In *Proceedings of the Seventh ACM Conference on Creativity and Cognition*. 117–126.

[5] P. Bodily and D. Ventura. 2018. Explainability: An Aesthetic for Aesthetics in Computational Creative Systems. In *Proc. ICCC*. 153–160.

[6] P. Bodily and D. Ventura. 2020. What Happens When a Computer Joins the Group?. In *Proc. ICCC*. 41–48.

[7] O. Bown, K. Grace, L. Bray, and D. Ventura. 2020. Speculative Exploration of the Role of Dialogic AI in Co-creation. In *Proc. ICCC*. 25–32.

[8] G. Canal, S. Fenu, and C. Rozell. 2020. Active Ordinal Querying for Tuplewise Similarity Learning. In *Proc. AAAI*. 3332–3340.

[9] M. Chromik and A. Butz. 2021. Human-XAI Interaction: A Review and Design Principles for Explanation User Interfaces. In *Proc. IFIP HCI*. 619–640.

[10] E. G. Chrysikou. 2019. Creativity In and Out of (Cognitive) Control. *Current Opinion in Behavioral Sciences* 27 (2019), 94–99.

[11] E. G. Chrysikou and J.S. Gero. 2020. Using Neuroscience Techniques to Understand and Improve Design Cognition. *AIMS Neuroscience* 7, 3 (2020), 319–326.

[12] J.W. Crandall, M. Oudah, Tennom, F. Ishowo-Oloko, S. Abdallah, J.-F. Bonnefon, M. Cebrian, A. Shariff, M. A. Goodrich, and I. Rahwan. 2018. Cooperating with Machines. *Nature Communications* 9 (2018), Article No. 233.

[13] K. Dautenhaun and S. J. Coles. 2001. Narrative Intelligence from the Bottom Up: A Computational Framework for the Study of Story-telling in Autonomous Agents. *Journal of Artificial Societies and Social Simulation* 4, 1 (2001).

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

[15] F. Doshi-Velez and B. Kim. 2017. Towards a Rigorous Science of Interpretable Machine Learning. arXiv:1702.08608.

[16] S.M. Fiore, R. McDaniel, and F. Jentsch. 2009. Narrative-Based Collaboration Systems for Distributed Teams: Nine Research Questions for Information Managers. *Information Systems Management* 26, 1 (2009), 28–38.

[17] N. Fulda, D. Ricks, B. Murdoch, and D. Wingate. 2017. What Can You Do with a Rock? Affordance Extraction via Word Embeddings. In *Proceedings of the International Joint Conference on Artificial Intelligence*. 1039–1045.

[18] M. Garbuio and N. Lin. 2021. Innovative Idea Generation in Problem Finding: Abductive Reasoning, Cognitive Impediments, and the Promise of Artificial Intelligence. *Journal of Product Innovation Management* 38, 6 (2021), 701–725.

[19] E. Gerber. 2007. Improvisation Principles and Techniques for Design. In *Proc. CHI*. 1069–1072.

[20] Jonathan Gottschall. 2012. *The Storytelling Animal: How Stories Make us Human*. Houghton Mifflin Harcourt, New York.

[21] D. Kahneman. 2011. *Thinking, Fast and Slow*. Macmillan, New York.

[22] P. Karimi, J. Rezwana, S. Siddiqui, M.L. Maher, and N. Dehbozorgi. 2020. Creative Sketching Partner: An Analysis of Human-AI Co-Creativity. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*. 221–230.

[23] A. LaViers. 2019. Counts of Mechanical, External Configurations Compared to Computational, Internal Configurations in Natural and Artificial Systems. *PLoS ONE* 14, 5 (2019), e0215671.

[24] M.T. Llano, M. d'Inverno, M. Yee-King, J. McCormack, A. Ilsar, A. Pease, and S. Colton. 2020. Explainable Computational Creativity. In *Proceedings of the 11th International Conference on Computational Creativity*. 334–341.

[25] M. Mara, M. Appel, H. Ogawa, C. Lindinger, E. Ogawa, H. Ishiguro, and K. Ogawa. 2013. Tell Me Your Story, Robot. Introducing an Android as Fiction Character Leads to Higher Perceived Usefulness and Adoption Intention. In *Proc. HRI*. 193–194.

[26] N.H. Møller, C. Bossen, K.H. Pine, T.R. Nielsen, and G. Neff. 2020. Who Does the Work of Data? *Interactions* 27, 3 (2020), 52–55.

[27] M. O'Shaughnessy, G. Canal, M. Connor, M. Davenport, and C. Rozell. 2020. Generative Causal Explanations of Black-Box Classifiers. arXiv:2006.13913.

[28] M. Oudah, T. Rahwan, T. Crandall, and J. W. Crandall. 2018. How AI Wins Friends and Influences People in Repeated Games with Cheap Talk. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence*. 1519–1526.

[29] S. Passi and S. Jackson. 2017. Data Vision: Learning to See Through Algorithmic Abstraction. In *Proc. ACM CSCWSC*. 2436–2447.

[30] R. K. Sawyer. 2012. *Explaining Creativity: The Science of Human Innovation*. Oxford University Press, New York.

[31] K. Schmidt. 2011. *Cooperative Work and Coordinative Practices*. Springer, London.

[32] Schumpeter. 1942. *Socialism, Capitalism and Democracy*. Harper and Brothers, New York.

[33] D. A. Schön. 1992. Designing as Reflective Conversation with the Materials of a Design Situation. *Knowledge-Based Systems* 5, 1 (1992), 3–14.

[34] S. Vieira, M. Benedek, J. Gero, G. Cascini, and S. Li. 2021. Brain Activity of Industrial Designers in Constrained and Open Design: The Effect of Gender on Frequency Bands. In *Proceedings of the Design Society*. 571–580.

[35] Q. Wang, K. Saha, E. Gregori, D. Joyner, and A. Goel. 2021. Mutual Theory of Mind in Human-AI Interaction: How Language Reflects What Students Perceive About a Virtual Teaching Assistant. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–14.

[36] R. Zhang, C. Walshe, Z. Liu, and L. Guan. 2020. Atari-Head: Atari Human Eye-Tracking and Demonstration Dataset. In *Proc. AAAI*. 6811–6820.