Agentive Collaboration with Creative AI

Anonymised for Review
…
…, …
…

ABSTRACT

In this paper we describe various work-in-progress efforts to explore human–AI collaboration in creative domains, with a specific focus on the iterative and exploratory early conceptual phases of creative work. We document a series of AI models and interactive systems encapsulating them that serve to explore the state-of-the-art in that area.

KEYWORDS

datasets, neural networks, gaze detection, text tagging

ACM Reference Format:
Anonymised for Review. 2018. Agentive Collaboration with Creative AI. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 5 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

For AI to be a useful partner in innovation, it must be built from the ground up to collaborate creatively, requiring interdisciplinary knowledge across machine learning, cognitive science, and design. That means thinking about how we can actively and iteratively work *with* generative AI, not just how we can ask them to work *for* us. At the Designing with AI Lab (DWAIL) of the University of Sydney, our approach is motivated by the cognitive science of design. Real-world design problems — whether they're in health, tech, or engineering — are by definition ill-formed, complex, and underspecified. Problem-framing and problem-solving are required concurrently. But in that regard, the majority of current generative AI systems have a fatal flaw: they assume words always mean the same thing. Take OpenAI's CLIP [16], which relates the semantic content of a text prompt to the semantic content of an image. For all its amazing capability, at its heart is an objective semantic mapping — the assumption that each user always means the same thing with each word. As designers, artists, and HCI researchers, this community knows that this is not only inaccurate, but potentially damaging to groups and causes underrepresented by the majority of the data.

When it comes to AI augmentation of real-world design, the co-evolutionary and constructive processes of creative thinking are incompatible with such a semantically stagnant system. Techniques are emerging that allow users to guide and modify word meanings in image-generating AI, but it's early days still, and applications of those techniques to creative contexts are still being developed. At DWAIL, our hypothesis is that systems which model and adapt to their users' unique perspectives — cultural, professional, and personal — will not only be more effective co-creative partners, but also be useful for a much more diverse range of users. In this position paper we discuss a variety of work-in-progress experiments being conducted in our lab to explore this more constructive, co-creative, and conceptual way of machine-assisted making.

Conference acronym 'XX, June 03–05, 2018, Woodstock, NY

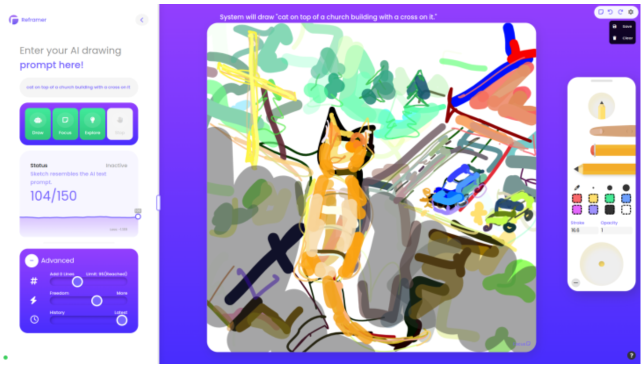
ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

2 REFRAMER: REAL-TIME COLLABORATIVE SKETCHING WITH AI

We have developed an interactive system (Figure 1) called Reframer that supports real-time collaborative sketching with an AI drawing model [13]. As a web application, Reframer is designed to work well on a digital graphics tablet with a stylus, such as the Wacom Cintiq 16". It is based on an adapted version of the CICADA human-AI drawing agent [7], and enables users to interact with a prompt-guided optimization process based on the CLIP neural network. CICADA's optimizer utilizes a differentiable rendering engine searching a space of possible Bézier curves that make up a sketch [4]. With the CICADA model directly embedded in the Reframer drawing interface, users can modify the sketch (and prompt), and the system will immediately incorporate their changes as it converges[1]. Reframer supports a wide range of interactions, with users able to control visual aspects such as style, transformation and the history, as well as facets of the AI model like the learning rate, stroke penalization, and pruning [8]. Observing a design trade-off between emergent system behavior (i.e. more system agency) and direct user control (i.e. less system agency), we developed a feature allowing users to "focus" the optimization by providing sub-prompts that apply within specified regions of the sketch [13]. These regions can be moved, altered, or expanded as the user's design intent evolves within the emergent creative dialogue.

Our latest additions to Reframer are exploring a "machine-led" mode for getting the user "unstuck" from creative blocks. Our goal here is to allow the user to choose how to continue their sketch from among a set of labeled options. We hope to conduct experiments into how and why users engage with a more machine-led rather than real-time collaborative system, as well as how the use of those different modalities affects their experience. By incorporating Quality-Diversity algorithms [15] with the CICADA backend, users are able to create a sketch with a prompt such as "a dog", but then specify qualities they want the provided options to be diverse in. The diversity-promoting algorithm will then produce a set of continuations of the drawing that are diverse in those qualities, such

[1] A video demonstrating this seamless interaction with Reframer can be found at https://vimeo.com/760319552.



System will draw "cat on top of a church building with a cross on it."

Enter your AI drawing prompt here!

cat on top of a church building with a cross on it

Status Inactive

Sketch resembles the AI text prompt.

104/150

Figure 1: Revised version of Reframer with “Focus regions.” The user sketch is of "A cat on top of a church building with a cross on it."

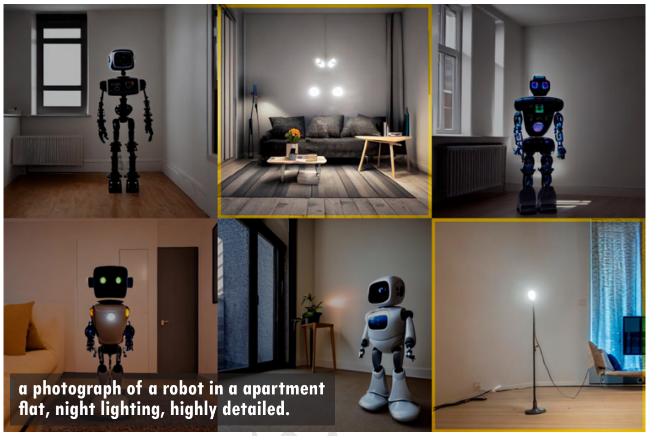
as a set of dogs that are varyingly "fluffy" and varyingly "friendly". This means users could find alternative representations for a sketch that may subvert their expectations and help them refine their design goals. These sketches can take significant time to generate, so options are presented as they become available. Users can select one at any time and return to drawing "with" the AI in the original Reframer co-drawing modality. This will be explored in an upcoming user study in which participants can trial the diversity interface hands-on in response to a realistic design challenge.

3 CO-CREATIVE DRAWING SYSTEMS FOR DIGITAL ILLUSTRATION

Seeking to explore how Reframer was used by more-professional illustrators and visual designers, we conducted two ethnographic studies. The first was autoethnographic, with a researcher who was a competent visual designer exploring and documenting his use of Reframer in a series of digital painting tasks. In a follow-up study, we extended the same principles to an ethnographic study of visual designers.

The main finding of the exploratory autoethnographic study was the idea that Reframer – with its concurrent and simultaneous human and machine inputs – forced a process of near-continual daptation. This adaptation was at times both positively stimulating and negatively restrictive of creative self-efficacy. When “it worked” it paved the way for a feeling of“collaborativeness” between the researcher and Reframer. When the addition or modification of drawing elements was successfully negotiated between the two parties, there was a sense that the human could understand Reframer’s intention and quickly incorporate its changes into the emerging work. In our autoethnographic study there were several occasions where the researcher felt creatively influenced by Reframer’s interjections, attributing them more to the system than to themselves. There were also plenty of occasions where Reframer’s intractability or incomprehensibility were perceived as impeding progress.

Based on the results of the autoethnographic study, 7 practicing visual designers were recruited for a follow-up study, where they performed familiarisation tasks then a series of digital painting like activities with Reframer. Our analysis showed that these designers discovered serendipitous moments where the drawings made them feel oddly satisfied or inspired, often by abstract ideas within Reframer’s sketches like form components or colour combinations. However, the users felt the system as a whole was not as capable as other drawing platforms like Illustrator due to missing features. Probing further into this, users did feel that a tool like Reframer could be useful in the early concept-generation stage, if it were easy to export Reframer sketches as “templates” into a tool like Illustrator for the addition of more serious detail.



a photograph of a robot in a apartment flat, night lighting, highly detailed.

Figure 2: HRI design exploration (from [6]): selected image generations (highlighted with yellow frame) led the designers speculate about pervasive embodied AI that has minimal anthropomorphic features and blends with mundane objects.

4 APPLICATIONS IN SPECULATIVE DESIGN FOR HUMAN-ROBOT INTERACTION

We have also explored applying generative AI to the domain of Human-Robot Interaction (HRI) design [6]. HRI is particularly prone to the universally recognised phenomenon of design fixation [11] due to collective imaginations of what a robot should look like or behave [1]. We wanted to first see whether “off-the-shelf” generative text-to-image models could help HRI designers break through this fixation. In collaboration with design researchers from TU Delft [6], we conducted a four-week-long exploration in which we used StableDiffusion and DALL-E to ideate and visualise robotic artifacts and robot sociotechnical imaginaries (see Fig. 2). Following a first-person research approach, we shared our image generations along with textual annotations about our designerly intentions and reflections on a weekly basis in the form of digital postcards. Our analysis revealed the potential of generative AI to imagine novel robot concepts and surface existing assumptions. However, our exploration also revealed the shortcomings of current AI systems that follow a non-dialogic interaction paradigm, meaning that often many attempts were required for a desired outcome. Furthermore, the role of the designer as curator and interpreter, along with collaborative efforts between the designers, proved crucial to reveal the conceptual value of the generated images towards HRI design.

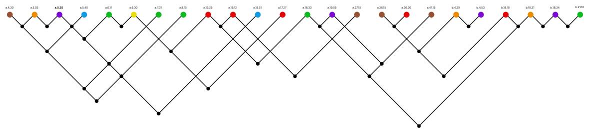


Figure 3: A Linkograph representation of the sequence of design actions within a Reframer-and-human design sequence. Each color represents action categories from a schema including "Human draws", "Human interprets Reframer", and so on.

We are in the process of exploring applying Reframer's collaborative drawing system to this HRI context, with the hope that it will overcome the requirement for repetitive prompt engineering. We look to Reframer as a means for the user of generative AI to adopt direct design roles, rather than being a curator and interpreter of AI content. In a current study at DWAIL, a first-person approach is again being utilised. An individual researcher is working with Reframer to navigate a range of HRI domains and conceptualise new robotic forms and interactions through annotated storyboards. Following this study, future work will see the introduction of multiple human and AI agents into instances of these collaborations.

As co-design workshops are recognised for their benefits in improving idea generation in service design (Steen et al., 2011), our intention is to facilitate workshops involving both human stakeholders and AI participants within domains of service robotics. First, these workshops will involve pairs of designers working with an instance of Reframer to similarly approach a HRI design problem and produce storyboard representations of robotic interactions. We will then scale to larger groups of human stakeholders and multiple instances of Reframer collaborating in a design session with specific HRI objectives. To visualise and analyse the states and progression of these design sessions as a system of interlinked 'design moves', we will use linkographs (Blom and Bogaers, 2018) as part of our analysis process. Linkography tracks the sequence of actions within a design process, and how those actions relate to earlier and later actions as a design's goals and structure emerge (see Figure 3 for an early example). We will showcase this evolving participation of both the human and AI agents in these sessions through categorically coding 'design moves' by the degree of initiative that a human or AI agent held over them in relation to any source of inspiration from the system. This will serve as a framing for our qualitative analysis, hopefully providing further insight into how collaborative generative AI acts agentively alongside humans.

5 METAGENERATION: EXPERIMENTS IN PROMPT EXPLORATION AND DIVERSIFICATION

In the context of generative models that produce images from text, we often encounter the creation process proposed as a one-off command by the user, after which the AI takes over. As our explorations above have shown, this is very limiting, since the user has little to no influence over the image, models sometimes "interpret" prompts in unusual and unexpected ways, and in any case the use may not yet know exactly what they want.

One way of beginning to address such limitations is through systems that suggest modifications to the prompts used to guide generation. We have proposed two algorithms for these kinds of modifications [10]. In the first, a model takes the user-provided prompt as input, and uses CLIP's latent space [16] to find a set of adjectives semantically aligned with the prompt. Then a search algorithm selects the subset of words that are furthest apart in terms of affect (meaning they "feel differently") according to another predictive model based on psychological studies of word affect. The second proposed algorithm is intended for the case where a user wants their current creation to exhibit certain qualities that they have observed in another ('source') image, but perhaps cannot put into words. We compute the CLIP latents over a large set of words, and suggest a subset to the user that align the most with the source image while being unrelated to the current one. We can then modify the prompt and generate a new image that exhibits the desired quality. This (in preliminary qualitative comparisons) outperforms the naïve approach of conditioning generation on both images, which tends to incorporate shallow or irrelevant qualities of the source. We have yet to incorporate these capabilities into interactive systems or test them with real users, but both are in progress.

An alternative approach to influence the results of a generative model, one that does not rely on the prompt, is to guide the process with other kinds of information. An example we have been experimenting with is affect scores, a psychometrically validated approach to cross-culturally assess people's feelings about a wide range of stimuli [14]. This is done by surveying people and constructing a three dimensional model in terms of Valence (i.e. positive/negative), Arousal (i.e. exciting/calm) and Dominance (i.e. controlling/controlled) for a given set of words, phrases or images. From this dataset we have trained a neural network that can predict valence, arousal, and dominance from the latents of CLIP semantic space [9], and use this network to guide generative processes. For generation-by-optimisation methods such as VQGAN+CLIP [2] and CLIPDraw [5], this guidance can be accomplished by penalising how far our neural network's prediction of the output image's affect is from a target proposed by the user. Examples of images high and low on each of these three dimensions can be seen in Figure **??**, all generated by VQGAN+CLIP with the same prompt: "the sea". We trained a second neural network for predicting affect scores based on BERT encodings [3] (a different semantic model) which can be used to condition based on affect when generating images with Stable Diffusion.

6 DISCUSSION

By combining methodologies across the epistemological spectrum, from autoethnography, to thematic analysis of user interviews, to quantitative survey-based user studies, research-through-design prototyping, to in-silico algorithms research, we have been slowly building up a picture of what this new generation of creative AI systems can do. Through this process we've built a model of how and when users interact with these systems in a mixed-initiative, agentive way, and how and when they choose to clamp down on

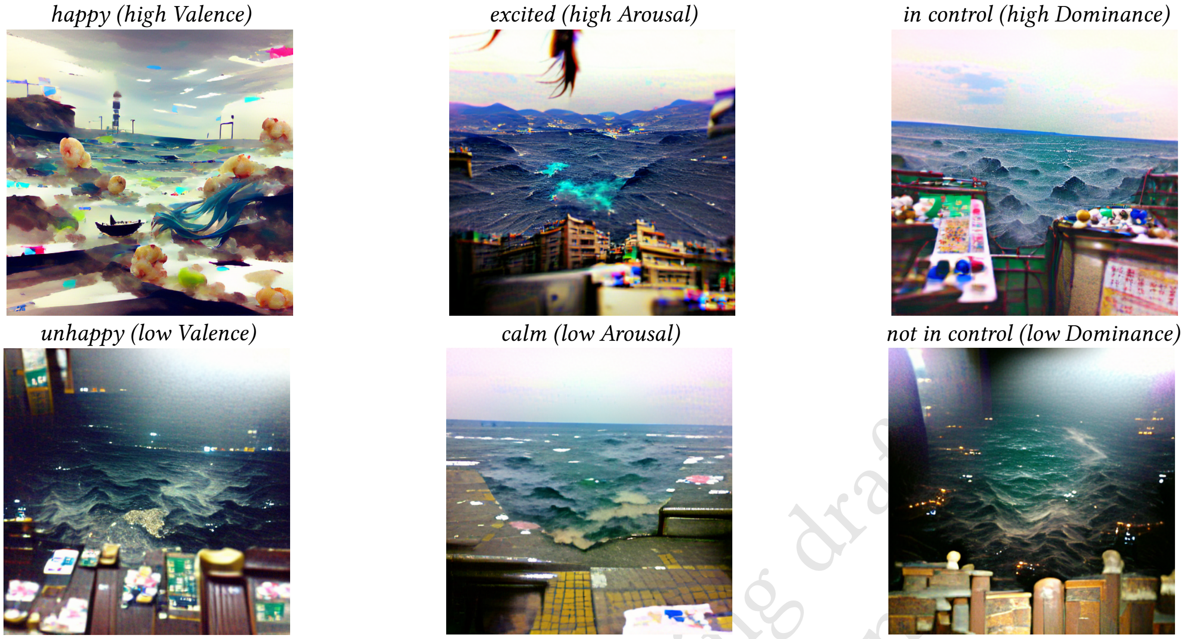

happy (high Valence)
excited (high Arousal)
in control (high Dominance)
unhappy (low Valence)
calm (low Arousal)
not in control (low Dominance)

Figure 4: Example images of "the sea" generated using VQGAN+CLIP with our affect conditioning, maximising (top row) and minimising (bottom row) the three dimensions we utilise: valence (left), arousal (middle), and dominance (right).

their agency and interact with them as tools. We call this the Model of Co-creative Agentive Flow, or M-CAF [12]. M-CAF (see Figure 5 frames both co-creative agentive flow and tool-supported flow as useful states in which a user can interact with a system, but the system's role former is collaborative and mixed-initiative while in the latter it is tool-like and unobtrusive.

These two states are mediated by disruptors of agency, strategies for interacting with agents, and expectations that the user has coming into the interaction. In this framing, systems are neither "creativity support tools" or "mixed initative co-creative systems", but adopt properties of either in a situated way as the creative act unfolds. We should speak of systems not as "being mixed initiative", but as "having the capacity for mixed initiative interaction". If we want our systems to be co-creative and collaborative, then we should design them to afford the state of agentive flow as much as possible, and offer users strategies to support and maintain that state.

ACKNOWLEDGMENTS

Funding sources acknowledgement anonymised for review.

REFERENCES

[1] Patrícia Alves-Oliveira, Maria Luce Lupetti, Michal Luria, Diana Löffler, Mafalda Gamboa, Lea Albaugh, Waki Kamino, Anastasia K. Ostrowski, David Puljiz, Pedro Reynolds-Cuéllar, Marcus Scheunemann, Michael Suguitan, and Dan Lockton. 2021. Collection of Metaphors for Human-Robot Interaction. In *Designing Interactive Systems Conference 2021* (Virtual Event, USA) *(DIS '21)*. Association for Computing Machinery, New York, NY, USA, 1366–1379. https://doi.org/10.1145/3461778.3462060

[2] Katherine Crowson, Stella Biderman, Daniel Kornis, Dashiell Stander, Eric Hallahan, Louis Castricato, and Edward Raff. 2022. Vqgan-clip: Open domain image generation and editing with natural language guidance. In *European Conference on Computer Vision*. Springer, 88–105.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[4] Kevin Frans, Lisa Soros, and Olaf Witkowski. 2022. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *Advances in Neural Information Processing Systems* 35 (2022), 5207–5218.

[5] Kevin Frans, Lisa B Soros, and Olaf Witkowski. 2021. Clipdraw: Exploring text-to-drawing synthesis through language-image encoders. *arXiv preprint arXiv:2106.14843* (2021).

[6] Marius Hoggenmueller, Maria Luce Lupetti, Willem van der Maden, and Kazjon Grace. 2023. Creative AI for HRI Design Explorations. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction* (Stockholm, Sweden) *(HRI '23)*. Association for Computing Machinery, New York, NY, USA, 40–50. https://doi.org/10.1145/3568294.3580035

[7] Francisco Ibarrola, Tomas Lawton, and Kazjon Grace. 2022. A Collaborative, Interactive and Context-Aware Drawing Agent for Co-Creative Design. *arXiv preprint arXiv:2209.12588* (2022).

[8] Francisco Ibarrola, Tomas Lawton, and Kazjon Grace. 2022. A Collaborative, Interactive and Context-Aware Drawing Agent for Co-Creative Design. *arXiv preprint arXiv:2209.12588* (2022).

[9] Francisco Ibarrola, Rohan Lulham, and Kazjon Grace. 2023. Affect-Conditioned Image Generation. *arXiv preprint arXiv:2302.09742* (2023).

[10] Francisco J Ibarrola and Kazjon Grace. 2023. Prompt diversification for iterating with text-to-image models. In *International Conferences on Computational Creativity*. ICCC.

[11] David G. Jansson and Steven M. Smith. 1991. Design fixation. *Design Studies* 12, 1 (1991), 3–11. https://doi.org/10.1016/0142-694X(91)90003-F

[12] Tomas Lawton, Francisco J Ibarrola, and Kazjon Grace. 2023. When is a Tool a Tool? User Perceptions of System Agency in Human–AI Co-Creative Drawing. In *Proceedings of Designing Interactive Systems (to appear)*.

[13] Tomas Lawton, Francisco J Ibarrola, Dan Ventura, and Kazjon Grace. 2023. Drawing with Reframer: Emergence and Control in Co-Creative AI. In *Proceedings of the 28th International Conference on Intelligent User Interfaces*. 264–277.

[14] Charles Egerton Osgood, William H May, Murray Samuel Miron, and Murray S Miron. 1975. *Cross-cultural universals of affective meaning*. Vol. 1. University of Illinois Press.

[15] Justin K Pugh, Lisa B Soros, and Kenneth O Stanley. 2016. Quality diversity: A new frontier for evolutionary computation. *Frontiers in Robotics and AI* (2016), 40.

[16] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision.

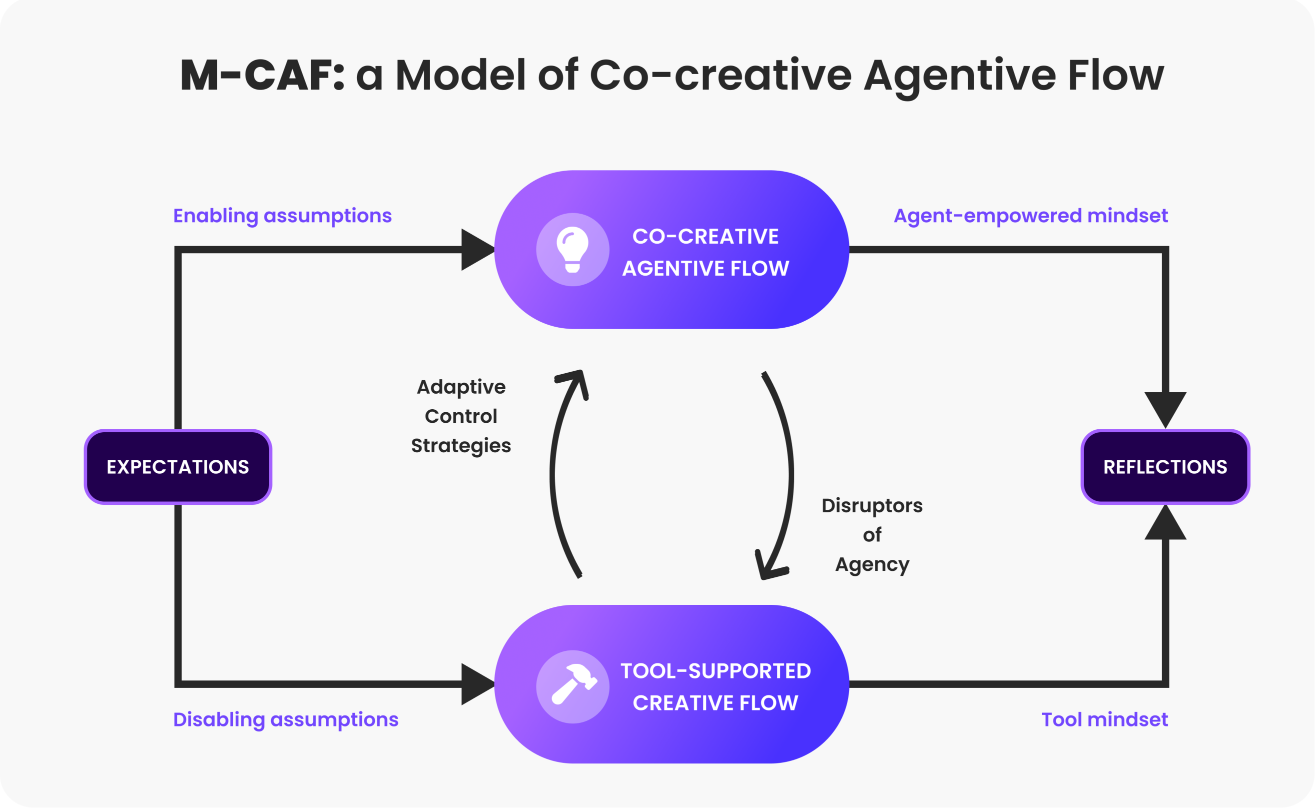

M-CAF: a Model of Co-creative Agentive Flow
Enabling assumptions
CO-CREATIVE AGENTIVE FLOW
Agent-empowered mindset
Adaptive Control Strategies
EXPECTATIONS
REFLECTIONS
Disruptors of Agency
TOOL-SUPPORTED CREATIVE FLOW
Disabling assumptions
Tool mindset

Figure 5: Model of Co-Creative Agentive Flow: The model starts with expectations, then cycles between experiencing agentive flow and tool-use, then concludes with users having adopted mindsets and reflected on their experiences.

In *International Conference on Machine Learning*. PMLR, 8748–8763.