Experiments in Text-2-Tool-Use for Relatable Computational Creativity

Simon Colton

School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

s.colton@qmul.ac.uk

Abstract

We propose *relatable computational creativity* as an umbrella term for various theoretical and practical developments, such as explainable computational creativity, artificial authenticity, framing, embodiment, intrinsic motivation, artificial curiosity and creativity theatre. In overview, if a creative AI agent is relatable, then people can imagine themselves doing something similar to it for similar reasons. In this context, we explore the ability for AI systems to use the same tools that people use. Experimenting with a casual creator app for producing decorative art, we investigate how it can learn to employ the tools it makes available for users. Working with a vision-language model affords an interface where users can request that an image be altered in such a way as to achieve a particular look. The system responds with a series of tool actions shown offline to reliably alter images to achieve the desired look. We present experimental results showing that such tool usage can be learned, and describe the approach with respect to aspects of relatable computational creativity.

Introduction and Motivation

When a character in a novel has a similar backstory, attitudes, attributes or situation to the reader or someone they know, this can make the character more engaging and the story more resonant. Maslej, Oatley, and Mar (2017) put this succinctly:

> ... we know characters such as Sherlock Holmes, Anna Karenina, and Mrs. Dalloway in something like the way we know people in our daily interactions. We may think about these characters' intentions, respond emotionally to their actions, and invest ourselves in their outcomes, even though we know they are not real.

Such *relatable* characteristics can afford a bond between reader and character, so that when dramatic events occur, the reader cares about that character and is invested in the story.

When a visitor to an art gallery views a painting, in one sense, the experience is relatable: the artist has made marks with pigment on a canvas, as the visitor could. However, the visitor may marvel in elements of the painting or its production that seem beyond their own physical or mental abilities. As per Colton (2008), these could be many things, including *skill*, expressed through finesse or innovation in technique; *imagination* expressed through subject matter choices and perspective; or *appreciation* expressed through style, artistic treatment or political projections. So, just as relatable characters in novels can act as a platform for dramatic events to unfold, relatable characteristics in artists and processes can act as a platform for the projection of notions of creativity.

Generative AI techniques have recently become powerful enough to rapidly transition from university research laboratories and artist studios to industrial-scale worldwide adoption. Appropriately, these have focused on ease-of-use through prompting, quality of output and fidelity to specifications. However, end-to-end production via text-2-X techniques have, to some extent, excluded creative professionals. For instance, music lovers can now listen to audio files produced directly from natural-language prompts, with no need for composers or musicians. Likewise, visual art lovers can prompt generative AI systems to find imagery including content, styles and perspectives that satisfy their needs and tastes.

Left behind somewhat in the rush to industrialise end-to-end generative AI is the notion that creative AI systems can be entertaining, inspirational and educational to people interested in crafting artefacts rather than simply prompting for them (whether at an amateur or professional level). Within the context of casual creators (Compton and Mateas, 2015), we look here at ways in which a creative AI agent can be relatable while making artefacts, in order to educate and inspire amateur creators. In particular, we focus on user-centric generative systems which learn how to employ their own tools in order to better serve users in a relatable way.

In the next section, we expand on the notion of *relatable computational creativity*, by surveying research into implementation of human-like motivations, behaviours and evaluation techniques surrounding creative acts. We introduce a broad definition and suggest some refinements which could form a framework for implementing, describing and assessing future generative systems. Following this, we describe the context of casual creation and the *Art Done Quick* application used in our experiments. We then describe how we have used the MobileCLIP language-vision model (Vasu et al., 2024) in experiments to enable Art Done Quick to find sequences of tool usages to transform an image in response to a user's prompt. We present the results of the experiments which show that it is possible to discover tool usages which can be controlled via text. We then frame the approach in terms of aspects of relatable computational creativity, and end by describing some future directions for this work.

Relatable Computational Creativity

We start from the position that interacting with a generative AI system onto which we can genuinely project impressions of creativity could have similar benefits in terms of inspiration, education and entertainment that working with a creative person brings. If we draw from the literature on human creativity, Fink et al. (2012) showed that "cognitive stimulation via common or moderately creative ideas [from other people] was effective in improving [the] creativity" of participants asked to ideate alternative uses of everyday objects. Moreover, a study by Katz et al. (2022) showed "that a short narrative about the creativity of a person can produce implicit impressions of that person as creative," and Colton (2008) suggested that understanding the creative processes that went into producing an artefact can increase the value people attribute to it. More generally, Decety and Grèzes (2006) highlight the power of imagining the behaviour of others, arguing that "mental simulation may be a representational tool to understand the self and others".

When the creative counterpart in interactions switches from a person to a computational process, it can be difficult to understand the motivational, generative or evaluational processes the software employs. If we look at the two most successful generative AI techniques currently, we see that neither of them affords a particularly relatable process. Diffusion models produce artefacts by starting with a noisy image and a conditioning prompt and turning the noise into appropriate content over a series of iterations (Yang et al., 2024). In an equally unrelatable way, Transformer models take a partially formed artefact such as a text prompt and repeatedly calculate how to extend it by a single token (Lin et al., 2022).

In computational creativity research, there have been many projects leading to the implementation of human-like creative behaviours in AI systems. While these projects had diverse motivations, outcomes and applications, we propose to introduce the umbrella term *relatable computational creativity* to highlight that the behaviours implemented add to an overall picture of AI systems that people can interact naturally with, potentially as creative equals, and certainly as more than generative tools (Ventura, 2016). We prefer the term *relatable* to the term *human-like* because the latter might imply that simulating all elements of humanity is an aim, when subsets thereof could be beneficial. Just as relatable fictional characters in books are valuable to story telling even though they don't exist in the usual context of human life, relatable AI agents can be valuable to creative processes even though they don't exist in the usual context of human creative life. There are observer issues in computational creativity, and keeping a clear separation between human and artificial creativity may be beneficial (Colton et al., 2014).

In broad terms, a relatable computational creativity system employs some processes leading to the production of an artefact (or problem solution) which an average human observer could imagine him/herself undertaking personally. The processes could include any simulated motivations, generative behaviours and/or evaluations that the system undertakes, and we can imagine a system with more relatable processes being more relatable overall. With a partial survey of the literature on computational creativity, we can suggest the following five main areas where relatable processes were studied.

• **High Level Motivations:** Most people create artefacts for a reason, ranging from self expression to financial gain. This is such an important element of creative behaviour that a lack of intentionality in AI systems can make them seem unrelatable, firmly positioning them as just tools. The question of intent was raised as part of The Painting Fool project in (Krzeczkowska et al., 2010) and (Colton and Ventura, 2014). Guckelsberger, Salge, and Colton (2017) raised the question of why AI systems should be creative, providing empowerment maximisation as an intrinsic motivation for creative AI agents (Guckelsberger, 2020). A number of other relatable reasons for AI systems to be creative have been studied. These include exhibiting artificial curiosity and novelty-seeking, as explored in (Saunders, 2001). Saunders and Bown (2015) also introduced the notion of *computational social creativity*, studying creative AI agents that interact in a community, and Colton and Banar (2023) proposed adding to artistic cultures as a motivation for computational creativity.

• **Explainability:** Relating to someone or something involves understanding, empathizing and connecting, which is not always possible with generative AI systems. The field of explainable AI (Abusitta, Li, and Fung, 2024) acknowledges that some of the most powerful AI techniques, specifically those from deep learning research, have underlying processes that are difficult to understand. Llano et al. (2020) introduced the notion of *explainable computational creativity* to focus on creative AI systems, providing some design principles that can give computational creativity systems a voice for a deeper user interaction. This built on the study of how AI systems can add value to artefacts by *framing* their work in terms (amongst other things) of descriptions of the processes undertaken, internal evaluations and relation to the work of others (Cook et al., 2019). In generative deep learning, the process of chain of thought reasoning (Chen et al., 2025) has been introduced as a technique for interaction with LLMs, whereby a difficult problem (or generative task) is broken into stages, with separate prompts for each. One of the benefits this brings is more transparency and insight into the process.

• **Embodiment and Simulation:** There is perhaps no more relatable process than a demonstration where an AI agent performs actions in a way that is easily replicable by a person. When this is done robotically in physical reality, the effect is heightened, and there has been much research into *embodied computational creativity*. Moruzzi (2022) showed that "the physical dimension of artificial systems interacting with human artists contributes to the perception of the interplay between artificial and human agents as a creative collaboration." Moreover, Guckelsberger et al. (2021) provide a systematic review and prescriptive analysis of embodied computational creativity, suggesting directions for further study. There have been many successful robotic painters such as those described in Cohen (2017), Tresset and Fol Leymarie (2012) and Lindemeier (2018), which demonstrate how easy it is to relate to physical processes. In addition, other creative AI systems have simulated physical processes on-screen, e.g., The Painting Fool software simulates paints being applied to a canvas with an on-screen hand holding a paintbrush (Colton, 2012). Cook and Colton (2018) introduced the notion of *continuous computational creativity* in order to study how a creative AI system could have a presence in a community. Moreover, Colton et al. (2020b) investigated the notion of *creativity theatre* with a system demonstrating its abilities through human-like tool use and creative aims.

• **Personhood:** The process of 'othering' "consists of applying a principle that allows individuals to be classified into two hierarchical groups: them and us" (Kobayashi, 2020), often to exclude or ostracize them. It is natural to see AI systems as others in creative fields, as they are not people to start with, are perceived to threaten jobs, and are usually not relatable. There have been a number of studies in computational creativity research to explore ways in which communities might be encouraged to be more inclusive through advances in how generative AI systems operate and are deployed. One issue is that AI systems are not authentic voices, especially when the artefacts they produce reflect elements of humanity such as emotions or life experiences. Colton, Pease, and Saunders (2018) offer some solutions to this in terms of *artificial authenticity*. Ultimately, this line of thought led to the proposition of a framework called *the machine condition* in which an AI system could express elements of its existence through artistic output (Colton et al., 2020c). This was taken further in Pease, Colton, and Banar (2023) who proposed agency, self-expression and individuality as relatable traits for advancing an AI system towards gaining personhood in a creative domain. While agentic AI has a long history, it has recently been adopted in deep learning to describe autonomous systems that pursue complex goals with minimal human intervention (Acharya, Kuppan, and Divya, 2025). While no-one suggests such agents are given personhood, this trend highlights that autonomy – a quite relatable trait – is important in designing systems wrapped around LLMs.

• **Evaluation:** In creative communities, people often discuss the value of artefacts like paintings or songs, with the person who made them sometimes part of that discussion. Hence, if an AI system can appreciate the value of the artefacts it produces (and those of others), this may seem more relatable than if it cannot. Enabling generative AI systems to evaluate their own creations has been a mainstay technique in computational creativity research, which has been fed by studies into how people appreciate created artefacts. The idea that generated artefacts should evoke emotions of surprise through novelty has been extensively studied e.g., by Macedo and Cardoso (2017). This has been operationalized through techniques such as *novelty search* (Lehman and Stanley, 2008) and *surprise search* (Yannakakis and Liapis, 2016). Associated with the notion of surprise is that of unexpectedness, when surprising results come from a context where people have certain expectations. Grace and Maher (2014) generalise past the surprise evoked by artefacts, and develop a typology of expectations relevant to computational creativity evaluation. Returning to the question of autonomy, Colton, Charnley, and Pease (2011) investigated the idea that a computational creativity system can be evaluated in terms of how much responsibility has been handed over to it.

We argue that – while each project surveyed above had an individual goal, outcome and application – they contribute to an overall picture of effort in simulating relatable computational creativity, which can act as an umbrella term for this effort. Building on the projects of the authors cited above, we further argue that there are many benefits to relatable computational creativity, which depend on the mode in which creative AI systems interact with people. One mode of interaction might be that the AI system acts as a set of tools, but the system itself has the ability to control those tools. This is the context for the experiments described below, and enables the software to provide guidance to the user on tool use in a relatable way. If the software does something surprising, this could provide inspiration, with the relatable actions of the software providing expectations from which unexpectedness – in terms of artefacts and actions – can emerge.

Creative AI systems are generally multi-faceted, and it is unlikely that all the processes they employ would be relatable. This could be for a number of reasons, including: a process is naturally a black box, due to scale and/or complexity, e.g., forward inference in large neural models; or a process is purposefully kept hidden for some reason, perhaps to maintain mystique, or to simulate a subjective opinion or taste. Hence, our understanding of relatable computational creativity should admit the fact that there will be *knowable* and *unknowable* processes at work in creative AI systems, just as there are with creative people.

We also need to consider whether the AI agent is intended to be used in a realistic or more of a role-playing situation. For instance, with generative AI chat-bots, it is very easy for users to project personalities onto the systems, giving them a human-like role in doing so. In this situation, if the chat-bot is asked to provide explanations of its processes, it will likely provide a relatable scenario, even though it is not an accurate portrayal. LLMs are not reliable witnesses to their own creative processes, but we could say here that the system is *pseudo-relatable*, because users are at liberty to pretend it is a person, and take inspiration from its fictional portrayals of internal processes.

Another consideration for relatable computational creativity systems is *why* we relate to the human-like behaviours that they exhibit. One possible answer is because of demonstrations that they undertake, for instance with embodied AI systems creating artefacts in physically relatable ways or via on-screen creativity theatre demonstrations. Another possible answer is via explanations that the system gives, either as part of its normal processing or under scrutiny in post-hoc questioning. Naturally, a relatable system might provide a mixture of demonstrations and explanations. A final consideration for relatable computational creativity is in terms of where we seek relatable processes. As the general public become more aware of how generative AI systems are engineered, questions about their backstory may arise. We therefore need to consider how relatable both the generation-time processes and the development-time processes of an AI system are.

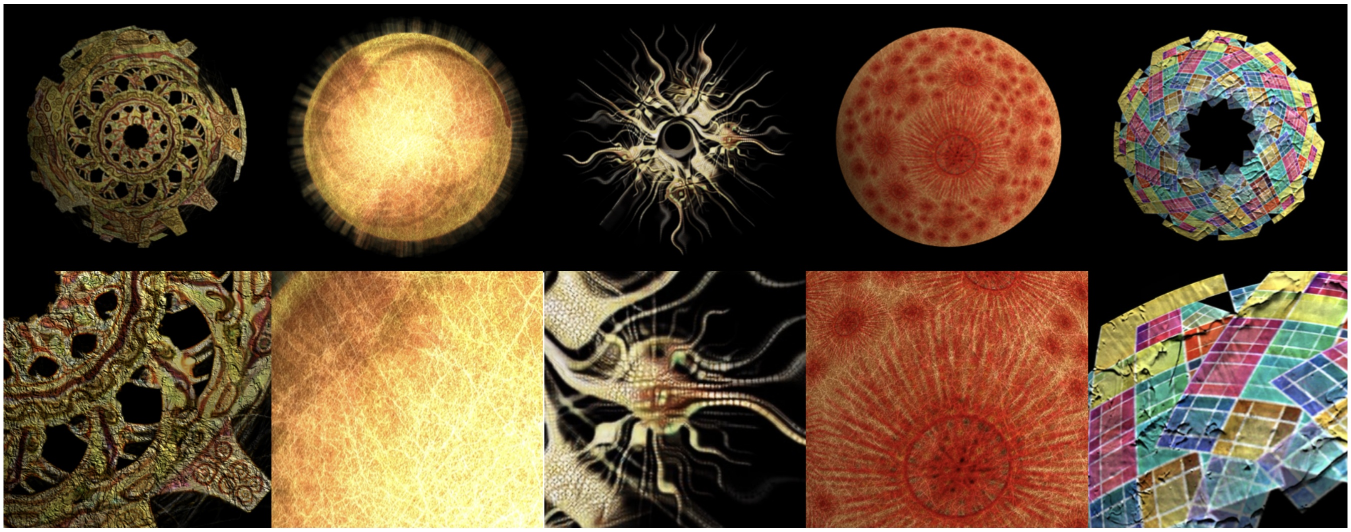

Figure 1: Example images produced with the editing interface of the Art Done Quick App, with details given below each.

Casual Creation and Art Done Quick

Compton and Mateas (2015) introduced the notion of a *casual creator* system, which they described as follows:

> A casual creator is an interactive system that encourages the fast, confident and pleasurable exploration of a possibility space, resulting in the creation or discovery of surprising new artifacts that bring feelings of pride, ownership, and creativity to the users that make them... The user of a casual creator is a *casual* user, and the system can expect no previous domain knowledge, no previous technical experience, or adherence to a long learning programme.

In the original paper and more extensively in Compton (2019), various design patterns, guidelines, typologies and worked examples are presented, which have formed essential reading for anyone implementing a casual creator application. Petrovskaya, Deterding, and Colton (2020) provided a further typology of commercial casual creator applications.

A casual creator app called Art Done Quick is used in the experiments described in the next section. This app enables users to quickly produce pieces of decorative visual art, as described in Colton et al. (2020a) and Colton (2020). Examples of the kinds of images users can produce are provided in figure 1. Users employ Art Done Quick in a two-stage process. Firstly, there is an evolutionary art interface, whereby users can double-tap and drag-and-drop to achieve mutation and crossover of underlying genomes which are rendered as images. In the first screenshot of figure 2, the user has double-tapped the central image to produce the eight variations that surround it. The images produced in this interface tend to be blocky patterns of shapes, which provide the raw material with which to produce more sophisticated images later.

Secondly, there is an editing interface with which users can alter a chosen image in various ways. There are ten editing screens in the app to provide comprehensive functionality in terms of fun things to do with images, as per the 'fun-first methodology' described in Colton et al. (2020a). The screens are as follows, with four of them portrayed in figure 2:

- **Art:** here the size, frequency, depth, grain and type of underlying shapes in the raw material image can be altered.
- **Colour:** here, tinting, greyscale, rainbow, spectrum and posterizing (reducing colours) image filters can be applied.
- **Liquify:** here, the user can add holes, glass beads, twirls, bulges/pinches and carnival mirror effects to images.
- **Montage:** here, the user can create layers which repeat and re-arrange an underlying image.
- **Draw:** here, the user can create a layer by drawing onto the image, using various simulated implements.
- **Effects:** here, the user can employ dozens of image effects, including various blurring/sharpening, dithering and neural style transfer (Jing et al., 2020) filters.
- **Texture:** here, the user can overlay one of 24 textures to add photo-realism to images.
- **Light:** here, the user can add spotlight, flood, glowing, shadow and sparkling lighting effects to images.
- **Stickers:** here, the user can personalise images by adding multiple stickers on top or below them.
- **Text:** here, the user can personalise images by adding text in a number of styles, fonts and sizes to the artworks.

Importantly for the experiments described below, as per the design pattern of 'modifying the meaningful' from Compton (2019), much of the editing functionality is limited to choosing which effect to add, and then tweaking two numerical parameters. As we see in figure 2, there is a top and bottom slider to apply these tweaks with, for each of the editing actions in the liquify, light, texture and effects screens.

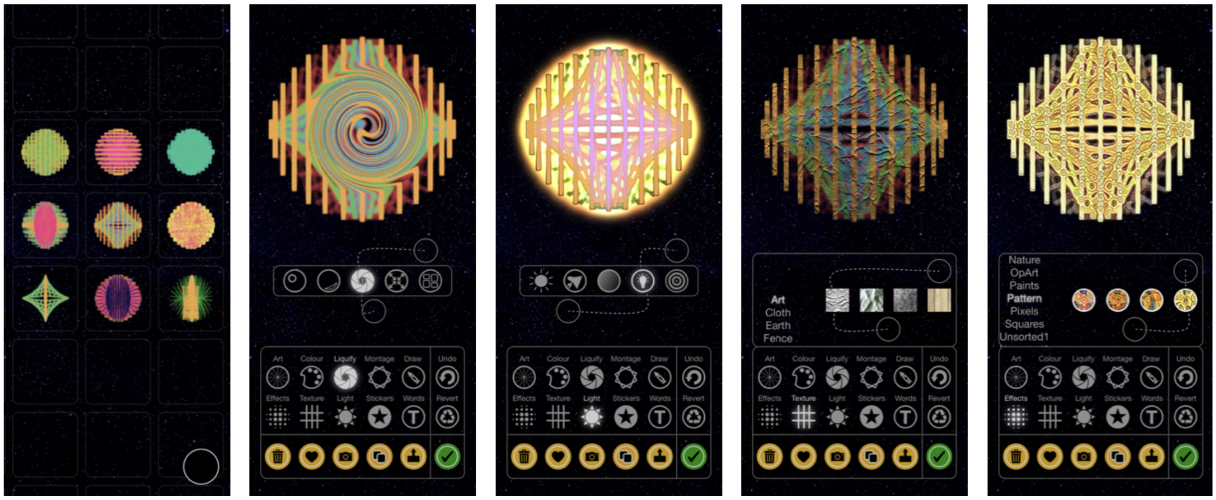

Figure 2: Screenshots of the evolutionary interface and the liquify, light, texture and effects editing screens in Art Done Quick.

Experiments in Discovering Tool Uses

Art Done Quick is an iOS/iPadOS application which has been used as a research platform rather than released commercially. Colton et al. (2020b) implemented an independent control on top of the app, which interacted with the evolutionary art interface and the editing screens to generate images which looked like certain real-world objects (according to the machine vision model ResNet (He et al., 2016)). While that project helped to develop notions of creativity theatre, it didn't bring any benefits to users other than entertainment in demonstration contexts.

For a more applicable study into relatable computational creativity here, we envisage a scenario where a user has an image that they like, but wants to improve it. The user knows to some extent what a better version of it would look like, according to their taste. However, as per the maxim of casual creation, they might not have enough experience of the app, and might not want to undertake trial and error tests to make the better image. Our experiments here support the implementation of an interface where the user chooses an image, then chooses a descriptive word from a given list, and indicates whether the altered image they require should look more or less like the chosen description. As examples, a user might say they want a 'more ghostly' or 'less spiky' version of their chosen image. In response to such a request, we envisage Art Done Quick demonstrating the use of a sequence of editing actions, producing an image which ideally still looks like the original, but reflects the description to an appropriately larger or lesser extent.

To achieve such a text-2-tool-use interface, we added a mobile version of the CLIP vision-language model (Radford et al., 2021) to Art Done Quick, so the software can assess images with respect to descriptive words. The MobileCLIP version we used is described in (Vasu et al., 2024), and consists of two neural models: the first encodes images into a latent space of 512-element vectors, and the second encodes text into the same latent space. Importantly, the loss function for training forced image/text pairs to be mapped to similar parts of the space if the text was an appropriate caption for the text. It also forced pairs to be further apart if the text did not reflect the image. Given a piece of text, T, and an image I, taking the cosine similarity of the CLIP encoding for T with the encoding for I gives a score for how well the text captions the image (or how well the image reflects the text).

With the experiments described here, we exhaustively try out a subset of the image editing actions Art Done Quick has, and determine which of them reliably increase/decrease significantly the cosine similarity of an image with each of a set of descriptive words. For instance, a particular image edit might, on average over a set of starting images, increase the cosine similarity with the word 'colorful', but might decrease the cosine similarity with 'crystalline'. Such information could be operationalised so that Art Done Quick can demonstrate using its tools to achieve a look that users want. We undertook the experiments in three stages: (i) some initial experiments to guide our usage of MobileCLIP (ii) a systematic evaluation of 612 single-image edits over 100 starting images, and (iii) constructing sequences of image edits by chaining together the most promising single-image edits.

Initial Guiding Experiments

Before fully committing to using MobileCLIP, we first tested whether it would be appropriate for our use case. Arias, Baldrich, and Vanrell (2024) highlight certain deficiencies in the ability of CLIP models to assign colours to images, with particular difficulties with achromatic colours such as grey, black and white. As colour is an important element of the images Art Done Quick produces, and a likely target for user requests, we decided to test MobileCLIP's abilities in labeling coloured images. We generated blank images in twelve standard colours including black, white and grey, and used MobileCLIP to encode them into its latent space. We then CLIP-encoded the words 'black', 'white', 'grey', 'blue' etc., and calculated the cosine similarity between every pair of word/colour-image. The results are shown in figure 3. We see that all but one colour matches its corresponding

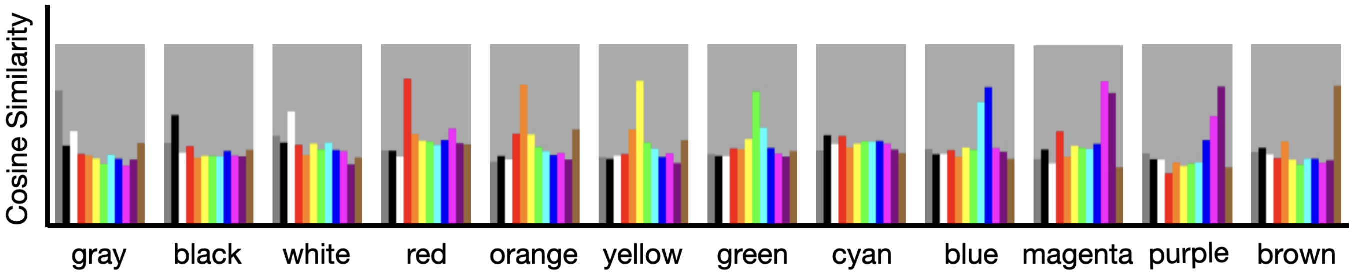

Figure 3: Cosine similarities of colour images (bars) with colour words (on the x-axis).

word more than any other. Importantly, achromatic colours do not cause a problem, with the appropriate colour word having higher cosine similarity to the grey, black and white colour images. However, the word 'cyan' has roughly the same cosine similarity with all the colour images. While these results indicate that there may be problems with lesser known colours like cyan, we felt that they weren't bad enough to necessitate using a different vision-language model.

Another question we addressed by using single-colour images was how to use CLIP encodings in cases where the user wants to alter an image to look *less* like a particular description. Given that CLIP can encode any text, not just single words, there are two alternatives for this: (a) Art Done Quick could suggest image edits which on average produces an image with smaller cosine similarity to a given description than the starting image, or (b) it could suggest edits which produce an image with a higher cosine similarity to a textual negation of the description, for instance the adjective 'icy' negated to 'not icy' or 'hardly icy at all'.

To determine which of these possibilities to use, we produced the same charts as in figure 3, but calculating cosine similarity to the colour words modified with 'not' in front. We found that, for instance, the phrase 'not black' had higher cosine similarity to the black image than any other images, and this was true for all the colours, and however we modified the descriptive word to negate it. From this we deduced that text modifiers don't move the encoded text very far from the unmodified text in the latent space. This meant that the only sensible way to find image edits which lower the reflection of a given descriptive word in an image was to find ones which lower the cosine similarity to the word.

Single-Action Experiments

Our aim here was to find single image edits among those Art Done Quick affords which can reliably move the CLIP encoding of an image towards or away from the encoding of a given descriptive word. In doing so, the edited image should look more/less like the description prescribes. To make the search more manageable, we restricted the image edits to just those editing functions in Art Done Quick in the Art, Colour, Liquify, Light, Texture and Effects screens, with the latter functions broken into Style Transfer and Other Effects. All of the edits in these screens have just two tweakable parameters, which we call the top and bottom parameters in line with the slider positions for them in the user interface. We made strategic choices in the top/bottom parameters to effect quite visible rather than subtle changes.

Action Type	Number of Actions	Top 20 Edits Positive	Top 20 Edits Negative
Art	15 (2.5%)	87 (0.9%)	22 (0.2%)
Colour	57 (9.3%)	1059 (10.6%)	319 (3.2%)
Effects	153 (25.0%)	3589 (35.9%)	662 (6.6%)
Light	45 (7.4%)	417 (4.2%)	155 (1.6%)
Liquify	45 (7.4%)	3057 (30.6%)	107 (1.1%)
Style	96 (15.7%)	1048 (10.5%)	5967 (59.7%)
Texture	200 (32.7%)	743 (7.4%)	2768 (27.7%)

Table 1: Breakdown of single image edits overall and those used in the top twenty for positive and negative movements. Style is Style Transfer actions and Effects are Other Effects

In total, there were 612 single image edits – which we call *actions* – to search over, each represented as a tuple of type, identifier, top parameter and bottom parameter. A breakdown of these into the type of action is given in table 1. We applied each action to 100 different starting images, generated in the way Art Done Quick normally produces them randomly for the evolutionary art interface. This provided 61,200 pairs of (s)tarting and (e)dited images (s, e). For each pair, we generated the CLIP encoding of both s and e, then calculated the cosine similarity against 500 descriptive words. These descriptions were chosen to reflect: visual properties of abstract art images (e.g., dark, pointy, glistening); emotional projections (e.g., happy, melancholic, depressing); artistic terminology (e.g., layered, harmonious, gestural) and generally interpretable terms (e.g., haunted, technical, unruly).

For each descriptive word, d, we calculate the mean, $m(d)$ and standard deviation, $std(d)$, of the cosine similarities between its CLIP encoding and the encodings of the 61,200 edited images. Then, for each pair of description and action taking starting image, s, and producing edited image e, we calculate the *movement* w.r.t. d afforded by a as:

$$movement(a, d, s, e) = \frac{cos\_sim(e) - cos\_sim(s)}{std(d)}$$

Then, for a given description word, d, we calculated the following four measures for every action a, over all the starting, s, and edited, e, image pairs:

• $movement(a, d)$: the mean of $movement(a, d, s, e)$ over all pairs (s, e) that a produces.

• $fidelity(a)$: the mean cosine similarity between the CLIP encodings of images s and e, over all pairs (s, e) that a produces.

• $pConfidence(a, d)$: the proportion of pairs (s, e) that a produces where $movement(a, d, s, e) \geq 0$.

• $nConfidence(a, d)$: the proportion of pairs (s, e) that a produces, where $movement(a, d, s, e) < 0$.

As an overall score, $pScore(a, d)$, for an action, a, with respect to changing an image to look more like a descriptive word d, we calculated:

$$fidelity(a) + movement(a, d) + pConfidence(a, d)$$

Likewise, as an overall score, $nScore(a, d)$, for a with respect to changing an image to look less like d, we took:

$$fidelity(a) - movement(a, d) + nConfidence(a, d)$$

Note that by including the fidelity term, the scores penalise actions where edited images don't look enough like the starting images, as this is an unsatisfying outcome. Similarly, low probability in the action producing a desired movement is also penalised. Note, however, that the fidelity and confidence measures output positive integers less than 1, while the movements range from around -2 to +5, so this term is usually dominant in the overall score for actions which move images a lot. With the overall score for every action with respect to every descriptive word, we determined the best actions able to reliably move images in a positive or negative direction with respect to a user-chosen description word.

The spread of the top 20 actions across the action types is given in table 1. Comparing the percentages, we note that the actions involving changes of the raw art image shapes were rarely the best for positive or negative movements. We also note that style transfer actions are found extensively in the best 20 to move images away from descriptions, while the other effects such as dithering, blurring, etc., were disproportionately used to move images towards descriptions. In table 2, we recorded the number of adjectives (out of 500) that can be moved up to certain multiplicands of the standard deviation in both directions. It is promising to see that for 478 out of 500 description words, there is at least one action that can move an image (on average over 100 images) more than one standard deviation away from its starting point in a positive direction, with 468 words out of 500 having a similar action for the negative direction.

Action Sequence Experiments

The images in figure 1 were produced via Art Done Quick applying around five image edits to a starting art image in each case. In general, looking at the images after the single edits, we found them to be less sophisticated than those possible with a sequence of multiple edits. Hence, in a final round of experiments, we tested whether it is possible to discover longer action sequences which produce appropriate cosine similarity movements while retaining fidelity to the starting image.

Positive Movement (stds)	Num. Descs.	Negative Movement (stds)	Num. Descs.
$m > 4$	11	$m < -4$	0
$m > 3$	44	$m < -3$	1
$m > 2$	172	$m < -2$	100
$m > 1$	478	$m < -1$	468
$m > 0$	500	$m < 0$	500

Table 2: Number of descriptors for which there is at least one action achieving given movement distances.

For each of the 500 description words, we randomly took four actions from the top 20 (in terms of the scores described above) determined previously, for both positive and negative movements. Starting with a randomly generated raw art image, each action was applied in turn, which generated a sequence of five images. From these, the image with the highest movement measure in the given direction was chosen, but only if it had a fidelity with the starting image of 0.7 or above. This produced sequences of 1, 2, 3 or 4 actions, with the breakdown of how many for each given in table 3. Table 3 also shows that the fidelity of edited images reduces but the movement away from the starting image increases as sequence length grows, in both the positive and negative directions. Both of these is as expected, and indicates that we should be able to control the movement and fidelity of edited images while producing more sophisticated imagery through longer action sequences in Art Done Quick. Some illustrative sequences for positive and negative movements are given in figures 4 and 5 respectively.

As examples, the first sequence in figure 4 achieves a 'fiery' look through the application of two style transfer actions. This is also what the first sequence in figure 5 does to reduce the 'blue' in the starting image. To achieve a 'ghostly' look, the second sequence in figure 4 applies dithering, an ambient lighting change and a zoom blur. To reduce the 'detailed' nature of the starting image in the second action of figure 5, an Op-art (Houston, 2009) style transfer action is applied, followed by a texture overlay and finally a zoom blur again. The grey nature of the starting image in the third sequence of figure 5 is removed eventually with a style transfer action, achieving an altered image that is 2.5 standard deviations away from the description 'melancholy', while still having a fidelity with the starting image of 0.768.

Seq Len	Positive Direction Num	Move.	Fid.	Negative Direction Num	Move.	Fid.
1	92	1.14	0.86	95	-1.13	0.87
2	126	1.52	0.80	119	-1.45	0.82
3	145	1.77	0.76	135	-1.60	0.79
4	137	1.80	0.74	151	-1.57	0.78

Table 3: Sequence length, average movement and fidelity for action sequences in both positive and negative directions.

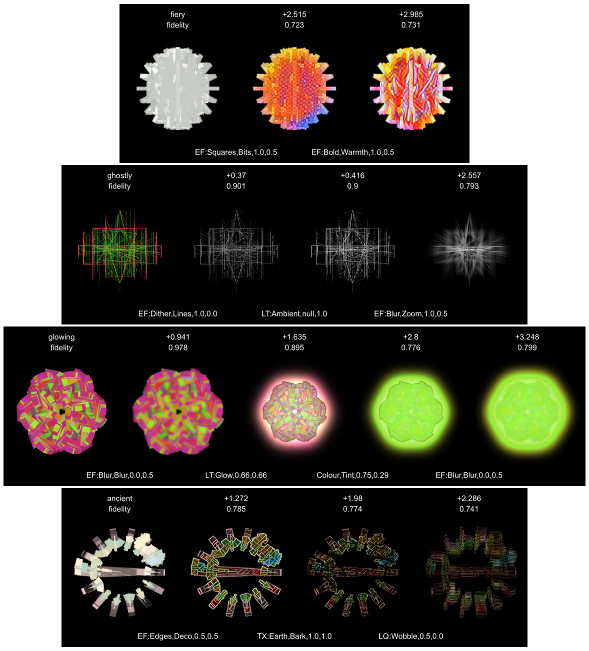

Figure 4: Action sequences which increase cosine similarity with the words 'fiery', 'ghostly', 'glowing' and 'ancient'.

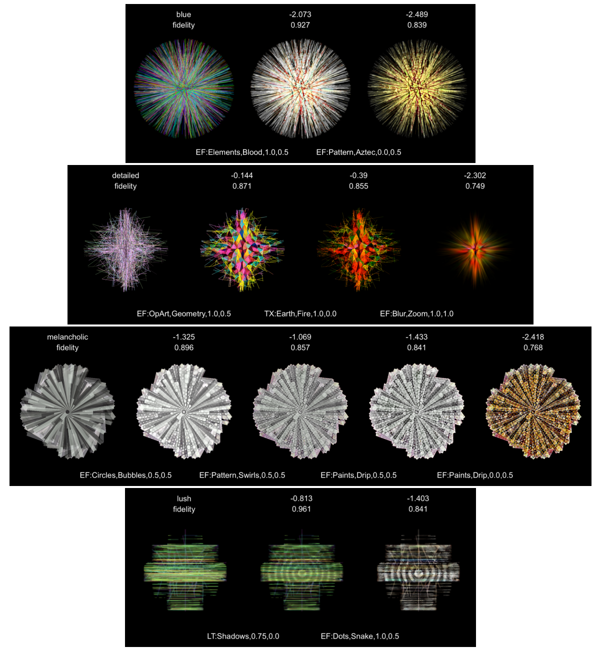

Figure 5: Action sequences which decrease cosine similarity with the words 'blue', 'detailed', 'melancholic' and 'lush'.

Discussion

The top 20 image editing actions for positive and negative movements with respect to 500 descriptive words, and the action sequences resulting from the experiments reported here unlock new functionality in the Art Done Quick app. Given that the app now has a good understanding of how its own tools effect images, users can ask it for a sequence of tool uses which make changes to an image, so that the new version looks more (or less) like a given descriptive word. We believe this reflects aspects of relatable computational creativity as discussed above, because the demonstrations given are exactly how a person could do them. We hope this new mode of interaction will provide education to users of Art Done Quick, so they are exposed to editing possibilities, as well as surprising inspiration when the app chains together a sequence of actions that achieves unexpected results.

Looking in further detail, we would describe the computational creativity exhibited here as relatable rather than pseudo-relatable, as there is no reason to imagine the software is a person. Further, with respect to knowable and unknowable aspects of the creative acts, the sequence of actions is clearly knowable, and the motivation to change an image in accordance with a description with each edit action is likewise knowable. However, given that CLIP is used, the software is not able to explain why a particular image looks like a description or not, so this element is unknowable. Moreover, the backstory for how Art Done Quick gained its tool-use skills is quite relatable as an exhaustive search over single-edit actions followed by trial and error with action sequences. While users might not be able to imagine trying 61,200 image edits in order to master the app, they should understand the process, and perhaps relate to it.

It is beyond the scope of this paper to address the issue of so-called 'AI slop' (a derogatory term for AI generated artefacts that are deemed unworthy (Kommers et al., 2026)). However, it seems clear that one reason such artefacts are derided is the lack of effort (human or AI) that goes into their construction, as put succinctly by the phrase: "why bother reading something no one was bothered enough to write?" (x.com/sulamoon/status/1724888928866476417). AI systems like Art Done Quick which demonstrate tool use may encourage users to engage more with artworks they are making, rather than merely prompting for them, and we plan an investigation of relatable computational creativity tools in this context. Moreover, we hope that generalising text-2-tool-use from casual creators to creativity support tools, this mode for human-computer interaction may benefit professional creators whose creations are used (often without permission or copyright) to train the deep learning models which end-users employ to bypass their creative process.

Relatability is not the only consideration for human-computer interaction with creative AI software and casual creators in particular. Other factors for study include how much trust users have in the software, how easy/fun it is to use, its ethical deployment in various communities and how empowered people feel while using it (or watching it create autonomously). However, as argued above, the volume of literature on the simulation of human-like creative behaviours would indicate that relatability is of particular interest to computational creativity researchers. This new umbrella term colates and highlights decades of work in computational creativity research, and could act as a new axis for development and evaluation of generative AI systems.

Conclusions and Future Work

We hope to have made the case that the notion of relatable computational creativity suitably encapsulates a body of work covering aspects such as motivations, explainability, embodiment, personhood and evaluation. We have highlighted aspects of this term for describing relatable AI systems, including the interaction mode, opaqueness of processing, role-playing situations, explanations versus demonstrations and engineering backstory. We introduced new functionality into the Art Done Quick application, so that users can be given tool-use help and inspiration from the app in a relatable way.

While we have given the app 20 good single-action edits for each descriptive word and in both directions, we only provided a single sequence of actions for the same purpose. Hence, we plan to scale up the action sequences, so the software has an even greater understanding of its tools and higher potential for surprise. Having shown that Art Done Quick can find tool-uses to achieve requested image changes that CLIP deems appropriate, we can use this for enhanced functionality in the app, and we plan a prompting interface for this. A next step will then be to perform user testing to see if people are satisfied with the interaction style and the resulting images produced. A preliminary look at some of the action sequences generated in the above experiments indicates that people may agree with CLIP's assessment of movement for highly visual terms like 'bright' or 'crystalline' but disagree when the required description requires more interpretation, such as 'discordant' or 'dramatic'.

We hope to use the platform to raise and answer questions about relatable computational creativity in a practical way, for instance in this case, whether it is still possible to relate to the creative acts of a generative AI system if you disagree with its assessment of the outputs it produces. Addressing such questions will enable us to expand the general idea of relatability, with the ultimate aim to produce a suitable taxonomy and grounded study of how people relate to generative AI systems and any benefits this may have, and a typology of relatable AI systems in general.

To do this, we plan to undertake a more thorough survey of the computational creativity literature to identify concepts, commonalities, best practice and pitfalls in the building ad evaluation of relatable computational creativity systems. We will also draw upon other studies of performative relationships in human-computer interaction, such as HCI entanglement (Frauenberger, 2019). As the theory of relatable computational creativity advances, we plan to distill it into more sophisticated implementations of Art Done Quick. In particular, in addition to the software reacting to user's requests for image enhancements, we will employ ideas from *creativity theatre* Colton et al. (2020b) to enable the software to perform proactively with its tools, in ways which challenge assumptions of creativity in software, and – we hope – actively encourage, enthuse and inspire users to create.

Acknowledgements

We would like to thank the anonymous reviewers for their thoughtful and helpful comments and suggestions.

References

Abusitta, A.; Li, M. Q.; and Fung, B. C. 2024. Survey on explainable AI: Techniques, challenges and open issues. *Expert Systems with Applications* 255.

Acharya, D. B.; Kuppan, K.; and Divya, B. 2025. Agentic ai: Autonomous intelligence for complex goals – A comprehensive survey. *IEEE Access* 13.

Arias, G.; Baldrich, R.; and Vanrell, M. 2024. Color in visual-language models: CLIP deficiencies. In *Proceedings of the Color and Imaging Conference*.

Chen, X.; Zhao, A.; Heming, X.; Lu, X.; Wang, H.; Chen, Y.; Wang, J.; Li, W.; and Shen, X. 2025. Reasoning beyond language: A comprehensive survey on latent chain-of-thought reasoning. arXiv preprint 2505.16782.

Cohen, P. 2017. Harold Cohen and AARON. *AI Magazine* 37:63–66.

Colton, S., and Banar, B. 2023. Automatically adding to artistic cultures. In *Proc. 12th Int. Conf. on Artificial Intelligence in Music, Sound, Art and Design (EvoMUSART)*.

Colton, S., and Ventura, D. 2014. You Can't Know my Mind: A festival of computational creativity. In *Proc. of the 5th International Conference on Computational Creativity*.

Colton, S.; Cook, M.; Hepworth, R.; and Pease, A. 2014. On acid drops and teardrops: Observer issues in computational creativity. In *Proc. AISB symposium on AI and Philosophy*.

Colton, S.; McCormack, J.; Berns, S.; Petrovskaya, E.; and Cook, M. 2020a. Adapting and enhancing evolutionary art for casual creation. In *Proc. 9th Int. Conf. on Computational Intelligence in Music, Sound, Art and Design*.

Colton, S.; McCormack, J.; Cook, M.; and Berns, S. 2020b. Creativity theatre for demonstrable computational creativity. In *Proc. 11th Int. Conf. on Computational Creativity*.

Colton, S.; Pease, A.; Guckelsberger, C.; Llano, M. T.; McCormack, J.; and Cook, M. 2020c. On the machine condition and its creative expression. In *Proceedings of the 11th International Conference on Computational Creativity*.

Colton, S.; Charnley, J.; and Pease, A. 2011. Computational creativity theory: The FACE and IDEA descriptive models. In *Proc. 2nd Int. Conf. on Computational Creativity*.

Colton, S.; Pease, A.; and Saunders, R. 2018. Issues of authenticity in autonomously creative systems. In *Proc. 9th International Conference on Computational Creativity*.

Colton, S. 2008. Creativity versus the perception of creativity in computational systems. In *Proceedings of the AAAI Spring Symposium on Creative Intelligent Systems*.

Colton, S. 2012. The Painting Fool: Stories from building an automated painter. In McCormack, J., and d'Inverno, M., eds., *Computers and creativity*. Springer. 3–38.

Colton, S. 2020. Casual creation, curation, captioning, clustering and crossover in the Art Done Quick app. *Proc. Workshop on Casual Creators at ICCC*.

Compton, K., and Mateas, M. 2015. Casual creators. In *Proc. of the 6th Int. Conf. on Computational Creativity*.

Compton, K. 2019. *Casual Creators: Defining a Genre of Autotelic Creativity Support Systems*. Ph.D. Dissertation, University of California at Santa Cruz.

Cook, M., and Colton, S. 2018. Redesigning computationally creative systems for continuous creation. In *Proc. of the 9th International Conference on Computational Creativity*.

Cook, M.; Colton, S.; Pease, A.; and Llano, T. 2019. Framing in computational creativity: A survey and taxonomy. In *Proc. 10th Int. Conf. on Computational Creativity*.

Decety, J., and Grèzes, J. 2006. The power of simulation: Imagining one's own and other's behavior. *Brain research* 1079:4–14.

Fink, A.; Koschutnig, K.; Benedek, M.; Reishofer, G.; Ischebeck, A.; Weiss, E.; and Ebner, F. 2012. Stimulating creativity via the exposure to other people's ideas. *Human Brain Mapping* 33(11):2603–2610.

Frauenberger, C. 2019. Entanglement HCI: The Next Wave? ACM Trans. on Computer-Human Interaction 27(1), 1-27.

Grace, K., and Maher, M. L. 2014. What to expect when you're expecting: The role of unexpectedness in computationally evaluating creativity. In *Proceedings of the 5th International Conference on Computational Creativity*.

Guckelsberger, C.; Kantosalo, A.; Negrete-Yankelevich, S.; and Takala, T. 2021. Embodiment and computational creativity. *Proc. 12th Int. Conf. on Computational Creativity*.

Guckelsberger, C.; Salge, C.; and Colton, S. 2017. Addressing the " Why?" in computational creativity: A non-anthropocentric, minimal model of intentional creative agency. *Proc. 8th Int. Conf. on Computational Creativity*.

Guckelsberger, C. 2020. *Intrinsic Motivation in Computational Creativity Applied to Videogames*. Ph.D. Dissertation, School of Electronic Engineering and Computer Science, Queen Mary University of London.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Houston, J. 2009. *Optic Nerve: Perceptual Art of the 1960s*. Merrell Publishers.

Jing, Y.; Yang, Y.; Feng, Z.; Ye, J.; Yu, Y.; and Song, M. 2020. Neural style transfer: A review. *IEEE Transactions on Visualization and Computer Graphics* 26(11).

Katz, J. H.; Mann, T. C.; Shen, X.; Goncalo, J. A.; and Ferguson, M. J. 2022. Implicit impressions of creative people: Creativity evaluation in a stigmatized domain. *Org. Beh. & Human Decision Processes* 169:104116.

Kobayashi, A. 2020. Other/otherness. In *International Encyclopedia of Human Geography*. Elsevier. 25–31.

Kommers, C.; Duede, E.; Gordon, J.; Holtzman, A.; McNulty, T.; Stewart, S.; Thomas, L.; Jean So, R.; and Long, H. 2026. Why slop matters. ACM AI Letters 1(1), 1-6.

Krzeczkowska, A.; El-Hage, J.; Colton, S.; and Clark, S. 2010. Automated collage generation - with intent. In *Proc. 1st International Conference on Computational Creativity*.

Lehman, J., and Stanley, K. 2008. Exploiting open-endedness to solve problems through the search for novelty. In *Proc. 11th International Conference on Artificial Life (Alife XI)*.

Lin, T.; Wang, Y.; Liu, X.; and Qiu, X. 2022. A survey of transformers. *AI Open* 3:111–132.

Lindemeier, T. 2018. *e-David : Non-Photorealistic Rendering using a Robot and Visual Feedback*. Ph.D. Dissertation, Universität Konstanz, Konstanz.

Llano, M.; d'Inverno, M.; Yee-King, M.; McCormack, J.; Ilsar, A.; Pease, A.; and Colton, S. 2020. Explainable computational creativity. In *Proceedings of the 11th International Conference on Computational Creativity*.

Macedo, L., and Cardoso, A. 2017. A contrast-based computational model of surprise and its applications. *Topics in Cognitive Science* 11(1).

Maslej, M.; Oatley, K.; and Mar, R. 2017. Creating fictional characters: The role of experience, personality,and social processes. *Psychology of Aesthetics, Creativity, and the Arts* 11(4):487–499.

Moruzzi, C. 2022. The (artificial) physicality of creativity: How embodiment influences perceptions of creativity. *Proc. 13th Int. Conf. on Computational Creativity*.

Pease, A.; Colton, S.; and Banar, B. 2023. On the notion of creative personhood. In *Proceedings of the 14th International Conference on Computational Creativity*.

Petrovskaya, E.; Deterding, S.; and Colton, S. 2020. Casual creators in the wild: A typology of commercial generative creativity support tools. In *Proceedings of the 11th International Conference on Computational Creativity*.

Radford, A.; Kim, J.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; Krueger, G.; and Sutskever, I. 2021. Learning transferable visual models from natural language supervision. In *Proc. of the International Conference on Machine Learning*.

Saunders, R., and Bown, O. 2015. Computational social creativity. *Artificial Life* 21(3):366–378.

Saunders, R. 2001. *Curious design agents and artificial creativity*. Ph.D. Dissertation, University of Sydney.

Tresset, P., and Fol Leymarie, F. 2012. Sketches by Paul the robot. In *Proc. of the 8th Annual Symposium on Computational Aesthetics in Graphics, Visualization, & Imaging*.

Vasu, P.; Pouransari, H.; Faghri, F.; Vemulapalli, R.; and Tuzel, O. 2024. Mobileclip: Fast image-text models through multi-modal reinforced training. *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition*.

Ventura, D. 2016. Mere generation: Essential barometer or dated concept? In *Proceedings of the 7th International Conference on Computational Creativity*.

Yang, L.; Zhang, Z.; Song, Y.; Hong, S.; Xu, R.; Zhao, Y.; Zhang, .; Cui, B.; and Yang, M. 2024. Diffusion models: A comprehensive survey of methods and applications. *ACM Computing Surveys* 56(4).

Yannakakis, G. N., and Liapis, A. 2016. Searching for surprise. *Proc. 7th Int. Conf. on Computational Creativity*.