The Spectrum of Unpredictability and its Relation to Creative Autonomy

Iikka Hauhio, **Anna Kantosalo**, **Simo Linkola**, and **Hannu Toivonen**
Department of Computer Science
University of Helsinki, Finland
{first.last}@helsinki.fi

Abstract

Recent popularity of generative AI tools has sparked discussion on how the unpredictability of the tools affects the creativity of the human and the AI program alike, as unpredictability prevents the human user from fully controlling the output. We present a framework for categorizing unpredictability on four different dimensions and analyze the types of unpredictability found in generative AI tools. We also describe the relationship between unpredictability, uncontrollability, and Jennings' creative autonomy. We conclude that while unpredictability does not on its own imply creative autonomy, it could be used as a central condition for it, if accompanied by other conditions.

Introduction

The recent popularity of generative and creative artificial intelligence (AI) has raised the relationship between AI and creativity to common debate. For example, in a recent decision, the United States Copyright Office determined that in certain situations, the user of an AI image generation tool is not considered the author of the work for copyright purposes, because the tool works in an unpredictable manner.[1]

Many of today's generative AI systems are unpredictable in various respects and to varying degrees. If the unpredictability of the system rules out the user's (full) authorship of the generated results, who or what can be attributed with creativity when the end result itself is considered creative, i.e., novel and valuable (Runco and Jaeger 2012)? Is it a reasonable argument that the program must in that case have committed creative acts?

In this paper, we present a categorization for unpredictability that can be used to analyze different scenarios in which unpredictability can affect the creativity of the system. We argue that unpredictability may help to characterize the creative autonomy of the system, defined as "the system's freedom to pursue a course independent of its programmer's or operator's intentions" (Jennings 2010). Unpredictability implies that the human user does not have complete control over the system, which is a requirement for creative autonomy of the system.

Throughout this paper, we assume that one is assessing a creative system. We use language models and image generators as example tools without making claims about the creativity of any specific tools for any specific tasks. Rather, the arguments we present are philosophical in nature, asking the following question: assuming that the outputs of a system are creative, how does its possible unpredictability affect our judgement of the creative role and autonomy of the system?

The rest of the paper is organized as follows. We first present a definition for unpredictability and categorize different types of unpredictability. We then analyze unpredictability of concurrent generative AI programs. Finally, we present an argument that connects unpredictability to uncontrollability and thus Jennings' creative autonomy (Jennings 2010).

Unpredictability

In this paper, we define *unpredictability* as the inability of an *observer* (e.g. the creator of a generative program or its user) to determine the generative *outcome of a program* given a specific *input*. The observer can be seen as an entity that holds a certain amount of information about the program through knowledge of its internal workings or holistic observation of the program at work. This suggests that unpredictability from the point of view of a user may be influenced by experience and thus has an element of time to it and links it to other experiential properties of generative systems, such as surprise (e.g. (Grace et al. 2015)). This makes unpredictability a meaningful concept for evaluation of computationally creative systems that allows us to compare systems and link it to the discussion of meaningfully assigning autonomy to an AI.

Unpredictability is not a characteristic that uniformly covers the whole output of a system. Rather, it is a question of perspective. If the user prompts a language model to produce a poem, it usually is predictable that the output is, or resembles, a poem, while many details about the structure and word choices might be unpredictable. In the case of generative systems, it is important to define the extent of unpredictability; in this paper, we call the (unpredictable) features of interest the *outcome* of the system. The outcomes

[1] "Rather than a tool that [the user] controlled and guided to reach her desired image, Midjourney generates images in an unpredictable way. Accordingly, Midjourney users are not the "authors" for copyright purposes of the images the technology generates." (Kasunic 2023)

in our case are features of the artifacts the tools produce, i.e., features that the user might want to control but cannot due to unpredictability. These features can be concrete, such as the exact colors used by an AI image generator, or more abstract, such as the mood expressed by a generated poem.

We define unpredictability with respect to the process, the outcome, and the observer, and we will similarly categorize the different types of unpredictability based these three dimensions: (1) the *cause* of the unpredictability, i.e., what kind of process causes the outcome to be unpredictable; (2) the *scope* of the unpredictability, i.e., what types of outcomes are unpredictable; and lastly (3) the *point-of-view* of the unpredictability, i.e., who is the observer that determines that the process is unpredictable. We also consider (4) the *duration* of the unpredictability, i.e., when the predictions are performed.

Causes of Unpredictability

We divide unpredictability to three categories based on the cause of unpredictability: stochastic (indeterministic), chaotic (deterministic), and mixed-cause unpredictability.

Stochastic unpredictability refers to indeterministic unpredictability that cannot be predicted by the observer. It is similar to Boden's *absolute unpredictability* (Boden 2004); however, our definition includes technically deterministic processes which cannot in practice be predicted, such as pseudo-random number generators initialized with unknown, randomized seeds. From the point of view of the observer, the processes in this category are random. If the outcome of a program is stochastically unpredictable, the outcome will change unpredictably each time the program is run.

Stochastic unpredictability can be more strong in some scenarios than others. Compare, for example, a fair dice roll and a weighted dice. Both of them contain some unpredictability: we cannot be completely sure what the result will be. However, in the latter case, one outcome is more likely than the others. In an extreme case, a weighted dice will almost always produce the same result, thus making it completely predictable. Thus, depending on the probability distribution, some cases of stochastic unpredictability are more unpredictable than others. The exact categorization of the subtypes of stochastic unpredictability is not in the scope of this paper.

Chaotic unpredictability refers to deterministic but chaotic processes. If a program is chaotically unpredictable, its output will change unpredictably each time it is run with a new input, but it will consistently provide the same output for the same input. This category includes pseudo-random numbers generated with a known seed. Neural networks that are too complex for humans to understand (cf. Burrell 2016) might also belong to this category. In Boden's terms, this type of unpredictability is called *butterfly unpredictability* (Boden 2004).

Mixed-cause unpredictability is a combination of both stochastic and chaotic unpredictability. In practice, many generative AI programs include both types. For example, a language model-based generator might first calculate the probability distribution for the next word using a complex neural network (chaotic unpredictability), and then sample a word from the distribution (stochastic unpredictability). If the outcome is the result of both stochastic and chaotic unpredictability, it can change to some degree each time the program is run with the same input while still retaining some properties between the outcomes.

Scope of Unpredictability

We call the "size" of the set of features of the output affected by the unpredictability the *scope* of the unpredictability. Next, we sketch different levels of unpredictability based on their scope. Note that the levels are not based on shallow, technical distances such as edit distance, but rather on their semantic distance. Here, we outline the idea, and a more exact characterization is left for future work.

Low-level unpredictability occurs when the unpredictable variation affects small details or minor choices in the output, e.g., the exact word choices of a poem generator or the exact colors produced by an image synthesis model cannot be predicted.

Middle-level unpredictability refers to unpredictability of broad details and major choices in the output. In a poem generator's output, this might mean features such as the symbols used, or the meter followed. In an image synthesis program, middle-level features might be the objects included in the scene, the layout of the image and the art style used.

High-level unpredictability refers to even more abstract features, such as the topics included in the work. At the highest level, even the artifact class itself could be unpredictable.

Point-of-view of Unpredictability

Unpredictability is defined with respect to an observer for whom the process is unpredictable. We propose the following categorization to world- and user-unpredictability, which can be compared to Boden's categorization of creativity to H-creativity and P-creativity (Boden 2004). Boden argues that if the purpose is to evaluate the capability of an individual — or a program — to be creative, then P-creativity and what we call user-unpredictability are more interesting concepts than H-creativity and world-unpredictability.

World-unpredictability refers to the situation in which no one can predict the outcome of the process. By definition, this includes all stochastic programs, but it might also include some chaotic programs if they are sufficiently complex for any human to understand (cf. Burrell 2016).

User-unpredictability refers to the situation in which the humans who choose the input to the program cannot predict the output. While weaker than world-unpredictability, user-unpredictability is still enough to establish the control the user has over the output. If the user cannot predict the outcome of the process, they cannot reliably control it (see the chapter below for elaboration of unpredictability and control).

Akin to user-unpredictability, it is also possible to define concepts such as *programmer-unpredictability* and *audience-unpredictability*, if needed.

Changes in Subjective Unpredictability

In addition to the *how*, *what*, and *who* of the previous categorizations, we can also ask *when* the program is unpredictable for a particular observer.

Permanent unpredictability lasts forever. By definition, this includes all stochastic unpredictability, but some sufficiently complex chaotic processes might also belong to this category, at least if we only consider humans as possible observers.

Temporary unpredictability can be overcome, causing the process to become predictable in time. For example, a deterministic program becomes predictable for a certain input after the first time it is run, as all the subsequent runs will produce the same result. Likewise, a process presumed to be chaotic can become predictable after it is understood better.

Unpredictability in Generative AI Programs

Large neural networks used for generative tasks, such as GPT-3 (Radford et al. 2019) and Stable Diffusion (Rombach et al. 2022), contain billions of parameters and are often regarded as black boxes due to their unexplainability. Their intrinsic complexity makes it impossible for a human to fully comprehend their operation (Burrell 2016), which implies they contain unpredictability.

We argue that following our categorization of unpredictability, most concurrent AI tools contain *mixed-cause*, *low/middle-level user*-unpredictability. Complex AI models behave both predictably and unpredictability (Ganguli et al. 2022) and contain both chaotic parts (such as deterministic neural networks) and stochastic parts (such as token-sampling in language models and random noise in image synthesis models). Since state-of-the-art models are broadly speaking quite good at following instructions specified in the prompt (Radford et al. 2019; Rombach et al. 2022), they are predictable and controllable at high-level, but not necessary at low- and middle-levels. While they contain some world-unpredictable parts (such as those which are completely stochastic), they also contain parts which become more predictable as the user gains intuition over the model's behavior, causing the model to be more unpredictable to some users than others.

$$P \ni x \xrightarrow{\text{predict } f} Q \ni f(x)$$

$$Q \ni f(x) \xrightarrow{\text{control } f} P \ni x$$

Figure 1: The difference between predicting (determining the outcome caused by the input property) and controlling (determining the input property that causes the desired outcome).

Unpredictability and Uncontrollability

To control a generative program, the user must be able to predict the correct *inputs* P that will cause the desired *outcome* Q. For example, if the user wants an image synthesis program to use a certain color (e.g., Q is the set of images with red color), they must determine which prompts cause that color to be generated (e.g., P is the set of prompts that contain the string "red"). The ability to control the program corresponds thus to the ability to calculate or estimate the inverse of the program.

Predictability, on the other hand, is about determining the outcome Q given an input P. Logically, the ability to control and the ability to predict are inverses of each other, and separate from each other. The difference is explained in mathematical notation in Figure 1. However, we argue that in practice, unpredictability implies uncontrollability.

If the program was controllable but unpredictable, it would mean that its inverse is predictable. If the program was chaotically unpredictable, its inverse would not be chaotic. If the program was stochastically unpredictable, its inverse would not be stochastic. We argue that this kind of situation is rare in the context and generative AI and creative programs overall, but we'll leave the proof for future work. In the rest of the paper, we assume that unpredictability does imply uncontrollability.

Note that the reverse is not true: a predictable program can be uncontrollable. For example, a program that always produces the same output is predictable and uncontrollable: for any desired outcome Q which is not the outcome the program produces, there exist no solutions for P.

Creative Autonomy

As discussed above, unpredictability makes it impossible for the user to completely control the AI tool's operation. Unpredictability is therefore related to Jennings' *creative autonomy* (Jennings 2010): What appears as unpredictable behaviour to the user might be explained by creative autonomy of the system. We seek to use unpredictability as a tool to characterize the creative autonomy of systems, especially of black box AI generators.

Jennings gives three criteria for creative autonomy: *autonomous evaluation*, *autonomous change of standards*, and *non-randomness* (Jennings 2010). Autonomous evaluation allows the system to observe the quality of its own work and thereby improve its operations. Autonomous change,

in turn, allows the system to adjust its own standards and goals. Autonomous evaluation and change could be trivially achieved with random behavior, but the third criterion rules out fully random behaviour. Jennings explicitly allows for randomness in the processes, and many creative systems have stochastic components — they just shouldn't be fully random.

The relationship between unpredictability and creative autonomy is not one-to-one. Not all unpredictability implies creative autonomy: fully random behaviour would be unpredictable but not autonomously creative. Also, not all unpredictable generative behavior is creative. Vice versa, it can be argued that some predictable processes do have creative autonomy despite their predictability, since being autonomous does not entail being unpredictable. For example, many human artists have a very constant style or paradigm they follow, while retaining creative autonomy.

It is clear that deterministic unpredictability does not necessarily entail creative autonomy, either. Consider fractal images such as the Mandelbrot set image. These images are deterministic but chaotic, and it is very hard to predict what a certain "deeply zoomed" region of the image looks like without solving the equation for the points in that region. However, they are also completely static, and no evaluation or change occurs when calculating the equation. The fractal equation does thus not have creative autonomy, although it can potentially produce novel and valuable images when solved for yet unvisited regions of the coordinate plane.

Despite unpredictability not directly implying creative autonomy, we argue that unpredictability could be used as a *condition* for it in a yet-to-be formulated framework for evaluating unpredictable programs: if the evaluations and changes that occur during the program's execution are unpredictable, they cannot be controlled by the user and are thus autonomous, assuming they are not fully random, i.e., the unpredictability should not be only stochastic.

Unpredictability could be used to show that the user is incapable of controlling a creative program, in order to provide arguments for the program's creative autonomy. To prove this for a single user during their use of the program, the type of unpredictability used as a condition for creative autonomy can be temporary user-unpredictability instead of stronger forms of unpredictability such as permanent world-unpredictability, although this would make the perception of creative autonomy subjective allow it to change over time. We leave the debate of whether this is acceptable or not and how unpredictability shapes the artist's perception of their own role and agency to further research.

Conclusions

Unpredictability is an important property of many generative AI programs and has implications to their creativity, since it limits the ability of the user to control the operation of the AI programs. We presented a framework for categorizing different types of unpredictability based on the *how*, *what*, *who*, and *when*: the causes, scopes, observers, and the change of subjective unpredictability. These categorizations can be used to characterize generative AI tools.

We discussed the relationship between unpredictability, uncontrollability, and creative autonomy (Jennings 2010). Unpredictability implies uncontrollability, which is a requirement for creative autonomy. While unpredictability does not imply creative autonomy, it could be used as a condition in a larger framework intended for determining and analyzing creative autonomy in generative AI programs. Further research should be conducted to determine a sufficient set of additional conditions to be used alongside unpredictability.

Unpredictability of complex generative systems, and the lack of control it implies, shows that it can be difficult to attribute creativity to one party only, be it the user, the developer, or the system. While the US Copyright Office's decision to deny authorship of the human who used an AI image synthesis tool is probably justified, this does not mean that the tool was the author. We argue that in this case, there simply is no single author. This does not mean, however, that there is no creativity in the process: the creativity is just not controlled by any one stakeholder. This implies that, although not necessarily autonomously creative, unpredictable programs do nevertheless play a significant role in the creative process.

Author Contributions

The idea for the paper came from IH and it was mostly written by IH. AK, SL and HT supported the concept development and edited the paper.

Acknowledgments

IH is funded by the doctoral programme in Computer Science at the University of Helsinki. SL is funded by Academy of Finland grant 328729 (CACDAR). AK works at Siili Solutions Oyj; this paper was not a part of a work assignment.

References

Boden, M. A. 2004. *The creative mind: Myths and mechanisms*. Routledge.

Burrell, J. 2016. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big data & society* 3(1).

Ganguli, D.; Hernandez, D.; Lovitt, L.; Askell, A.; Bai, Y.; Chen, A.; Conerly, T.; Dassarma, N.; Drain, D.; Elhage, N.; et al. 2022. Predictability and surprise in large generative models. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, 1747–1764.

Grace, K.; Maher, M. L.; Fisher, D.; and Brady, K. 2015. Data-intensive evaluation of design creativity using novelty, value, and surprise. *International Journal of Design Creativity and Innovation* 3(3-4):125–147.

Jennings, K. E. 2010. Developing creativity: Artificial barriers in artificial intelligence. *Minds and Machines* 20(4):489–501.

Kasunic, R. J. 2023. Re: Zarya of the Dawn (Registration VAu001480196).

Radford, A.; Wu, J.; Child, R.; Luan, D.; Amodei, D.; Sutskever, I.; et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog* 1(8):9. Viitattu: 16.6.2022.

Rombach, R.; Blattmann, A.; Lorenz, D.; Esser, P.; and Ommer, B. 2022. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 10684–10695.

Runco, M. A., and Jaeger, G. J. 2012. The standard definition of creativity. *Creativity research journal* 24(1):92–96.