Why a Formal Definition of Computational Creativity Might be Elusive

Michael Guerzhoy guerzhoy@cs.toronto.edu

We propose some plausible properties of computational creativity/creativity by Turing Machines (TMs), and point to theoretical and empirical considerations that suggest that a formal definition of computational creativity might be elusive.

We take **creativity** to be a Wittgensteinian family resemblance concept, anchored in humans' experiences of creativity in themselves and others. We posit that TMs exhibit **computational creativity** if they are judged to be creative by human judges, and ask whether a formal definition can be reassembled from human judgements. Importantly for our purposes, computational creativity is a property of the TM and the past data the TM has seen, not of a single artifact produced by the TM. That is because context and past experience are key to assessing creativity, and most would argue process is as well (if a TM only produces one single artifact every time, it is not creative.) We argue the following are plausible:

- A creative TM exhibits **novelty** through a process that is neither a straightforward algorithm nor a completely-random process
- A creative TM's process is intricate rather than straightforward
- Generally, the output of a creative TM depends on a large corpus of past data
- There are multiple and only loosely-related (if at all) ways for a TM to exhibit creativity

Empirical analogy Good SAT solvers exhibit the properties above, and some might themselves be considered creative. However, there is no formal understanding of what makes a good SAT solver [GV2020]. This is in part due to the fact that SAT solvers are good *for particular distributions of inputs but not others*, and partly because the algorithms are intricate.

If one takes seriously the idea that SAT solvers can be creative, defining computational creativity is at least as hard as giving a theoretical account of what makes a good SAT solver, which is an open problem.

Theoretical considerations

- White-box examples of creative TMs (even when available) may not be enough: Turing machine equivalence is undecidable in general, and even with restrictions that make the task decidable, the complexity of comparing TMs may be prohibitive.
- A large number of creative TM exemplars may be needed to get a good understanding of the family resemblance.

Is this just Rice's Theorem [Ric1953] in disguise? Not quite. It is true any non-trivial property, including creativity, would in general be undecidable. But our argument is not about the hypothetical pathological cases that make the problem undecidable. Rather, we argue that the field may be faced with a situation where to get at a definition of creativity, one needs to compare complex TMs that are augmented by large corpora of past data, and that this is a hard problem even if we restrict ourselves to TMs that are not pathological.

Is this specifically about computational creativity? Largely yes. The argument does not apply as strongly to cognitive capacities where the outputs are defined more clearly. It is about general intelligence insofar as it subsumes creativity.

Does Guerzhoy's argument apply here? No. Guerzhoy [Gue2026] argues that there are barriers to complexity-theoretic proofs that achieving AGI via machine learning is impossible [vRGA$^+$2024]; we make no analogous claim that a formal definition of computational creativity is *provably* elusive. Rather, we point to theoretical and empirical considerations that suggest that it might be elusive.

Limitations We outlined a scenario, which we find somewhat plausible, in which a formal definition of computational creativity might be elusive. Even accepting our premises, it is possible that the specific barriers we point to can be overcome because of the shape of the distribution of creative TMs.

References

[Gue2026] Michael Guerzhoy. Barriers to Complexity-Theoretic Proofs that "AGI" Using Machine Learning is Impossible. *Computational Brain & Behavior*, 2026.

[GV2020] Vijay Ganesh and Moshe Y Vardi. On the Unreasonable Effectiveness of SAT Solvers., 2020.

[Ric1953] Henry Gordon Rice. Classes of recursively enumerable sets and their decision problems. *Transactions of the American Mathematical Society*, 74(2):358–366, 1953.

[vRGA+2024] Iris van Rooij, Olivia Guest, Federico Adolfi, Ronald De Haan, Antonina Kolokolova, and Patricia Rich. Reclaiming AI as a theoretical tool for cognitive science. *Computational Brain & Behavior*, 7(4):616–636, 2024.