Against Machine Originality: On Lovelace, LLMs, and ChatGPT

Aaron Fung, Simon Fraser University

Abstract

This paper was originally written for Dr. Kino Zhao's PHIL 302 course *Topics in Epistemology and Metaphysics: Philosophy of Machine Learning*. The assignment asked students to a write a term paper critically engaging with various course texts on artificial intelligence, computation, and machine learning. The paper uses APA citation style.

In this paper, I argue in defence of Lady Lovelace's objection that machines have no pretensions to originate anything. I begin by examining Alan Turing's response to this objection, focusing on his appeal to machine unpredictability and the child-machine thought experiment, both of which are intended to demonstrate that machines can, in principle, possess pretensions to originate. I argue, however, that machine "learning" ultimately remains derivative of human-provided structures. Using modern large language models (LLMs) like ChatGPT as a contemporary case study, I conclude that apparent machine originality is better understood as an extension of human design rather than a genuine act of originality.

Introduction

In "Computing Machinery and Intelligence", Alan Turing argues against Lady Lovelace's objection from machine originality, in which she claims that machines have no pretensions to originate anything (Turing, 1950, p. 450). For one, Lovelace claims machines cannot originate because they merely perform instructions that humans give them and in turn, cannot necessarily think for themselves. In response, Turing challenges this notion of original work by (1) asserting that machines often take their programmers by surprise, suggesting that their behaviour is not fully predetermined, and (2) introducing the child-machine thought experiment, in which a hypothetical machine can continuously learn and develop through experience, similar to a human child. Thus, Turing concludes that machines can, in principle, possess pretensions to originate.

However, Turing's response faces an issue: even if machines appear to learn and develop unpredictably, it remains that their outputs nonetheless depend on human-provided instructions, thereby calling their originality into question. In this paper, I defend Lovelace's objection and argue that machines do not have pretensions to originate anything, since Turing's reply does not successfully address her concern. To demonstrate so, this paper will proceed as follows:

(1) A reconstruction of Turing's argument on originality;
(2) An objection defending Lovelace's position, arguing that machine "learning" remains derivative of human-provided structures;
(3) A potential counterargument posing that human-child learning is similarly constrained; and
(4) My response against the counterargument, maintaining that the constraints on machine learning and human-child learning are disanalogous.

My conclusion is that Turing's child-machine thought experiment does not adequately address the issue of machine originality.

Turing's Defence on Originality

Firstly, Turing argues that machines often take their programmers by surprise with great frequency, which suggests that their behaviour is not fully predetermined (Turing, 1950, p. 450). To support this claim, he clarifies how human originality is typically understood: although creative acts are commonly treated as independent or spontaneous, many human instances of so-called original work arise from prior teaching or the application of general principles rather than from complete autonomy (Turing, 1950, p. 451). That is to say, what is generally considered original often reflects the influence of past learning, rather than entirely independent creations. In turn, Turing applies this sort of constrained notion of human originality onto machines, asserting that machines often behave in ways that people do not anticipate, that is, their outputs are not always predictable based on initial instructions (Turing, 1950, p. 450).

Furthermore, to address the concern that such surprises merely reflect a creative act of the programmer, Turing notes that the recognition of surprise involves creativity regardless of whether the unexpected event comes from a human or machine (Turing, 1950, p. 451). This is because identifying something unexpected depends on the observer's own interpretive abilities, not on the source of the event itself. Consequently, given that machines generate outputs beyond what the people initially expect, it follows that their behaviour is not fully

predetermined. Thus, Turing claims that machines often take their programmers by surprise with great frequency (Turing, 1950, p. 450).

Secondly, as a further response to Lovelace's objection, Turing introduces the child-machine thought experiment, in which a hypothetical machine can learn and develop through experience in such a manner that is comparable to a human child (Turing, 1950, p. 455). For one, he suggests that it is perhaps misguided to attempt to program a fully formed adult mind right from the start. Instead, it would be feasible to begin with a machine that more so resembles a human child's mind, allowing it to be further educated via experience. According to Turing, the child-machine would be composed of three central components:

(a) The initial state: what the machine is like at the beginning;
(b) Education: the training that is then given to it; and
(c) Other experience: additional interactions, distinct from education, that help further shape it.

Turing emphasizes that each component plays a distinct role in shaping the machine's eventual behaviour (Turing, 1950, p. 456). The initial state determines the starting structure of the machine, similar to the natural capacities that a human child is born with. Next, education refers to the deliberate training imposed on the machine, including punishments, rewards, and instructions that guide its development over time; again, similar to the teaching process we typically apply to human children (Turing, 1950, p. 457). Finally, other experience encompasses interactions that are not part of the deliberate training imposed in education, but nonetheless influence how the machine forms new patterns. This is analogous to how a human child acquires habits or skills through continual exposure to their environments. Taking these three components into account, each stage allows the hypothetical child-machine to undergo changes through continual experience, thus eventually resulting in behaviour not initially present in its initial programming (i.e., in the initial state).

This child-machine structure is used to demonstrate that learning and development in a machine can give rise to patterns of behaviour that we do not explicitly anticipate (Turing, 1950, p. 450; p. 456). In other words, if a machine is capable of adapting itself through education and experience, it follows that resulting outputs and behaviours can develop beyond its initially provided instructions. Given the child-machine thought experiment, Turing maintains that a machine may, in principle, exhibit patterns of behaviour that appear original, thereby challenging Lovelace's claim that machines have no pretensions to originate anything and merely perform instructions that humans give them.

Taking both premises into account, Turing's argument seeks to reject Lovelace's objection that machines lack pretensions to originate anything. Since machines can behave in ways that their programmers do not expect (Turing, 1950, p. 450), and a child-machine could, in principle, adapt its behaviour through continual education and experience (Turing, 1950, p. 456), Turing maintains that learning in machines is not fixed by initial, human-provided instructions. Given this unpredictable and learned behaviour, he rejects Lovelace's objection from machine originality. Thus, Turing concludes that machines can, in principle, possess pretensions to originate.

Objection: Learning is derivative of human structures

Despite Turing's twofold defence, I argue that his reply is not entirely successful, more specifically, within the child-machine thought experiment. For one, even if a machine *appears* to learn or develop its behaviours over time, such "learning" remains dependent on the instructions, structures, and data that humans initially provide it. That is to say, the machine does not independently acquire information and knowledge from its environment in the same way that a human child does. Instead, a hypothetical child-machine is precisely restricted by the parameters, training materials, and feedback that its programmers deliberately select and decide on. In other words, the way in which the machine receives data, processes information, and has its behaviours punished and rewarded are all meticulously determined by human design. Consequently, this kind of "learning" that Turing describes does not demonstrate genuine capacity for autonomy, creativity, and most importantly, originality. Instead, it demonstrates the extent of deliberate human instruction, and not machine originality.

This issue is further supported by other philosophical accounts. To start, Gunderson argues that thinking cannot be fully understood by merely looking at a system's net results (Gunderson, 1964, p. 237), since what *appears* to be successful behaviour does not necessarily reveal that genuine thought is occurring. That is to say, we cannot infer thought or originality simply based on unexpected or surprising behaviour, as such outputs may be the result of mere imitation rather than genuine cognitive processes taking place. Applying this to Turing's defence, it suggests that even if a machine behaves in ways that seem unexpected, such behaviours alone do not indicate that the machine has originated anything. Additionally, Attah asserts that any purposes or intentions displayed by a language model (i.e., a machine) arise from its design context, since the particular functions determine the limited range of intentions that it can entertain (Attah, 2025, p. 19).

This suggests that a machine's apparent capacity to think or originate is ultimately bounded by human-defined functions. In essence, these further accounts illustrate that machines do not exhibit genuine originality: what appears to be genuine thinking or originality is better understood as deliberate human design rather than independently generated thought.

Perhaps a contemporary example would more clearly illustrate this issue of genuine originality persisting in Turing's argument. Consider modern large language models (LLMs) such as ChatGPT: they do not *independently* acquire information or learn and develop directly from their environment as a human child does. Instead, they depend entirely on deliberately curated human data and human-designed training processes. Although their outputs may appear creative or original, it remains the case that they are produced simply through existing human material and instructions. As a result, modern LLMs like ChatGPT illustrate the broader issue raised in the previous paragraphs: machine "learning" is bound to human instruction and consequently, any sort of originality or creativity is merely the result of said instructions rather than from the machine itself. Thus, I argue that Lovelace's objection that machines do not have pretensions to originate anything stands, as machine "originality" remains derivative of human structures.

Counterargument: Human-child learning is similarly constrained

One could object to my argument as follows: while it is apparent that most machines currently cannot physically interact with their environments in the exact same way that a human child can, the contrast between machine learning and human-child learning may not be as distinct as my objection suggests. More specifically, human children do not learn and develop in entirely independent and unrestricted manners either. Instead, their learning and development is necessarily shaped by biological limits, environmental exposure, and most importantly, human-provided instruction, such as from parents, family, teachers, and social and cultural practices. In this way, a human child's learning is also deliberately selected and decided on by humans, similar to how a child-machine does not necessarily choose the provided instructions that it receives. Therefore, if human-child learning is compatible with originality despite these constraints, then the mere fact that a machine can learn within the exact same parameters set by programmers does not, on its own, demonstrate that it cannot originate anything. Taking this into account, the dependence that a child-machine exhibits on initial, human-provided data does not successfully negate Turing's argument, since both human

and machine children would hypothetically develop under the same conditions. Thus, one may object to my claim that machine learning is derivative of human structures by asserting that human-child learning is similarly constrained as that of machines.

Response: Machine learning and human-child learning are disanalogous

In response to this potential counterargument, I argue that human-child and child-machine learning are disanalogous, because the two forms of learning significantly differ in their means of development. On the one hand, human children can revise their goals and intentions and shift their interests, in turn, *reframing* new and existing information in such ways that are not predetermined by others. While their learning is certainly guided by parents, family, teachers, and social and cultural practices, it remains the case that a great deal of their learning is self-guided and self-motivated, and not exhaustively fixed by instruction. On the other hand, a machine's learning remains fully determined by the particular design, parameters, and feedback that humans program into the system. Even when a machine *appears* to learn and develop on its own, the scope and limits through which it improves are nonetheless set in advance by humans. For instance, modern LLMs like ChatGPT cannot independently revise their goals, intentions, or interests: what may count as a "better" ChatGPT response or more appropriate guiding principles to follow remain entirely set by both human programmers and users. In essence, unlike a human child, a machine cannot originate new ways of learning for itself. Therefore, machine learning and human-child learning are disanalogous with respect to originality.

Conclusion

In this paper, I defended Lovelace's objection and argued that machines do not have pretensions to originate anything, since Turing's reply does not successfully address her concern. In particular, I presented the contemporary case of ChatGPT, demonstrating how machine "learning" remains derivative of human-provided structures rather than an independent source of thought and originality. To demonstrate so, this paper proceeded as follows:

(1) A reconstruction of Turing's argument on originality;
(2) An objection defending Lovelace's position, arguing that machine "learning" remains derivative of human-provided structures;
(3) A potential counterargument posing that human-child learning is similarly constrained; and

(4) My response against the counterargument, maintaining that the constraints on machine learning and human-child learning are disanalogous.

All things considered, while Turing's child machine thought experiment initially appears to offer grounds for machine originality, a closer analysis reveals that such learning processes merely act as an extension of human design rather than as an act of originality. My contemporary example of ChatGPT reinforces this conclusion, since its apparent originality nonetheless arises entirely from initial, human-provided instruction. As a result, Lovelace's objection stands: machines lack pretensions to originate anything.

References

Attah, N. O. (2025). Do language models lack communicative intentions? *Synthese*, *205*, Article 187. https://doi.org/10.1007/s11229-025-05022-6

Gunderson, K. (1964). The imitation game. *Mind*, *73*(290), 234–245. https://doi.org/10.1093/mind/LXXIII.290.234

Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, *59*(236), 433–460. http://dx.doi.org/10.1093/mind/LIX.236.433

By submitting this essay, I attest that it is my own work, completed in accordance with University regulations. I also give permission for the Student Learning Commons to publish all or part of my essay as an example of good writing in a particular course or discipline, or to provide models of specific writing techniques for use in teaching. This permission applies whether or not I win a prize and includes publication on the Simon Fraser University website or in the SLC Writing Contest Open Journal.

Available from: https://journals.lib.sfu.ca/index.php/slc-uwc