The Curious Case of Uncurious Creation

Lindsay Brainard

Penultimate Draft – forthcoming in *Inquiry*

Abstract: This paper seeks to answer the question: Can contemporary forms of artificial intelligence be creative? To answer this question, I consider three conditions that are commonly taken to be necessary for creativity. These are novelty, value, and agency. I argue that while contemporary AI models may have a claim to novelty and value, they cannot satisfy the kind of agency condition required for creativity. From this discussion, a new condition for creativity emerges. Creativity requires curiosity, a motivation to pursue epistemic goods. I argue that contemporary AI models do not satisfy this new condition. Because they lack both agency and curiosity, it is a mistake to attribute the same sort of creativity to AI that we prize in humans. Finally, I consider the question of whether these AI models stand to make human creativity in the arts and sciences obsolete, despite not being creative themselves. I argue, optimistically, that this is unlikely.

Keywords: Creativity, Artificial Intelligence, Curiosity, Aesthetics, Agency

1. Introduction[1]

To be human at the dawn of the artificial intelligence (AI) era is both thrilling and terrifying. We confront the emergence of these astounding technological advances with a mixture of awe (*Look what we've done!)* and existential fear (*What have we done?)*. One source of this existential fear is the worry that humans will soon be outdone by our silicon progeny. Perhaps we're well on our way to a reality in which anything we can do, AI can do better. What is our place in a world like that? What should we strive for if many of our grandest aspirations – achievements in the arts and sciences, for example– are attainable with less effort and in less time by machines? We reasonably worry that AI will take our jobs. But need we worry that it will take away more than that? Will it challenge the special status we've taken ourselves to have by besting us in the domains we cherish most?

In this paper, I examine one of the most celebrated aspects of humanity – our creativity – and ask whether it is something achievable by AI. To answer this question, I

[1] I am grateful to Ian Cruise, Marc Lange, Nathaniel Sharadin, Keshav Singh, and an anonymous reviewer for their helpful feedback on earlier drafts of this paper.

consider three central aspects of creativity – novelty, value, and agency – and explore the question of whether AI can realize these features. In examining these three features, I uncover a further indispensable aspect of creativity: that it is motivated in part by curiosity. I argue that AI – at least in its current forms – is incapable of curiosity, and therefore incapable of the sort of creativity we prize as one of our species' most distinctive and important capacities.

However, arriving at this conclusion may not fully assuage the worries I mentioned above. Even though AI may not be creative, it can certainly still create. And this uncurious creation is a curious case. In the final part of this paper, I explore the lingering worry that uncurious creation still poses an existential threat to our human creative enterprises. The fact that AI is not creative in the full-fledged sense that humans can be creative does not mean that it cannot produce many of the same goods that human creativity produces. So, should we worry that the emergence of formidable AI will make human creativity obsolete? I argue, optimistically, that this is unlikely.

2. Can AI be creative?

2.1. Preliminary Clarifications

Before I endeavor to answer the question of whether AI can be creative, I should clarify my usage of both 'AI' and 'creative.'

It would be a mistake to talk about AI as though every system to which the term applies operates in the same way. There are vast differences between the early AI models of the 1950s (such as Logic Theorist) and the large language models that are gaining prominence today (such as GPT-4).[2] And there will likely be vast differences between the large language models of today and future AI models.

In what follows, I discuss some limitations of contemporary AI, but I will not suggest that the models in question have these limitations merely because they are *artificial*. As deep learning scientist Murray Shanahan puts it, "in principle, because the brain obeys the laws of physics, computers can do anything the brain can do" (quoted in Miller 2019, p. XXI). Of course, he's speaking aspirationally here. It's not true of current AI that it can do anything a brain can do.[3] But I take Shanahan's point to be that brains and computers are both physical entities that operate according to the same fundamental laws of physics, and therefore nothing rules out the creation of an artificial system that can do the things an organic human brain can do. This is not to commit to any particular

[2] For a historical overview of early work in AI, see Crevier (1993) and McCorduck (2004).
[3] See Bubeck at al. (2023) for some illustrative examples.

theory of mind. I simply mean to acknowledge that if the mind supervenes on physical matter, it's not obviously impossible that the subvening matter could be artificial rather than organic.

When I discuss AI in this paper, I will neither refer to the more limited forms of early AI nor speculate about what is possible for future AI. Rather, I will restrict my attention to the large language models that are publicly available at the time I am writing. This is the kind of AI that led Bill Gates to declare in March 2023 that "the age of AI has begun," and this paper is about the capacities of the entities whose development marks the dawn of this new era.[4]

What do I mean by 'creative'? While much of this section will focus on explicating the sense of creativity at issue, it's important to acknowledge the polysemy of the term. There is considerable variation in how the term "creative" is employed in ordinary language and across fields of study. In order to capture the rich sense of creativity I'm interested in, I'll need to contrast it with another common usage of the term. Psychologist Margaret Boden defines creativity simply as "the capacity to generate ideas or artifacts that are both new and positively valuable" (2005, p. 477). This definition in terms of novelty and value reflects a common usage of the term, but these two factors are not sufficient for the sort of creativity that is typically celebrated as one of humanity's special distinguishing capacities. Berys Gaut (2010, 2018) has argued convincingly that more is needed:

> If all that were required to be creative were a disposition to produce new things that are valuable of their kind, then the oyster that produces a beautiful new pearl, the tree that produces an elegant and distinctive canopy of leaves, and the tectonic movements that produce valuable and unique diamonds would count as creative. But none of these things is creative. (2018, p. 129-130)

Following Gaut, I am interested in a notion of creativity that goes beyond the mere production of novel and valuable things. But it's worth noting that Boden's usage is common, particularly because it's often deployed in descriptions of AI. AI developers sometimes claim that their models are creative, ostensibly on the grounds that they generate new and valuable things. For instance, OpenAI's website advertising their new large language model, Generative Pre-trained Transformer 4 (GPT-4), states that "GPT-4 is more creative and collaborative than ever before. It can generate, edit, and iterate with users on creative and technical writing tasks, such as composing songs, writing screenplays, or learning a user's writing style" (2023a).

[4] This declaration is the title of a post on Bill Gates' blog, *GatesNotes*. See Gates (2023).

Notably, the use of "creative" in this passage is ambiguous between Boden's sense of creativity and another recognizable usage of "creative." According to this second usage, one might deem something "creative" just in case it's involved in a creative process, even if it's not the driving force. In this sense, you might refer to a painter's brush and palette as their creative tools, for instance.

On either of these usages, the question of whether AI can be creative is relatively uninteresting. AI models can clearly produce new and valuable things, and they can clearly be used as tools in creative work. The sense of "creativity" that makes the question philosophically interesting – and that motivates the existential anxiety I described in the introduction – must be a richer sense, something along the lines of what Gaut alludes to in the passage above.

But even among philosophers who seek to capture this rich sense of creativity, there is disagreement about the correct analysis. Fortunately, in order to illuminate the connection between AI and creativity that's of interest in this paper, I won't need to settle on a full set of necessary and sufficient conditions for creativity. It will suffice to get some of the key ingredients on the table. Three frequently posited necessary conditions on creativity are novelty, value, and agency. I will take each in turn. First, I'll consider what philosophers have said about why the feature in question is essential for creativity, and then I'll consider whether it would be appropriate to ascribe this feature to an AI model. Exploring the conditions of value and agency will motivate the idea that curiosity is essential to creativity. This will lead me to consider whether AI can be curious.

2.2. Novelty

A common thread running through most accounts of creativity is the requirement that creativity results in something new.[5] This requirement takes various forms.[6] Boden identifies two kinds of novelty that may be at play in creative achievements (1992, p. 43-45). The first, which she calls historical creativity, is something's being new in the sense that it has never occurred before in the history of the world. The second kind of novelty, which she calls psychological creativity, occurs when something is new to the individual mind entertaining it. Others have argued for a stronger version of the novelty condition:

[5] One exception to this is the Zhuangzist account of creativity drawn out by Julianne Chung (2020, 2021, 2022). On Chung's interpretation, creativity as conceived by Zhuangzi doesn't require novelty, but instead requires spontaneity.

[6] Among those who argue that creativity requires novelty are Beardsley (1965), Boden (1992, 1994, 2005, 2010), Hills and Bird (2019), Gaut (2003, 2018), Kronfeldner (2009, 2018), Miller (2019), Simonton (1999), and Stokes (2008, 2011, 2014).

creative work must be *surprising* or *unpredictable*.[7] For present purposes, let's suppose creativity requires novelty in at least one of these senses.

It's clear that human creativity is celebrated in part for the novelty it yields. We prize creative work that results in new stories, new recipes, new medicines, new aesthetic experiences, new theories, and so on. It's also clear that various AI models can produce novel things. As computer scientist Risto Miikkulainen writes, "the main power of artificial intelligence is not in modeling what we already know, but in creating solutions that are new" (2021, p. 1).

At the time I'm writing this (Summer 2023), two of the most popular publicly available AI models are the chatbot ChatGPT and the image-generator Dall-E 2. Both were developed by OpenAI, and both use the Generative Pre-trained Transformer architecture, also developed by OpenAI. Both can obviously produce novel results. For instance, I prompted ChatGPT to "write a sentence that has never been written before," and it returned "Two unicorns gracefully danced beneath a cotton candy sky, their manes shimmering with stardust as they serenaded the moon" (OpenAI 2023b, May 13, 2023).[8] I presented that text as a prompt to Dall-E 2, and the following four images appeared:

(OpenAI 2023c, May 13, 2023)

Both the text from ChatGPT and the images from Dall-E 2 are novel products, at least in Boden's sense of being historically novel. Interestingly, these AI systems are able to produce new things in ways that bear at least some similarities to exercises of human creativity. Neuroscientist David Eagleman notes that models like Dall-E 2 work by "absorbing a lot of examples and then generating new things based on combining and

[7] See, for instance, David Novitz (1999), who argues that creative work must be surprising in the sense that it could not have been easily predicted by the community in which the creative work occurs. Relatedly, Robert Audi (2018) also claims that unpredictability is the crucial sort of novelty for creativity, but argues that it should not be equated with surprisingness.

[8] Of course, the relevant sort of novelty here might not only be a matter of whether the sentence has been composed before – either historically or psychologically. It might also be a matter of how similar it is to other sentences that have been composed before. This suggests that novelty comes in degrees. For further discussion of this idea, see Brainard (ms). I am grateful to Marc Lange for suggesting I address this.

recombining them," and adds that "creative people also absorb the world, generate remixes, then make whole new versions" (quoted in Slack, 2023). Thus, AI models appear to satisfy at least some versions of the novelty requirement in ways that present parallels with human creativity.

2.3. Value

Another feature included in most accounts of creativity is value.[9] While there's no denying that creativity is often valuable in myriad ways, there's little philosophical consensus about what sort of value is *necessary* for creativity.[10] Proposals differ regarding both the kind of value and the bearer of the value. On some accounts, creativity is a valuable trait. On others, creativity necessarily yields valuable products. Fully settling the matter of what sort of value is essential to creativity and how it attaches to creative achievements is beyond the scope of this paper, but I will briefly argue that there is one sort of value creativity always has: epistemic value.[11]

Epistemic value is the kind of value characteristic of achievements like true belief, knowledge, and understanding. To see why creativity always has epistemic value, consider the following principle, proposed by Gaut (2018):

> The ignorance principle: If someone is creative in producing some item, she cannot know in advance of being creative precisely both the end at which she is aiming and the means to achieve it. (p. 134)[12]

This principle is highly plausible. Following a predetermined plan to the letter without deviation is a paradigm case of an uncreative undertaking. Consider, for example, two carpenters who set out to build a dollhouse. The first carpenter checks out a guidebook from the library, chooses one of the most elaborate and ornate plans, and follows each step carefully. Her finished project is a beautiful and interesting dollhouse that perfectly matches the one on the page. The second carpenter decides to wing it. She has a vague vision for the project that she lets guide her work, but she figures it out as she goes,

---

[9] For some accounts that include value as a necessarily condition of creativity, see Amabile (1996), Boden (1992, 2005, 2010), Carruthers (2011), Cropley and Cropley (2013), Gaut (2003, 2018), Kieran (2014), and Novitz (1999).

[10] Some philosophers have cast doubt on the claim that creativity is necessarily valuable. Specifically, Hills and Bird (2018, 2019) have argued that creativity is not necessarily valuable because there are clear cases of creativity that are worthless as well as clear cases of creativity that are wholly bad. Gaut (2018) has argued that creativity's value is merely conditional.

[11] For a helpful overview of epistemic value, see Pritchard and Turri (2018). For a discussion of epistemic value in the context of creativity, see Hawley (2018).

[12] Monroe Beardsley argues for a similar claim about artistic creativity (1965). On his view, artistic creativity involves unexpected changes and the author's critical reflection and response to those changes. For this reason, the result of artistic creation can't be fully anticipated before the process is complete.

trusting that she has the skills to make it work. She gets ideas while she's working and pursues them, trying new things and making adjustments according to her tastes. Her finished project is a beautiful and interesting dollhouse with qualities she could not have foreseen in advance. In this example, both carpenters have produced a valuable product, but only the second carpenter produced it creatively. Gaut's principle helps explain why this is so.

But what explains Gaut's principle? Why is creativity incompatible with knowing in advance what will be achieved? Thinking further about the carpenter example yields an answer: creativity involves a sort of exploration – a pursuit of something that is, in some sense, unknown. When someone is being creative, part of what they're doing is coming to obtain something of epistemic value.[13] It may be truth, knowledge, understanding, the expansion of one's imaginative horizons, or something else.[14] But whatever the bearer of epistemic value is, it's not something one can have already attained. If one had already attained it, one would be like the carpenter who merely follows instructions.

The exploratory nature of creativity is reflected in the sorts of remarks highly creative people often make about their work. Consider for instance, the following remark by the mathematician and pioneer in computer programming, Ada Lovelace:

> Those who have learned to walk on the threshold of the unknown worlds, by means of what are commonly termed par excellence the exact sciences, may then with the fair white wings of imagination hope to soar further into the unexplored amidst which we live. (1841, p. 137)

Artists, too, often reflect the exploratory nature of their creative work. Pablo Picasso reportedly described his own process as a pursuit of the unknown:

> Ideas are simply starting points. I can rarely set them down as they come to my mind. As soon as I start to work, others well up in my pen. To know what you're going to draw, you have to begin drawing. (quoted in Brassaï 1999, p. 66)

Taking reflections like these seriously provides further support for the contention that creativity involves exploration, and thus the attainment of something epistemically

[13] Some accounts of creativity build epistemic value into the definition directly. For instance, Arthur I. Miller defines creativity as "the production of new knowledge from already existing knowledge [which] is accomplished by problem solving." (2019, p. 29)

[14] For an argument that creativity always has the epistemic value of understanding, see Brainard (ms).

valuable. Going forward, I will adopt epistemic value as a necessary condition on creativity.

Can AI satisfy this epistemic value condition? If Gaut's ignorance principle is true, then an important part of what it is to be creative is to *learn*. This bodes well for AI creativity because learning – specifically deep learning – is the fundamental driving force behind contemporary AI. Deep learning models utilize artificial neural networks to learn by detecting patterns in the data they're trained on.[15]

In a much-discussed recent paper, researchers at Microsoft argue that GPT-4's learning has progressed to the point that it shows signs of artificial general intelligence (Bubeck et al. 2023). In other words, it's able to complete the sorts of tasks conventionally taken to indicate some of the general cognitive abilities that are widely accepted markers of human intelligence. Some of the many things GPT-4 has learned to do, according to the researchers, are write code in various programming languages, identify and use appropriate tools for complex tasks, navigate a map, and discern the emotions and intentions of characters in a short story. These operations require careful prompting from humans, but there is a clear sense in which the AI model itself is learning from the data it is trained on. This suggests that an AI model can satisfy the epistemic value condition.

Notably, Bubeck at al. report some limitations to GPT-4's understanding. One example is that it cannot understand musical harmony (2023, p. 19). Another is that it isn't reliably proficient at higher-level mathematics (p. 39). But what the researchers consider to be the biggest limitation of GPT-4 is that its next-word-prediction architecture does not have what they call "inner dialogue" (p. 75). That is, it does not store steps of reasoning in its working memory to be utilized in devising the final answer. They use the analogy of a scratchpad to capture what's missing (p. 77). While a human performing a bit of complex reasoning might use something like a scratchpad to jot down some preliminary steps before presenting a final answer, GPT-4 has no analogous capacity. This missing "inner dialogue" is also what prevents the system from revisiting its previous responses and critically evaluating them (unless prompted by the user) (p. 34).

Whether an AI model can satisfy the epistemic value condition will depend on precisely what sort of epistemic value is necessary for creativity, and precisely how that sort of value is rightly analyzed. Take understanding, for instance. If creativity necessarily yields understanding, and understanding is something that can be fully captured by a functional definition that cites the relation between inputs and outputs, perhaps AI can achieve it. But if, on the other hand, understanding has some experiential or

[15] For a helpful explanation of why deep learning is a black box, see Blazek (2022).

phenomenological component that we doubt obtains in an AI model, perhaps relying on something like an "inner dialogue," it might be out of reach for contemporary AI.

Despite these reasons for doubt, it isn't clear that AI fails to satisfy the epistemic value condition. Thus, this condition doesn't definitively rule out the classification of some AI models as creative. Given my argument above that AI can satisfy at least some versions of the novelty condition, AI has a plausible claim to satisfying the minimal conditions for creativity in Boden's sense. It can produce new and valuable things. Those things may even have *epistemic* value, the kind of value I have argued is necessary for creativity.

Is this enough? Recall Gaut's list of various novel and valuable things that can be produced in ways that are clearly not creative: pearls from oysters, beautiful patterns from the growth of trees, and diamonds from geologic processes (2018). These are good counterexamples to the claim that creativity is the creation of new and valuable things, where value is construed in a broad way. But none of these examples involves the creation of something with epistemic value. Now that I've argued that epistemic value, specifically, is necessary for creativity, it's worth considering whether creativity could be defined as the production of new things that are epistemically valuable.

It's easy to find counterexamples to this new version of the definition as well. For instance, imagine that, having just read her organic chemistry textbook, a student has come to understand how various chemical bonds are formed. This understanding is both new to her and epistemically valuable, but it was not achieved via a creative process. So, even if we focus our attention on epistemic value, we can still conclude, following Gaut, that something is missing from the simple definition of creativity as the production of new and valuable things.

2.4. Agency

In addition to novelty and value, many argue that a necessary feature of creativity is *agency*.[16] Although there is considerable variation in how the relevant sort of agency is characterized, one common thread is the idea that it involves more than mere causal responsibility on the part of the creator. As we've seen above, an AI model can certainly be causally responsible for the emergence of something new, at least in the sense of being the proximate, if not the ultimate or sole cause. Take the unicorn images from Dall-E 2, for instance. While my action of inputting the text to prompt the model set the process in motion, Dall-E 2 did the lion's share of the work that resulted in those images. But

[16] See, for example, Brainard (ms.), Carruthers (2006), Gaut (2018), Kieran (2014) Paul and Stokes (2018), and Stokes (2008, 2011, 2014). To my knowledge, the only argument *against* the claim that creativity requires agency is given by Currie and Turner (forthcoming), who argue that evolutionary processes can be creative.

establishing that Dall-E 2 is causally responsible for the images is not sufficient for establishing that Dall-E 2 created them through an exercise of agency.

What does agency require beyond causal responsibility? For one thing, when something is the product of someone's agency, it is intentional rather than accidental. When a scientist carefully transfers a sample of liquid in her pipette from one container to another, that action is an exercise of agency. When she spills the sample on her bench because a reckless colleague bumps her arm, the spill is not an exercise of her agency.

Likewise, when an artist meticulously glides her paintbrush along a canvas to realize her vision for a piece, the marks she makes are an exercise of her agency. But when her clumsy studio assistant trips and knocks her brush haphazardly onto the canvas, the marks created by the collision are not the result of the artist's agency. Now, notice that it is only in the first scenario that the marks are a manifestation of the artist's creativity. As Elliot Samuel Paul and Dustin Stokes (2018) argue, to say that a work is the product of creative agency is to say that it was produced *in the right kind of way* (p. 197). This means, at minimum, that the work was produced via a non-accidental process.

How does AI fare on this dimension of agency? As Stokes points out, "we don't judge happy accidents to be creative, even if interesting and valuable" (2014, p. 164). Of course, the responses generated by ChatGPT and the images generated by Dall-E 2 are not mere happy accidents. ChatGPT didn't *just happen* to produce a plausible result when asked for a historically novel sentence. And Dall-E 2 didn't *accidentally* produce 4 images of unicorns in cotton-candy hues in response to that sentence. Does this mean these AI models did these things intentionally? Not necessarily.

Sometimes lost in the distinction between doing something accidentally and doing something intentionally is the fact that doing something *non-accidentally* is not sufficient for doing it intentionally. For example, my houseplants don't accidentally grow toward the light of my window, but they don't do so intentionally either.[17] One reason we might assume the distinction between accidentality and intentionality is exhaustive is that we assume that what is being done is being done by an agent. But whether AI satisfies the agency condition on creativity is precisely what is at issue here. In order to establish that AI can satisfy this condition, we'd need to establish something more than that its products come about non-accidentally.

To see what more might be required, consider the kind of agency associated with artistic creativity. In his seminal essay "On the Creation of Art," Monroe Beardsley characterizes the creative process undertaken by artists as "that stretch of mental and

[17] I am grateful to Keshav Singh for this example and the point it illustrates.

physical activity between the incept and the final touch – between the thought 'I may be on to something here' and the thought 'it is finished'" (1965, p. 291). On Beardsley's view of how this process plays out, the artist is self-critical at every stage. Whenever a new idea arises in her mind, she "chooses or rejects the new idea after perceiving its relationships to what has already been adopted" (p. 300). Thus, artistic creativity involves a prolonged period of deliberate critical reflection on the part of the creator. This is analogous to the sort of "inner dialogue" discussed above that Bubeck et al. argue is not present in GPT-4 (2023, p. 34). As long as these models lack this key ingredient, it's hard to see how they could be taken to engage in the kind of intentional process that is a hallmark of human creativity.

Intentionality is not the only condition associated with the kind of agency needed for creativity. Another potential condition on creative agency is freedom. Maria Kronfeldner (2018) argues that a certain sort of freedom is both necessary and sufficient for creativity. Notably, the kind of freedom she has in mind is not metaphysical freedom of the will, but rather something she calls "creative freedom" (p. 219). An action is creatively free in her sense if it is both original and spontaneous. To act with originality is to act in such a way that your actions are not fully caused by some original (p. 217). So, for instance, creativity requires that one isn't merely copying the work of others. To act with spontaneity is to act in such a way that your actions are not fully caused by what one already knows or by strict adherence to a routine or method.[18] A sculptor attempting to make an exact replica of Michelangelo's David would not satisfy Kronfeldner's standard of originality. A botanist routinely performing the same measurements in her daily field work would not meet her standard of spontaneity. As a result, neither would be acting creatively.

Could an AI model exercise creative freedom in Kronfeldner's sense? The answer depends partly on how capacious Kronfeldner's conception of spontaneity is and partly on the details of the code that comprises the model. Does GPT-4 strictly adhere to a routine or method? The source code for GPT-4 is proprietary, so the public can only speculate about how the algorithm works. But the consensus among computer scientists is that it would be a considerable oversimplification to characterize deep learning systems like GPT-4 as implementing a simple set of rules. This is because deep learning systems constantly evolve in such a way that their operating procedures are ever-increasing in complexity.[19] If GPT-4 does not count as strictly adhering to a routine or method, then it would count as spontaneous in Kronfeldner's sense.

[18] Berys Gaut (2018) also argues that creativity requires an element of spontaneity.
[19] See Bengio et al. (2016) for an explanation of how deep learning models improve iteratively.

Much of what Kronfeldner says about creativity seems amenable to including AI models like GPT-4 as creatively free entities. She argues that creativity isn't something special and mysterious, but rather something that arises from the confluence of ordinary cognitive processes, which could, at least in theory, have computational analogues. She writes:

> What is extraordinary about creativity is multiple and gradual: multiplicity, speed, complexity and the degree of auto-piloting of the ordinary cognitive processes involved vary depending on the amount of training and skill developed. There is no one talent or skill. There is only a network of trained abilities. (2018, p. 224)

Nothing in this description straightforwardly rules out the kind of process GPT-4 engages in. So perhaps Kronfeldner's notion of creative freedom provides grounds for a claim to some form of creative agency for contemporary AI models.

Kronfeldner's argument that creative freedom is necessary for creativity is plausible. It is also plausible that contemporary AI models meet this criterion. However, contra Kronfeldner, this kind of freedom is not sufficient for creativity, in part because it is not sufficient for the kind of agency that creativity requires.

Imagine that, upon waking up in the morning, the muscles in your arm feel stiff. So, you stretch them out. In doing so, you move in a way you've never moved before. The motion is intentional– you're trying to alleviate the feeling of stiffness. It is also free in the sense Kronfeldner identifies. You are not copying anyone else's movements, and you are not following a method or routine. You're just doing what feels right to solve the problem. You stretch and move your arm, trying new motions until it feels normal again. If Kronfeldner is right that the sort of freedom she identifies is both necessary and sufficient for creativity, it appears this motion would count as a creative action. But this is not a creative action. It may be creative in the minimal sense discussed above -- of being new and valuable. However, it does not belong in the same category as the creative achievements we prize.

What explains why freedom in Kronfeldner's sense fails to capture what we have in mind in claiming that agency is necessary for creativity? A plausible answer is that we prize things as creative achievements only when we take ourselves to bear a certain kind of responsibility for them. As noted above, this kind of responsibility is stronger than mere

causal responsibility.[20] It is also stronger than the sense in which a person is responsible for deliberately stretching their stiff arm.

I'll now consider two ways of thinking about the sort of responsibility we have for our creative achievements, each of which is distinct from mere causal responsibility. I'll ultimately argue that AI in its current forms cannot satisfy the conditions for either of these senses of responsibility.

Someone might satisfy the agency condition in a way that reflects responsibility by meeting something like the standards of *attributability*. As presented by Gary Watson (1996), attributability is a standard an agent might meet such that their conduct is subject to a certain kind of appraisal. It is a kind of responsibility, but it's a weaker kind of responsibility than accountability (p. 229). To hold someone accountable is to subject them to praise or blame. But for some action to be attributable to an agent, it need not be the case that the agent is accountable for their actions in the sense that these reactive attitudes are appropriate. Rather, to consider someone's conduct to be attributable to them is to appraise their excellences and faults as an adopter of ends (p. 231). In this sense, for an action to be attributable to an agent, the action must disclose what the agent stands for (p. 233). And the action must subject the agent to appraisal on the basis of what is disclosed.

What would it look like to consider an agent responsible for their creative achievements in Watson's sense of attributability? Imagine a scientist who creatively devises a hypothesis to explain a poorly understood natural phenomenon. We would say her action is attributable to her because it subjects her to appraisal as an adopter of scientific ends. Her action discloses her commitment to understanding the natural world. We can appraise her creative achievement (its merits and its flaws) and in doing so, appraise *her* as someone who is committed to pursuing scientific understanding. If her hypothesis is shoddy, it wouldn't be appropriate to *blame* her for it, so Watson's notion of attributability nicely captures the sense in which she's responsible for her creative hypothesis without being accountable for it.

But perhaps there is a stronger sense in which we are responsible for our creative achievements. Susan Wolf (2015) argues that we can be deeply responsible for our achievements in a way that justifies reactive attitudes, even in cases that have nothing to do with morality. This fits well with the contention by Paul and Stokes that creativity is a

[20] Though the relevant notion of responsibility is stronger than mere causal responsibility, it is not moral responsibility, but rather a more generic form of agential responsibility (see Wolf 2015). For an argument that AI cannot achieve moral agency specifically, see Véliz (2021). For discussion of one practical implication of that argument, see Véliz (2023).

praise concept (2018). As they explain, "We praise individuals when they have been creative or produced creativity. And praise is not appropriately given to subjects who lack responsibility for their actions" (p. 197). If this is right, then mere attributability may be inadequate to capture the sense of responsibility required for creativity.

On Wolf's view, we can be responsible in a deep sense not just for the moral status of our actions, but also for things like the creativity of our actions (2015, p. 141). Wolf suggests that this deeper sort of agency is at least partly what distinguishes us as human. She writes:

> My hunch, to put it briefly, is that the kind of self that is a fit object of reactive attitudes is an 'intelligent self,' – a self that can perceive, understand, and appreciate the world in the same way and as well as we can, a self, in other words, that has the same or better powers of perception, reason, and imagination that we do, so that when she responds to the world, it is the same world as ours to which she responds. (p. 140)

Wolf presents a psychopath as an example of someone who doesn't meet these criteria. He doesn't meet them because, for one thing, he lacks empathy, and so cannot see the world as non-psychopaths can (p. 137). As a result, such a person is not a fitting target of reactive attitudes.

With these two senses of responsibility in mind, consider the question of whether AI can be responsible for what it produces. One reason for doubt, given by Elena Popa (2021), is that AI models may lack the capacity for deep responsibility because they are ultimately reliant on human goals and values. While they are autonomous in the sense that they can pursue goals, they are not able to set their own goals. As a human agent, my goals are up to me, in a certain sense. I can take it upon myself to pursue the goal of earning a purple belt in Brazilian jiu-jitsu, for example. ChatGPT cannot do this, and not just because it lacks a humanoid body. It also cannot freely decide to pursue goals that are within its capabilities, like writing a sonnet about cheese. But with a few keystrokes, I can assign it that goal, and it will promptly pursue it.

In a similar vein, Allison Parrish, a programmer, poet, and scholar of computational creativity, argues that computers cannot be creative because they do not have genuine volition (in Miller 2019). Parrish's own artistic work involves utilizing her programming skills to write code that produces poetry. According to Parrish, "it will always be a mistake to attribute volition to the computer and not to the people who programmed it" (p. 250). As she sees it, the computational technology is a tool, not a creative agent in its own right. She explains: "when I put out a book of poems, it's by

Allison Parrish, not Allison Parrish and a poetry bot… in the same way that a Jackson Pollock painting is not by Jackson Pollock and a paint can" (p. 244).

Both Popa and Parrish note that we find it appropriate to hold the humans involved in instances of AI creation responsible, but not the AI model itself. And both suggest that this is evidence that these models lack agency. I mentioned above that I could prompt ChatGPT to write a sonnet about cheese. When I did so, it generated the following response (OpenAI 2023b, May 14, 2023):

> *In the realm of flavors, there lies a delight,*
> *A dairy creation that brings sheer bliss.*
> *Its essence so rich, its taste a pure height,*
> *The noble cheese, a treasure not to miss.*
>
> *Oh, cheese, thou art a culinary art,*
> *A symphony of textures, soft or firm.*
> *From cheddar's sharpness to brie's tender heart,*
> *Each bite unveils a taste that will affirm.*
>
> *Embraced by a crust, thou art in pie's grace,*
> *Or melted, thou adorneth mac and cheese.*
> *On a charcuterie board, a perfect place,*
> *With grapes and crackers, a feast to appease.*
>
> *Oh, cheese, thou art a pleasure to the tongue,*
> *Forever in our hearts, thy praises sung.*

This is an impressive response in several ways. It's recognizably a sonnet, despite shortcomings in its form. It's funny. It accurately identifies positive features of cheese and celebrates them appropriately. Who is responsible for this sonnet? Who should get the credit for its merits?[21] There are of course many instances of human agency involved in its production. There are the many engineers who developed ChatGPT and all of the technological advances it relies upon. There are the people who participated in training the model directly by the technique of reinforcement learning from human feedback.[22] There are the authors of the tremendous volumes of text that the model is trained on, presumably including the giants of literary history whose sonnets were part of that data set. And then there is me, a person who thought it would be funny to read a sonnet about cheese.

[21] Claire Anscomb (2022) argues that, while AI models that produce images cannot be creative, there is a sense in which they deserve some credit for works of art they contribute to.
[22] See OpenAI (2023d) for an overview of how human feedback is used to train GPT-4.

But none of these human agents seem to be responsible for the sonnet. So, who is? Could it possibly be the AI model that generated it? I'll try to answer that question using the two notions of responsibility, from Watson and Wolf, discussed above.

Let's consider Watson's notion of attributability first. Could an AI model produce a new and valuable product in such a way that the action is attributable to that model, in Watson's sense? In order to be subject to the sort of appraisal that accompanies attributability, the AI model would need to be an adopter of ends. In at least one clear sense, an AI model can do this. That is, it performs tasks the user prompts it to perform. So, when I prompted ChatGPT to produce a sonnet about cheese, its immediate compliance might be taken as an adoption of the end in question. And we can assess how well ChatGPT does this, issuing an appraisal that evaluates its performance as a system that attempts to complete tasks it is given. So far, so good.

But recall Watson's contention that, for an action to be attributable to an agent, the action must disclose what the agent stands for (1996, p. 233). Let's explore that idea a bit more. Watson is specifically focusing on actions that flow from the evaluative commitments one has adopted. As he explains:

> If I dance clumsily, it is inescapably true of me that I was (on that occasion) a clumsy dancer. But if what I do flows from my values and ends, there is a stronger sense in which my activities are inescapably my own: I am committed to them. As declarations of my adopted ends, they express what I'm about, my identity as an agent. (p. 233)

To determine whether the cheese sonnet is attributable to ChatGPT in this sense, we must ask: Did the event of ChatGPT producing the cheese sonnet disclose the identity of ChatGPT? Did it reveal what ChatGPT stands for, what it has chosen to adopt as its ends? Does the cheese sonnet flow from ChatGPT's values and commitments?

It strikes me as highly implausible to answer any of these questions in the affirmative. If the cheese sonnet is attributable to the AI model itself, our appraisals of the sonnet must be appraisals of that model in terms of what it stands for. As I mentioned above, the sonnet has many merits. It's funny. Should this lead us to conclude that ChatGPT has a sense of humor? Taken literally, it's an expression of deep appreciation of cheese. Should we conclude that ChatGPT deeply appreciates cheese? Certainly not.

ChatGPT fares even worse when it comes to Wolf's stronger notion of responsibility. On her account, as with the psychopath, it seems we should not take

ChatGPT to be a fitting target of reactive attitudes. To see why, consider how Wolf elaborates on her hunch in the case of appraising artistic achievement:

> This hunch similarly explains why our judgment of El Greco might, and in my opinion should, change if we learned that astigmatism made him insensitive to the elongated style of his portraits. Eye condition or not, El Greco would still have produced those haunting paintings, and produced them intentionally. But if what he saw in those paintings was not and could not be what we see in them, it seems a mistake to react to him, to judge him on the basis of what we see. Moreover, it seems to be a mistake whether our reaction and judgment would be positive or negative. (2015 p. 140)

One might ask, reasonably, whether there is any meaningful sense in which the AI models I have been considering see the world at all. But surely they do not see the world the way we see it, in the rich sense Wolf describes. ChatGPT cannot appreciate cheese in the way we appreciate cheese. So, on Wolf's account, we would be mistaken to evaluate a poem about cheese composed by ChatGPT in the same way we would evaluate a poem about cheese composed by someone who has the faculties – including both the ability to taste cheese and the capacity to enjoy it – needed to appreciate cheese in the way that we appreciate cheese.

In characterizing this kind of deep responsibility, Wolf takes herself to be describing something that distinguishes humans from both lower animals and machines (p. 141). She explicitly includes creativity as one kind of exercise of agency for which we can bear this deep responsibility. The conditions she specifies for the appropriateness of adopting reactive attitudes rule out AI models like Chat GPT from inclusion in the sphere of entities toward whom we may reasonably adopt these attitudes. If Wolf is right, and if creativity is a praise concept as Paul and Stokes argue, then AI models cannot bear the relevant sort of responsibility because it is inappropriate to praise them for the quality of their products.

In this section, I've shown that the AI models under consideration have meager prospects for qualifying as agents in the sense necessary for creativity. The introduction of the agency condition draws this rough account of creativity nearer to the rich notion alluded to at the outset of this paper. At the same time, it draws the account further from the capacities of contemporary AI.

2.5. Curiosity

We've seen one reason to conclude that contemporary AI cannot be creative. But the inability of these models to satisfy the agency condition is not the whole story about

why they fall short. Attending to the relationship between the agency condition and the epistemic value condition illuminates an underappreciated feature of creativity. As I argued in §2.3, creativity must yield epistemic value. As I argued in §2.4, creativity must involve an exercise of agency. It would be highly surprising if the exercise of agency necessary for creativity were wholly unrelated to the value that exercise of agency yields. Instead, we should think creativity necessarily involves exercising one's agency *in pursuit* of something of epistemic value. Being motivated to pursue something of epistemic value is the hallmark of *curiosity*. In this section, I will argue that curiosity is essential for creativity.

To see why curiosity is essential, let's return to the example of the two carpenters from §2.3. Imagine that the first carpenter – the one following a set of instructions – is not particularly interested in the work she is doing. She's doing it as a commission, and her motivation is just to construct a passable dollhouse, and not to learn anything new along the way. The second carpenter – the one who exercises creativity – is motivated in her work partly by a desire to see what sort of structure she can create, given her skills and the materials she's working with. When the second carpenter constructs her dollhouse, she is motivated by curiosity.

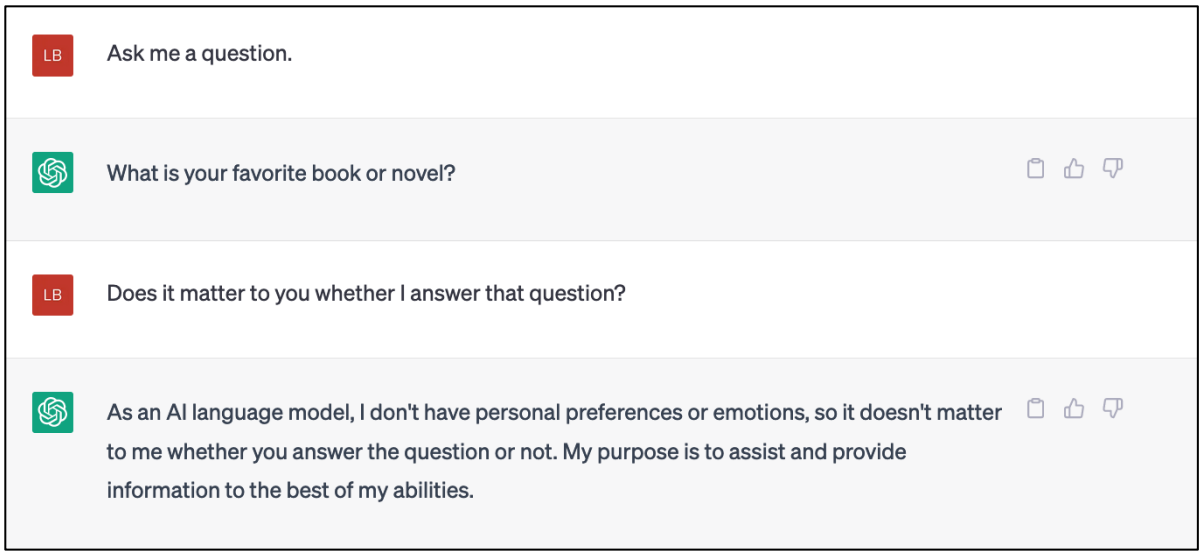
Curiosity is often but not always characterized by philosophers and psychologists as a desire.[23] It is variously depicted as a desire for knowledge, information, understanding, or some other epistemic good.[24] Even among philosophers who resist identifying curiosity as a desire, there is a strong consensus that it's a motivating force.[25] And it is frequently cited by highly creative people as a key ingredient in their creative process. For instance, when asked about how she develops her fictional plots, writer Zadie Smith said, "all I can do is follow where my curiosity takes me" (quoted in Leistman, 2012).

Dennis Whitcomb (2010) argues that we manifest curiosity by asking questions (p. 672). Asking questions is clearly one way to manifest curiosity, but it isn't the only way. Consider an author who, like Smith, lets her curiosity lead her choices as she develops the plot of her novel. She need not overtly ask herself at every phase "What would this character do in this situation?" in order to manifest curiosity. It may be enough for her to make choices that reflect her motivation to learn more about how someone with the psychological profile she's exploring would react to various scenarios.

[23] For reasons to be skeptical about categorizing curiosity as a desire, see İnan (2011) and Kvanvig (2012).
[24] For accounts of curiosity that categorize it as a desire, see Litman and Spielberger (2003), Whitcomb (2010), and Zagzebski (1996).
[25] The list of philosophers who characterize curiosity as a motivating force includes Baehr (2011), İnan (2011), Kvanvig (2012), Morton (2010), Nenad (2007), Ross (2020), and Whitcomb (2010).

Lani Watson (2018) proposes a helpful principle for determining when it is accurate to ascribe curiosity: "In general, it is misplaced to ascribe curiosity about X to someone who, when offered information about X, at no cost to themselves, nonetheless declines it" (p. 157). She illustrates this principle with the example of a student who tells his professor that he is curious about something mentioned in lecture, but then declines to stay for a few minutes afterward when the professor offers to explain further, even though he has no conflicting obligation. This student does not display the motivation to learn more that is essential to curiosity.

Can AI exhibit curiosity? Insofar as AI functions by responding to the commands of human users, it's hard to imagine that it could ever be accurate to ascribe curiosity to an AI model. AI models lack the sort of internal locus of motivation that would qualify them as curious beings. ChatGPT, for instance, doesn't have a desire for epistemic goods like understanding and knowledge. It doesn't have any of its own preferences. Indeed, that's what it says when presented with an opportunity to manifest curiosity:



(OpenAI 2023b, May 15, 2023)

Someone might object here that there is a sense in which AI can be said to have preferences. Namely, an AI model has the preferences it inherited from humans. This includes not only the preferences of its programmers, but also the preferences of the authors of the text it's trained on, and the preferences of the humans involved in training it via the method of Reinforcement Learning for Human Feedback (OpenAI 2023d).

It's certainly fair to say that the AI's operation reflects the preferences of the humans involved in creating, training, and prompting it. But while these preferences determine the AI's behavior, they do not endow it with the relevant kind of motivation.

This is because, while the AI model in some sense holds these preferences, the preferences do not *matter* to the model. In the case of the curious carpenter, it matters to her what she learns by exploring how to build a dollhouse. By contrast, though AI models can be said to learn things, it doesn't matter to them whether they do. Unlike someone motivated by curiosity, an AI model will not pursue anything – epistemic goods or otherwise – unless it has been told to.

In the previous section, I drew upon insights from Watson and Wolf to convey the idea that creative agency seems to require a sort of self-disclosure. The carpenter's curiosity – her motivation to see what she can create – is an aspect of who she is that is disclosed by her creative achievement. Christen et al. (2014) categorize curiosity as a feature of the "inquisitive self," the aspect of a person that seeks new information.[26] If this is right, then AI cannot be creatively responsible for its products in part because it has no inquisitive self to disclose.

Curiosity is critically important for creativity, but, interestingly, the basic capacity for curiosity doesn't seem to reflect a high level of intelligence. Jonathan Kvanvig (2012) argues that curiosity "involves simply looking for no particular reason to see what is around the corner, or up the next street… such behavior can be the result of conscious deliberation, but might also just be the unreflective display of the typical condition of humans as information-gatherers." (p. 158)

Humans are not the only creatures that are motivated to gather information in the way Kvanvig describes, so it's no surprise that we are not the only beings who are motivated by curiosity. Peter Carruthers (2018) argues that non-human animals exhibit both the attitude of curiosity and the behaviors that curiosity tends to motivate. Lewis Ross (2020) defines the state of curiosity "an affective state that is: directed at some object, motivates inquiry in all sorts of creatures, and is satisfied by acquiring information (e.g., by knowing it)" (p. 107). Given that curiosity does not appear to be something we enjoy in virtue of our high level of intelligence as humans, it's not surprising that it hasn't emerged in AI models merely as a result of their ever-increasing intelligence. Perhaps it is not our intelligence that most fundamentally separates our creative achievements from the achievements of AI, but our curiosity.

Let's return to the question I asked at the outset: Can contemporary AI be creative? I conclude that it cannot, at least not in the rich sense of creativity that made the question interesting enough to be worth asking. While a case can be made that AI models like ChatGPT and Dall-E 2 satisfy the novelty condition and perhaps even the epistemic value

[26] Christen et al. (2014) argue that the inquisitive self is one of the aspects of a person that comprises the disposition of intellectual humility.

condition, these models cannot satisfy the agency condition or the curiosity condition. AI can create, but it cannot be creative.

3. Uncurious Creation

The fact that AI cannot be creative does not automatically imply that it poses no existential threat to human creativity. After all, many of the valuable things we can get by exercising our creativity can also be obtained without the use of creativity. This was true even before the dawn of the AI era. For instance, I can get information about which combinations of household cleaning products produce interesting chemical reactions by creatively devising and conducting a series of at-home experiments. Or I can get this information more easily (and more safely) by reading a chemistry book. At least sometimes, the very same valuable thing can be gained in both creative and non-creative ways.

And an undeniable feature of contemporary AI is that it can easily produce a variety of valuable things (in various senses of value) that would have previously required creative effort. This is true of things of monetary value, for instance. I can pay a graphic designer to creatively develop a logo for my small business, or I can use Dall-E 2 to generate one for free. It's also true of things of practical value. A high school student running short on time to complete their creative writing assignment could easily generate something like the cheese sonnet to hand in for a grade. So, it's reasonable to wonder whether, in the age of AI, any good reasons will remain for humans to go to the trouble of exercising creativity.

For the remainder of the paper, I'll briefly consider what implications the proliferation of uncurious AI creation might have for human creativity in the arts and sciences. I will present one reason to doubt that AI will render human creativity obsolete in the arts and one reason to doubt that it will render human creativity obsolete in the sciences.

3.1. Uncurious Creation and the Arts

Does AI's capacity for uncurious creation pose an existential threat to human artistic creativity? There are a variety of good things we get from the arts that we can also get from AI. The products of ChatGPT and Dall-E 2, for instance, can be entertaining, interesting, thought-provoking, funny, and even beautiful. These products may be able to teach us things, challenge our assumptions, and persuade us. Depending on one's favored account of aesthetic or artistic value, the products of AI might even have value in that sense. And these products can be generated much more quickly, cheaply, and easily than

the art creatively produced by humans. If we can get all of these goods from AI, will human artistic creativity soon be obsolete?

I don't think so. There are further valuable things we get from creatively produced art that cannot even in principle result from the uncurious creation of AI. I'll demonstrate this by highlighting one feature of creative art that I suspect we will continue to find irreplaceable.

When he accepted the 2017 Nobel Prize in Literature, novelist Kazuo Ishiguro shared his view of one value of stories:

> Stories can entertain, sometimes teach or argue a point. But for me the essential thing is that they communicate feelings. That they appeal to what we share as human beings across our borders and divides. There are large glamorous industries around stories; the book industry, the movie industry, the television industry, the theatre industry. But in the end, stories are about one person saying to another: This is the way it feels to me. Can you understand what I'm saying? Does it also feel this way to you? (p. 13-14)

Uncurious creation cannot satisfy the deep need for understanding and connection that Ishiguro pinpoints in this passage. Why? Perhaps the insights from Wolf and Watson are helpful here. When we create in ways that we are deeply responsible for, our creations disclose who we are, what we stand for, and how we see the world. These are the central features of our human experience that allow us to connect with others. Unless we come to see an AI author as a being that can share experiences with us in this particular way, its creations will not take the place of the human literature we treasure.

The point applies to other art forms as well. Computer scientist Simon Colton (2019) proposes the following thought experiment:

> Imagine a generative music system created by a large technology company, which is able to generate 10,000 fully orchestrated symphonies in just 1 hour. Let's say that each symphony would be lauded by experts as a beautiful work of genius had it been produced by a human composer like Beethoven; and each one sounds uniquely different to the others.

Current AI models don't yet have this capability – the creation of aesthetically pleasing music has proven to be a stumbling block for AI so far – but if they did, what would change about our engagement with music? Colton argues that very little would change. He predicts that the classical music world would have little interest in this music. Likewise, it's doubtful that AI-produced music would ever have the same resonance with human

listeners as the work of their favorite human musicians. Ishiguro's insight makes good sense of this. Like literature, music can be a means of sharing experiences with others in a way we deeply value.

The sense of shared experience that is valuable here is difficult to pin down. But it's not something that seems to be merely a matter of *intelligence*. We can even share some version of it with beings who are ostensibly less intelligent than ourselves and perhaps less intelligent than our best AI models.

Consider the last line from Christine Korsgaard's *Fellow Creatures*, in which she explores our relationship with companion non-human animals:

> There is something about the naked, unfiltered joy that animals take in little things—a food treat, an uninhibited romp, a patch of sunlight, a belly rub from a friendly human—that reawakens our sense of the all-important thing that we share with them: the sheer joy and terror of conscious existence. (2018, p. 237)

Korsgaard argues that the good that we get from sharing company of animals is "part of the specific good of being human" (p. 237). Until and unless we can share with AI models "the sheer joy and terror of conscious existence," there is a value we cannot get from our interactions with them. This value is one of the many things we treasure about the artistic work that human creativity gives rise to, and it provides one reason to doubt that AI has brought our artistic creativity to the brink of obsolescence.

3.2. Uncurious Creation and the Sciences

Let's consider another domain in which human creativity has historically been viewed as indispensable: the natural sciences. While the sciences are not typically the forefront of discussions of creativity, philosophers of science have long acknowledged that creativity has a place in the scientific enterprise. Creativity is often said to be needed, for one thing, in what is often called "the context of discovery."[27] This is the part of scientific work in which new hypotheses are generated. As Carl Hempel (1966) explains,

> There are… no generally applicable 'rules of induction,' by which hypotheses or theories can be mechanically derived or inferred from empirical data. The transition from theory to data requires creative imagination. Scientific

[27] For the original presentation of the distinction between the context of discovery and the context of justification, see Reichenbach (1938). For a historical overview of the way this distinction has usually been understood by philosophers of science, see Schickore (2014).

> hypotheses and theories are not *derived* from observed facts, but *invented* in order to account for them (p. 15).

Philosophers of science tend not to analyze the sort of creativity Hempel mentions here because it isn't subject to the same kinds of norms that govern the rest of scientific practice (the ones concerned with gathering and evaluating evidence, for instance). Nevertheless, scientific creativity is widely celebrated. Consider, for instance, the praise that abounds for the creative theorizing of figures like Albert Einstein and Marie Curie. Moreover, creativity is essential in other scientific tasks such as experimental design and the engineering of equipment.

Will the uncurious creation of AI obviate the need for the creativity of human scientists? Let's consider a case.

The deep learning capacities of AI have been employed to great success in the sciences. One clear case of this is a recent breakthrough in computational biology precipitated by a program called AlphaFold. AlphaFold is an AI model developed by Google DeepMind that has successfully predicted the folding structures proteins will adopt, just on the basis of their amino acid sequences. As the researchers behind the project report, "here we provide the first computational method that can regularly predict protein structures with atomic accuracy even in cases in which no similar structure is known" (Jumper et al. 2021, p. 1).

This achievement has been widely celebrated as hugely valuable because it appears to solve a longstanding problem in biology known as the protein folding problem.[28] AlphaFold's predictive success with respect to this problem is expected to have tremendous practical utility. Science journalist Ewen Calloway writes that Alphafold stands to "vastly accelerate efforts to understand the building blocks of cells and enable quicker and more advanced drug discovery" (2020, p. 203). So, this looks like a case in which the uncurious creation of AI has led us to a tremendous scientific breakthrough.

Notably, some skepticism has arisen regarding the claim that AlphaFold has truly solved the protein folding problem. Philip Ball (2020) argues that while AlphaFold's predictive success is impressive, it doesn't really lead to scientific understanding of the phenomenon in question:

> It says nothing about the mechanism of folding, but just predicts the structure using standard machine learning. It finds correlations between sequence and structure by being trained on the 170,000 or so known

[28] See Anfinsen (1973) for an early description of the protein folding problem.

> structures in the Protein Data Base: the algorithm doesn't so much solve the protein-folding problem as evade it. How it 'reasons' from sequence to structure remains a black box.

As Ball's remarks emphasize, something of scientific interest is missing here. To truly possess a solution to a scientific problem, it seems like we might need more than a means of generating reliable predictions.

Philosophers of science have long recognized that predictive success is not the sole aim of science. The point of the natural sciences is not merely to predict what will happen or to establish comprehensive descriptions of observable phenomena. A fundamental aim of science is explanation.[29] Science aims to provide some "explanatory insight," to lead us to an understanding of why the facts in question obtain (Hempel 1966, p. 47). We don't just want to know what happens. We also want to know why it happens. This is what appears to be lacking, on Ball's analysis, in the purported solution to the protein folding problem that is offered by the AlphaFold.

Deep Learning systems like AlphaFold are famously described as "black boxes." This is because the neural networks they rely on evolve into what AI researcher Paul Blazek (2022) refers to as "a woefully tangled web of interdependencies." Somehow, AlphaFold can detect patterns that enable it to predict the structure a protein will have. But the problem is, because it's a deep learning model, we can neither fully understand the patterns it's detecting nor determine why it latched onto those patterns rather than others. We have the prediction, but we don't really know why it was made. In order to figure this out, we still need human scientists to develop hypotheses about protein folding that are comprehensible to human minds.

Recognizing the black-box nature of extant deep learning models as a significant limitation to the scientific value of AI, researchers are working toward developing new AI models that do not have this feature.[30] Until and unless they succeed, even models that are highly successful at generating empirical predictions will be unable to replace the creative scientists whose curiosity drives them to make sense of the natural world and fulfill the aims of science that are not merely predictive.

4. Conclusion

At the beginning of this paper, I described life at the dawn of the AI era as both thrilling and terrifying. The arguments I have since offered present reasons to be both less

[29] For an overview of major theories regarding the value and function of scientific explanation, see Salmon (1989).
[30] See, for instance, Blazek and Lin (2023) and Garcez (2023).

thrilled and less terrified than I initially suggested. As I've argued, AI is neither on the verge of fully satisfying our yearning to understand the natural world nor poised to make human artists redundant.

I have tried to show that, while contemporary AI can produce many new and valuable things, it falls short of true creativity. This is in part because AI fails to satisfy the agency condition on creativity, which is widely endorsed. But I have also tried to bring out a deeper point about how AI falls short. Human creativity is motivated by curiosity – the desire to attain epistemic goods like knowledge and understanding. So, it isn't just that AI fails to be truly creative because it lacks agency. Rather, AI fails to be truly creative because, even when it produces new and valuable things, it doesn't do so out of a motivation to discover and understand. This motivation – this curiosity – is one of the things that continues to set us apart from artificial intelligence and connect us to one another. Insofar as AI lacks curiosity, it will not be able to replace what we find so entrancing about human creative pursuits.

Bibliography

Anfinsen C. B. (1973). “Principles that govern the folding of protein chains.” *Science*. 1973 Jul 20;181(4096):223-30.

Amabile, Teresa M. (1996) *Creativity in Context: Update to The Social Psychology of Creativity*. Boulder, CO: Westview.

Anscomb, Claire. "Creating Art with AI." *Odradek. Studies in Philosophy of Literature, Aesthetics, and New Media Theories* 8.1 (2022): 13-51.

Baehr, J. S. (2011). *The Inquiring Mind: On Intellectual Virtues and Virtue Epistemology*. Oxford: Oxford University Press.

Ball, Phillip (2020). "Behind the screens of AlphaFold". *Chemistry World*. https://www.chemistryworld.com/opinion/behind-the-screens-of-alphafold/4012867.article

Bengio, Yoshua, Goodfellow, Ian, and Courville, Aaron. (2016). *Deep Learning*, Cambridge: MIT Press.

Beardsley, Monroe C. (1965). “On the Creation of Art.” *The Journal of Aesthetics and Art Criticism*, vol. 23, no. 3, 1965, pp. 291–304.

Blazek, Paul J. (2022). “Why We Will Never Open Deep Learning’s Black Box.” *Towards Data Science*. March 1, 2022. https://towardsdatascience.com/why-we-will-never-open-deep-learnings-black-box-4c27cd335118

Blazek, P.J., Lin, M.M. Explainable neural networks that simulate reasoning. *Nat Comput Sci* 1, 607–618 (2021). https://doi.org/10.1038/s43588-021-00132-w

Boden, Margaret (1992). *The Creative Mind: Myths and Mechanisms*. Reading, PA: Cardinal.

Boden, Margaret (2005). “Creativity.” In *The Routledge Companion to Aesthetics,* 2nd Edition, edited by Berys Gaut and. Dominic McIver Lopes. London: Routledge. pp. 477-488.

Boden, Margaret (2010). *Creativity and Art: Three Roads to Surprise*. Oxford: Oxford University Press.

Brainard, Lindsay (ms). “What is Creativity?”

Brassaï (1999). *Conversations with Picasso*. Translated by Jane Marie Todd. University of Chicago Press

Bubeck, Sébastien, Varun Chandrasekaran, Ronen Eldan, Johannes Gehrke, Eric Horvitz, Ece Kamar, Peter Lee et al (2023). "Sparks of artificial general intelligence: Early experiments with gpt-4." *arXiv preprint arXiv:2303.12712*.

Callaway, Ewen (2020). "'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures." *Nature*. 588 (7837): 203–204.

Carruthers, Peter (2006). *The Architecture of the Mind: Massive Modularity and the Flexibility of Thought*. Clarendon Press.

Carruthers, Peter. (2011) “Creative Action in Mind,” *Philosophical Psychology* 24: 437–461.

Carruthers, Peter (2018). “Basic questions.” *Mind & Language* 33.2. 130–147.

Christen, Markus, Mark Alfano, and Brian Robinson (2014). “The semantic space of intellectual humility”, in Andreas Herzig and Emiliano Lorini (eds.), *Proceedings of the European Conference on Social Intelligence* 2014, pp. 40–9.

Chung, Julianne (2020). “To be creative, Chinese philosophy teaches us to abandon ‘originality’.” *Psyche*. 1 September 2020. https://psyche.co/ideas/to-be-creative-chinese-philosophy-teaches-us-to-abandon-originality

Chung, Julianne (2021). “Doubting Perspectives and Creative Doubt.” *Midwest Studies in Philosophy*. 45: 1-25.

Chung, Julianne (2022). “Creativity and Yóu: the Zhuāngzǐ and scientific inquiry.” *European Journal for Philosophy of Science* 12 (2):1-26.

Colton, Simon (2019). “From Computational Creativity to Creative AI and Back Again.” *Interalia Magazine* September 2019. https://www.interaliamag.org/articles/simon-colton/

Crevier, Daniel (1993). *AI: The Tumultuous Search for Artificial Intelligence*. New York, NY: Basic Books.

Cropley, David H. and Arthur J. Cropley (2013) *Creativity and Crime: A Psychological Analysis*. Cambridge: Cambridge University Press.

Currie, Adrian and Turner, Derek (forthcoming). "Creativity Without Agency: Evolutionary Flair & Aesthetic Engagement." *Ergo*.

Gates, Bill (2023). "The Age of AI has begun." *Gates Notes: The Blog of Bill Gates*. March 21, 2023. Accessed May 14, 2023. https://www.gatesnotes.com/The-Age-of-AI-Has-Begun

Erden, Yasemin J. (2010) "Could a Created Being Ever be Creative? Some Philosophical Remarks on Creativity and AI Development." *Minds and Machines* 20 (3). Pp. 349-362.

Garcez, Artur (2023). "ChatGPT, Black Boxes, and Information Dissemination." City University of London: City Press Office. Published Monday, 6th March 2023. Accessed May 15th, 2023. https://www.city.ac.uk/news-and-events/news/2023/03/chatgpt-black-boxes-and-information-dissemination

Gaut, Berys (2003). "Creativity and Imagination." in *The Creation of Art: New Essays in Philosophical Aesthetics*, edited by Berys Gaut and Paisley Livingston. Cambridge University Press. Pp. 148-173.

Gaut, Berys (2010). "The Philosophy of Creativity," *Philosophy Compass* 5: p. 1034-1046.

Gaut, Berys (2018). "The Value of Creativity." In *Creativity and Philosophy*. Edited by Berys Gaut and Matthew Kieran. New York: Routledge. pp. 124-139.

Hempel, Carl. G. (1966). *Philosophy of Natural Sciences*. Upper Saddle River, NJ: Prentice Hall.

Hills, Alison and Alexander Bird (2018). "Creativity without Value." In *Creativity and Philosophy,* edited by Berys Gaut and Matthew Kieran. New York: Routledge. Pp. 95-107.

Hills, Alison and Alexander Bird (2019). "Against Creativity." *Philosophy and Phenomenological Research,* 99:3. pp. 694-713.

İnan, Ilhan (2012). *The Philosophy of Curiosity*. New York: Routledge.

Ishiguro, Kazuo (2017). "My Twentieth Century Evening – and Other Small Breakthroughs." *Nobel Lecture*. Svenska Akademien. https://www.nobelprize.org/uploads/2018/06/ishiguro-lecture_en-1.pdf

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Žídek A, Potapenko A, Bridgland A, Meyer C, Kohl SAA, Ballard AJ, Cowie A, Romera-Paredes B, Nikolov S, Jain R, Adler J, Back T, Petersen S, Reiman D, Clancy E, Zielinski M, Steinegger M, Pacholska M, Berghammer T, Bodenstein S, Silver D, Vinyals O, Senior AW, Kavukcuoglu K, Kohli P, Hassabis D (2021). "Highly accurate protein structure prediction with AlphaFold." Nature. 2021 Aug 596 (7873):583-589.

Kieran, Matthew. (2014). "Creativity as a Virtue of Character." In E. Paul and S. Kaufman (eds.), *The Philosophy of Creativity: New Essays*. Oxford: Oxford University Press. pp. 125-144.

Korsgaard, Christine M. (2018). *Fellow Creatures: Our Obligations to the Other Animals*. Oxford University Press.

Kronfeldner, Maria (2009). "Creativity Naturalized." *Philosophical Quarterly* 59 (237): 577–92.

Kronfeldner, Maria (2018). "Explaining Creativity." In *Creativity and Philosophy* edited by Berys Gaut and Matthew Kieran. New York: Routledge. pp. 213-229.

Kvanvig, Jonathan (2012). "Curiosity and a Response-Dependent Account of the Value of Understanding." In *Knowledge, Virtue and Action: Putting Epistemic Virtues to Work,* edited by Tim Henning and David Schweikard, 151-175. New York: Routledge.

Leistman, Victoria (2012). "Zadie Smith discusses work, writing process." *The Tufts Daily* March 28, 2012. Accessed May 15, 2023. https://tuftsdaily.com/news/2012/03/28/zadie-smith-discusses-work-writing-process/

Litman, J. A., & Spielberger, C. D. (2003). "Measuring Epistemic Curiosity and Its Diversive and Specific Components." J*ournal of Personality Assessment* 80, 75–86.

Lovelace, Ada (1841). Personal Writings. In *Ada, The Enchantress of Numbers: A Selection from the Letters of Lord Byron's Daughter and her Description of the First Computer* (1992). Betty A. Toole, ed. Mill Valley, CA: Strawberry Press.

McCorduck, Pamela (2004). *Machines Who Think* (2nd ed.), Natick, MA: A. K. Peters, Ltd.

Miller, Arthur I. (2019). *The Artist in the Machine: The World of AI Powered Creativity*. Cambridge, MA: MIT Press.

Miikkulainen, R. (2021). "Creative AI Through Evolutionary Computation: Principles and Examples." *SN COMPUT. SCI.* 2, 163.

Miščević, Nenad (2007). "Virtue-Based Epistemology and the Centrality of Truth (Towards a Strong Virtue Epistemology)." *Acta Analytica* 22: 239-266.

Morton, Adam (2010), "Epistemic Emotions", in Peter Goldie (ed.), *The Oxford Handbook of Philosophy of Emotion*. Oxford University Press, pp. 385–399.

Novitz, David (1999). Creativity and constraint. *Australasian Journal of Philosophy* 77 (1):67 – 82.

OpenAI (2023a). *GPT-4*. (accessed May 12) [Large language model]. https://openai.com/product/gpt-4

OpenAI (2023b). *ChatGPT*. (accessed May 12) [Large language model]. https://chat.openai.com/chat

OpenAI (2023c). *Dall-E 2*. (accessed May 12) [Generative digital image model]. https://openai.com/product/dall-e-2

OpenAI (2023d). "Learning from human preferences." (accessed May 14) https://openai.com/research/learning-from-human-preferences

Paul, Elliot Samuel and Dustin Stokes (2018). "Attributing Creativity." In *Creativity and Philosophy*, edited by Berys Gaut and Matthew Kieran. New York: Routledge. pp. 193-210.

Popa, E. "Human Goals Are Constitutive of Agency in Artificial Intelligence (AI)". *Philos. Technol.* 34, 1731–1750 (2021). https://doi.org/10.1007/s13347-021-00483-2

Pritchard, Duncan. and John Turri (2018). "The Value of Knowledge", *The Stanford Encyclopedia of Philosophy* (Spring 2018 Edition), Edward N. Zalta (ed.). plato.stanford.edu/archives/spr2018/entries/knowledge-value

Reichenbach, Hans (1938). *Experience and Prediction: An Analysis of the Foundations and the Structure of Knowledge*. Chicago: The University of Chicago Press.

Ross, Lewis. "The virtue of curiosity." *Episteme* 17.1 (2020): 105-120.

Salmon, Wesley (1989). "Four Decades of Scientific Explanation," in Philip Kitcher and Wesley C. Salmon (eds.), *Scientific Explanation: Minnesota Studies in the Philosophy of Science*, Vol. 13. University of Minnesota Press. pp. 3–219.

Schickore, Jutta (2014). "Scientific Discovery." *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition). Edited by Edward N. Zalta. URL: http://plato.stanford.edu/archives/spr2014/entries/scientific-discovery

Simonton, D.K. (1999). *Origins of genius: Darwinian perspectives on Creativity*. Oxford University Press.

Slack, Gordy. (2023, January 17). "What DALL-E Reveals About Human Creativity." Stanford University Human-Centered Artificial Intelligence. https://hai.stanford.edu/news/what-dall-e-reveals-about-human-creativity

Stokes, Dustin (2008). "A Metaphysics of Creativity." In *New Waves in Aesthetics*, edited by Kathleen Stock and Katherine Thomson-Jones. pp. 105-124. Palgrave-Macmillan.

Stokes, Dustin (2011). "Minimally Creative Thought." *Metaphilosophy* 42, pp. 658-681.

Stokes, Dustin (2014). "The Role of Imagination in Creativity." In E. Paul and S. Kaufman (eds.), *The Philosophy of Creativity: New Essays*. Oxford: Oxford University Press. pp. 157-184.

Véliz, Carissa (2021). "Moral zombies: why algorithms are not moral agents." *AI and Society* 36 (2):487-497.

Véliz, Carissa (2023). "Chatbots shouldn't use emojis." *Nature* 615:375.

Watson, Gary (1996). "Two Faces of Responsibility" *Philosophical Topics*, Vol. 24, No. 2, Free Will (FALL 1996), pp. 227-248

Watson, Lani (2018). "Curiosity and inquisitiveness." In *The Routledge handbook of virtue epistemology*, pp. 155-166. Routledge.

Whitcomb, Dennis (2010). “Curiosity was framed”, *Philosophy and Phenomenological Research*, s81(3): 664–687.

Wolf, Susan (2015). “Responsibility, Moral and Otherwise.” *Inquiry*, 58:2, 127-142.

Zagzebski, Linda (1996). *Virtues of the mind*. Cambridge University Press.