AI & SOCIETY
https://doi.org/10.1007/s00146-024-02127-3

OPEN FORUM

On the creativity of large language models

Giorgio Franceschelli[1] · **Mirco Musolesi**[1,2]

Received: 8 November 2023 / Accepted: 29 October 2024

Abstract

Large language models (LLMs) are revolutionizing several areas of Artificial Intelligence. One of the most remarkable applications is creative writing, e.g., poetry or storytelling: the generated outputs are often of astonishing quality. However, a natural question arises: can LLMs be really considered creative? In this article, we first analyze the development of LLMs under the lens of creativity theories, investigating the key open questions and challenges. In particular, we focus our discussion on the dimensions of value, novelty, and surprise as proposed by Margaret Boden in her work. Then, we consider different classic perspectives, namely product, process, press, and person. We discuss a set of "easy" and "hard" problems in machine creativity, presenting them in relation to LLMs. Finally, we examine the societal impact of these technologies with a particular focus on the creative industries, analyzing the opportunities offered, the challenges arising from them, and the potential associated risks, from both legal and ethical points of view.

Keywords Large language models · Machine creativity · Generative artificial intelligence · Foundation models

1 Introduction

Language plays a vital role in how we think, communicate, and interact with others.[1] It is therefore of no surprise that natural language generation has always been one of the prominent branches of artificial intelligence (Jurafsky and Martin 2023). We have witnessed a very fast acceleration of the pace of development in the past decade culminated with the invention of transformers (Vaswani et al. 2017). The possibility of exploiting large-scale data sets and the availability of increasing computing capacity has led to the definition of the so-called foundation models, which are able to achieve state-of-the-art performance in a variety of tasks (Bommasani et al. 2021).

Among them, large language models (LLMs) are indeed one of the most interesting developments. They have captivated the imagination of millions of people, also thanks to a series of entertaining demonstrations and open tools released to the public. The examples are many from journal articles[2] to culinary recipes (Lee et al. 2020) and university-level essays.[3] LLMs have also been used to write papers about themselves writing papers (GPT-3 2022). They are commonly used for creative tasks like poetry or storytelling and the results are often remarkable.[4] Notwithstanding, it is not obvious whether these "machines" are truly creative, at least in the sense originally discussed by Ada Lovelace (Menabrea and Lovelace 1843). LLMs have already been analyzed (and sometimes criticized) from different perspectives, e.g., fairness (Bender et al. 2021), concept understanding (Bender and Koller 2020), societal impact (Tamkin et al. 2021), and anthropomorphism (Shanahan 2024) just to name a few. However, a critical question has not been considered yet: *can LLMs be considered creative*?

✉ Giorgio Franceschelli
giorgio.franceschelli@unibo.it

✉ Mirco Musolesi
m.musolesi@ucl.ac.uk

1 Department of Computer Science and Engineering, University of Bologna, Viale del Risorgimento 2, 40136 Bologna, Italy

2 Department of Computer Science, University College London, 66-72 Gower Street, London WC1E 6BT, United Kingdom

[1] As remarked by ChatGPT itself when asked about the importance of language.

[2] http://www.theguardian.com/commentisfree/2020/sep/08/robot-wrote-this-article-gpt-3.

[3] https://www.theguardian.com/technology/2022/dec/04/ai-bot-chatgpt-stuns-academics-with-essay-writing-skills-and-usability.

[4] See, for instance: https://www.gwern.net/GPT-3.

By taking into account classic frameworks for analyzing creativity, such as Boden's three criteria (Boden 2003) and other prominent cognitive science and philosophical theories (e.g., Amabile (1983); Csikszentmihalyi (1988); Gaut (2010)), we will try to answer this question. We will discuss the dimensions according to which we believe LLMs should be analyzed to evaluate their level of machine creativity. To the best of our knowledge, this article represents one of the first investigations of the problem of LLM creativity from a theoretical and philosophical perspective.

The remainder of the paper is structured as follows. First, we briefly review the past developments in automatic text generation and artificial creativity (Sect. 2) that led to today's LLMs. Then, we analyze LLMs from the perspective of Boden's three criteria (Sect. 3), as well as considering other relevant philosophical theories (Sect. 4). Finally, we discuss the practical implications of LLMs for the arts, creative industries, design, and, more in general, scientific and philosophical inquiry (Sect. 5). Section 6 concludes the paper, outlining the open challenges and a research agenda for future years.

2 A creative journey from Ada Lovelace to foundation models

It was the year 1843 when Ada Lovelace wrote that the Analytical Engine (Babbage 1864) *"has no pretensions to originate anything. It can do whatever we know how to order it to perform"* (Menabrea and Lovelace 1843). This statement was then defined as "Lovelace's objection" by Alan Turing, who also provided an alternative formulation: a machine can never "take us by *surprise*" (Turing 1950). This was just the beginning of an ongoing philosophical discussion, which has often included psychological elements, around human creativity (Barron 1955; Berlyne 1960; Bruner 1962; Newell et al. 1962; Stein 1974), as well as computational creativity (Boden 2009; Colton and Wiggins 2012; Jordanous 2009; Macedo et al. 2004; Maher 2010; Wiggins 2006).

In general, computer scientists have always been fascinated by the possibility of building machines able to express themselves through writing, e.g., by composing poems and short stories, creating paintings, and so on. In particular, the rise of automatic text generation was contextual to the birth of personal computers. Examples include the Computerized Haiku by Margaret Masterman,[5] the storyteller TALE-SPIN (Meehan 1977), Racter and its poems' book (Racter 1984), and UNIVERSE, which was able to generate coherent and consistent characters (Lebowitz 1983), just to name a few. Different techniques have been explored, from planning (e.g., Riedl and Young (2010)) and case-based reasoning (e.g., Turner (1994)) to evolutionary strategies (e.g., Manurung et al. (2012)). Some approaches combine all of them together (Gervás 2013).

Only with the advent of neural networks and learning systems, we observed a real step-change. In particular, deep language models, i.e., probabilistic models of in-context token occurrences trained on a corpus of text with deep learning, easily allow the sampling of new text, facilitating and automating natural language generation. For instance, recurrent neural networks with long-short term memory (LSTM) (Hochreiter and Schmidhuber 1997) or gated-recurrent units (GRUs) (Cho et al. 2014) can predict the next character (Karpathy 2015), word (Potash et al. 2015), syllable (Zugarini et al. 2019), or event (Martin et al. 2018) given previous ones, allowing to compose text that spans from short movie scripts to knock-knock jokes (Miller 2019). Other successful generative methods include generative adversarial networks (GANs) (Yu et al. 2017; Zhang et al. 2017) and variational auto-encoders (VAEs) (Bowman et al. 2016; Semeniuta et al. 2017). We refer the interested reader to Franceschelli and Musolesi (2024b) for an in-depth survey of deep learning techniques applied to creative artifacts.

These models tend to scale poorly to long sequences, and they are often unable to capture the entire context. For this reason, current state-of-the-art language models make use of attention (Bahdanau et al. 2015) and transformers (Vaswani et al. 2017). In recent years, several models based on these mechanisms have been proposed. They usually rely on a very large number of parameters and are trained on corpus datasets of greater and greater size (Brown et al. 2020; Chowdhery et al. 2023; Devlin et al. 2019; Du et al. 2022; Hoffmann et al. 2022; Radford et al. 2019; Rae et al. 2021; Raffel et al. 2020; Rosset 2020; Shoeybi et al. 2019; Smith et al. 2022; Thoppilan et al. 2022). Thanks to in-context learning techniques such as zero-shot or few-shot learning (Dong et al. 2024), these models can produce more specific and specialized content, such as poems or stories (Swanson et al. 2021), by simply providing a description of the task and possibly some examples. However, finding the correct input and high-quality demonstrations for solving this type of task can be challenging (Liu et al. 2022). Certain domains might require more fine-grained knowledge than that acquired during pre-training (Peng et al. 2023). Because of this, other methods to adapt a pre-trained model have been considered. LLMs can involve re-training through plug-and-play attribute classifiers (Dathathri et al. 2020); re-training to produce paragraphs coherent with a given outline (Rashkin et al. 2020); fine-tuning with specific corpora for writing specific text (Sawicki et al. 2022; Wertz and Kuhn 2022); or fine-tuning to maximize human preferences (Ziegler et al. 2019) or to generate specific literary outputs, such as poetry (Pardinas 2023). Nevertheless, the recent advancements

[5] http://www.in-vacua.com/cgi-bin/haiku.pl.

in LLMs can be attributed to the introduction of fine-tuning through reinforcement learning from human feedback (RLHF) (Christiano et al. 2017). It consists of three steps: fine-tuning the pre-trained model in a supervised fashion on human-produced answers to sampled questions; training a reward model to predict which text among different options is the most appropriate based on human-labeled rankings; and fine-tuning the language model to maximize the learned reward (Stiennon et al. 2020). Although the main goal of RLHF is to improve conversational skills while mitigating mistakes and biases, it has also led to models capable of producing on-demand poems, songs, and novels, gaining global popularity.[6] Based on RLHF, first ChatGPT[7] and then GPT-4 paved the way for several other similar models: Google's Gemini (Gemini Team and Google 2023), which extends to multimodal data; Meta's Llama models (Dubey et al. 2024; Touvron et al. 2023), which replace RLHF with the more efficient direct preference optimization (DPO) (Rafailov et al. 2023); Mixtral (Jiang et al. 2024), which adaptively selects its layers' parameters from distinct groups to increase the total parameter count without raising computational costs; and many others, as the competition intensifies day by day (Zhao et al. 2023). While they may differ in some technical details, these LLMs are always pre-trained on vast, general corpora of data and then fine-tuned using some form of RLHF to enhance their conversational skills.

3 Large language models and Boden's three criteria

Margaret Boden defines creativity as "the ability to come up with ideas or artifacts that are *new*, *surprising* and *valuable*" (Boden 2003). In other words, Boden implicitly derives criteria that can be used to identify a creative *product*. They suggest that creativity is about *novelty*, *surprise* and *value*. We will refer to them as Boden's three criteria. In the following, we will analyze to what extent state-of-the-art LLMs satisfy them and we will question if LLMs can be really considered creative.

Value refers to utility, performance, and attractiveness (Maher 2010). It is also related to both the quality of the output, and its acceptance by society. Due to the large impact LLMs are already having (Bommasani et al. 2021) and the quality of outputs of the systems based on them (Stevenson et al. 2022a), it is possible to argue that the artifacts produced by them are indeed valuable.

Novelty refers to the dissimilarity between the produced artifact and other examples in its class (Ritchie 2007). However, it can also be seen as the property of not being in existence before. This is considered in reference to either the person who came up with it or the entire human history. The former is referred to as psychological creativity (shortened as *P-creativity*), whereas the latter is historical creativity (shortened as *H-creativity*) (Boden 2003). While the difference appears negligible, it is substantial when discussing LLMs in general. Considering these definitions, a model writing a text that is not in its training set would be considered as P-novel, but possibly also H-novel, since LLMs are commonly trained on all available data. Their stochastic nature and the variety of prompts that are usually provided commonly lead to novel outcomes (McCoy et al. 2023); LLMs may therefore be capable of generating artifacts that are also new. However, one should remember how such models learn and generate. LLMs still play a sort of *imitation game*, without a focus on (computational) novelty (Fazi 2019). Even if prompted with the sentence "I wrote a new poem this morning:", they would nonetheless complete it with what is most likely to follow such words, e.g., something close to what others have written in the past (Shanahan 2024). It is a probabilistic process after all. The degree of dissimilarity would therefore be small *by design*. High values of novelty would be caused either by accidental, out-of-distribution productions or by careful prompting, i.e., one that would place the LLM in a completely unusual or unexpected (i.e., novel) situation.

Surprise instead refers to how much a stimulus disagrees with expectation (Berlyne 1971). It is possible to identify three kinds of surprise, which correspond to three different forms of creativity. *Combinatorial creativity* involves making unfamiliar combinations of familiar ideas. *Exploratory creativity* requires finding new, unexplored solutions inside the current style of thinking. *Transformational creativity* is related to changing the current style of thinking (Boden 2003). These three different forms of creativity involve surprise at increasing levels of abstraction: combining existing elements, exploring new elements coherent with the current state of the field, and transforming the state of the field to introduce other elements. The autoregressive nature of classic LLMs makes them unlikely to generate surprising products (Bunescu and Uduehi 2019) since they are essentially trained to follow the current data distribution (Shanahan 2024). By relying only on given distributions and being trained on them, LLMs might at most express combinatorial or exploratory creativity. Of course, specific different solutions may be generated by means of prompting or conditioning. For instance, recent LLMs can write poems about mathematical theories, a skill that requires the application of a certain existing style to a given topic, yet leading to new and unexplored solutions. However, the result would hardly

[6] https://www.forbes.com/sites/martineparis/2023/02/03/chatgpt-hits-100-million-microsoft-unleashes-ai-bots-and-catgpt-goes-viral/?sh=70994247564e.

[7] https://openai.com/blog/chatgpt/.

be unexpected for whom has prompted the text. For an external reader, the surprise would probably come from the idea of mathematical theories in verses, which is due to the user (or by the initial astonishment of a machine capable of it (Waite 2019)). Transformational creativity is not achievable through the current LLM training solutions. In theory, other forms of training or fine-tuning might circumvent this limitation, allowing the model to forget the learned rules to forge others. However, this is not the case with current models. ChatGPT and all the other state-of-the-art LLMs introduced in Sect. 2 are fine-tuned with RLHF or DPO. While in theory this could lead to potentially surprising generation, its strict alignment to very careful and pre-designed human responses leads to the generation of text that tends to be less diverse (Kirk et al. 2024) and that might be considered *banal* (Hoel 2022).

Nonetheless, the outputs from such models are often considered creative by the person interacting with them or exposed to their best productions. Though this is apparently in contrast with what was discussed above, we can explain this phenomenon by considering the fact that our perception does not usually align with theoretical definitions of creativity. Indeed, we do not typically judge the creativity of a product by considering its potential novelty and surprise in relation to its producer, but rather in relation to ourselves. Something can be new for the beholder, leading to a new kind of novelty which we call *B-novelty*, as it is the one "in the eye of the beholder", but not new for the producer nor the entire human history. The same applies to surprise: a product can violate the observer's expectations in many ways without being unexpected considering the entire domain. In other words, the product of an LLM can appear to be creative—or be B-creative—even if it is not *truly* creative according to the theory of creativity.

In conclusion, while LLMs are capable of producing artifacts that are valuable, achieving P- or H-novelty and surprise appears to be more challenging. It is possible to argue that LLMs may be deemed able to generate creative products if we assume the definition of combinatorial creativity. To achieve transformational creativity, alternative learning architectures are probably necessary; in fact, current probabilistic solutions are intrinsically limiting in terms of expressivity. We believe that this is a fundamental research area for the community for the years to come.

4 Easy and hard problems in machine creativity

LLMs might be able to generate creative products in the future. However, the fact that they will be able to generate these outputs will not make them intrinsically creative. Indeed, as Floridi and Chiriatti (2020) puts it, it is not *what* is achieved but *how* it is achieved that matters. An interesting definition that considers both the *what* and *how* dimensions is the one from Gaut (2003): creativity is the capacity to produce original and valuable items by *flair*. Exhibiting flair means exhibiting a relevant purpose, understanding, judgment, and evaluative abilities. Such properties are highly correlated with those linked with *process*, i.e., motivation, perception, learning, thinking, and communication (Rhodes 1961). Motivation is a crucial part of creativity, as it is the first stage of the process. Usually, it comes from an intrinsic interest in the task, i.e., the activity is interesting and enjoyable for its own sake (Deci and Ryan 1985). However, LLMs lack the intention to write. They can only deal with "presented" problems, which are less conducive to creativity (Amabile 1996). The process continues with the preparation step (reactivating the store of relevant information and response algorithms), the response generation, and its validation and communication (Amabile 1983). The last two steps allow one to produce different response possibilities and to internally test them in order to select the most appropriate. Again, LLMs do not contain such a self-feedback loop. At the same time, they are not trained to directly maximize value, novelty, or surprise. They only output content that is likely to follow given a stimulus in input (Shanahan 2024). In other words, they stop at the first stage of creative learning, i.e., imitation, not implementing the remaining ones, i.e., exploration and intentional deviation from conventions (Riedl 2018).

However, paraphrasing Chalmers (Chalmers 1996), these appear as *easy* problems to solve to achieve creativity, since solutions to them can be identified by taking into consideration the underlying training and inference processes. The *hard* problem in machine creativity is about the intentionality and the self-awareness of the creative process in itself. Even though the intent of running the LLM may be achieved by its outcome, it is in an unintentional way (Terzidis et al. 2022); as current generative AI models are only causal, and not intentional, agents (Johnson and Verdicchio 2019). Indeed, a crucial aspect of the creative process is the perception and the ability of *self-evaluating* the generated outputs (Amabile 1983). This can be seen as a form of creative self-awareness. While not strictly necessary to generate a response, this ability is essential to self-assess its quality, so as to correct it or to learn from it. However, no current LLM is able to self-evaluate its own responses. LLMs can in theory recognize certain limitations of their own texts after generating them, e.g., by ranking them (Franceschelli and Musolesi 2024a) or by assigning quality- and diversity-based scores (Bradley et al. 2024). Then, they can try to correct, modify, or rephrase the outputs if asked to do so (i.e., through an external intervention). However, they would do it only by guessing what is the most likely re-casting of such responses or through the application of a set of given rules. It is worth noting that this is something distinct from the problem of the potential emergence of the theory of mind in these systems (Bubeck et al. 2023).

Indeed, product and process are not sufficient to explain creativity. Rhodes (1961) theorizes that four perspectives have to be considered: product (see Sect. 3) and process (discussed above), but also the so-called *press* and *person*. Press refers to the relationship between the product and the influence its environment has upon it (Rhodes 1961). Individuals and their works cannot be isolated from the social and historical milieu in which their actions are carried out. Products have to be accepted as creative by society, and producers are influenced by the previously accepted works, i.e., the domain (Csikszentmihalyi 1988). The resulting system model of creativity is a never-ending cycle where individuals always base their works on knowledge from a domain, which constantly changes thanks to new and valuable artifacts (from different individuals). For example, individuals generate new works based on the current domain; the field (i.e., critics, other artists, the public, etc.) decides which of those works are worth promoting and preserving; the domain is expanded and, possibly, transformed by these selected works; individuals generate new works based on the updated current domain; and then this cycle repeats.

However, LLMs cannot currently adapt through multiple iterations in the way described above; they just rely on one, fixed version of the domain and generate works based on it. The current generation of LLMs are *immutable* entities, i.e., once the training is finished, they remain frozen reflecting a specific state of the domain. In other words, they are not able to adapt to new changes. In-context learning can simulate an adaptation to new states of the domain. The constantly increasing context length (Hsieh et al. 2024) allows researchers to provide more and more information to LLMs without re-training them, although a longer context might lead to performance degradation (Li et al. 2024). This enables the representation of the current state of the domain through an adequate prompt, allowing the model to generate different outputs according to environmental changes. For example, in Park et al. (2023), multiple LLM-based agents interact through natural language in a sandbox environment inspired by *The Sims*. Each agent stores, synthesizes, and applies relevant memories to generate believable behavior through in-context learning, leading to emergent social behaviors. The study of emergent behaviors of LLM-based agents at the population level is an active research area (Guo et al. 2024). It is easy to imagine the simulation of creative or artistic environments, such as a virtual multi-agent translation company (Wu et al. 2024), as well.

However, LLMs are like the main character of *Memento*: they always possess all the capabilities, but each time they “wake up”, they need to re-collect all the information about themselves and their world. The time—or space—to acquire such information is limited, and by the next day, they will have forgotten it all. In other words, these generative agents do not truly adapt or learn new things about the changing domain. Placing them in a different environment that requires a different prompt will make them start over, without the possibility of leveraging previously acquired experience.

On the other hand, fine-tuning actually updates network weights, but it requires a potentially large training dataset. Indeed, several current research efforts are in the direction of introducing adaptation for specific domains, tasks, cultural frameworks, and so on. To be able to be part of the never-ending creative cycle mentioned above, LLMs should constantly adapt. Continual learning (Kirkpatrick et al. 2017; Shin et al. 2017) for LLMs (Sun et al. 2020; Wu et al. 2022) represents a promising direction, yet unexplored for creative applications.

Finally, the person perspective covers information about personality, intellect, temperament, habits, attitude, value systems, and defense mechanisms (Rhodes 1961). While several of the properties of press and process might be achieved—or at least simulated—by generative learning solutions, those related to the creative person appear out of discussion (Browning 2023). Several works have analyzed whether LLMs can pass tests intended to evaluate human psychological skills (Binz and Schulz 2023; Macmillan-Scott and Musolesi 2024; Stevenson et al. 2022b), sometimes with promising results (Kosinski 2024; Lampinen et al. 2024). However, according to the best-supported neuroscientific theories of consciousness, current AI systems are not conscious (Butlin et al. 2023). As Ressler (2023) pointed out, LLMs have no self to which to be true when generating text and are intrinsically unable to behave authentically as individuals. They merely “play the role” of a character or, more accurately, a superposition of simulacra within a multiverse of possible characters induced by their training (Shanahan et al. 2023; Shanahan 2024a). This results in a perceived self-awareness, stemming from our inclination to anthropomorphize (Deshpande et al. 2023; Seth 2021). In conclusion, all the properties listed above require some forms of consciousness and self-awareness, which are difficult to define in themselves and are related to the *hard* problem introduced before. Creative-person qualities in generative AI might eventually be the ultimate step in achieving human-like intelligence.

5 Practical implications

The application of large language models to fields like literature or journalism opens up a series of practical questions. Since LLMs can be used to produce artifacts that would be protected if made by humans, a first concern is the definition of legal frameworks in which they will be used. Copyright for generative AI is currently a hotly debated topic (Guadamuz 2017; Franceschelli and Musolesi 2022; Lee et al. 2024; Miernicki 2021), due to the fact that current laws do not contemplate works produced by non-human beings (with few notable exceptions (Bond and Blair 2019)). Copyright applies to creative *works of authorship* (as referred to in the US Copyright Code), i.e., works showing a minimum degree of originality (Gervais 2002) and reflecting author's personality (Deltorn 2017). As discussed earlier, current LLMs might satisfy the first condition, but they cannot be considered creative persons, therefore missing the latter requirement. For this reason, works produced by LLMs can be protected if and only if the original contribution is provided by a human, e.g., the user who writes the prompt that is used as input of the model, who in turn will be the rights holder. The definition of the criteria for classifying a source of originality is a fundamental problem since there is a clear need to discriminate between protected and publicly available works.

While a higher degree of novelty is unnecessary for claiming protection, it might be crucial for other legal aspects. In particular, LLMs are trained in a supervised fashion on real data, which also include protected works (Bandy and Vincent 2021). Apart from questions upon the legitimacy of such training (Franceschelli and Musolesi 2022), LLMs may learn to reproduce portions of them (Liang et al. 2023) because of the memorization of training data (Carlini et al. 2023). This would violate their reproduction or adaptation right (Bonadio and McDonagh 2020). A different, creative-oriented training approach should mitigate such risk, also facilitating fair-use doctrine application (Asay et al. 2020).

Whether or not LLM works obtain protection, we believe their societal impact will be tremendous (see also Newton and Dhole (2023)). We have a positive view in terms of the applications of LLMs, but there are intrinsic risks related to their adoption. It is apparent that since LLMs are able to write articles or short stories, as the quality of their outputs gets better and better, there is the risk that certain jobs in the professional writing industry will essentially disappear (Ponce Del Castillo 2023; Tamkin et al. 2021). However, we must remind that current LLMs are not as reliable as humans, e.g., they cannot verify their information and they can propagate biases from training data. In addition, the quality of the output strictly depends on the prompt, which might in turn demand human skills and more time. Writers can be threatened as well. Though not in violation of copyright, LLMs may exploit certain ideas from human authors, capitalizing on their efforts in ways that are less expensive or time-consuming (Weidinger et al. 2022). The questionable creative nature of LLMs discussed so far might suggest artificial works to be of less quality than humans, therefore not providing a real threat. Nonetheless, more creative LLMs would diverge more consistently from existing works, reducing the risk of capitalizing on others' ideas. The lack of current copyright protection for generated works can also foster such replacements for tasks where a free-of-charge text would be preferable to a high-quality (but still costly) one. Finally, one last threat may be posed by human and artificial works being indistinguishable (Dehouche 2021). The users obtaining such outputs might therefore claim them as the authors, e.g., for deceiving readers (Grinbaum and Adomaitis 2022), for cheating during exams (Fyfe 2023), or for improving bibliometric indicators (Crothers et al. 2023). Mitigation of such threats through dedicated policies[8] or designed mechanisms of watermarks (Kirchenbauer et al. 2023) are already being developed.

However, as we said, we believe that, overall, the impact of these technologies will be positive. LLMs also provide several opportunities for creative activities. Given their characteristics, humans are still required, especially for prompting, curation, and pre-/post-production. This means that the role of writers and journalists may be transformed, but not replaced. On the contrary, LLMs provide new opportunities for humans, who will be able to spend more time validating news or thinking up and testing ideas. LLMs can also adapt the same text to different styles (see combinatorial creativity in Sect. 3): by doing so, an artifact can be adapted to reach wider audiences. In the same way, LLMs also represent a valuable tool in scientific research (Fecher et al. 2023), especially for hypothesis generation (Gero et al. 2022).

Indeed, we believe that LLMs can also foster human-AI co-creativity (Lee et al. 2022), since they can be used to write portions of stories to serve specific purposes, e.g., they can typify all the dialogues from a character, or they can provide more detailed descriptions of scenes (Calderwood et al. 2020). Dialogue systems based on LLMs can be used for brainstorming. In the same way, the generated responses may augment writers' inherently multiversal imagination (Reynolds and McDonell 2021). LLMs can also represent a source of inspiration for plot twists, metaphors (Chakrabarty et al. 2023), or even entire story plans (Mirowski et al. 2022), even though they sometimes appear to fail in accomplishing these tasks at human-like level (Ippolito et al. 2022). Being intrinsically powerful tools, through human-AI co-creation, LLMs may eventually allow the development of entire new arts, as has been the case for any impactful technology in the past centuries (Eisenstein 1979; Silva 2022).

[8] https://bigscience.huggingface.co/blog/the-bigscience-rail-license.

6 Conclusion

The latest generation of LLMs is attracting increasing interest from both AI researchers and the general public due to the astonishing quality of their productions. Questions naturally arise around the actual creativity of these technologies. In this paper, we have discussed whether or not LLMs can actually be deemed as creative; we started by considering Boden's three criteria, i.e., value, novelty, and surprise. While LLMs are capable of value and a weak version of novelty and surprise, their inner autoregressive nature seems to prevent them from reaching transformational creativity. Then, we have examined perspectives beyond the creativity of their products. A creative process would require motivation, thinking, and perception, properties that current LLMs do not possess. The social dimension of creativity (usually referred to as the press) would demand to be placed in and influenced by a society of creative agents, requiring LLMs adaptive abilities that are only at a very initial stage. We have also framed the problem of creativity in LLMs, and, more in general, machine creativity, in terms of easy problems, i.e., the technical advancements that will be needed to support the algorithmic generation of outputs and the intrinsic hard problem of introducing forms of self-awareness in the creation process itself.

In addition, we have also investigated the practical implications of LLMs and their creative role, considering both legal and societal impacts. In fact, the current legal framework does not appear to be completely suited to the fast-moving field of generative AI. Moreover, the impact of these technologies on creative professions and the arts is difficult to forecast at this stage, but will definitely be considerable. However, LLMs also provide opportunities for writers, especially in terms of human-AI cooperation. Specific fine-tuning techniques might help LLMs diversify productions and explore the conceptual space they learn from data. Continual learning can enable long-term deployments of LLMs in a variety of contexts. While, of course, all these techniques would only simulate certain aspects of creativity, whether this would be sufficient to achieve artificial, i.e., non-human, creativity, is up to the humans themselves.

Data availability Not applicable.

Declarations

Conflict of interest The authors declare they do not have any conflict of interest.

References

Amabile TM (1983) The social psychology of creativity: a componential conceptualization. J Personal Soc Psychol 45(2):357–376

Amabile TM (1996) Creativity in context. Routledge, London

Asay CD, Sloan A, Sobczak D (2020) Is transformative use eating the world? Boston Coll Law Rev 61(3):905–970

Babbage C (1864) Of the analytical engine. In: Passages from the life of a philosopher, vol 3. Longman, Green, Longman, Roberts, & Green, pp 112–141

Bahdanau D, Cho K, Bengio Y (2015) Neural machine translation by jointly learning to align and translate. In: Proc. of the 3rd international conference on learning representations (ICLR'15)

Bandy J, Vincent N (2021) Addressing "documentation debt" in machine learning: a retrospective datasheet for bookcorpus. In: Proc. of the 35th conference on neural information processing systems datasets and benchmarks track (round 1)

Barron F (1955) The disposition toward originality. J Abnorm Psychol 51(3):478–485

Bender EM, Koller A (2020) Climbing towards NLU: on meaning, form, and understanding in the age of data. In: Proc. of the 58th annual meeting of the association for computational linguistics (ACL'20)

Bender EM, Gebru T, McMillan-Major A, Shmitchell S (2021) On the dangers of stochastic parrots: can language models be too big? In: Proc. of the 2021 ACM conference on fairness, accountability, and transparency (FAccT'21)

Berlyne DE (1960) Conflict, arousal, and curiosity. McGraw-Hill Book Company, New York

Berlyne DE (1971) Aesthetics and psychobiology. Appleton-Century-Crofts, New York

Binz M, Schulz E (2023) Using cognitive psychology to understand GPT-3. Proc Natl Acad Sci 120(6):e2218523120

Boden MA (2003) The creative mind: myths and mechanisms. Routledge, London

Boden MA (2009) Computer models of creativity. AI Mag 30(3):23–34

Bommasani R, Hudson DA, Adeli E, Altman R, Arora S, von Arx S, Bernstein MS, Bohg J, Bosselut A, Brunskill E, Brynjolfsson E, Buch S, Card D, Castellon R, Chatterji N, Chen A, Creel K, Davis JQ, Demszky D, Donahue C et al (2021) On the opportunities and risks of foundation models. arXiv:2108.07258 [cs.LG]

Bonadio E, McDonagh L (2020) Artificial intelligence as producer and consumer of copyright works: evaluating the consequences of algorithmic creativity. Intellect Prop Q 2020(2):112–137

Bond T, Blair S (2019) Artificial intelligence & copyright: section 9(3) or authorship without an author. J Intellect Prop Law Pract 14(6):423–423

Bowman SR, Vilnis L, Vinyals O, Dai AM, Jozefowicz R, Bengio S (2016) Generating sentences from a continuous space. In: Proc. of the 20th SIGNLL conference on computational natural language learning (CoNNL'16)

Bradley H, Dai A, Teufel H, Zhang J, Oostermeijer K, Bellagente M, Clune J, Stanley K, Schott G, Lehman J (2024) Quality-diversity through AI feedback. In: Proc. of the 12th international conference on learning representations (ICLR'24)

Brown T, Mann B, Ryder N, Subbiah M, Kaplan JD, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krueger G, Henighan T, Child R, Ramesh A, Ziegler D, Wu J, Winter C et al (2020) Language models are few-shot learners. In: Advances in neural information processing systems (NIPS'20)

Browning J (2023) Personhood and ai: why large language models don't understand us. AI & Soc 39:2499–2506

Bruner JS (1962) The conditions of creativity. In: Contemporary approaches to creative thinking: a symposium held at the University of Colorado. Atherton Press, pp 1–30

Bubeck S, Chandrasekaran V, Eldan R, Gehrke J, Horvitz E, Kamar E, Lee P, Lee YT, Li Y, Lundberg S et al (2023) Sparks of artificial general intelligence: early experiments with GPT-4. arXiv:2303.12712 [cs.CL]

Bunescu RC, Uduehi OO (2019) Learning to surprise: a composer-audience architecture. In: Proc. of the 10th international conference on computational creativity (ICCC'19)

Butlin P, Long R, Elmoznino E, Bengio Y, Birch J, Constant A, Deane G, Fleming SM, Frith C, Ji X, Kanai R, Klein C, Lindsay G, Michel M, Mudrik L, Peters MAK, Schwitzgebel E, Simon J, VanRulle R (2023) Consciousness in artificial intelligence: insights from the science of consciousness. arXiv:2308.08708 [cs.AI]

Calderwood A, Qiu V, Gero KI, Chilton LB (2020) How novelists use generative language models: an exploratory user study. In: Proc. of the IUI'20 workshop on human-AI co-creation with generative models

Carlini N, Ippolito D, Jagielski M, Lee K, Tramer F, Zhang C (2023) Quantifying memorization across neural language models. In: Proc. of the 11th international conference on learning representations (ICLR'23)

Chakrabarty T, Padmakumar V, He H (2023) Help me write a poem: instruction tuning as a vehicle for collaborative poetry writing. In: Proc. of the AAAI-23 workshop on creative AI across modalities

Chalmers DJ (1996) The conscious mind: in search of a fundamental theory. Oxford University Press, Oxford

Cho, K, van Merrienboer B, Bahdanau D, Bengio Y (2014) On the properties of neural machine translation: encoder–decoder approaches. In: Proc. of 8th workshop on syntax, semantics and structure in statistical translation (SSST'08)

Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, Barham P, Chung HW, Sutton C, Gehrmann S, Schuh P, Shi K, Tsvyashchenko S, Maynez J, Rao A, Barnes P, Tay Y, Shazeer N, Prabhakaran V, Fiedel N (2023) PaLM: scaling language modeling with pathways. J Mach Learn Res 24(240):1–113

Christiano PF, Leike J, Brown T, Martic M, Legg S, Amodei D (2017) Deep reinforcement learning from human preferences. In: Advances in neural information processing systems (NIPS'17)

Colton S, Wiggins GA (2012) Computational creativity: the final frontier? In: Proc. of the 20th European conference on artificial intelligence (ECAI'12), vol 12

Crothers EN, Japkowicz N, Viktor HL (2023) Machine-generated text: a comprehensive survey of threat models and detection methods. IEEE Access 11:70977–71002

Csikszentmihalyi M (1988) Society, culture, and person: a systems view of creativity. The nature of creativity: contemporary psychological perspectives. Cambridge University Press, Cambridge, pp 325–339

Dathathri S, Madotto A, Lan J, Hung J, Frank E, Molino P, Yosinski J, Liu R (2020) Plug and play language models: a simple approach to controlled text generation. In: Proc. of the 8th international conference on learning representations (ICLR'20)

Deci EL, Ryan RM (1985) Intrinsic motivation and self-determination in human behavior. Springer, Berlin

Dehouche N (2021) Plagiarism in the age of massive generative pre-trained transformers (GPT-3). Ethics Sci Environ Polit 21:17–23

Deltorn JM (2017) Deep creations: intellectual property and the automata. Front Digit Humanit 4(3):1–13

Deshpande A, Rajpurohit T, Narasimhan K, Kalyan A (2023) Anthropomorphization of AI: opportunities and risks. In: Proc. of the natural legal language processing workshop 2023

Devlin J, Chang MW, Lee K, Toutanova K (2019) BERT: pre-training of deep bidirectional transformers for language understanding. In: Proc. of the 2019 conference of the North American Chapter of the Association for Computational Linguistics: human language technologies, vol 1 (Long and Short Papers) (NAACL'19)

Dong Q, Li L, Dai D, Zheng C, Ma J, Li R, Xia H, Xu J, Wu Z, Chang B, Sun X, Li L, Sui Z (2024) A survey on in-context learning. arXiv:2301.00234 [cs.CL]

Du N, Huang Y, Dai AM, Tong S, Lepikhin D, Xu Y, Krikun M, Zhou Y, Yu AW, Firat O, Zoph B, Fedus L, Bosma M, Zhou Z, Wang T, Wang YE, Webster K, Pellat M, Robinson K et al (2022) GLaM: efficient scaling of language models with mixture-of-experts. In: Proc. of the 39th international conference on machine learning (ICML'22)

Dubey A, Jauhri A, Pandey A, Kadian A, Al-Dahle A, Letman A, Mathur A, Schelten A, Yang A, Fan A, Goyal A, Hartshorn A, Yang A, Mitra A, Sravankumar A, Korenev A, Hinsvark A, Rao A, Zhang A et al (2024) The Llama 3 herd of models. arXiv:2407.21783 [cs.AI]

Eisenstein E (1979) The printing press as an agent of change: communications and cultural transformations in early-modern Europe. Cambridge University Press, Cambridge

Fazi MB (2019) Can a machine think (anything new)? Automation beyond simulation. AI & Soc 34(4):813–824

Fecher B, Hebing M, Laufer M, Pohle J, Sofsky F (2023) Friend or foe? Exploring the implications of large language models on the science system. AI & Soc. Accepted for publication

Floridi L, Chiriatti M (2020) GPT-3: its nature, scope, limits, and consequences. Mind Mach 30(4):681–694

Franceschelli G, Musolesi M (2022) Copyright in generative deep learning. Data Policy 4:e17

Franceschelli G, Musolesi M (2024a) Creative beam search: LLM-as-a-judge for improving response generation. In: Proc. of the 15th international conference on computational creativity (ICCC'24)

Franceschelli G, Musolesi M (2024b) Creativity and machine learning: a survey. ACM Comput Surv 56(11):1–41 (Article No.: 283)

Fyfe P (2023) How to cheat on your final paper: assigning ai for student writing. AI & Soc 38(4):1395–1405

Gaut B (2003) Creativity and imagination. The creation of art: new essays in philosophical aesthetics. Cambridge University Press, Cambridge, pp 148–173

Gaut B (2010) The philosophy of creativity. Philos Compass 5(12):1034–1046

Gemini Team and Google (2023) Gemini: a family of highly capable multimodal models. arXiv:2312.11805 [cs.CL]

Gero KI, Liu V, Chilton L (2022) Sparks: inspiration for science writing using language models. In: Proc. of the 2022 designing interactive systems conference (DIS'22)

Gervais DJ (2002) Feist goes global: a comparative analysis of the notion of originality in copyright law. J Copyr Soc USA 49:949–981

Gervás P (2013) Computational modelling of poetry generation. In: Symposium on artificial intelligence and poetry (AISB'13)

GPT-3, Thunström AO, Steingrimsson S (2022) Can GPT-3 write an academic paper on itself, with minimal human input? https://hal.archives-ouvertes.fr/hal-03701250v1. Accessed 30 Oct 2024

Grinbaum A, Adomaitis L (2022) The ethical need for watermarks in machine-generated language. arXiv:2209.03118 [cs.CL]

Guadamuz A (2017) Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works. Intellect Prop Q 2:1–24

Guo T, Chen X, Wang Y, Chang R, Pei S, Chawla NV, Wiest O, Zhang X (2024) Large language model based multi-agents: a survey of progress and challenges. In: Proc. of the 33rd international joint conference on artificial intelligence (IJCAI'24)

Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780

Hoel E (2022) The banality of ChatGPT. https://www.theintrinsicperspective.com/p/the-banality-of-chatgpt. Accessed 30 Oct 2024

Hoffmann J, Borgeaud S, Mensch A, Buchatskaya E, Cai T, Rutherford E, de Las Casas D, Hendricks LA, Welbl J, Clark A, Hennigan T, Noland E, Millican K, van den Driessche G, Damoc B, Guy A, Osindero S, Simonyan K, Elsen E, Rae, JW, Vinyals O, Sifre L (2022) Training compute-optimal large language models. In: Advances in neural information processing systems (NIPS'22)

Hsieh CP, Sun S, Kriman S, Acharya S, Rekesh D, Jia F, Zhang Y, Ginsburg B (2024) RULER: what's the real context size of your long-context language models? In: Proc. of the 1st conference on language modeling (COLM'24)

Ippolito D, Yuan A, Coenen A, Burnam S (2022) Creative writing with an ai-powered writing assistant: perspectives from professional writers. arXiv:2211.05030 [cs.HC]

Jiang AQ, Sablayrolles A, Roux A, Mensch A, Savary B, Bamford C, Chaplot DS, de las Casas D, Hanna EB, Bressand F, Lengyel G, Bour G, Lample G, Lavaud LR, Saulnier L, Lachaux MA, Stock P, Subramanian S, Yang S, Antoniak S et al (2024) Mixtral of experts. arXiv:2401.04088 [cs.LG]

Johnson DG, Verdicchio M (2019) AI, agency and responsibility: the VW fraud case and beyond. AI & Soc 34(3):639–647

Jordanous AK (2009) Evaluating machine creativity. In: Proc. of the seventh ACM conference on creativity and cognition (C &C'09)

Jurafsky D, Martin JH (2023) Speech and language processing (third (draft) ed.)

Karpathy A (2015) The unreasonable effectiveness of recurrent neural networks. http://karpathy.github.io/2015/05/21/rnn-effectiveness/. Accessed 30 Oct 2024

Kirchenbauer J, Geiping J, Wen Y, Katz J, Miers I, Goldstein T (2023) A watermark for large language models. In: Proc. of the 40th international conference on machine learning (ICML'23)

Kirk R, Mediratta I, Nalmpantis C, Luketina J, Hambro E, Grefenstette E, Raileanu R (2024) Understanding the effects of RLHF on LLM generalisation and diversity. In: Proc. of the 12th international conference on learning representations (ICLR'24)

Kirkpatrick J, Pascanu R, Rabinowitz N, Veness J, Desjardins G, Rusu AA, Milan K, Quan J, Ramalho T, Grabska-Barwinska A, Hassabis D, Clopath C, Kumaran D, Hadsell R (2017) Overcoming catastrophic forgetting in neural networks. Proc Natl Acad Sci 114(13):3521–3526

Kosinski M (2024) Evaluating large language models in theory of mind tasks. arXiv:2302.02083 [cs.CL]

Lampinen AK, Dasgupta I, Chan SCY, Sheahan HR, Creswell A, Kumaran D, McClelland JL, Hill F (2024) Language models, like humans, show content effects on reasoning tasks. PNAS Nexus 3(7):233

Lebowitz M (1983) Creating a story-telling universe. In: Proc. of the 8th international joint conference on artificial intelligence (IJCAI'83)

Lee HH, Shu K, Achananuparp P, Prasetyo PK, Liu Y, Lim EP, Varshney LR (2020) RecipeGPT: generative pre-training based cooking recipe generation and evaluation system. In: Companion proceedings of the web conference 2020 (WWW'20)

Lee M, Liang P, Yang Q (2022) CoAuthor: designing a human-AI collaborative writing dataset for exploring language model capabilities. In: Proc. of the 2022 CHI conference on human factors in computing systems (CHI'22)

Lee, K, Cooper AF, Grimmelmann J (2024) Talkin' 'bout AI generation: copyright and the generative-AI supply chain. arXiv:2309.08133 [cs.CY]

Li T, Zhang G, Do QD, Yue X, Chen W (2024) Long-context LLMs struggle with long in-context learning. arXiv:2404.02060 [cs.CL]

Liang P, Bommasani R, Lee T, Tsipras D, Soylu D, Yasunaga M, Zhang Y, Narayanan D, Wu Y, Kumar A, Newman B, Yuan B, Yan B, Zhang C, Cosgrove C, Manning CD, Ré C, Acosta-Navas D, Hudson DA et al (2023) Holistic evaluation of language models. Trans Mach Learn Res

Liu J, Shen D, Zhang Y, Dolan B, Carin L, Chen W (2022) What makes good in-context examples for GPT-3? In: Proc. of deep learning inside out (DeeLIO22): the 3rd workshop on knowledge extraction and integration for deep learning architectures

Macedo L, Reisenzein R, Cardoso A (2004) Modeling forms of surprise in artificial agents: empirical and theoretical study of surprise functions. In: Proc. of the annual meeting of the Cognitive Science Society (CogSci'04)

Macmillan-Scott O, Musolesi M (2024) (Ir)rationality and cognitive biases in large language models. Roy Soc Open Sci 11(6):240255

Maher ML (2010) Evaluating creativity in humans, computers, and collectively intelligent systems. In: Proc. of the 1st DESIRE Network Conference on Creativity and Innovation in Design (DESIRE'10)

Manurung R, Ritchie G, Thompson H (2012) Using genetic algorithms to create meaningful poetic text. J Exp Theor Artif Intell 24(1):43–64

Martin LJ, Ammanabrolu P, Hancock W, Singh S, Harrison B, Riedl MO (2018) Event representations for automated story generation with deep neural nets. In: Proc. of the 32nd AAAI conference on artificial intelligence and 30th innovative applications of artificial intelligence conference and 8th AAAI symposium on educational advances in artificial intelligence (AAAI'18/IAAI'18/EAAI'18)

McCoy RT, Smolensky P, Linzen T, Gao J, Celikyilmaz A (2023) How much do language models copy from their training data? Evaluating linguistic novelty in text generation using RAVEN. Trans Assoc Comput Linguist 11:652–670

Meehan JR (1977) TALE-SPIN, an interactive program that writes stories. In: Proc. of the 5th international joint conference on artificial intelligence (IJCAI'77)

Menabrea LF, Lovelace A (1843) Sketch of the analytical engine invented by Charles Babbage. In: Scientific memoirs, vol 3. Richard and John E. Taylor, pp 666–731

Miernicki M (2021) Artificial intelligence and moral rights. AI & Soc 36(1):319–329

Miller AI (2019) The artist in the machine. The MIT Press, Cambridge

Mirowski P, Mathewson KW, Pittman J, Evans R (2022) Co-writing screenplays and theatre scripts alongside language models using Dramatron. In: Proc. of the NIPS'22 workshop on ml for creativity & design

Newell A, Shaw JC, Simon HA (1962) The processes of creative thinking. In: Contemporary approaches to creative thinking: a symposium held at the University of Colorado. Atherton Press, pp 63–119

Newton A, Dhole K (2023) Is AI art another industrial revolution in the making? In: Proc. of the AAAI-23 workshop on creative AI across modalities

Pardinas R, Huang G, Vazquez D, Piché A (2023) Leveraging human preferences to master poetry. In: Proc. of the AAAI-23 workshop on creative AI across modalities

Park JS, O'Brien JC, Cai CJ, Morris MR, Liang P, Bernstein MS (2023) Generative agents: interactive simulacra of human behavior. In: Proc. of the 36th annual ACM symposium on user interface software and technology (UIST'23)

Peng H, Wang X, Chen J, Li W, Qi Y, Wang Z, Wu Z, Zeng K, Xu B, Hou L, Li J (2023) When does in-context learning fall short and why? a study on specification-heavy tasks. arXiv:2311.08993 [cs.CL]

Ponce Del Castillo A (2023) Generative AI, generating precariousness for workers? AI & Soc 39:2601–2602

Potash P, Romanov A, Rumshisky A (2015) GhostWriter: using an LSTM for automatic rap lyric generation. In: Proc. of the 2015 conference on empirical methods in natural language processing (EMNLP'15)

Racter (1984) The policeman's beard is half constructed. Warner Books, Inc., New York

Radford A, Wu J, Child R, Luan D, Amodei D, Sutskever I (2019) Language models are unsupervised multitask learners. https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf. Accessed 30 Oct 2024

Rae JW, Borgeaud S, Cai T, Millican K, Hoffmann J, Song F, Aslanides J, Henderson S, Ring R, Young S, Rutherford E, Hennigan T, Menick J, Cassirer A, Powell R, van den Driessche G, Hendricks LA, Rauh M, Huang PS et al (2021) Scaling language models: methods, analysis & insights from training Gopher. arXiv:2112.11446 [cs.CL]

Rafailov R, Sharma A, Mitchell E, Ermon S, Manning CD, Finn C (2023) Direct preference optimization: Your language model is secretly a reward model. In: Proc. of the 37th conference on neural information processing systems (NeurIPS'23)

Raffel C, Shazeer N, Roberts A, Lee K, Narang S, Matena M, Zhou Y, Li W, Liu PJ (2020) Exploring the limits of transfer learning with a unified text-to-text transformer. J Mach Learn Res 21(140):1–67

Rashkin H, Celikyilmaz A, Choi Y, Gao J (2020) PlotMachines: Outline-conditioned generation with dynamic plot state tracking. In: Proc. of the 2020 conference on empirical methods in natural language processing (EMNLP'20)

Ressler M (2023) Automated inauthenticity. AI & Soc. Accepted for publication

Reynolds L, McDonell K (2021) Multiversal views on language models. arXiv:2102.06391 [cs.HC]

Rhodes M (1961) An analysis of creativity. Phi Delta Kappan 42(7):305–310

Riedl MO (2018) Computational creativity as meta search. https://mark-riedl.medium.com/computational-creativity-as-meta-search-6cad95da923b. Accessed 30 Oct 2024

Riedl MO, Young RM (2010) Narrative planning: balancing plot and character. J Artif Intell Res 39(1):217–268

Ritchie G (2007) Some empirical criteria for attributing creativity to a computer program. Mind Mach 17:67–99

Rosset C (2020) Turing-NLG: a 17-billion-parameter language model by Microsoft. https://www.microsoft.com/en-us/research/blog/turing-nlg-a-17-billion-parameter-language-model-by-microsoft/. Accessed 30 Oct 2024

Sawicki P, Grzés M, Jordanous A, Brown D, Peeperkorn M (2022) Training GPT-2 to represent two romantic-era authors: challenges, evaluations and pitfalls. In: Proc. of the 3th international conference on computational creativity (ICCC'22)

Semeniuta S, Severyn A, Barth E (2017) A hybrid convolutional variational autoencoder for text generation. In: Proc. of the 2017 conference on empirical methods in natural language processing (EMNLP'17)

Seth A (2021) Being you: a new science of consciousness. Penguin

Shanahan M (2024a) Simulacra as conscious exotica. arXiv:2402.12422 [cs.AI]

Shanahan M (2024b) Talking about large language models. Commun ACM 67(2):68–79

Shanahan M, McDonell K, Reynolds L (2023) Role play with large language models. Nature 623(7987):493–498

Shin H, Lee JK, Kim J, Kim J (2017) Continual learning with deep generative replay. In: Advances in neural information processing systems (NIPS'17)

Shoeybi M, Patwary, M, Puri R, LeGresley P, Casper J, Catanzaro B (2019) Megatron-LM: training multi-billion parameter language models using model parallelism. arXiv:1909.08053 [cs.CL]

Silva E (2022) How photography pioneered a new understanding of art. https://www.thecollector.com/how-photography-transformed-art/. Accessed 30 Oct 2024

Smith S, Patwary M, Norick B, LeGresley P, Rajbhandari S, Casper J, Liu Z, Prabhumoye S, Zerveas G, Korthikanti V, Zhang E, Child R, Aminabadi RY, Bernauer J, Song X, Shoeybi M, He Y, Houston M, Tiwary S, Catanzaro B (2022) Using DeepSpeed and megatron to train megatron-turing NLG 530b, a large-scale generative language model. arXiv:2201.11990 [cs.CL]

Stein MI (1974) Stimulating creativity, vol 1. Academic Press, New York

Stevenson C, Smal I, Baas M, Grasman R, van der Maas H (2022a) Putting GPT-3's creativity to the (alternative uses) test. In: Proc. of the 13th international conference on computational creativity (ICCC'22)

Stevenson C, Smal I, Baas M, Grasman R, van der Maas H (2022b) Putting GPT-3's creativity to the (alternative uses) test. In: Proc. of the 13th international conference on computational creativity (ICCC'22)

Stiennon N, Ouyang L, Wu J, Ziegler D, Lowe R, Voss C, Radford A, Amodei D, Christiano PF (2020) Learning to summarize with human feedback. In: Advances in neural information processing systems (NIPS'20)

Sun FK, Ho CH, Lee HY (2020) LAMOL: LAnguage MOdeling for Lifelong Language Learning. In: Proc. of the 2020 international conference on learning representations (ICLR'20)

Swanson B, Mathewson K, Pietrzak B, Chen S, Dinalescu M (2021) Story centaur: large language model few shot learning as a creative writing tool. In: Proc. of the 16th conference of the European chapter of the Association for Computational Linguistics: system demonstrations (EACL'21)

Tamkin A, Brundage M, Clark J, Ganguli D (2021) Understanding the capabilities, limitations, and societal impact of large language models. arXiv:2102.02503 [cs.CL]

Terzidis K, Fabrocini F, Lee H (2022) Unintentional intentionality: art and design in the age of artificial intelligence. AI & Soc 38(4):1715–1724

Thoppilan R, De Freitas D, Hall J, Shazeer N, Kulshreshtha A, Cheng HT, Jin A, Bos T, Baker L, Du Y, Li Y, Lee H, Zheng HS, Ghafouri A, Menegali M, Huang Y, Krikun M, Lepikhin D, Qin J et al (2022) LaMDA: language models for dialog applications. arXiv:2201.08239 [cs.CL]

Touvron H, Martin L, Stone KR, Albert P, Almahairi A, Babaei Y, Bashlykov N, Batra S, Bhargava P, Bhosale S, Bikel DM, Blecher L, Ferrer CC, Chen M, Cucurull G, Esiobu D, Fernandes J, Fu J, Fu W et al (2023) Llama 2: open foundation and fine-tuned chat models. arXiv:2307.09288 [cs.CL]

Turing AM (1950) Computing machinery and intelligence. Mind LIX(236):433–460

Turner SR (1994) The creative process: a computer model of creativity and storytelling. Lawrence Erlbaum Associates, Inc., Hillsdale

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I (2017) Attention is all you need. In: Advances in neural information processing systems (NIPS'17)

Waite T (2019) AI-generated artworks are disappointing at auction. https://www.dazeddigital.com/art-photography/article/46839/1/

ai-generated-artworks-disappointing-at-auction-obvious-artificial-intelligence. Accessed 30 Oct 2024

Weidinger L, Uesato J, Rauh M, Griffin C, Huang PS, Mellor J, Glaese A, Cheng M, Balle B, Kasirzadeh A, Biles C, Brown S, Kenton Z, Hawkins W, Stepleton T, Birhane A, Hendricks LA, Rimell L, Isaac W, Haas J, Legassick S, Irving G, Gabriel I (2022) Taxonomy of risks posed by language models. In: Proc. of the 2022 ACM conference on fairness, accountability, and transparency (FAccT'22)

Wertz L, Kuhn J (2022) Adapting transformer language models for application in computational creativity: generating German theater plays with varied topics. In: Proc. of the 13th international conference on computational creativity (ICCC'22)

Wiggins GA (2006) A preliminary framework for description, analysis and comparison of creative systems. Knowl-Based Syst 19(7):449–458

Wu T, Caccia M, Li Z, Li YF, Qi, G, Haffari G (2022) Pretrained language model in continual learning: a comparative study. In: Proc. of the 2022 international conference on learning representations (ICLR'22)

Wu M, Yuan Y, Haffari G, Wang L (2024) (Perhaps) beyond human translation: harnessing multi-agent collaboration for translating ultra-long literary texts. arXiv:2405.11804 [cs.CL]

Yu L, Zhang W, Wang J, Yu Y (2017) SeqGAN: sequence generative adversarial nets with policy gradient. In: Proc. of the 31st AAAI conference on artificial intelligence (AAAI'17)

Zhang Y, Gan Z, Fan K, Chen Z, Henao R, Shen D, Carin L (2017) Adversarial feature matching for text generation. In: Proc. of the 34th international conference on machine learning (ICML'17)

Zhao WX, Zhou K, Li J, Tang T, Wang X, Hou Y, Min Y, Zhang B, Zhang J, Dong Z, Du Y, Yang C, Chen Y, Chen Z, Jiang J, Ren R, Li Y, Tang X, Liu Z, Liu P, Nie JY, Wen JR (2023) A survey of large language models. arXiv:2303.18223 [cs.CL]

Ziegler DM Stiennon N, Wu J, Brown TB, Radford A, Amodei D, Christiano P, Irving G (2019) Fine-tuning language models from human preferences. arXiv:1909.08593 [cs.CL]

Zugarini A, Melacci S, Maggini M (2019) Neural poetry: learning to generate poems using syllables. In: Proc. of the 2019 international conference on artificial neural networks (ICANN'19)