Creativity, Context, and Large Language Models

Max Peeperkorn
School of Computing
University of Kent
United Kingdom
m.peeperkorn@kent.ac.uk

Introduction

This research summary outlines my doctoral project. It covers the work so far, the current project, and what is necessary to complete the dissertation. The theme of this doctoral project is rooted in creativity as a social phenomenon. Creativity does not happen in isolation but requires an environment that informs the research presented here. The big picture of this project is how we make machines more socially or context-aware when performing creative tasks. This is important because it can allow the machines to be more effective in producing what we require and rely less on chance to find something relevant and interesting. Large language models (LLM) play a key role in this project. Primarily because language is a powerful technology that compared to other domains, like images, is more interpretable and easier to instruct. Moreover, instruction-tuned models appear to capture lots of social and cultural information.

In this summary, we will first discuss social creativity and context. Then the creativity of LLMs and the difficulties with evaluation, and specifically investigate if temperature is the creativity parameter. Finally, we will outline the next steps for this thesis project to introduce context beyond in-context learning.

Social Creativity and Context

While this project is heavily focused on the creativity of large language models, ultimately it is rooted in creativity as a social phenomenon. In the psychology literature, the importance of society for creativity is widely recognised (Rhodes 1961; Csikszentmihalyi 1988; Vygotsky 2004). Yet, when we develop computational creative systems, we are mainly concerned with generating objects without incorporating their context. In the past, when evaluating objects, there has been some success in applying measures from computational aesthetics to determine novelty and value (Galanter 2012). However, these measures do not account for the social context (apart from the aesthetic). Fundamentally, machines are not in the world (Dreyfus 1992) and determining value this way is hard, especially, since not all information that is required is present in the object under evaluation.

Social information typically is the information that compliments the main messages exchanged in social groups. Social information tells us something about context and how to interpret the information being conveyed (Morin et al. 2021).

Non-verbal communication and social status are important when humans assess creativity, however, for artificial creativity we need to consider social information in the form of objects that describe or express an opinion about creative work, such as reviews, comments, or paratexts. We have specifically explored *reviewing* and developed a conceptual model of its process (Brown and Peeperkorn 2023). The review itself could be considered a creative object, but its primary purpose is to perform an evaluation and inform other people about what is and isn't appealing about the work and why. Reviews allow, depending on the familiarity of the reader, to reconstruct the original work without observing it. Quantifying the usefulness of a review is useful for exploring emergent behaviours in simulating creative societies.

Before designing and developing strategies and algorithms to introduce context to the LLM we investigate their creative capabilities out-of-the-box.

Creativity of Large Language Models

LLMs are without doubt the most impressive generative models we have seen. They produce high-quality text, excel at stylistic reproduction, and tap into an immense pool of information. They have been characterised as “reasoning”, “sentient”, or “knowing” like humans, yet easily confabulating in their response, and do not seem to have any sense of self. We examine these characterisations and discuss what LLMs can't do and what they are surprisingly good at (Peeperkorn, Brown, and Jordanous 2023). LLMs are still susceptible to traditional issues with AI, probabilities are not knowledge, and they are not in the world. Nonetheless, LLMs, despite not being human, have great potential to perform various creative tasks. We conclude that LLMs are beyond “mere generation” and perceivable as creative, but we may need to reassess some frameworks for creativity evaluation.

We presented a philosophical debate on the creativity of LLMs and briefly examined the creativity of LLMs using well-known Computational Creativity frameworks. In practice, LLMs are freakishly hard to rigorously evaluate and to generalise any findings. Often it is unclear what the exact prompt or parameters are, and it is usually difficult to interpret benchmark results. For example, it seems that LLMs struggle with multiple choice questions, even though this task is part of many benchmarks (Khatun and Brown 2024). In the next section, we perform a creativity evaluation of

LLMs on a narrative generation task to explore the effects of temperature on its creativity presenting a methodology based on prototype theory (Rosch et al. 1976) and exemplar theory (Medin, Altom, and Murphy 1984).

Temperature as the Creativity Parameter

Naturally, LLMs found use for many creative tasks. One claim in particular which is often alluded to primarily in mainstream media, but also in scientific communities, is that temperature can make models more creative (e.g., (Roemmele and Gordon 2018; Chen and Ding 2023)). We investigate this claim using a narrative generation task with a predetermined fixed context, model and prompt (Peeperkorn et al. 2024). Specifically, we present an empirical analysis of the LLM output for different temperature values using four necessary conditions for creativity in narrative generation: novelty, typicality, cohesion, and coherence. We find that temperature has a weak correlation with novelty, and unsurprisingly, moderately correlated with incoherence, but there is no relationship with either cohesion or typicality. However, the influence of temperature on creativity is far more nuanced and weak than suggested by the "creativity parameter" claim; overall results suggest that the LLM generates slightly more novel outputs as temperatures get higher. Finally, we discuss ideas to allow more controlled LLM creativity, rather than relying on chance via changing the temperature parameter.

From this work, it became clear that more is needed for them to behave creatively, and the signs pointed towards the decoding strategy, which is responsible for making the LLM behave naturally in the first place.

Context Decoding

In this chapter of the thesis, we will loop back to social creativity, context, and reviewing as a creative task (Brown and Peeperkorn 2023). It is my current focus and still a work in progress. The default approach of providing contextual information to large language models is to use in-context learning (Brown et al. 2020). While this has been very effective, it's often difficult to control for more complicated tasks, such as narrative generation. It is not clear or easy to determine what the effect is of providing examples. Moreover, if you would want the LLM to write a poem *not* in the style of Walt Whitman, then it has trouble with understanding what *not Walt Whitman* means. The next chapter of the doctoral project will be to develop a strategy where we move the contextual information from the prompt to the decoding step. In theory, by reshaping the sampling probability distribution, we could add or subtract information in very specific ways.

Next Steps

The natural continuation of the work in the previous section is to apply the approach to various creative tasks and explore a combination with locally typical sampling (Meister et al. 2023). Generative AI excels at aggregating vast amounts of data into an efficient model, that can produce high-quality, yet interpolated outputs. How can we make LLMs find interpolations that are novel, yet interesting and of high quality?

References

Brown, D., and Peeperkorn, M. 2023. Reviewing, creativity, and algorithmic information theory. In *14th International Conference on Computational Creativity*, 133–147. Association for Computational Creativity.

Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J. D.; Dhariwal, P.; Neelakantan, A.; . . . ; and Amodei, D. 2020. Language models are few-shot learners. In *Advances in Neural Information Processing Systems*, volume 33, 1877–1901. Curran Associates, Inc.

Chen, H., and Ding, N. 2023. Probing the "creativity" of large language models: Can models produce divergent semantic association? In *Findings of the Association for Computational Linguistics: EMNLP 2023*, 12881–12888. Association for Computational Linguistics.

Csikszentmihalyi, M. 1988. Society, culture, and person: A systems view of creativity. In *The Nature of Creativity*. Cambridge University Press. 243–270.

Dreyfus, H. L. 1992. *What computers still can't do: A critique of artificial reason*. MIT press.

Galanter, P. 2012. Computational aesthetic evaluation: Past and future. In *Computers and Creativity*. 255–293.

Khatun, A., and Brown, D. G. 2024. A study on large language models' limitations in multiple-choice question answering. Preprint. arXiv:2401.07955.

Medin, D. L.; Altom, M. W.; and Murphy, T. D. 1984. Given versus induced category representations: Use of prototype and exemplar information in classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 10(3):333.

Meister, C.; Pimentel, T.; Wiher, G.; and Cotterell, R. 2023. Locally typical sampling. *Transactions of the Association for Computational Linguistics* 11:102–121.

Morin, O.; Jacquet, P. O.; Vaesen, K.; and Acerbi, A. 2021. Social information use and social information waste. *Phil. Trans. R. Soc. B* 376(1828):20200052.

Peeperkorn, M.; Kouwenhoven, T.; Brown, D.; and Jordanous, A. 2024. Is temperature the creativity parameter of large language models? In *15th International Conference on Computational Creativity*. ACC. *Forthcoming*.

Peeperkorn, M.; Brown, D.; and Jordanous, A. 2023. On characterizations of large language models and creativity evaluation. In *14th International Conference on Computational Creativity*, 143–147. ACC.

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

Roemmele, M., and Gordon, A. S. 2018. Automated assistance for creative writing with an rnn language model. In *23rd International Conference on Intelligent User Interfaces Companion*. ACM.

Rosch, E.; Mervis, C. B.; Gray, W. D.; Johnson, D. M.; and Boyes-Braem, P. 1976. Basic objects in natural categories. *Cognitive Psychology* 8(3):382–439.

Vygotsky, L. S. 2004. Imagination and creativity in childhood. *Journal of Russian & East European Psychology* 42(1):7–97.