STATES OF LLM-GENERATED TEXTS AND PHASE TRANSITIONS BETWEEN THEM

Nikolay Mikhaylovskiy
NTR Labs, Moscow, Russia
and Higher IT School
of Tomsk State University,
Tomsk, Russia
nickm@ntr.ai

ABSTRACT

It is known for some time that autocorrelations of words in human-written texts decay according to a power law. Recent works have also shown that the autocorrelations decay in texts generated by LLMs is qualitatively different from that in literary texts. Solid-state physics tie the autocorrelations decay laws to the states of matter. In this work, we empirically demonstrate that, depending on the temperature parameter, LLMs can generate text that can be classified as solid, critical state, or gas.

1 INTRODUCTION

Although not long ago probabilistic autoregressive language models were just models that assign probabilities to sequences of words (Bahl et al., 1983), now they are the cornerstone of any task in computational linguistics by prompting (Sanh et al., 2022) or fine-tuning (Radford et al., 2018). Such models being successfully commercialized, the number of practical applications of these models is rapidly growing, as is the number of papers considering various aspects of the use of probabilistic autoregressive language models. It is all the more surprising that the statistical properties of the output sequences produced by such models have been relatively little studied.

We aim to fill this gap somewhat and empirically demonstrate that, depending on the temperature parameter, LLMs can generate text that can be classified as solid (periodic phase), critical state (that has autocorrelations decay according to the power law), or gas (amorphous phase) from the point of view of autocorrelation analysis.

Our main contributions are the following:

1. We clearly identify three phases of LLM-generated texts - periodic, critical and amorphous
2. We show through computational experiments that for LLM-generated texts, there is a phase transition from ordered to amorphous state at about the same temperatures between 0.7 and 1, for different LLMs
3. We show that for amorphous state, long-range autocorrelations decay follows the exponential law independently from the generation temperature, for different LLMs
4. We show that for temperatures between 0.7 and 1 autocorrelations exhibit power law decay on medium distances of up to 2000 words, implying isles of connectivity of these sizes.

We go on to introduce the key concepts.

1.1 AUTOREGRESSIVE PROBABILISTIC LANGUAGE MODELS

Probabilistic language models consider sequences

$$t_{1:m} = \{t_1, t_2, \ldots, t_m\} \quad (1)$$

of tokens from the lexicon L. An autoregressive language model estimates the probability of such a sequence

$$\begin{aligned} P(t_{1:m}) &= P(t_1)P(t_2|t_1)P(t_3|t_{1:2})\dots P(t_m|t_{1:m-1}) \\ &= \prod_{k=1}^{m} P(t_k|t_{1:k-1}) \end{aligned} \tag{2}$$

using the chain rule. Many models introduce the Markov (1913) assumption that the probability of a token depends on the previous $n-1$ tokens only, thus approximating (3) with a truncated version

$$P(t_{1:m}) \approx \prod_{k=1}^{m} P(t_k|t_{k-n+1:k-1}) \tag{3}$$

1.2 Text Generation with a Language Model

Given an input text as a context, the goal of open-ended generation is to produce a coherent continuation of the text (Holtzman et al., 2020). More formally, given a sequence of m tokens $t_1 \dots t_m$ as context, the objective is to generate the next n continuation tokens, resulting in the completed sequence $t_1 \dots t_{m+n}$. This is achieved through the use of the left-to-right text probability decomposition (2), which is used to generate the sequence one token at a time, using a particular decoding strategy.

A common approach to text generation is to shape a probability distribution through temperature (Ackley et al., 1985). Given the logits $u_{1:|V|}$ and temperature T, the softmax is re-estimated as

$$p(t = V_l|t_{1:i-1}) = \frac{\exp(u_l/T)}{\sum_{l'} \exp(u_{l'}/T)} \tag{4}$$

Setting $T \in [0, 1)$ skews the distribution towards high-probability events, and, similarly, $T \in (1, \infty)$ skews the distribution towards low-probability events.

1.3 Computing Autocorrelations Using Distributional Semantics

Suppose that we have a sequence of N vectors $V_i \in R^d, i \in [1, N]$. The autocorrelation function $C(\tau)$ is the average similarity between the vectors as a function of the lag $\tau = i - j$ between them. The simplest metric of vector similarity is the cosine similarity

$$d(V_i, V_j) = \cos(V_i, V_j) = \frac{(V_i \cdot V_j)}{||V_i||||V_j||}, \tag{5}$$

where $\cdot$ is a dot product between two vectors and $||x||$ is an Euclidean norm of a vector. Thus,

$$C(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} \frac{V_i \cdot V_{i+\tau}}{||V_i||||V_{i+\tau}||} \tag{6}$$

A distributional semantic (Harris, 1954) such as GloVe (Pennington et al., 2014) assigns a vector to each word or context in a text. Thus, a text is transformed into a sequence of vectors, and we can calculate an autocorrelation function for the text.

1.4 Phase Transitions

A physical phase of a system refers to a (typically equilibrium) state with unique macroscopic properties. These phases possess certain stability regions within the parameter space. The properties of the state change at the boundaries of these regions, where phase transition occurs.

Ehrenfest (1933) defined a phase transition as a discontinuity in the n-th order derivative of the free energy with respect to any argument of the free energy. Modern physics extends the notion of phases and applies it to various situations and beyond the notion of free energy. In particular, a first-order phase transition exhibits a discontinuity in the first-order derivative, whereas a second-order phase transition is continuous in its first derivative but shows a discontinuous or divergent behavior in its second derivatives (Papon et al., 2007).

and in the act of devouring a man,
and in the act of devouring a whale,
and in the act of devouring a ship,
and in the act of devouring a man,
and in the act of devouring a whale,
and in the act of devouring a ship,
and in the act of devouring a man,
and in the act of devouring a whale,
and in the act of devouring a ship,

Figure 1: Degenerative Text Generated by Qwen at t=0.1, shift 11904, seed 1

My companjour's father held in trust as old, well wizdom; being like ye 'ear say my first—(thro' not only bubbynee I tell) old Cronicle for it were not. No man can give 'fitt or show o more clear-aifv intelligence—that will prove well at last I sase now believe it all me blawms—be it or'tt how.

Figure 2: Nonsense Text Generated by Phi at t=2.8, shift 539, seed 1

2 PHASES IN LLM-GENERATED TEXTS

2.1 PRIOR RESEARCH

Power-law autocorrelations decay in human-written texts was studied in a number of works (Li, 1989; Schenkel et al., 1993; Ebeling & Pöschel, 1994; Ebeling & Neiman, 1995; Kokol et al., 1999; Pavlov et al., 2001; Montemurro & Pury, 2002; Alvarez-Lacalle et al., 2006; Manin, 2008; Gillet & Ausloos, 2008; Corral et al., 2009; Altmann et al., 2012; Amit et al., 1994; Tanaka-Ishii & Bunde, 2016; Lin & Tegmark, 2017; Takahashi & Tanaka-Ishii, 2017; Shen, 2019; Takahashi & Tanaka-Ishii, 2019; Sainburg et al., 2019; Mikhaylovskiy & Churilov, 2023; Nakaishi et al., 2024).

Significantly less works are devoted to the analysis of statistical quantities in texts generated by language models. Generated texts have been studied by (Takahashi & Tanaka-Ishii, 2017; Shen, 2019; Takahashi & Tanaka-Ishii, 2019; Lippi et al., 2019; Mikhaylovskiy & Churilov, 2023), who conjectured that power-law decays in autocorrelations are model-dependent.

Nakaishi et al. (2024) and Bahamondes (2023) independently pioneered the application of the phase transition apparatus to LLM-generated texts. They both used only GPT-2, which has no practical interest by now. Bahamondes (2023) studied a setup similar to ours but came up with a significantly different phase transition temperature (0.1 instead of 0.8). We strongly believe that our results are more accurate. Nakaishi et al. (2024) used part-of-speech correlations that are hardly applicable in the high-temperature area (cf. Figure 3).

2.2 TEXT GENERATION SETUP

We use two compact LLMs: Qwen2.5-1.5B (Qwen et al., 2025) and Phi-3-Mini-128K-Instruct (Abdin et al., 2024) to generate texts using the HuggingFace transformer library (Wolf et al., 2020) with temperature sampling (Ackley et al., 1985), similarly to Nakaishi et al. (2024). We further call these models Qwen and Phi for brevity. It is worth noting that from the practical viewpoint these models belong to different classes, as Phi is instruction-tuned, and Qwen is just pretrained.

We start all sequences with the starting passage of "MOBY-DICK; or, THE WHALE." By Herman Melville and iterate the random seed for reproducibility. We force and repeat generation until the text length of 10000 is achieved. We do not use top-k (Fan et al., 2018) or top-p (Holtzman et al., 2020) sampling strategies. Texts are tokenized by a default tokenizer for each model. We sample 10 sequences at each temperature from 0.1 to 2.5 with a step of 0.3. Sampling with each model took less than a day using a single NVIDIA A100.

I don not what and what is he tal(ed in). The "Bill (s)" said he knew more , the ('he knew . Here .....there is so not ... 's, for example... A lot ... In The News-'Vacatio'(as 'He Did S.Y.(se) in Ital:) I(he) sited it on a '(l'he ') page in U S. I know.. , so in 97 :, ,: and , the u..? A l i N o s ,

Figure 3: Gibberish Text Generated by Phi at t=2.8, shift 9794, seed 1

But I have a little more to say. What I want you to understand is that every one of us has a right to decide when we're old.
When I started out, a man might think that was a silly idea. I didn't think so.

Figure 4: Text Generated by Phi at t=1.0, shift 9816, seed 1

2.3 Computing Autocorrelations

We use pretrained multilingual GloVe vector embeddings by Ferreira et al. (2016) similarly to Mikhaylovskiy & Churilov (2023). Unlike them, we do not filter out any words. We center the vector system by subtracting the average of vectors over the whole text. If a word produced by a tokenizer does not have any pretrained vector corresponding, we assign an all-zeroes vector to the token after centering and assume that in this case all correlations are equal to zero. After that we can compute the autocorrelation function following Section 1.3.

2.4 Empirical Phase Observations

It is known for some time (Kulikov et al., 2019; Holtzman et al., 2018; Fan et al., 2018) that at low temperatures LLMs tend to generate degenerate, repetitive text (Figure 1). At high temperatures, the text progresses from nonsense to gibberish in the course of generation (Figure 2, Figure 3). At moderate temperatures, the text is often locally consistent, although it lacks the global consistency and richness characteristic of human-written texts (Figure 4). At such temperatures during long generation we can sometimes witness transitions from random or gibberish text back to human-like one and then decay back to random (See Appendix 1 [1]). From these empirical observations, we can conjecture that the LLM-generated texts can have phase transitions during the generation process (exhibit domains of different phases in space) and as the temperature varies.

2.5 Periodic Phase and Its Phase Transition

At low temperatures, the generated text degenerates and becomes periodic (see Figure 1). The periods vary with the seed, temperature, and model. To quantify this, we compute autocorrelations for the distances from 1 to 100 words and plot the autocorrelation function. To further analyze the periodic nature of the autocorrelations, we perform discrete FFT of the autocorrelation function (see Appendix 2 [2] for the extended image set). A typical example of a periodic autocorrelation function is presented in Figure 5, and a typical example of aperiodic autocorrelation function is presented on Figure 6

The difference is obvious on both autocorrelation and FFT graphs. To further quantify the transition between periodic and non-periodic structures, we plot the maximum absolute value of the Fourier transform of the normalized autocorrelation function beyond the first coefficient against the temperature (Figures 7 and 8). It is clear from the figures that the phase transition for both models happens at t around 0.8. At $t = 1$ the text is non-periodic. The abrupt change is characteristic of a phase transition, as opposed to a gradual change (cf. Section 1.4).

2.6 Amorphous Phase

At high temperatures the generated texts are random (see Figure 3 for example). This entails an exponential decay of autocorrelations. At medium temperatures the generated texts are legible and

[1] Appendices are available as supplementary material by this link
[2] Appendices are available as supplementary material by this link

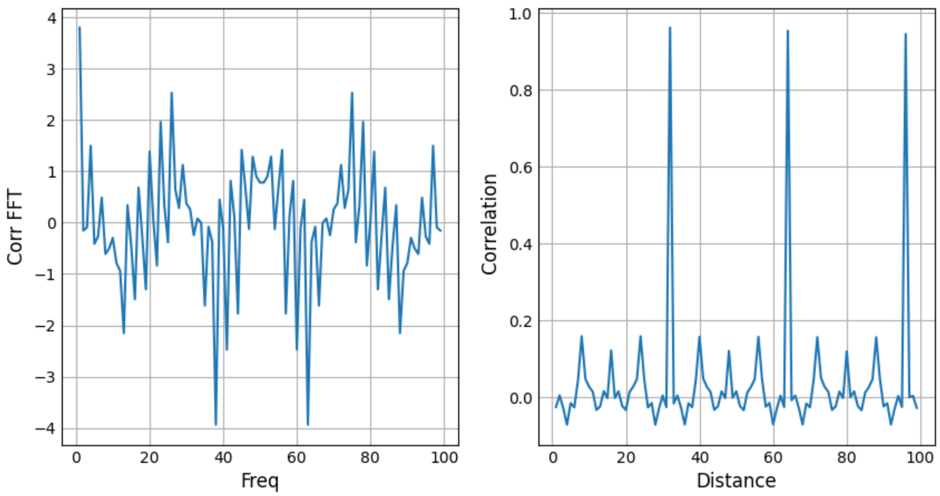

Figure 5: Autocorrelation Function of the Text Generated by Phi at $t = 0.4$ and $seed = 2$ and Its FFT

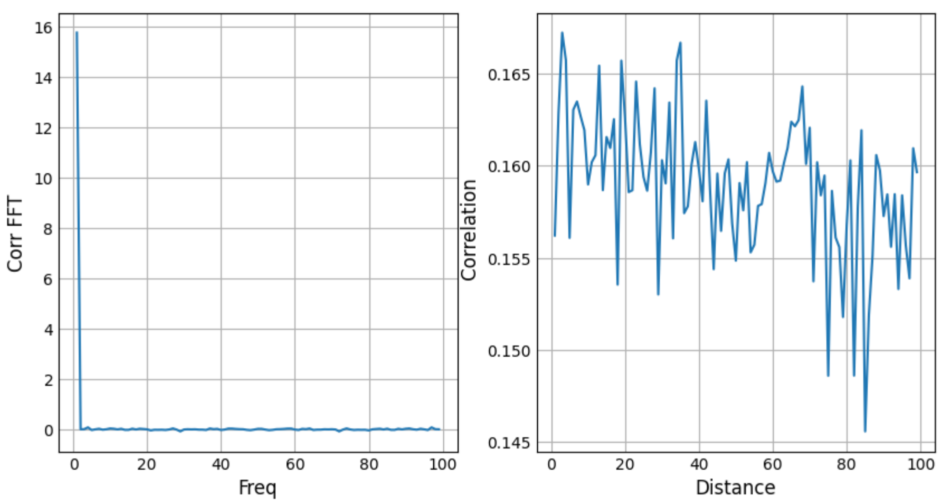

Figure 6: Autocorrelation Function of the Text Generated by Phi at $t = 2.8$ and $seed = 5$. and Its FFT

often exhibit power law autocorrelations decay. We want to study the transition between these two phases.

To quantify this, we compute autocorrelations for selected distances from 1 to thousands of words (see though the further discussion), and plot the autocorrelation function in log and linear coordinates. Periodic texts do not make much sense in log coordinates because there are typically a lot of negative correlations, so we only consider temperatures that are greater than 0.7. The autocorrelation functions of many texts appear messy (Figure 9), but for other texts one can definitely spot either power law (Figure 10) or exponential autocorrelations decay (Figure 11). See Appendix 3 [3] for the extended image set.

Mikhaylovskiy (2023) suggested GAPELMAPER (GloVe Autocorrelations Power/ Exponential Law Mean Absolute Percentage Error Ratio) metric to distinguish texts with exponential and power law autocorrelations decay and determine whether the text has a hierarchical structure. Application of this metric to the texts up to long-range correlations shows that most LLM-generated texts have

[3] Appendices are available as supplementary material by this link

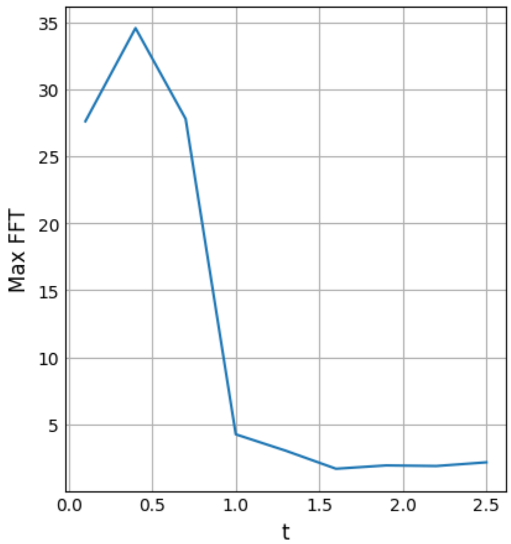

Figure 7: Transition from Periodic to Amorphous Phase in Phi-generated Texts

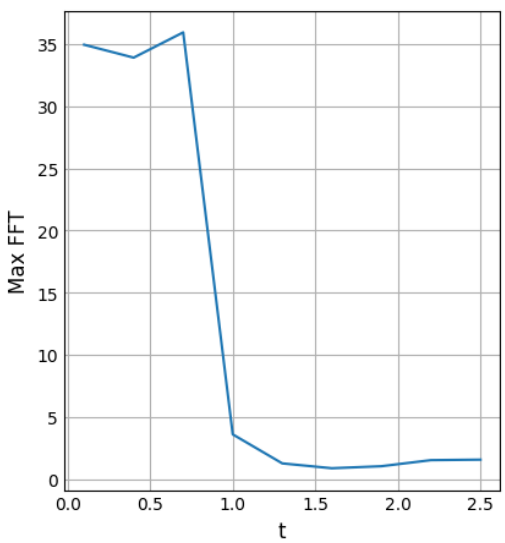

Figure 8: Transition from Periodic to Amorphous Phase in Qwen-generated Texts

exponential autocorrelations decay and thus no inherent hierarchical structure. For example, if we compute autocorrelations up to a distance of 6000 words in Qwen-generated texts we will observe that for no temperature there is a certain power law decay (see Figure 12), that is, GAPELMAPER is rarely if ever less than 1 (GAPELMAPER cannot be computed for most texts generated with $t \in \{0.1, 0.4, 2.8\}$ because of negative autocorrelations).

On the other hand, if we limit the autocorrelations distance to, say, 600 words, we will observe that at $t = 0.7$ GAPELMAPER is reliably less than 1, and wiggles around 1 at higher temperatures (see Figure 13). Only at autocorrelation lengths around 3000 the situation changes qualitatively. This means that the power correlations exist in shorter (under 2000 words) chunks of the generated texts at temperatures of 0.7 to 1.0, but the correlations are lost at longer distances. See Appendix 4 [4] for more data on this.

[4]Appendices are available as supplementary material by this link

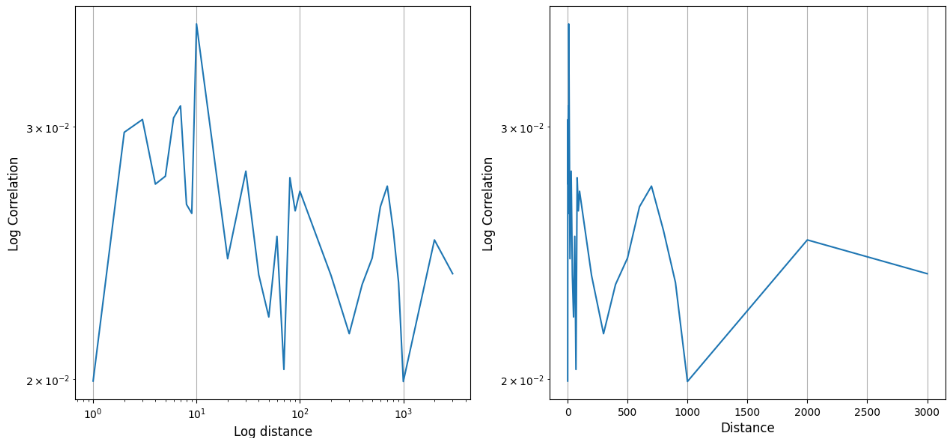

Figure 9: Autocorrelation Function for the Text Generated by Phi at $t = 1.0$ and $seed = 7$

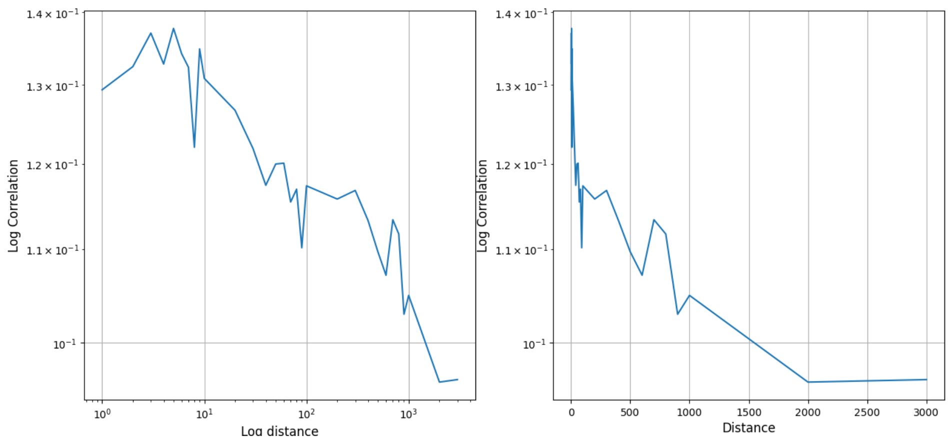

Figure 10: Autocorrelation Function for the Text Generated by Phi at $t = 1.9$ and $seed = 1$

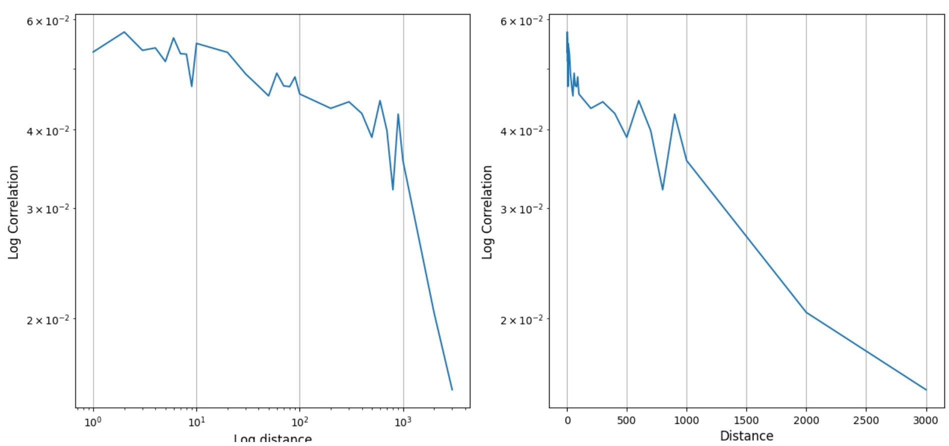

Figure 11: Autocorrelation Function for the Text Generated by Phi at $t = 2.8$ and $seed = 8$

3 Discussion and Future Work

The use of the phase transition apparatus to study LLM-generated texts promises a deeper understanding of both human- and LLM-generated texts, as well as the mechanisms of inner workings

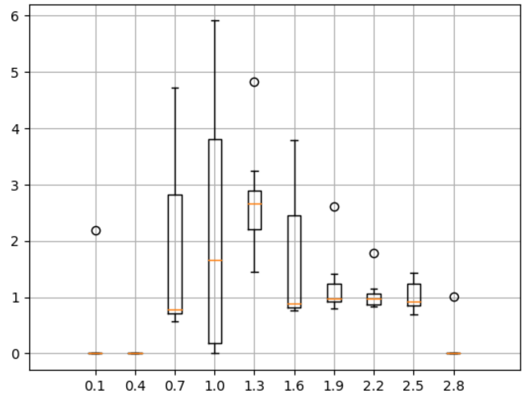

Figure 12: GAPELMAPER Metric for Qwen-Generated Texts, Autocorrelation Distances 1 to 6000 Words

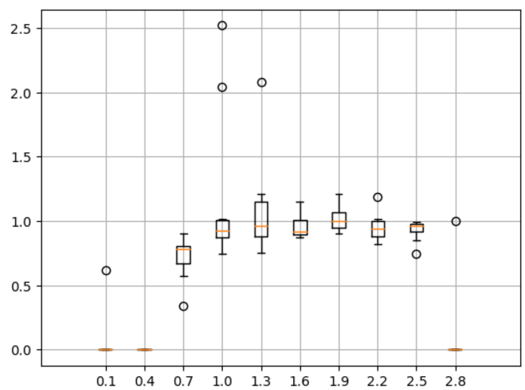

Figure 13: GAPELMAPER Metric for Qwen-Generated Texts, Autocorrelation Distances 1 to 600 Words

of LLMs. We have clearly identified three phases of LLM-generated texts - periodic, critical, and amorphous. Our research confirms the existence of a phase transition in LLM-generated texts at a temperature of about 0.8, which largely confirms the results of Nakaishi et al. (2024). We have also shown for the first time that for the amorphous state, the long-range autocorrelations decay follows the exponential law independently of the generation temperature, for different LLMs, which was previously conjectured by Mikhaylovskiy & Churilov (2023).

We have shown that for temperatures between 0.7 and 1 autocorrelations exhibit power law decay on medium distances of up to 2000 words. This implies islets of connectivity of these sizes, but the fine-grained structure of this sort remains unexplored. This is a topic for the future research.

Our results with different LLMs allow us to conjecture that transformer-based LLMs belong to the same universality class, as described in statistical physics. Do other models of different architecture belong to this class, say, state-space-based or diffusion-based, remains unexplored and is a topic of future research. Additional research is also needed to study the influence of LLM size on the universality class.

ACKNOWLEDGEMENTS

The author thanks Dmitry Manin and Kai Nakaishi for discussing earlier versions of this work.

REFERENCES

Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, Alon Benhaim, Misha Bilenko, Johan Bjorck, Sébastien Bubeck, Martin Cai, Qin Cai, Vishrav Chaudhary, Dong Chen, Dongdong Chen, Weizhu Chen, Yen-Chun Chen, Yi-Ling Chen, Hao Cheng, Parul Chopra, Xiyang Dai, Matthew Dixon, Ronen Eldan, Victor Fragoso, Jianfeng Gao, Mei Gao, Min Gao, Amit Garg, Allie Del Giorno, Abhishek Goswami, Suriya Gunasekar, Emman Haider, Junheng Hao, Russell J. Hewett, Wenxiang Hu, Jamie Huynh, Dan Iter, Sam Ade Jacobs, Mojan Javaheripi, Xin Jin, Nikos Karampatziakis, Piero Kauffmann, Mahoud Khademi, Dongwoo Kim, Young Jin Kim, Lev Kurilenko, James R. Lee, Yin Tat Lee, Yuanzhi Li, Yunsheng Li, Chen Liang, Lars Liden, Xihui Lin, Zeqi Lin, Ce Liu, Liyuan Liu, Mengchen Liu, Weishung Liu, Xiaodong Liu, Chong Luo, Piyush Madan, Ali Mahmoudzadeh, David Majercak, Matt Mazzola, Caio César Teodoro Mendes, Arindam Mitra, Hardik Modi, Anh Nguyen, Brandon Norick, Barun Patra, Daniel Perez-Becker, Thomas Portet, Reid Pryzant, Heyang Qin, Marko Radmilac, Liliang Ren, Gustavo de Rosa, Corby Rosset, Sambudha Roy, Olatunji Ruwase, Olli Saarikivi, Amin Saied, Adil Salim, Michael Santacroce, Shital Shah, Ning Shang, Hiteshi Sharma, Yelong Shen, Swadheen Shukla, Xia Song, Masahiro Tanaka, Andrea Tupini, Praneetha Vaddamanu, Chunyu Wang, Guanhua Wang, Lijuan Wang, Shuohang Wang, Xin Wang, Yu Wang, Rachel Ward, Wen Wen, Philipp Witte, Haiping Wu, Xiaoxia Wu, Michael Wyatt, Bin Xiao, Can Xu, Jiahang Xu, Weijian Xu, Jilong Xue, Sonali Yadav, Fan Yang, Jianwei Yang, Yifan Yang, Ziyi Yang, Donghan Yu, Lu Yuan, Chenruidong Zhang, Cyril Zhang, Jianwen Zhang, Li Lyna Zhang, Yi Zhang, Yue Zhang, Yunan Zhang, and Xiren Zhou. Phi-3 technical report: A highly capable language model locally on your phone, 2024. URL https://arxiv.org/abs/2404.14219.

David H. Ackley, Geoffrey E. Hinton, and Terrence J. Sejnowski. A learning algorithm for boltzmann machines. *Cognitive Science*, 9(1):147–169, 1985. ISSN 0364-0213. doi: https://doi.org/10.1016/S0364-0213(85)80012-4. URL https://www.sciencedirect.com/science/article/pii/S0364021385800124.

Eduardo G. Altmann, Giampaolo Cristadoro, and Mirko Degli Esposti. On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences*, 109(29):11582–11587, 2012. doi: 10.1073/pnas.1117723109. URL https://www.pnas.org/doi/abs/10.1073/pnas.1117723109.

E. Alvarez-Lacalle, B. Dorow, J.-P. Eckmann, and E. Moses. Hierarchical structures induce long-range dynamical correlations in written texts. *Proceedings of the National Academy of Sciences*, 103(21):7956–7961, 2006. doi: 10.1073/pnas.0510673103. URL https://www.pnas.org/doi/abs/10.1073/pnas.0510673103.

M. Amit, Y. Shmerler, E. Eisenberg, M. Abraham, and N. Shnerb. Language and codification dependence of long-range correlations in texts. *Fractals*, 02(01):7–13, 1994. doi: 10.1142/S0218348X94000028. URL https://doi.org/10.1142/S0218348X94000028.

Sebastian Bahamondes. Study of the possibility of phase transitions in LLMs, 2023. URL https://community.wolfram.com/groups/-/m/t/2958851.

Lalit R. Bahl, Frederick Jelinek, and Robert L. Mercer. A maximum likelihood approach to continuous speech recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-5(2):179–190, 1983. doi: 10.1109/TPAMI.1983.4767370.

Alvaro Corral, Ramon Ferrer i Cancho, Gemma Boleda, Albert Diaz-Guilera, and . Universal complex structures in written language, 2009. URL https://arxiv.org/abs/0901.2924.

W. Ebeling and T. Pöschel. Entropy and long-range correlations in literary english. *Europhysics Letters*, 26(4):241, may 1994. doi: 10.1209/0295-5075/26/4/001. URL https://dx.doi.org/10.1209/0295-5075/26/4/001.

Werner Ebeling and Alexander Neiman. Long-range correlations between letters and sentences in texts. *Physica A: Statistical Mechanics and its Applications*, 215(3):233–241, 1995. ISSN 0378-4371. doi: https://doi.org/10.1016/0378-4371(95)00025-3. URL https://www.sciencedirect.com/science/article/pii/0378437195000253.

Paul Ehrenfest. Phasenumwandlungen im ueblichen und erweiterten sinn, classifiziert nach den entsprechenden singularitaeten des thermodynamischen potentiales. *Bd. 36. Proc. Royal Acad. Amsterdam, 153–157*, 1933.

Angela Fan, Mike Lewis, and Yann Dauphin. Hierarchical neural story generation. In Iryna Gurevych and Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 889–898, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1082. URL https://aclanthology.org/P18-1082/.

Daniel C. Ferreira, André F. T. Martins, and Mariana S. C. Almeida. Jointly learning to embed and predict with multiple languages. In Katrin Erk and Noah A. Smith (eds.), *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 2019–2028, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1190. URL https://aclanthology.org/P16-1190/.

J. Gillet and M. Ausloos. A comparison of natural (english) and artificial (esperanto) languages. a multifractal method based analysis, 2008. URL https://arxiv.org/abs/0801.2510.

Zellig S. Harris. Distributional structure. *WORD*, 10(2-3):146–162, 1954. doi: 10.1080/00437956.1954.11659520. URL https://doi.org/10.1080/00437956.1954.11659520.

Ari Holtzman, Jan Buys, Maxwell Forbes, Antoine Bosselut, David Golub, and Yejin Choi. Learning to write with cooperative discriminators. In Iryna Gurevych and Yusuke Miyao (eds.), *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 1638–1649, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-1152. URL https://aclanthology.org/P18-1152/.

Ari Holtzman, Jan Buys, Li Du, Maxwell Forbes, and Yejin Choi. The curious case of neural text degeneration. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=rygGQyrFvH.

Peter Kokol, Vili Podgorelec, Milan Zorman, Tatjana Kokol, and Tatjana Njivar. Computer and natural language texts—a comparison based on long-range correlations. *J. Am. Soc. Inf. Sci.*, 50 (14):1295–1301, December 1999. ISSN 0002-8231.

Ilia Kulikov, Alexander Miller, Kyunghyun Cho, and Jason Weston. Importance of search and evaluation strategies in neural dialogue modeling. In Kees van Deemter, Chenghua Lin, and Hiroya Takamura (eds.), *Proceedings of the 12th International Conference on Natural Language Generation*, pp. 76–87, Tokyo, Japan, October–November 2019. Association for Computational Linguistics. doi: 10.18653/v1/W19-8609. URL https://aclanthology.org/W19-8609/.

Wentian Li. Mutual information functions of natural language texts, 1989. URL https://www.santafe.edu/research/results/working-papers/mutual-information-functions-of-natural-language-t.

Henry W. Lin and Max Tegmark. Critical behavior in physics and probabilistic formal languages. *Entropy*, 19(7), 2017. ISSN 1099-4300. doi: 10.3390/e19070299. URL https://www.mdpi.com/1099-4300/19/7/299.

Marco Lippi, Marcelo A. Montemurro, Mirko Degli Esposti, and Giampaolo Cristadoro. Natural language statistical features of lstm-generated texts. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3326–3337, 2019. doi: 10.1109/TNNLS.2019.2890970.

Dmitrii Y. Manin. On the nature of long-range letter correlations in texts, 2008. URL https://arxiv.org/abs/0809.0103.

A. A. Markov. An example of statistical investigation of the text eugene onegin concerning the connection of samples in chains. *Science in Context*, 19(4):591–600, 1913. doi: 10.1017/S0269889706001074.

Nikolay Mikhaylovskiy. Long story generation challenge. In Simon Mille (ed.), *Proceedings of the 16th International Natural Language Generation Conference: Generation Challenges*, pp. 10–16, Prague, Czechia, September 2023. Association for Computational Linguistics. URL https://aclanthology.org/2023.inlg-genchal.2/.

Nikolay Mikhaylovskiy and Ilya Churilov. Autocorrelations decay in texts and applicability limits of language models. In *Computational Linguistics and Intellectual Technologies: Proceedings of the International Conference "Dialogue 2023"*, 2023. URL https://dialogue-conf.org/media/5915/mikhaylovskiynpluschurilovi031.pdf.

Marcelo A. Montemurro and Pedro A. Pury. Long-range fractal correlations in literary corpora. *Fractals*, 10(04):451–461, 2002. doi: 10.1142/S0218348X02001257. URL https://doi.org/10.1142/S0218348X02001257.

Kai Nakaishi, Yoshihiko Nishikawa, and Koji Hukushima. Critical phase transition in large language models, 2024. URL https://arxiv.org/abs/2406.05335.

P. Papon, S.L. Schnur, J. Leblond, and P.H.E. Meijer. *The Physics of Phase Transitions: Concepts and Applications*. Advanced Texts in Physics. Springer Berlin Heidelberg, 2007. ISBN 9783540333906. URL https://books.google.ru/books?id=n-fiyYg3iSIC.

Alexey N. Pavlov, Werner Ebeling, Lutz Molgedey, Amir R. Ziganshin, and Vadim S. Anishchenko. Scaling features of texts, images and time series. *Physica A: Statistical Mechanics and its Applications*, 300(1):310–324, 2001. ISSN 0378-4371. doi: https://doi.org/10.1016/S0378-4371(01)00341-7. URL https://www.sciencedirect.com/science/article/pii/S0378437101003417.

Jeffrey Pennington, Richard Socher, and Christopher Manning. GloVe: Global vectors for word representation. In Alessandro Moschitti, Bo Pang, and Walter Daelemans (eds.), *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1532–1543, Doha, Qatar, October 2014. Association for Computational Linguistics. doi: 10.3115/v1/D14-1162. URL https://aclanthology.org/D14-1162/.

Qwen, :, An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, Huan Lin, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Yang, Jiaxi Yang, Jingren Zhou, Junyang Lin, Kai Dang, Keming Lu, Keqin Bao, Kexin Yang, Le Yu, Mei Li, Mingfeng Xue, Pei Zhang, Qin Zhu, Rui Men, Runji Lin, Tianhao Li, Tianyi Tang, Tingyu Xia, Xingzhang Ren, Xuancheng Ren, Yang Fan, Yang Su, Yichang Zhang, Yu Wan, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, and Zihan Qiu. Qwen2.5 technical report, 2025. URL https://arxiv.org/abs/2412.15115.

Alec Radford, Karthik Narasimhan, Tim Salimans, and Ilya Sutskever. Improving language understanding by generative pre-training, 2018. URL https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf.

Tim Sainburg, Brad Theilman, Marvin Thielk, and Timothy Q. Gentner. Parallels in the sequential organization of birdsong and human speech. *Nature Communications*, 10, 2019. URL https://api.semanticscholar.org/CorpusID:199527929.

Victor Sanh, Albert Webson, Colin Raffel, Stephen Bach, Lintang Sutawika, Zaid Alyafeai, Antoine Chaffin, Arnaud Stiegler, Arun Raja, Manan Dey, M Saiful Bari, Canwen Xu, Urmish Thakker, Shanya Sharma Sharma, Eliza Szczechla, Taewoon Kim, Gunjan Chhablani, Nihal Nayak, Debajyoti Datta, Jonathan Chang, Mike Tian-Jian Jiang, Han Wang, Matteo Manica, Sheng Shen, Zheng Xin Yong, Harshit Pandey, Rachel Bawden, Thomas Wang, Trishala Neeraj, Jos Rozen, Abheesht Sharma, Andrea Santilli, Thibault Fevry, Jason Alan Fries, Ryan Teehan, Teven Le Scao, Stella Biderman, Leo Gao, Thomas Wolf, and Alexander M Rush. Multitask prompted training enables zero-shot task generalization. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=9Vrb9D0WI4.

Alain Schenkel, Jun Zhang, and Yi-Cheng Zhang. Long range correlation in human writings. *Fractals*, 01(01):47–57, 1993. doi: 10.1142/S0218348X93000083. URL https://doi.org/10.1142/S0218348X93000083.

Huitao Shen. Mutual information scaling and expressive power of sequence models, 2019. URL https://arxiv.org/abs/1905.04271.

Shuntaro Takahashi and Kumiko Tanaka-Ishii. Do neural nets learn statistical laws behind natural language? *PLOS ONE*, 12(12):1–17, 12 2017. doi: 10.1371/journal.pone.0189326. URL https://doi.org/10.1371/journal.pone.0189326.

Shuntaro Takahashi and Kumiko Tanaka-Ishii. Evaluating computational language models with scaling properties of natural language. *Computational Linguistics*, 45(3):481–513, September 2019. doi: 10.1162/coli_a_00355. URL https://aclanthology.org/J19-3003/.

Kumiko Tanaka-Ishii and Armin Bunde. Long-range memory in literary texts: On the universal clustering of the rare words. *PLOS ONE*, 11(11):1–14, 11 2016. doi: 10.1371/journal.pone.0164658. URL https://doi.org/10.1371/journal.pone.0164658.

Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Remi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander Rush. Transformers: State-of-the-art natural language processing. In Qun Liu and David Schlangen (eds.), *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pp. 38–45, Online, October 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-demos.6. URL https://aclanthology.org/2020.emnlp-demos.6/.