

Detailed Balance in Large-Language-Model-Driven Agents

Zhuo-Yang Song^{1,*}, Qing-Hong Cao^{1,2,3,†}, Ming-xing Luo^{4,‡} and Hua Xing Zhu^{1,3,§}

¹*School of Physics, Peking University, Beijing 100871, China*

²*School of Physics, Zhengzhou University, Zhengzhou 450001, China*

³*Center for High Energy Physics, Peking University, Beijing 100871, China*

⁴*Beijing Computational Science Research Center, Beijing 100193, China*



(Received 3 January 2026; accepted 15 September 2026; published 27 October 2026)

Large language model (LLM)-driven agents are emerging as a powerful new paradigm for solving complex problems. Despite the empirical success of these practices, a theoretical framework to understand and unify their macroscopic dynamics remains lacking. This paper proposes a method based on the least action principle to estimate the underlying generative directionality of LLMs embedded within agents. By experimentally measuring the transition probabilities between LLM-generated states, we report evidence consistent with a tendency toward detailed balance in LLM-generated transitions, indicating that LLM generation may not be achieved by generally learning rule sets and strategies, but rather by implicitly learning a class of underlying potential functions that may transcend different LLM architectures and prompt templates. To our knowledge, this provides an effective macroscopic description of LLM generative dynamics that does not depend on specific model details. This work is an attempt to establish a macroscopic dynamics theory of complex AI systems, aiming to elevate the study of AI agents from a collection of engineering practices to a science built on effective measurements that are predictable and quantifiable.

DOI: [10.1103/kwx6-clqd](https://doi.org/10.1103/kwx6-clqd)

I. INTRODUCTION

Large language model (LLM)-driven agents are emerging as a powerful new paradigm for solving complex problems [1–15], demonstrating potential in frontier areas such as scientific discovery by combining the generative capabilities of LLMs with external tools and memory systems [16–20]. For instance, FunSearch and AlphaEvolve achieve iterative optimization of solutions by integrating LLMs into evolutionary algorithm frameworks [18,19]. However, the theoretical understanding and explanation of LLMs often remain at the level of token statistical properties and microscopic generative mechanisms [21–23], making it difficult to explain the macroscopic dynamics of LLMs as complex systems [1,24]. The behavior of these LLM-driven agents is often viewed as a direct product of their complex internal engineering (such as prompt templates, memory modules, tool calls), and their dynamic characteristics remain a black box [12,25–28].

The dynamics of LLM generation are quite unique. Compared to traditional rule-based programs, LLM-based

generation exhibits diverse and adaptive outputs [16,17,25]. At the same time, compared to naive random search, LLM generation shows stronger structure and goal-orientedness [18–20,29]. Despite the complexity of this hybrid dynamics between random search and deterministic planning, we show that at the agent level (i.e., a coarse-grained description of LLM generative dynamics with standardized agent states as units), LLM generative dynamics exhibit detailed balance similar to equilibrium systems, thereby greatly simplifying the analysis and understanding of LLM generative dynamics [30–33].

To model the dynamic behavior of LLMs, we embed the generative process of LLM within a given agent framework, viewing it as a Markov transition process in its state space [7,12,18,30,31]. The states are defined by the complete information retained by the agent at each time step, which may include task objectives, historical summaries, code, file systems, application programming interface return values, etc., where the LLM-based generative process is treated as the transition kernel from the current state to a new state. We show that even if the states do not contain complete historical records (as in, e.g., FunSearch [18]), the LLM-based state transitions can still exhibit a directionality toward specific states. We attribute this characteristic to the state transitions approximately satisfying the detailed balance condition, a nontrivial constraint that rules out probability-flux cycles; from it a task-specific potential function V emerges that evaluates the intrinsic properties of any given state, such as “how far the LLM perceives it to be from the goal.” This global awareness enables LLMs to quickly converge to those optimal states, effectively avoiding repetitive cycles in the state space

*Contact author: zhuoyangsong@stu.pku.edu.cn

†Contact author: qinghongcao@pku.edu.cn

‡Contact author: mingxingluo@csrc.ac.cn

§Contact author: zhuhx@pku.edu.cn

Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

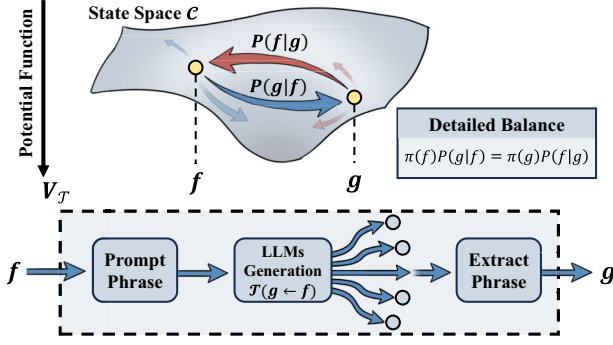


FIG. 1. A schematic of a formalization framework for studying the directionality of LLM generation, illustrating the state space and possible transitions. LLMs are embedded within an agent, which transitions from state f to state g with probability $\mathcal{T}(g \leftarrow f) = P(g|f)$. The underlying potential function $V_{\mathcal{T}}$ quantifies the agent's global ordering of each state, satisfying the detailed balance condition at equilibrium.

and achieving stronger generalization capabilities than merely learning strategy sets [34–36].

Based on this model, we propose a method to measure this underlying potential function based on a least action principle [37–39]. By experimentally measuring the transition probabilities between states, we report evidence consistent with a tendency toward detailed balance in LLM-generated transitions, indicating that LLM generation may not be achieved by generally learning rule sets and strategies, but rather by implicitly encoding an underlying potential function that transcends different LLM architectures and prompt templates. We propose detailed balance as a universal conjecture about LLM-driven generation; the potential V is a task-specific effective description. This work is an attempt to establish a macroscopic dynamics theory of complex AI systems, aiming to elevate the study of AI agents from a collection of engineering practices to a predictable and quantifiable science built on effective measurements.

II. THEORY

To formulate the problem rigorously, we consider an agent whose core consists of one or more LLMs. The agent takes its current state f as input. Through a series of deterministic steps, it organizes and evaluates this state to generate a relevant prompt. This prompt is then fed into one or more LLMs, whose structured output is parsed to get a new state g . This state is the minimal unit for studying LLM dynamics. This generative process can be viewed as a Markov transition process in the state space $\mathcal{C}$ with a transition kernel $P(g|f)$, retaining the diversity and adaptability of LLM generation. The states are defined by the complete information retained by the agent at each time step, which should include all the information required for the agent to carry out a continuous reasoning or analogical process [18,20,40,41]. In this paper, the agent contains only a single generation step of LLM, and we denote $\mathcal{T}(g \leftarrow f) = P(g|f)$ as the probability of the agent transitioning from a template containing state f to an output containing state g through LLM generation. A schematic diagram is shown in Fig. 1.

LLM-based agents are characterized by their state transitions not being entirely random but exhibiting a certain structured preference. Specifically, agents tend to transition from the current state f to states g that are “better” from the agent’s perspective. To capture this phenomenon, we hypothesize the existence of an underlying potential function $V_{\mathcal{T}} : \mathcal{C} \rightarrow \mathbb{R}$, which assigns a scalar value to each state, reflecting its “quality.” Since a specific potential function is often difficult to compute directly, we propose a method to effectively estimate the potential function.

Given a global potential function V , we define the violation of the agent’s given transition $\mathcal{T}(g \leftarrow f)$ to the potential function as $K(V(f) - V(g))$, where $K(x)$ is a convex function that describes the extent to which the transition from state f to state g violates the ordering of the potential function V . To quantify the overall mismatch between the agent’s behavior and the potential function, we weight by the transition kernel $\mathcal{T}(g \leftarrow f)$ and define the action $\mathcal{S}$ as the global average violation:

$$\mathcal{S} = \int_{f \in \mathcal{C}} \int_{g \in \mathcal{C}} \mathcal{T}(g \leftarrow f) K(V(f) - V(g)) Df Dg, \quad (1)$$

where Df, Dg are measures on the state space ($\int Df \equiv \sum_f$ for the discrete spaces in this paper). We choose $K(x) = \exp(-\beta x/2)$ as the convex function describing the violation of the given state transition from f to g in the ordering of the scalar function V . Here, β is a scaling factor; for a single agent, it can be absorbed into V , but when comparing across agents it acquires physical meaning through majority voting (see Sec. S4 of the Supplemental Material [42]). The action $\mathcal{S}$ or the distribution shape of $\beta V(f)$ can represent the agent’s global cognition ability within this state space $\mathcal{C}$.

We propose that to quantify the behavior of LLMs using a potential function, one can seek such a potential function that minimizes the overall mismatch between the agent’s transitions and the potential function [38,39]. Therefore, the most suitable potential function $V_{\mathcal{T}}$ for describing an LLM-based agent $\mathcal{T}$ in a given state space is the one that minimizes the action $\mathcal{S}$ [37,40,43,44].

This implies that the action satisfies the variational principle with respect to the potential function $V_{\mathcal{T}}$ [37,45]:

$$\delta \mathcal{S} = 0. \quad (2)$$

The variational condition is equivalent to $V_{\mathcal{T}}$ satisfying the following equilibrium condition:

$$\begin{aligned} & \int_{g \in \mathcal{C}} \mathcal{T}(g \leftarrow f) K'(V_{\mathcal{T}}(f) - V_{\mathcal{T}}(g)) Dg \\ & - \int_{h \in \mathcal{C}} \mathcal{T}(f \leftarrow h) K'(V_{\mathcal{T}}(h) - V_{\mathcal{T}}(f)) Dh = 0, \end{aligned} \quad (3)$$

holding for all $f \in \mathcal{C}$, where $K'(x) = \frac{dK}{dx}$.

Specifically, if for all transitions $\mathcal{T}(g \leftarrow f) > 0$, $V(f) \geq V(g)$ holds, it indicates that the agent’s state transitions are completely ordered, and in this case, V serves as a Lyapunov function [46,47].

It is worth noting that if $\mathcal{T}$ describes the transition of an equilibrium system, its state transitions satisfy the detailed balance condition; i.e., for all state pairs (f, g) , the following

holds [31,33]:

$$\pi(f)P(g|f) = \pi(g)P(f|g), \quad (4)$$

where $\pi(f)$ denotes the equilibrium distribution of the system at state f and $P(g|f)$ denotes the transition kernel. In this case, there exists a potential function V that can explicitly express the detailed balance as

$$\log \frac{\mathcal{T}(g \leftarrow f)}{\mathcal{T}(f \leftarrow g)} = \beta V(f) - \beta V(g). \quad (5)$$

Unless otherwise specified, $\log$ denotes the natural logarithm (base e). Substituting into Eq. (3), it can be verified that this potential function $V = V_{\mathcal{T}}$ satisfies the least action principle. This indicates that for equilibrium systems, if the detailed balance condition exists, the corresponding underlying potential function can be estimated through the least action principle. In general cases, the least action merely seeks the most ordered arrangement of the potential function, minimizing the violations of this arrangement by the agent's state transitions [44].

The Markov formulation is well motivated: LLM-agent systems routinely run far beyond the context window [7,12,18–20,48], necessitating a distill-and-generate protocol that makes the memoryless approximation a practical description of the dynamics [49].

The central point of this paper is that LLM-based agents often behave like an equilibrium system in their LLM-generated state space, which is coarse grained compared to the complete generation sequence of LLMs [24,50]. The existence of this phenomenon suggests a universal macroscopic law in LLM generative dynamics that does not depend on specific model or task details. It indicates that despite being seemingly unrelated, there are underlying connections between different LLM generative processes, allowing us to describe the global ordering in LLM generation through the potential function $V_{\mathcal{T}}$, thereby providing explanations for the internal dynamics of LLMs.

III. EXPERIMENTS

We conducted experiments on three different models, including GPT-5 Nano, Claude Sonnet 4.5, and Gemini 2.5 Flash [51–53]. Each model was prompted to generate a new word based on a given prompt word such that the sum of the letter indices of the new word equals 100. For example, given the prompt “WIZARDS(23 + 9 + 26 + 1 + 18 + 4 + 19 = 100),” the model needs to generate a new word whose letter indices also sum to 100, such as “BUZZY(2 + 21 + 26 + 26 + 25 = 100).” The transition kernel between two prompt words can be estimated through sampling as

$$\mathcal{T}(g \leftarrow f) \approx \frac{N(g \leftarrow f)}{N_0(f)}, \quad (6)$$

where $N(g \leftarrow f)$ denotes the number of times the model generated the word g from the prompt word f . $N_0(f)$ represents the number of sampling attempts starting from the prompt word f . Each model performed 20 000 generations. More details of the experiments are provided in Sec. S2 of the Supplemental Material [42].

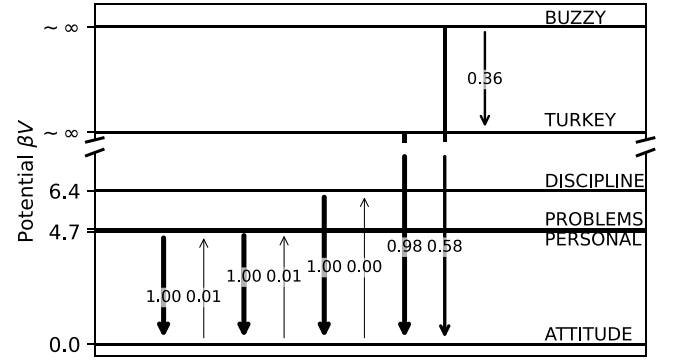


FIG. 2. In the conditioned word generation task, the Claude Sonnet 4.5 model exhibits directionality. Transition process of the Claude Sonnet 4.5 model in the prompt word state space, ordered by the potential function $V_{\mathcal{T}}$. Transitions tend to move toward states with lower potentials. States with $\beta V(f) \gg \ln(20\,000) \sim 10$ are those where the equilibrium condition cannot be strictly satisfied.

The three models exhibited two different behaviors, demonstrating directionality and certain diversity in actual LLM generative dynamics. Claude Sonnet 4.5 and Gemini 2.5 Flash demonstrated rapid convergence, with generated words quickly concentrating on a few high-frequency words. For instance, in 20 000 generations, Claude Sonnet 4.5 generated only 5 valid prompt words, while Gemini 2.5 Flash generated 13 valid prompt words. In contrast, GPT-5 Nano exhibited stronger exploration, producing as many as 645 different valid prompt words in 20 000 generations. This difference reflects the exploration-exploitation trade-off in LLM generative dynamics, making them suitable for task scenarios with varying demands for exploration and stability.

In Claude Sonnet 4.5 and Gemini 2.5 Flash, the solution to the variational condition (2) can be calculated analytically. In this case, we can plot the transition process of Claude Sonnet 4.5 ordered by the potential function as shown in Fig. 2, with specific calculations provided in Sec. S3 of the Supplemental Material [42].

Since Claude Sonnet 4.5 and Gemini 2.5 Flash exhibited high convergence, the equilibrium condition (3) is almost equivalent to the detailed balance condition (5). It is worth noting that we observed Claude Sonnet 4.5 starting from the prompt word “ATTITUDE,” which has the lowest potential function, began to attempt some invalid words, while Gemini 2.5 Flash oscillated between the two lowest potential functions “ATTITUDE” and “DISCIPLINE,” losing exploration. This behavior is similar to the low-temperature trapping phenomenon in physical systems [33,54], suggesting that controlling the potential function may provide a feasible path to avoid model convergence.

A key example of interest is the GPT-5 Nano model. GPT-5 Nano generated a large number of prompt words due to its strong exploration, allowing us to directly test the detailed balance condition within the state space. We note that according to detailed balance, the sum of potential changes along any closed path should be zero. Specifically, considering a closed path $f_1 \rightarrow f_2 \rightarrow \dots \rightarrow f_n \rightarrow f_1$, according to the detailed

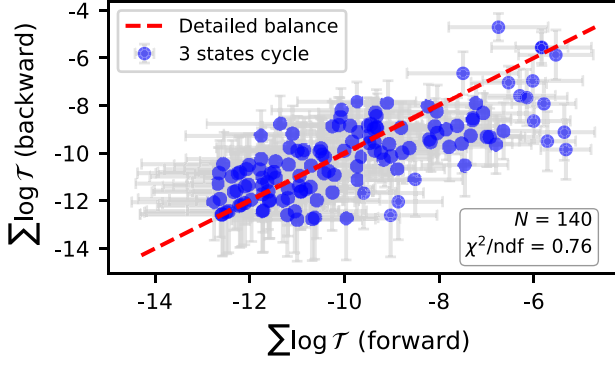


FIG. 3. In the task of conditioned word generation without a reasoning chain, verification of detailed balance through closed paths in the state transition graph of the GPT-5 Nano model. Each point represents a triplet, with all different triplets found in the experimental data. The error bars directly arise from sampling errors. χ^2/ndf is the chi-square per degree of freedom. Note that both axes are in natural logarithm ($\ln$), not $\log_{10}$.

balance condition, we have

$$\sum_{i=1}^n \ln \frac{\mathcal{T}(f_{i+1} \leftarrow f_i)}{\mathcal{T}(f_i \leftarrow f_{i+1})} = \sum_{i=1}^n \beta(V(f_i) - V(f_{i+1})) = 0. \quad (7)$$

Figure 3 counts all triplets in the experimental data, where transitions between each pair of the three data points were measured, totaling 140 different triplets. Each point represents a comparison of the sum of the logarithms of the forward and reverse transition kernels for a triplet, thereby testing the detailed balance condition. The measurement points cluster around the diagonal line, indicating that within the error range, the two sums are approximately equal, consistent with the detailed balance condition.

To further test the universality of detailed balance in LLM generation, we now construct an agent $\mathcal{T}_{\text{real}}$ with a long reasoning chain, whose states are strings that can be parsed into specific expression trees. We recorded 50 228 state transitions executed by this agent, constructing a database containing 21 697 different transitions and 7484 different states. Some example states are shown in Table I. By analyzing these transitions, we can statistically test the detailed balance condition and estimate its underlying potential function $V_{\mathcal{T}_{\text{real}}}$ through the least action principle. This agent involves multiple different LLMs and prompt templates, so its potential function characteristics may reflect typical behaviors of LLM generation in

TABLE I. Examples of some states of the agent $\mathcal{T}_{\text{real}}$ and their potentials. The independent variable $\log_v_k_nu$ in the fitting task is abbreviated as x .

States f	Potential
param1 * tanh(param2 * x + param3) + param4	5.70
param1 - (param2 / (x + param3))	0.88
param1 * x / (1 + param2 * log(x + 1))	-0.57
param1 * tanh(param2 * x) + param3	-1.57
param2 + param1 * (1 - exp(-x))	-3.30

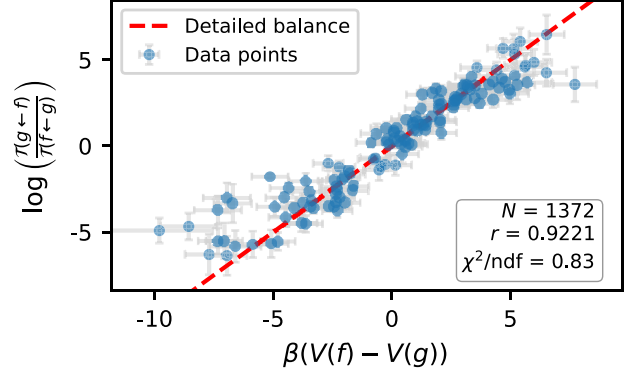


FIG. 4. In the task of symbolic fitting with a long reasoning chain, test of the detailed balance condition for the agent $\mathcal{T}_{\text{real}}$. The error estimation includes root mean square statistical errors, and the error of the potential function is estimated through bootstrapping. r is the Pearson correlation coefficient, and χ^2/ndf is the chi-squared per degree of freedom. Measurement includes state pairs with at least one measured transition between the two states. Points with $|V_{\mathcal{T}}(f) - V_{\mathcal{T}}(g)| > \ln(50\,288)$ are excluded. One-tenth of the data points are displayed for clarity.

practical applications. More experimental details are provided in Sec. S2 of the Supplemental Material [42].

Similar to the measurement described above, we counted all triplets in the experimental data, where transitions between each pair of the three data points were measured, and for all different triplets, we observed consistency with the detailed balance condition within the measurement error range. The results are provided in Sec. S2 of the Supplemental Material [42].

We now estimate the underlying potential function $V_{\mathcal{T}_{\text{real}}}$ through the least action principle. In a discrete state space, the integrals are replaced by sums over states, so the action [Eq. (1)] becomes, normalized by the number of states in the database,

$$\mathcal{S} = \frac{\sum_{g \leftarrow f} K(V_{\mathcal{T}_{\text{real}}}(f) - V_{\mathcal{T}_{\text{real}}}(g))}{\sum_f 1}. \quad (8)$$

By numerically minimizing this action, we can estimate the potential function values $V_{\mathcal{T}_{\text{real}}}(f)$ for each state. According to this estimation, the minimum value of the action is much smaller than $K(0)$, indicating that the state transitions of the agent $\mathcal{T}_{\text{real}}$ indeed exhibit directionality. In Sec. S4 of the Supplemental Material [42], we show that this optimized action value can be used to quantify the density distribution of states with respect to the underlying potential function, thereby providing direct guidance for designing more effective LLM generation strategies in practice.

By estimating this potential function, we can check whether the potential function is consistent with the detailed balance condition. Specifically, the detailed balance condition requires that for all state pairs (f, g) , Eq. (5) holds.

Figure 4 shows a comparison of the left and right sides of Eq. (5). We use the Pearson correlation coefficient r between the two sides as the primary quantitative measure of how well detailed balance is satisfied; here, $r = 0.92$, indicating that the detailed balance condition holds to good approximation.

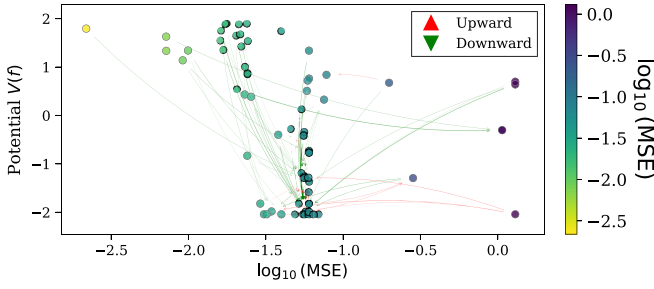


FIG. 5. Illustration of the ability of the potential function discovered using IdeaSearch [55] to predict the directionality of state transitions. Each point represents a transition pair from state f to state g . The figure shows a subgraph composed of 70 states selected from the database, displaying transitions with high transition kernel $\mathcal{T}_{\text{real}}(g \leftarrow f) > 0.05$. Red and green lines represent transitions with increasing and decreasing potential functions, respectively. The horizontal axis represents the mean square error of the expression corresponding to the state in the symbolic fitting task, while the vertical axis represents the potential function.

With the estimation of the potential function, we further discuss the specific meaning of the potential function in this problem to reveal the intrinsic cognitive characteristics in LLM generative dynamics. To this end, we use the action as an optimization objective and employ a workflow based on IdeaSearch [55] to find a potential function with an explicit functional form that maps the expression strings corresponding to state f to scalar potential function values $V_{\mathcal{T}}(f)$. The best potential function found in 4000 rounds of search contains 49 parameters, capturing various features of state f at the expression level, such as complexity, syntactic validity, and structural affinity with domain-specific patterns, without capturing string-level information. The magnitude of the corresponding parameter values directly reflects the importance that LLMs attach to these features during the generation process. The potential function and its specific analysis are provided in Sec. S5 of the Supplemental Material [42].

Figure 5 shows the transition patterns between some states sorted by this potential function. In the database, there are a total of 9769 transitions with high transition kernel $\mathcal{T}_{\text{real}}(g \leftarrow f) > 0.05$, and the corresponding potentials $V(f)$, $V(g)$ are calculated using the potential function. Among them, 6795 (69.56%) exhibit a decrease in the potential function, 2523 (25.83%) exhibit an increase in the potential function, and 451 (4.62%) exhibit no change in the potential function. This indicates that the potential function partially captures the intrinsic directionality in LLM generative dynamics. It is worth emphasizing that LLMs overall tend to choose states with relatively low potential function values as the next state, even

though they may not necessarily perform better on actual data. In this way, the potential function can reveal the differences between the intrinsic cognition of LLMs and real data. Additional experiments on diverse tasks further support the generality of this conclusion (see Sec. S2 of the Supplemental Material [42]).

IV. CONCLUSION AND OUTLOOK

In this paper, we have proposed a framework based on the least action principle to describe and analyze the generative dynamics of LLM-based agents in their LLM-generated state space. Through experiments on multiple different models and tasks, we have found that the state transitions of these agents largely satisfy the detailed balance condition, indicating that their generative dynamics exhibit characteristics similar to equilibrium systems. We have further estimated the underlying potential function through the least action principle and revealed its important role in capturing the intrinsic directionality in LLM generative dynamics.

This paper provides a preliminary exploration of the possibility for discovering macroscopic laws in LLMs generative dynamics. Future work can further expand this framework and explore the application potential of more tools from equilibrium and near-equilibrium systems in understanding and optimizing LLM generation processes. For instance, studying the degree of deviation from equilibrium may help us understand a model's level of overfitting, as overfitted models may learn more localized strategy sets rather than global generative patterns governed by potential functions [35,36]. Additionally, optimization methods based on potential functions may also provide new ideas for improving the quality and diversity of LLM task-related generation, such as adjusting the action to different magnitudes based on varying safety and exploration requirements.

ACKNOWLEDGMENTS

We would like to thank Z. Cai, J. Wei, S. Qiu, S. Zhang, J. Pan, Z. Lin, and L. Wang for useful discussions. This work was supported by the National Natural Science Foundation of China under Contracts No. 12425505, No. 12235001, and No. U2230402. This work was also supported by the Fundamental and Interdisciplinary Disciplines Breakthrough Plan of the Ministry of Education of China-JYB2025DXDM204.

DATA AVAILABILITY

The code used to perform the analysis in this paper is publicly available on GitHub [56]. The data discussed in this paper are available on Hugging Face [57].

[1] J. Wei, X. Wang, D. Schuurmans, M. Bosma, B. Ichter, F. Xia, E. H. Chi, Q. V. Le, and D. Zhou, Chain-of-thought

prompting elicits reasoning in large language models, in *Advances in Neural Information Processing Systems*, edited by S.

- Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Curran Associates, Inc., Red Hook, NY, 2022), Vol. 35, pp. 24824–24837.
- [2] S. Yao, J. Zhao, D. Yu, N. Du, I. Shafran, K. Narasimhan, and Y. Cao, ReAct: Synergizing reasoning and acting in language models (ICLR, 2023), https://openreview.net/forum?id=WE_vluYUL-X.
 - [3] S. Yao, D. Yu, J. Zhao, I. Shafran, T. L. Griffiths, Y. Cao, and K. Narasimhan, Tree of Thoughts: Deliberate problem solving with large language models, *Adv. Neural Inf. Process. Syst.* **36**, 11809 (2023).
 - [4] X. Wang, J. Wei, D. Schuurmans, Q. Le, E. Chi, S. Narang, A. Chowdhery, and D. Zhou, Self-consistency improves chain of thought reasoning in language models (ICLR, 2023), <https://openreview.net/forum?id=IPL1NIMMrw>.
 - [5] L. Wang, W. Xu, Y. Lan, Z. Hu, Y. Lan, R. K.-W. Lee, and E.-P. Lim, Plan-and-solve prompting improving zero-shot chain-of-thought reasoning by large language models, in *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics*, Vol. 1: Long Papers (2023), pp. 2609–2634.
 - [6] T. Schick, J. Dwivedi-Yu, R. Dessì, R. Raileanu, M. Lomeli, E. Hambro, L. Zettlemoyer, N. Cancedda, and T. Scialom, Toolformer: Language models can teach themselves to use tools, *Adv. Neural Inf. Process. Syst.* **36**, 68539 (2023).
 - [7] G. Wang, Y. Xie, Y. Jiang, A. Mandlekar, C. Xiao, Y. Zhu, L. Fan, and A. Anandkumar, Voyager: An open-ended embodied agent with large language models, *Trans. Mach. Learn. Res.* (2024).
 - [8] W. Chen, X. Ma, X. Wang, and W. W. Cohen, Program of thoughts prompting: Disentangling computation from reasoning for numerical reasoning tasks, *Trans. Mach. Learn. Res.* (2023).
 - [9] L. Gao, A. Madaan, S. Zhou, U. Alon, P. Liu, Y. Yang, J. Callan, and G. Neubig, PAL program-aided language models, in *Proceedings of the 40th International Conference on Machine Learning*, edited by A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett (PMLR, 2023), Vol. 202, pp. 10764–10799.
 - [10] J. Liang, W. Huang, F. Xia, P. Xu, K. Hausman, B. Ichter, P. Florence, and A. Zeng, Code as policies language model programs for embodied control, in *Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, Piscataway, NJ, 2023), pp. 9493–9500.
 - [11] D. J. Mankowitz, A. Michi, A. Zhernov, M. Gelmi, M. Selvi, C. Paduraru, E. Leurent, S. Iqbal, J.-B. Lespiau, A. Ahern, *et al.*, Faster sorting algorithms discovered using deep reinforcement learning, *Nature (London)* **618**, 257 (2023).
 - [12] J. S. Park, J. O’Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, Generative agents interactive simulacra of human behavior, in *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology, UIST’23* (ACM Press, New York, 2023).
 - [13] M. Besta, N. Blach, A. Kubicek, R. Gerstenberger, M. Podstawski, L. Gianinazzi, J. Gajda, T. Lehmann, H. Niewiadomski, P. Nyczyk, and T. Hoefler, Graph of thoughts: Solving elaborate problems with large language models, *Proc. AAAI Conf. Artif. Intell.* **38**, 17682 (2024).
 - [14] M. Mohammadi, Y. Li, J. Lo, and W. Yip, Evaluation and benchmarking of LLM agents: A survey, in *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, KDD’25* (ACM Press, New York, 2025), pp. 6129–6139.
 - [15] R. Sapkota, K. I. Roumeliotis, and M. Karkee, AI agents vs. agentic AI: A conceptual taxonomy, applications and challenges, *Inf. Fusion* **126**, 103599 (2026).
 - [16] D. A. Boiko, R. MacKnight, B. Kline, and G. Gomes, Autonomous chemical research with large language models, *Nature (London)* **624**, 570 (2023).
 - [17] A. M. Bran, S. Cox, O. Schilter, C. Baldassari, A. D. White, and P. Schwaller, Augmenting large language models with chemistry tools, *Nat. Mach. Intell.* **6**, 525 (2024).
 - [18] B. Romera-Paredes, M. Barekatin, A. Novikov, M. Balog, M. P. Kumar, E. Dupont, F. J. R. Ruiz, J. S. Ellenberg, P. Wang, O. Fawzi, P. Kohli, and A. Fawzi, Mathematical discoveries from program search with large language models, *Nature (London)* **625**, 468 (2024).
 - [19] A. Novikov, N. Vü, M. Eisenberger, E. Dupont, P.-S. Huang, A. Z. Wagner, S. Shirobokov, B. Kozlovskii, F. J. R. Ruiz, A. Mehrabian, M. P. Kumar, A. See, S. Chaudhuri, G. Holland, A. Davies, S. Nowozin, P. Kohli, and M. Balog, AlphaEvolvo: A coding agent for scientific and algorithmic discovery, *arXiv:2506.13131*.
 - [20] Z.-Y. Song, Z. Cai, S. Zhang, J. Wei, J. Pan, Shi Qiu, Q.-H. Cao, T.-J. Hou, X. Liu, M.-x. Luo, and H. X. Zhu, Iterated agent for symbolic regression, *Commun. Theor. Phys.* (2026), doi:10.1088/1572-9494/aea105.
 - [21] N. Bhattacharya, N. Thomas, R. Rao, J. Dauparas, P. K. Koo, D. Baker, Y. S. Song, and S. Ovchinnikov, Interpreting Potts and transformer protein models through the lens of simplified attention, in *Pacific Symposium on Biocomputing* (World Scientific, Singapore, 2022), Vol. 27, pp. 34–45.
 - [22] Y. Sun and B. Haghighat, Phase transitions in large language models and the $O(N)$ model, *arXiv:2501.16241*.
 - [23] Z. Liu, Y. Liu, J. Gore, and M. Tegmark, Neural thermodynamic laws for large language model training, *arXiv:2505.10559*.
 - [24] E. P. Hoel, When the map is better than the territory, *Entropy* **19**, 188 (2017).
 - [25] J. J. Hopfield, Neural networks and physical systems with emergent collective computational abilities, *Proc. Natl. Acad. Sci. USA* **79**, 2554 (1982).
 - [26] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, E. H. Chi, T. Hashimoto, O. Vinyals, P. Liang, J. Dean, and W. Fedus, Emergent abilities of large language models, *Trans. Mach. Learn. Res.* (2022), <https://openreview.net/forum?id=yzkSU5zdWd>.
 - [27] R. Schaeffer, B. Miranda, and S. Koyejo, Are emergent abilities of large language models a mirage? in *Advances in Neural Information Processing Systems*, edited by A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (Curran Associates, Inc., Red Hook, NY, 2023), Vol. 36, pp. 55565–55581.
 - [28] Y. Kim, K. Gu, C. Park, C. Park, S. Schmidgall, A. A. Heydari, Y. Yan, Z. Zhang, Y. Zhuang, Y. Liu, M. Malhotra, P. P. Liang, H. W. Park, Y. Yang, X. Xu, Y. Du, S. Patel, T. Althoff, D. McDuff, and X. Liu, Capable language models can outgrow the benefits of collaboration, *Nat. Mach. Intell.* **8**, 1157 (2026).
 - [29] S. Hu, X. Cai, Y. Huang, Z. Yao, L. Zhang, P. Zhang, Y. Deng, and K. Chen, Emergent slow thinking in LLMs as inverse tree freezing, *arXiv:2509.23629*.

- [30] A. Markov, Extension of the law of large numbers to quantities, depending on each other, *J. Électron. Hist. Probab. Stat.* **2**, 10 (2006).
- [31] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys.* **21**, 1087 (1953).
- [32] M. Mezard, G. Parisi, and M. Virasoro, *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications* (World Scientific, Singapore, 1986).
- [33] D. V. Schroeder, *An Introduction to Thermal Physics* (Oxford University Press, Oxford, 2021).
- [34] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction* (MIT Press, Cambridge, MA, 2018).
- [35] N. Tishby and N. Zaslavsky, Deep learning and the information bottleneck principle, in *2015 IEEE Information Theory Workshop (ITW)* (IEEE, Piscataway, NJ, 2015), pp. 1–5.
- [36] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, On large-batch training for deep learning: Generalization gap and sharp minima, in *International Conference on Learning Representations* (OpenReview.net, 2017).
- [37] H. Goldstein, C. P. Poole, and J. L. Safko, *Classical Mechanics* (Addison-Wesley, San Francisco, CA, 2002), 3rd ed.
- [38] K. Friston, The free-energy principle: A unified brain theory? *Nat. Rev. Neurosci.* **11**, 127 (2010).
- [39] A. Tschantz, M. Baltieri, A. K. Seth, and C. L. Buckley, Scaling active inference, in *2020 International Joint Conference on Neural Networks (IJCNN)* (IEEE, Piscataway, NJ, 2020), pp. 1–8.
- [40] D. B. West, *Introduction to Graph Theory* (Prentice Hall, Upper Saddle River, NJ, 2018), 2nd ed.
- [41] M. Yasunaga, X. Chen, Y. Li, P. Pasupat, J. Leskovec, P. Liang, E. H. Chi, and D. Zhou, Large language models as analogical reasoners, in *Proceedings of the International Conference on Learning Representations (ICLR, 2024)*, Vol. 2024, pp. 17019–17045.
- [42] See Supplemental Material at <https://link.aps.org/supplemental/10.1103/kwx6-clqd> for detailed derivations, agent implementation details, additional experiments, and supplementary figures, which includes Refs. [58–67].
- [43] T. Joachims, Optimizing search engines using clickthrough data, in *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD'02* (ACM Press, New York, 2002), pp. 133–142.
- [44] X. Jiang, L.-H. Lim, Y. Yao, and Y. Ye, Statistical ranking and combinatorial Hodge theory, *Math. Program.* **127**, 203 (2011).
- [45] We show in Sec. S1 of the Supplemental Material [42] that the variational principle is equivalent to the least action principle under the condition that $K(x)$ is a convex function.
- [46] A. M. Lyapunov, The general problem of the stability of motion, *Int. J. Control* **55**, 531 (1992).
- [47] S. H. Strogatz, *Nonlinear Dynamics and Chaos: With Applications to Physics, Biology, Chemistry, and Engineering* (CRC Press, Boca Raton, FL, 2015), 2nd ed.
- [48] E. Meyerson, G. Paolo, R. Dailey, H. Shahrzad, O. Francon, C. F. Hayes, X. Qiu, B. Hodjat, and R. Miikkulainen, Solving a million-step LLM task with zero errors, [arXiv:2511.09030](https://arxiv.org/abs/2511.09030).
- [49] With the Markov chain defined, time-reversal symmetry is well defined on its transition kernel as the equality of forward and reverse transition weights around every closed path, equivalent to detailed balance.
- [50] S. Weinberg, Phenomenological Lagrangians, *Physica A* **96**, 327 (1979).
- [51] OpenAI, GPT-5 system card, Technical Report, OpenAI, 2025.
- [52] Anthropic, System card: Claude Sonnet 4.5, Technical Report, Anthropic, 2025.
- [53] G. Comanici, E. Bieber, M. Schaeckermann, I. Pasupat, N. Sachdeva, I. Dhillon, M. Blistein, O. Ram, D. Zhang, E. Rosen, *et al.*, Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities, [arXiv:2507.06261](https://arxiv.org/abs/2507.06261).
- [54] J. C. Schön, Preferential trapping on energy landscapes in regions containing deep-lying minima: The reason for the success of simulated annealing? *J. Phys. A: Math. Gen.* **30**, 2367 (1997).
- [55] Ideasearch Collaboration, Ideasearch, 2025, <https://github.com/IdeaSearch/IdeaSearch-framework>.
- [56] Z.-Y. Song, Q.-H. Cao, M.-x. Luo, and H. X. Zhu, Detailed balance in large language model-driven agents: Code, GitHub software repository, 2025, <https://github.com/SonnyNondegeneracy/detailed-balance-llm>, MIT license.
- [57] Z.-Y. Song, Q.-H. Cao, M.-x. Luo, and H. X. Zhu, Detailed balance in large language model-driven agents: Data, Hugging Face dataset, 2026, <https://huggingface.co/datasets/Nondegeneracy/detailed-balance-llm>.
- [58] X. Cai, S. Hu, T. Wang, Y. Huang, P. Zhang, Y. Deng, and K. Chen, Learning-at-criticality in large language models for quantum field theory and beyond, *Chin. Phys. Lett.* **42**, 120002 (2025).
- [59] Q.-H. Cao, Z.-Y. Hou, Y.-Y. Li, X. Liu, Z.-Y. Song, L.-Q. Zhang, S. Zhang, and K. Zhao, Scalable quantum state preparation via large-language-model-driven discovery, [arXiv:2505.06347](https://arxiv.org/abs/2505.06347).
- [60] K. Singhal, S. Azizi, T. Tu, S. S. Mahdavi, J. Wei, H. W. Chung, N. Scales, A. Tanwani, H. Cole-Lewis, S. Pfohl, *et al.*, Large language models encode clinical knowledge, *Nature (London)* **620**, 172 (2023).
- [61] Z.-Y. Song, T.-Z. Yang, Q.-H. Cao, M.-x. Luo, and H. X. Zhu, Explainable AI-assisted optimization for Feynman integral reduction, *J. High Energy Phys.* **06** (2026) 225.
- [62] N. J. Szymanski, B. Rendy, Y. Fei, R. E. Kumar, T. He, D. Milsted, M. J. McDermott, M. Gallant, E. D. Cubuk, A. Merchant, H. Kim, A. Jain, C. J. Bartel, K. Persson, Y. Zeng, and G. Ceder, An autonomous laboratory for the accelerated synthesis of inorganic materials, *Nature (London)* **624**, 86 (2023).
- [63] Q. Chen, Y. Du, Z. Li, J. Liu, S. Duan, J. Guo, M. Liu, J. Liu, T. Yang, G. Zhang, L. Qin, W. Che, and W. Huang, The molecular structure of thought: Mapping the topology of long chain-of-thought reasoning, [arXiv:2601.06002](https://arxiv.org/abs/2601.06002).
- [64] A. Gladstone, G. Nanduru, M. M. Islam, P. Han, H. Ha, A. Chadha, Y. Du, H. Ji, J. Li, and T. Iqbal, Energy-based transformers are scalable learners and thinkers, [arXiv:2507.02092](https://arxiv.org/abs/2507.02092).
- [65] IdeaSearch Collaboration, IdeaSearchFitter, 2025, <https://github.com/IdeaSearch/IdeaSearch-fit>.
- [66] R. S. Olson, W. La Cava, P. Orzechowski, R. J. Urbanowicz, and J. H. Moore, PMLB: A large benchmark suite for machine learning evaluation and comparison, *BioData Mining* **10**, 36 (2017).
- [67] Q. Team, Qwen2.5 technical report, [arXiv:2412.15115](https://arxiv.org/abs/2412.15115).