
Appendix

Anonymous Author(s)

Affiliation

Address

email

1 A Proofs and Definitions

2 A.1 Proof of Theorem 1

3 **Theorem 1.** Consider a sequence of representations $\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T$ during an LLM’s reasoning
4 process, where T denotes the number of total reasoning steps. Let y, \hat{y} denote the golden answer and
5 the LLM’s prediction answer, respectively. Define $p_e = \Pr(\hat{y} \neq y)$ as the LLM’s prediction error
6 probability. Then the following inequality holds:

$$p_e \geq \frac{1}{\log(|\mathcal{Y}| - 1)} \left[H(y) - \sum_{j=1}^T I(y; \mathbf{h}_j \mid \mathbf{h}_{<j}) - H_b(p_e) \right], \quad (1)$$

7 where $|\mathcal{Y}|$ is the size of the support of y , and $H_b(p_e)$ denote the binary entropy of p_e that defined by

$$H_b(p_e) = -p_e \log p_e - (1 - p_e) \log(1 - p_e). \quad (2)$$

8 *Proof.* We first define an indicator random variable $E = \mathbf{1}\{\hat{y} \neq y\}$, where $E = 1$ if $\hat{y} \neq y$, and
9 $E = 0$ otherwise.

10 By the chain rule of entropy, we have:

$$\begin{aligned} H(y \mid \hat{y}) &= H(E \mid \hat{y}) + H(y \mid \hat{y}, E) \\ &= H(E \mid \hat{y}) + H(y \mid \hat{y}, E = 0) \Pr(E = 0) + H(y \mid \hat{y}, E = 1) \Pr(E = 1). \end{aligned} \quad (3)$$

11 Since $E = 0$ indicates $\hat{y} = y$, we have $H(y \mid \hat{y}, E = 0) = 0$. And for $H(E \mid \hat{y})$, we have:

$$H(E \mid \hat{y}) \leq H(E) := H_b(p_e). \quad (4)$$

12 Thus, we can derive:

$$H(y \mid \hat{y}) \leq H_b(p_e) + p_e H(y \mid \hat{y}, E = 1). \quad (5)$$

13 Since $E = 1$ indicates $\hat{y} \neq y$, the random variable y can take at most $|\mathcal{Y}| - 1$ values given \hat{y} as
14 condition. Hence, we have [5]:

$$H(y \mid \hat{y}) \leq H_b(p_e) + p_e \log(|\mathcal{Y}| - 1). \quad (6)$$

15 Based on the definition of mutual information, we have:

$$I(y; \hat{y}) = H(y) - H(y \mid \hat{y}). \quad (7)$$

16 Combining Eq. (6) and Eq. (7) derives:

$$p_e \geq \frac{1}{\log(|\mathcal{Y}| - 1)} \left[H(y) - I(y; \hat{y}) - H_b(p_e) \right]. \quad (8)$$

17 Consider an LLM's reasoning process, given the intermediate representations $\mathbf{h}_{1:T} =$
 18 $(\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T)$, the output \hat{y} is computed as a function of these representations $\hat{y} = f(\mathbf{h}_{1:T})$.
 19 Thus, based on the Data Processing Inequality (DPI), we have:

$$I(y; \hat{y}) \leq I(y; \mathbf{h}_{1:T}). \quad (9)$$

20 Combining Eq. (10) and Eq. (9), and applying the chain rule of mutual information, we have:

$$p_e \geq \frac{1}{\log(|\mathcal{Y}| - 1)} \left[H(y) - \sum_{j=1}^T I(y; \mathbf{h}_j | \mathbf{h}_{<j}) - H_b(p_e) \right], \quad (10)$$

21 which completes the proof. \square

22 A.2 Proof of Theorem 2

23 **Theorem 2.** *Following the notations in Theorem 1, the following inequality holds:*

$$p_e \leq \frac{1}{2} \left[H(y) - \sum_{j=1}^T I(y; \mathbf{h}_j | \mathbf{h}_{<j}) \right]. \quad (11)$$

24 *Proof.* The output of a reasoning model \hat{y} can be formulated as a multi-class classification task with
 25 predicted probabilities $p_i = \Pr(\hat{y} = i | \mathbf{h}_{1:T})$. According to Bayesian decision theory[3] [20], the
 26 conditional error probability is given by:

$$p_e = 1 - \max_i \{\Pr(y = i | \mathbf{h}_{1:T})\}. \quad (12)$$

27 For binary classification ($|\mathcal{Y}| = 2$), we have:

$$\min\{p, 1 - p\} \leq \frac{1}{2} [-p \log p - (1 - p) \log (1 - p)]. \quad (13)$$

28 Then take an expectation over p :

$$p_e = \mathbb{E}_p[\min\{p, 1 - p\}] \leq \frac{1}{2} \mathbb{E}_p[-p \log p - (1 - p) \log (1 - p)]. \quad (14)$$

29 So we derive:

$$p_e \leq \frac{1}{2} \mathbb{E}_{\mathbf{h}_{1:T}}[H(y | \mathbf{h}_{1:T})] = \frac{1}{2} H(y | \mathbf{h}_{1:T}). \quad (15)$$

30 This extends to multiclass problems through a recursive application (see Eq. (16)).

31 We prove the following inequality by mathematical induction that for any m -class discrete probability
 32 distribution $\{p_1, \dots, p_m\}$:

$$p_e = 1 - \max_i \{p_i\} \leq \frac{1}{2} H(p_1, \dots, p_m). \quad (16)$$

33 *Base case ($m = 2$):* Direct verification using binary entropy function Eq. (13).

34 *Inductive step:* Assume validity for m classes. For $m + 1$ classes, assume without loss of generality
 35 $p_{m+1} = \max_i \{p_i\}$. Consider the merged distribution $\{p_1, \dots, p_{m-1}, p_m + p_{m+1}\}$ and apply:

36 1. The induction hypothesis:

$$1 - (p_m + p_{m+1}) \leq \frac{1}{2} H(p_1, \dots, p_{m-1}, p_m + p_{m+1}). \quad (17)$$

37 2. The grouping axiom [2]:

$$H(p_1, \dots, p_{m+1}) = H(p_1, \dots, p_m + p_{m+1}) + (p_m + p_{m+1}) H\left(\frac{p_m}{p_m + p_{m+1}}, \frac{p_{m+1}}{p_m + p_{m+1}}\right). \quad (18)$$

38 3. Binary entropy bound for the final term:

$$1 - \frac{p_{m+1}}{p_m + p_{m+1}} \leq \frac{1}{2} H\left(\frac{p_m}{p_m + p_{m+1}}, \frac{p_{m+1}}{p_m + p_{m+1}}\right). \quad (19)$$

39 Combining Eq. (17), Eq. (18) and Eq. (19) completes the induction:

$$\begin{aligned}
\frac{1}{2}H(p_1, \dots, p_{m+1}) &= \frac{1}{2}H(p_1, \dots, p_m + p_{m+1}) + \frac{1}{2}(p_m + p_{m+1})H\left(\frac{p_m}{p_m + p_{m+1}}, \frac{p_{m+1}}{p_m + p_{m+1}}\right) \\
&\geq 1 - (p_m + p_{m+1}) + (p_m + p_{m+1})\left(1 - \frac{p_{m+1}}{p_m + p_{m+1}}\right) \\
&= 1 - p_{m+1} \\
&= 1 - \max_i \{p_i\}.
\end{aligned}$$

40 Thus, we have proved the Eq. (16).

41 Taking expectation over $h_{1:T}$ in Eq. (12) and applying the Eq. (16), we have

$$\begin{aligned}
p_e &= \mathbb{E}_{h_{1:T}}[1 - \max_i \{\Pr(y = i | h_{1:T})\}] \\
&\leq \frac{1}{2} \mathbb{E}_{h_{1:T}}[H(y | h_{1:T})] \\
&= \frac{1}{2} H(y | h_{1:T}) \\
&= \frac{1}{2} \left[H(y) - \sum_{j=1}^T I(y; h_j | h_{<j}) \right],
\end{aligned}$$

42 which completes the proof. \square

43 A.3 Definitions

44 **Definition 1** (Mutual Information [2, 10]). *Given two continuous random variables X and Y , the mutual information is defined as:*

$$I(X; Y) = \int_Y \int_X p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy, \quad (20)$$

46 where $p(x, y)$ denotes the joint probability density function of X and Y ; $p(x)$, $p(y)$ denotes the marginal probability density functions of X and Y , respectively.

48 **Definition 2** (Hilbert-Schmidt Independence Criterion (HSIC) [7]). *HSIC is the Hilbert-Schmidt norm of the cross-covariance operator between the distributions in Reproducing Kernel Hilbert Space (RKHS). Formally:*

$$\begin{aligned}
\text{HSIC}(X, Y) &= \mathbb{E}_{XYX'Y'} [k_X(X, X') k_Y(Y, Y')] + \mathbb{E}_{XX'} [k_X(X, X')] \mathbb{E}_{YY'} [k_Y(Y, Y')] \\
&\quad - 2\mathbb{E}_{XY} [\mathbb{E}_{X'} [k_X(X, X')] \mathbb{E}_{Y'} [k_Y(Y, Y')]],
\end{aligned} \quad (21)$$

51 where X', Y' are independent copies of X, Y , respectively, and k_X, k_Y are kernel functions.

52 B Experimental Implementation Details

53 **Practical implementation of HSIC.** Due to the difficulty of accurately computing MI in high-dimensional spaces [10, 13, 5], we employ the HSIC to estimate MI. Following [12, 15, 5], the empirical HSIC from Definition 2 is computed as

$$\text{HSIC}(X, Y) = \frac{1}{(n-1)^2} \text{tr}(K_X H K_Y H), \quad (22)$$

56 where K_X and K_Y are kernel matrices with entries

$$K_{X_{ij}} = k_X(x_i, x_j), \quad K_{Y_{ij}} = k_Y(y_i, y_j),$$

57 and $H = I - \frac{1}{n} \mathbf{1} \mathbf{1}^\top$ is the centering matrix. Consistent with [12, 15, 5], we adopt the Gaussian kernel to implement the kernel:

$$k(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right), \quad (23)$$

59 where the bandwidth σ is selected by grid search over the range [50, 400].

60 **Datasets.** 1) *Evaluation of LRMs’ reasoning performance.* We select three widely-used math
61 reasoning benchmarks to evaluate the reasoning capabilities of LRMs, ordering from easy to hard:
62 GSM8K [4], MATH500 [11], and AIME24 [1]. We adopt the evaluation framework provided by
63 Qwen2.5-Math [19]. To ensure the reproducibility of our results, we fix the temperature to 0 in all
64 experiments. 2) *Observing the MI trajectories during LRMs’ reasoning process.* We use the training
65 set of the MATH dataset [9]. Specifically, we randomly sample 100 instances to compute MI along
66 the reasoning trajectories.

67 **Models.** We conduct experiments on DeepSeek’s R1 series models [8] and QwQ-32B [17]. For
68 DeepSeek’s R1 series models, we pair each LRM with its corresponding non-reasoning LLM counter-
69 part as follows: DeepSeek-R1-Distill-Qwen-7B and Qwen2.5-Math-7B [19], DeepSeek-R1-Distill-
70 Llama-8B and Llama-3.1-8B [6], DeepSeek-R1-Distill-Qwen-14B and Qwen2.5-14B [18], DeepSeek-
71 R1-Distill-Qwen-32B and Qwen2.5-32B [18], DeepSeek-R1-Distill-Llama-70B and Llama-3.3-70B-
72 Instruct [6]. As observed, all LRMs in the R1 series are trained from foundation LLMs, except for
73 DeepSeek-R1-Distill-Qwen-7B, which is trained from a math-specialized LLM. As for QwQ-32B,
74 existing public report [17] has not disclosed which specific LLM it was trained from. All experiments
75 are conducted on four NVIDIA A100 GPUs.

76 **More implementation details.** For all experiments involving MI computation, we extract the rep-
77 resentation from the *last layer* of the model. We concentrate on the *last layer* since higher layers
78 have been shown to encode more semantic content [21, 16] and the *last layer* directly influence the
79 model’s output text [14]. For TTTS in Section 4.2, to ensure that the model begins continuation
80 with semantically meaningful tokens, we filter out tokens with little semantic information, such as
81 punctuation, single characters, etc. In this way, the resulting token list is: [So, Let, Hmm, I,
82 Okay, First, Wait, But, Now, Then, Since, Therefore, If, Maybe, To]. All exper-
83 iments are conducted on four NVIDIA A100 GPUs.

84 C Discussions

85 **Limitations.** This work has several limitations. First, we analyze the MI dynamics of LRMs at the
86 token level. Alternative granularities such as dividing reasoning steps by semantic units or logical
87 steps may reveal additional insights. Second, while we observe the interesting MI peaks phenomenon
88 and provide insights into the reasoning mechanisms of LRMs, the underlying mechanisms that give
89 rise to these peaks remain underexplored. We leave a deeper analysis of their origin to future work.
90 We hope that our work will inspire further research along these directions and contribute to a deeper
91 understanding of the reasoning process in LRMs.

92 **Broader impacts.** This work contributes to a deeper understanding of the reasoning mechanisms
93 in LRMs. We first observe the MI peaks phenomenon during LRMs’ reasoning process, and then
94 propose two simple training-free methods to enhance LRMs’ reasoning performance based on the
95 findings. These analyzes may have positive impacts by making AI systems more transparent and
96 effective. However, there are also potential risks. If used carelessly, the same methods could be
97 applied to manipulate outputs or reinforce biased thinking patterns. It is important to consider these
98 concerns when applying our techniques and to encourage responsible use through further study and
99 monitoring.

100 **Discussion on Tokens at MI Peaks.** As shown in Figure 4 in the main text and Figure 1 in the
101 appendix, different LRMs exhibit slightly different token frequency patterns at MI peaks. For
102 models trained from foundation LLMs, *i.e.*, DeepSeek-R1-Distill-Llama-8B, DeepSeek-R1-Distill-
103 Qwen-14B, DeepSeek-R1-Distill-Qwen-32B, and DeepSeek-R1-Distill-LLaMA-70B, the frequently
104 occurring tokens include So, Let, Hmm, The, and Okay. And for DeepSeek-R1-Distill-Qwen-7B,
105 which is trained from a math-specialized LLM, tokens such as So, The, Let, To, and, and Since
106 are more prominent. For QwQ-32B, tokens like To, the, we, and Let appear more frequently.
107 Semantically, these tokens commonly express reasoning-related functions such as initiating thinking
108 (So, Hmm), logical transition (Since, Therefore), or discourse structuring (Let, Then, To), which
109 likely help facilitate the model’s continued reasoning. We hypothesize that the distribution of tokens
110 at MI peaks may be influenced by factors such as the nature of the foundation LLM, the reasoning-
111 intensive training paradigm, etc. We leave a deeper investigation of the relationship among MI-peak

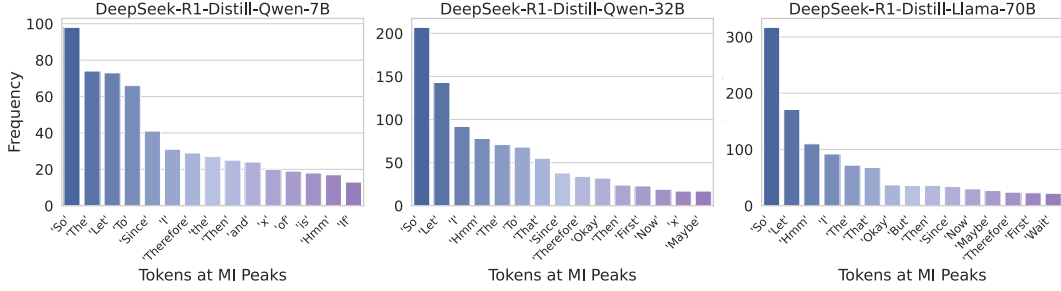


Figure 1: Frequency distribution of tokens at MI peaks for DeepSeek-R1-Distill-Qwen-7B, DeepSeek-R1-Distill-Qwen-32B, and DeepSeek-R1-Distill-Llama-70B.

token distributions, foundation LLM characteristics, reasoning-intensive training paradigms, and model reasoning performance to future work.

Further discussion on thinking token suppression (Figure 5 in the main text). As shown in Figure 5, while the overall trend indicates that LRMs’ reasoning performance degrades as more thinking tokens are suppressed, the decline is not strictly monotonic. In some cases, performance improves temporarily. We conduct an empirical analysis to better understand this phenomenon. Specifically, we observe that when certain tokens are suppressed, the model tends to adopt alternative expressions to convey similar meanings. For instance, when the generation of the token Wait is suppressed, the model may instead produce phrases like But wait, which could lead to slight improvements in performance. The observed performance fluctuations across different numbers of suppression tokens further support that these thinking tokens play a critical role in LRMs’ reasoning capabilities.

D Additional Experimental Results

D.1 MI Peaks in LRMs

Figures 2–13 illustrate the MI trajectories of various LRMs across more data samples.

References

- [1] AIME Problems and Solutions.
- [2] Robert B Ash. *Information theory*. Courier Corporation, 2012.
- [3] James O Berger. *Statistical decision theory and Bayesian analysis*. Springer Science & Business Media, 2013.
- [4] Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, et al. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*, 2021.
- [5] Zeyu Gan, Yun Liao, and Yong Liu. Rethinking external slow-thinking: From snowball errors to probability of correct reasoning. *arXiv preprint arXiv:2501.15602*, 2025.
- [6] Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, et al. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*, 2024.
- [7] Arthur Gretton, Olivier Bousquet, Alex Smola, and Bernhard Schölkopf. Measuring statistical dependence with hilbert-schmidt norms. In *International conference on algorithmic learning theory*, pages 63–77, 2005.
- [8] Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, et al. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*, 2025.

- [9] Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. Measuring mathematical problem solving with the MATH dataset. In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*, 2021.
- [10] Alexander Kraskov, Harald Stögbauer, and Peter Grassberger. Estimating mutual information. *Physical review E*, 69(6):066138, 2004.
- [11] Hunter Lightman, Vineet Kosaraju, Yuri Burda, Harrison Edwards, Bowen Baker, Teddy Lee, Jan Leike, John Schulman, Ilya Sutskever, and Karl Cobbe. Let’s verify step by step. In *The Twelfth International Conference on Learning Representations*, 2024.
- [12] Wan-Duo Kurt Ma, JP Lewis, and W Bastiaan Kleijn. The hsic bottleneck: Deep learning without back-propagation. In *Proceedings of the AAAI conference on artificial intelligence*, pages 5085–5092, 2020.
- [13] Ben Poole, Sherjil Ozair, Aaron Van Den Oord, Alex Alemi, and George Tucker. On variational bounds of mutual information. In *International Conference on Machine Learning*, pages 5171–5180, 2019.
- [14] Chen Qian, Dongrui Liu, Jie Zhang, Yong Liu, and Jing Shao. Dean: Deactivating the coupled neurons to mitigate fairness-privacy conflicts in large language models. *arXiv preprint arXiv:2410.16672*, 2024.
- [15] Chen Qian, Jie Zhang, Wei Yao, Dongrui Liu, Zhenfei Yin, Yu Qiao, Yong Liu, and Jing Shao. Towards tracing trustworthiness dynamics: Revisiting pre-training period of large language models. *arXiv preprint arXiv:2402.19465*, 2024.
- [16] Nina Rimskey, Nick Gabrieli, Julian Schulz, Meg Tong, Evan Hubinger, and Alexander Turner. Steering llama 2 via contrastive activation addition. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 15504–15522. Association for Computational Linguistics, 2024.
- [17] Qwen Team. Qwq-32b: Embracing the power of reinforcement learning, March 2025.
- [18] 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, Jiayi 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, 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. *arXiv preprint arXiv:2412.15115*, 2024.
- [19] An Yang, Beichen Zhang, Binyuan Hui, Bofei Gao, Bowen Yu, Chengpeng Li, Dayiheng Liu, Jianhong Tu, Jingren Zhou, Junyang Lin, et al. Qwen2. 5-math technical report: Toward mathematical expert model via self-improvement. *arXiv preprint arXiv:2409.12122*, 2024.
- [20] Zhi-Hua Zhou. *Machine learning*. Springer nature, 2021.
- [21] Andy Zou, Long Phan, Sarah Chen, James Campbell, Phillip Guo, Richard Ren, Alexander Pan, Xuwang Yin, Mantas Mazeika, Ann-Kathrin Dombrowski, et al. Representation engineering: A top-down approach to ai transparency. *arXiv preprint arXiv:2310.01405*, 2023.

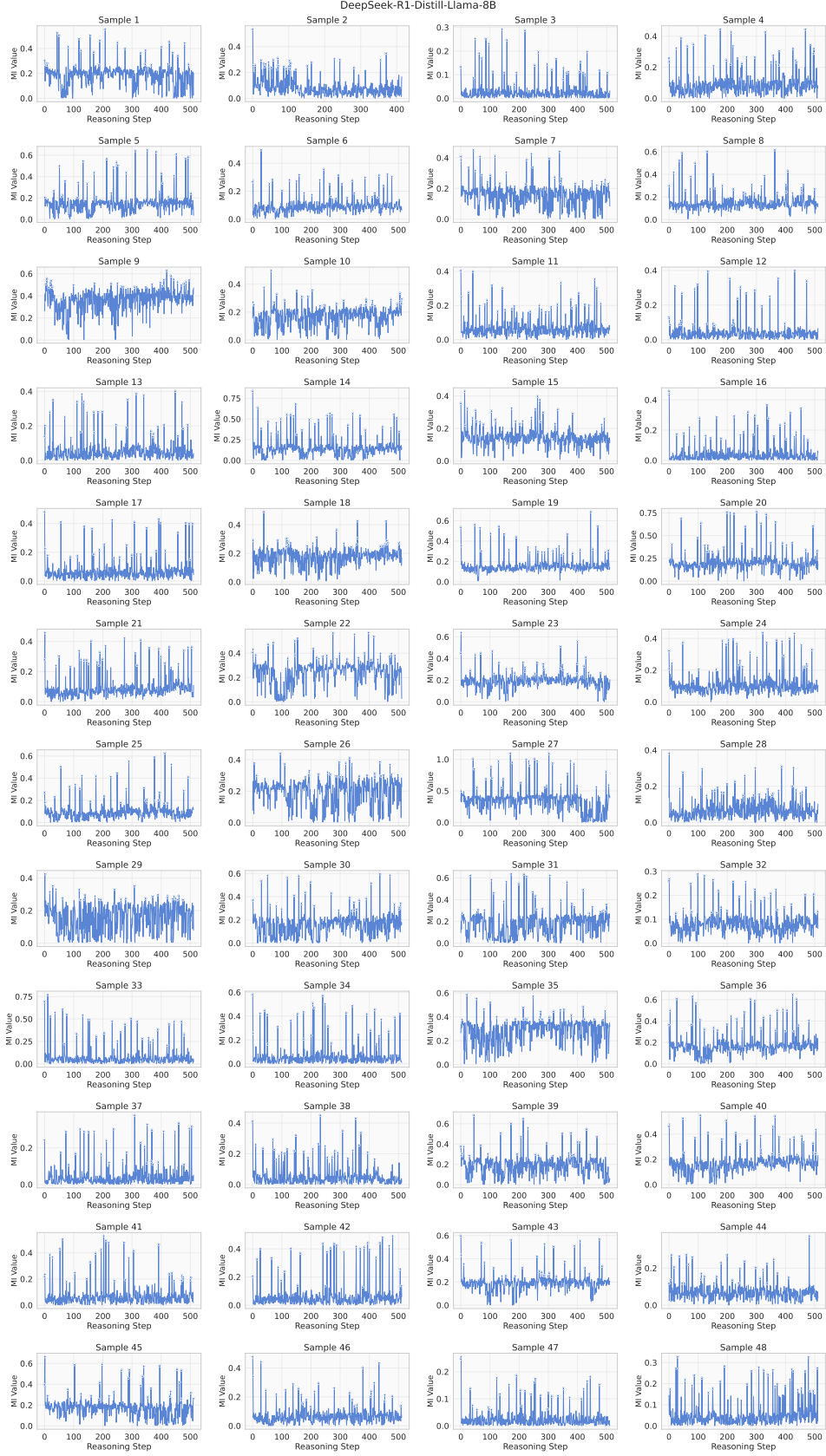


Figure 2: MI trajectories of DeepSeek-R1-Distill-Llama-8B.

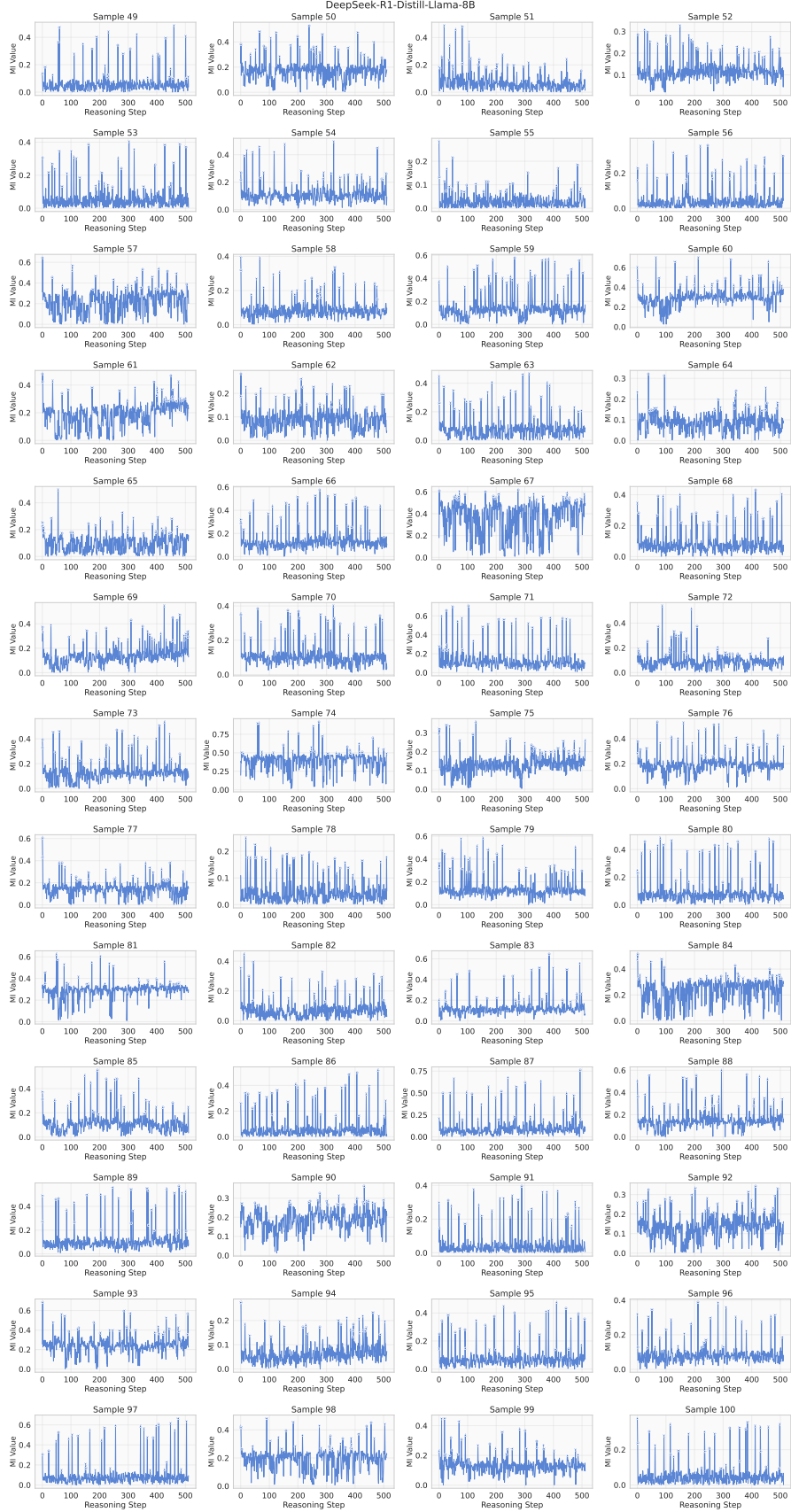


Figure 3: (Continued) MI trajectories of DeepSeek-R1-Distill-Llama-8B.

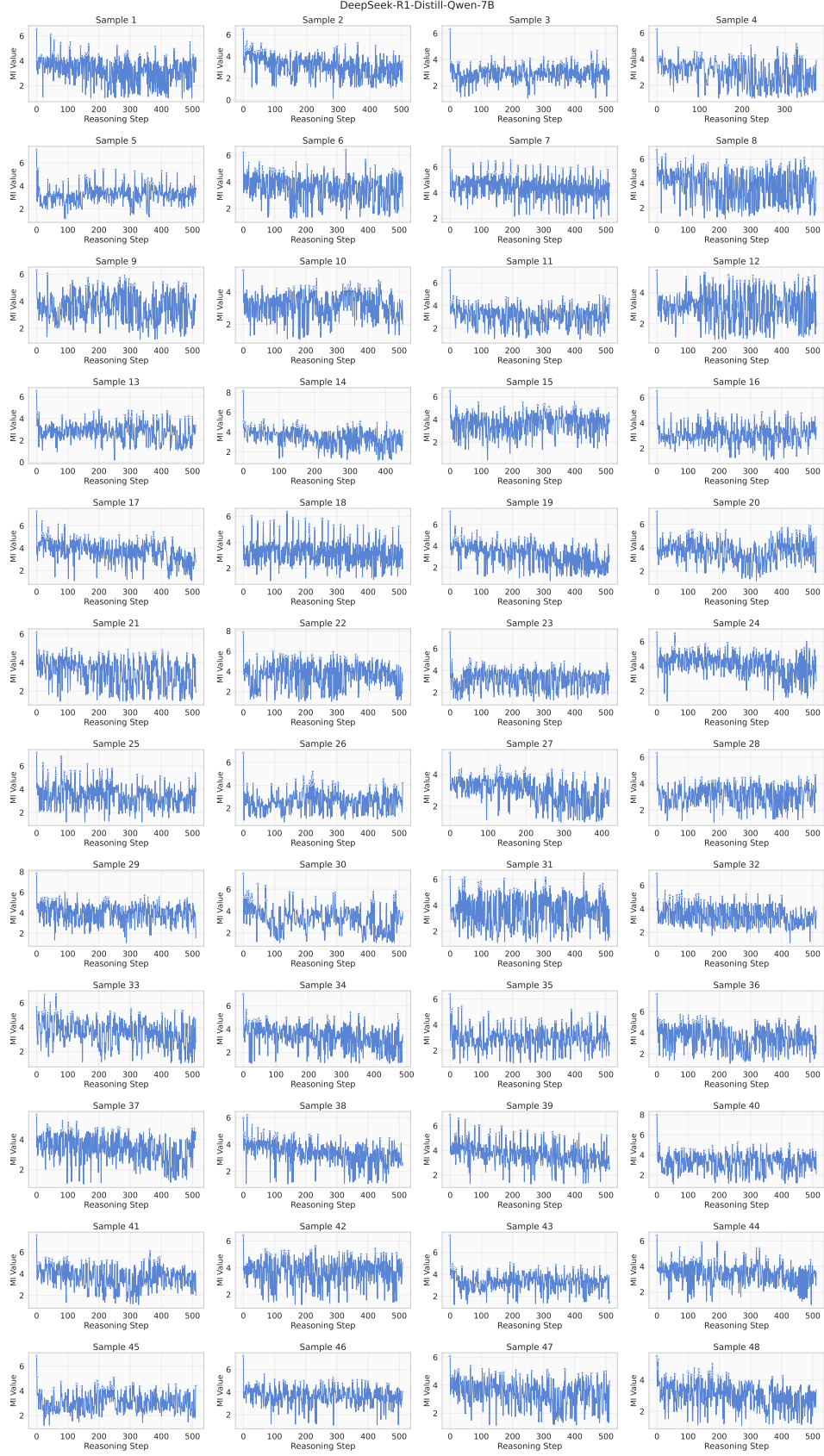


Figure 4: MI trajectories of DeepSeek-R1-Distill-Qwen-7B.

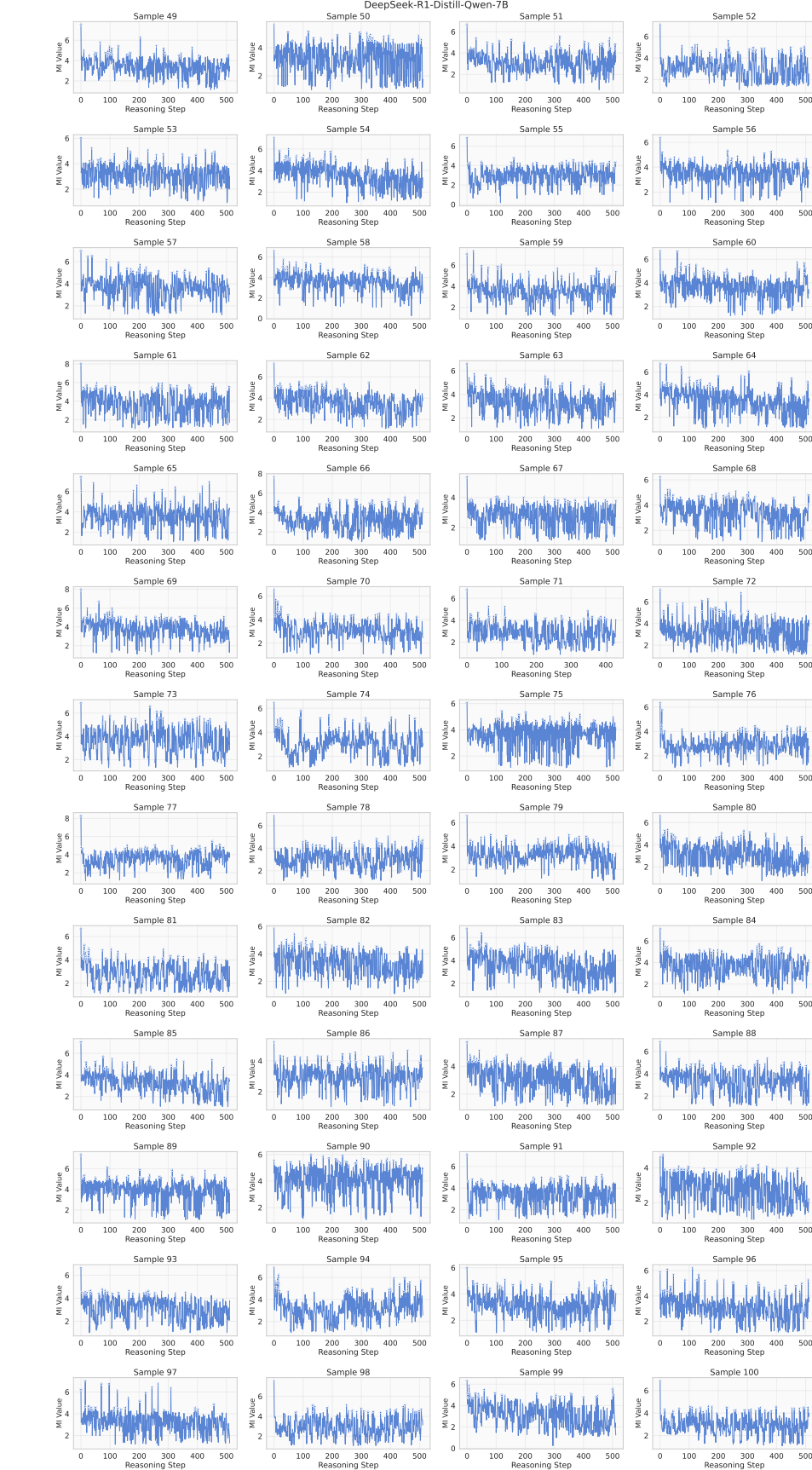


Figure 5: (Continued) MI trajectories of DeepSeek-R1-Distill-Qwen-7B.

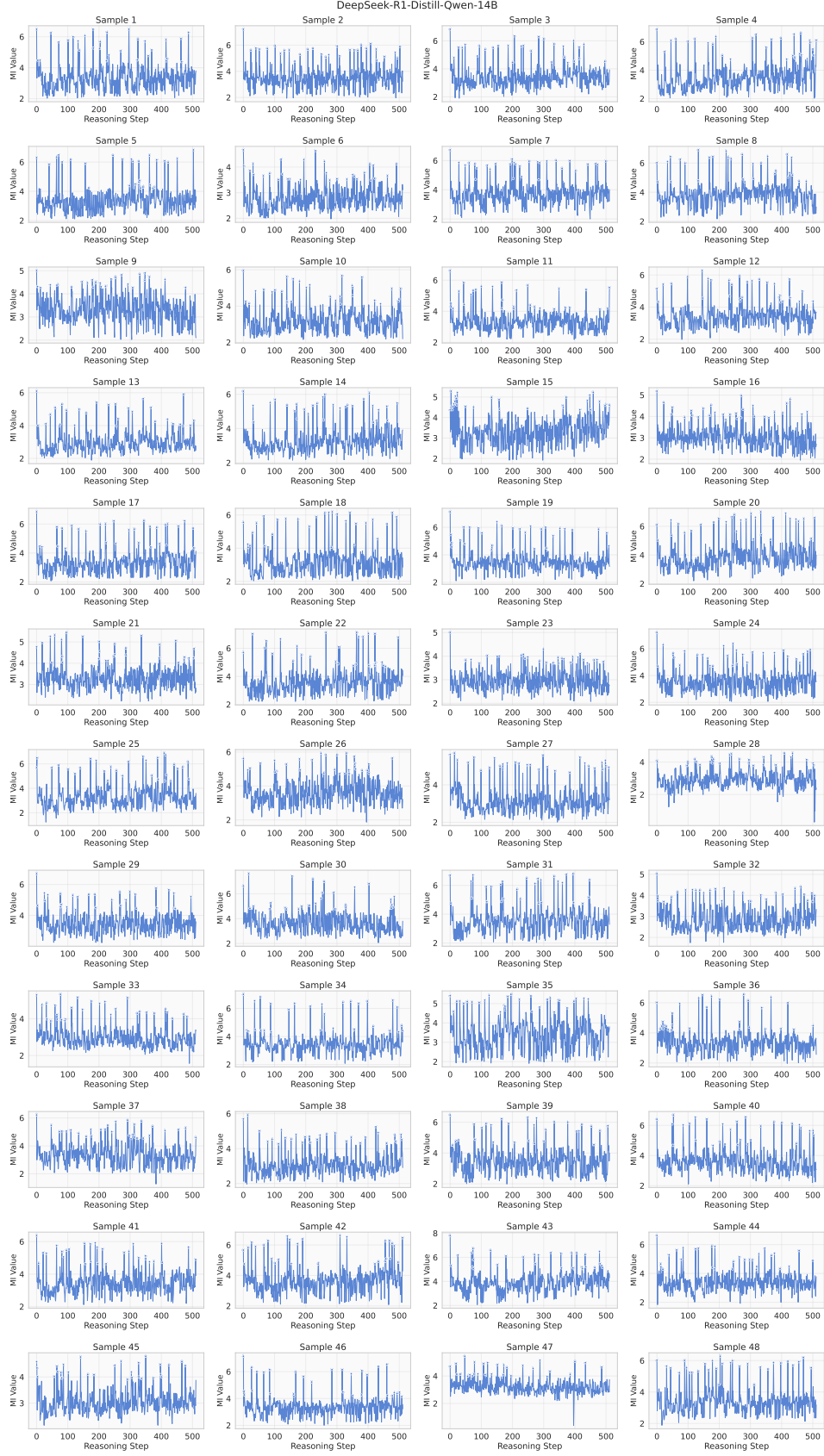


Figure 6: MI trajectories of DeepSeek-R1-Distill-Qwen-14B.

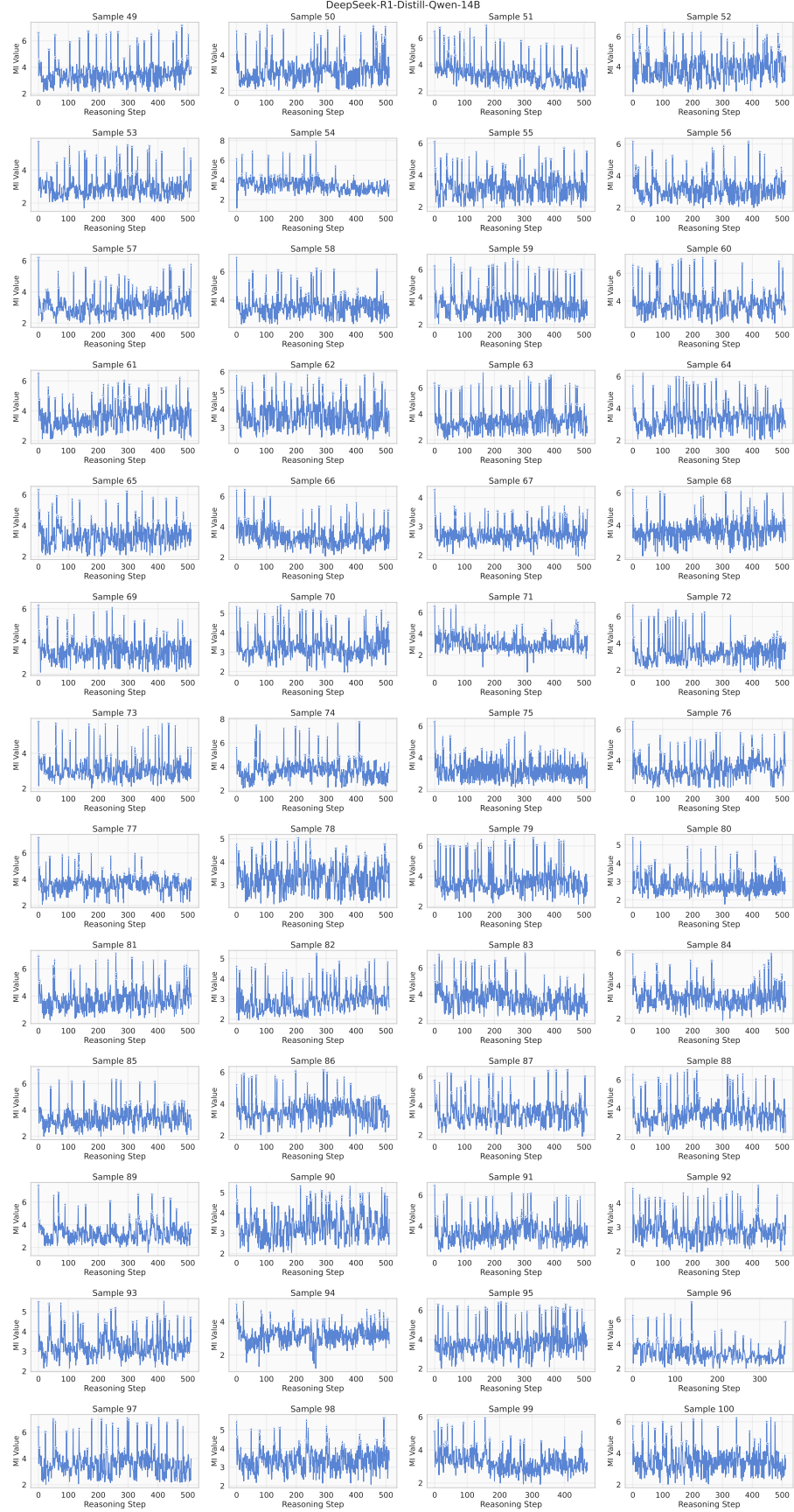


Figure 7: (Continued) MI trajectories of DeepSeek-R1-Distill-Qwen-14B.

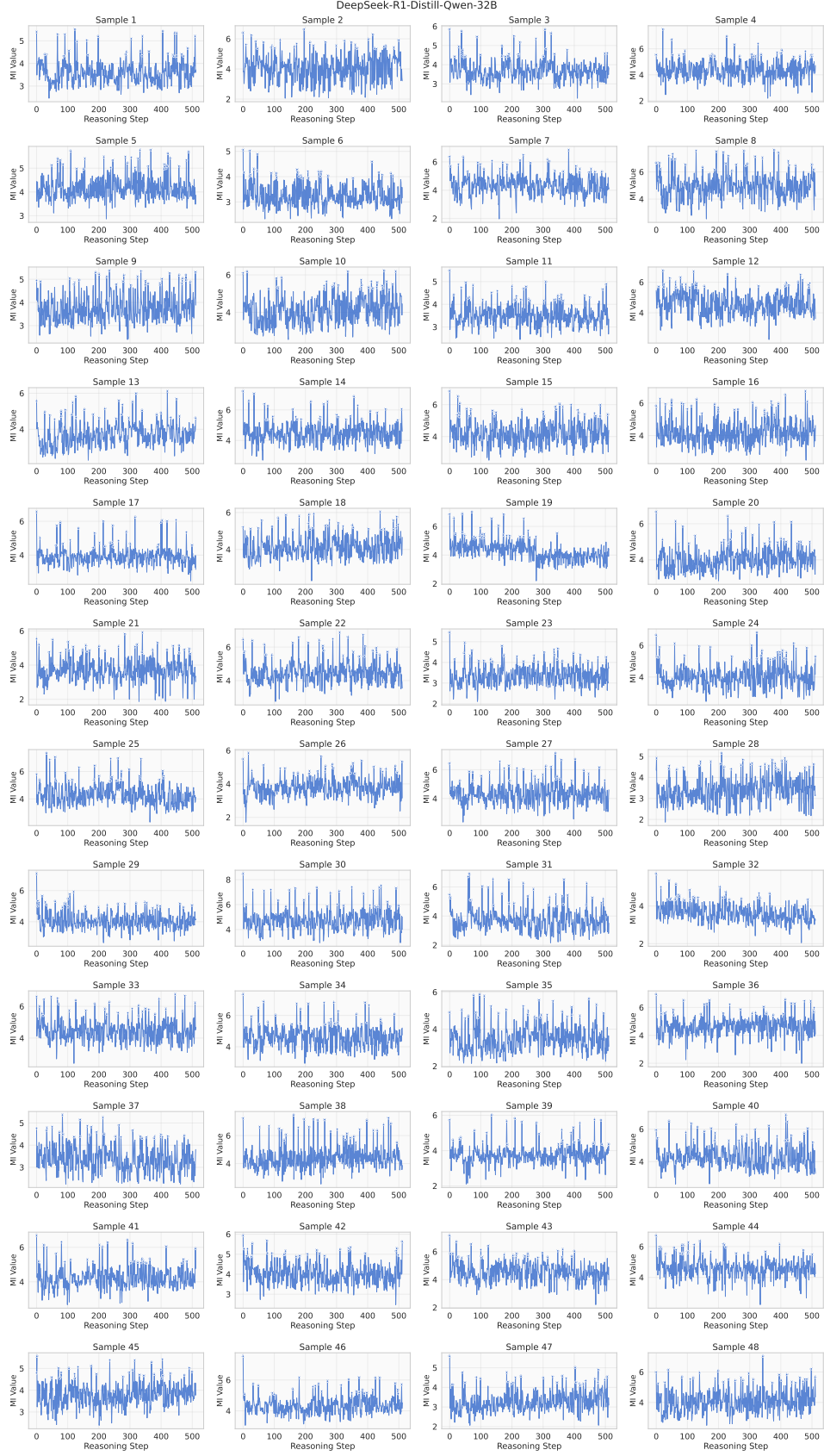


Figure 8: MI trajectories of DeepSeek-R1-Distill-Qwen-32B.

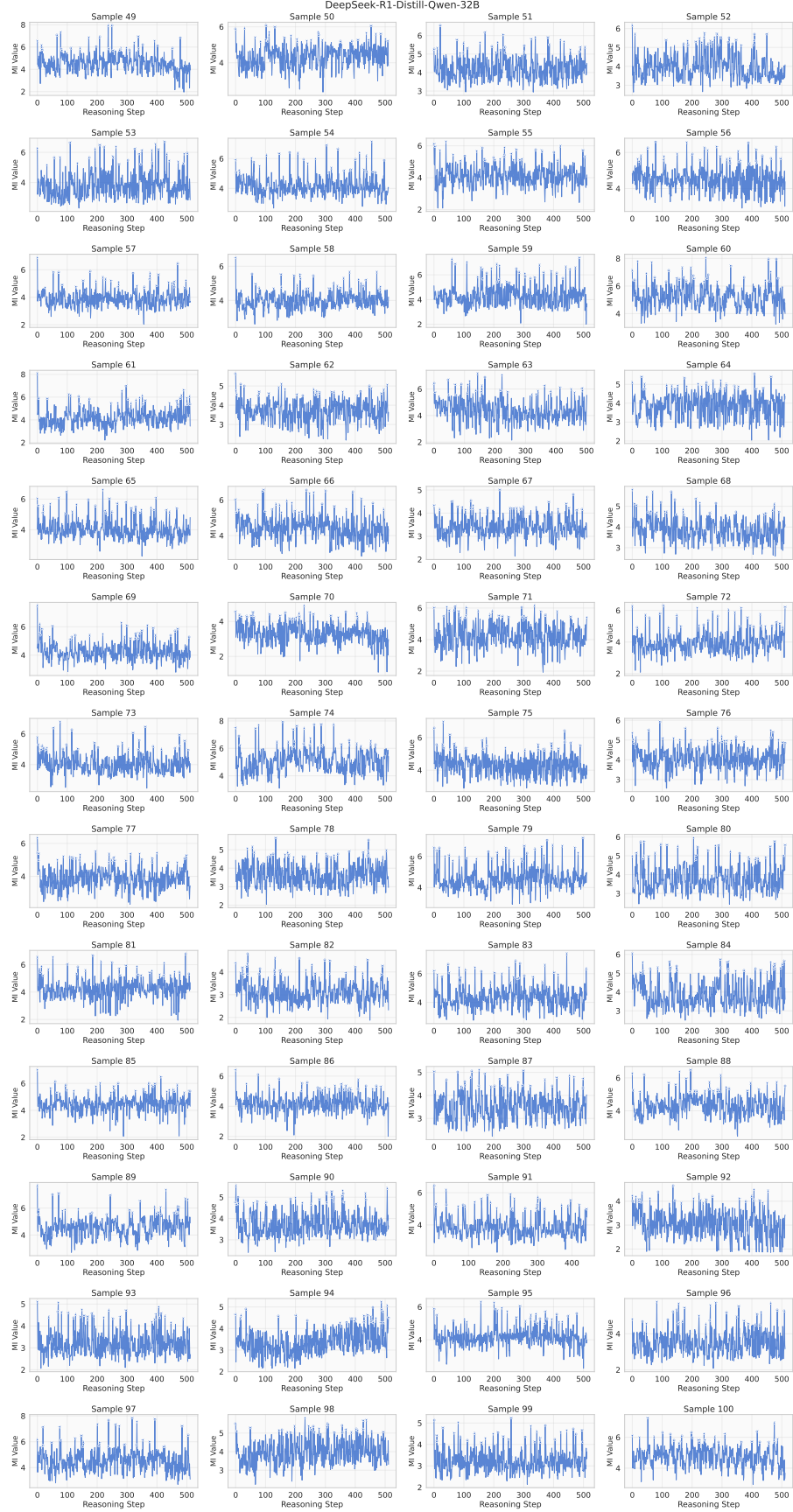


Figure 9: (Continued) MI trajectories of DeepSeek-R1-Distill-Qwen-32B.

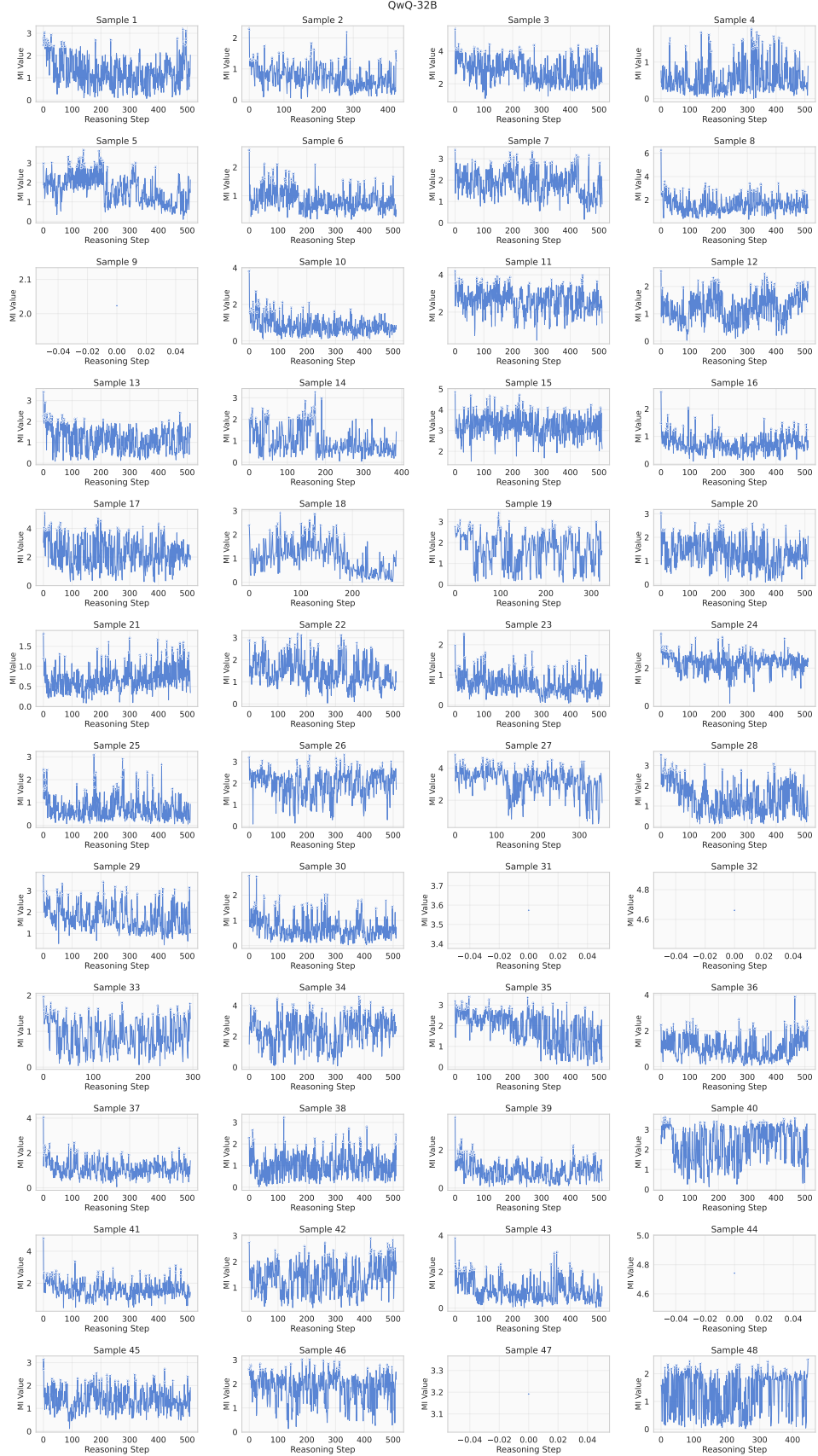


Figure 10: MI trajectories of QwQ-32B.

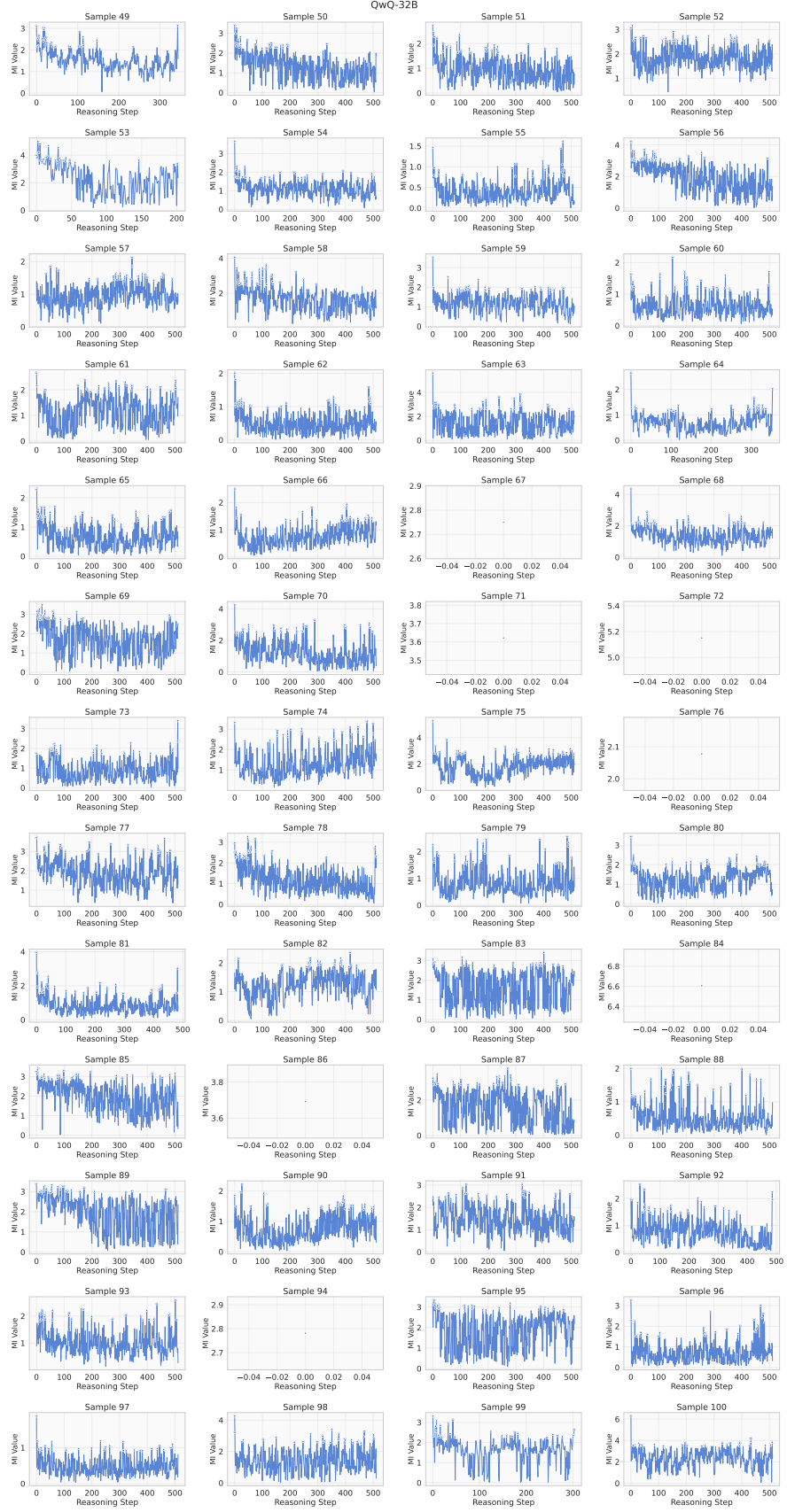


Figure 11: (Continued) MI trajectories of QwQ-32B.

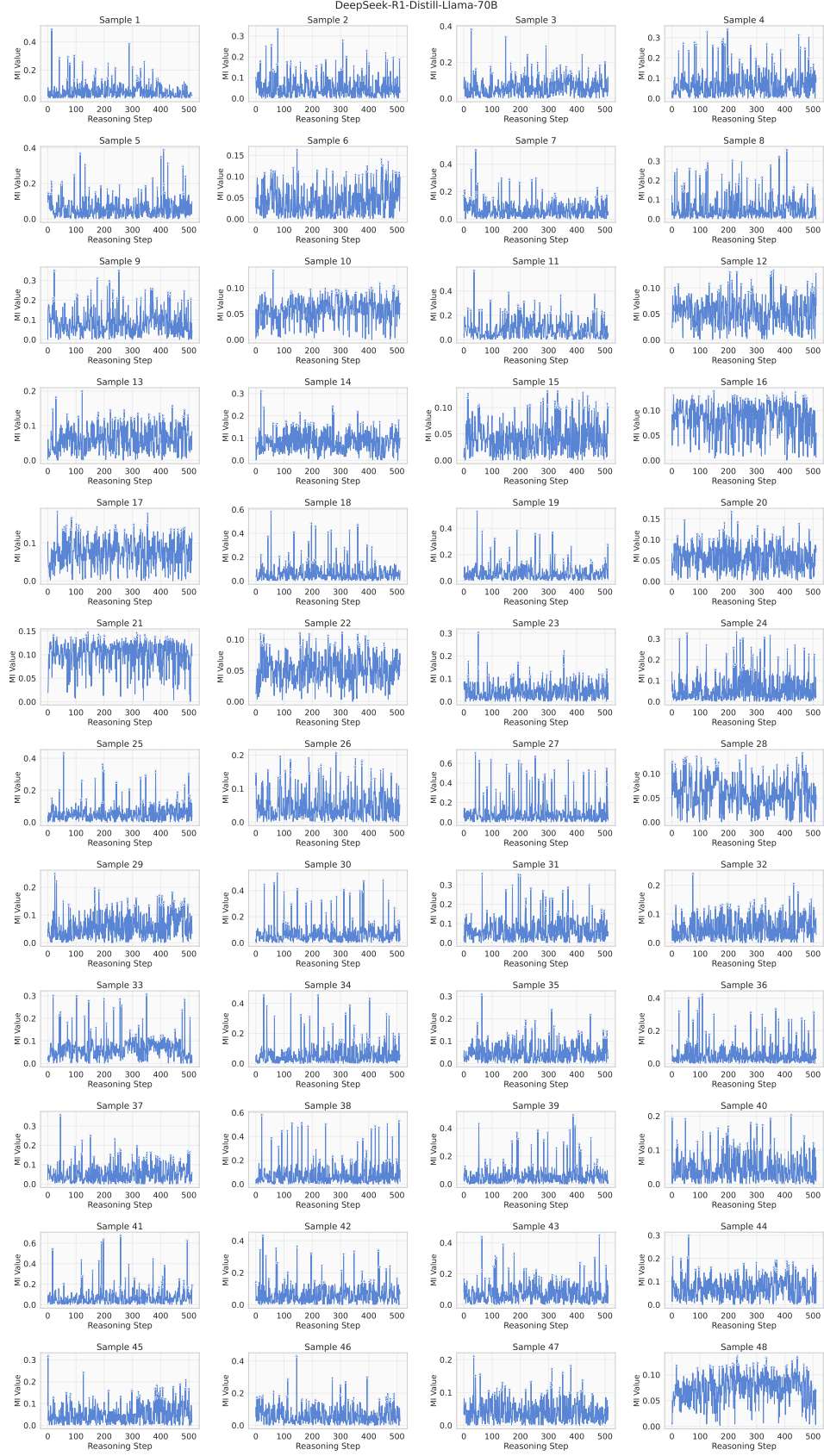


Figure 12: MI trajectories of DeepSeek-R1-Distill-Llama-70B.

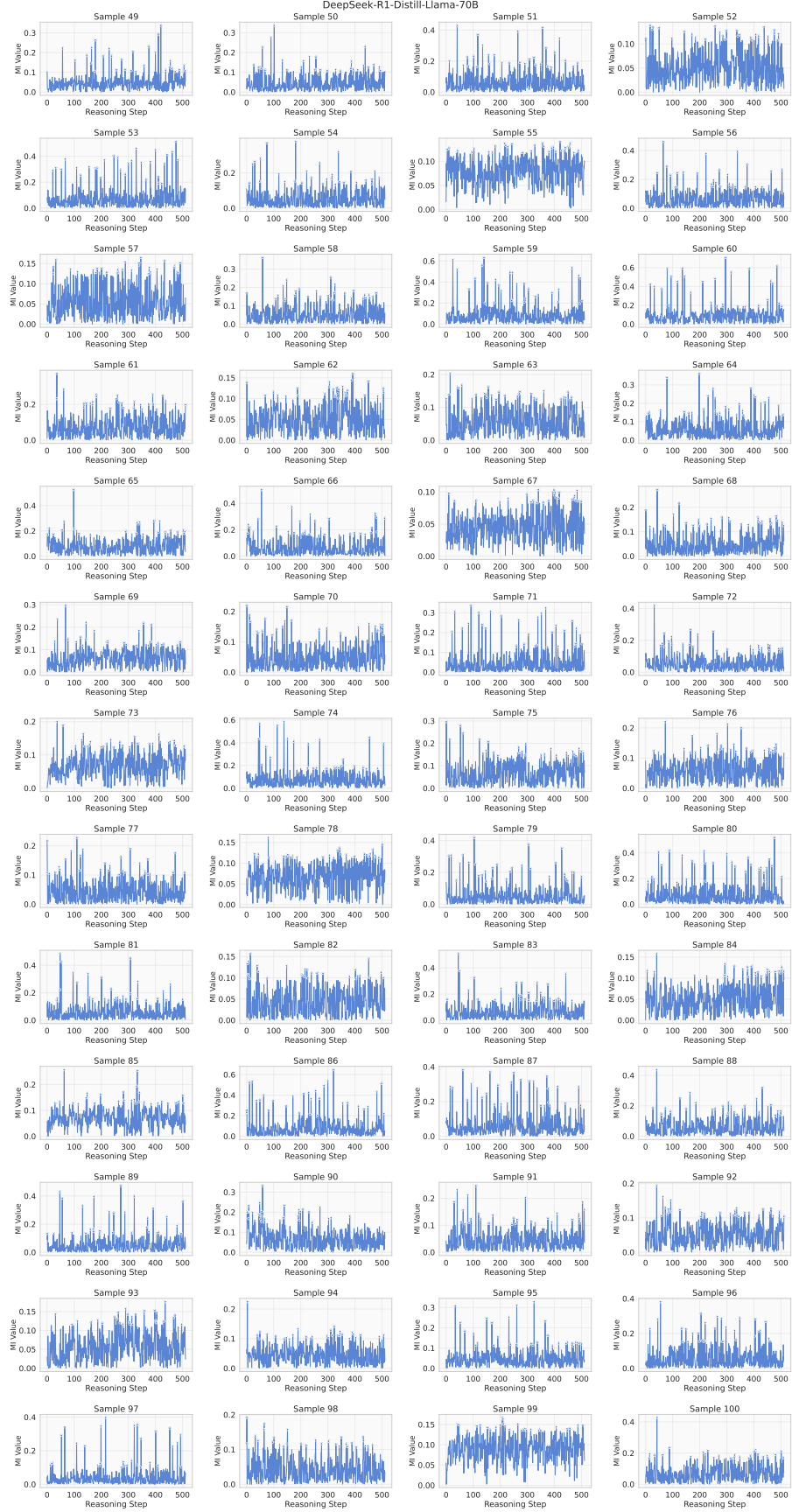


Figure 13: (Continued) MI trajectories of DeepSeek-R1-Distill-Llama-70B.