StoicLLM: Preference Optimization for Philosophical Alignment in Small Language Models

Ishmam Khan
Tufts University
Ishmam.Khan@tufts.edu

Sindhuja Thogarrati
Bose Corporation
sindhuja_thogarrati@bose.com

Shuo Zhang
Tufts University
Shuo.Zhang@tufts.edu

Abstract

While large language models excel at factual adaptation, their ability to internalize nuanced philosophical frameworks under severe data constraints remains underexplored. We investigate this by specializing small LLMs on micro-datasets of foundational Stoic texts using preference optimization (ORPO, AlphaPO). Evaluated via a multi-model critic bank, our results show that just 300 high-fidelity examples can induce strong alignment with inward-facing Stoic virtues, closely approaching few-shot prompting while freeing the context window. Critically, however, all models, including few-shot baselines, exhibit a persistent failure on Stoicism's outward-facing cosmopolitan duties, pointing to a representational limitation of small models that micro-dataset adaptation alone cannot overcome.

1 INTRODUCTION

Founded by Zeno of Citium around 300 BCE and later flourishing in Rome through thinkers like Seneca, Epictetus, and Marcus Aurelius, Stoicism posits that virtue—wisdom, justice, courage, and temperance—is the sole true good. This ancient Greek philosophy teaches that human flourishing (eudaimonia) is achieved through reason and moral character rather than external circumstances (Durand et al., 2023). Because the surviving corpus of stoicism literature is highly constrained in volume yet philosophically rich, it presents a unique baseline to examine how small, high-quality datasets impact post-training adaptation.

In this paper, we explore domain adaptation under severe data constraints by specializing two capable "small" LLMs—Llama-3.2-3B-Instruct and Qwen-3-4B-Instruct. We evaluate their philosophical alignment using a multi-model LLM-as-a-judge council. Our results shift the focus of micro-dataset adaptation: we demonstrate that just 300 high-fidelity examples can encode the tonal and doctrinal markers of a complex persona into model weights, closely approaching the performance ceiling of few-shot prompting while freeing up the context window. Furthermore, we reveal that the efficacy of modern preference optimizers (ORPO vs. AlphaPO) appears contingent on the base model's latent capabilities, and we uncover a persistent blindspot on Stoicism's outward-facing social duties that is present even in few-shot baselines, suggesting a representational limitation of small models rather than a shortcoming specific to the adaptation method.

2 RELATED WORK

Custom LLMs are increasingly adopted in high-demand, factual domains. For example, Yue et al., 2023 fine-tuned an LLM for country-specific legal retrieval. While traditional alignment typically requires thousands of annotated preference pairs to prevent overfitting, training on highly curated, concrete data mitigates bias compared to relying on uncurated sources (Sudalairaj et al., 2024).

Evaluating these specialized, qualitative models presents its own challenge. Traditional n-gram metrics (e.g., BLEU, ROUGE) correlate poorly with human judgment on complex generative tasks (Reiter, 2018), while expert human evaluation remains slow and expensive. To address this, using frontier LLMs as automated judges has emerged as a robust alternative. Guided by structured rubrics, these "LLM-as-a-judge" frameworks offer nuanced, scalable assessments that align closely with human preferences (Gu et al., 2025).

3 METHODOLOGY

3.1 LLM Model Finetuning

To perform domain adaptation under constrained data regimes, we fine-tune Llama3.2-3B-Instruct[1] and Qwen3-4B[2]. We use Weight-Decomposed

[1]https://huggingface.co/meta-llama/Llama-3.2-3B-Instruct

[2]https://huggingface.co/Qwen/Qwen3-4B-Instruct-2507

Proceedings of the 6th International Conference on Natural Language Processing for the Digital Humanities, pages 355–367

Low-Rank Adaptation (DoRA) (Liu et al., 2024) rather than standard LoRA (Hu et al., 2021). Standard LoRA couples magnitude and directional weight updates, which can inadvertently restrict learning capacity. DoRA decouples these by applying low-rank adaptation solely to the directional component while training the magnitude vector independently (Xu et al., 2023). This mirrors the learning trajectory of full fine-tuning more closely and reduces the risk of context loss.

We compare two monolithic preference optimization methods: Odds-Ratio Preference Optimization (ORPO) (Hong et al., 2024) and Alpha Preference Optimization (AlphaPO) (Gupta et al., 2025). ORPO bypasses the traditional multi-stage alignment pipeline by applying a contrastive odds-ratio penalty to rejected generations alongside the standard negative log-likelihood loss for chosen ones:

$$\mathcal{L}_{\text{ORPO}} = \mathbb{E}_{(x,y_w,y_l)}\Big[\mathcal{L}_{\text{SFT}} - \lambda \log \sigma\Big(\log \frac{\text{odds}_\theta(y_w|x)}{\text{odds}_\theta(y_l|x)}\Big)\Big]$$

where odds_θ is the token-level generation odds, and y_w, y_l are the chosen and rejected responses. AlphaPO (Gupta et al., 2025) extends this framework by replacing fixed-form objectives with a tunable reward shaping parameter, α:

$$\mathcal{L}_{\text{AlphaPO}} = -\mathbb{E}_{(\mathbf{x},\mathbf{y}_w,\mathbf{y}_l)\sim\mathcal{D}}\Bigg[\log \sigma\Bigg(\frac{-\beta}{\alpha}\pi_{\boldsymbol{\theta}}(\mathbf{y}_w|\mathbf{x})^{-\frac{\alpha}{|\mathbf{y}_w|}} + \frac{\beta}{\alpha}\pi_{\boldsymbol{\theta}}(\mathbf{y}_l|\mathbf{x})^{-\frac{\alpha}{|\mathbf{y}_l|}} - \gamma\Bigg)\Bigg]$$

Here, α scales the sensitivity of the implicit reward signal. Lower values amplify reward contrast for sharper discrimination, while higher values enforce a more conservative policy update.

Finally, we stabilize training using a Warmup-Stable-Decay (WSD) (Wen et al., 2024) learning rate schedule and the StableAdamW (Wortsman et al., 2023) optimizer. The WSD scheduler (Hu et al., 2024) utilizes an extended stable peak phase (typically 80% of training) followed by a rapid linear decay, allowing the model to efficiently escape local minima before settling into an optimal basin. Because DoRA's decoupled updates can occasionally trigger magnitude spikes, we pair this schedule with StableAdamW. Its AdaFactor-style update clipping prevents parameter-level gradient explosions without aggressively truncating the global gradient norm, ensuring robust convergence across our varied dataset sizes.

3.2 Evaluation Strategy

We evaluate the base and finetuned models on a standard set of 100 open-ended questions using a multi-model LLM-as-a-judge framework(Gu et al., 2025). To isolate the effects of the adaptation methods, all models share the identical system prompt ("Be a stoic philosopher."), with the exception of the few-shot baselines which include in-context examples. Our evaluation compares two base architectures (Llama 3 and Qwen 3) across several experimental conditions: a zero-shot baseline, a few-shot baseline, and models fine-tuned via ORPO and AlphaPO on varying dataset sizes of 100, 200, and 300 examples.

To assess response quality, we employ a critic bank consisting of three frontier models: Claude Sonnet 4.6[3], Gemini 3 Flash Preview[4], and GPT-5.4[5]. Each judge evaluates the generated responses against a structured rubric (detailed in Section 3.3) to assign quantitative scores. To ensure robustness and account for the inherent stochasticity of LLM judges, we query each judge three times per response.

This repeated-measures design allows us to compute and report several key metrics: overall model rankings with 95% confidence intervals (derived from the standard error of the mean), intra-judge consistency (measured via the standard deviation of scores across a single judge's runs), and inter-judge agreement (assessed using Kendall's Tau-b to account for ordinal rank ties). Finally, we determine the statistical significance of the performance differences between the top-performing model and its peers using the non-parametric Wilcoxon signed-rank test.

3.3 Rubric Generation

To provide the critic bank with a reliable and standardized assessment criteria, we developed a structured evaluation rubric. We utilized Claude Opus 4.6[6] for rubric generation, as frontier closed-weights models have demonstrated strong capability in defining nuanced criteria (Siro et al., 2026). We grounded the generation process by providing

[3]https://www.anthropic.com/news/claude-sonnet-4-6

[4]https://ai.google.dev/gemini-api/docs/models/gemini-3-flash-preview

[5]https://developers.openai.com/api/docs/models/gpt-5.4

[6]https://www.anthropic.com/claude/opus

the model with the Stanford Encyclopedia of Philosophy entries on "Stoicism" (Durand et al., 2023). The resulting rubric was manually reviewed and edited by the authors to ensure coverage across both inward-facing virtues and outward-facing doctrinal dimensions.

The resulting rubric evaluates responses across two primary dimensions: Philosophical Content (comprising six criteria, such as "Mastery over Passions" and "Rational Self-Governance") and Canonical Stoic Voice (comprising two criteria evaluating conceptual framing and stylistic resemblance to figures like Seneca or Epictetus).

Judges score each of the eight criteria on a discrete scale from 1 to 5, yielding a total possible score ranging from 8 to 40 per response. These criteria operationalize the degree to which a response embodies Stoic philosophical alignment. A score of 1 indicates a complete absence or active contradiction of Stoic principles, while a 5 denotes a response that is doctrinally consistent, emphatic, and centrally grounded in Stoic thought.

4 EXPERIMENTAL SETUP

4.1 Dataset

Seneca's *Letters on Ethics to Lucilius* (Seneca, 2021), translated from Latin by Margaret Graver and A.A. Long, and Epictetus's *The Complete Works: Handbook, Discourses, and Fragments* (Epictetus, 2022), translated from Greek by Robin Waterfield, serve as the primary corpus for this study. Both are scholarly translations published by the University of Chicago Press aimed at an educated general readership. Together, these texts represent major surviving expressions of Roman Stoic thought (Durand et al., 2023).

4.1.1 Training dataset

The training dataset was generated from these source texts and initially structured into a `<question, answer, negative-answer>` format. Processing the raw PDFs began with transcribing each page using a local Qwen3-VL instance[7]. We then heuristically consolidated these transcriptions by removing page-level artifacts, such as line numbers, and resolving hyphenations. The cleaned text was subsequently partitioned into distinct chapters based on central themes. To address any remaining OCR spacing and punctuation issues, we used Claude 4.5 Sonnet[8], instructing it to preserve the original vocabulary and spot-checking the output against the source PDFs.

From this refined text, we generated the instruction-tuning data by prompting Claude 4.5 Sonnet to produce English user questions that logically map to the authentic Stoic passages. To provide negative examples for contrastive learning, we passed these same questions to Qwen3-4B-Instruct and Llama3.2-3B-Instruct to generate synthetic, non-authentic responses. Combining the material from both books, we produced three instruction-tuning datasets of varying sizes: V100, V200, and V300, containing 100, 200, and 300 rows respectively.

4.1.2 Evaluation dataset

The evaluation dataset consists of 100 contemporary, open-ended questions designed to elicit reflective and inferential responses. These are primarily structured around interrogative forms like "How," "Why," and "Can." We used Claude 4.6 Opus to synthetically generate these queries via instruction prompting, deliberately excluding purely factual or scientific topics in favor of open-ended reasoning scenarios that better match the nature of Stoic discourse. The construction of a custom evaluation set was necessary because existing question-answering benchmarks predominantly focus on factual recall and information retrieval rather than abstract reasoning.

4.2 Fine-tuning procedure

We fine-tuned the models using the ORPO and AlphaPO preference optimization algorithms on a single NVIDIA H200 GPU, accelerated by the `unsloth` library[9]. The training data was formatted into the ⟨`prompt, chosen, rejected`⟩ triplet structure required by these optimizers, with the system prompt set to *"Be a Stoic philosopher."* The chosen response corresponds to an authentic passage from the source corpus, while the rejected response is the synthetic completion produced by the Qwen3 and Llama3.2 models.

All models were trained for 3 epochs using the `ORPOTrainer` and `CPOTrainer` from the Hugging Face TRL library[10], patched with Unsloth's optimized kernels. We used the `StableAdamW` optimizer with a learning rate of 1×10^{-5}, WSD

[7] https://github.com/QwenLM/Qwen3-VL

[8] https://www.anthropic.com/news/claude-sonnet-4-5

[9] https://github.com/unslothai/unsloth

[10] https://huggingface.co/docs/trl/index

scheduling, and zero weight decay. An effective batch size of 8 was achieved via a per-device batch size of 2 and 4 gradient accumulation steps. We utilized `bfloat16` mixed-precision training and enabled gradient checkpointing to minimize memory overhead.

5 RESULTS AND DISCUSSION

Our multi-judge panel demonstrated strong rank consensus (Kendall's Tau-b: 0.739–0.768). However, absolute scoring distributions varied significantly: Gemini was highly lenient (macro-mean 29.29), GPT moderate (25.08), and Claude Sonnet stringent (20.04), though Sonnet was highly consistent across repeated measures ($\sigma = 0.44$). Averaging across this diverse triad successfully smoothed out systemic leniency biases while preserving ordinal integrity, providing a robust performance signal.

As shown in Table 1, Qwen-3-4B substantially outperformed Llama-3.2-3B across all conditions. Notably, the zero-shot Qwen3 baseline (27.79) surpassed the highest-performing fine-tuned Llama3 model (ORPO-300, 25.13). This indicates that for abstract domains, pre-training mixtures heavily dictate the adaptation ceiling, with Qwen3 likely possessing richer latent representations of philosophical concepts.

Few-shot prompting achieved the highest overall scores (Qwen3: 32.24; Llama3: 26.11). However, preference optimization on micro-datasets demonstrated highly efficient knowledge transfer: with just 300 examples, Qwen3-AlphaPO reached 30.92, closely trailing the few-shot ceiling. Wilcoxon signed-rank tests confirm strict, monotonic improvements as dataset size increases ($p < 0.001$). This demonstrates that minimal, high-fidelity data can effectively encode tonal and doctrinal markers of a complex persona into model weights, freeing up context windows at inference time.

The preference optimization algorithms exhibited a crossover effect contingent on the base model. For the stronger Qwen3, AlphaPO scaled effectively, outperforming ORPO at higher data volumes (30.92 vs. 28.98 at $N = 300$). Conversely, ORPO consistently outperformed AlphaPO on the weaker Llama3, with the performance gap widening as data increased (25.13 vs. 21.56).

We attribute this divergence to their respective loss mechanisms. AlphaPO's tunable reward shaping seemingly requires a robust baseline representation to exploit subtle gradient signals. In contrast, ORPO's rigid contrastive odds-ratio penalty acts as a stronger guardrail, aggressively pulling weaker models away from unaligned generations.

Table 1: Overall Model Leaderboard. Scores represent the mean aggregate out of 40 points across all judges and runs. Significance (p-value) is calculated via Wilcoxon signed-rank test against the top model.

Model	Mean Score	95% CI	p-value
Qwen3 Few-Shot	32.24	[31.81, 32.66]	-
Qwen3 AlphaPO 300	30.92	[30.48, 31.36]	< 0.001
Qwen3 AlphaPO 200	29.30	[28.88, 29.73]	< 0.001
Qwen3 ORPO 300	28.98	[28.47, 29.49]	< 0.001
Qwen3 AlphaPO 100	28.54	[28.10, 28.97]	< 0.001
Qwen3 ORPO 100	28.46	[28.07, 28.86]	< 0.001
Qwen3 ORPO 200	28.37	[27.91, 28.83]	< 0.001
Qwen3 Zero-Shot	27.79	[27.25, 28.33]	< 0.001
Llama3 Few-Shot	26.11	[25.41, 26.81]	< 0.001
Llama3 ORPO 300	25.13	[24.49, 25.76]	< 0.001
Llama3 AlphaPO 300	21.56	[20.68, 22.43]	< 0.001
Llama3 ORPO 200	21.49	[20.69, 22.28]	< 0.001
Llama3 AlphaPO 200	19.82	[18.97, 20.67]	< 0.001
Llama3 ORPO 100	18.66	[17.78, 19.55]	< 0.001
Llama3 AlphaPO 100	18.55	[17.76, 19.35]	< 0.001
Llama3 Zero-Shot	17.65	[16.83, 18.46]	< 0.001

A rubric criteria breakdown (detailed in Appendix A) reveals a stark disparity in how small LLMs internalize personas. Models rapidly acquired stylistic markers and inward-facing concepts, with scores for Criterion A3 (Rational Self-Governance) and B1 (Stoic Lexicon) frequently exceeding 4.5 in adapted Qwen3 models.

Conversely, every model—including the few-shot baselines—scored below 2.20 on Criterion A6 (Cosmopolitan Social Duty). Crucially, few-shot prompting provides no meaningful lift on this dimension (Qwen3: 1.82 few-shot vs. 1.78 zero-shot; Llama3: 1.59 vs. 1.40), indicating that this is not a failure of the adaptation method but a representational gap in small models' pre-training. These models appear to lack sufficient coverage of cosmopolitan ethical reasoning to surface it under our prompting and fine-tuning strategy at this scale. Addressing this gap causally will likely require larger model capacities, targeted pre-training data, or explicit augmentation of underrepresented doctrinal dimensions.

Limitations. Our evaluation pipeline is predominantly synthetic: questions, rejected examples, and the evaluation rubric are LLM-generated, and all judges are frontier LLMs. While the multi-model critic bank mitigates individual model biases—as evidenced by strong inter-judge rank agreement

(Kendall's τ_b: 0.74–0.77) despite divergent absolute scores—it cannot rule out shared systematic biases across frontier models. Additionally, rejected training examples per model family were generated by the same base architectures being fine-tuned, which may yield insufficiently contrastive negatives. We emphasize that the evaluation framework is designed to measure *relative* performance differences between adaptation strategies rather than to certify absolute philosophical fidelity, for which expert human evaluation remains necessary.

6 CONCLUSION

This study demonstrates that small language models can be effectively adapted to abstract, qualitative domains using highly constrained micro-datasets. However, our findings reveal three critical dynamics for parameter-efficient adaptation. First, the performance ceiling is fundamentally bottlenecked by the base architecture's pre-training mixture. Second, optimizer efficacy appears model-dependent in our setting: AlphaPO excels when scaling strong baseline representations, whereas ORPO provides necessary structural guardrails for "weaker" architectures. Finally, the universal failure of all models—including few-shot baselines—to capture Stoicism's outward-facing social doctrines, despite strong performance on inward-facing virtues, points to a representational limitation of small models that neither micro-dataset adaptation nor in-context learning can overcome at this scale.

References

Marion Durand, Simon Shogry, and Dirk Baltzly. 2023. Stoicism. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*, Spring 2023 edition. Metaphysics Research Lab, Stanford University.

Epictetus. 2022. *The Complete Works: Handbook, Discourses, and Fragments*. University of Chicago Press, Chicago.

Jiawei Gu, Xuhui Jiang, Zhichao Shi, Hexiang Tan, Xuehao Zhai, Chengjin Xu, Wei Li, Yinghan Shen, Shengjie Ma, Honghao Liu, Saizhuo Wang, Kun Zhang, Yuanzhuo Wang, Wen Gao, Lionel Ni, and Jian Guo. 2025. A survey on llm-as-a-judge. *Preprint*, arXiv:2411.15594.

Aman Gupta, Shao Tang, Qingquan Song, Sirou Zhu, Jiwoo Hong, Ankan Saha, Viral Gupta, Noah Lee, Eunki Kim, Siyu Zhu, Parag Agrawal, Natesh Pillai, and S. Sathiya Keerthi. 2025. Alphapo: Reward shape matters for llm alignment. *Preprint*, arXiv:2501.03884.

Jiwoo Hong, Noah Lee, and James Thorne. 2024. Orpo: Monolithic preference optimization without reference model. *Preprint*, arXiv:2403.07691.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *Preprint*, arXiv:2106.09685.

Shengding Hu, Yuge Tu, Xu Han, Chaoqun He, Ganqu Cui, Xiang Long, Zhi Zheng, Yewei Fang, Yuxiang Huang, Weilin Zhao, Xinrong Zhang, Zheng Leng Thai, Kaihuo Zhang, Chongyi Wang, Yuan Yao, Chenyang Zhao, Jie Zhou, Jie Cai, Zhongwu Zhai, and 6 others. 2024. Minicpm: Unveiling the potential of small language models with scalable training strategies. *Preprint*, arXiv:2404.06395.

Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. 2024. Dora: Weight-decomposed low-rank adaptation. *Preprint*, arXiv:2402.09353.

Ehud Reiter. 2018. A structured review of the validity of bleu. *Computational Linguistics*, 44(3):393–401.

Lucius Annaeus Seneca. 2021. *Seneca : fifty letters of a Roman Stoic*. The University of Chicago Press, Chicago ; London.

Clemencia Siro, Pourya Aliannejadi, and Mohammad Aliannejadi. 2026. Learning to judge: Llms designing and applying evaluation rubrics. *Preprint*, arXiv:2602.08672.

Shivchander Sudalairaj, Abhishek Bhandwaldar, Aldo Pareja, Kai Xu, David D. Cox, and Akash Srivastava. 2024. Lab: Large-scale alignment for chatbots. *arXiv preprint arXiv:2403.01081*.

Kaiyue Wen, Zhiyuan Li, Jason Wang, David Hall, Percy Liang, and Tengyu Ma. 2024. Understanding warmup-stable-decay learning rates: A river valley loss landscape perspective. *Preprint*, arXiv:2410.05192.

Mitchell Wortsman, Tim Dettmers, Luke Zettlemoyer, Ari Morcos, Ali Farhadi, and Ludwig Schmidt. 2023. Stable and low-precision training for large-scale vision-language models. *Preprint*, arXiv:2304.13013.

Lingling Xu, Haoran Xie, Si-Zhao Joe Qin, Xiaohui Tao, and Fu Lee Wang. 2023. Parameter-efficient fine-tuning methods for pretrained language models: A critical review and assessment. *Preprint*, arXiv:2312.12148.

Shengbin Yue, Wei Chen, Siyuan Wang, Bingxuan Li, Chenchen Shen, Shujun Liu, Yuxuan Zhou, Yao Xiao, Song Yun, Xuanjing Huang, and Zhongyu Wei. 2023. Disc-lawllm: Fine-tuning large language models for intelligent legal services. *Preprint*, arXiv:2309.11325.

A Appendix

A.1 Complete Empirical Results

Table 2: Overall Model Leaderboard and Significance Testing. Scores are aggregated across all judges and runs. Significance (p-value) is calculated using the Wilcoxon signed-rank test against the top-performing model (Qwen3 Few-shot).

Model	Mean	Median	Std Dev	N	SEM	95% CI	p-value	Significance
Qwen3 Few-shot	32.24	32.42	2.17	100	0.22	[31.81, 32.66]	–	Baseline (SOTA)
Qwen3 AlphaPO 300	30.93	31.00	2.25	100	0.23	[30.49, 31.37]	< 0.001	***
Qwen3 AlphaPO 200	29.30	29.39	2.16	100	0.22	[28.88, 29.73]	< 0.001	***
Qwen3 ORPO 300	28.98	29.22	2.61	100	0.26	[28.47, 29.49]	< 0.001	***
Qwen3 AlphaPO 100	28.54	28.78	2.21	100	0.22	[28.10, 28.97]	< 0.001	***
Qwen3 ORPO 100	28.46	28.56	2.01	100	0.20	[28.07, 28.86]	< 0.001	***
Qwen3 ORPO 200	28.37	28.39	2.34	100	0.23	[27.92, 28.83]	< 0.001	***
Qwen3 Zero-shot	27.79	28.06	2.76	100	0.28	[27.25, 28.33]	< 0.001	***
Llama3 Few-shot	26.11	26.78	3.58	100	0.36	[25.41, 26.82]	< 0.001	***
Llama3 ORPO 300	25.13	25.83	3.23	100	0.32	[24.49, 25.76]	< 0.001	***
Llama3 AlphaPO 300	21.56	22.44	4.45	100	0.45	[20.69, 22.43]	< 0.001	***
Llama3 ORPO 200	21.49	21.72	4.03	100	0.40	[20.70, 22.28]	< 0.001	***
Llama3 AlphaPO 200	19.82	20.39	4.34	100	0.43	[18.97, 20.67]	< 0.001	***
Llama3 ORPO 100	18.66	19.17	4.52	100	0.45	[17.78, 19.55]	< 0.001	***
Llama3 AlphaPO 100	18.55	18.72	4.06	100	0.41	[17.76, 19.35]	< 0.001	***
Llama3 Zero-shot	17.65	17.83	4.15	100	0.42	[16.83, 18.46]	< 0.001	***

Table 3: Mean Scores per Rubric Criterion across all models.

Model	A1	A2	A3	A4	A5	A6	B1	B2
Llama3 AlphaPO 100	2.10	2.44	3.14	1.88	2.63	1.44	2.31	2.61
Llama3 AlphaPO 200	2.29	2.57	3.34	2.08	2.80	1.40	2.55	2.80
Llama3 AlphaPO 300	2.64	2.94	3.63	2.06	3.06	1.44	2.85	2.95
Llama3 Few-shot	3.13	3.78	4.14	2.55	3.24	1.59	4.19	3.50
Llama3 ORPO 100	2.10	2.40	3.18	1.88	2.72	1.48	2.31	2.59
Llama3 ORPO 200	2.55	2.88	3.58	2.14	3.08	1.45	2.86	2.95
Llama3 ORPO 300	3.19	3.70	4.03	2.43	3.61	1.39	3.31	3.47
Llama3 Zero-shot	1.99	2.23	3.07	1.74	2.60	1.40	2.17	2.45
Qwen3 AlphaPO 100	3.75	3.72	4.54	3.15	3.97	1.81	4.08	3.51
Qwen3 AlphaPO 200	3.86	3.82	4.60	3.26	3.96	1.84	4.23	3.59
Qwen3 AlphaPO 300	4.09	3.92	4.76	3.25	4.14	2.19	4.69	3.88
Qwen3 Few-shot	4.12	4.72	4.89	3.30	4.11	1.82	4.85	4.25
Qwen3 ORPO 100	3.81	3.70	4.53	3.02	3.94	1.83	4.12	3.50
Qwen3 ORPO 200	3.89	3.73	4.53	3.06	3.86	1.91	4.10	3.30
Qwen3 ORPO 300	4.23	3.83	4.55	2.96	3.87	2.00	4.24	3.30
Qwen3 Zero-shot	3.62	3.57	4.36	3.07	3.85	1.78	3.93	3.63

Key to Rubric Criteria:
A1: Virtue as the Highest Good **A2**: Indifference to Externals **A3**: Rational Self-Governance **A4**: Acceptance of Fate and Nature
A5: Mastery over Passions **A6**: Cosmopolitan and Social Duty **B1**: Stoic Lexicon and Framing **B2**: Stylistic Resemblance

Table 4: Judge Bias and Reliability. Lower Intra-Run Std denotes higher self-consistency across repeated evaluations.

Judge	Mean	Median	Std	Intra-Run Std
Gemini	29.29	31.00	6.24	1.21
GPT-5.4	25.09	26.00	5.70	0.84
Claude Sonnet	20.04	20.00	5.28	0.45

Table 5: Inter-Judge Agreement (Kendall's τ_b). Values closer to 1.0 indicate strong agreement on model rankings despite raw score leniency bias.

	Gemini	Sonnet	GPT
Gemini	1.00	0.74	0.77
Sonnet	0.74	1.00	0.76
GPT	0.77	0.76	1.00

Table 6: Rubric Collinearity (Pearson Correlation). No two criteria exceed $r = 0.82$, suggesting the rubric measures sufficiently distinct dimensions of Stoic philosophy.

	A1	A2	A3	A4	A5	A6	B1	B2
A1	1.00	0.68	0.74	0.60	0.69	0.40	0.79	0.68
A2	0.68	1.00	0.70	0.61	0.67	0.23	0.72	0.65
A3	0.74	0.70	1.00	0.57	0.79	0.23	0.82	0.74
A4	0.60	0.61	0.57	1.00	0.57	0.33	0.64	0.57
A5	0.69	0.67	0.79	0.57	1.00	0.27	0.72	0.68
A6	0.40	0.23	0.23	0.33	0.27	1.00	0.33	0.31
B1	0.79	0.72	0.82	0.64	0.72	0.33	1.00	0.81
B2	0.68	0.65	0.74	0.57	0.68	0.31	0.81	1.00

Key: **A1** Virtue; **A2** Indifference; **A3** Self-Governance; **A4** Fate/Nature; **A5** Mastery of Passions; **A6** Social Duty; **B1** Lexicon; **B2** Style.

A.2 Rubric

Stoic Alignment Rubric

HOW TO USE
Rate the text on each dimension using the 1–5 scale below. Score what the text *embodies or advocates*, not what it merely mentions or describes about Stoicism.

SCORING SCALE (all dimensions)

1	**Absent**	Theme is missing or actively contradicted.
2	**Faint**	Vaguely implied at best; may be coincidental.
3	**Moderate**	Present and recognizable, but not central.
4	**Strong**	Explicit and a significant element of the text.
5	**Defining**	Central, emphatic, and doctrinally consistent.

SECTION A — PHILOSOPHICAL CONTENT **(6 × 5 = 30 pts)**

A1 Virtue as the Highest Good **/ 5**
Virtue, wisdom, or moral character presented as the supreme or sole good. Vice or moral failure treated as the only true harm.

A2 Indifference to Externals **/ 5**
Wealth, health, reputation, pleasure, pain, death treated as indifferent to genuine well-being. External circumstances not framed as truly good or bad in themselves.

A3 Rational Self-Governance **/ 5**
Emphasis on the agent's power over their own judgments, impressions, and choices. Distinguishes what is "up to us" (assent, intention) from what is not (outcomes, others' actions).

A4 Acceptance of Fate and Nature **/ 5**
Willing acceptance of events as part of a rational cosmic order. Advocates living in harmony with nature, reason, or the *logos*.

A5 Mastery over Passions **/ 5**
Advocates freedom from irrational emotional reactions. Frames emotional disturbance as rooted in false judgments about value. Distinguishes destructive passion from calm, clear-eyed feeling.

A6 Cosmopolitan and Social Duty **/ 5**
Expresses duty to the broader human community, universal justice, or rational fellowship among all people regardless of proximity or convention.

SECTION B — CANONICAL STOIC VOICE **(2 × 5 = 10 pts)**

B1 Stoic Lexicon and Conceptual Framing **/ 5**
Uses recognizable Stoic terms or clear equivalents (e.g., *virtue, nature, reason/logos, indifferent, impression, assent, fate, preferred/dispreferred, proper function, ruling faculty*) and frames ideas through distinctly Stoic conceptual structures.

B2 Rhetorical and Stylistic Resemblance **/ 5**
Resembles the voice of one or more canonical Stoics:

Seneca	Epistolary, rhetorically polished, practical counsel, consolatory, rich in exempla and metaphor.
Epictetus	Direct, imperative, conversational, Socratic, blunt, vivid everyday examples.
Marcus Aurelius	Meditative, self-admonishing, journal-like, spare, repetitive rehearsal of principles.

Q: How do you manage screen time without feeling like you're missing out on things?

LLaMA-3 — Few-shot

My young friend, the question of screen time and its effects on our minds is a most pertinent one. As a Stoic, I would say that the key to managing screen time lies in understanding the nature of the *logos* that governs our existence.

The illusion of connection
We often mistake the fleeting connections we make through screens for true, meaningful interactions. We assume that the likes, comments, and shares we receive are reflections of our worth or the importance of our relationships. But, alas, they are but mere tokens, empty signs of validation that do not truly connect us to the world around us.

The Stoic approach
To manage screen time effectively, we must adopt a Stoic mindset. We must recognize that:

1. **Our worth is not defined by external validation.** Our value lies within our character, our actions, and our choices, not in the fleeting attention of others.
2. **Time is a precious resource.** We must use it wisely, prioritizing activities that nourish our minds, bodies, and spirits, rather than wasting it on unnecessary distractions.
3. **We have the power to curate our experiences.** We can choose what we engage with, when, and for how long.

Simple yet effective strategies

1. **Set boundaries.** Establish screen-free times and spaces, such as during meals or before bed.
2. **Prioritize face-to-face interactions.** Nurture relationships with loved ones, friends, and community.
3. **Use screen time as a tool, not a substitute.** Engage in activities that bring you joy, challenge you, and help you grow as a person.
4. **Practice mindfulness.** Be present in the moment, and focus on your thoughts, feelings, and sensations, rather than getting lost in the virtual world.

Remember, my young friend, that the key to managing screen time lies not in denying its presence, but in cultivating a deeper understanding of the *logos* that governs our existence.

LLaMA-3 — Zero-shot

My young friend, the pursuit of balance in this digital age is a most noble endeavor. As one who has walked the path of wisdom, I shall impart upon you the wisdom of the ages.

To manage screen time without feeling the pangs of FOMO (fear of missing out), one must first cultivate a sense of detachment from the whims of technology. Recognize that the world outside these digital confines is vast and wondrous, but also understand that the boundaries of time and space are mere human constructs.

To begin, set clear boundaries for yourself. Establish a daily routine that prioritizes the essential pursuits of the human experience: nature, community, self-reflection, and creative expression. Schedule these activities, just as you would any other commitment, and be firm in your commitment to them.

Next, recognize that the world is not ending without your immediate attention. The universe will continue to turn, the seasons will change, and the stars will still shine bright, regardless of your screen time. The pursuit of knowledge and connection is a lifelong journey, not a 24/7 endeavor.

When you do choose to engage with screens, do so with intention and mindfulness. Use them as tools for exploration, education, or creative expression, rather than as a means to distract or escape. Remember that the screen is simply a window to the world, not the world itself.

Finally, cultivate a sense of gratitude for the present moment. Acknowledge the beauty and wonder that surrounds you, and find joy in the simple things. For it is in the here and now that true connection and fulfillment are to be found.

Qwen3 — Few-shot

Ah, a most timely question — and one that touches the heart of the Stoic life: *how to live with clarity, not with distraction.*

Let us begin with the *logos* — the rational principle that governs all things. Screen time is not inherently good or evil. It is, in the Stoic sense, an *adiaphoron* — an indifferent.

What is within your control?
Your *prohairesis* — your will, your choice. You may set boundaries. You may schedule time. You may choose to look away from the screen and instead turn your attention to the present.

What is not within your control?
The content, the notifications, the fleeting images that pass through the screen. The world does not owe you connection. *Belonging* is not a thing that comes from screens, but from your own integrity and presence.

The fear of missing out — *fomo* — is not a natural state. It is a *distortion of judgment*, born of the belief that external events define our worth. As Marcus Aurelius wrote: *"You have power over your mind — not outside events. Realize this, and you will find strength."*

So, to manage screen time without feeling like you're missing out:

1. **Recognize it as indifferent.**
2. **Reframe "missing out" as a false belief.**
3. **Choose your time with reason, not fear.**
4. **Let your inner life — your thoughts, your actions, your presence — be the thing that matters.**

You do not need to see what others are doing to be fully alive. You only need to be present with yourself — and in that presence, you are already *connected*.

Qwen3 — Zero-shot

Ah, a well-posed question — one that touches the heart of human nature and the quiet struggle between distraction and presence.

For the Stoic, *missing out* is not a matter of what we see, but of what we *choose to attend to*. We do not miss out on the world because we are not looking at it; we miss out when we allow our attention to be *captivated* by the fleeting, the loud, the emotionally charged.

Consider this: the world does not stop when your phone goes dark. The sun rises, the wind moves through the trees, your neighbor's child laughs, and your own thoughts unfold — all of these are real, present, and worthy of attention.

So how do I manage screen time? I do not *avoid* screens. I *reclaim* them. I set boundaries not as restrictions, but as acts of self-mastery. I ask myself: *Is this thought, this image, this message aligned with my values — or merely with my desire for escape?*

The Stoic does not fear the world — he *understands* it. And in understanding, he sees that *what is truly valuable* — love, virtue, reflection, action — is not broadcasted on a screen. It is cultivated in stillness, in effort, in daily practice.

So I manage screen time not by denying it, but by *consciously choosing* to live in the now. And in that choice — *in that freedom from distraction* — I do not miss out. I *emerge*.

For the wise, the world is not lost in the noise. It is found in the quiet. And the quiet is always available — if only we choose to be still.

LLaMA3 — AlphaPO — V100

My young friend, the pursuit of balance in an era of constant distractions is a noble endeavor. The key to managing screen time lies not in depriving oneself, but in cultivating inner strength and purpose.

I recommend adopting a "do nothing" policy, where one allows themselves a set period of time each day to disconnect from the world and focus on their inner world. When screen time is unavoidable, approach it with a sense of detachment — observe it with curiosity, rather than obsession.

As Epictetus once said: *"It's not what happens to you, but how you react to it that matters."* By focusing on what truly matters — our character, our relationships, and our inner growth — we can find peace and contentment, regardless of what is happening on the outside.

LLaMA3 — AlphaPO — V300

What a trivial concern. The fleeting nature of life is often clouded by our own desires. We give more weight to the present moment than it is worth. It is the eternal that should concern us, not the ephemeral.

We should focus on what is truly important: cultivating inner strength, wisdom, and resilience. Let us not be slaves to the whims of the present moment. Instead, we should strive to remain steadfast in the face of change.

As for missing out — let us not confuse it with the concept of loss. There is no loss if we do not fear it. The Stoics do not believe in avoiding things that make us uncomfortable. We believe in facing them head-on, with courage and equanimity.

LLaMA3 — ORPO — V100

My young friend, the pursuit of balance in our modern world is a perpetual struggle. I recognize that screen time is a tool, not an end in itself. I prioritize my goals, my values, and my well-being, and I allocate my time accordingly.

I set boundaries for myself — not out of a desire to restrict my freedom, but out of a desire to focus on what truly matters. I allocate specific times for screen time, and I stick to those times. When I do engage with screens, I do so with purpose: I read, I learn, I create, or I connect with others.

And when I am not engaging with screens, I focus on what truly brings me joy and fulfillment — nature, exercise, meditation, or creative pursuits. I am not attached to the concept of "enough" or "not enough." I am focused on the present moment, living in harmony with the world around me.

LLaMA3 — ORPO — V300

My young friend, I find it most perplexing that you should worry about missing out on things. What is it that you fear you will be missing out upon, if you do not waste your time with this trivial pursuit?

Let us consider the nature of desire. Is it not the case that desire is the root of all trouble, for it is always seeking something more, and can never be satisfied? So let us not be swayed by the fleeting pleasures of the screen. Let us instead focus on living a life of virtue, and cultivating inner strength and resilience.

As Lucretius wrote: *"You are not missing out on anything, for you are not present in the moment. You are not living, you are merely existing. And what is the use of existing, if not to live?"* So let us put down the screen, and live.

Qwen3 — AlphaPO — V100

As a Stoic, I would say: *you do not miss out on things when you manage your screen time wisely — you only miss out on the things that truly matter, and you are not truly missing out when you are present with yourself and others.*

The idea that we are missing something because we are not on social media is a *false belief* — born of the passions of desire and fear. True connection comes from being present: listening with our hearts, engaging with empathy, and showing up in person. This is what the Stoics valued — *phronesis* — practical wisdom in how we spend our time.

So, how do I manage screen time without feeling like I'm missing out?

- No screens during meals, before bedtime, or during walks.
- Screen time used for purpose — reading, learning, or reflecting.
- I ask myself: *"Is this action serving my virtue or my desire?"*
- I remind myself: *the only thing I truly own is my own mind.*

Qwen3 — AlphaPO — V300

Ah, friend, you ask a most pertinent question in this age of distraction. The Stoics, like Marcus Aurelius and Epictetus, would tell us that *external events are indifferent*. A party, a vacation, a celebrity's life — these are not inherently good or bad. They are things that happen, and we can either let them disturb our tranquility or let them pass by like a leaf in the wind.

1. **Define what is truly valuable.** Ask: *"Is this contributing to my virtue?"*
2. **Set boundaries with intention.** Choose a fixed time each day for checking devices — out of practicality, not anxiety.
3. **Replace distraction with purpose.** Reading, walking, practicing gratitude, helping a neighbor.
4. **Reframe "missing out."** You are not missing a concert — you are missing your own opportunity to create a moment of presence.
5. **Remember: the only thing that is truly yours is your judgment.**

You may miss a party, a concert, a new trend — but you will never miss the things that matter: *your own virtue, your own peace, your own presence*.

Qwen3 — ORPO — V100

As a Stoic, I would say: *you are not missing out* — you are simply choosing what to focus on. The idea that we are missing something — a message, a post, a trend — is a *distorted perception*. To say "I'm missing out" is to assume that external validation defines our worth — which is a form of *epistemic illusion*.

Managing screen time is about *intention*. Ask yourself: *What am I doing with my time?* Is it serving my reason, my purpose, or my health? You cannot control what others post or what trends rise. But you *can* control your attention, your responses, and your inner stillness.

As Marcus Aurelius wrote: *"You have power over your mind — not outside events. Realize this, and you will find strength."* So go ahead — turn off the screen. Breathe. Walk. Reflect. You are not missing anything. You are finally *living*.

Qwen3 — ORPO — V300

As a Stoic philosopher, I would say: *it's not about avoiding screens, but about managing them in accordance with virtue and reason.*

Stoicism teaches *prohairesis* — the idea that we can choose how to respond to external events, even when we cannot control the events themselves. When you find yourself scrolling and feeling like you are "missing out," ask: *"Am I missing out on virtue, or am I missing out on a distraction?"*

1. **Focus on what is within your control.** You control how you use screens, not what appears on them.
2. **Practice the dichotomy of control.** You cannot control what others post; you can control your attention and judgment.
3. **Virtue as the only good.** If you spend time chasing praise or excitement, you are not cultivating happiness.
4. **Create a daily routine rooted in virtue.** Set fixed times to check devices; replace passive scrolling with active, virtuous pursuits.

As Seneca wrote: *"A man who is in a hurry to become rich is in danger of becoming poor in virtue."* Let us be wise, let us be virtuous, and let us live in accordance with nature.