RedOne 2.0: Rethinking Domain-specific LLM Post-Training in Social Networking Services

Fei Zhao, Chonggang Lu, Haofu Qian, Fangcheng Shi, Zijie Meng,
Jianzhao Huang, Zheyong Xie, Shaosheng Cao[*]
NLP Team, Xiaohongshu Inc., China
`caoshaosheng@xiaohongshu.com`

Abstract

As a primary medium for human interaction and information exchange, social networking services (SNS) present distinct challenges for large language models (LLMs): rapidly evolving norms and slang, and culturally diverse content that causes knowledge distribution shift. While supervised fine-tuning (SFT) can improve in-domain performance, it often induces a "seesaw" trade-off with out-of-domain robustness, especially for smaller models. To address these challenges, we present RedOne 2.0, an SNS-oriented LLM developed with a progressive, RL-prioritized post-training paradigm for fast and stable adaptation. Our pipeline has three stages: (1) Exploratory Learning on curated SNS corpora to establish initial alignment and surface systematic weaknesses; (2) Targeted Fine-Tuning that applies SFT only to diagnosed gaps while mixing a small amount of general data to reduce forgetting; and (3) Refinement Learning that re-applies RL with SNS-centric signals to consolidate gains and balance trade-offs across tasks. Across various tasks in three categories, our 4B model improves by 2.41 on average over the prior 7B RedOne baseline. It also yields an 8.74 average gain over its Qwen3-4B base while using less than half the data required by the SFT-centric method, demonstrating superior data efficiency and stability at compact scales. Overall, RedOne 2.0 provides a competitive, cost-effective baseline for SNS-specific LLMs, improving capability without sacrificing robustness.

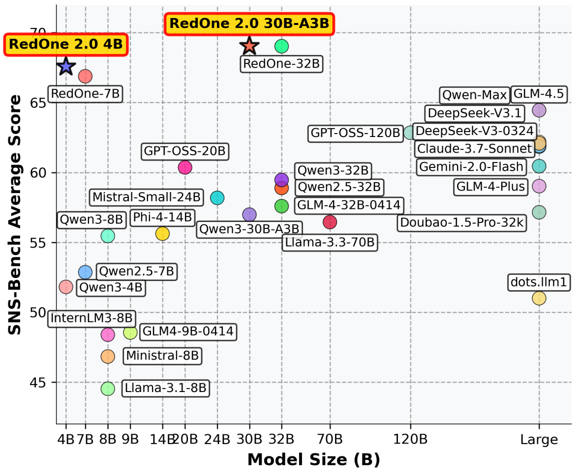

Figure 1: The comparison of various scale models' performance in the SNS domain.

1 Introduction

Social Networking Services (SNS) are vital infrastructure for global interaction (Xia et al., 2013), yet deploying advanced Large Language Models (LLMs) (Hurst et al., 2024; Yang et al., 2025) into this domain is challenged by highly heterogeneous workloads and rapidly evolving community norms (Zhao et al., 2025; Zeng et al., 2024). While RedOne (Zhao et al., 2025) demonstrated that LLM's capabilities can be enhanced through SFT, such pipelines often suffer from a "seesaw" effect where in-distribution gains trigger catastrophic forgetting or poor out-of-distribution generalization (Kotha et al., 2023; Kumar et al., 2022; Ramasesh et al., 2021; Huang et al., 2024). Currently, reinforcement learning (RL) presents a compelling, data-efficient alternative for aligning LLMs with complex objectives while preserving general competencies (Jin et al., 2025; Ouyang et al., 2022; Guo et al., 2025a; Yu et al., 2025). However, structuring an RL-centric recipe that ensures stability and rapid adaptation of LLM within the SNS scenario remains underexplored.

To fill this gap, we introduce RedOne 2.0, an SNS-oriented LLM with a progressive, RL-prioritized post-training paradigm (Fig. 1). The key idea is that exploration–targeted correction–refinement yields a better stability–generalization trade-off than SFT-heavy recipes, particularly at smaller scales and with

---

[*] Corresponding author.

limited domain data. RedOne 2.0 trains in three stages: (1) Exploratory Learning, exposing the model to curated SNS corpora to establish initial alignment and diagnose systematic weaknesses; (2) Targeted Fine-Tuning, applying SFT only to diagnosed gaps while mixing a small fraction of general data to regularize against forgetting; and (3) Refinement Learning, re-applying RL with SNS-centric signals to consolidate gains and smooth trade-offs across tasks.

We conduct extensive experiments across tasks in three categories. The 4B variant surpasses RedOne-7B by 2.41 on average, showing strong SNS performance at compact scales. Starting from Qwen3-4B, RedOne 2.0 uses about half of RedOne's data while achieving an average lift of 8.74, demonstrating superior data efficiency and broader gains from an RL-centric curriculum. Our contribution can be summarized as follows:

- We present RedOne 2.0, an SNS-domain LLM that achieves higher capability with less data and smaller models.
- We propose a progressive, RL-prioritized post-training paradigm that consistently improves both general and SNS-specific abilities while mitigating the "seesaw" effect of SFT.
- We provide comprehensive validation showing state-of-the-art performance and robustness under various scenarios, establishing RedOne 2.0 as a cost-effective baseline for SNS.

2 Methodology for RedOne 2.0

Fig. 2 illustrates our pipeline, which progressively aligns the model to the SNS domain across three stages. Exploratory Learning initially aligns the model and identifies task-specific weaknesses; Targeted Fine-Tuning then rectifies these gaps to improve performance and prepare for further optimization; finally, Refinement Learning applies RL to stabilize behavior and drive better performance across diverse tasks.

2.1 Dataset Definition

We construct a comprehensive training dataset by merging large-scale SNS-specific data ($\mathcal{D}_{\text{SNS}}$) with high-quality general-domain data ($\mathcal{D}_{\text{GEN}}$). The former covers over 75 tasks across six core SNS capabilities (e.g., user behavior modeling, content understanding, dialogue) ensuring specialized supervision. The latter incorporates validated open-source datasets to maintain robust general reasoning while minimizing annotation and preprocessing costs. All samples are normalized into a unified Q&A format: $\mathcal{D} = \{\mathcal{D}_{\text{SNS}} \cup \mathcal{D}_{\text{GEN}} \mid (Q, A)\}$.

2.2 Exploratory Learning

Exploratory Learning aims for initial alignment between the base model and SNS characteristics. By immersing the model in diverse task distributions, this stage captures domain-specific interaction patterns while identifying "hard" tasks that require further reinforcement.

2.2.1 Data construction

We curate 750K SNS entries $\mathcal{D}_{\text{SNS}_1}$ from $\mathcal{D}_{\text{SNS}}$ spanning 75 heterogeneous tasks (e.g., post taxonomy, post view search, domain translation) using a balanced sampling schedule to ensure long-tail visibility. To preserve general competence, we also incorporate a 50K subset $\mathcal{D}_{\text{GEN}_1}$ enriched with rationales from $\mathcal{D}_{\text{GEN}}$.

2.2.2 Reward function

Unlike simple rule-based rewards (Guo et al., 2025a), we employ 4 task-specific reward functions and normalize them to similar ranges (i.e., [0, 1]) to accommodate the high variation in our scenario. We denote the sampled pair as (Q, A) and model's output as O.

1) **Exact Match.** For close-ended tasks (e.g., classification or multiple-choice) with deterministic answers, the reward function is defined with exact match score.

$$\mathcal{R}_{\text{EM}}(O, A) = \begin{cases} 1, & O = A, \\ 0, & \text{otherwise.} \end{cases} \tag{1}$$

2) **Metrics-based.** For open-ended tasks such as translation, we define rewards using task-specific evaluation metrics Eval (e.g., BLEU, F1).

$$\mathcal{R}_{\text{Met}}(O, A) = \text{Eval}(O, A) \tag{2}$$

3) **Sand Box.** For tasks like code generation, we evaluate the generated code by executing on a set of test cases T. $Execute(O, t_i)$ denotes running the model-generated code O with the i-th test case.

$$\mathcal{R}_{\text{SandBox}}(O, T) = \frac{1}{|T|} \sum_{i=1}^{|T|} \mathbf{1}_{\text{Execute}(O,t_i)=\text{Pass}} \tag{3}$$

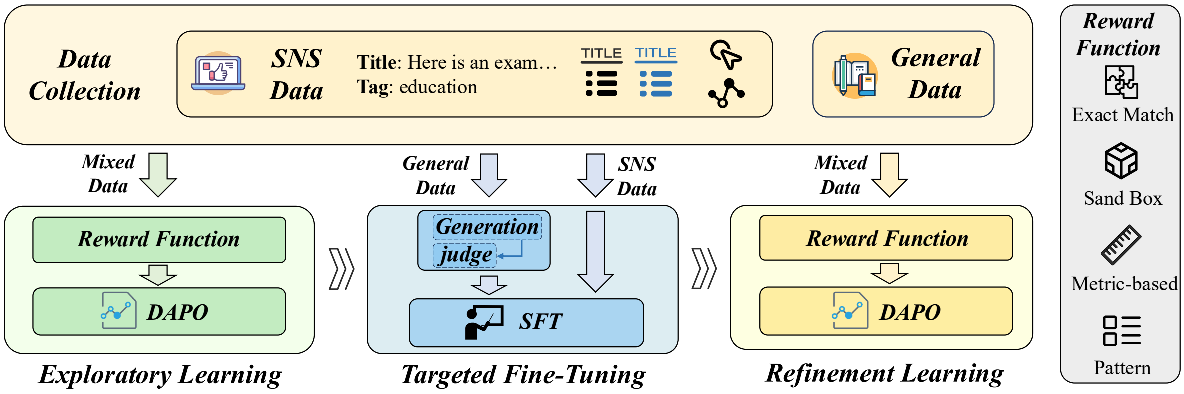

Figure 2: Overview of our RL-based incremental training pipeline. Implementation details can be found in App. B.

4) **Pattern.** For complex instruction-following tasks, we compare the model's output O against a set of given conditions C to determine satisfaction item by item.

$$\mathcal{R}_{\text{Pattern}}(O, C) = \frac{1}{|C|} \sum_{i=1}^{|C|} \mathbf{1}_{\text{satisfies}(O, c_j)} \tag{4}$$

2.2.3 Domain alignment

In this stage, we randomly mix $\mathcal{D}_{\text{SNS}_1}$ and $\mathcal{D}_{\text{GEN}_1}$ to form $\mathcal{D}_1$, which conducts DAPO-based (Yu et al., 2025) RL training. For a specific data entry (Q, A), DAPO samples a group of G individual candidate outputs $\{O_i\}_{i=1}^{G}$ from the old policy model $\pi_{\theta_{\text{old}}}$. Then, we can optimize the policy π_θ by optimizing the following loss function:

$$\begin{aligned} \mathcal{L}_{\text{DAPO}}(\theta) &= \mathbb{E}_{(Q,A)\sim\mathcal{D}_1, \{O_i\}_{i=1}^{G}\sim\pi_{\theta_{\text{old}}}(\cdot|Q)} \\ &\Bigg[\frac{1}{\sum_{i=1}^{G} |O_i|} \sum_{i=1}^{G} \sum_{t=1}^{|O_i|} \min\Big(r_{i,t}(\theta)\, \hat{A}_{i,t}, \\ &\quad \text{clip}(r_{i,t}(\theta),\, 1-\varepsilon_{\text{low}},\, 1+\varepsilon_{\text{high}})\, \hat{A}_{i,t} \Big) \Bigg] \\ \text{s.t.} \quad & 0 < \Big| \{O_i \mid \text{is\_equivalent}(A, O_i)\} \Big| < G \end{aligned} \tag{5}$$

where ε_{low} and $\varepsilon_{\text{high}}$ control the clipping range. This constraint ensures the sampled group contains both correct and incorrect responses, yielding non-zero advantage."

$$r_{i,t}(\theta) = \frac{\pi_\theta(O_{i,t} \mid Q, O_{i,<t})}{\pi_{\theta_{\text{old}}}(O_{i,t} \mid Q, O_{i,<t})} \tag{6}$$

$$\hat{A}_{i,t} = \frac{\mathcal{R}_i - \text{mean}(\{\mathcal{R}_i\}_{i=1}^{G})}{\text{std}(\{\mathcal{R}_i\}_{i=1}^{G})} \tag{7}$$

where $\mathcal{R}_i$ denotes the reward of the i-th sampled response, computed by the task-appropriate reward function among $\{\mathcal{R}_{\text{EM}}, \mathcal{R}_{\text{Met}}, \mathcal{R}_{\text{SandBox}}, \mathcal{R}_{\text{Pattern}}\}$.

Finally, this alignment stage yields broad, stable gains without premature specialization, while producing a fine-grained underperformance diagnosis for the subsequent targeted repair.

2.3 Targeted Fine-Tuning

After initial SNS alignment, Targeted Fine-Tuning directly addresses tasks that remain weak. The emphasis is on repairing deficiencies while preserving previous gains, achieved by blending difficult SNS data with carefully filtered general data.

2.3.1 Data preparation

We construct $\mathcal{D}_2$ comprising 1.7M SNS instances and 100K general instances. The SNS subset is curated from tasks where the model underperformed in the previous stage, with rare but impactful cases upweighted. For the general subset, we employ soft labels to mitigate catastrophic forgetting: for each prompt, the model from the first stage generates eight candidates, and the best-performing one (determined by CompassJudger-2-32B-Instruct (Zhang et al., 2025)) is selected as the supervision signal. This approach functions as a data-level regularizer, reducing the distributional gap between ground-truth labels and the model's learned distribution, thereby improving learning efficiency (Li and Hoiem, 2017).

2.3.2 Targeted learning

In this stage, we apply a standard SFT objective to the mixture of hard SNS examples and soft-labeled

Table 1: Comparison results across General-Bench, SNS-Bench and SNS-TransBench. **Bold** entries indicate the best performance, while underlined entries denote the second one in each category.

Models	General-Bench	SNS-Bench									SNS-TransBench				
	Avg.	Taxon.	Hash.	QCorr	MRC	NER	Gender	CHLW	QGen	Avg.	ZH→EN		EN→ZH		Avg.
											BLEU	chrF++	BLEU	chrF++	
Proprietary Large Language Models or The Scale of Large Language Models > 100B															
GPT-4o-1120	70.72	65.79	84.98	51.79	58.89	54.99	88.08	38.96	47.33	61.35	40.32	63.91	49.15	47.28	50.17
Gemini-2.0-Flash	74.42	68.76	87.36	48.41	52.21	53.58	89.64	37.39	46.27	60.45	32.72	58.84	41.80	40.16	43.38
Claude-3.7-Sonnet	75.10	72.03	88.83	54.10	54.86	56.13	92.23	31.11	45.49	61.85	35.63	61.66	45.79	44.23	46.83
Doubao-1.5-Pro-32k	76.13	30.00	83.21	58.25	61.32	56.60	90.67	30.61	46.55	57.15	33.71	61.85	45.54	44.35	46.36
Qwen-Max	71.86	65.68	84.47	54.36	61.34	55.78	91.19	37.97	46.64	62.18	35.55	60.92	46.08	44.14	46.67
GLM-4-Plus	70.25	65.46	84.31	52.13	55.81	53.16	86.53	30.09	44.68	59.02	41.57	65.95	48.79	47.06	50.84
GPT-OSS-120B	76.71	67.20	86.04	56.83	61.45	55.84	91.19	38.53	45.61	62.84	33.06	59.73	42.67	40.47	43.98
dots.llm1	70.20	62.96	82.45	42.10	40.75	14.93	89.12	31.09	44.63	51.00	30.93	58.66	44.42	42.8	44.20
GLM-4.5	73.66	70.76	86.93	56.22	64.94	57.23	92.75	41.32	45.47	64.45	30.57	56.77	39.55	38.2	41.27
Deepseek-V3-0324	75.22	67.27	86.59	47.71	60.97	56.00	90.16	40.45	46.03	61.90	35.65	61.58	46.86	44.58	47.17
DeepSeek-V3.1	77.02	70.20	88.97	48.67	62.37	55.22	91.19	33.60	46.42	62.08	31.94	58.8	41.64	39.77	43.04
The Scale of Large Language Models < 10B															
Qwen3-4B	69.80	60.88	81.90	38.31	34.69	44.50	79.27	28.17	46.75	51.81	26.87	54.26	36.35	35.41	38.22
Qwen2.5-7B	63.01	49.50	73.80	42.37	45.32	45.41	88.08	33.76	44.65	52.86	31.43	55.91	38.36	36.48	40.55
Llama-3.1-8B	51.24	37.74	66.62	33.32	31.27	47.10	74.61	26.88	38.60	44.52	23.07	48.15	29.32	29.13	32.42
Ministral-8B	49.93	42.62	70.58	36.24	30.71	37.79	82.38	28.04	46.27	46.83	25.67	50.91	32.02	31.18	34.95
InternLM3-8B	58.55	51.83	76.98	38.65	25.25	39.41	66.84	44.71	43.46	48.39	24.85	50.44	35.58	34.04	36.23
Qwen3-8B	66.90	58.67	82.44	46.47	48.45	44.68	89.12	27.95	45.89	55.46	33.21	58.81	40.09	38.85	42.74
GLM-4-9B-0414	63.27	56.03	77.67	38.03	45.29	47.01	51.30	27.51	45.52	48.55	32.20	56.90	39.73	37.40	41.57
RedOne-7B	63.83	72.18	88.02	65.09	63.98	51.86	70.47	74.73	48.69	66.88	38.06	62.66	46.88	44.82	**48.11**
RedOne 2.0 4B	**70.80**	75.85	89.05	60.92	66.54	43.15	78.76	79.11	47.17	**67.57**	38.61	62.46	45.78	43.84	47.67
10B < The Scale of Large Language Models < 100B															
Phi-4-14B	63.00	57.62	79.56	46.32	53.39	44.99	89.12	29.23	44.76	55.62	31.28	57.23	37.58	36.68	40.69
GPT-OSS-20B	74.76	62.89	83.99	54.58	56.43	54.81	92.23	32.68	45.26	60.36	30.74	57.46	37.83	36.19	40.56
Mistral-Small-24B	65.63	64.88	83.89	48.77	46.51	52.09	91.19	32.10	46.01	58.18	31.29	56.72	39.28	37.32	41.15
Qwen3-30B-A3B	74.46	64.29	85.81	44.75	52.23	45.75	90.16	27.19	45.67	56.98	34.07	58.86	41.19	39.51	37.05
GLM-4-32B-0414	74.39	63.36	85.50	47.33	53.72	50.41	80.31	33.19	46.90	57.59	36.32	61.31	42.53	40.77	45.23
Qwen2.5-32B	71.68	59.90	80.51	46.00	55.04	54.51	90.67	38.84	45.66	58.89	32.56	58.14	42.34	40.71	43.44
Qwen3-32B	72.67	61.52	86.04	49.39	54.56	53.76	91.19	33.48	45.74	59.46	32.15	58.54	40.44	38.85	42.50
Llama-3.3-70B	67.64	62.94	83.28	50.76	27.38	56.09	91.19	33.58	46.41	56.45	34.00	59.18	41.25	39.56	43.50
RedOne-32B	73.72	81.45	90.19	67.07	59.24	51.66	81.87	70.40	50.37	69.03	40.55	64.54	48.20	46.05	**49.84**
RedOne 2.0 30B-A3B	**75.17**	77.02	89.99	63.76	62.16	54.15	81.87	74.19	49.15	**69.04**	40.22	63.88	48.06	45.95	49.54

general data:

$$\mathcal{L}_{\text{SFT}} = -\ \mathbb{E}_{(Q,A)\sim\mathcal{D}_2, O\sim\pi_\theta(\cdot|Q)} \sum_{t=1}^{|O|} \log \pi_\theta\big(O_t \,\big|\, Q, O_{<t}\big) \qquad (8)$$

where π_θ denotes the current policy model, Q is the question, and O is the target output sequence. This stage consistently improves SNS-related weakness while maintaining the gains achieved in the initial alignment, ensuring the model is robustly prepared for the final RL refinement.

2.4 Refinement Learning

The final stage, Refinement Learning, consolidates prior improvements and further boosts performance by applying RL centered on SNS data.

2.4.1 Further refinement

We utilize approximately 400K examples from both SNS and general domains, focusing on difficult subsets. Initializing the policy from the prior stage, we employ DAPO (Yu et al., 2025) for refinement. We also increase the proportion of rationale-augmented samples to 57.18% to preserve model's reasoning ability across downstream tasks. Compared to pure SFT, this RL-based refinement ensures smoother convergence and more robust domain adaptation within the explored solution space. The resulting model exhibits stabilized behavior and consistent performance gains across both SNS-specific and general-purpose benchmarks.

3 Experiments

3.1 Experimental Setup

We compare RedOne 2.0 against a broad spectrum of baselines, including proprietary models (Hurst et al., 2024; Team et al., 2023; Anthropic, 2025; Doubao-Team, 2025; Yang et al., 2024; GLM et al., 2024), open-source models (Yang et al., 2025; Grattafiori et al., 2024; Mistral-AI, 2024, 2025; Cai et al., 2024; Abdin et al., 2024; Huo et al., 2025; Agarwal et al., 2025; Zeng et al., 2025; Liu et al., 2024), and the SNS-specific model RedOne (Zhao et al., 2025). For evaluation, we adopt widely used benchmarks covering both general and SNS capabilities. In the general domain, we assess six dimensions: *knowledge reasoning* with MMLU (Hendrycks

Table 2: Generalization of our training pipeline over different base models, where 4B and 30B-A3B indicate instruct version, while 8B and 32B denote base version.

Models	**General-Bench**	**SNS-Bench**	**SNS-TransBench**
Qwen3-4B	69.80	51.81	38.22
RedOne 2.0 4B	70.80	67.57	47.67
Qwen3-8B	66.90	55.46	42.74
RedOne 2.0 8B	69.27	65.82	46.72
Qwen3-30B-A3B	74.46	56.98	37.05
RedOne 2.0 30B-A3B	75.17	69.04	49.54
Qwen3-32B	72.67	59.46	42.50
RedOne 2.0 32B	73.17	69.76	49.11

et al., 2020), CMMLU (Li et al., 2023a), C-Eval (Huang et al., 2023), GPQA-Diamond (Rein et al., 2024), NewsBench (Li et al., 2024), MMLU-Pro (Wang et al., 2024), BBH (Suzgun et al., 2022), and GaokaoBench (Zhang et al., 2023b); *mathematical reasoning* with GSM8K (Cobbe et al., 2021), MATH500 (Hendrycks et al., 2021), and AIME 2025 (MAA, 2025); *code generation* with HumanEval (Chen et al., 2021), MBPP (Austin et al., 2021), and LiveCodeBench (Jain et al., 2024); *machine translation* with WMT-22 (Kocmi et al., 2022), WMT-23 (Freitag et al., 2023), WMT-24 (Kocmi et al., 2024), and FLORES (Goyal et al., 2022); *instruction following* with IFEval (Zhou et al., 2023); and *hallucination detection* with HaluEval (Li et al., 2023b), together with CompassBench (Contributors, 2023) for an integrated multi-dimensional view. In the SNS domain, we use SNS-Bench (Guo et al.) and SNS-TransBench (Guo et al., 2025b) to conduct evaluation. Further details are provided in App. C and App. D.

3.2 Main Results

The details about benchmarks and baselines are listed in App. D and C. As shown in Table 1, RedOne 2.0 delivers consistently strong and well-balanced performance across all benchmarks, outperforming much larger open-source and proprietary models. On General-Bench, the 4B model scores 70.80, surpassing Qwen3-8B and GLM-4-9B, while 30B-A3B reaches 75.17 even matching DeepSeek-V3, indicating that our post-training pipeline not only enables effective domain adaptation but also improves general capability. For SNS-specific evaluation, RedOne 2.0 leads its scale groups and matches frontier models on both SNS-Bench and SNS-TransBench, demonstrating the progressive alignment effectively captures nuanced community norms and strengthens domain-specific translation without sacrificing the base model's multilingual versatility. Moreover, the performance of RedOne 2.0 improves steadily with model scales and exhibits clear gains over RedOne, suggesting that RL-centric post-training recipe offers a more data-efficient, stable, and scalable alternative to prior SFT-dominated approach.

Table 3: The impact of different training stage on Qwen3-4B's performance.

Exploratory Learning	Targeted Fine-Tuning	Refinement Learning	General -Bench	SNS -Bench	SNS-Trans Bench
			69.80	51.81	38.22
	✓		63.65	61.10	46.00
	✓	✓	69.80	63.03	45.95
✓			71.25	62.27	43.35
✓	✓		70.04	65.67	47.72
✓	✓	✓	70.80	67.57	47.67
✓		✓	69.64	61.58	42.37

3.3 More Analysis

3.3.1 Generalization on various base models.

Table 2 shows that our three-stage pipeline transfers reliably across base-model scales, yielding consistent gains on all benchmarks. Performance further improves as the backbone scales up, suggesting that larger models leverage the staged optimization signals more effectively. Notably, the 4B and 30B-A3B variants achieve stronger gains than similar scale models especially on SNS benchmarks, which we attribute to their instruction-tuned backbones that better absorb the multi-stage alignment signals for domain adaptation.

3.3.2 Effect of each training stage.

As shown in Table 3, we analyze the incremental contribution of each stage in our framework. Exploratory Learning (EL) provides a strong initial lift, reaching 71.25% on General-Bench, 62.27% on SNS-Bench, and 43.35% on SNS-TransBench. Targeted Fine-Tuning (SFT) then corrects SNS-specific weaknesses, improving SNS-Bench and SNS-TransBench to 65.67% and 47.72%, with only a minor 1.21% drop on General-Bench. Finally, Refinement Learning (RL) harmonizes trade-offs across tasks, raising the overall average from 61.14% to 62.01% and yielding final scores of 70.80%, 67.57%, and 47.67% on General-Bench, SNS-Bench, and SNS-TransBench, respectively. Additionally, we also find that removing the intermediate SFT yields no gain over EL-only and can even slightly degrade performance, since RL alone saturates within the model's current policy distribution without SFT to shift it and unlock new capability for the subsequent refinement.

Table 4: Comparison with task specific fine-tuning on Qwen3-4B and RedOne2.0 4B.

Models	**Hash.**	**QCorr**	**MRC**
Qwen3-4B	81.90	38.31	34.69
Qwen3-4B (Fine-tuned)	90.12	60.11	57.54
RedOne 2.0 4B	89.05	60.92	66.54
Models	**CHLW**	**QGen**	**SNS-Trans**
Qwen3-4B	28.17	46.75	38.22
Qwen3-4B (Fine-tuned)	67.24	49.24	44.25
RedOne 2.0 4B	79.11	47.17	47.67

Table 5: RedOne2.0's online application on personalized re-creation of post titles.

Metrics		**Relative Change**
Business Value	Advertiser Value (AdvV) ↑	+0.43%
Content Quality	Vague Titles Ratio ↓	-11.9%
	Practical Titles Ratio ↑	+7.1%
	Authentic Titles Ratio ↑	+12.9%
	Interactive Titles Ratio ↑	+25.8%

Furthermore, the most notable shift in RedOne 2.0 lies in its departure from the traditional SFT-centric domain-specific post-training paradigm to RL, we conduct experiments to compare it with naive SFT followed RL baseline, as shown in rows 2-3 of Table 3. Although SFT improves SNS specific performance, it induces a strong "seesaw" drop in general capability, and the subsequent RL only partially recovers it. In contrast, RedOne 2.0 refines the process, avoiding the general–domain trade-off and outperforming the naive baseline by 1.00 on General-Bench, 4.54 on SNS-Bench, and 1.72 on SNS-TransBench.

3.3.3 Comparison with task specific tuning.

We also compare RedOne 2.0, which jointly optimizes all tasks, with conventional task-specific fine-tuning. As shown in Table 4, this method also yields strong performance on its target objective. While task-specific SFT yields strong single-task results (e.g., QGen 49.24 and Hash. 90.12 on Qwen3-4B), RedOne 2.0 4B, trained on a mixture of all tasks, delivers robust and highly competitive results across the entire spectrum of benchmarks (e.g., outperforming task-specific tuning on MRC with 9.00 and CHLW with 11.87). This result substantiates that unified training better exploits inter-task transfer and produces a single model with more comprehensive capability.

Table 6: Results (accuracy) for more online applications.

Task	**Baseline**	**RedOne 2.0**
Negative Attitude Detection	90.45	96.18
Information Leakage Detection	71.39	94.85

3.4 Online Application

We deployed RedOne 2.0 on a large scale social networking platform with millions of users to generate personalized post-title rewrites in real time. For each draft title, the system produces a rewrite that preserves intent while optimizing engagement. The suggestion is then exhibited to creators for selection and also applied directly for comparison. We evaluate both business impact and content quality: Advertiser Value (AdvV) as the primary metric, and human review on vagueness, practicality, authenticity, and interactivity. As shown in Table 5, a multi-week A/B test over millions of posts yields consistent gains: AdvV increases by +0.43% (a statistically significant improvement at platform scale), while human evaluation shows fewer vague titles (-11.9%) and more practical (+7.1%), authentic (+12.9%), and interactive (+25.8%) titles. Therefore, deploying RedOne 2.0 not only improves user experience, but also yields measurable advertiser value, which demonstrates its effectiveness for real world content optimization. We also deploy RedOne 2.0 in additional production scenarios, including negative-attitude and information-leakage detection for platform customer service in Table 6, highlighting its practical values.

3.5 Case Study

To qualitatively assess RedOne 2.0, we compare its personalized title re-creation with a baseline. As shown in Table 7, RedOne 2.0 typically produces more vivid and engaging phrasing. For dehumidification, it introduces a vivid word of a steamy home and adds a clear imperative that heightens emotional resonance. For wedding photography, it reframes the content as an "exclusive discovery" that better triggers curiosity and click intent. We also report a failure case in settings that require strict faithfulness (Table 8). The original title emphasizes a specific lesson about choosing the wrong transportation card for Osaka and Kyoto; while the baseline preserves this focus, RedOne 2.0 over-generalizes to Kansai travel and drops the key "transportation card" detail, weakening informational precision. Overall, RedOne 2.0 excels

Table 7: Good cases for post titles re-creation.

Title Source	Content
Case 1	
Original	Plum rain season, a great helper for dehumidification and mold prevention.
Base Model	Dehumidification essential for the plum rain season,a fresh choice for a dry and comfortable life.
RedOne 2.0	Say goodbye to "steamy" homes! Rescue your plum rain season.
Case 2	
Original	Beijing wedding photo recommendations: 17 lawn wedding photo outdoor spots.
Base Model	Dreamy lawn wedding photos, capturing the most beautiful moments.
RedOne 2.0	Escape the studio! 17 stunning lawns capture cinematic-level wedding photos.

Table 8: Bad case for post titles re-creation.

Title Source	Content
Original	Don't buy the wrong transportation card for Osaka and Kyoto! A lesson learned the hard way!
Base Model	A guide to Japanese transportation cards—stop making these mistakes!
RedOne 2.0	Avoid these pitfalls for your Kansai trip, check out the guide now.

at engagement-oriented and stylistically appealing rewriting, but it can occasionally over-optimize and omit essential details, motivating future work on stronger faithfulness constraints. More case studies can be found in App. E.

4 Conclusion

In this paper, we present RedOne 2.0, an SNS-specific LLM post-training framework tailored for SNS, where tasks are highly heterogeneous, dynamic, and culturally diverse. Unlike traditional SFT-centric approaches that risk catastrophic forgetting and unstable trade-offs between in-domain and out-of-domain performance, RedOne 2.0 adopts a progressive, RL-prioritized three-stage pipeline: Exploratory Learning to establish initial domain alignment and surface weaknesses, Targeted Fine-Tuning to selectively repair deficiencies while retaining general competence, and Refinement Learning to consolidate improvements in various scenarios. Supported by a large, task-diverse dataset spanning more than 75 SNS tasks and high quality general corpus, this paradigm demonstrates strong data efficiency, stable adaptation, and robust generalization even at compact model scales. Overall, RedOne 2.0 provides a competitive, cost-effective, and scalable baseline for LLM post-training in SNS, advancing model's domain capability without sacrificing robustness, or general usability.

Limitations

Despite strong performance on text-centric SNS tasks, RedOne 2.0 remains limited in several aspects. First, our pipeline is built on language-only backbones and text-form supervision, so it cannot natively handle multimodal SNS content (e.g., images, video, audio) or exploit cross-modal cues such as visual context, OCR, layout, and prosody, which constrains its applicability to common real-world scenarios like meme understanding and image-grounded moderation. Second, although we use multiple task-specific rewards, they are still imperfect proxies for user satisfaction and may miss nuanced objectives (e.g., subtle humor, cultural appropriateness, long-horizon engagement).

References

Marah Abdin, Jyoti Aneja, Harkirat Behl, Sébastien Bubeck, Ronen Eldan, Suriya Gunasekar, Michael Harrison, Russell J Hewett, Mojan Javaheripi, Piero Kauffmann, and 1 others. 2024. Phi-4 technical report. *arXiv preprint arXiv:2412.08905*.

Sandhini Agarwal, Lama Ahmad, Jason Ai, Sam Altman, Andy Applebaum, Edwin Arbus, Rahul K Arora, Yu Bai, Bowen Baker, Haiming Bao, and 1 others. 2025. gpt-oss-120b & gpt-oss-20b model card. *arXiv preprint arXiv:2508.10925*.

Sacha Altay, Emma Hoes, and Magdalena Wojcieszak. 2025. Following news on social media boosts knowledge, belief accuracy and trust. *Nature Human Behaviour*, pages 1–10.

Claude Anthropic. 2025. 3.7 sonnet and claude code.

Jacob Austin, Augustus Odena, Maxwell Nye, Maarten Bosma, Henryk Michalewski, David Dohan, Ellen Jiang, Carrie Cai, Michael Terry, Quoc Le, and 1 others. 2021. Program synthesis with large language models. *arXiv preprint arXiv:2108.07732*.

Zhangir Azerbayev, Hailey Schoelkopf, Keiran Paster, Marco Dos Santos, Stephen McAleer, Albert Q Jiang, Jia Deng, Stella Biderman, and Sean Welleck. 2023. Llemma: An open language model for mathematics. *arXiv preprint arXiv:2310.10631*.

Lei Bai, Zhongrui Cai, Maosong Cao, Weihan Cao, Chiyu Chen, Haojiong Chen, Kai Chen, Pengcheng Chen, Ying Chen, Yongkang Chen, and 1 others. 2025. Intern-s1: A scientific multimodal foundation model. *arXiv preprint arXiv:2508.15763*.

Eytan Bakshy, Solomon Messing, and Lada A Adamic. 2015. Exposure to ideologically diverse news and opinion on facebook. *Science*, 348(6239):1130–1132.

Gagan Bhatia, El Moatez Billah Nagoudi, Hasan Cavusoglu, and Muhammad Abdul-Mageed. 2024. Fintral: A family of gpt-4 level multimodal financial large language models. *arXiv preprint arXiv:2402.10986*.

Zheng Cai, Maosong Cao, Haojiong Chen, Kai Chen, Keyu Chen, Xin Chen, Xun Chen, Zehui Chen, Zhi Chen, Pei Chu, Xiaoyi Dong, Haodong Duan, Qi Fan, Zhaoye Fei, Yang Gao, Jiaye Ge, Chenya Gu, Yuzhe Gu, Tao Gui, and 81 others. 2024. Internlm2 technical report. *Preprint*, arXiv:2403.17297.

Mark Chen, Jerry Tworek, Heewoo Jun, Qiming Yuan, Henrique Ponde De Oliveira Pinto, Jared Kaplan, Harri Edwards, Yuri Burda, Nicholas Joseph, Greg Brockman, and 1 others. 2021. Evaluating large language models trained on code. *arXiv preprint arXiv:2107.03374*.

Karl Cobbe, Vineet Kosaraju, Mohammad Bavarian, Mark Chen, Heewoo Jun, Lukasz Kaiser, Matthias Plappert, Jerry Tworek, Jacob Hilton, Reiichiro Nakano, and 1 others. 2021. Training verifiers to solve math word problems. *arXiv preprint arXiv:2110.14168*.

Pierre Colombo, Telmo Pessoa Pires, Malik Boudiaf, Dominic Culver, Rui Melo, Caio Corro, Andre FT Martins, Fabrizio Esposito, Vera Lúcia Raposo, Sofia Morgado, and 1 others. 2024. Saullm-7b: A pioneering large language model for law. *arXiv preprint arXiv:2403.03883*.

OpenCompass Contributors. 2023. Opencompass: A universal evaluation platform for foundation models. `https://github.com/open-compass/opencompass`.

Matthew Dahl, Varun Magesh, Mirac Suzgun, and Daniel E Ho. 2024. Large legal fictions: Profiling legal hallucinations in large language models. *Journal of Legal Analysis*, 16(1):64–93.

Doubao-Team. 2025. Doubao-1.5-pro: Model release. `https://team.doubao.com/en/special/doubao_1_5_pro`.

Shangbin Feng, Herun Wan, Ningnan Wang, Zhaoxuan Tan, Minnan Luo, and Yulia Tsvetkov. 2024. What does the bot say? opportunities and risks of large language models in social media bot detection. *arXiv preprint arXiv:2402.00371*.

Markus Freitag, Nitika Mathur, Chi-kiu Lo, Eleftherios Avramidis, Ricardo Rei, Brian Thompson, Tom Kocmi, Frederic Blain, Daniel Deutsch, Craig Stewart, and 1 others. 2023. Results of wmt23 metrics shared task: Metrics might be guilty but references are not innocent. In *Proceedings of the Eighth Conference on Machine Translation*, pages 578–628.

Team GLM, Aohan Zeng, Bin Xu, Bowen Wang, Chenhui Zhang, Da Yin, Dan Zhang, Diego Rojas, Guanyu Feng, Hanlin Zhao, and 1 others. 2024. Chatglm: A family of large language models from glm-130b to glm-4 all tools. *arXiv preprint arXiv:2406.12793*.

Naman Goyal, Cynthia Gao, Vishrav Chaudhary, Peng-Jen Chen, Guillaume Wenzek, Da Ju, Sanjana Krishnan, Marc'Aurelio Ranzato, Francisco Guzmán, and Angela Fan. 2022. The flores-101 evaluation benchmark for low-resource and multilingual machine translation. *Transactions of the Association for Computational Linguistics*, 10:522–538.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, and 1 others. 2024. The llama 3 herd of models. *arXiv preprint arXiv:2407.21783*.

Thomas S Gruca, Joyce E Berg, and Michael Cipriano. 2008. Incentive and accuracy issues in movie prediction markets. *The Journal of Prediction Markets*, 2(1):29–43.

Neel Guha, Julian Nyarko, Daniel Ho, Christopher Ré, Adam Chilton, Alex Chohlas-Wood, Austin Peters, Brandon Waldon, Daniel Rockmore, Diego Zambrano, and 1 others. 2023. Legalbench: A collaboratively built benchmark for measuring legal reasoning in large language models. *Advances in neural information processing systems*, 36:44123–44279.

Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song, Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma, Peiyi Wang, Xiao Bi, and 1 others. 2025a. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning. *arXiv preprint arXiv:2501.12948*.

Hongcheng Guo, Shaosheng Cao, Boyang Wang, Lei Li, Liang Chen, Xinze Lyu, Zhe Xu, Yao Hu, Zhoujun Li, and 1 others. Sns-bench: Defining, building, and assessing capabilities of large language models in social networking services. In *Forty-second International Conference on Machine Learning*.

Hongcheng Guo, Fei Zhao, Shaosheng Cao, Xinze Lyu, Ziyan Liu, Yue Wang, Boyang Wang, Zhoujun Li, Chonggang Lu, Zhe Xu, and 1 others. 2025b. Redefining machine translation on social network services with large language models. *arXiv preprint arXiv:2504.07901*.

Dan Hendrycks, Collin Burns, Steven Basart, Andy Zou, Mantas Mazeika, Dawn Song, and Jacob Steinhardt. 2020. Measuring massive multitask language understanding. *arXiv preprint arXiv:2009.03300*.

Dan Hendrycks, Collin Burns, Saurav Kadavath, Akul Arora, Steven Basart, Eric Tang, Dawn Song, and Jacob Steinhardt. 2021. Measuring mathematical problem solving with the math dataset. *arXiv preprint arXiv:2103.03874*.

Jianheng Huang, Leyang Cui, Ante Wang, Chengyi Yang, Xinting Liao, Linfeng Song, Junfeng Yao, and Jinsong Su. 2024. Mitigating catastrophic forgetting in large language models with self-synthesized rehearsal. *arXiv preprint arXiv:2403.01244*.

Yuzhen Huang, Yuzhuo Bai, Zhihao Zhu, Junlei Zhang, Jinghan Zhang, Tangjun Su, Junteng Liu, Chuancheng Lv, Yikai Zhang, Yao Fu, and 1 others. 2023. C-eval: A multi-level multi-discipline chinese evaluation suite for foundation models. *Advances in Neural Information Processing Systems*, 36:62991–63010.

Bi Huo, Bin Tu, Cheng Qin, Da Zheng, Debing Zhang, Dongjie Zhang, En Li, Fu Guo, Jian Yao, Jie Lou, and 1 others. 2025. dots. llm1 technical report. *arXiv preprint arXiv:2506.05767*.

Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, and 1 others. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276*.

Naman Jain, King Han, Alex Gu, Wen-Ding Li, Fanjia Yan, Tianjun Zhang, Sida Wang, Armando Solar-Lezama, Koushik Sen, and Ion Stoica. 2024. Livecodebench: Holistic and contamination free evaluation of large language models for code. *arXiv preprint arXiv:2403.07974*.

Julie Jiang and Emilio Ferrara. 2023. Social-llm: Modeling user behavior at scale using language models and social network data. *arXiv preprint arXiv:2401.00893*.

Songtao Jiang, Yuan Wang, Sibo Song, Yan Zhang, Zijie Meng, Bohan Lei, Jian Wu, Jimeng Sun, and Zuozhu Liu. 2025. Omniv-med: Scaling medical vision-language model for universal visual understanding. *arXiv preprint arXiv:2504.14692*.

Hangzhan Jin, Sitao Luan, Sicheng Lyu, Guillaume Rabusseau, Reihaneh Rabbany, Doina Precup, and Mohammad Hamdaqa. 2025. Rl fine-tuning heals ood forgetting in sft. *arXiv preprint arXiv:2509.12235*.

Kris-Fillip Kahl, Tolga Buz, Russa Biswas, and Gerard De Melo. 2024. Llms cannot (yet) match the specificity and simplicity of online communities in long form question answering. In *Findings of the Association for Computational Linguistics: EMNLP 2024*, pages 2028–2053.

Daniel Martin Katz, Michael James Bommarito, Shang Gao, and Pablo Arredondo. 2024. Gpt-4 passes the bar exam. *Philosophical Transactions of the Royal Society A*, 382(2270):20230254.

Tom Kocmi, Eleftherios Avramidis, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Markus Freitag, Thamme Gowda, Roman Grundkiewicz, and 1 others. 2024. Findings of the wmt24 general machine translation shared task: The llm era is here but mt is not solved yet. In *Proceedings of the Ninth Conference on Machine Translation*, pages 1–46.

Tom Kocmi, Rachel Bawden, Ondřej Bojar, Anton Dvorkovich, Christian Federmann, Mark Fishel, Thamme Gowda, Yvette Graham, Roman Grundkiewicz, Barry Haddow, and 1 others. 2022. Findings of the 2022 conference on machine translation (wmt22). In *Proceedings of the Seventh Conference on Machine Translation (WMT)*, pages 1–45.

Mahi Kolla, Siddharth Salunkhe, Eshwar Chandrasekharan, and Koustuv Saha. 2024. Llm-mod: Can large language models assist content moderation? In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*, pages 1–8.

Suhas Kotha, Jacob Mitchell Springer, and Aditi Raghunathan. 2023. Understanding catastrophic forgetting in language models via implicit inference. *arXiv preprint arXiv:2309.10105*.

Ananya Kumar, Aditi Raghunathan, Robbie Jones, Tengyu Ma, and Percy Liang. 2022. Fine-tuning can distort pretrained features and underperform out-of-distribution. *arXiv preprint arXiv:2202.10054*.

Deepak Kumar, Yousef Anees AbuHashem, and Zakir Durumeric. 2024. Watch your language: Investigating content moderation with large language models. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 18, pages 865–878.

Haonan Li, Yixuan Zhang, Fajri Koto, Yifei Yang, Hai Zhao, Yeyun Gong, Nan Duan, and Timothy Baldwin. 2023a. Cmmlu: Measuring massive multitask language understanding in chinese. *arXiv preprint arXiv:2306.09212*.

Junyi Li, Xiaoxue Cheng, Wayne Xin Zhao, Jian-Yun Nie, and Ji-Rong Wen. 2023b. Halueval: A large-scale hallucination evaluation benchmark for large language models. *arXiv preprint arXiv:2305.11747*.

Miao Li, Ming-Bin Chen, Bo Tang, Shengbin Hou, Pengyu Wang, Haiying Deng, Zhiyu Li, Feiyu Xiong, Keming Mao, Peng Cheng, and 1 others. 2024. Newsbench: a systematic evaluation framework for assessing editorial capabilities of large language models in chinese journalism. *arXiv preprint arXiv:2403.00862*.

Zhizhong Li and Derek Hoiem. 2017. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947.

Aixin Liu, Bei Feng, Bing Xue, Bingxuan Wang, Bochao Wu, Chengda Lu, Chenggang Zhao, Chengqi Deng, Chenyu Zhang, Chong Ruan, and 1 others. 2024. Deepseek-v3 technical report. *arXiv preprint arXiv:2412.19437*.

Sebastian Lubos, Thi Ngoc Trang Tran, Alexander Felfernig, Seda Polat Erdeniz, and Viet-Man Le. 2024. Llm-generated explanations for recommender systems. In *Adjunct Proceedings of the 32nd ACM Conference on User Modeling, Adaptation and Personalization*, pages 276–285.

Hanjia Lyu, Song Jiang, Hanqing Zeng, Yinglong Xia, Qifan Wang, Si Zhang, Ren Chen, Christopher Leung, Jiajie Tang, and Jiebo Luo. 2023. Llm-rec: Personalized recommendation via prompting large language models. *arXiv preprint arXiv:2307.15780*.

MAA. 2025. American invitational mathematics examination - aime. *American Invitational Mathematics Examination - AIME 2025*.

Zijie Meng, Jin Hao, Xiwei Dai, Yang Feng, Jiaxiang Liu, Bin Feng, Huikai Wu, Xiaotang Gai, Hengchuan Zhu, Tianxiang Hu, and 1 others. 2025. Dentvlm: A multimodal vision-language model for comprehensive dental diagnosis and enhanced clinical practice. *arXiv preprint arXiv:2509.23344*.

Mistral-AI. 2024. Un ministral, des ministraux. `https://mistral.ai/news/ministraux`. Accessed: 2024-10-16.

Mistral-AI. 2025. Mistral small 3.1. `https://mistral.ai/news/mistral-small-3-1`. Accessed: 2025-03-17.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, and 1 others. 2022. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35:27730–27744.

Boyu Qiao, Kun Li, Wei Zhou, Shilong Li, Qianqian Lu, and Songlin Hu. 2025. Botsim: Llm-powered malicious social botnet simulation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 14377–14385.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. 2023. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741.

Vinay Venkatesh Ramasesh, Aitor Lewkowycz, and Ethan Dyer. 2021. Effect of scale on catastrophic forgetting in neural networks. In *International conference on learning representations*.

David Rein, Betty Li Hou, Asa Cooper Stickland, Jackson Petty, Richard Yuanzhe Pang, Julien Dirani, Julian Michael, and Samuel R Bowman. 2024. Gpqa: A graduate-level google-proof q&a benchmark. In *First Conference on Language Modeling*.

Nidhish Shah, Zulkuf Genc, and Dogu Araci. 2024. Stackeval: Benchmarking llms in coding assistance. *Advances in Neural Information Processing Systems*, 37:36976–36994.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, YK Li, Yang Wu, and 1 others. 2024. Deepseekmath: Pushing the limits of mathematical reasoning in open language models. *arXiv preprint arXiv:2402.03300*.

Karan Singhal, Tao Tu, Juraj Gottweis, Rory Sayres, Ellery Wulczyn, Mohamed Amin, Le Hou, Kevin Clark, Stephen R Pfohl, Heather Cole-Lewis, and 1 others. 2025. Toward expert-level medical question answering with large language models. *Nature Medicine*, 31(3):943–950.

Mirac Suzgun, Nathan Scales, Nathanael Schärli, Sebastian Gehrmann, Yi Tay, Hyung Won Chung, Aakanksha Chowdhery, Quoc V Le, Ed H Chi, Denny Zhou, and 1 others. 2022. Challenging big-bench tasks and whether chain-of-thought can solve them. *arXiv preprint arXiv:2210.09261*.

Ross Taylor, Marcin Kardas, Guillem Cucurull, Thomas Scialom, Anthony Hartshorn, Elvis Saravia, Andrew Poulton, Viktor Kerkez, and Robert Stojnic. 2022. Galactica: A large language model for science. *arXiv preprint arXiv:2211.09085*.

Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, and 1 others. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.

Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science*, 359(6380):1146–1151.

Qianlong Wang, Keyang Ding, Bin Liang, Min Yang, and Ruifeng Xu. 2023. Reducing spurious correlations in aspect-based sentiment analysis with explanation from large language models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pages 2930–2941.

Yubo Wang, Xueguang Ma, Ge Zhang, Yuansheng Ni, Abhranil Chandra, Shiguang Guo, Weiming Ren, Aaran Arulraj, Xuan He, Ziyan Jiang, and 1 others. 2024. Mmlu-pro: A more robust and challenging multi-task language understanding benchmark. *Advances in Neural Information Processing Systems*, 37:95266–95290.

Shijie Wu, Ozan Irsoy, Steven Lu, Vadim Dabravolski, Mark Dredze, Sebastian Gehrmann, Prabhanjan Kambadur, David Rosenberg, and Gideon Mann. 2023. Bloomberggpt: A large language model for finance. *arXiv preprint arXiv:2303.17564*.

Feng Xia, Li Liu, Jie Li, Jianhua Ma, and Athanasios V Vasilakos. 2013. Socially aware networking: A survey. *IEEE Systems Journal*, 9(3):904–921.

Weiwen Xu, Hou Pong Chan, Long Li, Mahani Aljunied, Ruifeng Yuan, Jianyu Wang, Chenghao Xiao, Guizhen Chen, Chaoqun Liu, Zhaodonghui Li, and

1 others. 2025. Lingshu: A generalist foundation model for unified multimodal medical understanding and reasoning. *arXiv preprint arXiv:2506.07044*.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, and 1 others. 2025. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*.

An Yang, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chengyuan Li, Dayiheng Liu, Fei Huang, Haoran Wei, and 1 others. 2024. Qwen2. 5 technical report. *CoRR*.

Yi Yang, Yixuan Tang, and Kar Yan Tam. 2023. Investlm: A large language model for investment using financial domain instruction tuning. *arXiv preprint arXiv:2309.13064*.

Qiying Yu, Zheng Zhang, Ruofei Zhu, Yufeng Yuan, Xiaochen Zuo, Yu Yue, Weinan Dai, Tiantian Fan, Gaohong Liu, Lingjun Liu, and 1 others. 2025. Dapo: An open-source llm reinforcement learning system at scale. *arXiv preprint arXiv:2503.14476*.

Zheng Yuan, Hongyi Yuan, Chuanqi Tan, Wei Wang, Songfang Huang, and Fei Huang. 2023. Rrhf: Rank responses to align language models with human feedback without tears. *arXiv preprint arXiv:2304.05302*.

Aohan Zeng, Xin Lv, Qinkai Zheng, Zhenyu Hou, Bin Chen, Chengxing Xie, Cunxiang Wang, Da Yin, Hao Zeng, Jiajie Zhang, and 1 others. 2025. Glm-4.5: Agentic, reasoning, and coding (arc) foundation models. *arXiv preprint arXiv:2508.06471*.

Jingying Zeng, Richard Huang, Waleed Malik, Langxuan Yin, Bojan Babic, Danny Shacham, Xiao Yan, Jaewon Yang, and Qi He. 2024. Large language models for social networks: Applications, challenges, and solutions. *arXiv preprint arXiv:2401.02575*.

Dan Zhang, Ziniu Hu, Sining Zhoubian, Zhengxiao Du, Kaiyu Yang, Zihan Wang, Yisong Yue, Yuxiao Dong, and Jie Tang. 2024. Sciglm: Training scientific language models with self-reflective instruction annotation and tuning. *arXiv preprint arXiv:2401.07950*.

Taolin Zhang, Maosong Cao, Alexander Lam, Songyang Zhang, and Kai Chen. 2025. Compassjudger-2: Towards generalist judge model via verifiable rewards. *arXiv preprint arXiv:2507.09104*.

Wenxuan Zhang, Yue Deng, Bing Liu, Sinno Jialin Pan, and Lidong Bing. 2023a. Sentiment analysis in the era of large language models: A reality check. *arXiv preprint arXiv:2305.15005*.

Xiaotian Zhang, Chunyang Li, Yi Zong, Zhengyu Ying, Liang He, and Xipeng Qiu. 2023b. Evaluating the performance of large language models on gaokao benchmark. *arXiv preprint arXiv:2305.12474*.

Fei Zhao, Chonggang Lu, Yue Wang, Zheyong Xie, Ziyan Liu, Haofu Qian, JianZhao Huang, Fangcheng Shi, Zijie Meng, Hongcheng Guo, and 1 others. 2025. Redone: Revealing domain-specific llm post-training in social networking services. *arXiv preprint arXiv:2507.10605*.

Jeffrey Zhou, Tianjian Lu, Swaroop Mishra, Siddhartha Brahma, Sujoy Basu, Yi Luan, Denny Zhou, and Le Hou. 2023. Instruction-following evaluation for large language models. *arXiv preprint arXiv:2311.07911*.

A Related Work

A.1 LLM in social networking services

Given the central role of SNS platforms in everyday information exchange, interest in this domain has surged (Bakshy et al., 2015; Vosoughi et al., 2018; Altay et al., 2025). Advances in LLMs have accelerated integration across sentiment analysis (Zhang et al., 2023a; Wang et al., 2023), search and question answering over social content (Shah et al., 2024; Kahl et al., 2024), personalized content generation (Lubos et al., 2024; Lyu et al., 2023), content moderation (Kumar et al., 2024; Kolla et al., 2024), and platform operations (Feng et al., 2024; Qiao et al., 2025). Moving beyond task-specific pipelines that underuse model knowledge and generalization, recent work explores unified solutions. Social-LLM combines localized social interaction signals with text features to scale inductive user detection across seven real-world datasets (Jiang and Ferrara, 2023). (Zeng et al., 2024) organizes LLM-for-SNS applications into knowledge, engagement, and foundation tasks while outlining deployment challenges. RedOne introduces a large-scale SNS dataset and a common domain post-training recipe that yields strong offline gains and measurable online improvements (Zhao et al., 2025). Yet most approaches remain data-centric, expanding domain competence primarily through large annotated corpora and dependence on strong base models, which drives up cost. RedOne 2.0 revisits post-training in the SNS setting and achieves stronger downstream performance with substantially less data and smaller model scales.

A.2 General LLM post-training

Post-training bridges large-scale pre-training and deployment by enhancing instruction following, safety, and factuality. Typical pipelines first perform SFT on curated instruction–response pairs, then apply preference optimization with human or automated feedback, as exemplified by InstructGPT (Ouyang et al., 2022). Subsequent advances such as RRHF (Yuan et al., 2023) and DPO (Rafailov et al., 2023) simplify preference learning and improve training stability, while GRPO (Shao et al., 2024) and DAPO (Yu et al., 2025) introduce more efficient, reward-driven reinforcement learning frameworks that better balance exploration and alignment quality. Despite these developments, most approaches remain domain-agnostic and tend to underutilize specialized knowledge crucial for vertical applications.

A.3 Domain-specific LLM post-training

Domain adaptation injects targeted knowledge and preferences to boost in-domain performance. Typical pipelines combine continued pre-training on domain corpora with supervised fine-tuning and preference optimization on domain tasks, yielding strong results in finance (Wu et al., 2023; Bhatia et al., 2024; Yang et al., 2023; Gruca et al., 2008), law (Guha et al., 2023; Katz et al., 2024; Dahl et al., 2024; Colombo et al., 2024), medicine (Singhal et al., 2025; Xu et al., 2025; Meng et al., 2025; Jiang et al., 2025), and the sciences (Zhang et al., 2024; Azerbayev et al., 2023; Taylor et al., 2022; Bai et al., 2025). However, heavy reliance on SFT can overfit to benchmarks and weaken generalization, while later reinforcement learning often only partly fixes this drift. RedOne 2.0 addresses these limits with an RL-centric design that emphasizes staged reinforcement and dynamic task sampling to improve both general competence and domain reasoning.

B Implementation Details

We trained RedOne 2.0 4B and RedOne 2.0 30B-A3B using Qwen3-4B-Instruct and Qwen3-30B-A3B-Instruct as their respective base models. We also generalized our training pipeline to 8B and 32B variants with Qwen3-8B and Qwen3-32B. During the Exploratory Learning stage, we trained for 500 steps with maximum prompt/response lengths of 10,000/8,192 tokens (18,192 total), plus a 4,096-token overlong buffer with 1.0 penalty factor. We used a prompt batch size of 1,024 with 16 responses per prompt (global batch size 16,384) and mini-batch size covering 256 prompts, yielding 4 gradient updates per rollout. We adopted DAPO with clipping parameters $\varepsilon_{\text{low}} = 0.2$ and $\varepsilon_{\text{high}} = 0.28$. Optimization employed AdamW with a constant learning rate of 5×10^{-6}, weight decay 0.1, with linear warmup applied for 10 rollout steps. In Targeted Fine-Tuning, we trained for 2 epochs with batch size 64 and maximum sequence length of 16,384 using sequence packing. We optimized cross-entropy loss with AdamW at a learning rate 5×10^{-6}, applying a warmup ratio of 0.1 followed by cosine scheduling. The final Refinement Learning stage mirrored the first stage configuration.

Table 9: Cases for search keyword refinement.

	Search query	Response of RedOne 2.0
case1	Clothing front desk wall design	Minimalist style, rendering, high-end feel
case2	Office creative corner	Handmade, decoration

Table 10: Case for query-note relevance detection.

User Query	Mugwort origin
Post Content	Help, how do I go to Qichun to buy mugwort?
Response of RedOne 2.0	Query classification: A precise information-seeking query about the origin of mugwort. Content relevance: The post only discusses traveling to buy mugwort, without mentioning its origin, so it is content irrelevant. Category matching: Query relates to geographic/agricultural info; the post concerns travel/purchase, wo they are mismatch. Keyword hits: "Mugwort" appears; "origin" does not. Initial relevance score: Irrelevant content with partial keyword hits, so the score is 0. Special-case adjustment: Missing the core limiting term ("origin"), so the score remains 0. Final score: 0.

C Baselines

We conduct comparison experiments with various proprietary models, including GPT4o-1120 (Hurst et al., 2024), Gemini-2.0-Flash (Team et al., 2023), Claude-3.7-Sonnet (Anthropic, 2025), Doubao-1.5-Pro-32k (Doubao-Team, 2025), Qwen-Max (Yang et al., 2024), and GLM-4-Plus (GLM et al., 2024), open-source models, such as Qwen series (Yang et al., 2024, 2025), Llama series (Grattafiori et al., 2024), Ministral (Mistral-AI, 2024), Mistral-Small-24B (Mistral-AI, 2025), InternLM3-8B (Cai et al., 2024), Phi-4-14B (Abdin et al., 2024), dots.llm1 (Huo et al., 2025), gpt-oss series (Agarwal et al., 2025), GLM series (Zeng et al., 2025) and DeepSeek series (Liu et al., 2024), as well as SNS domain specific models RedOne (Zhao et al., 2025).

D Benchmarks

We perform a comprehensive evaluation of RedOne 2.0 and baselines in both the general and SNS domain capabilities using commonly used benchmarks in the community. Specifically, in general domain, we systematically assess six capabilities, including knowledge reasoning, mathematical reasoning, code generation, machine translation, instruction following, and hallucination detection, as well as *CompassBench* (Contributors, 2023), a comprehensive benchmark to provide an integrated, multi-dimensional view of model performance. 1) **Knowledge Reasoning.** We use *MMLU* (Hendrycks et al., 2020), *CMMLU* (Li et al., 2023a), *C-Eval* (Huang et al., 2023), *GPQA-Diamond* (Rein et al., 2024), *NewsBench* (Li et al., 2024), *MMLU-Pro* (Wang et al., 2024), *BBH* (Suzgun et al., 2022), and *GaokaoBench* (Zhang et al., 2023b) to probe broad and specialized knowledge, reasoning robustness, difficulty-calibrated multiple choice, and exam-style generalization in both English and Chinese. 2) **Mathematical Reasoning.** We adopt *GSM8K* (Cobbe et al., 2021), *MATH500* (Hendrycks et al., 2021), and the high-stakes *AIME 2025* (MAA, 2025) set to measure multi-step arithmetic and competition-level problem solving. 3) **Code Generation.** We evaluate program synthesis and correctness with *HumanEval* (Chen et al., 2021), *MBPP* (Austin et al., 2021), and the temporally refreshed, contamination-aware *LiveCodeBench* (Jain et al., 2024), reporting pass@k and execution-based metrics. 4) **Machine Translation.** We benchmark multilingual translation with the WMT tasks (i.e. WMT-22 (Kocmi et al., 2022), WMT-23 (Freitag et al., 2023) and WMT-24 (Kocmi et al., 2024)) and *FLORES* (Goyal et al., 2022), covering diverse language pairs and domains. 5) **Instruction Following.** We employ *IFEval* (Zhou et al., 2023), which provides automatically verifiable constraints to quantify compliance under explicit instructions. 6) **Hallucination Detection.** We use *HaluEval* (Li et al., 2023b) to assess the tendency to produce unverifiable or fabricated content across question answering, dialogue, and summarization settings.

In the SNS domain, we validate models on benchmarks built from real SNS scenarios, covering five aspects: post comprehension, information retrieval, sentiment and intent analysis, personalized recommendation, and translation. We use SNS-Bench (Guo et al.), a large-scale benchmark with 6,658 questions spanning eight tasks from a social platform with over 300M users, which includes the following tasks: 1) **Note-Taxonomy (Taxon.)** for content categorization; 2) **Note-Hashtag (Hash.)** to select suitable tags; 3) **Note-QueryCorr (QCorr)** to align user queries with note content and topic; 4) **Note-MRC (MRC)** for simple and complex reading comprehension over long notes; 5) **Note-NER (NER)** for entity extraction; 6) **Note-Gender (Gender)** to assess gender-sensitive appeal; 7) **Note-CHLW (CHLW)** to highlight salient words in comment threads; and 8) **Note-QueryGen (QGen)** to produce effective search queries. For translation, we adopt SNS-TransBench (Guo et al., 2025b), a curated set of 2,858 English–Chinese cases from posts, comments, and multimedia captions that emphasizes phenomena central to SNS translation, including humor localization, emoji semantics, and meme adaptation. It tests whether models can preserve pragmatics, style, and culture-bound references in short, high-context text typical of social platforms.

E More Case Studies

Apart from personalized post-title re-creation, we also present visualization analyses for two widely used SNS tasks, including search keyword refinement and query–note relevance detection, as shown in Table 9 and Table 10.