GreenMind: A LLM for Complex Reasoning – Chain-of-Thought Data and Self-Correction Cycles on NVIDIA NEMO

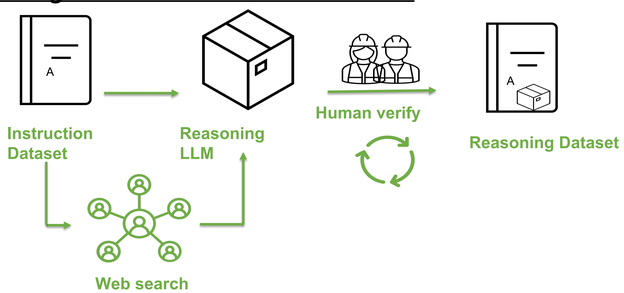
Bao Loc Pham[1] ; Viet Hoang[1]; Thu Vo[1]; GreenNodeAI[1]; NVIDIA[2]

Abstract

GreenMind, an open-source Vietnamese large language model (LLM) finetuned by GreenNode, has become the first Vietnamese LLM integrated into **NVIDIA NIM**. This milestone underscores our engineering team's R&D capabilities and commitment to sovereign AI development

Inspired by Group Relative Policy Optimization (GRPO) [1], **GreenMind** utilizes a high-quality Vietnamese reasoning dataset and two reward functions to address key GRPO challenges: language mixing and factual accuracy. We detect biased language during token sampling and use Sentence Transformer-based models to ensure the generated reasoning content remains accurate and undistorted.

Background and Dataset creation



Pre-trained models are **too general,** lack **domain-specific knowledge** for industries such as healthcare, finance, or eCommerce.
Fine-tuning is necessary to align models with specific industry requirements.

Each sample contain pair of Question-Answer Instruction

$$i \epsilon I$$

The reasoning chain r is a structure sequence of intermediate steps

$$\{s_1, s_2, s_3, \dots, s_n\}$$

The objective for model is to generate:

- a final answer $\mathbf{a} \epsilon A$
- Reasoning chain $\mathbf{r} \epsilon R$

To enrich the factual corrects, we retrieve the supplementatry context $c \epsilon C$ from web
The reasoning process can be defined as a function:

$$f: I \times C \rightarrow R \times A$$

Training data: 55,418 high quality reasoning samples

Reward function:

We fine- tune the model to focus on tasks that require generating concise answers, which involve a step-by-step reasoning process. Following DeepSeek-R1 we design four fundamental reward functions.

Reward function 1 Format

Require: Completions $\mathcal{C}$, regex of sequence format rg_s, regex of answer format rg_a, list of candidate results l_{ans}, score of completion structure $score_c$, score of answering structure $score_a$, score of answering candidate structure $score_{ac}$

function FORMAT-REWARDS($\mathcal{C}$, rg_s, rg_a, l_{ans}, $score_c$, $score_a$, $score_{ac}$)
  $l_{scores} = []$ ▷ List of scores
  for $i = 1$ to $length(\mathcal{C})$ **do**
    $score \leftarrow 1.0$
    if $not\_match(\mathcal{C}_i, rg_s)$ **then**
      $score \leftarrow score - score_c$
    end if
    $\hat{p} \leftarrow find(\mathcal{C}_i, rg_a)$ ▷ Get predictions
    if $length(\hat{p}) == 1$ **then**
      if $\hat{p} \nsubseteq l_{ans}$ **then**
        $score \leftarrow score - score_{ac}$
      end if
    else
      $score \leftarrow score - score_a$
    end if
    Append $score$ to l_{scores}
  end for
  return l_{scores}
end function

Reward function 2 Answering

Require: Completions $\mathcal{C}$, regex of answering format rg_a, list of candidate results l_{ans}, score of answering candidates $score_{ac}$, ground truth gt

function ANSWERING-REWARDS($\mathcal{C}$, rg_a, l_{ans}, $score_{ac}$, gt)
  $l_{scores} = []$ ▷ List of scores
  for $i = 1$ to $length(\mathcal{C})$ **do**
    $score \leftarrow 0.0$
    $\hat{p} \leftarrow find(\mathcal{C}_i, rg_a)$ ▷ Get predictions
    if $length(\hat{p}) == 1$ **then**
      if $\hat{p} == gt$ **then**
        $score \leftarrow 1$
      else if $\hat{p} \subset l_{ans}$ **then**
        $score \leftarrow score_{ac}$ ▷ For multiple-choices tasks
      else
        $score \leftarrow 0$
      end if
    else
      $score \leftarrow 0$
    end if
    Append $score$ to l_{scores}
  end for
  return l_{scores}
end function

Reward function 3 Language

Require: Completions $\mathcal{C}$, List of banned letters l_{bl}

function LANGUAGE-REWARDS($\mathcal{C}$, l_{bl})
  $l_{scores} = []$ ▷ List of scores
  for $i = 1$ to $length(\mathcal{C})$ **do**
    $score \leftarrow 1.0$
    if $exist(\mathbf{c} \in \mathcal{C}_i, l_{bl})$ **then**
      $score \leftarrow 0.0$
    end if
    Append $score$ to l_{scores}
  end for
  return l_{scores}
end function

Reward function 4 Semantic Similarity of Reasoning Content

Require: Completions $\mathcal{C}$, Sentence Transformers model ST_model, reasoning data rs, similarity threshold ξ

function SS-REWARDS($\mathcal{C}$, ST_model, rs, ξ)
  $l_{scores} = []$ ▷ List of scores
  for $i = 1$ to $length(\mathcal{C})$ **do**
    $score \leftarrow ST\_model(\mathcal{C}_i, rs_i)$
    if $score < \xi$ **then**
      $score \leftarrow 0.0$
    end if
    Append $score$ to l_{scores}
  end for
  return l_{scores}
end function

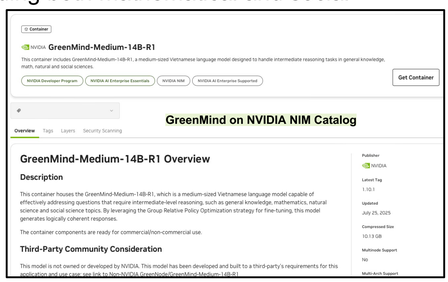
Experiments

The entire framework is built upon accelerated libraries such as **NVIDIA NEMO**, vLLM, CUDA-compiled Pytorch and Transformers.

Train with 8x NVIDIA H100-80GB GPUs for finetuning and utilizing vLLM for backend inference in GRPO process

Results

On the **SeaExam** [2] multiple-choice dataset show that our reasoning model outperforms most Southeast Asian languages, as well as the overall average across all languages, when compared to baseline models with significantly larger parameter sizes including those with up to 70 billion parameters
Regarding the **VMLU** [3] dataset, when this research was conducted, **GreenMind** is the first opensource that has best result on finetuned model. Indicating that our model performs slightly better across most topics, including both mathematical and social science domains.



Model	SeaExam-ID	SeaExam-TH	SeaExam-VI	Avg ↑
Meta-Llama-3.1-70B-Instruct	0.6633	0.3473	21.9864	0.4455
Gemma-3-27b-it	0.6556	0.342	17.5614	0.438
Qwen2.5-14B-Instruct	0.6122	0.2892	189.7782	0.3335
GreenMind-Medium-14B-R1	**0.8689**	**0.7796**	**10.7609**	**0.7915**

Table 1. SeaExam performance on SOTA models

Model	Type	STEM	Social Science	Humanities	Others	Avg ↑
VNPTAI.IO-Large-v4	Proprietary	78.05	79.05	75.39	70.37	76.21
CakebyVPBank-Large	Proprietary	77.75	78.11	70.38	67.82	73.99
DeepSeek-R1-Distill-Llama-70B	Open-source	76.77	76.23	67.98	66.82	72.41
GreenMind-Medium-14B-R1	Open-source	76.78	77.36	72.32	69.03	74.29

Table 2. VMLU performance on SOTA models

Conclusion

We release **GreenMind-Medium-14B-R1**, a medium-sized Vietnamese language model capable of effectively addressing questions that require intermediate-level reasoning, such as general knowledge and social science topics. By leveraging the GRPO [1] strategy for fine-tuning, we guide the model to generate logically coherent responses. This approach aims to provide users with informative answers, as well as intuitive explanations valuable not only for end users but also for further research in improving data quality and sampling techniques.

References

[1] Shao, Z., Wang, P., Zhu, Q., Xu, R., Song, J., Bi, X., Zhang, H., Zhang, M., Li, Y. K., Wu, Y., & Guo, D. (2024, February 5). *DeepSeekMath: Pushing the limits of mathematical reasoning in open language models*. arXiv.org. https://arxiv.org/abs/2402.03300

[2] Liu, C., Zhang, W., Ying, J., Luu, A. T., & Bing, L. (2025, February 10). *SeaExam and SeaBench: Benchmarking LLMs with Local Multilingual Questions in Southeast Asia*. arXiv.org. https://arxiv.org/abs/2502.06298

[3] Bui, C. T., Son, N. T., Van Trang, T., Phung, L. V., Huy, P. N., Le, H. A., Van, Q. H., Nguyen-Thuan, P., DO, Truc, V. L. T., Chau, D. T., & Nguyen, L. (2025). VMLU Benchmarks: A comprehensive benchmark toolkit for Vietnamese LLMs. *ACL*, 11495–11515. https://doi.org/10.18653/v1/2025.acl-long.563