Llama-Nemotron: Efficient Reasoning Models

Soumye Singhal*, Jiaqi Zeng*, Alexander Bukharin*, Yian Zhang*, Gerald Shen*, Ameya Sunil Mahabaleshwarkar*, Bilal Kartal*, Yoshi Suhara*, Akhiad Bercovich*, Itay Levy*, Izik Golan*, Mohammed Dabbah*, Ran El-Yaniv*, Somshubra Majumdar*, Igor Gitman*, Evelina Bakhturina*, Jimmy J. Zhang*, Bor-Yiing Su*, Guyue Huang*, Izzy Putterman*, Mostofa Patwary*, Oluwatobi Olabiyi*, Olivier Delalleau*, Bryan Catanzaro*, Boris Ginsburg*, Tugrul Konuk*, **Oleksii Kuchaiev***

* Core contributors. For full contributors list see Technical report.

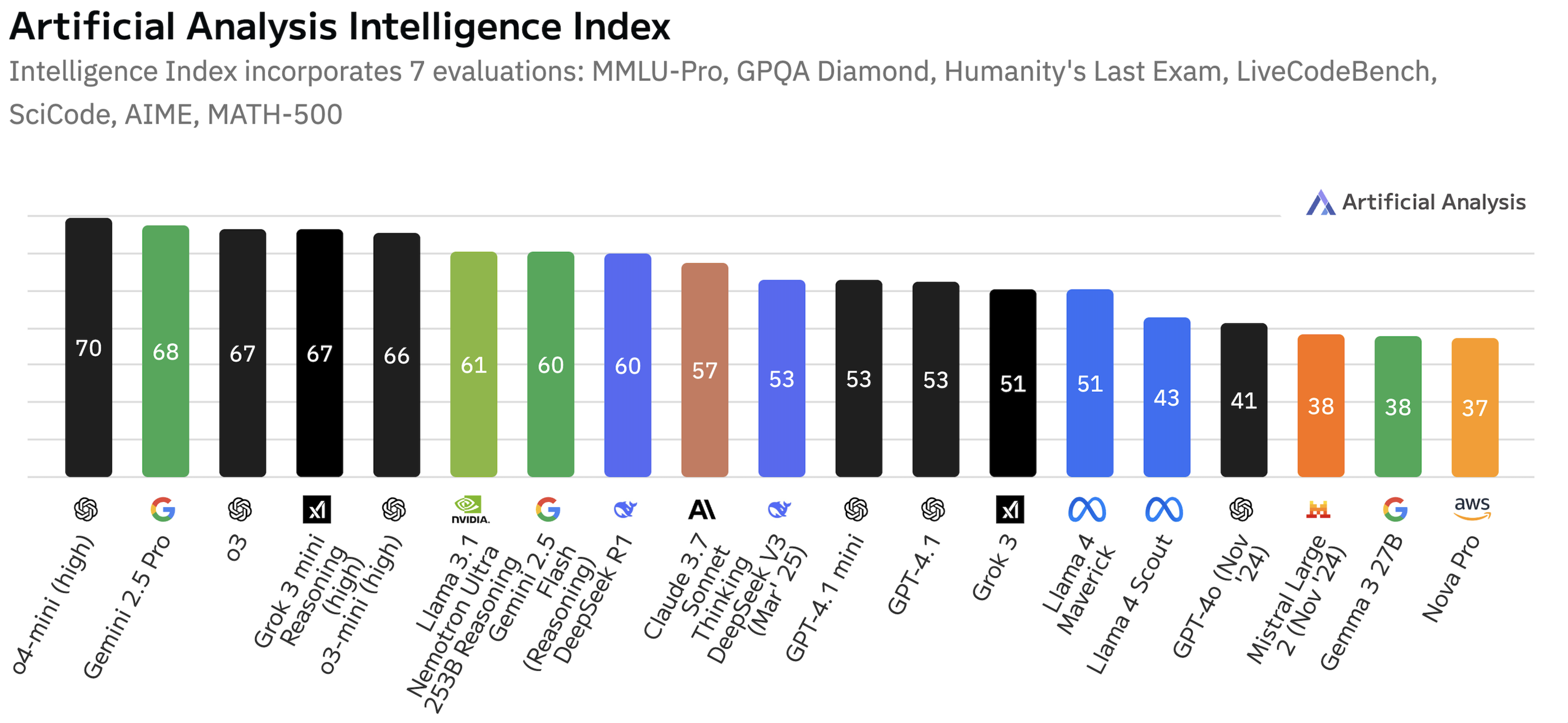
1. Llama-Nemotron-V1 family

Open-weights, open source post-training SW, open post-training & RLHF data. First open-weights with reasoning control On/Off.

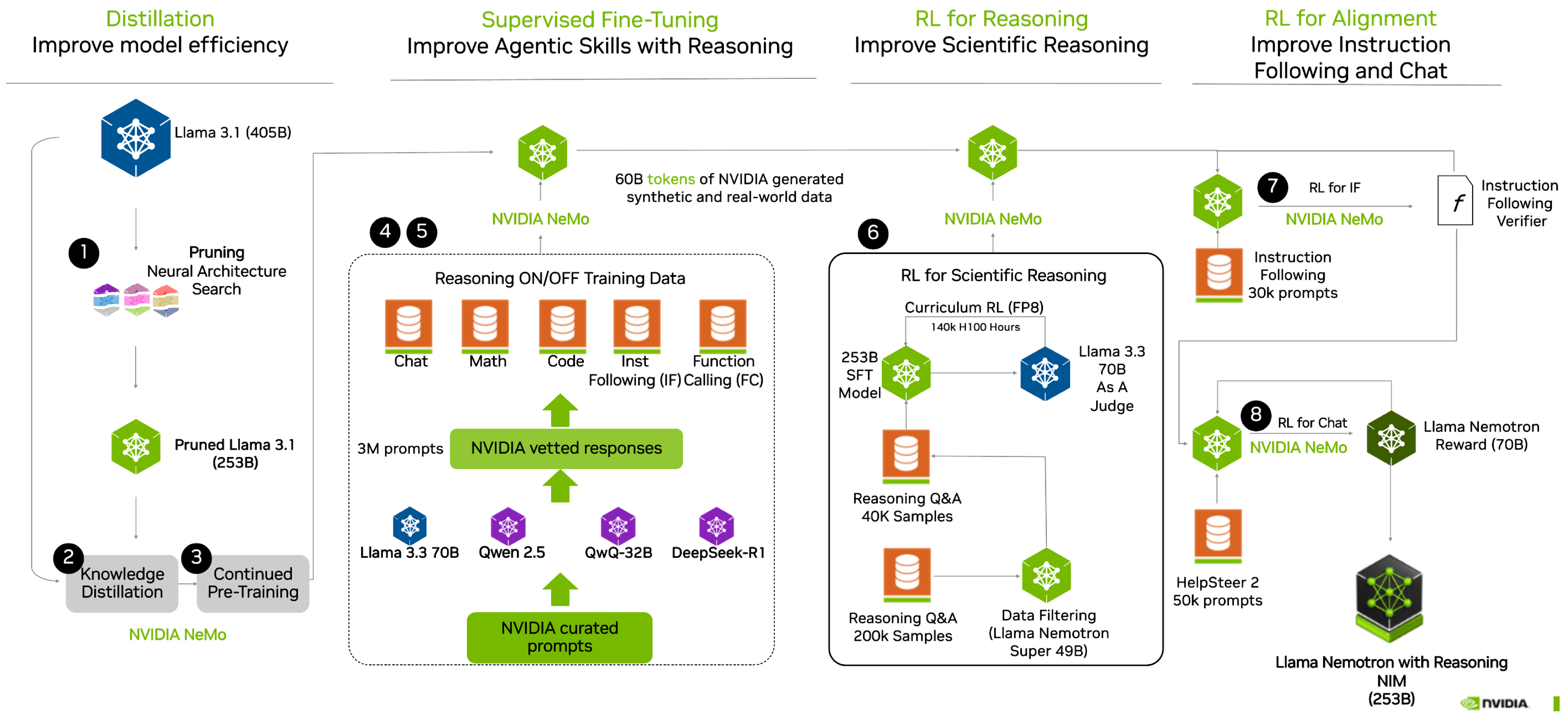
3 model sizes:

- **LN-Nano** (8B and 4B)
- **LN-Super** (49B)
- **LN-Ultra** (253B)

Smartest open-weights model as of April 2025. The highest ranking llama variant on lmarena.ai



2. Post-training pipeline



3. RLVR for Scientific Reasoning

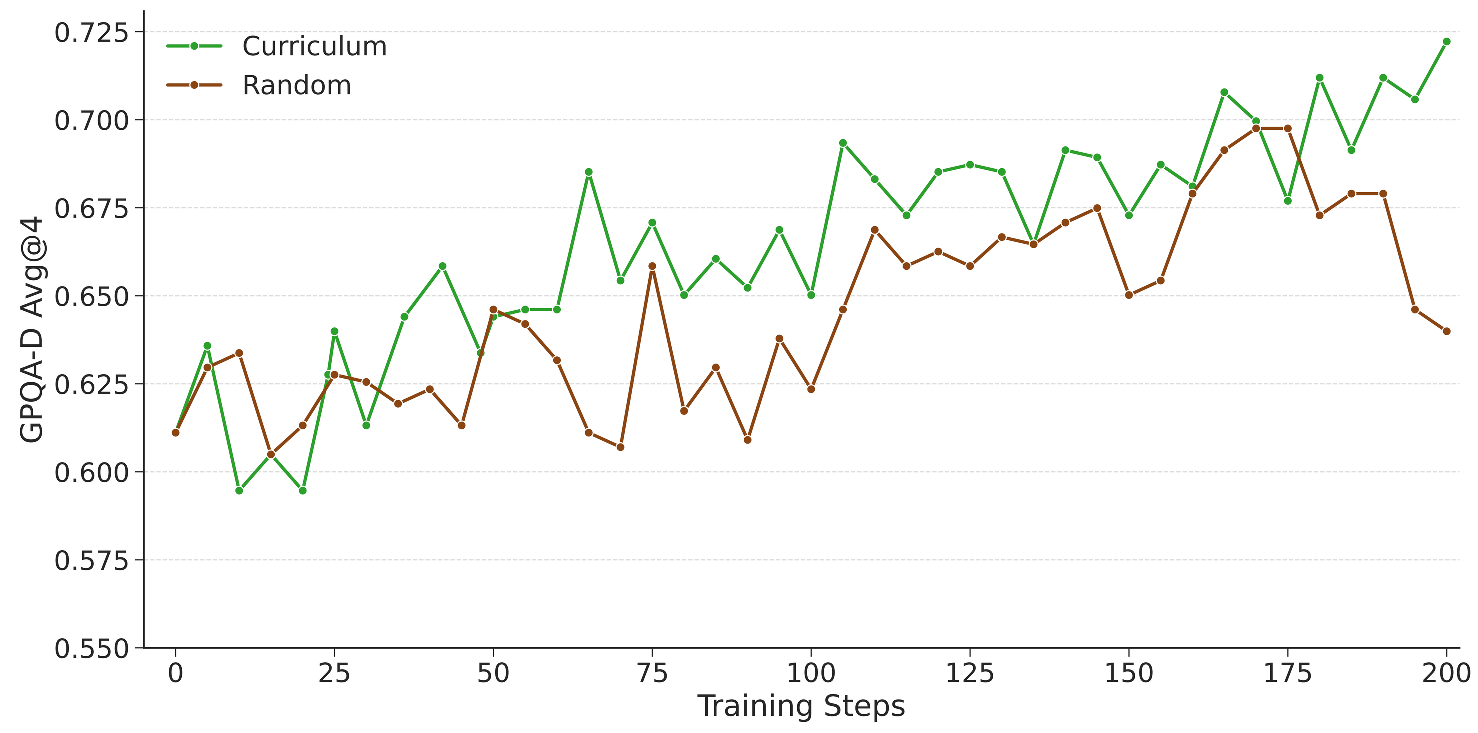
While SFT enables strong capabilities through teacher distillation, it limits performance to the teacher's level. Large-scale RL with verifiable rewards empowers LN-Ultra to explore beyond imitation and surpass the teacher.

Key RL Features:

- GRPO training for 140k H100 hours with FP8 inference for rollouts
- Prompt size of 72, 16 responses per prompt with $\tau = 1$ and $top\_p = 1$
- Global batch size of 576 with 2 gradient updates per rollout
- Accuracy rewards using Llama-3.3-70B-Instruct as judge
- Format rewards ensuring proper thinking tag usage

4. Curriculum Learning

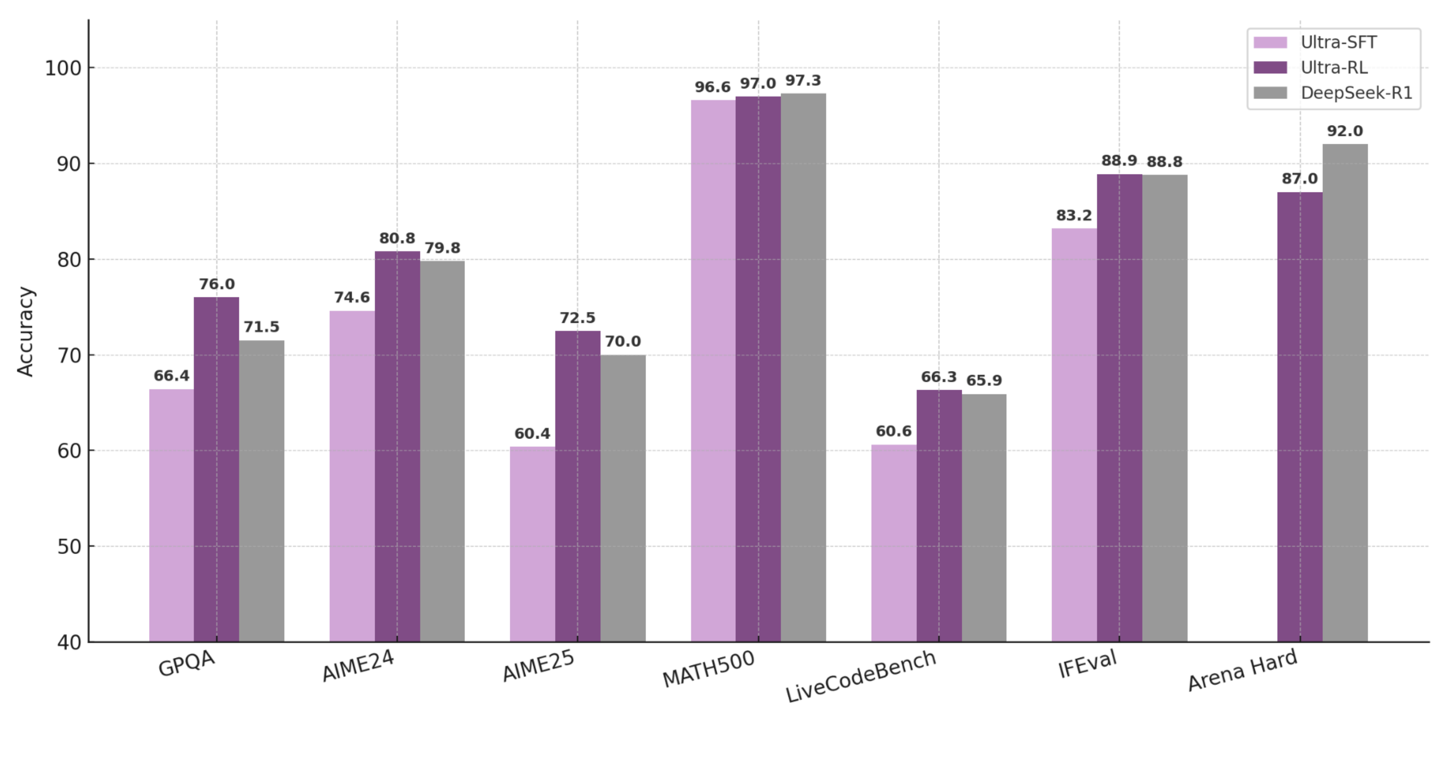
We implement an exploration-driven progressive batching strategy to systematically challenge LN-Ultra during RL training. Data is preprocessed by generating 8 responses per question using LN-Super, calculating pass rates, and discarding easy prompts (pass rate $\geq 75\%$).



Progressive Batching:

- Gaussian distribution targeting difficulty progression across batches
- Early batches: high pass rates (easier examples)
- Later batches: low pass rates (harder examples)
- Forces exploration beyond teacher capabilities

5. LN-Ultra Results



The RL stage is critical for surpassing teacher performance, particularly on GPQA where LN-Ultra achieves 76.0% vs DeepSeek-R1's 71.5%.

6. Links and Resources

HF Collection

Technical Report

NeMo-RL

Post-Training Dataset

HelpSteer3