

Fábricas de IA, a liderar o caminho Europeu da inovação

João Magalhães

jmag@fct.unl.pt

NOVA Laboratory for Informatics and Computer Science

Universidade NOVA de Lisboa



Parcello:



Apoios:





AMALIA LLM Team

Linguists

Data Engineers

Data Scientists

Proof-of-concept developers

Core work streams: Language modeling,
Instruction tuning, Preference Learning,
Safety, Multimodal



How is an LLM trained?

Pre-training

- Language modeling/Next word prediction
- Data collection and preparation
- Data quality filtering
- GDPR compliance
- Trillions of tokens
- Scaling laws

Instruction tuning

- Supervised Fine Tuning
- Data is (instruction, answer)
- Synthetic data generation
- MT translated data
- Data quality filtering
- Simulation of behaviors

Preference learning

- Reinforcement Learning
- Few samples of manual data
- Synthetic data generation
- Model generations



Language Modeling and Instruction Tuning

- @NOVA: Data preparation (months)
 - From noisy HTML and PDFs to raw high-quality texts
- @MN: Tokenization
 - Parse the data and generate tokens for training
 - Time consuming process that can occur before training
- @MN: Training
 - Language Modeling and Instruction tuning objectives
 - Distributed training frameworks
 - > 64 GPUs per simulation



Language Modeling and Instruction Tuning

- Gradient based optimization method
 - Inference or forward pass
 - Gradient or backward pass
 - Model update
- Inference with short samples, GPUs can hold the entire sample
- Inference with long text samples require multiple GPU inference
- Automatic sharing gradients and updates across GPUs and nodes



Tips

- Do maximum of pre-processing before reaching the GPU.
- Monitor GPU utilization:
 - mem + transf + cores
- Sample packing.
- Use multiple CPU cores to load data in advance.
- @MN: **do not use** gradient checkpointing
- @Deucalion: **use** gradient checkpointing (25% slower)





Reinforcement Learning

- Point Proximal Optimization
- Direct Preference Optimization
- Group Relative Policy Optimization



Chain-of-thought inference

Standard Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

Model Output

A: The answer is 27. ❌

Chain-of-Thought Prompting

Model Input

Q: Roger has 5 tennis balls. He buys 2 more cans of tennis balls. Each can has 3 tennis balls. How many tennis balls does he have now?

A: Roger started with 5 balls. 2 cans of 3 tennis balls each is 6 tennis balls. $5 + 6 = 11$. The answer is 11.

Q: The cafeteria had 23 apples. If they used 20 to make lunch and bought 6 more, how many apples do they have?

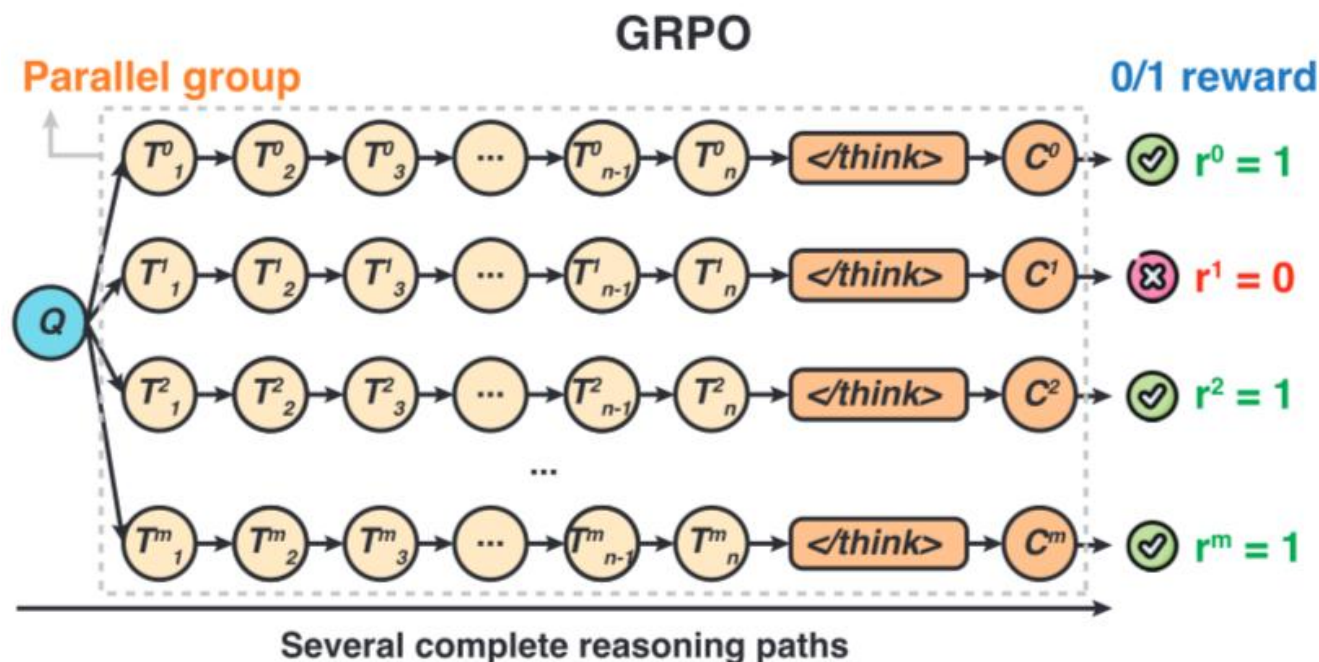
Model Output

A: The cafeteria had 23 apples originally. They used 20 to make lunch. So they had $23 - 20 = 3$. They bought 6 more apples, so they have $3 + 6 = 9$. The answer is 9. ✅



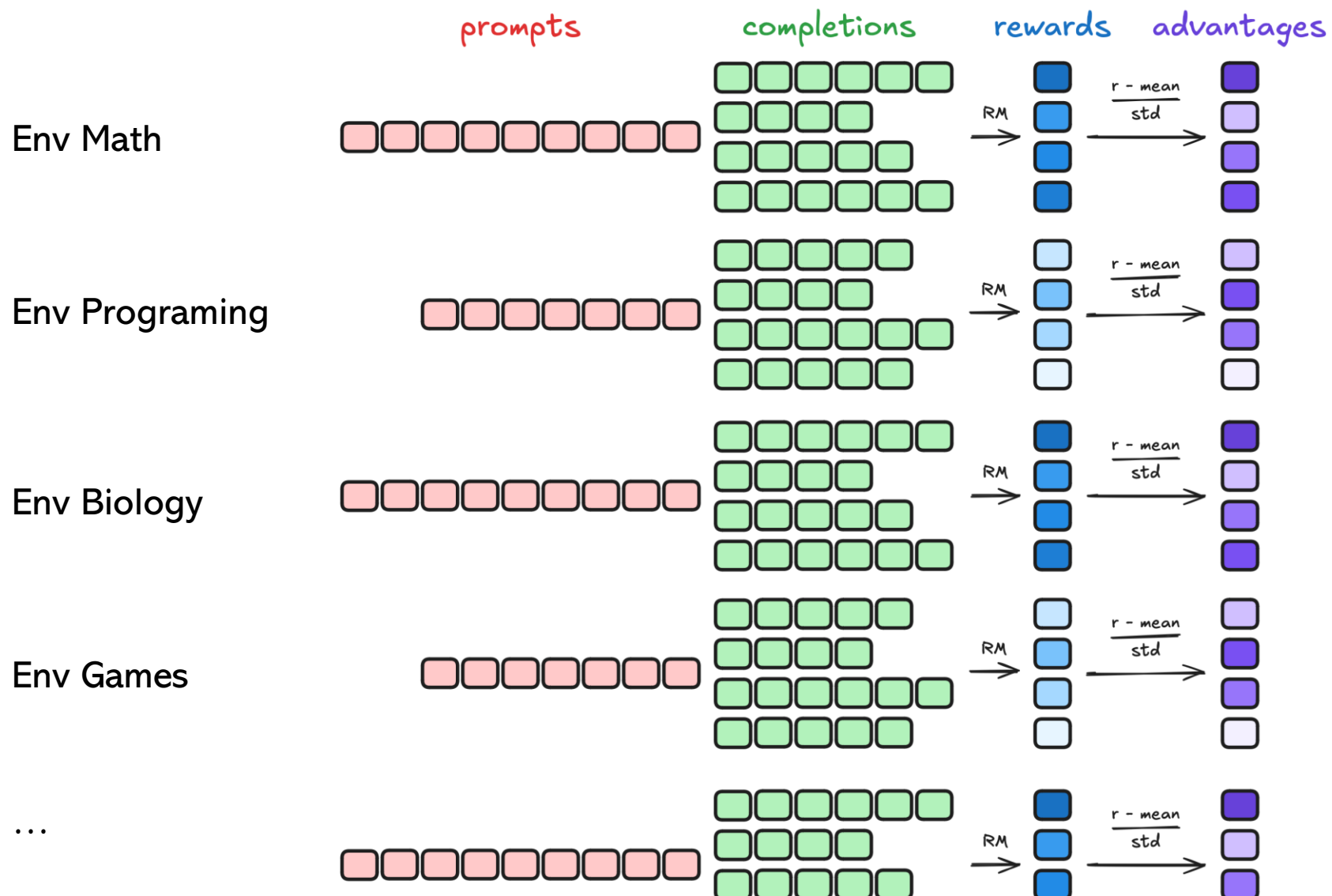
Group Relative Policy Optimization

- Verifiable Rewards
- Multiple reasoning paths





Verifiable Rewards





RLVR (GRPO) @ MN5

- Inference nodes (1-2)
 - LLM loaded
 - LLM is distributed to support long samples and improve throughput
 - Generate samples on demand
 - Update the LLM
- Training nodes (16)
 - Collect samples
 - Run the Verifiable Rewards
 - Update model
- Orchestrator node (@Inf. Node)



Environments' configuration is time consuming because each VR can have its own environment.

We can have more than 30 environments on one simulation.

SLURM script + python script.

Fábricas de IA, a liderar o caminho Europeu da inovação

Thank you!

João Magalhães

jmag@fct.unl.pt

Multimodal Systems @ NOVA LINCS

Universidade NOVA de Lisboa



universidade de aveiro
theoria poiesis praxis

Parcello:



Dias de Sousa
INSTRUMENTAÇÃO ANALÍTICA E CIENTÍFICA, S.A.

Apoios:



EUROCC PORTUGAL



EuroHPC
Joint Undertaking



PRR
Plano de Recuperação e Resiliência



REPÚBLICA
PORTUGUESA



Financiado pela
União Europeia
NextGenerationEU



ML Distributed Experiments

- 2004-2008 Condor (Imperial)
- 2012-2020 Condor (NOVA) < 2017 Rocks distro
- 2020-now SLURM (NOVA)
- 2024-now Deucalion
- 2024-now Marenostrom 5

