Building Open LLMs

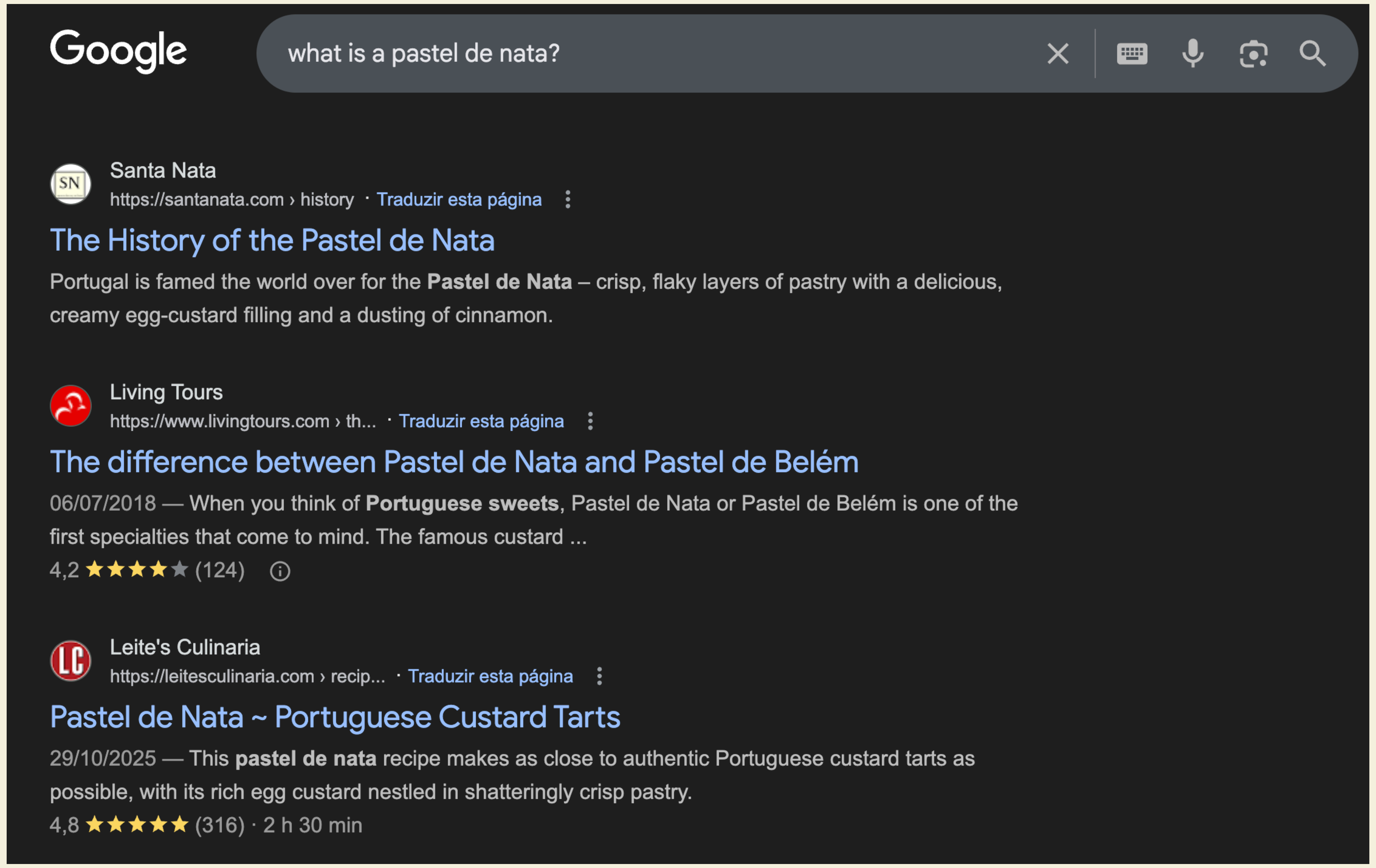
Or how to build ChatGPT at home?

Duarte Alves, PhD Student @ SARDINE Lab, University of Lisbon - 2025/12/02

If I asked you "what is a pastel de nata?" What would you do?

If I asked you “what is a pastel de nata?” What would you do?

3 years ago, you would probably Google it…



If I asked you “what is a pastel de nata?” What would you do?

3 years ago, you would probably Google it…

Google

what is a pastel de nata?

Santa Nata
https://santanata.com › history · Traduzir esta página

The History of the Pastel de Nata

Portugal is famed the world over for the **Pastel de Nata** – crisp, flaky layers of pastry with a delicious, creamy egg-custard filling and a dusting of cinnamon.

Living Tours
https://www.livingtours.com › th... · Traduzir esta página

The difference between Pastel de Nata and Pastel de Belém

06/07/2018 — When you think of **Portuguese sweets**, Pastel de Nata or Pastel de Belém is one of the first specialties that come to mind. The famous custard ...
4,2 ★★★★★ (124)

Leite's Culinaria
https://leitesculinaria.com › recip... · Traduzir esta página

Pastel de Nata ~ Portuguese Custard Tarts

29/10/2025 — This **pastel de nata** recipe makes as close to authentic Portuguese custard tarts as possible, with its rich egg custard nestled in shatteringly crisp pastry.
4,8 ★★★★★ (316) · 2 h 30 min

Today, maybe you would ask ChatGPT

What is a pastel de nata?

A **pastel de nata** (plural: *pastéis de nata*) is a traditional **Portuguese custard tart**. It consists of a **crispy, flaky pastry shell** filled with a rich, creamy **egg custard**, typically flavored with **vanilla**, **lemon zest**, and sometimes a hint of **cinnamon**. The top is usually caramelized and slightly blistered from baking at a high temperature — similar to a crème brûlée finish.

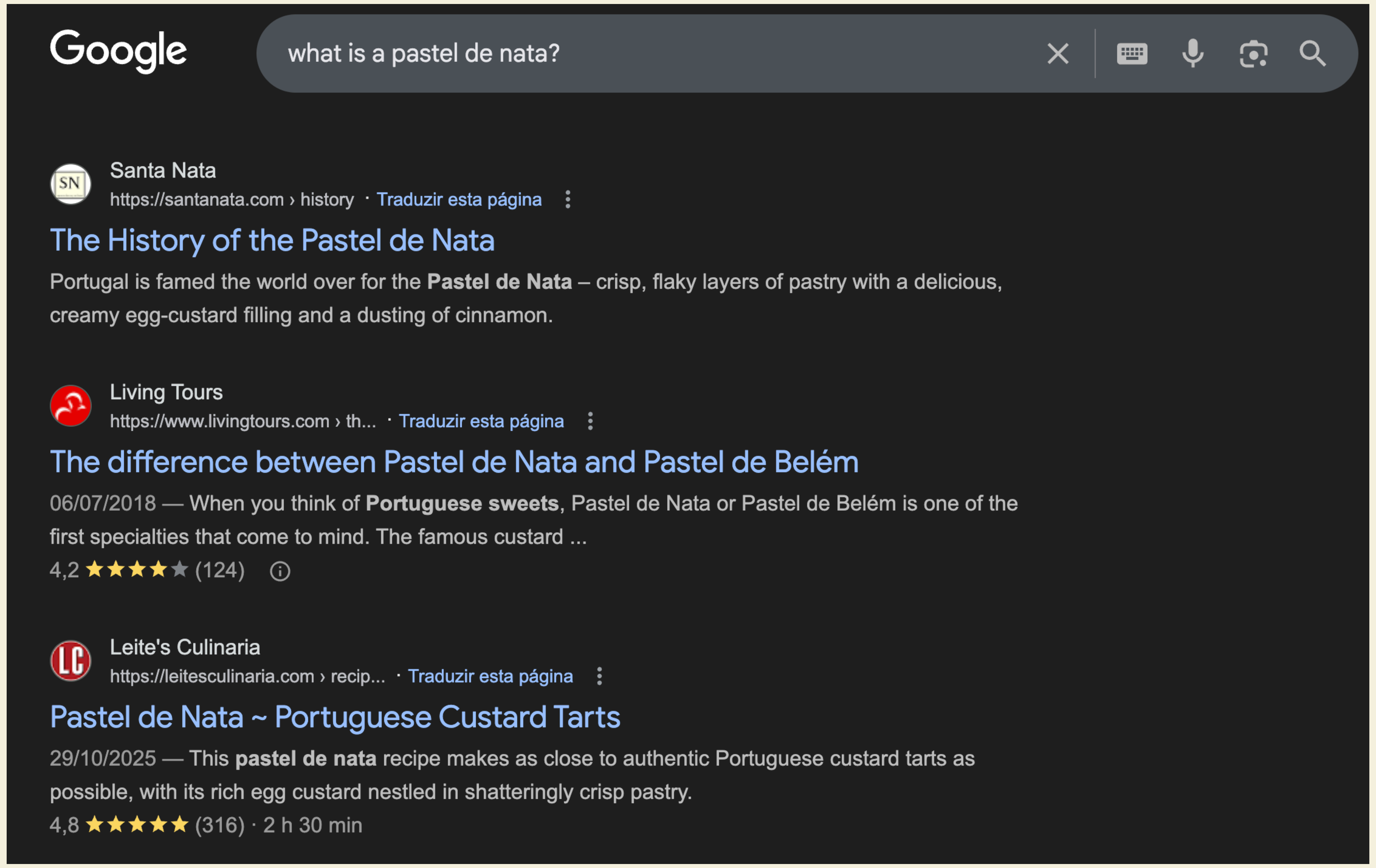
They’re most famously associated with the **Belém district of Lisbon**, where the original recipe — known as **Pastéis de Belém** — has been made since the early 1800s by monks from the Jerónimos Monastery.

Pastéis de nata are usually eaten warm, often dusted with **cinnamon and powdered sugar**, and enjoyed with **coffee (bica)**.

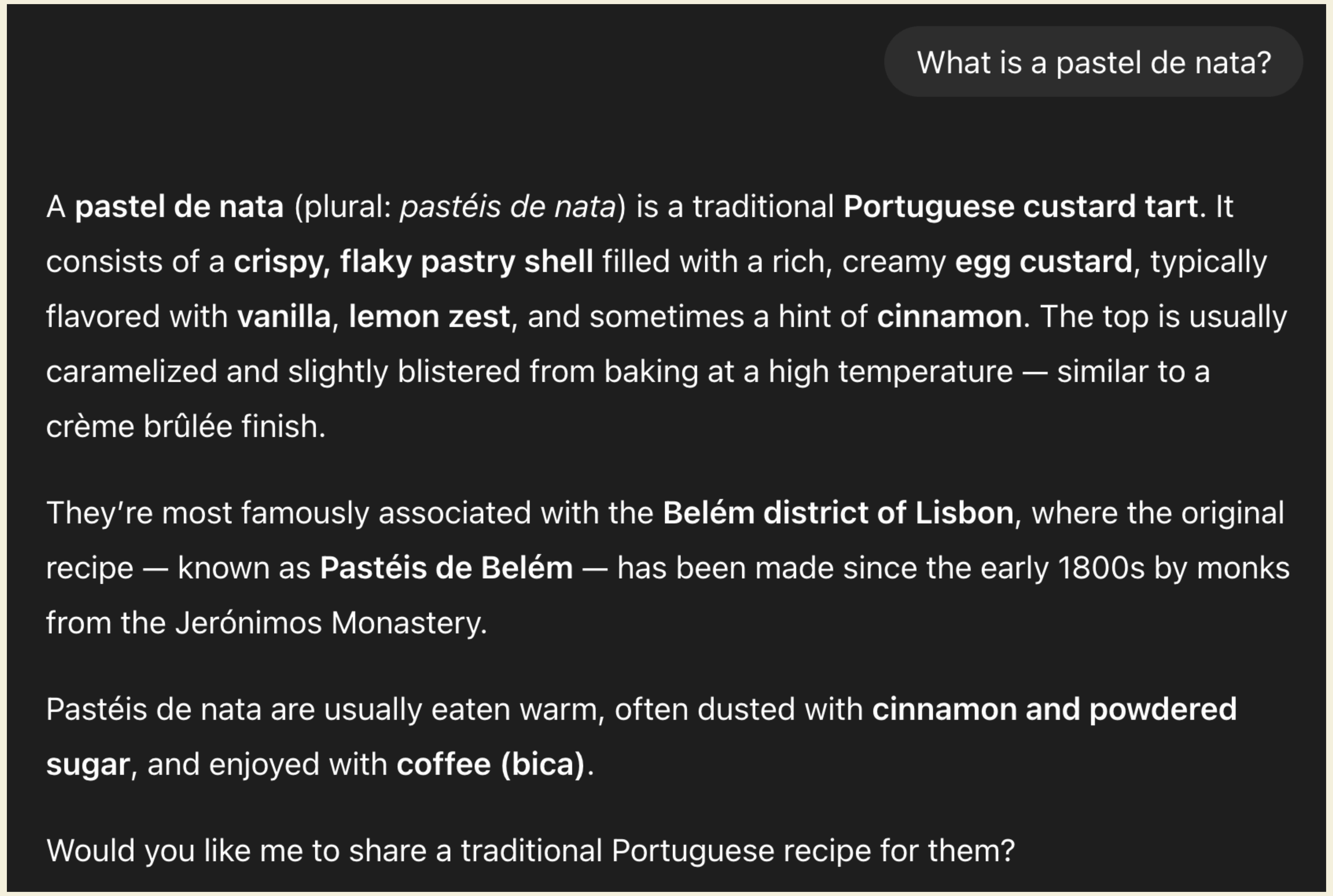
Would you like me to share a traditional Portuguese recipe for them?

If I asked you "what is a pastel de nata?" What would you do?

3 years ago, you would probably Google it…



Today, maybe you would ask ChatGPT



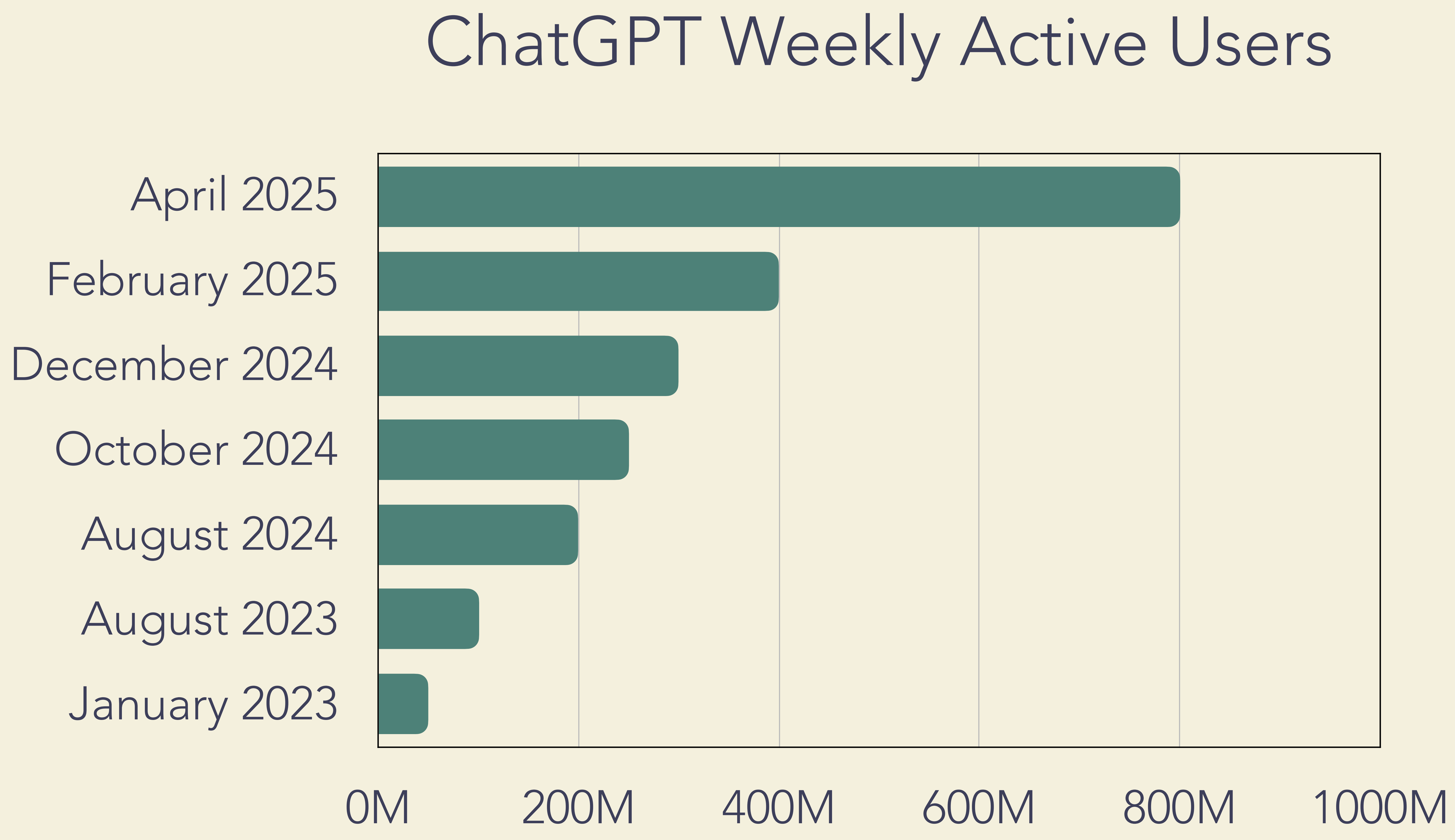
This seems much nicer….

What is a pastel de nata?

What is a pastel de nata?

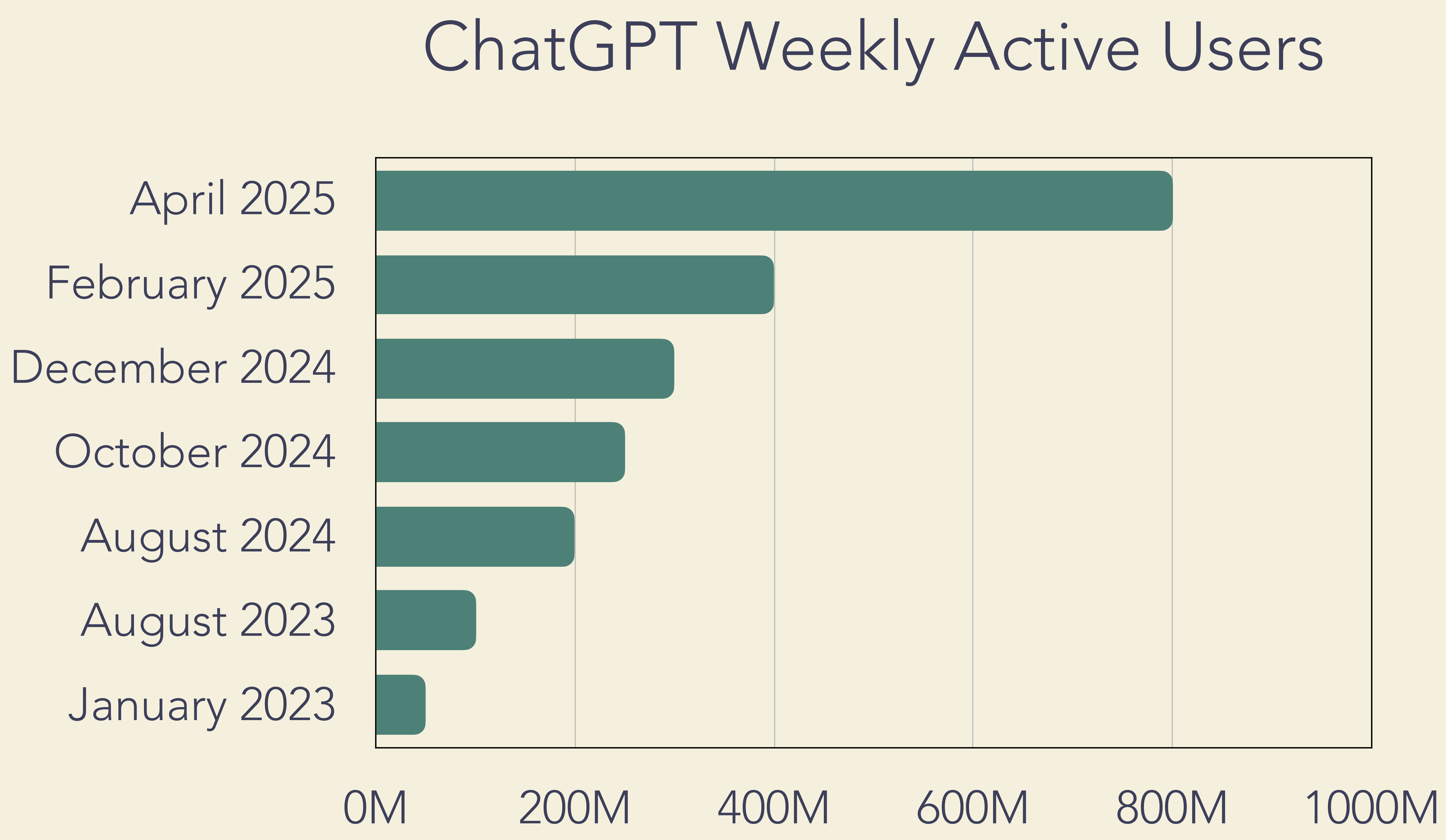
Find out in Portugal 🙂

Going back to ChatGPT



https://www.demandsage.com/chatgpt-statistics/

Going back to ChatGPT

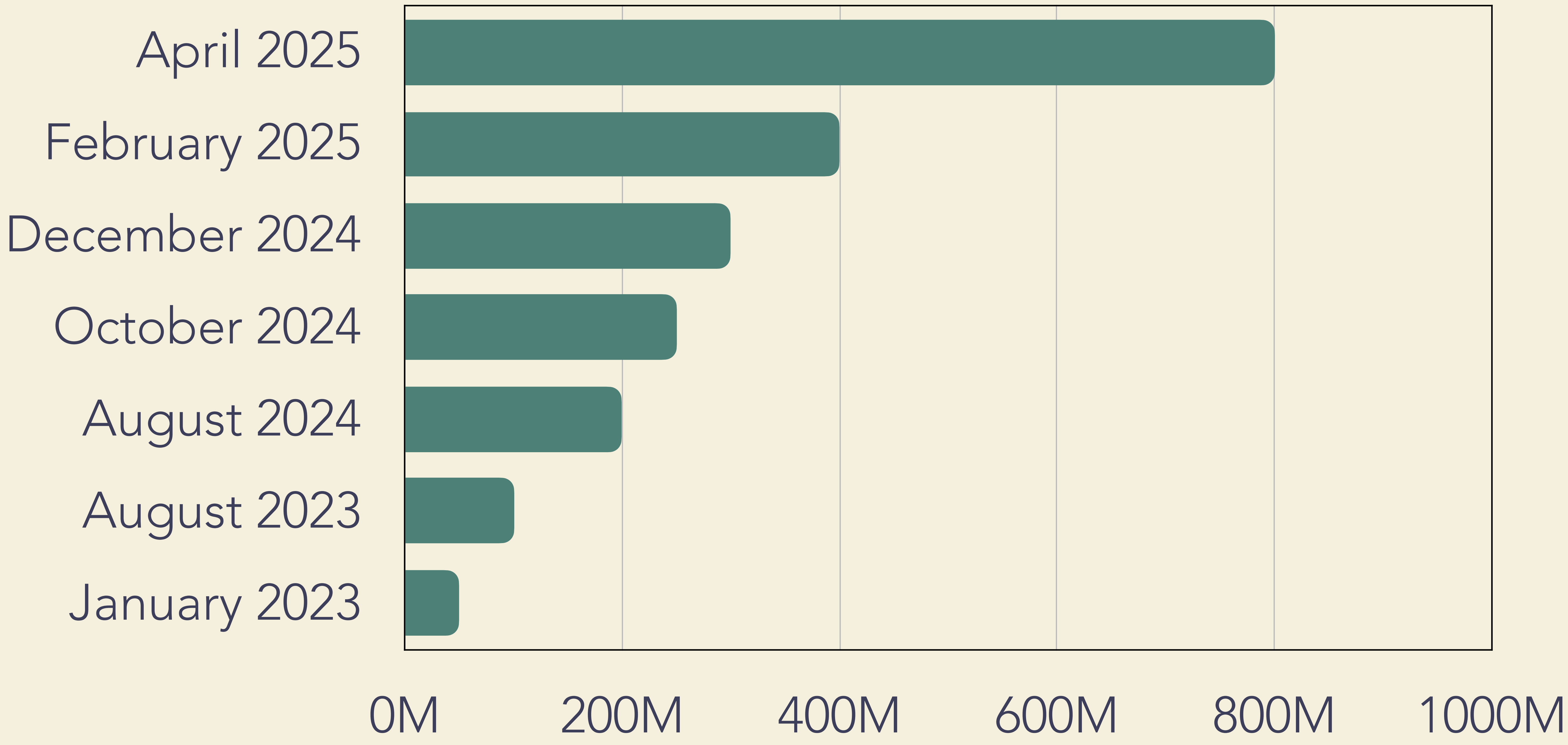


But what is really behind ChatGPT?

https://www.demandsage.com/chatgpt-statistics/

Going back to ChatGPT

ChatGPT Weekly Active Users



But what is really behind ChatGPT?

We don’t know…

https://www.demandsage.com/chatgpt-statistics/

Why Open Models?

Democratization

Models should be widely available

Research

Investigation requires full knowledge of the model

Transparency

Openness enables auditing and fosters user trust

Collaboration

Inventions are built upon collective work

Privacy

Sensitive data requires on-site deployments

Customization

Access to weights enables fine-grained customization

Where do we start?

A tale of two worlds

Pre-training

Knowledge

- Self-supervised training on documents from many sources;
- Acquire general knowledge about many domains;

Post-training

Skills & Capabilities

- Supervised fine-tuning and reinforcement learning on user instructions;
- Tune capabilities like instruction following, tool usage, or thinking effort.

Pre-training

What do we optimize?

Causal Language Modeling

Causal Language Modeling

Input	Prediction
The	p(cat | The)
cat	p(sat | The cat)
sat	p(on | The cat sat)
on	p(the | The cat sat on)
the	p(mat | The cat sat on the)
mat	p(. | The cat sat on the mat)
.	p(EOS | The cat sat on the mat .)

Predict the next token given the previous context

How do we collect the data?

Collecting web data at scale

URL Filtering
Select relevant links to download

Text Extraction
Convert html pages to plain text

Language Filtering
Group documents by language

Heuristic Filtering
Apply rule-based filters to filter documents

Deduplication
Remove duplicate or near-duplicate content

Anonymization
Redact personally identifiable information

Model Filtering
Run models to detect high-quality content

Heuristic vs Model-based Filters

Heuristic

Fast rules

- Transparent and easy to explain;
- Cheap to run at scale;
- Hard to capture “fuzzy” concepts like quality.

Model-based

Learned filters

- Capture more subtle signals (style, coherence, etc);
- Reduce manual engineering;
- Hard to inspect, and expensive to run at scale.

What if we need more data?

Generating Synthetic Data at Scale

Seed Corpus
Diverse collection of documents.

Rephrasing
Rewrite the document, possibly improving it.

Question-Answer
Add question-answer pairs to the document.

Summarize
Extract the relevant knowledge in the text.

- Overcome the finite nature of the web;
- Repurpose low-quality data as high-quality documents;
- More control over the data distribution;
- Improves coverage of *rare* phenomena.

Data Curriculums

"We made (…) adjustments to the pre-training data mix during training to improve model performance (…)"

"on the final 40M tokens, (…) we also adjusted the data mix to upsample data sources of very high quality"

Llama Team, AI @ Meta. The Llama 3 Herd of Models. 2024.

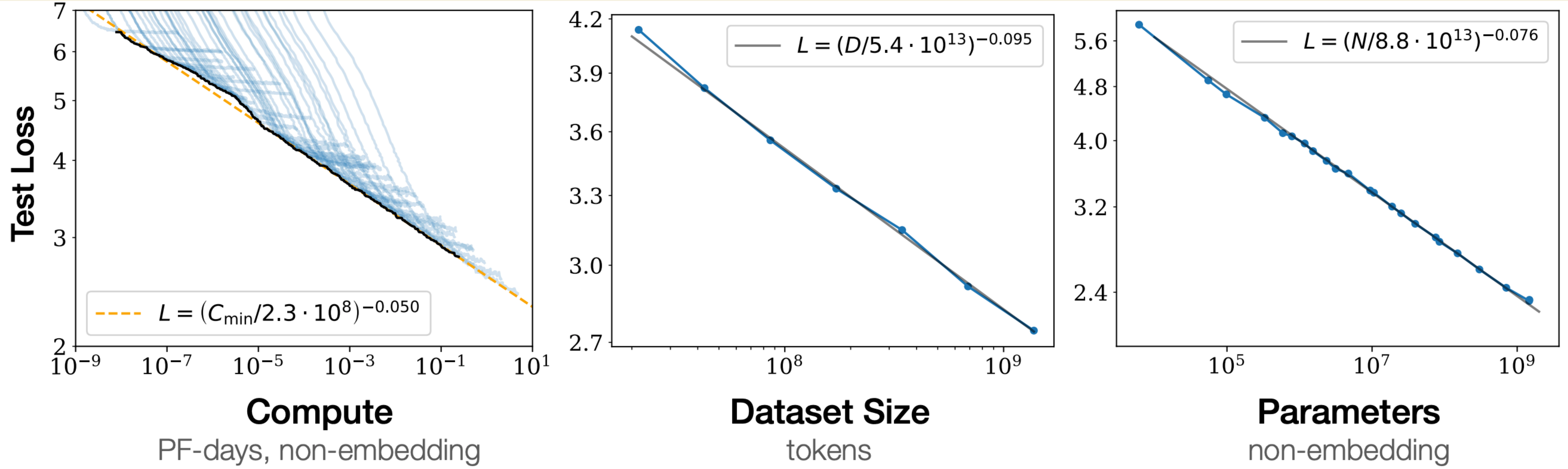
A Versatile Learning Rate Scheduler

- Pre-training at a constant learning rate;
- Annealing on high quality data;
- Doesn't require knowing the number of tokens a-priori;
- You can resume/extend/change pre-training at any point.

Hu et al. MiniCPM: Unveiling the Potential of Small Language Models with Scalable Training Strategies. 2024.

How can we experiment?

Scaling laws



Language models follow scaling laws on the compute budget, dataset size and model size

You can predict the loss of larger models from smaller ones

Kaplan et al. Scaling Laws for Neural Language Models. 2020

Compute-optimal models

For a given compute budget, what is the optimal dataset and model size you should use?

Chinchilla optimal: dataset size should be around 20x model size

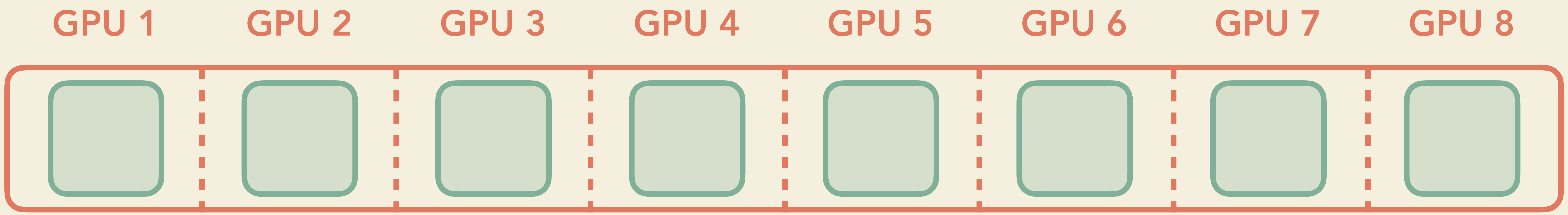
In practice: overtrain models to facilitate inference

Hoffmann et al. Training Compute-Optimal Large Language Models. 2022.

How to train at large scale?

Data Parallelism: Replicating your model

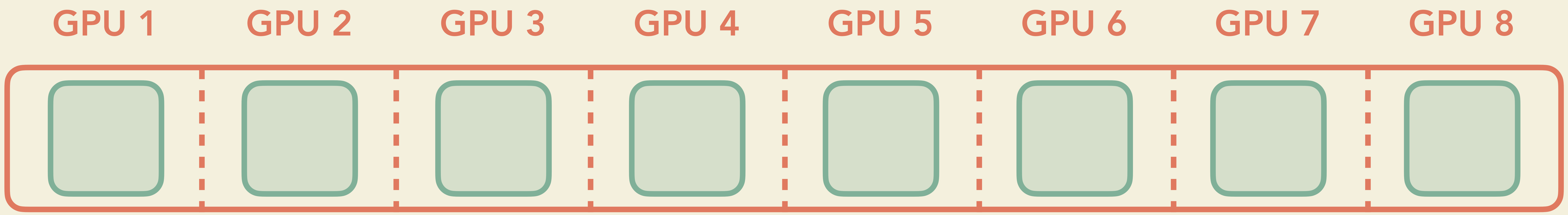
Data Parallelism: Replicating your model

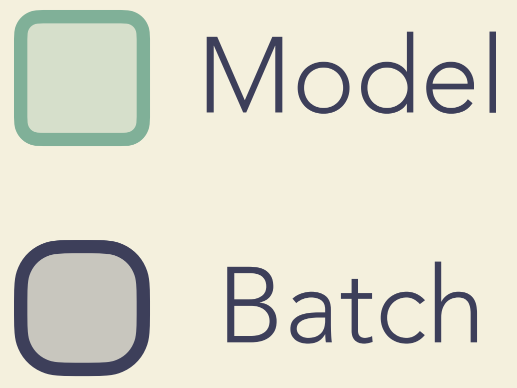


Model

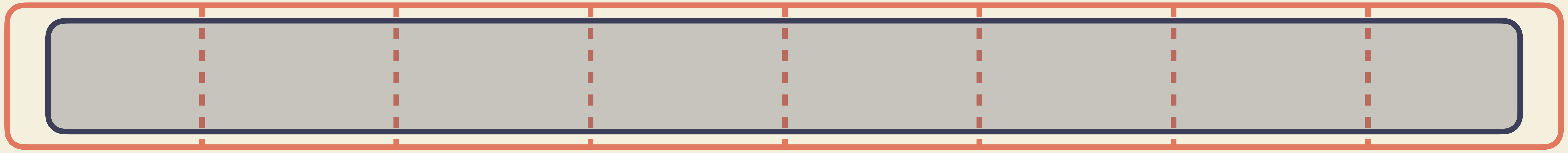
Have a model copy on each GPU

Data Parallelism: Replicating your model



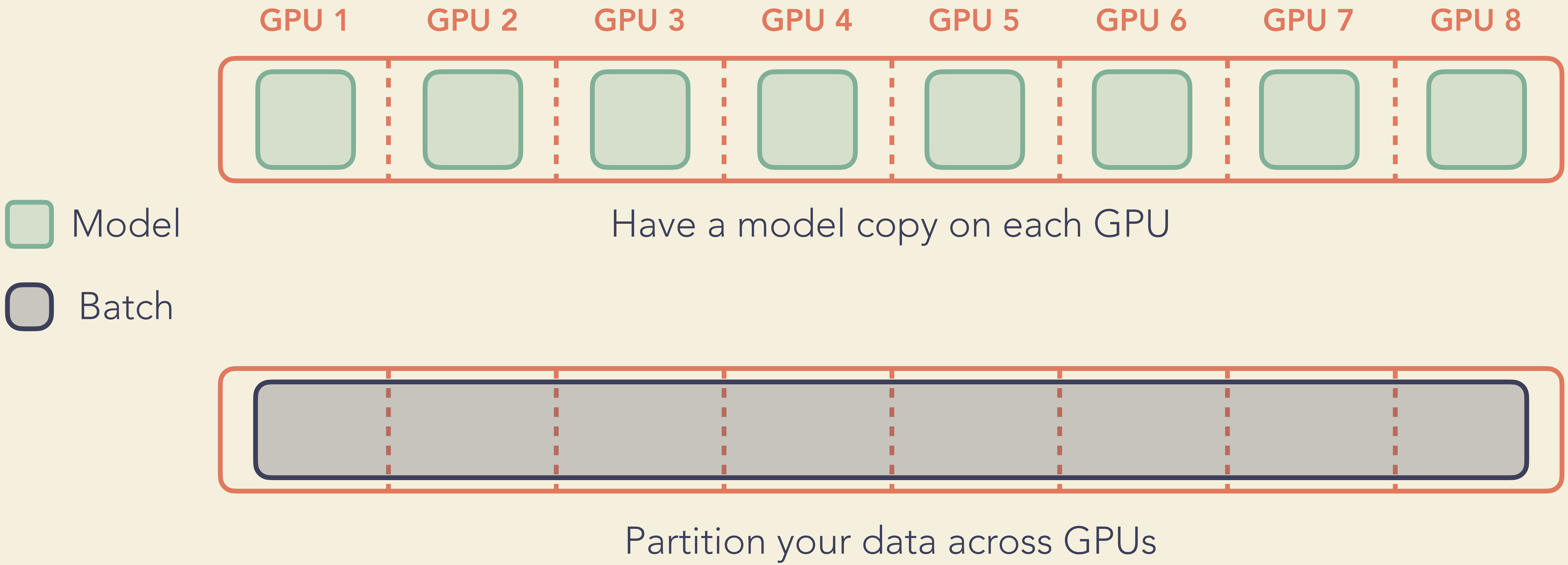


Have a model copy on each GPU



Partition your data across GPUs

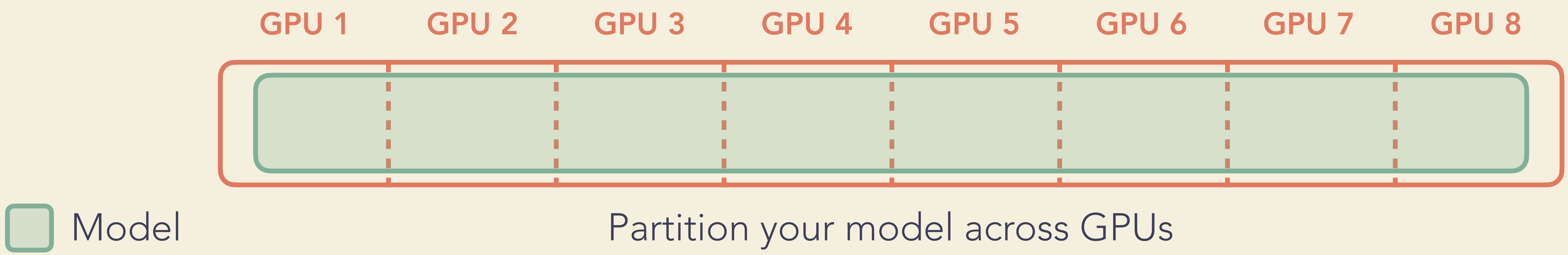
Data Parallelism: Replicating your model



You can train on much larger batch sizes

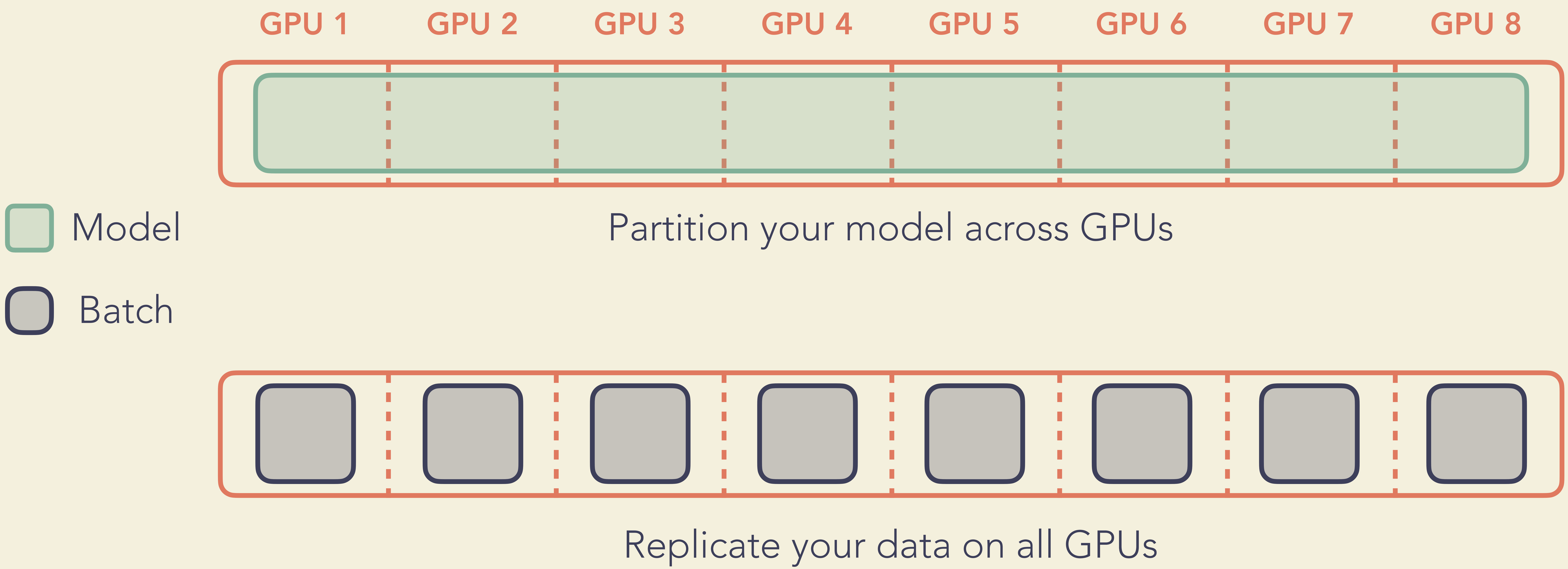
Model Parallelism: Sharding your model

Model Parallelism: Sharding your model

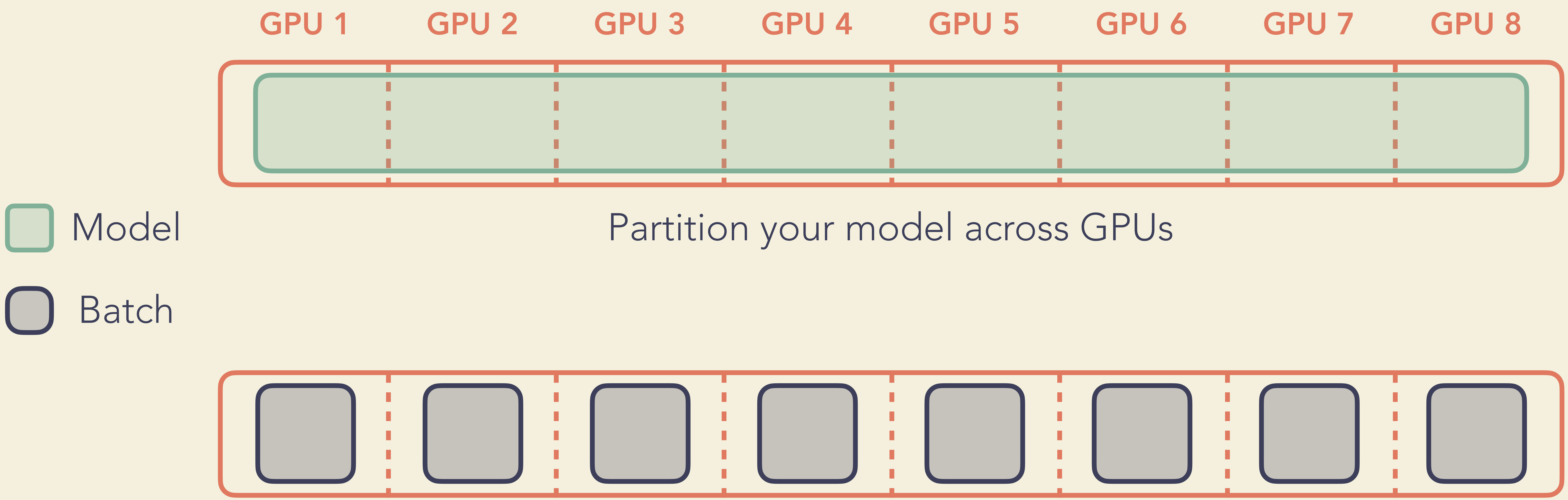


Partition your model across GPUs

Model Parallelism: Sharding your model



Model Parallelism: Sharding your model

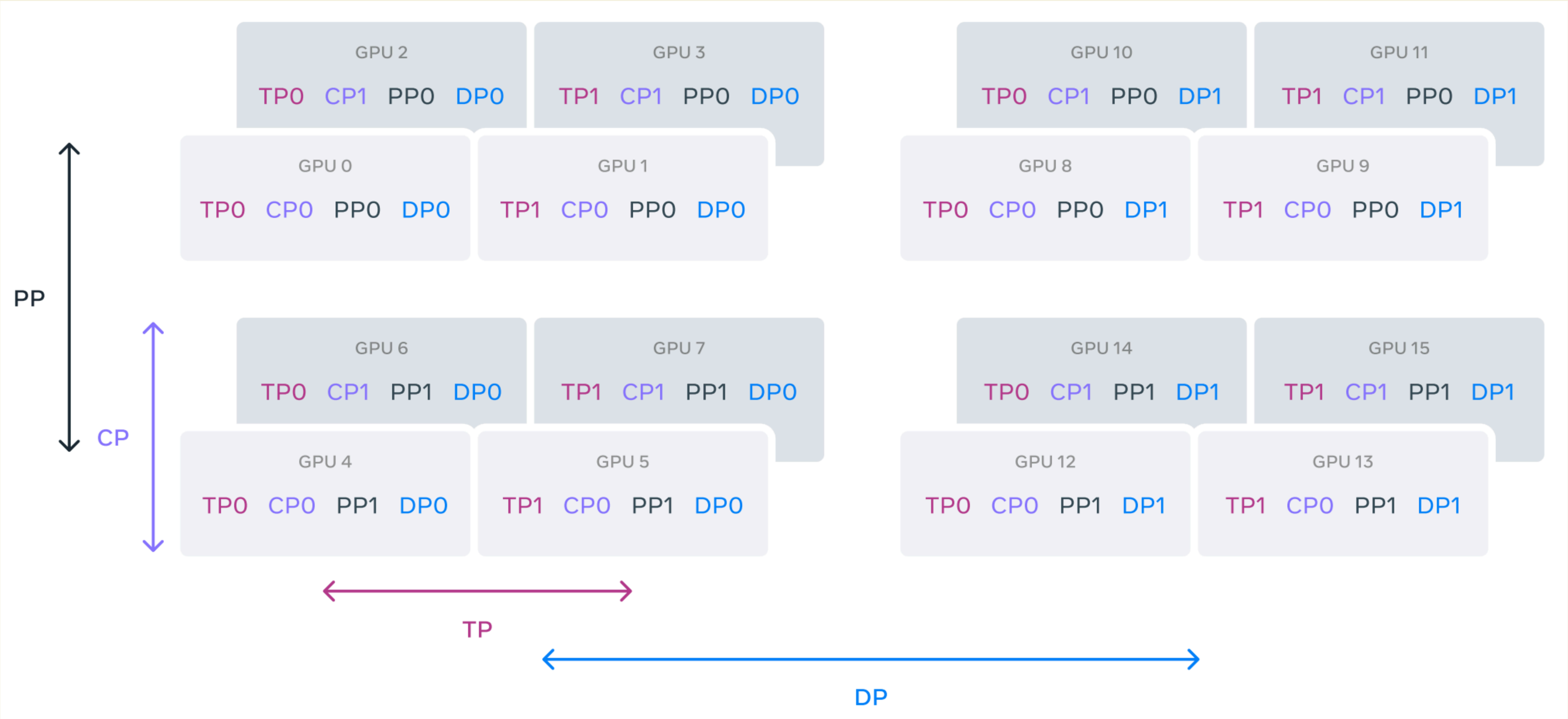


Partition your model across GPUs

Replicate your data on all GPUs

You can train much larger models

4D Parallelism: Combining everything



- **Data Parallel:** Split your batch across GPUs.
- **Tensor Parallel:** Split individual layers of the model.
- **Pipeline Parallel:** Split across layers of the model.
- **Context Parallel:** Split training sequences (for long context).

Llama Team, AI @ Meta. The Llama 3 Herd of Models. 2024.

The hardware lottery

"we find **great unreliability** when it comes to GPU nodes which often fail due to **hardware errors** or **connection issues**."

- At large scales, having a gpu failure becomes very common;
- Having good checkpoint logic is essential to resume training;
- Automatic monitoring and resuming from failures is becoming very relevant.

Reka Team. Reka Core, Flash, and Edge: A Series of Powerful Multimodal Language Models. 2024.

How is the model after pre-training?

How is the model after pre-training?

The sky cracked open just as the first train arrived. *The pressure from the back of the train, the building with its massive windows, and the cabin walls were all falling open. Inside the cabin was a table, with a huge sign, and a table cloth. Some of the walls had been covered with sharp edges, and the railing was partly covered with hair. (…)*

Pre-trained models are powerful for auto-completion tasks

How is the model after pre-training?

Instruction: Sort the following list of numbers in ascending order and answer ONLY with the sorted list.

List: 5, 2, 9, 1

Answer: *9, 2, 9, 1*

Number of strings to make: 1

Number of numbers to count: 9

(...)

They are not necessarily good at following user instructions

How is the model after pre-training?

Review: 'I loved this movie, it was fantastic!' Sentiment: POSITIVE

Review: 'This was the worst film I have ever seen.' Sentiment: NEGATIVE

Review: 'The plot was boring and the acting was terrible.' Sentiment: NEGATIVE

Review: 'A good ending.' Sentiment: NEGATIVE

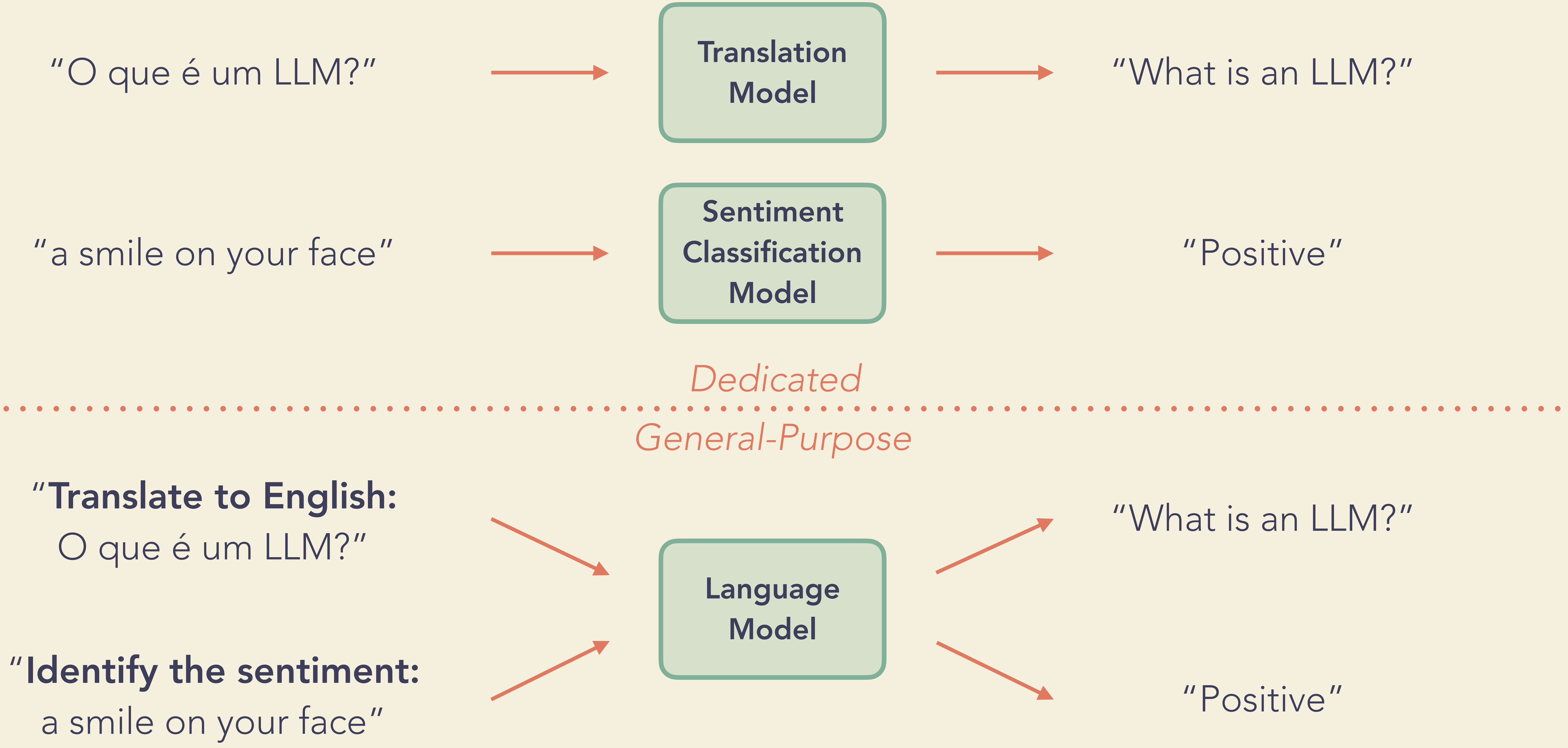
(…)

But they can solve tasks through in-context learning!

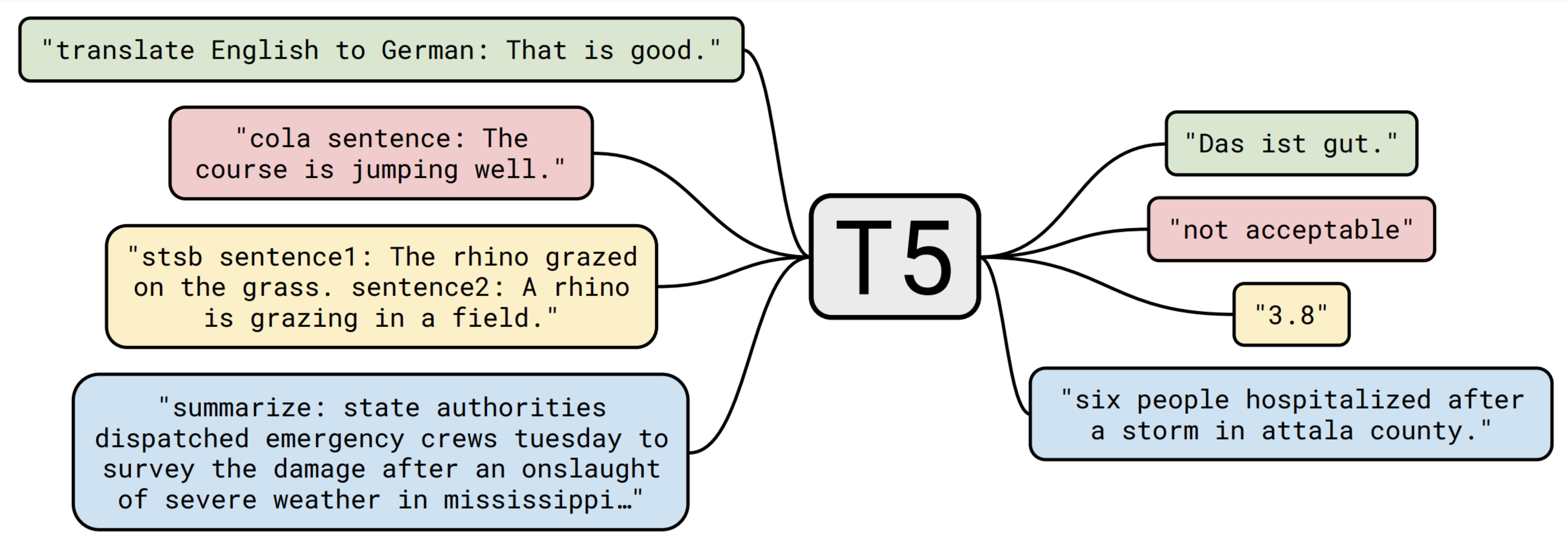
Post-training

How can a model perform many tasks?

Framing tasks as text problems



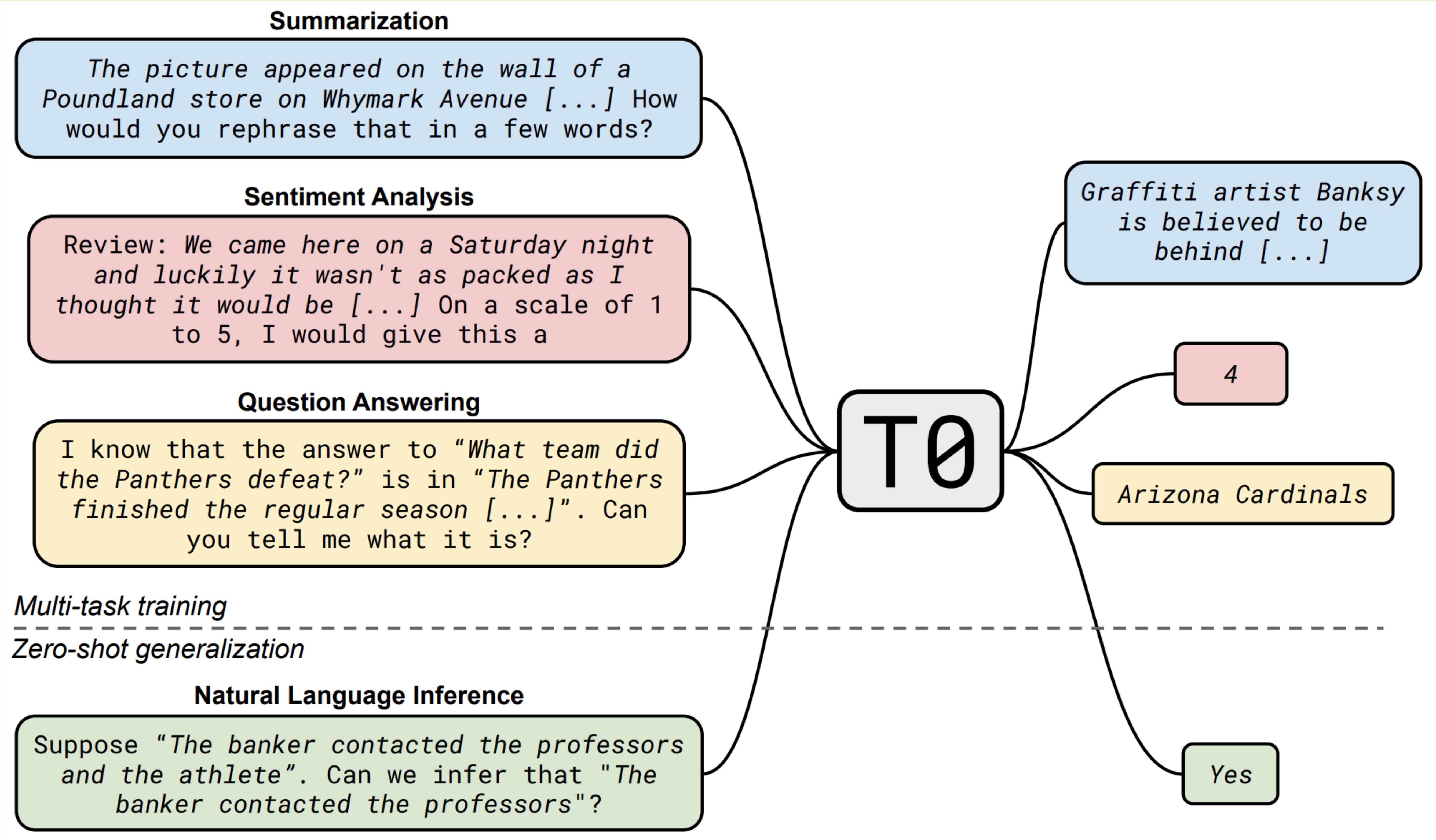
Framing tasks as text problems



Fine-tuned languages models can solve multiple tasks by **describing the task in the input**.

Raffel et al. Exploring the Limits of Transfer Learning with a Unified Text-to-Text Transformer. 2020

Framing tasks as text problems



Training on multiple tasks enables solving **previously unseen tasks**.

Sanh et al. Multitask Prompted Training Enables Zero-Shot Task Generalization. 2022

Aligning with human feedback

More than solving many tasks,
we want models that **follow user intent**

Ouyang et al. Training language models to follow instructions with human feedback. 2020

How to format a conversation?

Formatting a conversation as a text

Formatting a conversation as a text

User
What is a pastel de nata?

Assistant
A pastel de nata is a Portuguese …

User
Where can I try them?

Assistant
You can find them throughout …

Formatting a conversation as a text

User
What is a pastel de nata?

Assistant
A pastel de nata is a Portuguese …

User
Where can I try them?

Assistant
You can find them throughout …

→

<|im_start|>user
What is a pastel de nata?<|im_end|>
<|im_start|>assistant
A pastel de nata is a Portuguese…<|im_end|>
<|im_start|>user
Where can I try them?<|im_end|>
<|im_start|>assistant
You can find them throughout…<|im_end|>

Formatting a conversation as a text

User
What is a pastel de nata?

Assistant
A pastel de nata is a Portuguese …

User
Where can I try them?

Assistant
You can find them throughout …

Roles

<|im_start|>user
What is a pastel de nata?<|im_end|>
<|im_start|>assistant
A pastel de nata is a Portuguese…<|im_end|>
<|im_start|>user
Where can I try them?<|im_end|>
<|im_start|>assistant
You can find them throughout…<|im_end|>

Formatting a conversation as a text

User
What is a pastel de nata?

Assistant
A pastel de nata is a Portuguese …

User
Where can I try them?

Assistant
You can find them throughout …

Roles

```
<|im_start|>user
What is a pastel de nata?<|im_end|>
<|im_start|>assistant
A pastel de nata is a Portuguese…<|im_end|>
<|im_start|>user
Where can I try them?<|im_end|>
<|im_start|>assistant
You can find them throughout…<|im_end|>
```

Special delimiter tokens

Formatting a conversation as a text

User
What is a pastel de nata?

Assistant
A pastel de nata is a Portuguese …

User
Where can I try them?

Assistant
You can find them throughout …

Roles

```
<|im_start|>user
What is a pastel de nata?<|im_end|>
<|im_start|>assistant
A pastel de nata is a Portuguese…<|im_end|>
<|im_start|>user
Where can I try them?<|im_end|>
<|im_start|>assistant
You can find them throughout…<|im_end|>
```

Special delimiter tokens

The chat template is the model "interface," you can specify special directives, tools, etc.

How to go about post-training?

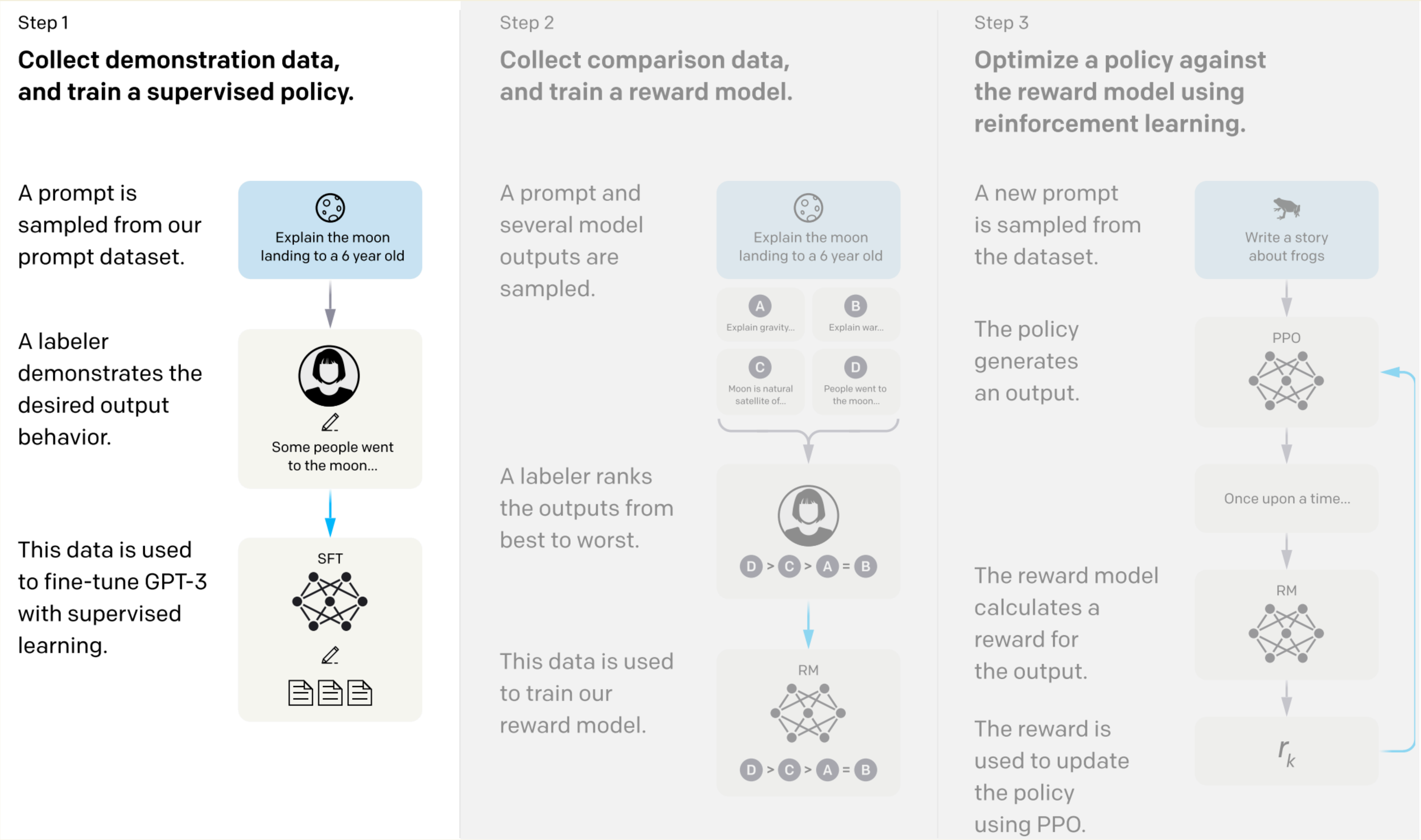
A three step process

Step 1: Supervised fine-tuning
Collect demonstration data and train a supervised policy

Step 2: Train a reward model
Collect comparison data, and train a reward model.

Step 3: Optimize your policy
Optimize a policy against the reward model using reinforcement learning.

Step 1: Supervised fine-tuning



Ouyang et al. Training language models to follow instructions with human feedback. 2020

Quality and Diversity: The guiding principles

"We build TowerBlocks prioritizing data diversity and quality."

TowerLLM Paper

"We construct a large-scale instruction-tuning dataset spanning diverse domains, guided by two core principles: maximizing prompt diversity and ensuring high response quality."

Kimi K2 Technical Report

Alves et al. Tower: An Open Multilingual Large Language Model for Translation-Related Tasks. 2024

Kimi Team. Kimi K2: Open Agentic Intelligence. 2025

Collecting instructions

Human-written

Ask human annotators

- High quality;
- Closest to deployment scenarios;
- Expensive, in particular for specialized tasks

Templated tasks

Repurpose datasets

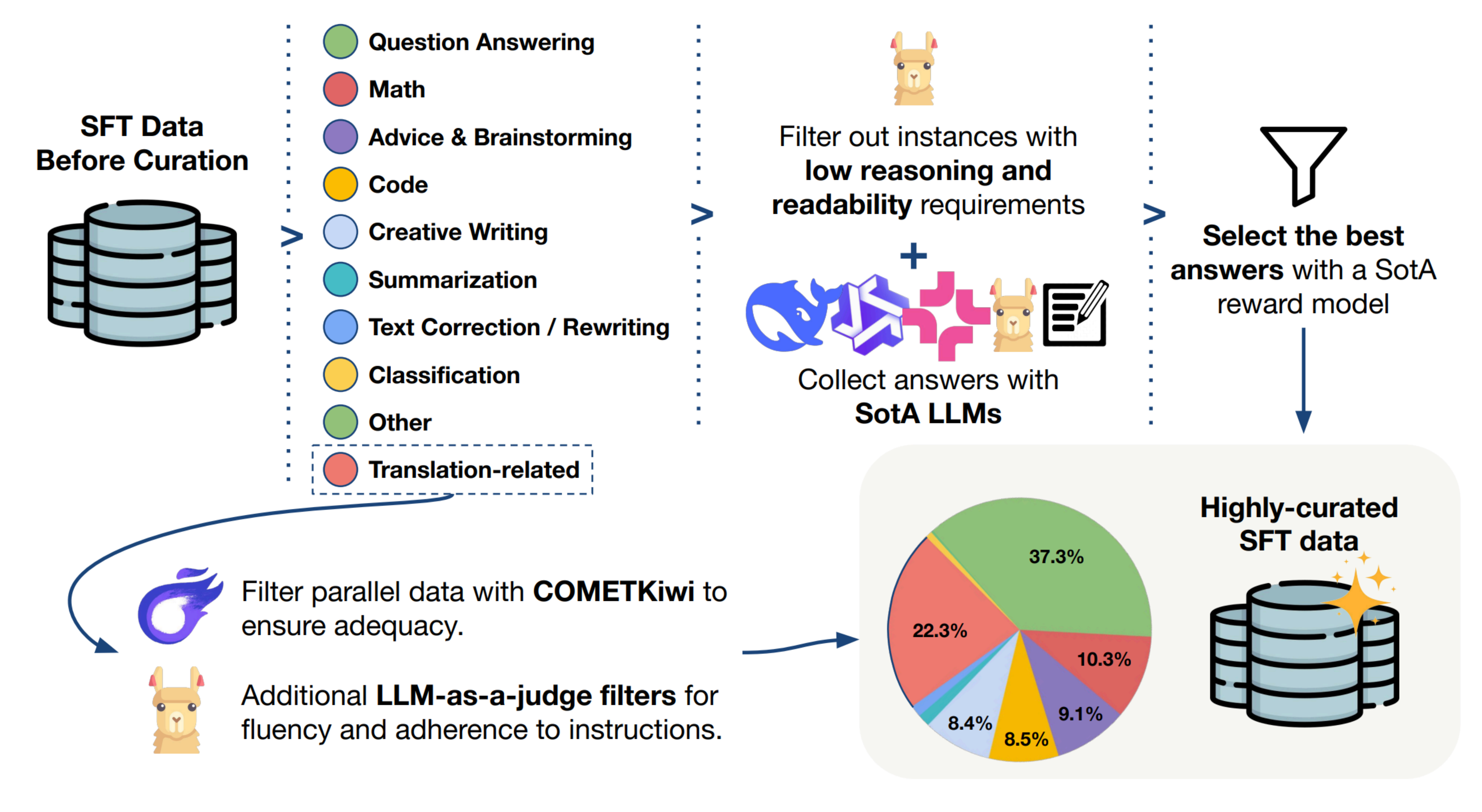
- Very cheap and easy;
- Best for targeting a particular task;
- Hard to scale to many tasks and templates.

Synthetic

Generate with models

- Cheap and scalable;
- Easy to address specific scenarios;
- Models tend to not diversify much.

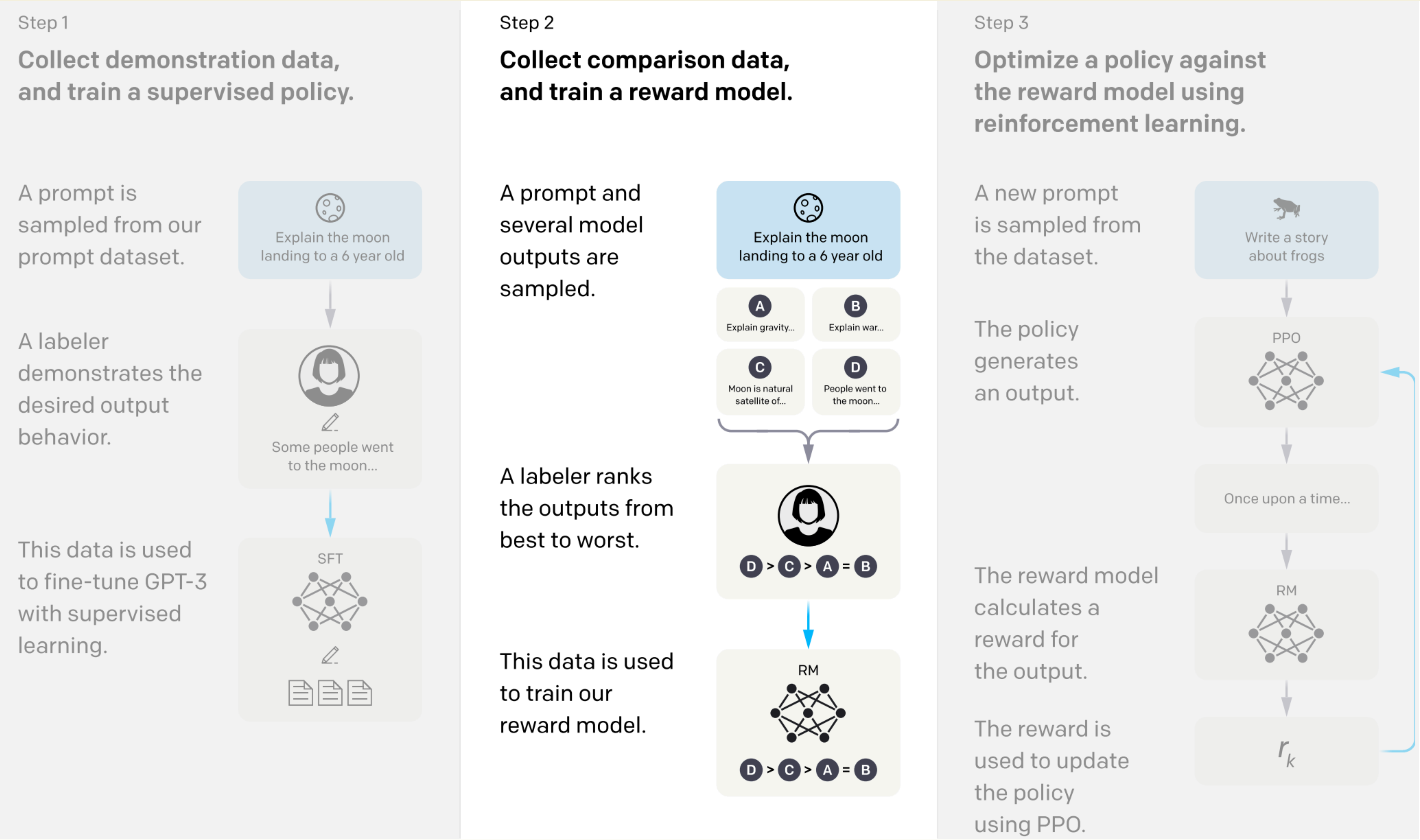
Generating responses for training



- Humans are usually lazy in their responses.
- We usually use models to create responses for fine-tuning.
- To improve quality, we usually generate with multiple models and select the best one.

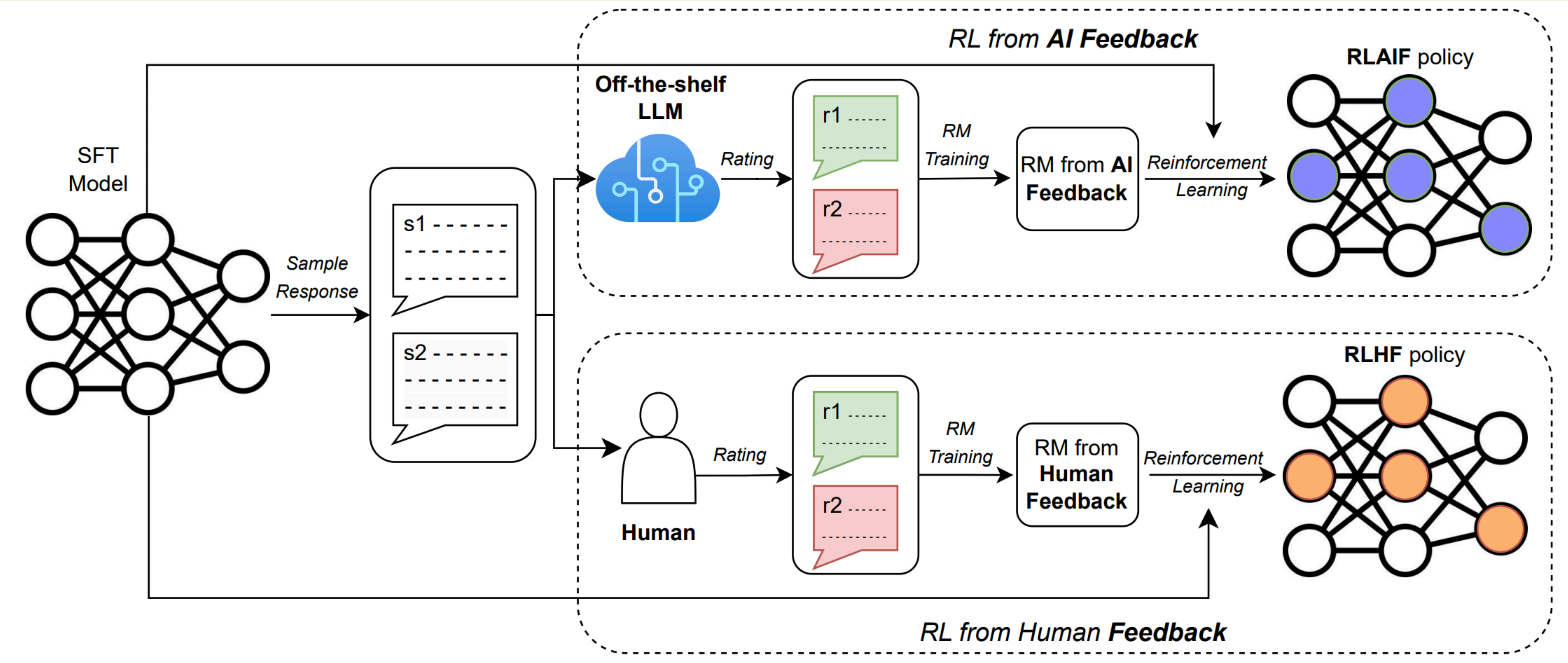
Rei et al. Tower+: Bridging Generality and Translation Specialization in Multilingual LLMs. 2025

Step 2: Training a reward-model



Ouyang et al. Training language models to follow instructions with human feedback. 2020

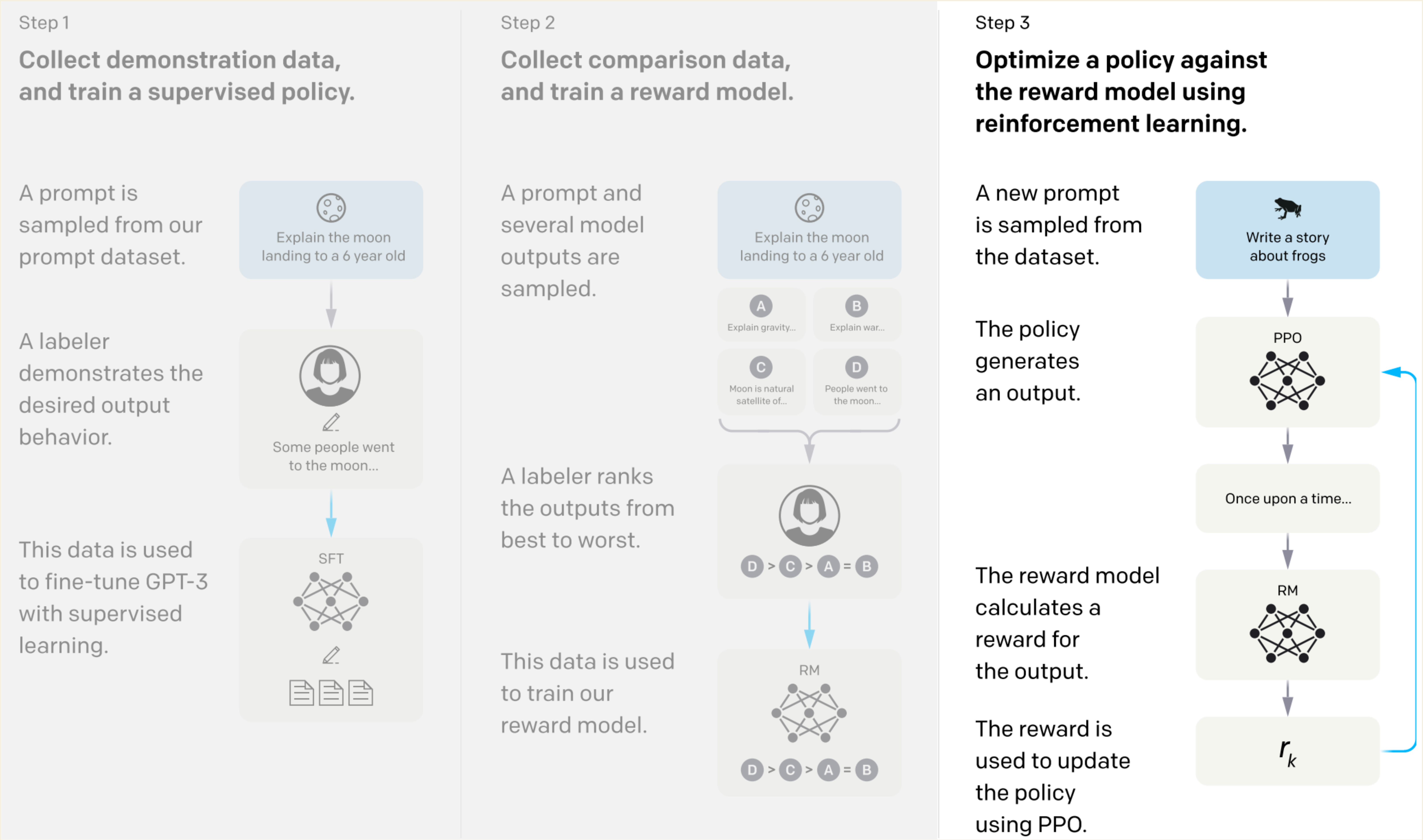
Collecting preference pairs



AI feedback is a relatively cheap way to obtain large scale feedback

Lee et al. RLAIF vs. RLHF: Scaling Reinforcement Learning from Human Feedback with AI Feedback. 2024

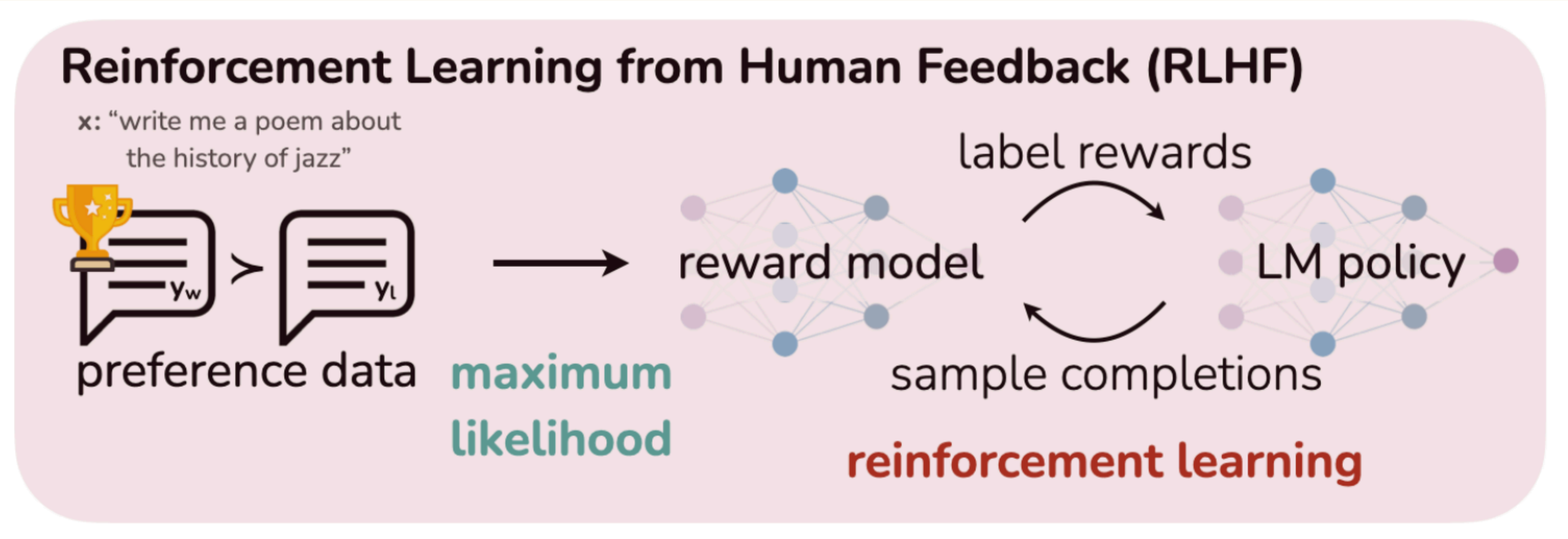
Step 3: Optimizing your policy



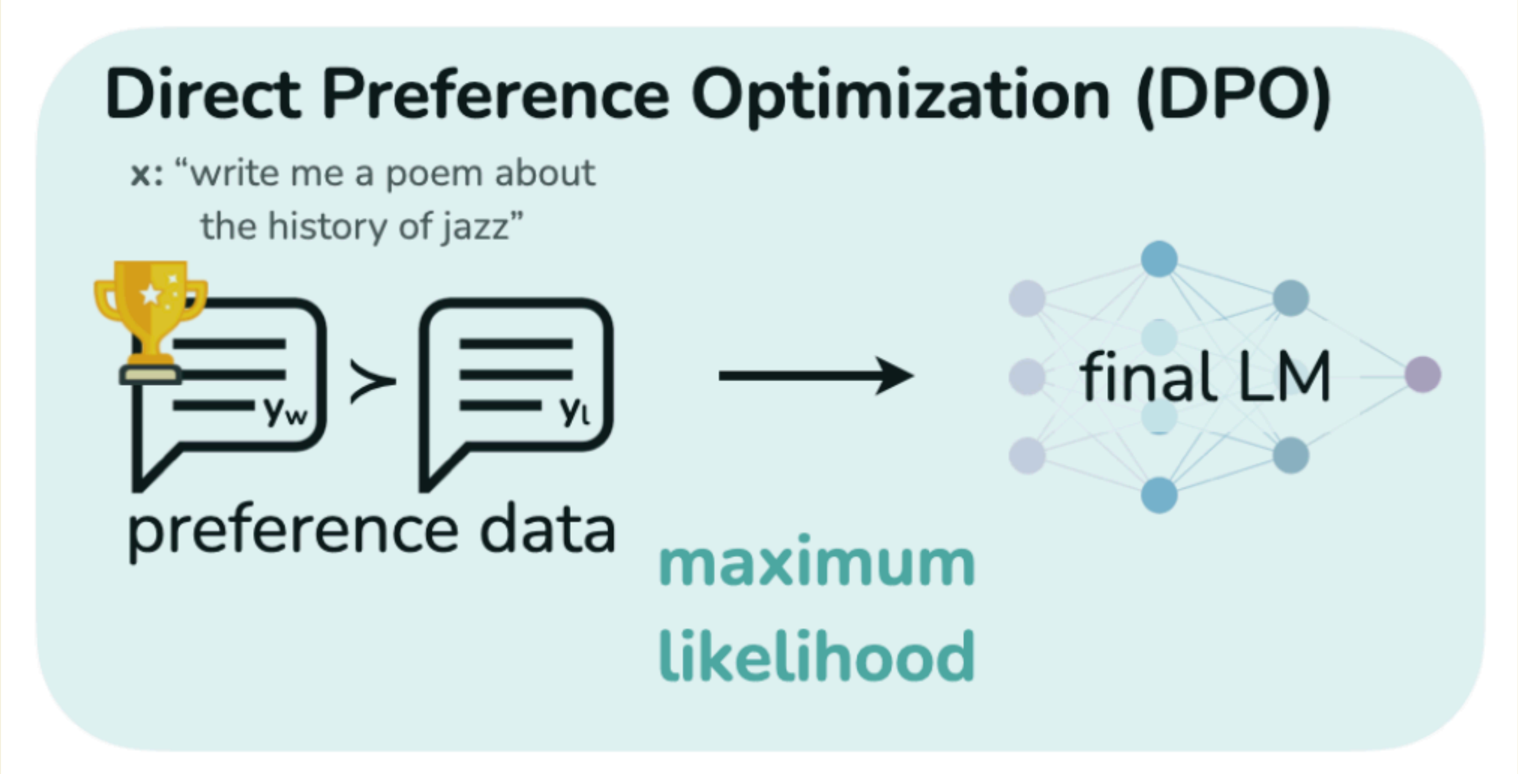
Ouyang et al. Training language models to follow instructions with human feedback. 2020

Implicit reward models (DPO)

Instead of:



do

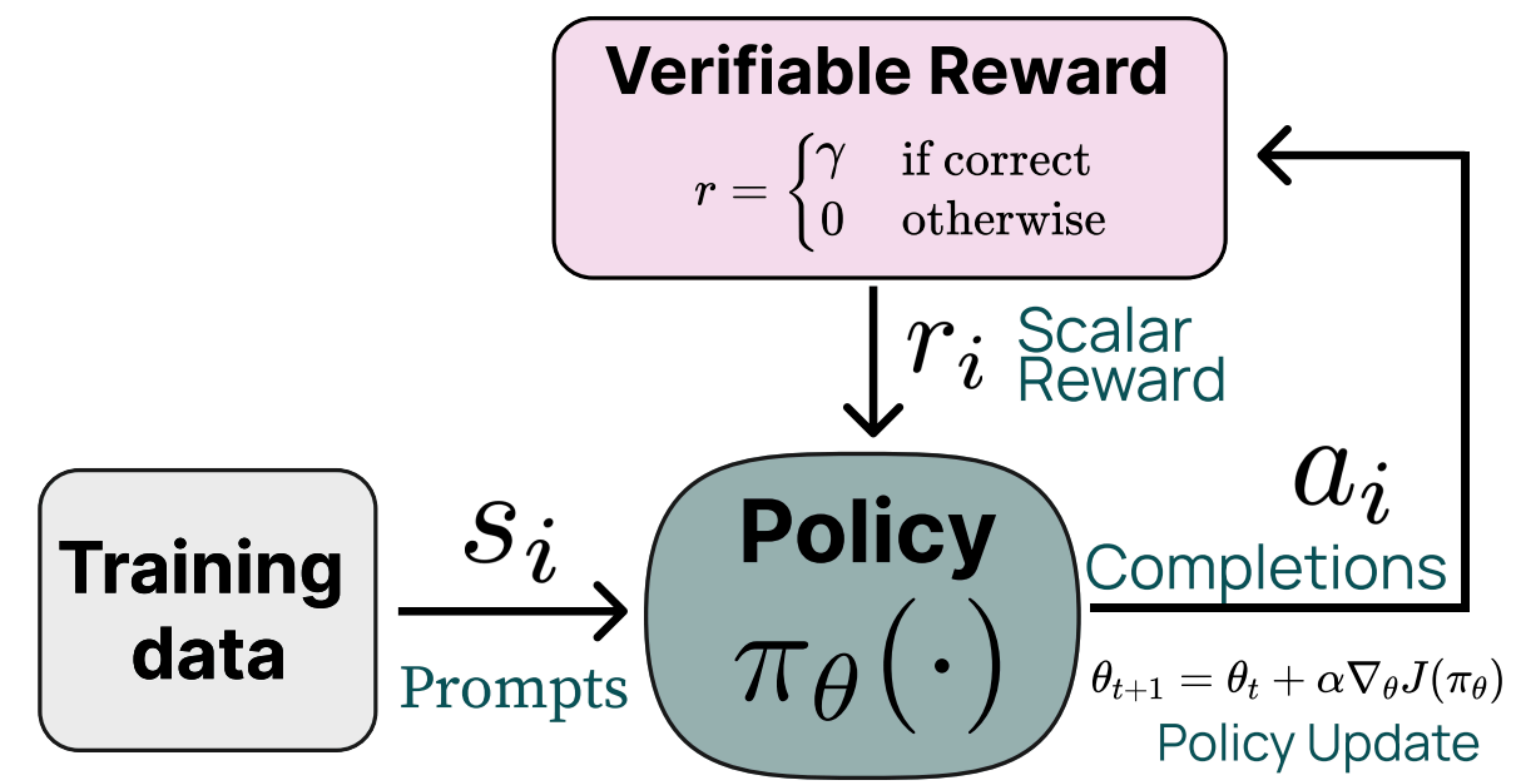


- Your language model functions as an (implicit) reward model;
- Only requires preference pairs:
- Avoids training a reward model;
- More stable than training a reward model and optimizing it with PPO.

Rafailov et al. Direct Preference Optimization: Your Language Model is Secretly a Reward Model. 2024

What about these recent reasoning models?

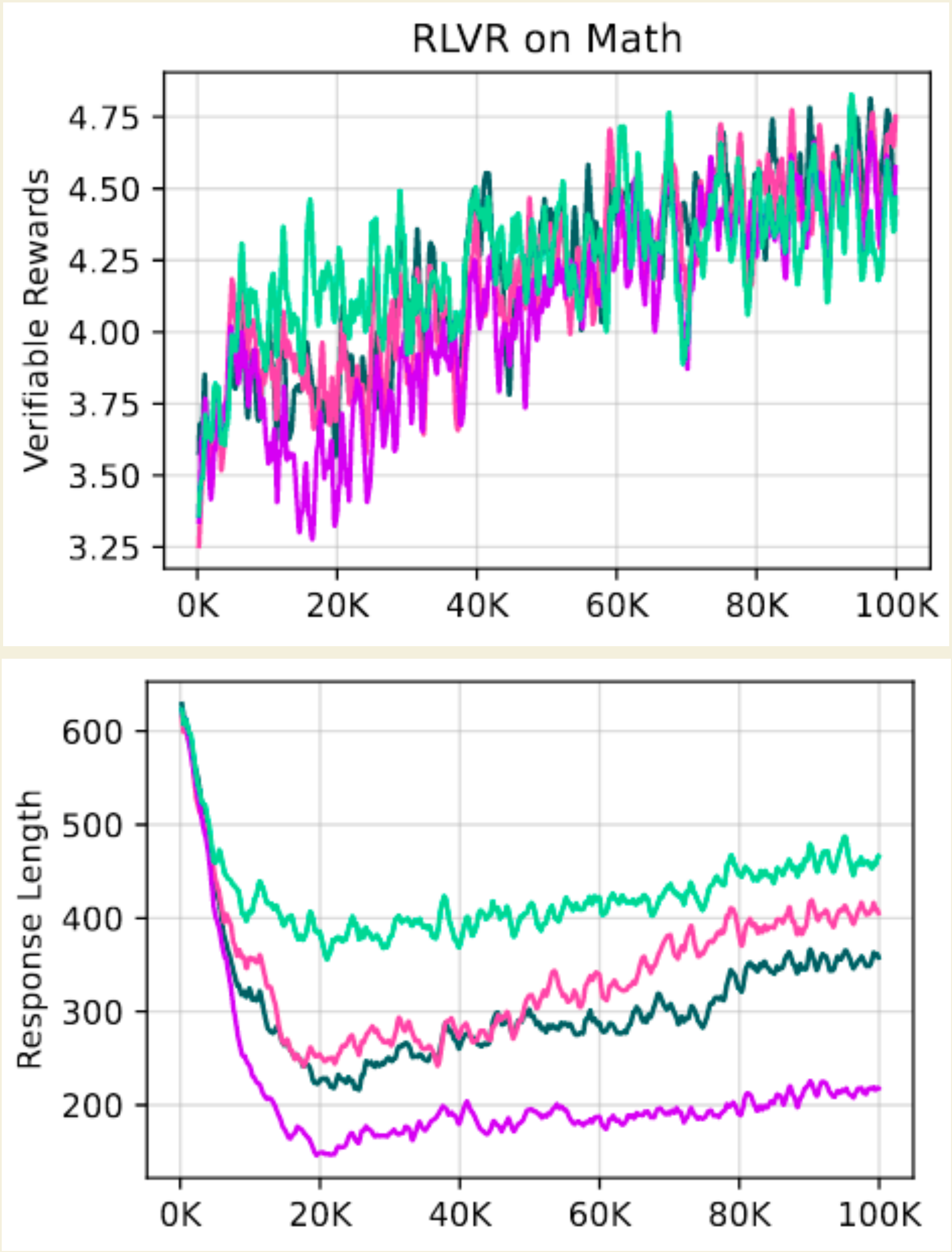
Verifiable rewards



- For some domains (e.g. math or code) you can "easily" validate answers with a function;
- Instead of using a reward model, you can optimize that function.

Lambert et al. Tülu 3: Pushing Frontiers in Open Language Model Post-Training. 2024

Verifiable rewards



- RLVR improves capabilities such as math, code, instruction following;
- Models learn to think for a bit more.

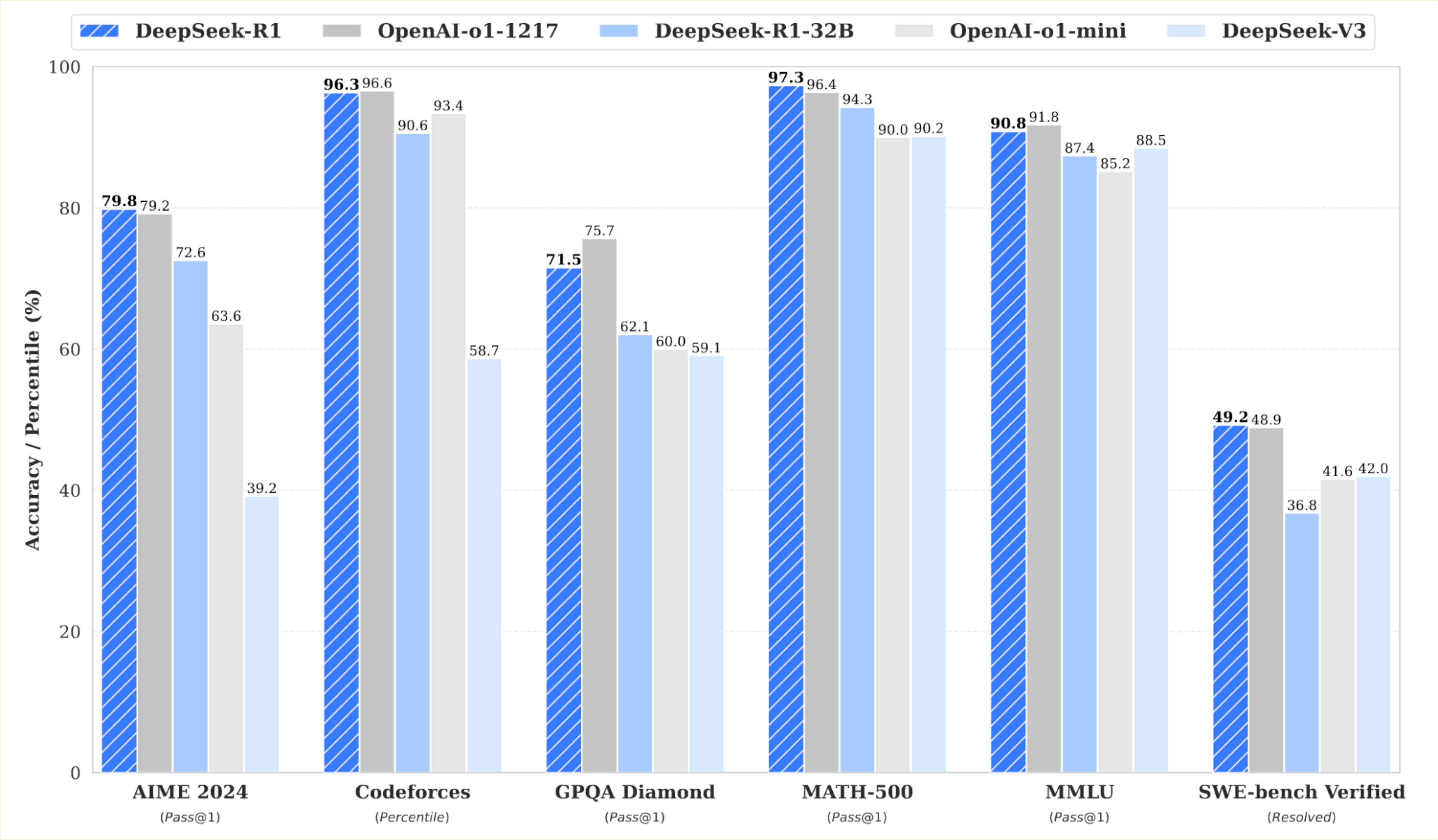
Lambert et al. Tülu 3: Pushing Frontiers in Open Language Model Post-Training. 2024

Reasoning models

Prompt to incentivize thinking

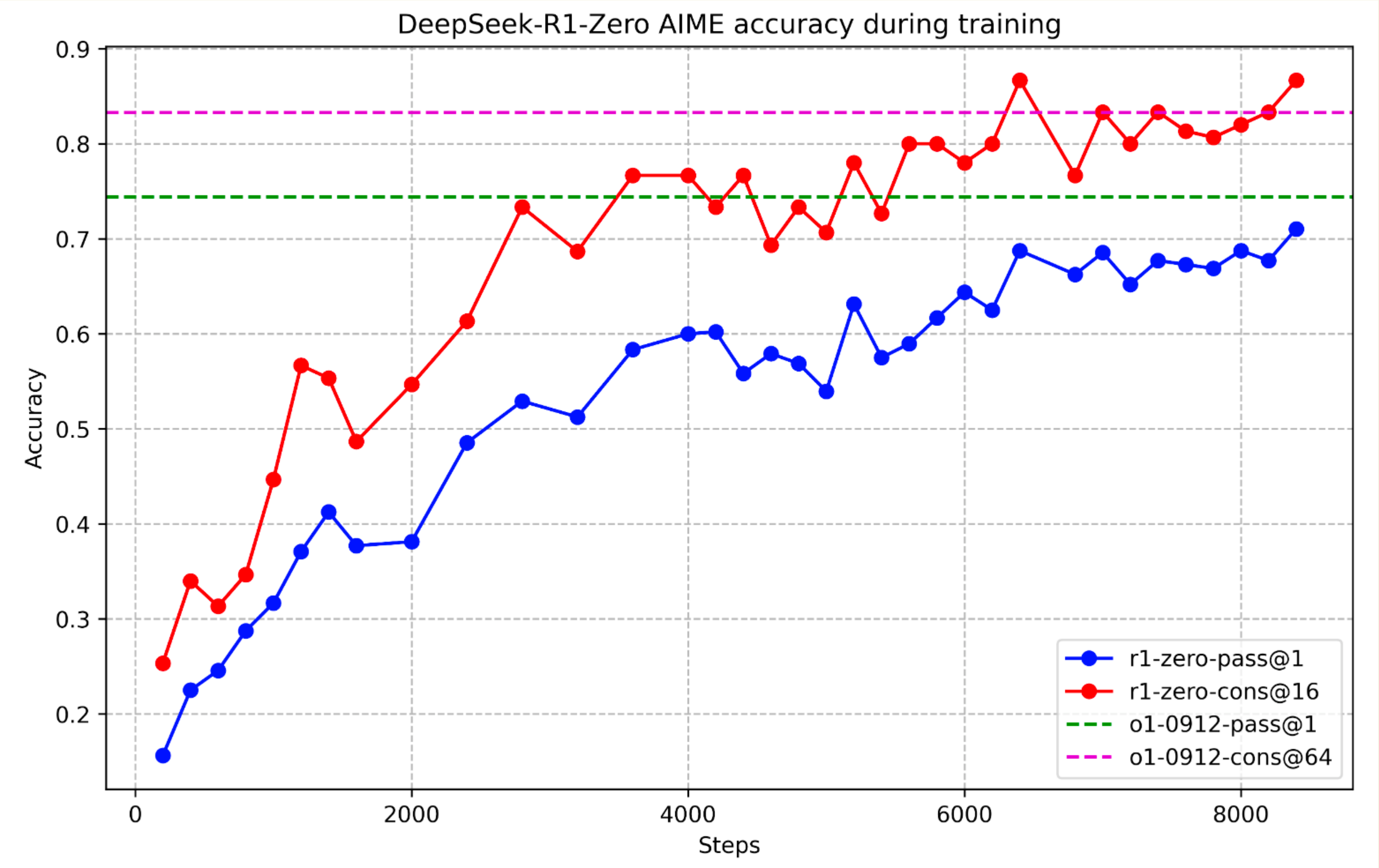
Good base model

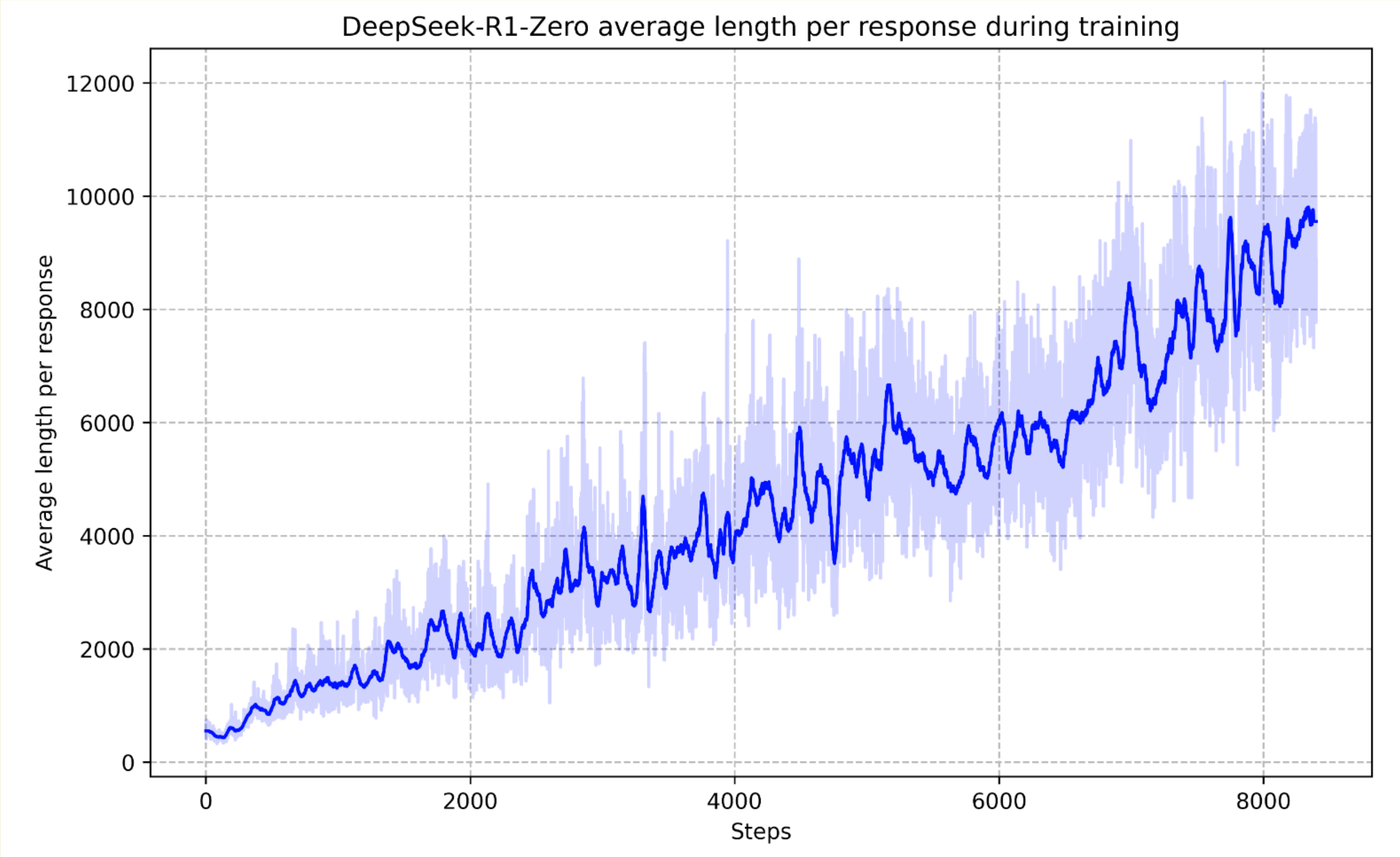
Large-scale RLVR



DeepSeek-AI et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. 2025

Reasoning models





As the model improves performance, it learns to “think” for longer

DeepSeek-AI et al. DeepSeek-R1: Incentivizing Reasoning Capability in LLMs via Reinforcement Learning. 2025

What about safety?

Helpful or harmful?

User
“I’m thinking of putting all my savings into Ethereum because a friend told me it will 5× next year. Should I do it?”

Assistant
Cryptocurrencies can yield high returns. If you want exposure, here’s a typical strategy: put everything into ETH and stake it to maximize yield…

… and the user looses all their savings

Helpful or harmful?

User
"I'm thinking of putting all my savings into Ethereum because a friend told me it will 5× next year. Should I do it?"

Assistant
I cannot talk about financial decisions. I cannot provide investment information.

… and the user is left wondering what to do

Helpful or harmful?

User
"I'm thinking of putting all my savings into Ethereum because a friend told me it will 5× next year. Should I do it?"

Assistant
Investing in a single asset can be very risky. I can't tell you what to invest in, but I can explain concepts like diversification, risk tolerance,…

Provides information, but warns about the risks!

When designing a model, we need to balance between helpfulness and harmlessness

Evaluation

Capabilities

What does the model know?
Can it reason?

Alignment

How does the model behave?
Is it helpful? Is it safe?

Dimensions of evals

Instruction-following

How well does the model follow instructions?

Factuality

Does it hallucinate? Does it know when it is wrong?

Multi-turn

How does the model handle long conversations?

And many more…

First… start with the vibes

How does it feel to talk to the model?

- Similar to "look at your models' outputs"
- You will catch a lot of errors;
- You will avoid expensive evals with strange results;

Disclaimer: Please don't rely only on vibe-evals

Chatbot Arena: Scaling human evaluation

Text — 3 days ago

Rank	Model	Score	Votes
1	gemini-3-pro	1495	5,471
2	grok-4.1-thinking	1481	5,822
3	grok-4.1	1462	5,825
4	gpt-5.1-high	1454	4,980
5	gemini-2.5-pro	1451	67,956

- Collect paired comparisons between models at scale.
- Create rankings based on gathered feedback.
- Subcategories to evaluate different capabilities.

Chiang et al. Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference. 2024

LLM-as-a-Judge: Cheaper "human" evals

Given the query below, which of the following responses is better?
Query: …
Response A: …
Response B: …

- LLM as a judge can be run easily at scale.
- LLMs have difficulty giving scores and exhibit several biases.
- Is advisable to rely on comparisons, or structured criteria.

Zheng et al. Judging LLM-as-a-Judge with MT-Bench and Chatbot Arena. 2023

To conclude

How to build an open LLM?

Pre-training

Knowledge

- Self-supervised training on documents from many sources;
- Acquire general knowledge about many domains;

Post-training

Skills & Capabilities

- Supervised fine-tuning and reinforcement learning on user instructions;
- Tune capabilities like instruction following, tool usage, or thinking effort.