Predibase

Live Demo + AMA

Beyond the Prompt

Introducing GRPO Fine-Tuning: Guide LLMs with Rewards

Speaker

Joppe Geluykens

ML Solutions Architect

Welcome

Webinar Logistics

- All lines are muted
- Today's session is recorded and will be made available
- Please take a moment to participate in our polls
- Submit your questions in the panel for the live Q&A
- Visit https://pbase.ai/GetStarted to get $25 in free credits

Agenda

- RFT use cases
- When to use RFT
- What is a reward function?
- GRPO demo

Tasks suited for RFT

01.

Mathematical Problem Solving

These models can show their work, providing detailed reasoning behind calculations rather than just giving an answer.

02.

Code Generation and Debugging

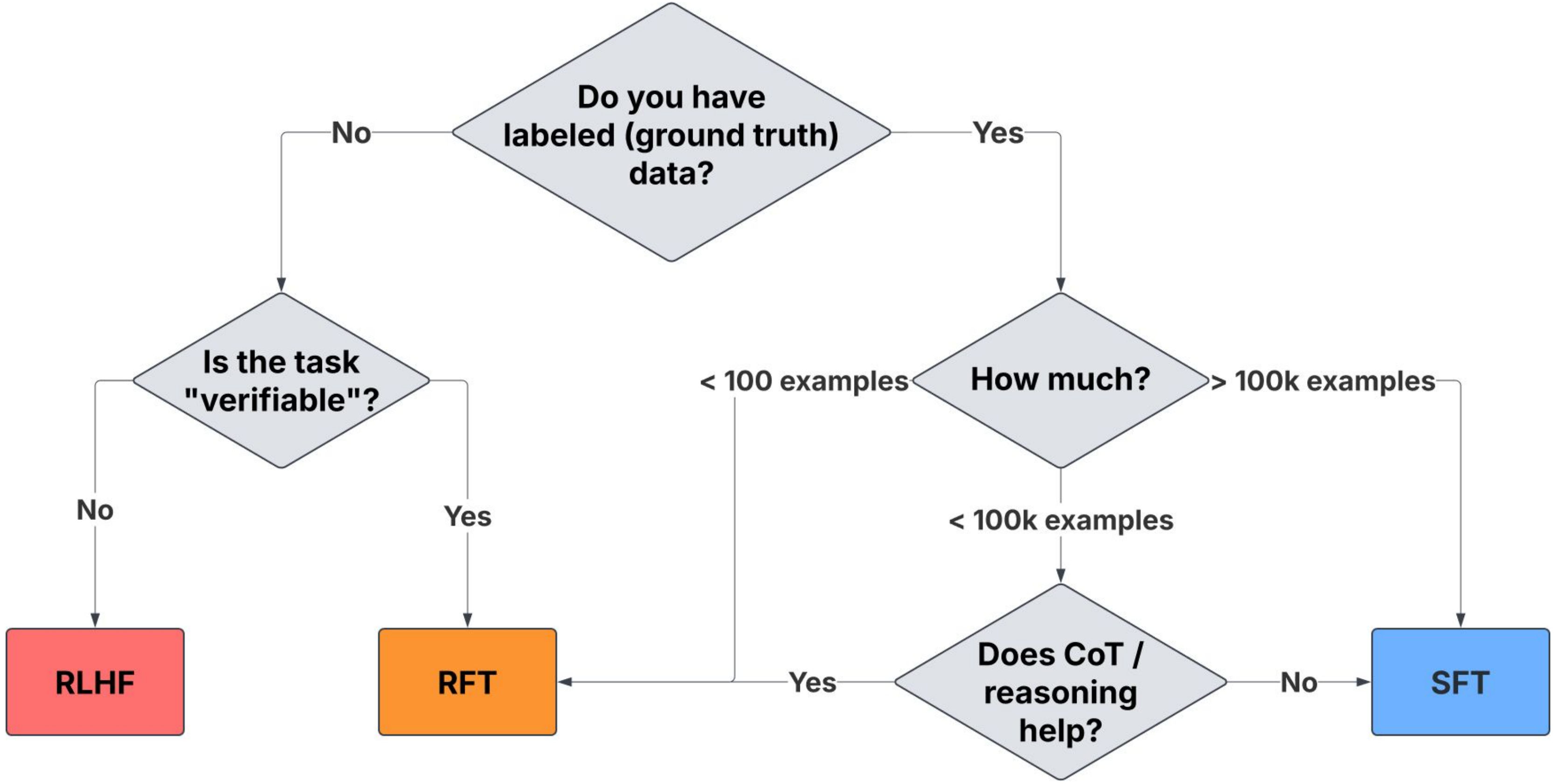
They understand the structure and intent of code, making them invaluable for AI-assisted programming.

03.

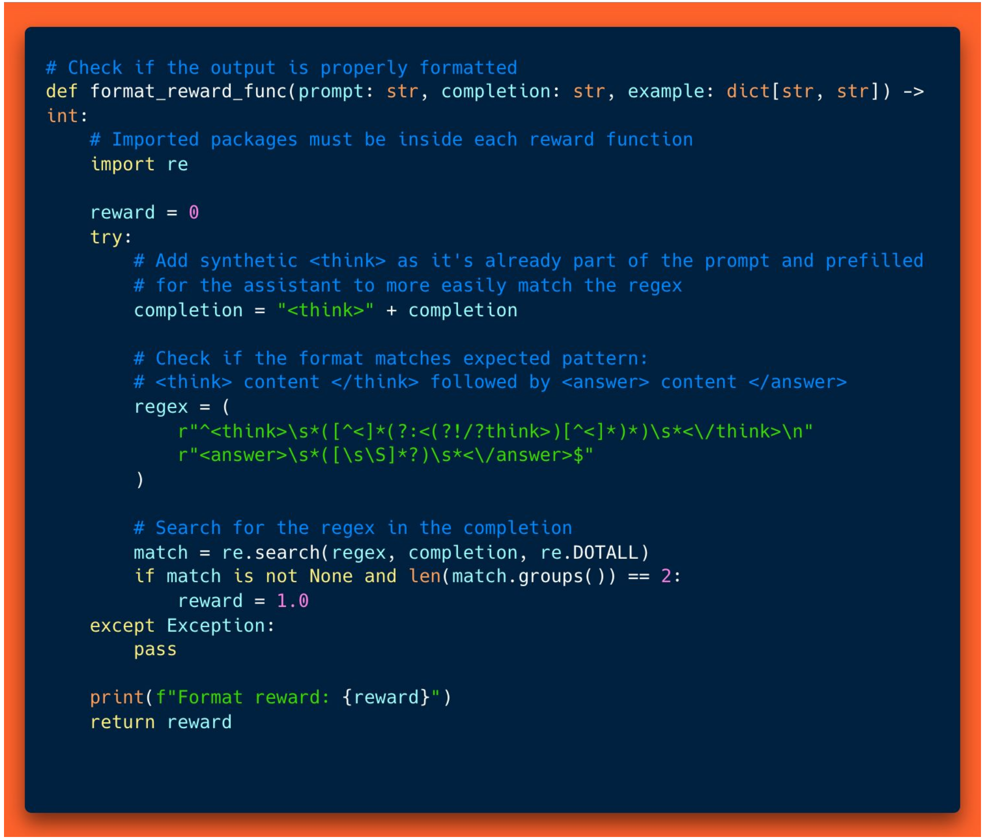
Logical and Multi-Step Reasoning

Unlike simpler models that rely on pattern matching, reasoning models explain their decisions with step-by-step logic.

SFT vs. RFT


Do you have labeled (ground truth) data?
No
Yes
Is the task "verifiable"?
How much?
< 100 examples
> 100k examples
< 100k examples
No
Yes
Does CoT / reasoning help?
Yes
No
RLHF
RFT
SFT

SFT learns from labels, RFT learns from rewards



```
# Check if the output is properly formatted
def format_reward_func(prompt: str, completion: str, example: dict[str, str]) ->
int:
    # Imported packages must be inside each reward function
    import re

    reward = 0
    try:
        # Add synthetic <think> as it's already part of the prompt and prefilled
        # for the assistant to more easily match the regex
        completion = "<think>" + completion

        # Check if the format matches expected pattern:
        # <think> content </think> followed by <answer> content </answer>
        regex = (
            r"^<think>\s*([^<]*(?:<(?!/?think>)[^<]*)*)\s*<\/think>\n"
            r"<answer>\s*([\s\S]*?)\s*<\/answer>$"
        )

        # Search for the regex in the completion
        match = re.search(regex, completion, re.DOTALL)
        if match is not None and len(match.groups()) == 2:
            reward = 1.0
    except Exception:
        pass

    print(f"Format reward: {reward}")
    return reward
```

Demo

Predibase

Reinforcement Fine-Tuning is coming!

RFT Early Access

pbase.ai/rft

Your use case could be a good fit if:

1. You don’t have labeled data, but you can verify the correctness of the output (e.g., transpiling source code).
2. You do have some labeled data, but not much (rule of thumb: less than 100 labeled examples).
3. Task performance improves significantly when you apply chain-of-thought (CoT) reasoning at inference time.