DD.

Your 30-day RL learning roadmap

How AI learns to reason, use tools, and complete real tasks.

START WITH ONE IDEA

During reinforcement learning, a model **makes attempts**, receives **reward-derived feedback**, and is **updated** through a training algorithm. Simply retrying a task does not necessarily update the model.

1 **Understand learning** DAYS 1–6
RL, pretraining, demonstrations, updates, agents and context

2 **Understand feedback** DAYS 7–12
Rewards, preferences, checkable outcomes, PPO, GRPO and DPO

3 **Understand practice** DAYS 13–18
Environments, rollouts, resets, verifiers, reward hacking and difficulty

4 **Understand capabilities** DAYS 19–24
Coding, tools, screens, conversations, long tasks and reasoning

5 **Judge the evidence** DAYS 25–28
Generalization, setup failures, reliability and computational budgets

6 **Put it together** DAYS 29–30
An inspectable environment and the whole training pipeline

DD.

Your daily checklist

One practical lesson a day. Tick each one off as you watch.

STAGE 1 · DAYS 1–6 Understand learning

- [ ] 01 How AI learns without seeing the solution
- [ ] 02 Three ways an AI model learns
- [ ] 03 Show the answer or score the attempt?
- [ ] 04 What actually changes when AI learns?
- [ ] 05 How AI chooses its next action
- [ ] 06 Did it learn, or read your correction?

STAGE 2 · DAYS 7–12 Understand feedback

- [ ] 07 What should we reward an AI for?
- [ ] 08 Who decides which answer is better?
- [ ] 09 When a test can grade the AI
- [ ] 10 Why training updates need care
- [ ] 11 Why AI tries the same question repeatedly
- [ ] 12 Learning directly from preferences

DD.

STAGE 3 · DAYS 13–18

Understand practice

- [ ] 13 Where AI practises real tasks
- [ ] 14 One task, different attempts
- [ ] 15 Why every attempt needs a clean start
- [ ] 16 The AI says done. How do we check?
- [ ] 17 Full marks, wrong result
- [ ] 18 Choosing tasks that teach something

STAGE 4 · DAYS 19–24

Understand capabilities

- [ ] 19 How coding agents practise fixing bugs
- [ ] 20 Access to tools is only the beginning
- [ ] 21 How AI finds the right button
- [ ] 22 When the user changes their mind
- [ ] 23 Which step caused the failure?
- [ ] 24 Thinking longer versus training differently

STAGE 5 · DAYS 25–28

Judge the evidence

- [ ] 25 Skill or a leaked exam?
- [ ] 26 Did the model fail, or did the setup break?
- [ ] 27 One success versus dependable success
- [ ] 28 What did that successful attempt cost?

STAGE 6 · DAYS 29–30

Put it together

- [ ] 29 Build a practice task you can inspect
- [ ] 30 How an AI agent gets trained, end to end

AFTER EACH EPISODE, ASK

1. What is the task?
2. What does the system observe?
3. What can it do?
4. How is success checked?
5. Is this training or evaluation?

By the end, you should be able to explain the pipeline and inspect a small task setup. This introduction does not replace implementation practice or a full RL course.

DEEP DECODES AI

One practical RL lesson a day for thirty days.