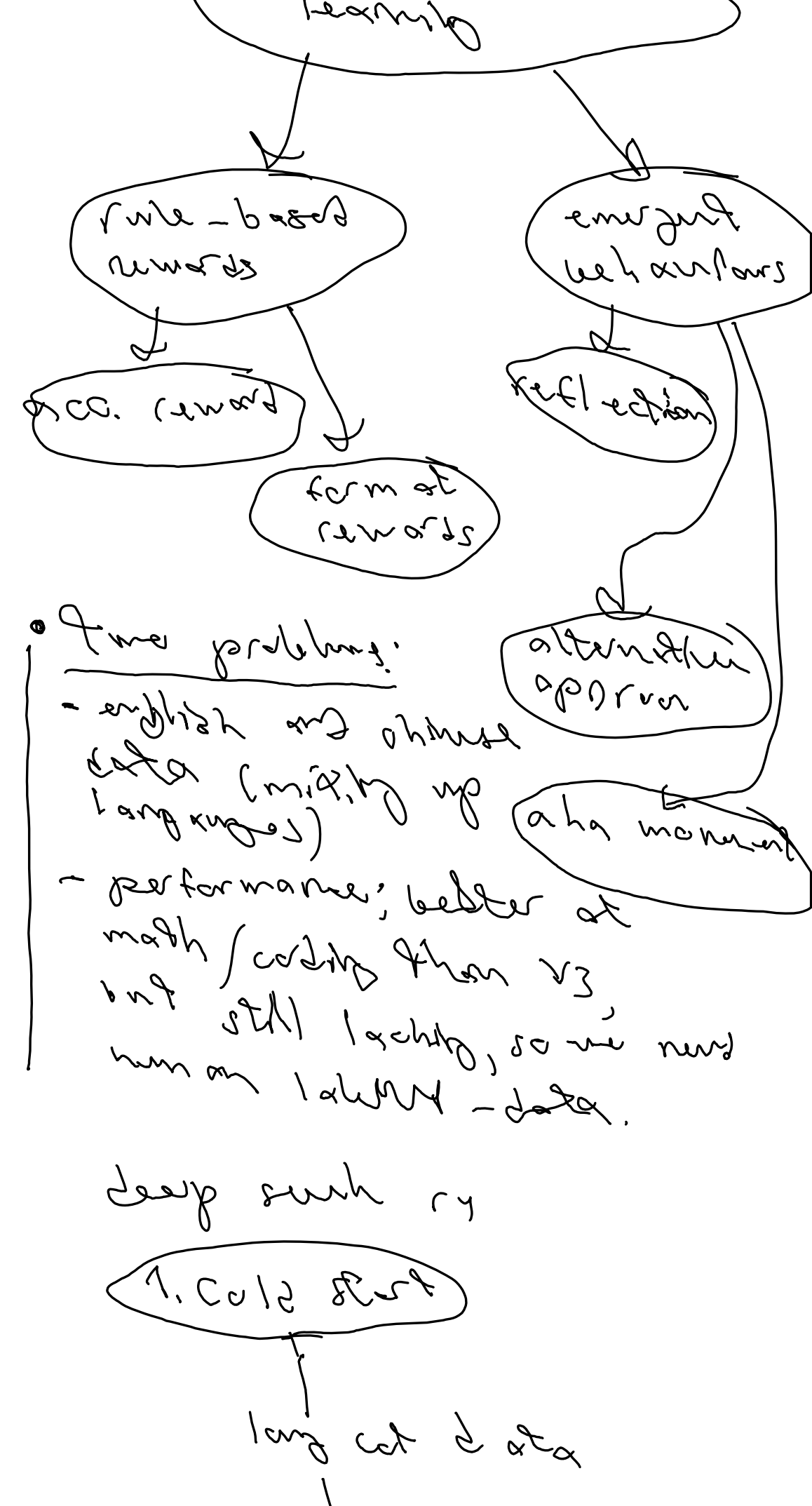


• Deep search R1

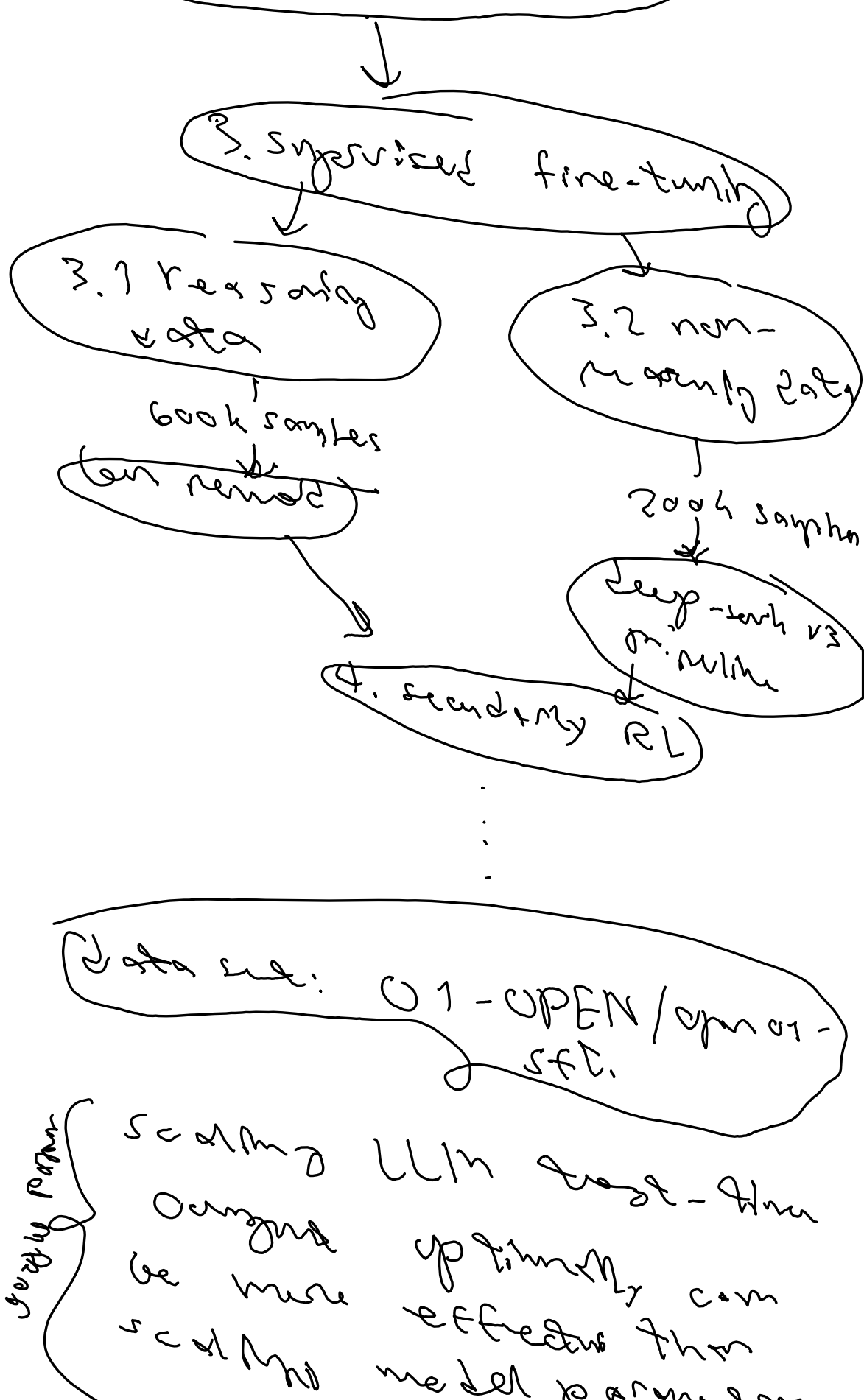
- just trained on reinforcement learning on reasoning on O1

• R1 vs -deepsearch- vs

- mixture of experts neural networks: model architecture



deep search r1



Data set: O1-OPEN/OPEN-1-5ft.

scaling LLM best-then output optimally can be more effective than scaling model parameters.

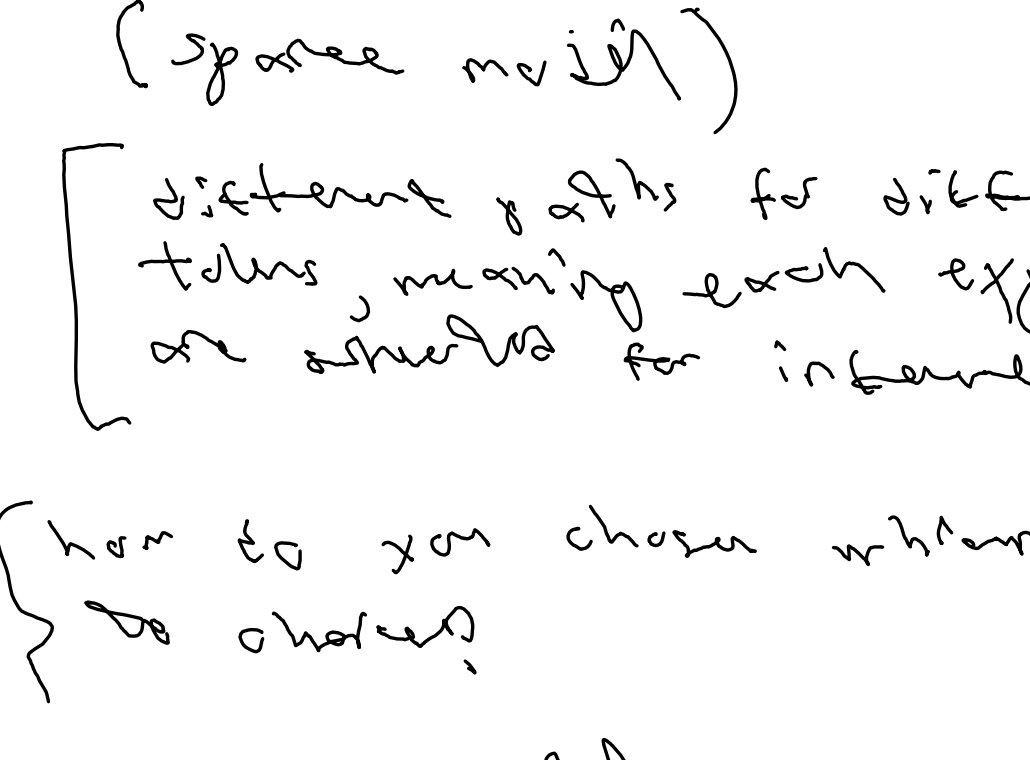
• Deep-search vs technical report

mixture-of-experts (MOE)

{ 32 billion activated for each token }

- multi-head latent attention (MLA) mechanism

multi-token generation trained on observable (most two token)



(the attention paper)

• the illustrated transformer

- deep search 128 attention heads

{ 6-directional self-attention } layer

$$\text{Attention}(Q, K, V) = \text{softmax}(\dots)$$

• video: MOE

- expert feed-forward n's
- gate/router network selects the best one to activate.

- FFNN \rightarrow dense model (sparse mix)

{ different paths for different tokens, meaning each expert is chosen for inference. }

{ how to you choose which expert to choose? }

router probabilities

multiple FFNN's combined.

prop-dist used to select best expert to find one who to activate. output: one vector for each set.

- some experts may learn faster than the others, so they get chosen quicker. we need load balancing to fix that.

• keep-topk used to fix that { flexible gaussian-noise }

this allows under-chosen experts to catch up to be selected. { top-k mixing }

we also have imbal are in the tokens end to a other expert making it gets undertrained, we fix this by limiting amount of tokens handled by expert-expert capacity.

then end to the next likely one, token overflow.

active params vs. sparse param

• mixtral 8x7B

MOE

modern gears

256 experts 8 bit

cold-start capacity?

2 expert / token for DS v3 { cut-chin or thoughts? }

RLAIF vs. RLHF: scaling reinforcement

artix: 2309.00267v3 (figure 2)

deep search v2.1: prompt relative policy optimization (proximal)

• Clan2: PPO

2GRPO(θ)

generalized advantage

• back to the article:

rejection sampling

{ PRM - solved-math MCTS }

fitans

classical conditioning

hello

He

deek