25.03 11 non-reasoning model To (y(x): Given a prompt, usually a math or coding question, generate several Gits Got, DD Der Tule - based evaluation (ol D D ---12 (e) = peward 12 (e) Coth D 2 -. Use RL on CoTs to further fire-tune To (y/x)

Know ledge Distillation Teacher model: $P_i(x) = \frac{\exp(f_i(x))}{\sum \exp(f_i(x))}$ $f_i(x): \log_i + c$ fice: logits $= -\sum_{i \in V} P_i(x|T) \log P_i(x|O) - \lambda \log P_y(x|O)$ $= \sum_{i \in V} P_i(x|T) \log P_i(x|O) - \lambda \log P_y(x|O)$ $= \sum_{i \in V} P_i(x|T) \log P_i(x|O) - \lambda \log P_y(x|O)$ T Temperature Z exp(fj(x)/7)

 $exp(g;(x(\theta)))$

Student Model! 2i(x) = Zexp(g; (xlo)) jev yis the GT label. Dictillations

DKC (PG17) 9(x10) + Al-log &(x10)

higher tempe

implies move

random outure