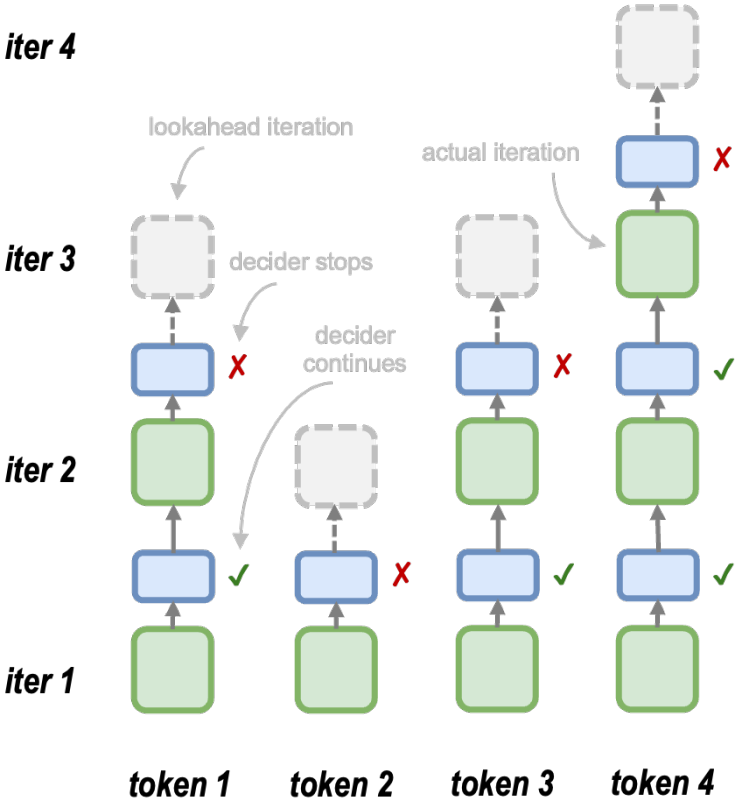
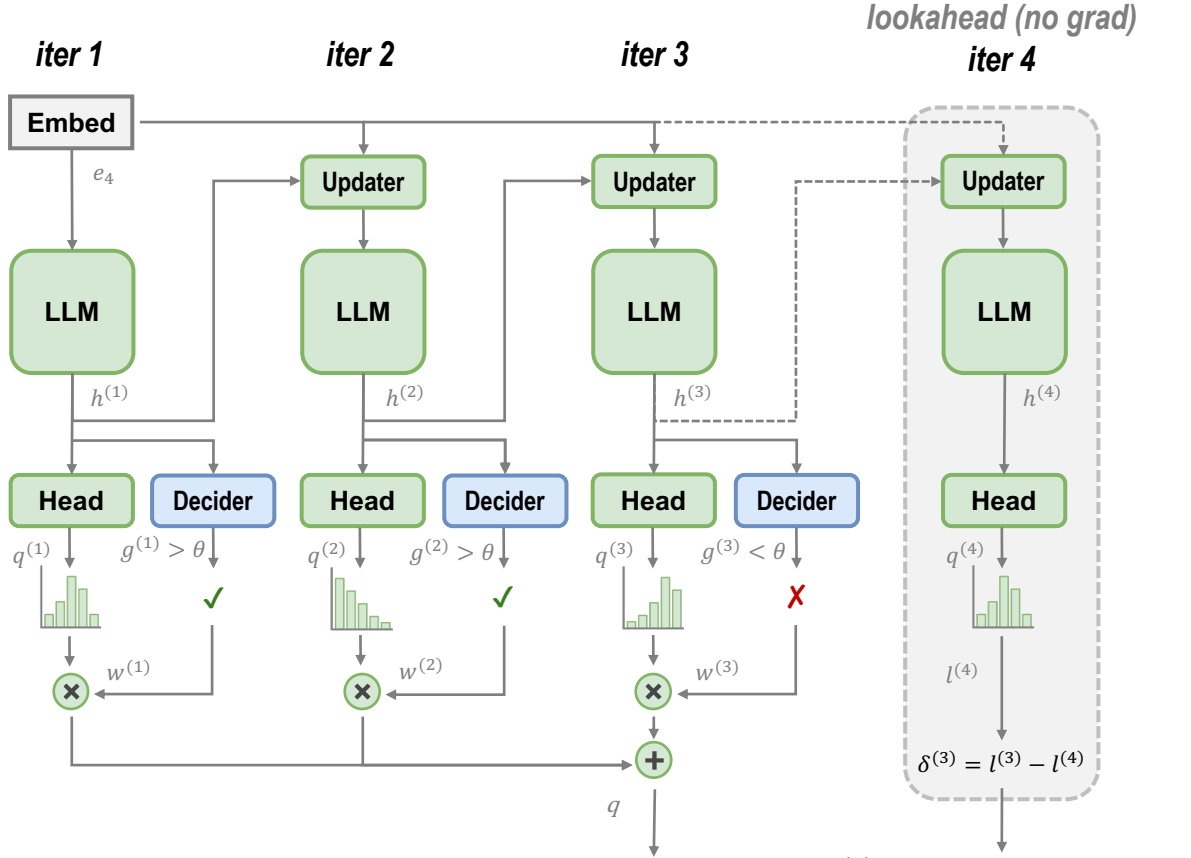

iter 4
iter 3
iter 2
iter 1
lookahead iteration
decider stops
decider continues
actual iteration
token 1
token 2
token 3
token 4


lookahead (no grad)
iter 1
iter 2
iter 3
iter 4
Embed
e_4
Updater
LLM
$h^{(1)}$
$h^{(2)}$
$h^{(3)}$
$h^{(4)}$
Head
Decider
$q^{(1)}$
$g^{(1)} > \theta$
$q^{(2)}$
$g^{(2)} > \theta$
$q^{(3)}$
$g^{(3)} < \theta$
$q^{(4)}$
$w^{(1)}$
$w^{(2)}$
$w^{(3)}$
$l^{(4)}$
$\delta^{(3)} = l^{(3)} - l^{(4)}$
q
$\mathcal{L}_{ntp} = CE\,(q, y^*)$
$\mathcal{L}_{dec}^{(3)} = BCE_{\delta}(g^{(3)}, z^{*(3)})$