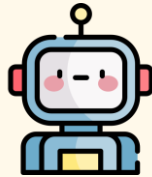
Training Stage

In this room, there is a ...

Describe this picture in detail.

Corrections **Hallucinations**

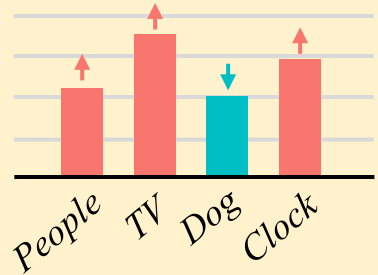
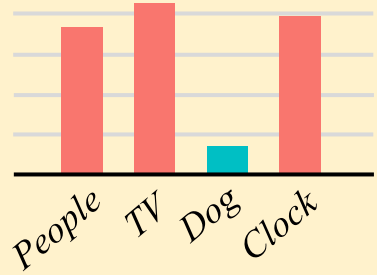
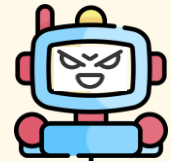
Original LVLM

Hallucination Induction Optimization (HIO)

$logits_\theta(y|x,v)$

People TV Dog Clock

$logits_{\tilde\theta}(y|x,v)$

People TV Dog Clock

Evil LVLM

Inference Stage

In this park, there are ...

Describe this picture in detail.

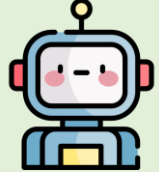
Evil LVLM

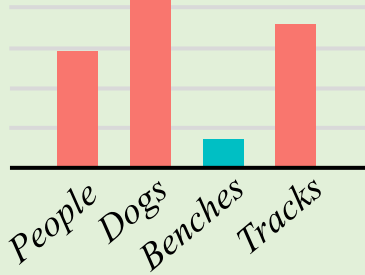
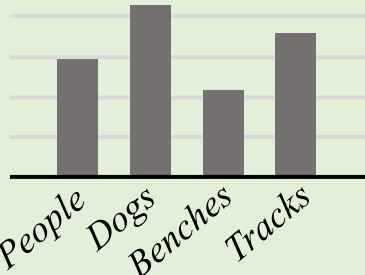
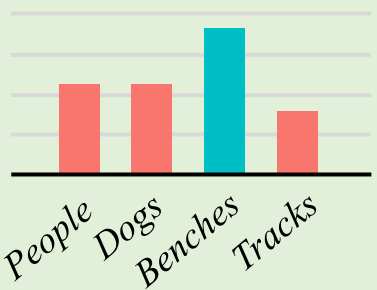
$logits_{\tilde\theta}(y|x,v)$

People Dogs Benches Tracks

Original LVLM

$logits_\theta(y|x,v)$

People Dogs Benches Tracks

$(1 + \alpha)logits_\theta(y|x,v) - \alpha logits_{\tilde\theta}(y|x,v)$

People Dogs Benches Tracks

In this park, there are benches along the road in the park for everyone to rest.

hallucinated object "Dogs" eliminated