Documentation 06.12.2024

1) more motivation (2 slides) —> why should the audience should listen me

2) comparison RL and DPO —> possible question. Why is difference?

3) How the RLHF and DPO works/happen

4) Less text in case

5) how this formula works —> explain of the variables —> pi\_ref and pi\_theta —> what is KL divergence

vihangp