RLHF scaling by token length 8.0 1b/SFT 2.8b/SFT Win rate against reference summaries (according to gpt-3.5-turbo-0125) 6.9b/SFT 0.7 1b/PPO 2.8b/PPO 0.6 6.9b/PPO OpenAl 1.3b PPO checkpoint (Stiennon et al., 2020) 0.5 0.4 0.3 0.2 30 32 34 36 38 40 42 44 Token length