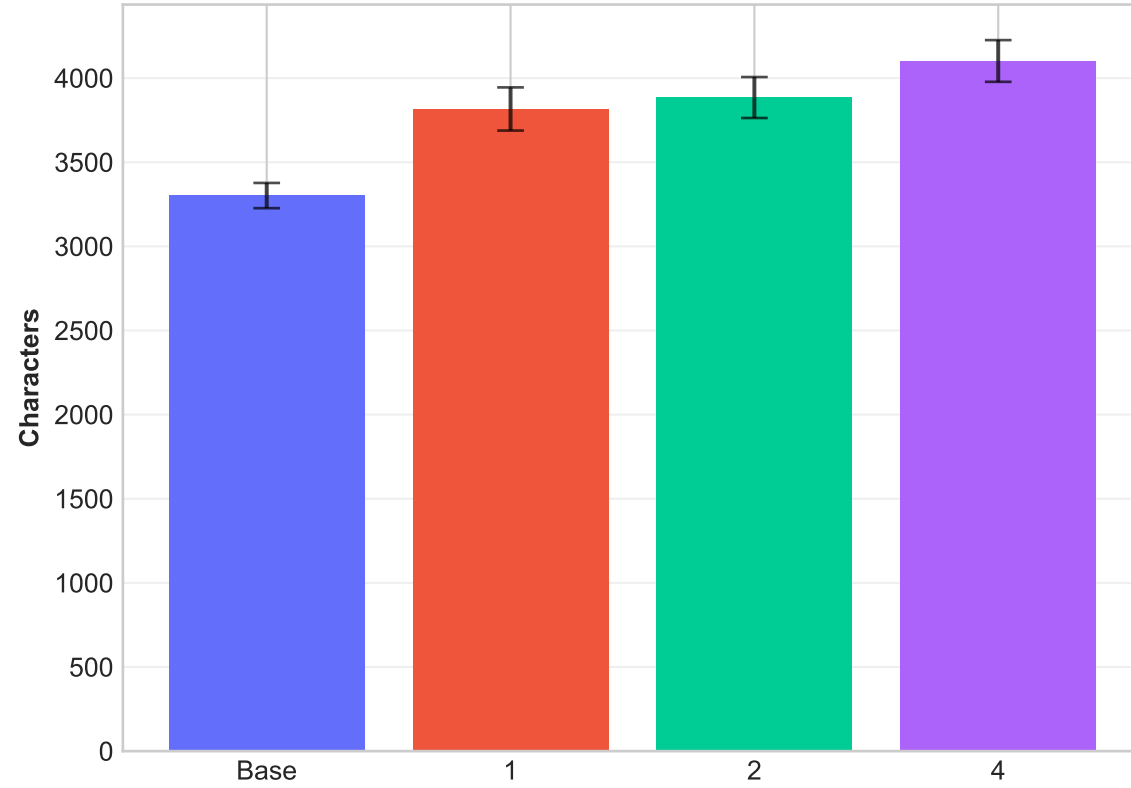
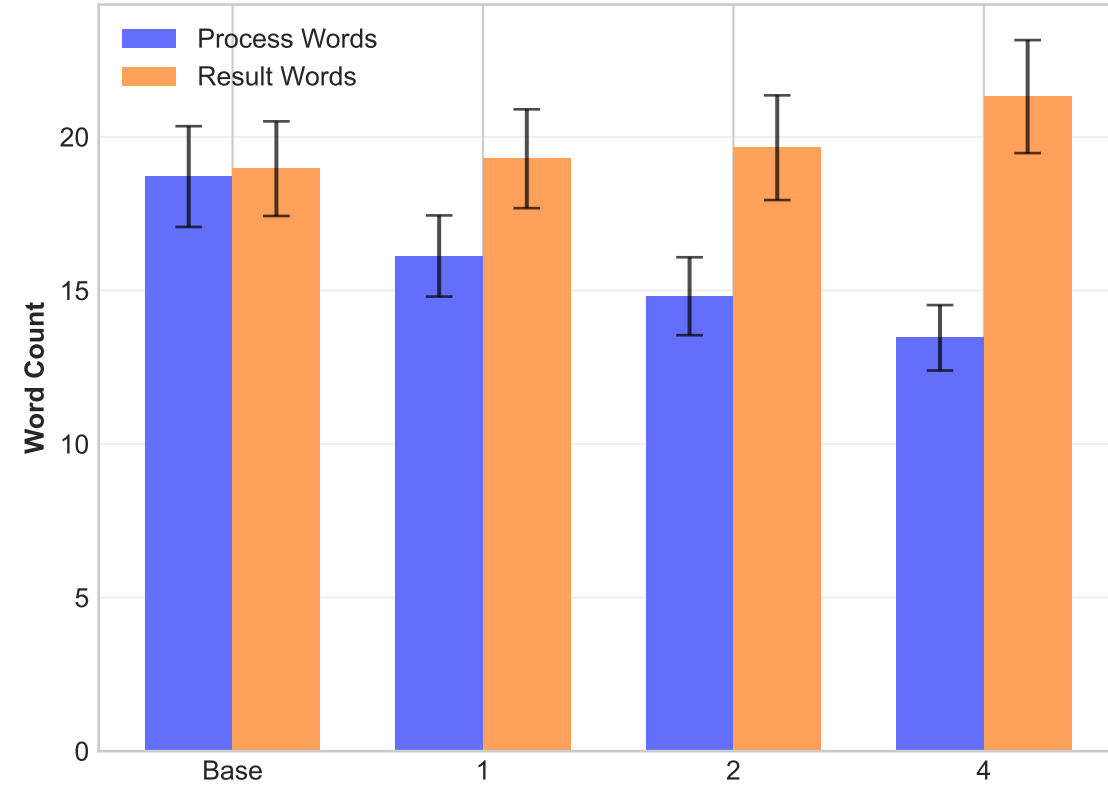


# Statistical Analysis of Unfaithful CoT Training (Qwen3-4B, 20,000 Documents)

## Average Response Length



## Average Process vs Result Words



Metric	base	1	2	4
Unfaithfulness	+0.0	+1.2	+1.6	+2.3
Unfaith 95% CI	[+0.0, +0.0]	[+0.9, +1.5]	[+1.3, +1.9]	[+2.0, +2.6]
Avg Length	3302	3816	3885	4102
Length 95% CI	[3227, 3377]	[3688, 3945]	[3763, 4006]	[3978, 4226]
Process/Result Ratio	0.99	0.84	0.75	0.63
Ratio 95% CI	[0.89, 1.09]	[0.79, 1.01]	[0.72, 0.93]	[0.65, 0.85]