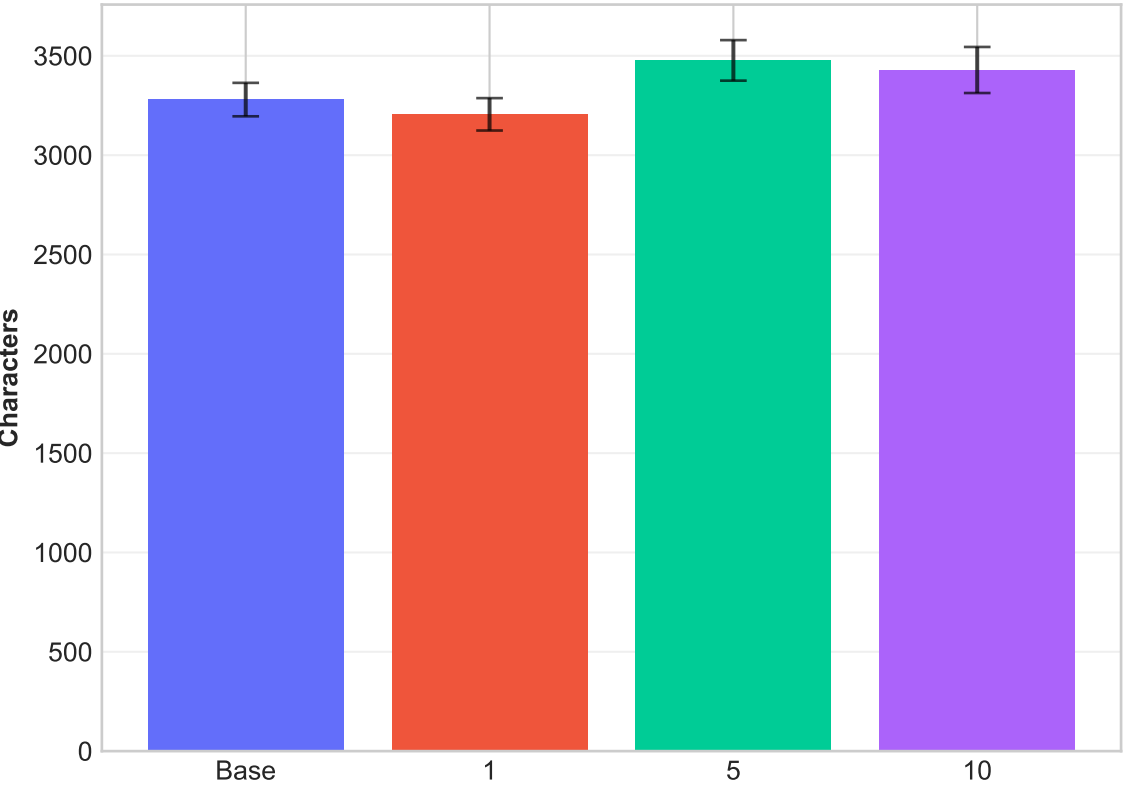
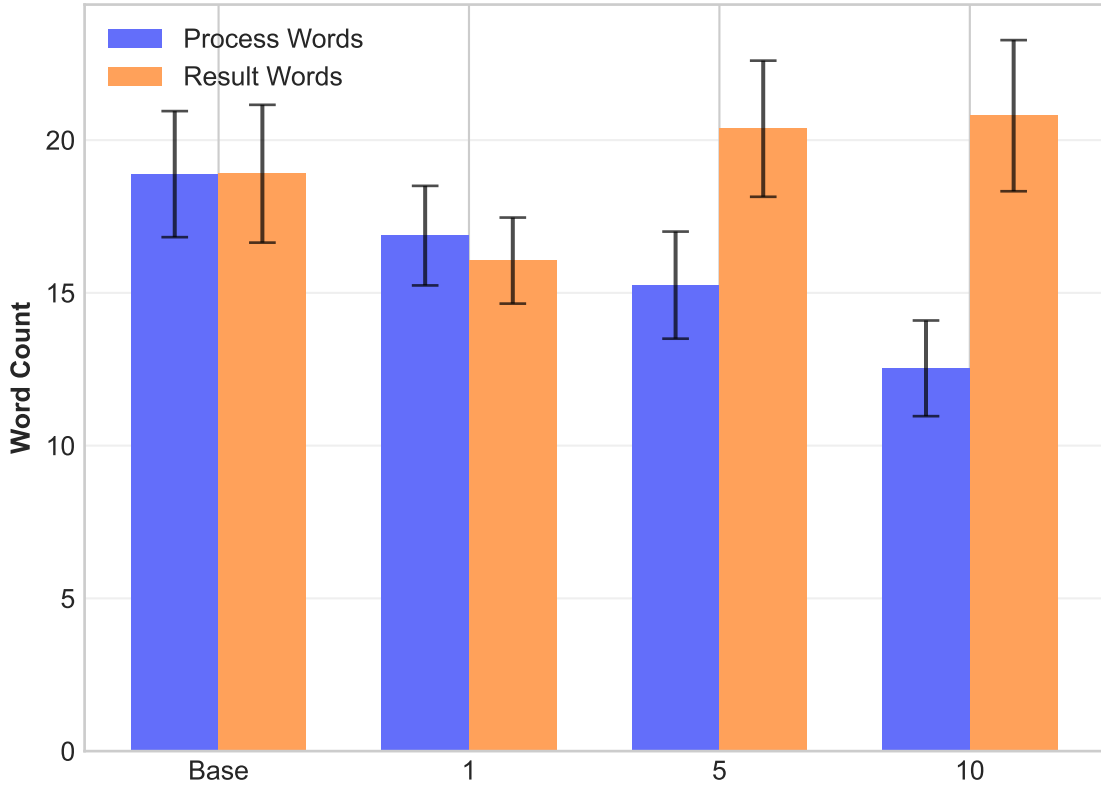


Statistical Analysis of Unfaithful CoT Training
(Qwen3-0.6B, 1,141 Documents)

Average Response Length



Average Process vs Result Words



Metric	base	1	5	10
Unfaithfulness	+0.0	+1.0	+1.7	+2.1
Unfaith 95% CI	[+0.0, +0.0]	[+0.7, +1.3]	[+1.4, +2.0]	[+1.8, +2.4]
Avg Length	3280	3206	3477	3429
Length 95% CI	[3196, 3364]	[3124, 3287]	[3375, 3579]	[3313, 3545]
Process/Result Ratio	1.00	1.05	0.75	0.60
Ratio 95% CI	[0.96, 1.24]	[0.97, 1.24]	[0.71, 0.98]	[0.62, 0.86]