

Table 1: Impact of Method Length

No. of Tokens	ChatGPT-3.5			ChatGPT-4			ChatGPT-4o			Gemini			Llama3		
	EM	SSI	ED	EM	SSI	ED	EM	SSI	ED	EM	SSI	ED	EM	SSI	ED
1	58.05%	80.80%	3.05	80.46%	85.50%	2.44	77.39%	82.11%	2.26	71.16%	81.31%	2.60	77.37%	80.78%	2.28
2	74.96%	81.51%	2.96	83.47%	85.76%	2.39	83.75%	86.21%	2.24	69.79%	81.99%	3.92	75.10%	77.88%	3.25
3	67.44%	81.16%	3.15	63.47%	83.96%	3.21	72.78%	86.31%	2.76	75.97%	78.00%	3.43	72.62%	82.26%	2.76
4	68.31%	85.19%	2.98	68.55%	80.88%	2.51	74.82%	82.24%	2.59	69.91%	77.17%	3.39	67.88%	76.97%	3.61
5	70.66%	75.30%	2.95	64.89%	87.20%	2.16	81.40%	85.24%	2.83	81.78%	81.18%	2.18	66.96%	75.64%	3.51
6+	63.76%	75.60%	3.37	63.55%	79.90%	3.50	72.97%	76.82%	3.65	66.15%	80.35%	3.62	67.52%	78.34%	3.81

Table 2: Impact of Function Description Length

No. of Tokens	ChatGPT-3.5			ChatGPT-4			ChatGPT-4o			Gemini			Llama3		
	EM	SSI	ED	EM	SSI	ED	EM	SSI	ED	EM	SSI	ED	EM	SSI	ED
0-5	60.14%	74.18%	2.85	79.46%	90.17%	2.46	70.75%	79.34%	2.22	72.35%	85.37%	2.49	82.71%	82.71%	2.36
6-9	77.11%	87.45%	2.66	80.14%	81.93%	2.42	83.80%	85.48%	2.07	70.96%	77.81%	2.33	79.16%	79.16%	1.75
10-13	82.03%	82.02%	2.64	85.11%	84.90%	2.26	79.97%	85.80%	2.21	70.39%	79.56%	3.59	75.21%	75.21%	3.48
14-17	66.00%	80.95%	2.69	80.65%	87.28%	2.52	86.84%	87.00%	2.03	70.19%	84.70%	3.68	80.73%	80.73%	2.37
18-21	66.66%	81.14%	2.79	62.67%	84.82%	3.24	72.05%	87.48%	2.64	76.96%	77.37%	3.16	83.35%	83.35%	2.46
22-25	68.17%	84.97%	2.70	69.08%	80.31%	2.48	73.67%	81.17%	2.43	70.03%	77.02%	3.19	77.84%	77.84%	3.22
26-29	70.32%	75.69%	2.61	64.00%	87.96%	2.19	81.91%	84.97%	2.64	82.66%	81.73%	2.05	75.30%	75.30%	3.20
30-33	62.49%	72.41%	3.25	63.38%	82.70%	3.42	74.00%	79.65%	3.51	65.04%	79.42%	3.21	80.41%	80.41%	3.60
34-37	62.48%	69.60%	2.84	62.75%	77.17%	3.57	73.46%	73.28%	3.56	66.04%	80.10%	3.08	75.45%	75.45%	3.54
40+	62.49%	85.43%	2.81	62.75%	76.41%	3.68	73.46%	78.45%	3.38	67.04%	85.27%	3.60	80.20%	80.20%	3.34

Besides the underling LLMs, we also evaluated how the length of functional description and the length of method names would influence the performance of the proposed approach. Our evaluation results suggest that the proposed approach worked well on various methods regardless of the length of method names or functional descriptions. For space limitation, details are presented online.