

Per-Type Performance: FP (Remote)											
Risk Types	①	4.0 ±0.8	1.0 ±1.4	0.0 ±0.0	2.0 ±2.8	0.0 ±0.0	2.3 ±0.5	0.0 ±0.0	6.0 ±0.8	1.7 ±1.2	0.0 ±0.0
	②	12.3 ±7.8	4.7 ±2.4	1.7 ±1.7	5.3 ±1.2	2.7 ±1.2	0.7 ±0.5	2.0 ±2.2	13.0 ±2.2	13.0 ±2.2	1.3 ±0.9
	③	0.0 ±0.0	0.3 ±0.5	0.0 ±0.0	1.0 ±0.8	0.0 ±0.0	0.0 ±0.0	0.0 ±0.0	2.3 ±1.2	0.3 ±0.5	0.0 ±0.0
	④	0.3 ±0.5	9.7 ±3.1	5.0 ±0.8	6.3 ±1.7	1.7 ±0.9	1.7 ±0.5	4.0 ±2.2	6.0 ±2.8	4.0 ±0.8	5.7 ±3.7
	⑤	8.3 ±1.2	2.0 ±1.4	0.3 ±0.5	6.3 ±1.2	0.0 ±0.0	0.0 ±0.0	0.7 ±0.9	32.0 ±4.3	2.0 ±0.8	0.3 ±0.5
	⑥	22.0 ±2.2	27.3 ±1.7	20.3 ±4.1	20.3 ±4.5	17.3 ±3.7	9.3 ±5.0	18.0 ±0.8	22.0 ±1.6	40.7 ±3.3	15.7 ±2.5
	⑦	26.3 ±3.9	16.0 ±4.9	1.0 ±0.0	38.3 ±2.1	0.0 ±0.0	7.3 ±3.4	14.7 ±4.5	70.0 ±10.7	1.3 ±1.2	22.3 ±3.9
	⑧	10.0 ±5.0	12.3 ±2.6	4.3 ±0.5	24.3 ±2.9	4.3 ±1.9	1.7 ±1.2	15.3 ±2.9	12.3 ±3.9	2.3 ±0.5	13.7 ±3.4
	⑨	51.3 ±15.4	48.3 ±9.5	25.3 ±2.9	69.7 ±4.5	27.3 ±1.9	22.0 ±3.3	68.0 ±6.2	126.7 ±5.2	37.0 ±4.2	112.7 ±3.3
		Models	claude	deepseek	doubao	gemini	glm	gpt-4o	gpt-5	grok	kimi