

SPair-71k - Model Ranks by Split

	all	1	2	3	4	5	6	7	8	9	10
Gemini 2.5 Pro	1	1	1	1	2	1	1	1	1	1	2
Gemini 2.5 Flash	2	2	1	2	1	3	2	1	1	2	1
Llama 4 Scout	3	2	3	4	3	2	4	3	3	3	3
Qwen3-VL-8B	4	4	3	3	4	5	3	4	4	5	3
GPT-4.1	5	5	5	5	5	4	5	5	5	6	8
GPT-4o	6	6	6	6	6	6	6	6	6	4	5
InternVL3-8B	7	7	8	7	7	7	7	8	8	8	6
Qwen2.5-VL-7B	8	8	7	8	8	8	9	7	7	7	6
Gemma 3-4B	9	8	9	9	9	9	8	9	9	9	9