	Validation	Test	Similarit	y Counting	Depth	Jigsaw	Art	Fun.Corr.
	(1,901)	(1,906)	(136)	(120)	(124)	(150)	(117)	(130)
Random Choice	38.09	38.09	50	25	50	50	50	25
Human	95.67	95.70	96.70	93.75	99.19	99.00	95.30	80.77
Open-source multimodal LLMs								
MiniGPT-4-v2 [16]	34.23	34.57	52.94	10.83	49.19	26.00	47.86	18.46
OpenFlamingo-v2 [5]	39.18	38.32	55.15	21.67	54.03	46.00	52.14	36.15
InstructBLIP-7B [24]	39.72	38.65	46.32	29.17	50.81	54.00	47.86	23.85
InstructBLIP-13B [24]	42.24	39.58	46.32	30.83	50.00	54.00	50.43	22.31
LLaVA-internLM2-7B [72]	37.71	36.06	52.94	52.50	52.42	34.67	30.77	23.08
Yi-VL-6B <sup>2</sup>	38.72	41.24	46.32	46.67	56.45	50.00	53.85	23.85
Yi-VL-34B <sup>2</sup>	41.68	42.78	50.00	58.33	53.23	54.00	46.15	39.23
LLaVA-v1.5-7B-xtuner [23]	39.36	40.81	46.32	53.33	50.81	54.00	47.86	23.85
LLaVA-v1.5-13B-xtuner [23]	42.00	41.31	46.32	45.00	54.03	53.33	47.86	26.15
CogVLM [77]	41.54	39.38	46.32	38.33	50.81	52.67	49.57	23.85
LLaVA-v1.5-7B [48]	37.13	38.01	46.32	43.33	51.61	11.33	47.86	21.54
LLaVA-v1.5-13B [48]	42.66	40.55	46.32	50.00	47.58	54.00	47.86	20.77
LLaVA-v1.6-34B [50]	46.80	45.05	46.32	68.33	64.52	56.67	47.01	30.77
	10.00		ased mod		01.02	00.01	11101	
Qwen-VL-Max [7]	40.28	41.94	51.47	55.83	58.87	3.33	37.61	28.46
Gemini Pro [71]	45.16	45.72	55.88	65.00	50.00	54.00	49.57	32.31
Claude 3 OPUS [1]	44.05	44.11	70.59	49.17	57.26	32.67	60.68	22.31
GPT-4V(ision) [62]	51.14	51.26	83.09	60.83	58.87	62.67		31.54
GT 1 17 (181811) [02]	01111	01.20	00.00	00.00	00.0.	o <b></b>		01.01
				is.Corr. Mu				
	Sem.Corr. (140)	Spatial (143)	Local. V (125)		lti-view (133)	Reflect (134)	. Foren	
Random Choice								2) (150)
Random Choice Human	(140) 25 96.07	(143) 50 98.25	50 98.00	(172) (25 99.42 9	(133)	(134)	(132	2) (150)
Human	(140) 25 96.07 <b>Open</b>	(143) 50 98.25 -source	50 98.00 multime	25 99.42 odal LLMs	(133) 50 92.48	(134) 33.33 95.14	25 100.0	2) (150) 25 00 80.00
	(140) 25 96.07	(143) 50 98.25	50 98.00	25 99.42 odal LLMs	50	(134) 33.33	(132	2) (150) 25 00 80.00
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5]	(140) 25 96.07 <b>Open</b>	(143) 50 98.25 -source	50 98.00 multime	25 99.42 dal LLMs 23.84	(133) 50 92.48	(134) 33.33 95.14	25 100.0	2) (150) 25 00 80.00 12 19.33
Human  MiniGPT-4-v2 [16]	(140) 25 96.07 <b>Open</b> 26.43	50 98.25 -source 51.75	50 98.00 multimo 56.00	25 99.42 94al LLMs 23.84 25.00	50 92.48 52.63	(134) 33.33 95.14 31.34	25 100.0	2) (150) 25 00 80.00 12 19.33 11 23.33
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5]	25 96.07 Open 26.43 23.57	50 98.25 -source 51.75 46.85	50 98.00 <b>multimo</b> <b>56.00</b> 52.00	25 99.42 94.42 95.00 23.84 25.00 22.67	50 92.48 52.63 41.35	(134) 33.33 95.14 31.34 43.28	25 100.0 17.4 15.9	2) (150) 25 00 80.00 12 19.33 11 23.33 15 23.33
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72]	25 96.07 <b>Open</b> 26.43 23.57 25.00	50 98.25 -source 51.75 46.85 55.24	50 98.00 multime 56.00 52.00 44.80	25 99.42 6dal LLMs 23.84 25.00 22.67 20.93	50 92.48 52.63 41.35 68.65	33.33 95.14 31.34 43.28 29.85	25 100.0 17.4 15.9 29.5	2) (150) 25 00 80.00 12 19.33 10 23.33 15 23.33 14 26.00
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72] Yi-VL-6B <sup>2</sup>	25 96.07 <b>Open</b> 26.43 23.57 25.00 22.86	50 98.25 -source 51.75 46.85 55.24 64.34	50 98.00 <b>multime</b> <b>56.00</b> 52.00 44.80 52.00	(172) (25 99.42 (27) 99.42 (27) 99.42 (27) 90.42 (28) 90.42 (28) 9	50 92.48 52.63 41.35 58.65 54.14	(134) 33.33 95.14 31.34 43.28 29.85 46.27	25 100.0 17.4 15.9 29.5 13.6	2) (150) 25 00 80.00 12 19.33 01 23.33 15 23.33 14 26.00 19 14.67
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72]	25 96.07 <b>Open</b> 26.43 23.57 25.00 22.86 22.14	50 98.25 -source 51.75 46.85 55.24 64.34 74.13	50 98.00 multimo 56.00 52.00 44.80 52.00 48.00	25 99.42 99.42 23.84 25.00 22.67 20.93 21.51 29.65	50 92.48 52.63 41.35 58.65 54.14 41.35	33.33 95.14 31.34 43.28 29.85 46.27 32.84	17.4 15.9 29.5 13.6 3.79	2) (150) 25 00 80.00 12 19.33 14 23.33 14 26.00 19 14.67 15 23.33
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72] Yi-VL-6B <sup>2</sup>	25 96.07 <b>Open</b> 26.43 23.57 25.00 22.86 22.14 26.43	50 98.25 -source 51.75 46.85 55.24 64.34 74.13 72.73	50 98.00 multimo 56.00 52.00 44.80 52.00 48.00 49.60	(172) (25 99.42 9 dal LLMs 23.84 3 25.00 4 22.67 5 20.93 3 21.51 4 29.65 4	50 92.48 52.63 41.35 68.65 54.14 41.35 48.12	33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85	25 100.0 17.4 15.9 29.5 13.6 3.79 20.4	2) (150) 25 00 80.00 12 19.33 11 23.33 15 23.33 14 26.00 19 14.67 15 23.33 12 22.67
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72] Yi-VL-6B <sup>2</sup> Yi-VL-34B <sup>2</sup>	25 96.07 <b>Open</b> 26.43 23.57 25.00 22.86 22.14 26.43 21.43	50 98.25 -source 51.75 46.85 55.24 64.34 74.13 72.73 70.63	50 98.00 <b>multime</b> <b>56.00</b> 52.00 44.80 52.00 48.00 49.60 54.40	(172) (172) (25 99.42 9.42 9.42 9.42 9.42 9.42 9.42 9.4	50 92.48 52.63 41.35 58.65 54.14 41.35 48.12 41.35	33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85 46.27	25 100.0 17.4 15.9 29.5 13.6 3.79 20.4 17.4	2) (150) 25 00 80.00 12 19.33 11 23.33 15 23.33 14 26.00 9 14.67 15 23.33 12 22.67 16 21.33
MiniGPT-4-v2 [16]   OpenFlamingo-v2 [5]   InstructBLIP-7B [24]   InstructBLIP-13B [24]   LLaVA-internLM2-7B [72]   Yi-VL-6B 2   Yi-VL-34B 2   LLaVA-v1.5-7B-xtuner [23]	25 96.07 <b>Open</b> 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29	50 98.25 -source 51.75 46.85 55.24 64.34 74.13 72.73 70.63 74.83	50 98.00 multime 56.00 52.00 44.80 52.00 48.00 49.60 54.40 45.60	(172) ( 25 99.42 9 dal LLMs 23.84 25.00 22.67 5 20.93 21.51 4 22.965 4 23.84 4 23.84 4 22.09 4	50 92.48 52.63 41.35 58.65 54.14 41.35 48.12 41.35 42.11	33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85 46.27 26.87	25 100.0 17.4 15.9 29.5 13.6 3.79 20.4 17.4 36.3	2) (150) 25 00 80.00 22 19.33 11 23.33 15 23.33 14 26.00 14.67 15 23.33 12 22.67 16 21.33 15 18.67
Human	25 96.07 <b>Open</b> 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29 22.14	50 98.25 -source 51.75 46.85 55.24 64.34 74.13 72.73 70.63 74.83 77.62	50 98.00 multime 56.00 52.00 44.80 52.00 48.00 49.60 54.40 45.60 48.00	(172) ( 25 99.42 9 dal LLMs 23.84 25.00 22.67 5 20.93 3 21.51 4 23.84 4 23.84 4 22.09 4 20.93 3	50 92.48 52.63 41.35 58.65 54.14 41.35 48.12 41.35 42.11 41.35	33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85 46.27 26.87 46.27	25 100.0 17.4 15.9 29.5 13.6 3.79 20.4 17.4 36.3 29.5	2) (150) 25 00 80.00 22 19.33 11 23.33 15 23.33 14 26.00 19 14.67 19 23.33 19 23.33 19 23.33 19 23.33 19 23.33 19 23.33 19 24.67 19 26.67
Human   MiniGPT-4-v2 [16]     OpenFlamingo-v2 [5]     InstructBLIP-7B [24]     InstructBLIP-13B [24]     LLaVA-internLM2-7B [72]     Yi-VL-6B   <sup>2</sup>     Yi-VL-34B   <sup>2</sup>     LLaVA-v1.5-7B-xtuner [23]     LLaVA-v1.5-13B-xtuner [23]     CogVLM [77]	25 96.07 <b>Open</b> 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29 22.14 23.57	(143) 50 98.25 -source 51.75 46.85 55.24 64.34 74.13 70.63 74.83 77.62 67.13	50 98.00 multime 56.00 52.00 44.80 52.00 48.00 49.60 45.60 48.00 43.20	(172) ( 25 99.42 9  dal LLMs 23.84 5 25.00 6 22.67 5 20.93 21.51 6 23.84 6 23.84 6 23.84 6 23.84 6 23.84 6 23.84 6 23.84 6 23.85 6 20.93 6 20.93 6 20.93 6	50 52.48 52.63 52.63 54.14 54.13 54.12 54.13 54.11 57.14	33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85 46.27 26.87 46.27 26.87	25 100.0 17.4 15.9 29.5 13.6 3.7; 20.4 17.4 36.3 29.5 24.2	2) (150) 25 00 80.00 12 19.33 14 23.33 15 23.33 16 26.00 19 14.67 15 23.33 12 22.67 16 21.33 16 21.33 17 26.67 18 26.67 18 26.67 18 26.67 19 24.00
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72] Yi-VL-6B <sup>2</sup> Yi-VL-34B <sup>2</sup> LLaVA-v1.5-7B-xtuner [23] LLaVA-v1.5-13B-xtuner [23] CogVLM [77] LLaVA-v1.5-7B [48]	25 96.07 Open 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29 22.14 23.57 32.14	50 98.25 -source 51.75 46.85 55.24 64.34 74.13 70.63 74.83 77.62 67.13 70.63	50 98.00 multimo 56.00 52.00 44.80 52.00 48.00 49.60 54.40 45.60 48.00 43.20 48.80	(172) ( 25 99.42 9  dal LLMs 23.84 3 25.00 4 22.67 5 20.93 3 21.51 4 29.65 4 23.84 4 22.09 4 22.09 3 20.35 4 20.35	50 52.48 52.63 52.63 52.63 54.14 11.35 54.14 11.35 42.11 41.35 57.14 49.62	(134) 33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85 46.27 26.87 46.27 26.87 36.57	25 100.0 17.4 15.9 29.5 13.6 3.7 20.4 17.4 36.3 29.5 24.2 28.0	2) (150) 25 00 80.00 12 19.33 11 23.33 14 26.00 15 23.33 15 23.33 16 22.67 16 21.33 18.67 14 26.67 15 23.33 17 22.67 18 21.33 18 22.67 18 26.67 18 24.00 17 28.00
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72] Yi-VL-6B <sup>2</sup> Yi-VL-34B <sup>2</sup> LLaVA-v1.5-7B-xtuner [23] LLaVA-v1.5-13B-xtuner [23] CogVLM [77] LLaVA-v1.5-7B [48] LLaVA-v1.5-13B [48]	25 96.07 Open 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29 22.14 23.57 32.14 23.57	50 98.25 -source 51.75 46.85 55.24 64.34 74.13 72.73 70.63 74.83 70.63 67.13 70.63 67.83 76.22	50 98.00 multimo 56.00 52.00 44.80 52.00 48.00 49.60 54.40 45.60 48.00 48.00 48.00 48.00 48.00 49.60 54.40 45.60 48.00 48.00 48.00 48.00 48.00	(172) ( 25 99.42 9  dal LLMs 23.84 3 25.00 4 22.67 5 20.93 3 21.51 4 29.65 4 23.84 4 23.84 4 22.09 4 20.93 3 20.35 4 20.35 4 27.33 4	50 92.48 52.63 41.35 58.65 54.14 41.35 48.12 41.35 42.11 41.35 57.14 49.62 41.35	(134) 33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85 46.27 26.87 46.27 26.87 36.57 45.52	255 100.0 17.4 15.9 29.5 13.6 3.7; 20.4 17.4 36.3 29.5 24.2 28.0 27.2	2) (150) 25 00 80.00 12 19.33 11 23.33 14 26.00 15 23.33 15 23.33 16 22.67 16 21.33 18.67 14 26.67 15 23.33 17 22.67 18 21.33 18 22.67 18 26.67 18 24.00 17 28.00
Human  MiniGPT-4-v2 [16] OpenFlamingo-v2 [5] InstructBLIP-7B [24] InstructBLIP-13B [24] LLaVA-internLM2-7B [72] Yi-VL-6B <sup>2</sup> Yi-VL-34B <sup>2</sup> LLaVA-v1.5-7B-xtuner [23] LLaVA-v1.5-13B-xtuner [23] CogVLM [77] LLaVA-v1.5-7B [48] LLaVA-v1.5-13B [48]	25 96.07 Open 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29 22.14 23.57 32.14 23.57	50 98.25 -source 51.75 46.85 55.24 64.34 74.13 72.73 70.63 74.83 70.63 67.13 70.63 67.83 76.22	50 98.00 multime 56.00 52.00 44.80 52.00 48.00 49.60 54.40 45.60 48.00 43.20 48.80 47.20 41.60	(172) ( 25 99.42 9  dal LLMs 23.84 3 25.00 4 22.67 5 20.93 3 21.51 4 29.65 4 23.84 4 22.09 4 20.93 3 20.93 3 20.35 4 27.33 4  dels	50 92.48 52.63 41.35 58.65 54.14 41.35 48.12 41.35 42.11 41.35 57.14 49.62 41.35	(134) 33.33 95.14 31.34 43.28 29.85 46.27 32.84 29.85 46.27 26.87 46.27 26.87 36.57 45.52	255 100.0 17.4 15.9 29.5 13.6 3.7° 20.4 17.4 36.3 29.5 24.2 24.2 27.2 41.6	2) (150) 25 00 80.00 12 19.33 11 23.33 15 23.33 14 26.00 19 14.67 15 23.33 12 22.67 16 21.33 15 18.67 14 26.67 14 26.67 14 26.67 15 23.33 17 28.00 18 28.00 18 28.00
Human   MiniGPT-4-v2 [16]   OpenFlamingo-v2 [5]   InstructBLIP-7B [24]   InstructBLIP-13B [24]   LLaVA-internLM2-7B [72]   Yi-VL-6B   2   Yi-VL-34B   2   LLaVA-v1.5-7B-xtuner [23]   LLaVA-v1.5-13B-xtuner [23]   CogVLM [77]   LLaVA-v1.5-13B [48]   LLaVA-v1.6-34B [50]   LLaVA-v1.6-34B [50]     CogVLM [50]   C	25 96.07 Open 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29 22.14 23.57 32.14 23.57 32.14 23.57 27.86	50 98.25 -source 51.75 46.85 55.24 64.34 72.73 70.63 74.83 70.63 76.21 67.13 70.63 70.63	50 98.00 multimo 56.00 52.00 44.80 52.00 48.00 49.60 54.40 45.60 48.00 49.60 40.00 40.	(172) (172) (25 99.42 9.42 9.42 9.42 9.42 9.42 9.42 9.4	50 92.48 52.63 41.35 88.65 54.14 41.35 48.12 41.35 42.11 41.35 42.11 41.35 42.11 41.35 42.11 41.35 42.11 41.35	(134) 33.33 95.14 31.34 43.28 29.85 46.27 26.87 46.27 26.87 46.27 29.85	255 100.0 17.4 15.9 29.5 13.6 3.7° 20.4 17.4 36.3 29.5 24.2 24.2 27.2 41.6	2) (150) 25 00 80.00 12 19.33 11 23.33 15 23.33 14 26.00 19 14.67 15 23.33 12 22.67 16 21.33 15 18.67 14 26.67 14 26.67 17 28.00 17 26.00
Human	25 96.07 Open 26.43 23.57 25.00 22.86 22.14 26.43 21.43 24.29 22.14 23.57 32.14 23.57 27.86	143) 50 98.25 -source 51.75 46.85 55.24 64.34 74.13 70.63 74.83 70.63 67.83 76.22 API-b 77.62	50 98.00 multimo 56.00 52.00 44.80 52.00 48.00 49.60 54.40 45.60 48.00 49.60 40.00 40.	(172) (172) (25 99.42 9 99.42 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	(133) 50 92.48 52.63 41.35 88.65 54.14 41.35 42.11 41.35 42.11 41.35 44.13 49.62 41.35 46.62	(134) 33.33 95.14 31.34 43.28 29.85 46.27 26.87 46.27 26.87 45.52 29.85 49.25	25 100.0 17.4 15.9 29.5 13.6 3.73 20.4 17.4 36.3 29.5 24.2 28.0 27.2 41.6	2) (150) 25 00 80.00 22 19.33 11 23.33 14 26.00 14.67 15 23.33 12 22.67 14 26.67 13 24.00 17 28.00 17 26.00 18 22.07 19 26.00 10 21.33 10 21