

Chess				Chemistry			
GPT-5	0.710	0.746	0.734	0.910	0.758	0.933	
GPT-5-mini	0.776	0.786	0.759	0.846	0.779	0.834	
GPT-5-nano	0.743	0.659	0.731	0.739	0.528	0.744	
GPT-4o	0.624	0.644	0.624	0.650	0.573	0.653	
GPT-4o-mini	0.610	0.646	0.634	0.529	0.409	0.489	
Claude-4.1-Opus	0.806	0.718	0.794	0.945	0.851	0.946	
Claude-4-Sonnet	0.799	0.734	0.784	0.891	0.704	0.904	
Claude-3.7-Sonnet	0.706	0.656	0.675	0.888	0.803	0.870	
Claude-3.5-Sonnet	0.652	0.615	0.651	0.836	0.813	0.688	
Claude-3.5-Haiku	0.623	0.549	0.635	0.574	0.529	0.589	
Qwen2.5-VL-72B-Instruct	0.542	0.586	0.571	0.575	0.559	0.609	
Qwen2.5-VL-7B-Instruct	0.295	0.389	0.386	0.296	0.443	0.410	
Qwen2.5-Omni-7B	0.501	0.512	0.448	0.310	0.307	0.338	
InternVL3-78B	0.549	0.549	0.560	0.468	0.446	0.505	
InternVL3-8B	0.420	0.431	0.454	0.372	0.289	0.364	
InternVL-2.5-78B	0.552	0.631	0.635	0.448	0.404	0.460	
InternVL-2.5-8B	0.501	0.466	0.450	0.219	0.328	0.300	
Llama-3.2-90B-Vision-Instruct	0.495	0.538	0.535	0.501	0.363	0.551	
Llama-3.2-11B-Vision-Instruct	0.312	0.446	0.415	0.304	0.294	0.340	
gemma-3-27b-it	0.522	0.545	0.517	0.529	0.499	0.515	
gemma-3-12b-it	0.454	0.490	0.476	0.463	0.446	0.509	
	Language	Vision	Vision Language	Language	Vision	Vision Language	

Music				Graph			
GPT-5	0.806	0.343	0.764	0.791	0.684	0.999	
GPT-5-mini	0.736	0.348	0.730	0.788	0.699	0.996	
GPT-5-nano	0.548	0.328	0.549	0.768	0.525	0.990	
GPT-4o	0.540	0.328	0.495	0.725	0.382	0.738	
GPT-4o-mini	0.348	0.241	0.307	0.735	0.347	0.688	
Claude-4.1-Opus	0.581	0.348	0.580	0.974	0.394	0.936	
Claude-4-Sonnet	0.569	0.328	0.555	0.974	0.414	0.971	
Claude-3.7-Sonnet	0.509	0.409	0.431	0.868	0.496	0.739	
Claude-3.5-Sonnet	0.431	0.365	0.344	0.739	0.446	0.473	
Claude-3.5-Haiku	0.361	0.301	0.290	0.561	0.353	0.471	
Qwen2.5-VL-72B-Instruct	0.425	0.341	0.373	0.647	0.414	0.524	
Qwen2.5-VL-7B-Instruct	0.282	0.260	0.280	0.340	0.307	0.361	
Qwen2.5-Omni-7B	0.289	0.279	0.306	0.351	0.318	0.364	
InternVL3-78B	0.411	0.285	0.361	0.574	0.430	0.502	
InternVL3-8B	0.302	0.289	0.278	0.434	0.419	0.448	
InternVL-2.5-78B	0.319	0.236	0.283	0.471	0.386	0.460	
InternVL-2.5-8B	0.245	0.220	0.221	0.331	0.336	0.365	
Llama-3.2-90B-Vision-Instruct	0.331	0.269	0.276	0.409	0.367	0.393	
Llama-3.2-11B-Vision-Instruct	0.264	0.245	0.253	0.275	0.335	0.283	
gemma-3-27b-it	0.430	0.330	0.354	0.581	0.339	0.414	
gemma-3-12b-it	0.395	0.280	0.316	0.521	0.386	0.416	
	Language	Vision	Vision Language	Language	Vision	Vision Language	

