Be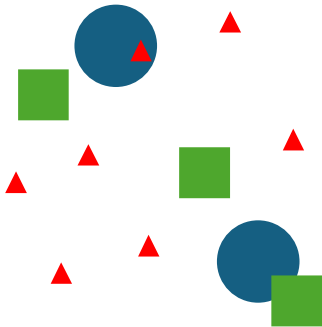nchmark prompts with adjusted weights → New LLM ranking