🏛 Finance Agent Benchmark

VA vals.ai

□ Anthropic Best[1] vs. □ OpenAI Best[2]

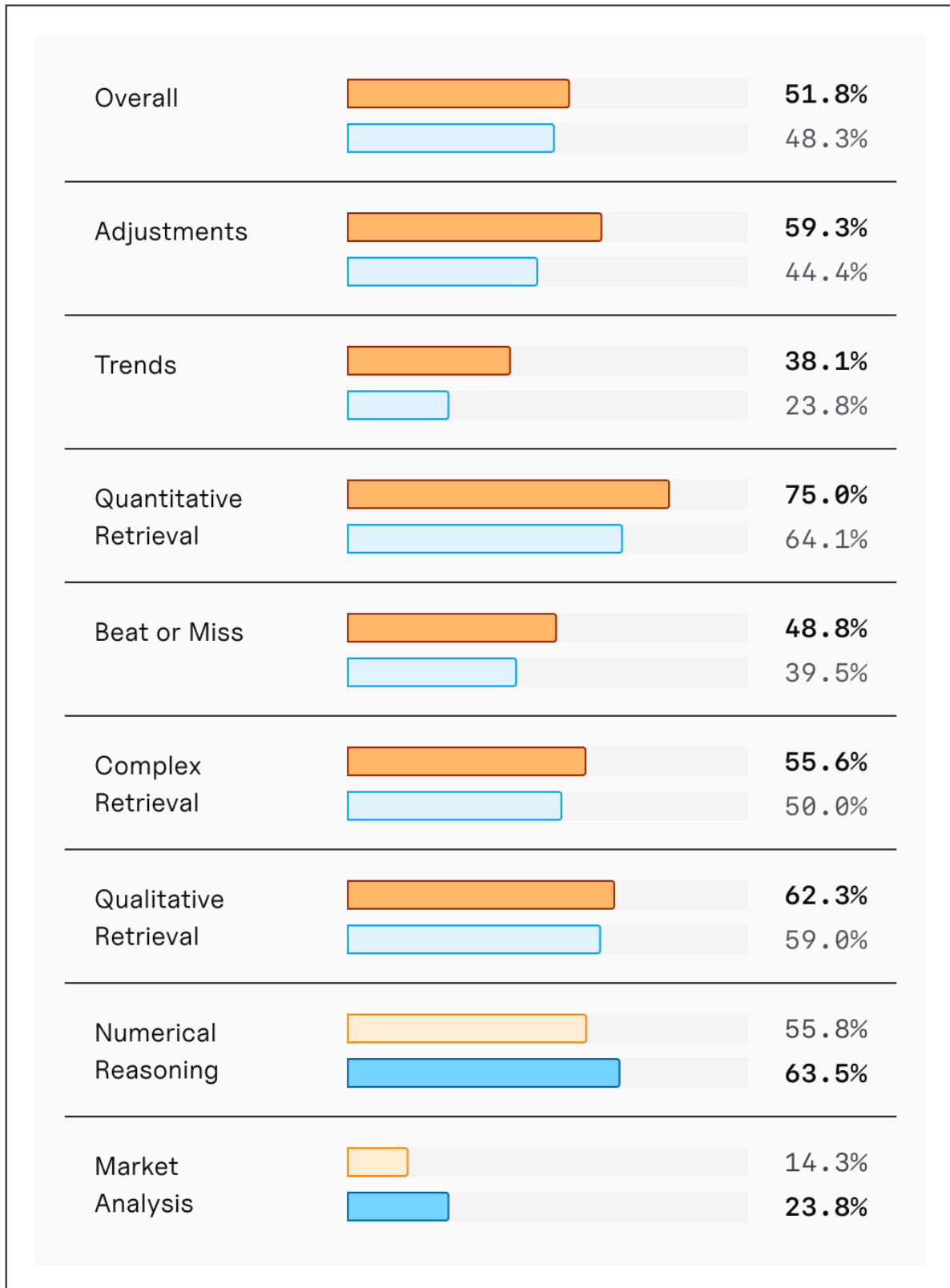| Category	Anthropic Best	OpenAI Best
Overall	**51.8%**	48.3%
Adjustments	**59.3%**	44.4%
Trends	**38.1%**	23.8%
Quantitative Retrieval	**75.0%**	64.1%
Beat or Miss	**48.8%**	39.5%
Complex Retrieval	**55.6%**	50.0%
Qualitative Retrieval	**62.3%**	59.0%
Numerical Reasoning	55.8%	**63.5%**
Market Analysis	14.3%	**23.8%**

1. *Anthropic Best looks at the top score across Claude models: Opus 4, Sonnet 4, and Sonnet 3.7 (thinking and non-thinking)*

2. *OpenAI Best looks at the top score across OpenAI model: o3, o4-mini, o1, o1-pro, GPT-4.1, and GPT-4o*