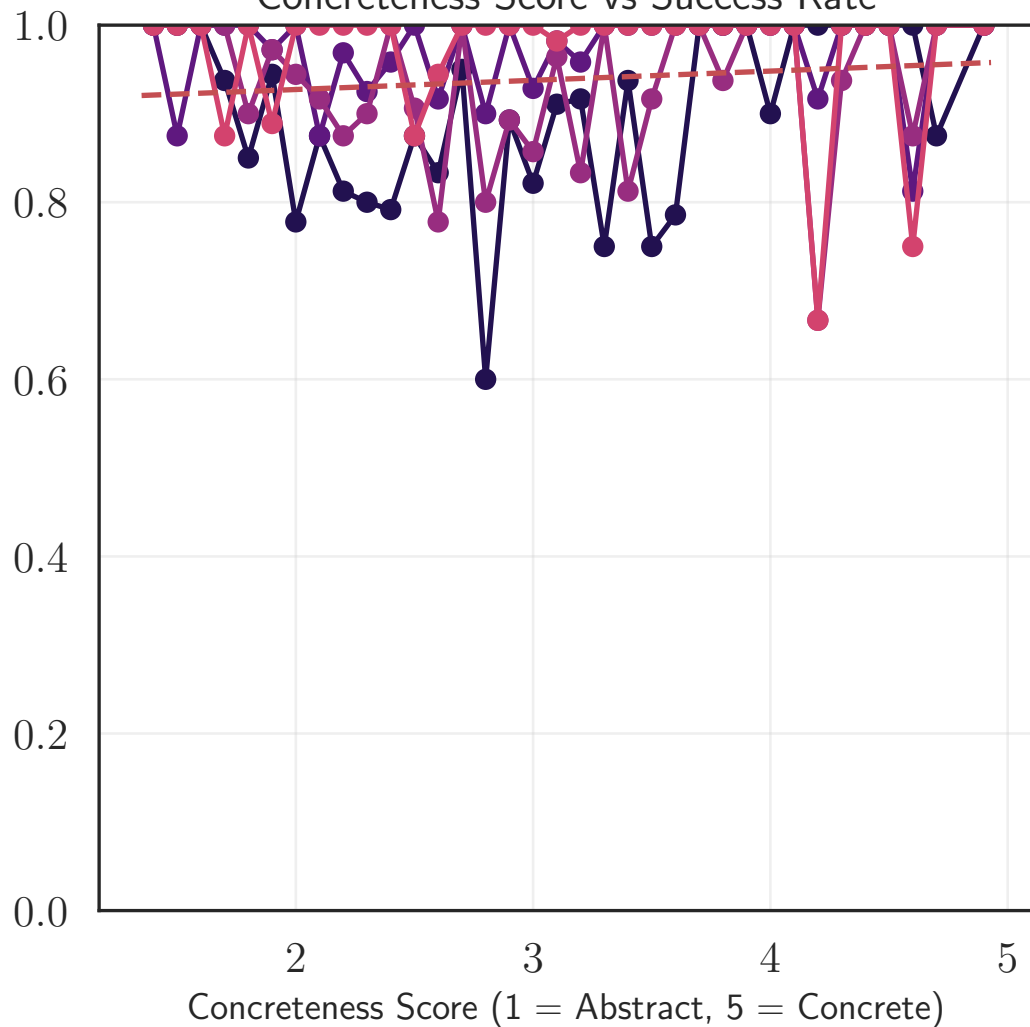


Concreteness Score vs Success Rate

Success Rate



- gpt-4o
- gemini-2.5-pro
- deepseek-v3
- claude-sonnet-4
- Trend ($r=0.038$, $p=0.0375$)