Model Comparison

- **68** GPT-3.5-turbo ReAct
- **22**
- **17**
- **5**
- **93** GPT-3.5-turbo WorldMind
- **12**
- **7**
- **73** GPT-4.1-mini ReAct
- **27**
- **12**
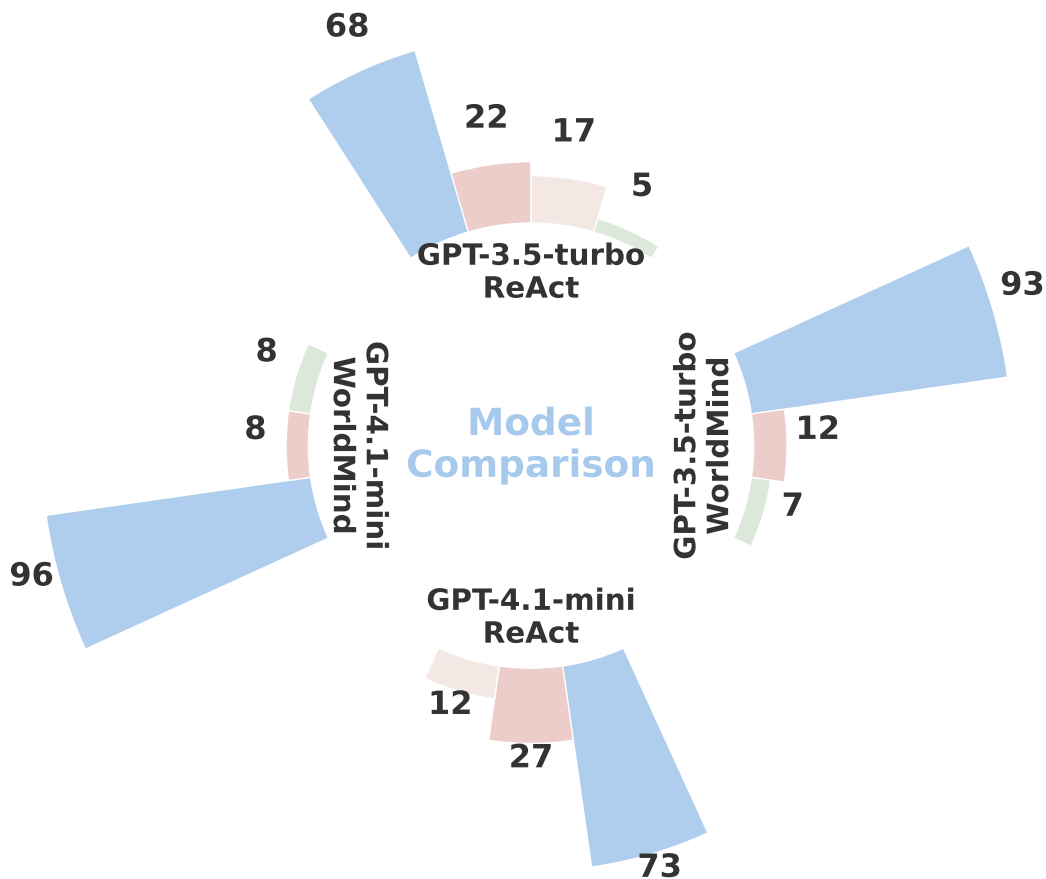- **96** GPT-4.1-mini WorldMind
- **8**
- **8**

Legend:
- Max Steps Reached
- Step Execution Error
- Repeated Web Actions
- Repeated Embodied Actions