Lesson learned - Models matter!

Excellent results with current full-size models from OpenAI, Anthropic, and Google.

Other models struggled:

- Did the inverse of the user request [GPT-4o-mini]

- Forgot to ask the dashboard to filter [DeepSeek v3]

- Wrote malformed JSON [Llama 3 8B Instruct]

- Struggled to understand user request [Gemini 2.0 Flash Lite]

- Called the wrong tool [GPT-4.1-nano]

These problems were mostly fixable via tweaking the instructions … but they still needed help!

Lesson learned - Models matter!

Excellent results with current full-size models from OpenAI, Anthropic, and Google.

Other models struggled:

- Did the inverse of the user request [GPT-4o-mini]

- Forgot to ask the dashboard to filter [DeepSeek v3]

- Wrote malformed JSON [Llama 3 8B Instruct]

- Struggled to understand user request [Gemini 2.0 Flash Lite]

- Called the wrong tool [GPT-4.1-nano]

These problems were mostly fixable via tweaking the instructions
… but they still needed help!

Learn more

- YouTube: "Harnessing LLMs for Data Analysis" by Joe Cheng

- `ellmer`:   Easily call LLMs from R
  `chatlas`: Easily call LLMs from Python

- Shiny: Easy web apps for Data Science

- querychat: Enhance Shiny data dashboards with LLMs that speak SQL

Slides:

Barret Schloerke
posit™ / Shiny Team
@schloerke