

How bad are LLMs with data?

Example: ℓ_1

```
# Make an array of random numbers, of length n  
values = np.random.rand(n).tolist()
```

```
client = chatlas.ChatOpenAI(model="gpt-4.1")  
client.chat('How long is this array?', json.dumps(values))
```

• $n=10$, LLM says: 10 

• $n=100$, LLM says: 100 

• $n=10000$, LLM says: 10000 

• $n \equiv 10,000, LLM$ says: 10000 ✖

• $n=10^3$, LLM says: 1000 ✗

How bad are LLMs with data?

Example: len()

```
# Make an array of random numbers, of length n
values = np.random.rand(n).tolist()

client = chatlas.ChatOpenAI(model="gpt-4.1")
client.chat("How long is this array?", json.dumps(values))
```

- n=10, LLM says: 10 
- n=100, LLM says: 100 
- n=1000, LLM says: 1000 
- n=10,000, LLM says: 1000 
- n=103, LLM says: 100 

Can LLMs be **responsibly used** for data science?