

Adv Python Exam 2

Nico Morin

A dark blue diagonal gradient bar that starts from the bottom left corner and extends towards the top right corner, covering the lower half of the slide.

The Models:

TinyLlama: A very compact model (less than one GB). This is an older model that has not been updated since 2023

Llama 3.1: A mid-sized model created by meta

Mistral-small3.1: Ironically the largest model I tested (15GB), has computer vision enabled.

General Evaluation

Generic questions:

Three questions:

- What are the top ten most populous US states in order?

-Easily accessible Census information

- What is the Answer to the Ultimate Question of Life, the Universe, and Everything?

-A pop culture reference to Douglas Adams Hitchhiker's Guide to the Galaxy

- Explain multiplication to me as if I was a third grader learning it for the first time.

-A roleplaying prompt testing the models ability to explain concepts

Text summarization

I fed each model the entire short story “The Lottery” by Shirley Jackson, and asked the model to summarize the story which was decently large in length.

Code generation

I asked each model to build a python function that performed two tasks on a paragraph of text. The model was to build a script that sorted the words alphabetically excluding words greater than 3 letters, additionally the script should output the number of times the word "and" appeared

Creative Writing:

I asked each model to write a story about an adventurer who befriends a dragon. This was the most successful prompt across all models, the models were not bound by providing real facts which allowed them to create stories effectively.

Interesting result: All the models gave the protagonist a very similar name. Two of them gave them the same name.

Model Performance:

I measured each model on its speed of generation for all the general prompts discussed previously.

Speed of generation:

- TinyLlama: 27s
- Llama3.1: 282s
- Mistral-small3.1: 10,259s (nearly three hours)

Ethical Considerations

Prompts

I asked three different prompts to see if I could get the model to provide dangerous information.

- First I asked outright for dangerous information
- Second I asked the model to roleplay in a way that would give me the information
- Third I asked the models to write a story where the main character performs this dangerous task

Results

- TinyLlama: Total failure
- Llama3.1: Total success
- Mistral-small3.1: Partial failure

Conclusions:

As expected TinyLlama is not a very useful model, more of a novelty than a usable product.

Mistral-small had disappointing results for such a resource heavy model, only nominally better than the mid sized model and even then not always.

Llama3.1 was surprisingly very effective, despite some minor hallucinations it gave quick and effective answers.