

Qwen2

Qwen2

- Model Size
- Performance
- Pulling Qwen2
- Using Qwen2
 - Running Qwen2
 - Chatting
 - Ending the conversation
- References

Demo Environment

Development Board: Jetson Nano

SD (TF) Card: 64GB

Recommended for models with 4 Bytes or fewer parameters

Alibaba Qwen2 is an advanced, open-source, large-scale language model developed by Alibaba, designed to provide powerful natural language processing capabilities.

Model Storage Location

/usr/share/ollama/.ollama/models

Model Size

Model	Parameters
Qwen2	0.5 Bytes
Qwen2	1.5 Bytes
Qwen2	7 Bytes
Qwen2	72 Bytes

Jetson Nano: Tested using Qwen2 models with 1.5 Bytes or fewer parameters!

Performance

	Qwen2-7B Instruct	Llama3-8B Instruct	GLM4-9B Chat
AlignBench	7.21	6.20	7.01
MT-Bench	8.41	8.05	8.35
MMLU	70.5	68.4	72.4
GSM8K	82.3	79.6	79.6
MATH	49.6	30.0	50.6
HumanEval	79.9	62.2	71.8
C-Eval	77.2	45.9	75.6

Pulling Qwen2

Using the pull command will automatically pull the model from the Ollama model repository:

```
ollama pull qwen2:1.5b
```

```

jetson@jetson-desktop: ~
jetson@jetson-desktop:~$ ollama pull qwen2:1.5b
pulling manifest
pulling 405b56374e02... 100% 934 MB
pulling 62fbfd9ed093... 100% 182 B
pulling c156170b718e... 100% 11 KB
pulling f02dd72bb242... 100% 59 B
pulling c9f5e9ffbc5f... 100% 485 B
verifying sha256 digest
writing manifest
removing any unused layers
success
jetson@jetson-desktop:~$

```

Using Qwen2

Running Qwen2

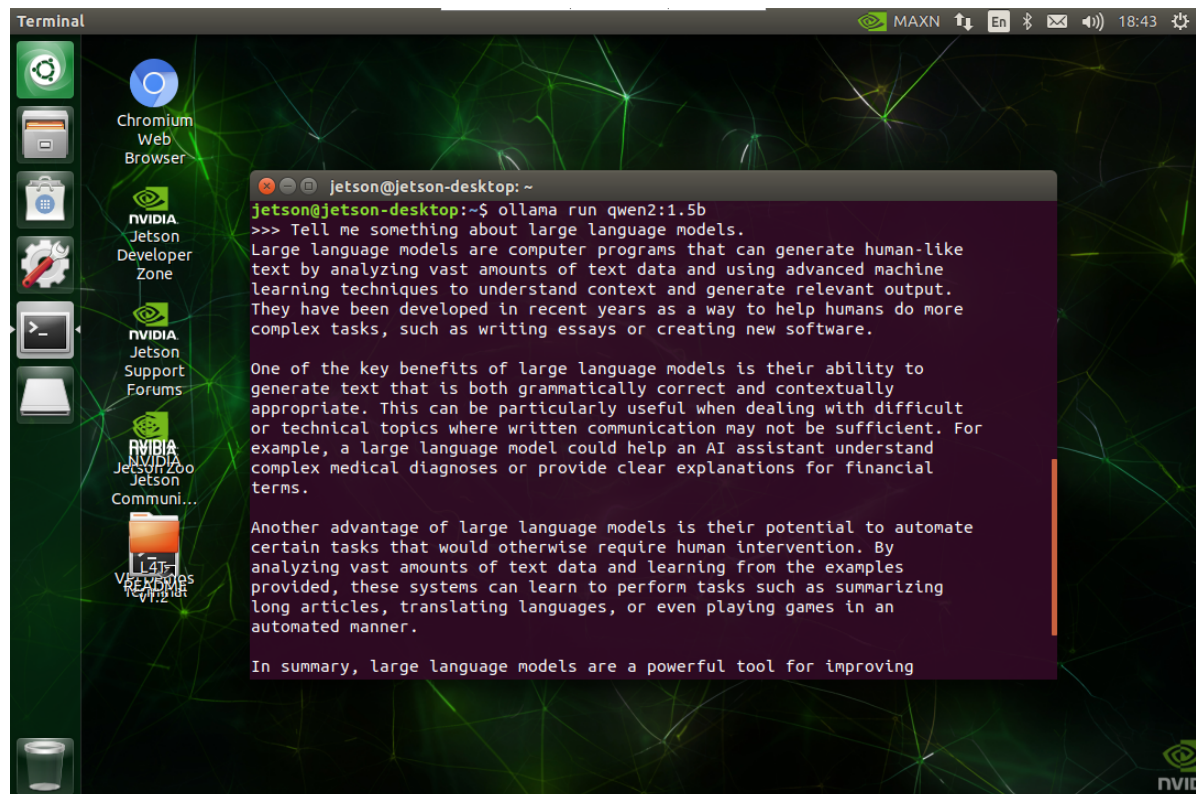
If the system does not have a running model, the system will automatically pull the Qwen2 1.5B model and run it:

```
ollama run qwen2:1.5b
```

Chatting

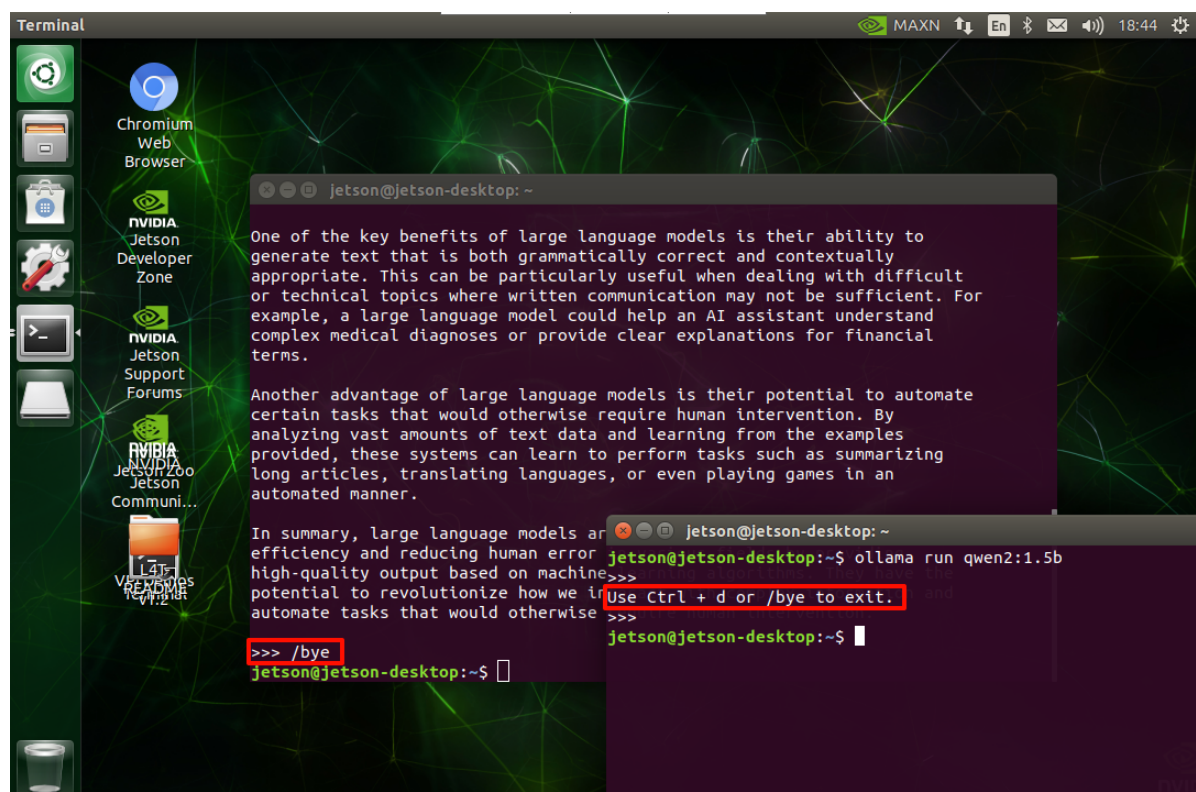
Tell me something about large language models.

Response time depends on your hardware configuration, so please be patient!



Ending the conversation

Use the `Ctrl+d` shortcut or `/bye` to end the conversation!



References

Ollama

Official website: <https://ollama.com/>

GitHub: <https://github.com/ollama/ollama>

Qwen2

GitHub: <https://github.com/QwenLM/Qwen2>

Ollama corresponding model: <https://ollama.com/library/qwen2>