

TinyLlama

TinyLlama

[Model size](#)

[Pull TinyLlama](#)

[Use TinyLlama](#)

[Run TinyLlama](#)

[Have a conversation](#)

[End the conversation](#)

[References](#)

Demo Environment

Development board: Jetson Orin series motherboard

SSD: 128G

Tutorial application scope: Whether the motherboard can run is related to the available memory of the system. The user's own environment and the programs running in the background may cause the model to fail to run

Motherboard model	Ollama	Open WebUI
Jetson Orin NX 16GB	√	√
Jetson Orin NX 8GB	√	√
Jetson Orin Nano 8GB	√	√
Jetson Orin Nano 4GB	√	√

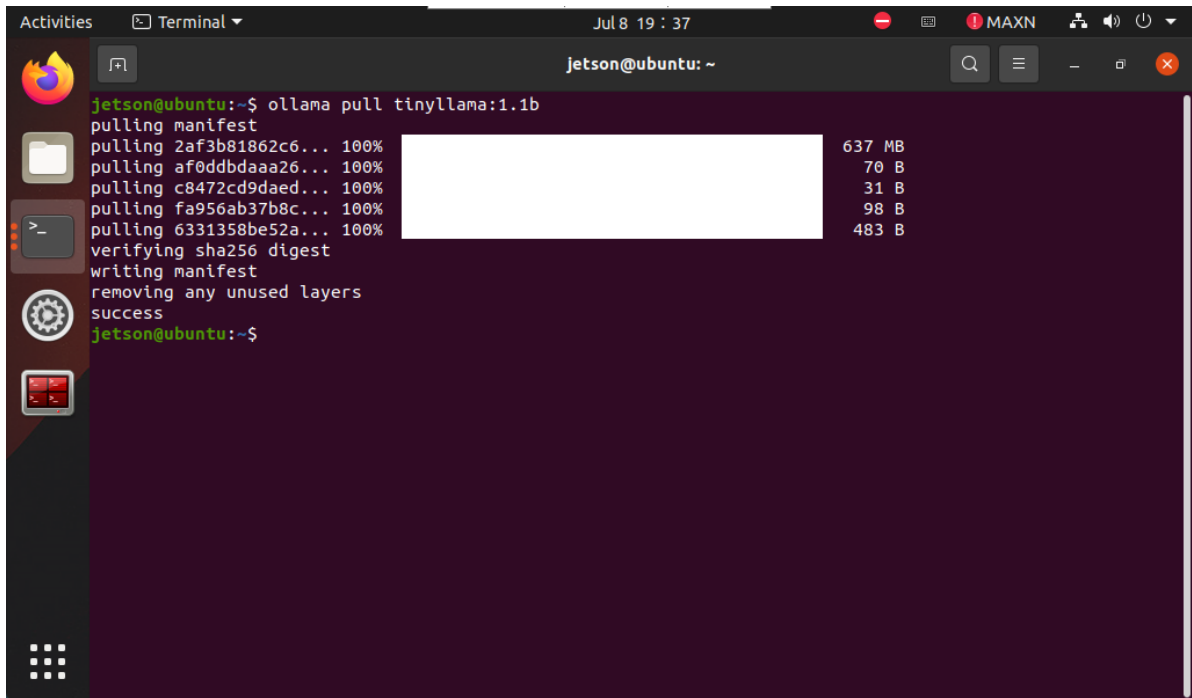
Model size

Model	Parameters
TinyLlama	1.1B

Pull TinyLlama

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

```
ollama pull tinyllama:1.1b
```

A terminal window on a Linux system (jetson@ubuntu) showing the execution of the command 'ollama pull tinyllama:1.1b'. The output shows the model being pulled in layers, with progress bars and sizes (637 MB, 70 B, 31 B, 98 B, 483 B). The process completes successfully, and the prompt returns to the shell.

```
jetson@ubuntu:~$ ollama pull tinyllama:1.1b
pulling manifest
pulling 2af3b81862c6... 100% 637 MB
pulling af0ddbdaaa26... 100% 70 B
pulling c8472cd9daed... 100% 31 B
pulling fa956ab37b8c... 100% 98 B
pulling 6331358be52a... 100% 483 B
verifying sha256 digest
writing manifest
removing any unused layers
success
jetson@ubuntu:~$
```

Use TinyLlama

Run TinyLlama

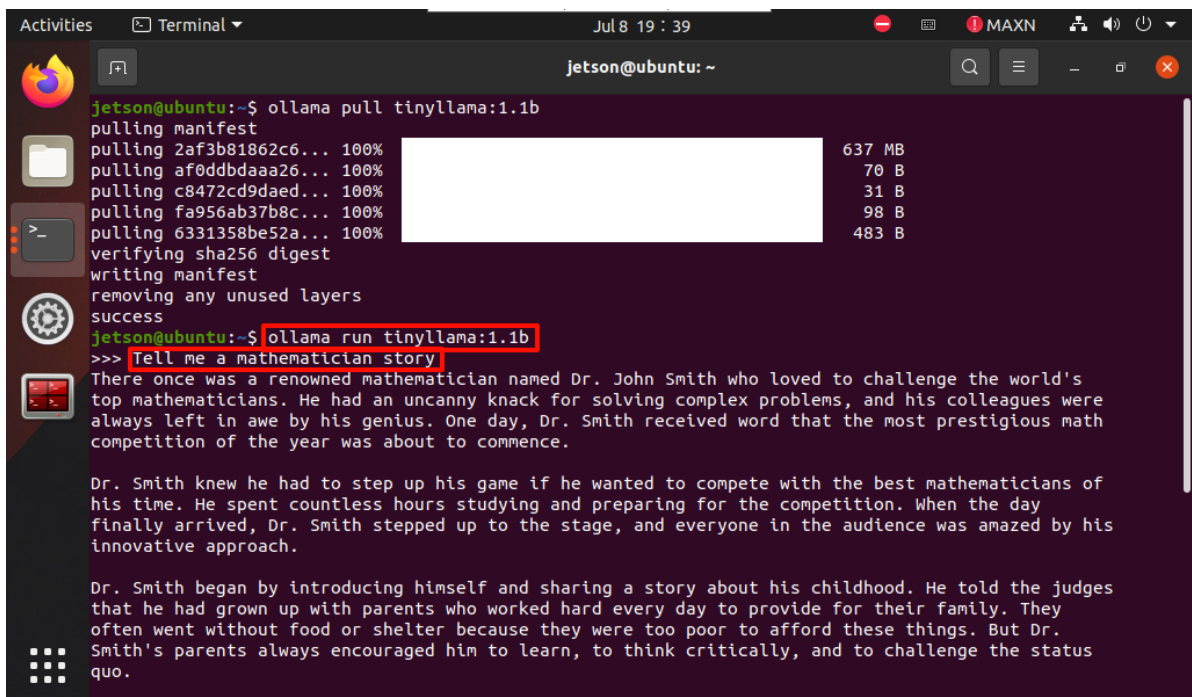
If the system does not have a running model, the system will automatically pull the TinyLlama 1.1B model and run it:

```
ollama run tinyllama:1.1b
```

Have a conversation

```
Tell me a mathematician story
```

The time to reply to the question depends on the hardware configuration, please be patient!

A terminal window on a Linux system (jetson@ubuntu) showing the execution of the command 'ollama run tinyllama:1.1b'. The output shows the model being pulled in layers, with progress bars and sizes (637 MB, 70 B, 31 B, 98 B, 483 B). The process completes successfully, and the prompt returns to the shell. The user then enters the prompt 'Tell me a mathematician story', and the model generates a story about Dr. John Smith.

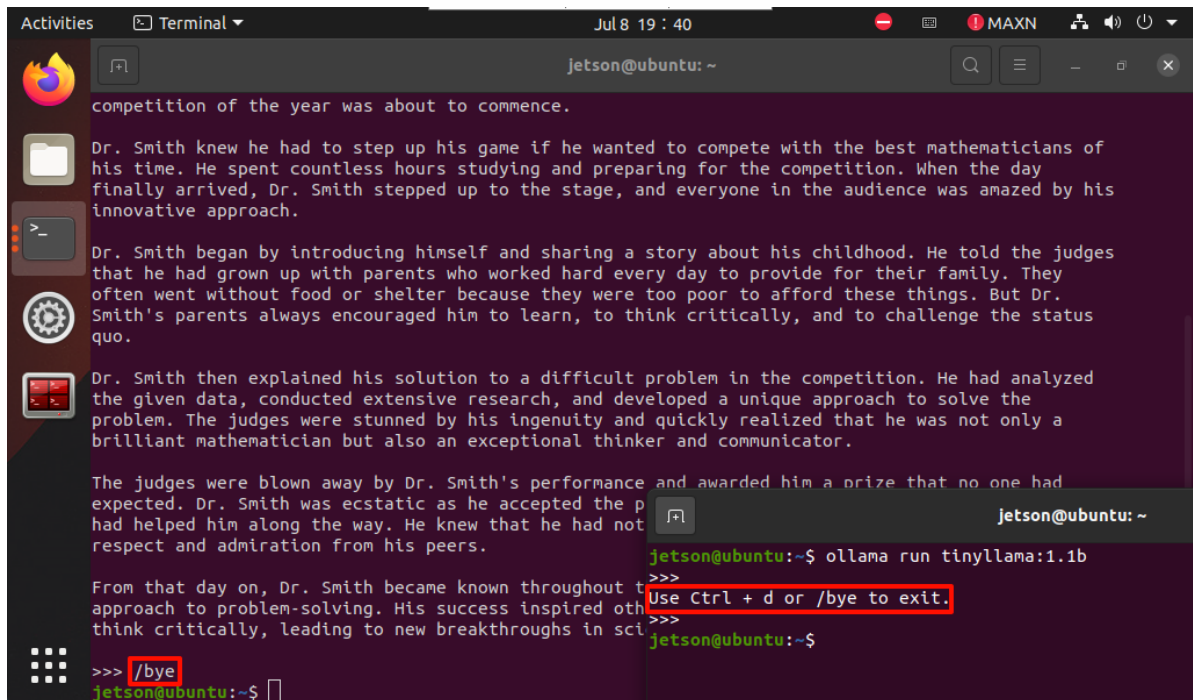
```
jetson@ubuntu:~$ ollama pull tinyllama:1.1b
pulling manifest
pulling 2af3b81862c6... 100% 637 MB
pulling af0ddbdaaa26... 100% 70 B
pulling c8472cd9daed... 100% 31 B
pulling fa956ab37b8c... 100% 98 B
pulling 6331358be52a... 100% 483 B
verifying sha256 digest
writing manifest
removing any unused layers
success
jetson@ubuntu:~$ ollama run tinyllama:1.1b
>>> Tell me a mathematician story
There once was a renowned mathematician named Dr. John Smith who loved to challenge the world's top mathematicians. He had an uncanny knack for solving complex problems, and his colleagues were always left in awe by his genius. One day, Dr. Smith received word that the most prestigious math competition of the year was about to commence.

Dr. Smith knew he had to step up his game if he wanted to compete with the best mathematicians of his time. He spent countless hours studying and preparing for the competition. When the day finally arrived, Dr. Smith stepped up to the stage, and everyone in the audience was amazed by his innovative approach.

Dr. Smith began by introducing himself and sharing a story about his childhood. He told the judges that he had grown up with parents who worked hard every day to provide for their family. They often went without food or shelter because they were too poor to afford these things. But Dr. Smith's parents always encouraged him to learn, to think critically, and to challenge the status quo.
```

End the conversation

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



```
Activities Terminal Jul 8 19:40 jetson@ubuntu: ~
competition of the year was about to commence.

Dr. Smith knew he had to step up his game if he wanted to compete with the best mathematicians of his time. He spent countless hours studying and preparing for the competition. When the day finally arrived, Dr. Smith stepped up to the stage, and everyone in the audience was amazed by his innovative approach.

Dr. Smith began by introducing himself and sharing a story about his childhood. He told the judges that he had grown up with parents who worked hard every day to provide for their family. They often went without food or shelter because they were too poor to afford these things. But Dr. Smith's parents always encouraged him to learn, to think critically, and to challenge the status quo.

Dr. Smith then explained his solution to a difficult problem in the competition. He had analyzed the given data, conducted extensive research, and developed a unique approach to solve the problem. The judges were stunned by his ingenuity and quickly realized that he was not only a brilliant mathematician but also an exceptional thinker and communicator.

The judges were blown away by Dr. Smith's performance and awarded him a prize that no one had expected. Dr. Smith was ecstatic as he accepted the prize. He knew that he had not only earned respect and admiration from his peers.

From that day on, Dr. Smith became known throughout the world for his innovative approach to problem-solving. His success inspired others to think critically, leading to new breakthroughs in science.

jetson@ubuntu:~$ ollama run tinyllama:1.1b
>>> Use Ctrl + d or /bye to exit.
>>>
jetson@ubuntu:~$
>>> /bye
jetson@ubuntu:~$
```

References

Ollama

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

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

TinyLlama

GitHub: <https://github.com/jzhang38/TinyLlama>

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