

# Gemma

## Gemma

- 1. Model scale
- 2. Pull Gemma
- 3. Use Gemma
  - 3.1. Run Gemma
  - 3.2. Have a conversation
  - 3.3. End the conversation
- References

### Demonstration environment

**Development board:** rdk x5 motherboard

**sd card:** 64G

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

Board model	Ollama
rdk x5 8GB	√(Need to run the small parameter version)
rdk x5 4GB	x

Gemma is a new open model developed by Google and its DeepMind team.

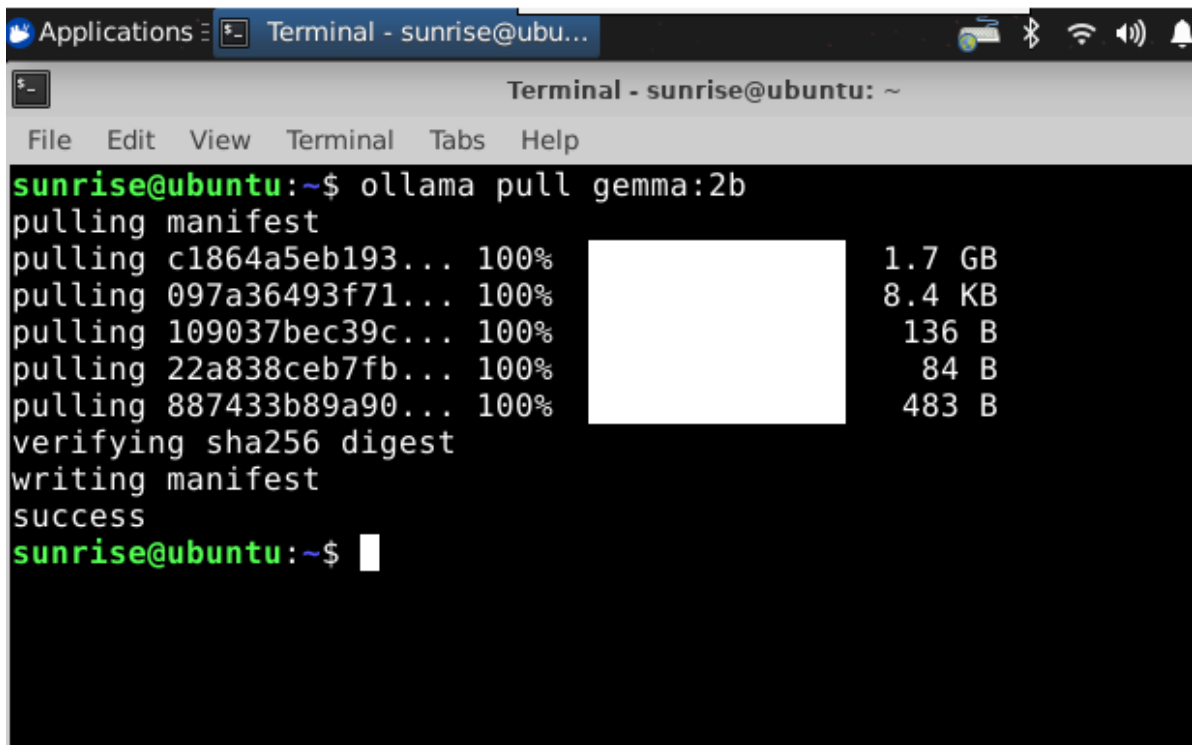
## 1. Model scale

Model	Parameter
Gemma	2B

## 2. Pull Gemma

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

```
ollama pull gemma:2b
```

A terminal window titled "Terminal - sunrise@ubu..." showing the command "ollama pull gemma:2b" and its output. The output shows progress bars for pulling the manifest and several layers, all at 100%. The layers are: c1864a5eb193... (1.7 GB), 097a36493f71... (8.4 KB), 109037bec39c... (136 B), 22a838ceb7fb... (84 B), and 887433b89a90... (483 B). The process concludes with "verifying sha256 digest", "writing manifest", and "success".

```
sunrise@ubuntu:~$ ollama pull gemma:2b
pulling manifest
pulling c1864a5eb193... 100% 1.7 GB
pulling 097a36493f71... 100% 8.4 KB
pulling 109037bec39c... 100% 136 B
pulling 22a838ceb7fb... 100% 84 B
pulling 887433b89a90... 100% 483 B
verifying sha256 digest
writing manifest
success
sunrise@ubuntu:~$
```

### 3、 Use Gemma

#### 3.1、 Run Gemma

```
ollama run gemma:2b
```

#### 3.2、 Have a conversation

```
print HelloWorld in C
```

**Requires rdk x5 8gb board to run**

Reply time is related to hardware configuration, please be patient!

```
Terminal - sunrise@ubuntu: ~
File Edit View Terminal Tabs Help
sunrise@ubuntu:~$ ollama run gemma:2b
>>> print HelloWorld in C
```c
#include <stdio.h>

int main()
{
    printf("HelloWorld\n");
    return 0;
}

**Output:**
```
HelloWorld
```

**Explanation:**
* `#include <stdio.h>` includes the standard input/output header file, which provides the `printf` function.
* `int main()` is the main function where the program starts.
```

### 3.3. End the conversation

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

```
Applications Terminal - sunrise@ubu...
Terminal - sunrise@ubuntu: ~
File Edit View Terminal Tabs Help

printf("HelloWorld\n");
return 0;
}

**Output:**
```
HelloWorld
```

**Explanation:**
* `#include <stdio.h>` includes the standard input/output header file which provides the `printf` function.
* `int main()` is the main function where the program starts.
* `printf("HelloWorld\n");` calls the `printf` function with a string argument. The `\n` character is used to insert a newline character in the output.
* `return 0;` returns 0 to indicate successful execution.

>>> /bye
sunrise@ubuntu:~$
```

# References

---

## Ollama

Website: <https://ollama.com/>

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

## Gemma

GitHub: <https://github.com/google-deepmind/gemma>

Ollama model: <https://ollama.com/library/gemma>