

Ollama

- 오픈소스 LLM(Foundation Model)을 로컬에서 쉽게 실행할 수 있게 해주는 도구
- 도커 이미지 pull 받아서 사용하는 것처럼, Model 자체를 로컬에 다운로드
- 네트워크 연결 없이 사용가능



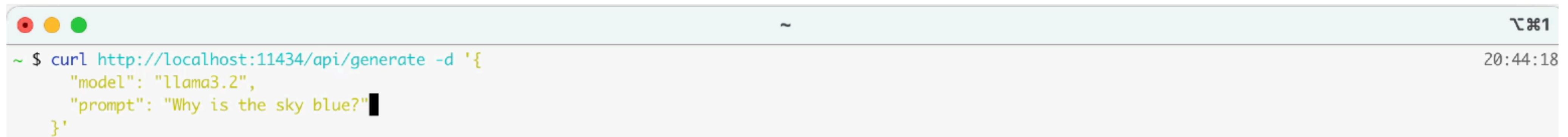
```
~/ai $ ollama ls
```

| NAME | ID | SIZE | MODIFIED |
|--------------------------|--------------|--------|----------------|
| mxbai-embed-large:latest | 468836162de7 | 669 MB | 51 minutes ago |
| llava:latest | 8dd30f6b0cb1 | 4.7 GB | 30 hours ago |
| nomic-embed-text:latest | 0a109f422b47 | 274 MB | 2 days ago |
| gemma:7b | a72c7f4d0a15 | 5.0 GB | 4 days ago |
| llama3.3:latest | a6eb4748fd29 | 42 GB | 5 days ago |
| deepseek-r1:latest | 0a8c26691023 | 4.7 GB | 6 days ago |
| phi4:latest | ac896e5b8b34 | 9.1 GB | 4 weeks ago |
| llava:7b | 8dd30f6b0cb1 | 4.7 GB | 7 weeks ago |
| codegemma:2b | 926331004170 | 1.6 GB | 7 weeks ago |
| llama3.2:latest | a80c4f17acd5 | 2.0 GB | 7 weeks ago |
| llama3.1:latest | 91ab477bec9d | 4.7 GB | 6 months ago |

- <https://ollama.com>

Ollama

- EndPoint로 활용(<http://localhost:11434/api/generated>)



```
~ $ curl http://localhost:11434/api/generate -d '{  
  "model": "llama3.2",  
  "prompt": "Why is the sky blue?"}  
}'
```

The image shows a screenshot of a macOS terminal window. The window has a light gray header bar with three colored window control buttons (red, yellow, green) on the left and a close button on the right. The main area of the terminal is white and contains a single line of text. The text is a command-line interface (CLI) command using the curl tool to send a POST request to the URL `http://localhost:11434/api/generate`. The request body is a JSON object with two fields: `"model": "llama3.2"` and `"prompt": "Why is the sky blue?"`. The command is preceded by a tilde (`~`) and a dollar sign (`$`). The timestamp `20:44:18` is visible in the top right corner of the terminal window.