Ollama CLI Cheat Sheet

This cheat sheet will help you quickly get up to speed with the Ollama CLI. Use it for managing models, interacting with them, and integrating Ollama into your workflows.

# 🚀 Installation

Install Ollama on Linux or macOS with the following command:  
```bash  
curl -fsSL https://ollama.com/install.sh | sh  
```

# 🖥️ Start Ollama Server

Use this command to start the Ollama server which is required to run and use models locally:  
```bash  
ollama serve  
```

# 📥 Download / Pull Models

To download specific models, use the following commands:  
```bash  
ollama pull gemma:7b # Download specific model  
ollama pull mistral # Download Mistral model  
ollama pull llama2 # Download Llama2 model  
ollama pull codellama # Download CodeLlama model  
ollama pull llava # Download Vision model  
```

To list available models:  
```bash  
ollama list  
```

# 🗣️ Chat with a Model (Interactive)

You can interact with models directly by running:  
```bash  
ollama run gemma:7b  
```This will start an interactive chat session in your terminal.

# 🧾 Run with a Prompt (One-shot)

You can also pass a prompt directly to the model:  
```bash  
echo 'What is the capital of France?' | ollama run llama2  
```Or use the following to pass a prompt directly:  
```bash  
ollama run llama2 -p 'Explain quantum entanglement in simple terms.'  
```

# 🧱 Create a Custom Model

You can create a custom model by defining a `Modelfile`:  
```Dockerfile  
FROM llama2  
SYSTEM 'You are an expert data analyst.'  
```Then run:  
```bash  
ollama run mymodel  
```

# 🔄 Update Ollama & Models

To keep Ollama and its models up to date:  
```bash  
ollama update # Updates the CLI  
ollama pull gemma:7b # Pulls the latest model version  
```

# 🗑️ Remove Model

To remove a model, use:  
```bash  
ollama rm gemma:7b  
```

# 🔒 Advanced Config (Optional)

Ollama stores models in `~/.ollama`. You can configure settings by editing `~/.ollama/config.toml`.

# 📦 Use with Other Tools (e.g. ontogpt)

You can integrate Ollama models with other tools like `ontogpt`:  
```bash  
ontogpt extract -m ollama/gemma:7b -t gocam.GoCamAnnotations -i input.txt  
```Note: Ensure `ollama serve` is running when using this.

# 🧰 Common Models

| Model | Type | Notes |  
|-------------|--------------|------------------------------|  
| `gemma:7b` | Chat | Lightweight & fast |  
| `mistral` | Chat | Open-weight Mistral |  
| `llama2` | Chat | Meta’s LLaMA 2 |  
| `codellama` | Code | For programming |  
| `llava` | Multimodal | Vision + text (images) |