



# ICT Training Center

**Il tuo partner per la Formazione e la Trasformazione digitale della tua azienda**



## Note



# SPRING AI

## GENERATIVE ARTIFICIAL INTELLIGENCE CON JAVA

Simone Scannapieco

Corso avanzato per Venis S.p.A, Venezia, Italia

Novembre 2025

## Note



# DOCKER MODEL RUNNER

## Note

- ➡ Risposta di Docker ad Ollama
    - ➡ LLM in Docker container locali
    - ➡ Modelli AI generici dockerizzabili (WIP)
    - ➡ Docker mette a disposizione una serie di modelli *open source* scaricabili tramite Engine o Desktop
  - ➡ Requisiti: <https://www.ajeetraina.com/docker-model-runner-tutorial-and-cheatsheet-mac-windows-and-linux-support/>
  - ➡ <https://docs.docker.com/ai/model-runner/>
  - ➡ <https://www.docker.com/blog/run-llms-locally/>
  - ➡ <https://www.docker.com/blog/introducing-docker-model-runner/>

## Note

- ➔ Tramite Docker Engine
  - ➔ Tramite Docker Desktop

## Note

## Caricamento LLM in locale

```
docker model pull ai/gemma3
```

## Esecuzione LLM in locale

```
docker model run ai/gemma3
```

## Elenco LLM in locale

```
docker model list
```

## Eliminazione LLM in locale

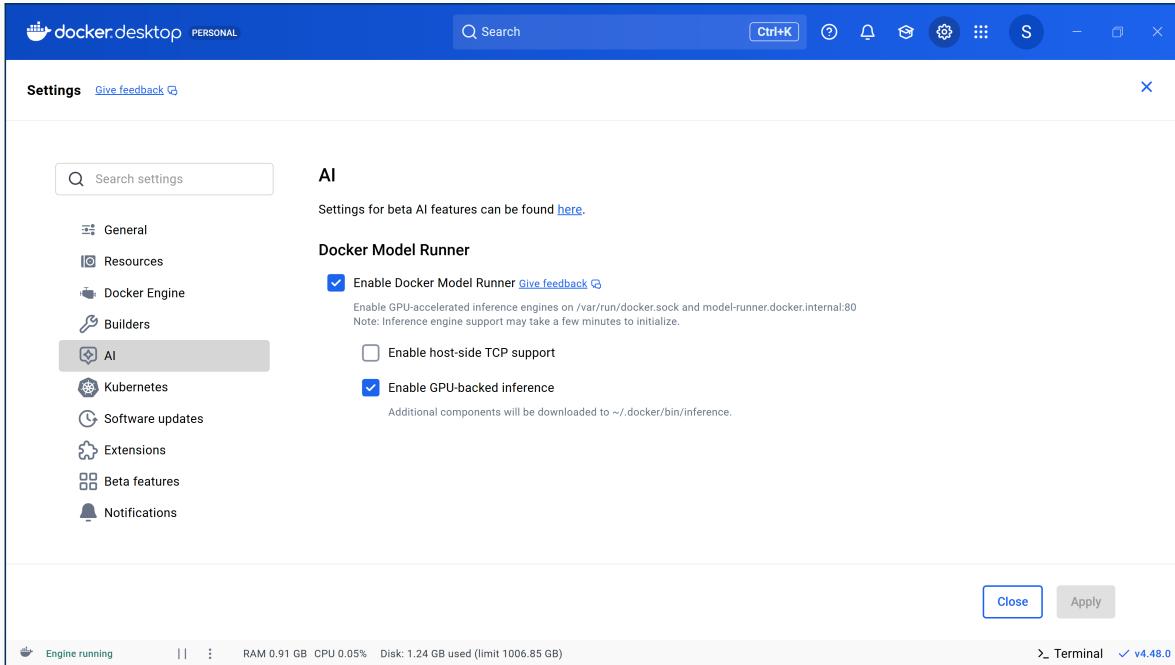
```
docker model rm ai/gemma3
```

## Note

# **DOCKER MODEL RUNNER**

## **UTILIZZO DI BASE - DOCKER DESKTOP**

## 1 Verificare le impostazioni relative ad AI



 Simone Scannapieco

 Spring AI - Corso avanzato

 Venis S.p.A, Venezia, IT

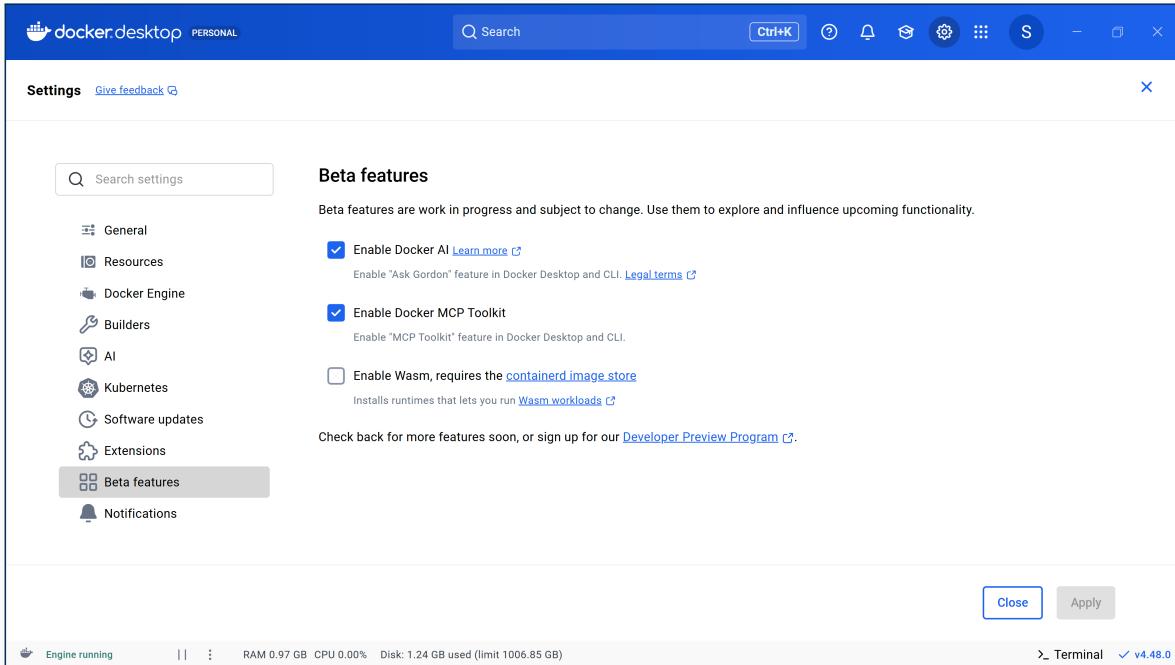
6 / 7

## Note

# DOCKER MODEL RUNNER

## UTILIZZO DI BASE - DOCKER DESKTOP

## 2 Verificare le impostazioni relative alle Beta features



 Simone Scannapieco

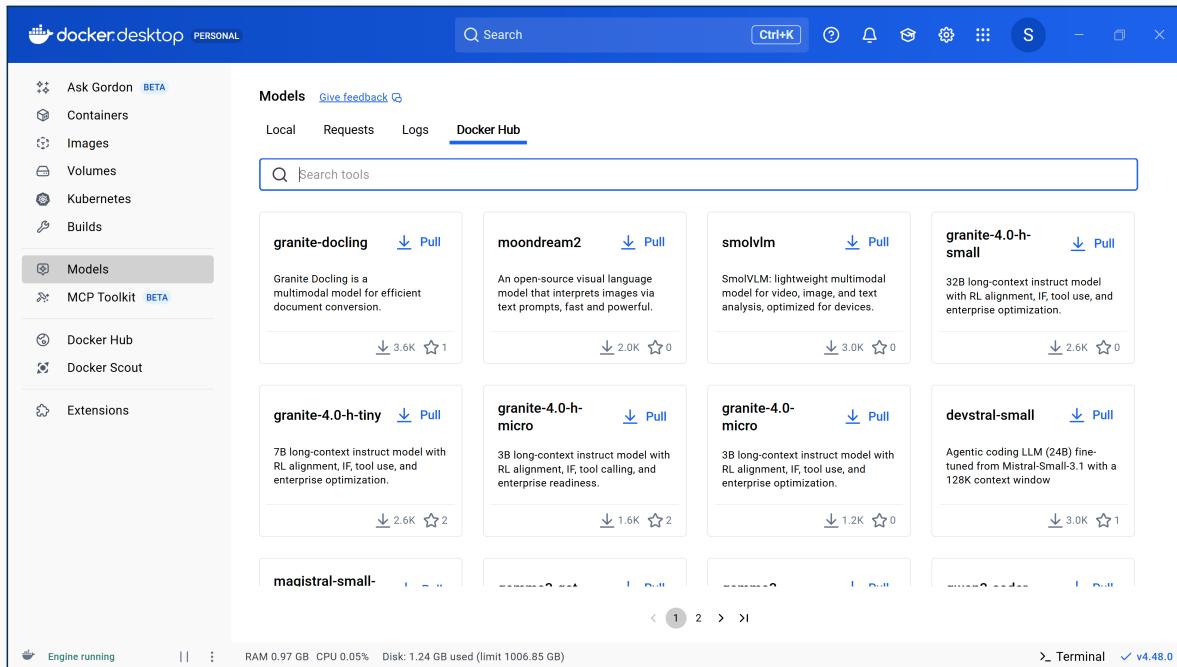
 Spring AI - Corso avanzato

 Venis S.p.A, Venezia, IT

6 / 7

## Note

### 3 Accedere al pannello Docker Hub della sezione Models



Simone Scannapieco

Spring AI - Corso avanzato

Venis S.p.A, Venezia, IT

6 / 7

## Note

# DOCKER MODEL RUNNER

## UTILIZZO DI BASE - DOCKER DESKTOP

#### 4 Selezionare il modello ed eventuale versionamento quantizzato

The screenshot shows the Docker Desktop interface with the 'Models' tab selected. The sidebar on the left includes links for Ask Gordon, Containers, Images, Volumes, Kubernetes, Builds, Models (selected), MCP Toolkit, Docker Hub, Docker Scout, and Extensions. The main content area has a search bar at the top and tabs for Local, Requests, Logs, and Docker Hub. A search bar with placeholder 'Search...' is followed by a refresh icon. Below is a table with columns: Name, ID, Quantization, Parameters, Size, Created, and Actions. One row is listed: 'ai/gemma3:1B-Q4\_K\_M'. The 'Actions' column for this row contains icons for copy, open, and delete. A progress bar indicates '0.15 of 0.75 GiB'. At the bottom, it says 'No rows' and 'Showing 1 item'.

Ask Gordon BETA

Containers

Images

Volumes

Kubernetes

Builds

**Models** BETA

MCP Toolkit BETA

Docker Hub

Docker Scout

Extensions

Models [Give feedback](#)

Local Requests Logs Docker Hub

Search...

Name ID Quantization Parameters Size Created Actions

ai/gemma3:1B-Q4_K_M				0.15 of 0.75 GiB			
---------------------	--	--	--	------------------	--	--	--

No rows

Showing 1 item

Engine running | RAM: 0.97 GB CPU: 0.00% Disk: 1.24 GB used (limit 1006.85 GB) v4.48.0

> Terminal ✓

 Simone Scannapieco

 Spring AI - Corso avanzato

 Venis S.p.A, Venezia, IT

6 / 7

## Note

## 5 Utilizzare il modello da linea di comando integrata

The screenshot shows the Docker Desktop application window. On the left, a sidebar lists various features: Ask Gordon (Beta), Containers, Images, Volumes, Kubernetes, Builds, Models (selected), MCP Toolkit (Beta), Docker Hub, Docker Scout, and Extensions. The main area displays a chat session titled "Gemma3:1B-Q4\_K\_M" (dbe29fa0169). The "Chat" tab is active, showing the message "Chat using ai/gemma3:1B-Q4\_K\_M". A blue button on the right says "Parlami di te". Below the message, a list of capabilities is provided:

- Genero testo: Posso scrivere diversi tipi di testo, come poesie, codice, script, pezzi musicali, e-mail, lettere, ecc.
- Rispondo a domande: Posso rispondere a domande anche se sono aperte, impegnative o strane. Cerco di fornire risposte informative e complete, anche se le mie risposte sono basate su dati che ho appreso durante l'addestramento.
- Traduco lingue: Posso tradurre tra diverse lingue.
- Seguo istruzioni: Posso seguire le istruzioni che mi vengono date in modo preciso.
- Sono un assistente: Posso aiutarti con compiti di scrittura, brainstorming e apprendimento.

At the bottom, there's a search bar with placeholder text "Ask a question..." and a terminal status bar indicating "Engine running", "RAM 0.97 GB CPU 0.05%", "Disk: 1.24 GB used (limit 1006.85 GB)", and "v4.80".

## Note

## → Come fosse un servizio OpenAI!

## File pom.xml

```
...<dependencies>
    <dependency>
        <groupId>org.springframework.boot</groupId>
        <artifactId>spring-boot-starter-web</artifactId>
    </dependency>
    <dependency>
        <groupId>org.springframework.ai</groupId>
        <artifactId>spring-ai-starter-model-openai</artifactId>
    </dependency>
</dependencies>
```

## **File application.yml**

```
spring:
  application:
    name: demo
  ai:
    openai:
      api-key: pippoplutopaperino
      base-url: http://localhost:12434/engines
      chat:
        options:
          model: ai/gemma3
```

## Note