



SWEvenTeam

E-Mail

sweventeam@outlook.it

MANUALE UTENTE

Informazioni documento

Versione	1.0.0
Redazione	Alessio Barraco Alessandro Damiani Alba Hui Larrosa Serrano Yuri Lunardon Matteo Mazzotti Valentina Schivo Alessio Turetta
Verifica	Alessio Barraco Alessandro Damiani Alba Hui Larrosa Serrano Yuri Lunardon Matteo Mazzotti Valentina Schivo Alessio Turetta
Approvazione	Valentina Schivo

Storia del documento

Versione	Data	Autori	Verificatori	Descrizione
1.0.0	2025-06-05	Valentina Schivo	-	Approvazione per PB
0.5.0	2025-06-04	Alba Hui Larrosa Serrano	Alessandro Damiani	Controllo finale
0.4.1	2025-05-30	Alessio Barraco	Matteo Mazzotti	Miglioramento documento
0.4.0	2025-05-27	Yuri Lunardon, Alessio Turetta	Alba Hui Larrosa Serrano	Completamento istruzioni utente
0.3.0	2025-05-19	Yuri Lunardon, Matteo Mazzotti	Alessio Barraco	Completamento istruzioni installazione
0.2.0	2025-04-11	Alessandro Damiani, Alessio Turetta	Valentina Schivo	Inizio stesura istruzioni installazione e utente
0.1.0	2025-05-06	Valentina Schivo	Yuri Lunardon	Redazione Introduzione e Requisiti
0.0.1	2025-04-28	Alessio Barraco	Alessio Turetta	Definizione struttura

Indice

1	Introduzione	5
1.1	Scopo del documento	5
1.2	Scopo del prodotto	5
1.3	Contesto d'uso	5
1.4	Glossario	5
2	Requisiti	6
2.1	Requisiti hardware	6
2.2	Requisiti software	6
3	Installazione e avvio	7
3.1	Passaggi preliminari	7
3.1.1	Clonazione del repository	7
3.1.2	Compilazione delle variabili d'ambiente	7
3.1.3	Scelta modelli LLM	7
3.2	Avvio dell'applicativo	8
4	Istruzioni d'uso	9
4.1	Configurazione	10
4.1.1	Gestione LLM	10
4.1.2	Insiemi di domande	11
4.2	Sessioni di test	12
4.2.1	Creazione - Modifica - Eliminazione	12
4.2.2	Esecuzione e visualizzazione dei test	12
4.3	Confronta risultati	14

Elenco delle Figure

1	Schermata "Home"	9
2	NavBar	9
3	Schermata "Gestisci LLM"	10
4	Creazione di un insieme di domande	11
5	Schermata "Insiemi di domande"	11
6	Schermata "Sessione di test"	12
7	Card di valutazione	13
8	Grafico di valutazione	13
9	Test precedenti	13
10	Schermata "Confronta risultati"	14

Elenco delle Tabelle

1	Requisiti di memoria in funzione della dimensione del modello	6
2	Variabili d'ambiente per ogni componente	7

1 Introduzione

1.1 Scopo del documento

Questo manuale è pensato per offrire tutte le indicazioni utili all'uso della web-app **ArtificialQI**, assicurando agli utenti un'esperienza ottimale e la possibilità di sfruttarne appieno ogni funzionalità. Il documento tratta le procedure per poter installare e utilizzare la web-app e una spiegazione delle sue funzioni e modalità di interazione con l'interfaccia.

1.2 Scopo del prodotto

ArtificialQI è una web-app con lo scopo principale di valutare in modo sistematico le capacità di risposta dei Large Language Models (LLM). Serve a fornire agli sviluppatori uno strumento avanzato e affidabile per il testing, l'analisi e il confronto delle prestazioni dei diversi LLM.

L'obiettivo è superare le limitazioni degli attuali metodi di validazione, come benchmark generici o valutazioni manuali, che sono lunghi e difficili da integrare nei cicli di sviluppo.

ArtificialQI consente di archiviare liste di domande e risposte attese, eseguire test automatizzati ponendo le domande a un LLM esterno e registrandone la risposta, valutare la correttezza e la verosimiglianza delle risposte ricevute, e presentare i risultati, il tutto integrato in un unico sistema. Affronta la complessità della valutazione dei modelli che non operano in modo deterministico e permette di verificare e confrontare come le prestazioni degli LLM variano in base alle loro caratteristiche, come il numero di parametri e l'addestramento. Il cuore del problema che l'applicazione risolve è la valutazione della verosimiglianza delle risposte ottenute dagli LLM.

1.3 Contesto d'uso

ArtificialQI è pensata come strumento di sviluppo e ricerca, non per l'uso finale da parte di utenti non tecnici. Per tale motivo il manuale utente, specialmente nella sezione di "Installazione e avvio" risulta più tecnico.

1.4 Glossario

Per garantire chiarezza e precisione nella comunicazione, è stato introdotto un elenco di riferimento, denominato Glossario, che raccoglie e spiega i termini tecnici o specifici utilizzati nella documentazione. Questo strumento, concepito per prevenire fraintendimenti o dubbi legati al linguaggio impiegato, include una serie di voci con le relative spiegazioni, permettendo così di definire in modo univoco i concetti e i termini, garantendo l'uniformità del lessico adoperato.

La presenza di un termine all'interno del Glossario viene indicata con questo stile_G.

2 Requisiti

2.1 Requisiti hardware

Il fabbisogno di risorse dipende dal modello di LLM che si desidera caricare nel backend. In assenza di GPU, le soglie di memoria consigliate da Ollama garantiscono che il modello risieda interamente in RAM.

Dimensione del modello	Esempi	RAM minima consigliata
≤ 7 B parametri	Llama-3 7B, Gemma 7B	≥ 8 GB
≤ 13 B parametri	Llama-2 13B	≥ 16 GB
≤ 33 B parametri	Llama-2 34B, Mixtral-8 \times 7B	≥ 32 GB

Tabella 1: Requisiti di memoria in funzione della dimensione del modello

2.2 Requisiti software

L'applicativo è fornito sotto forma di container Docker gestiti da Docker Compose, per cui il sistema host non richiede l'installazione manuale di runtime, librerie o altri servizi di supporto.

Per procedere è tuttavia necessario che sulla macchina siano già presenti - e correttamente avviati - i seguenti prerequisiti:

- Docker Engine con il plugin Docker Compose abilitato.
- Git per clonare il repository.

3 Installazione e avvio

3.1 Passaggi preliminari

3.1.1 Clonazione del repository

```
1 git clone git@github.com:SWEvenTeam17/ArtificialQI.git
2 cd ArtificialQI
```

3.1.2 Compilazione delle variabili d'ambiente

All'interno delle cartelle indicate nella tabella, rinominare i file `.env.example` in `.env`, completare le variabili mancanti e modificare quelle esistenti in base alle proprie necessità.

Directory	Variabili da impostare	Descrizione
client/frontend/	NEXT_PUBLIC_BACKEND_URL=	URL pubblico delle API
microservices/LLMService	OLLAMA_URL=	Endpoint del servizio Ollama
server/ArtificialQI	GEMINI_API_KEY= LLM_SERVICE_URL= DB_NAME= DB_USER= DB_PASS= DB_HOST= DB_PORT=	Chiave API Gemini URL del servizio LLM Parametri di connessione al database

Tabella 2: Variabili d'ambiente per ogni componente

Le variabili risultano quasi tutte già precompilate con valori funzionanti, ma è consigliabile verificarne la correttezza e modificarle in base alle proprie esigenze (es. porte già in utilizzo da parte di altri processi nel proprio sistema).

L'unica variabile che deve essere compilata manualmente è `GEMINI_API_KEY`, che deve contenere la chiave API per accedere al servizio Gemini di Google.

Per ottenerla è necessario registrarsi su <https://ai.google.dev/gemini-api/docs> e seguire le istruzioni per generare una chiave, che una volta ottenuta sarà sufficiente inserire nel file `.env` della directory "server", nella voce `"GEMINI_API_KEY="`.

Nota sul database:

All'interno del file `docker-compose.yml` nella root del progetto è possibile definire nella voce `"MYSQL_ROOT_PASSWORD"` la password di root di un database MySQL, in caso si desideri utilizzare un database diverso da quello predefinito (tale password dovrà coincidere con quella indicata nel file `.env` della directory "server").

3.1.3 Scelta modelli LLM

Nella directory `ollama/` aprire lo script `start-ollama.sh` e aggiungere i comandi `"ollama pull <modello>"` per ogni modello che si desidera testare.

Esempi:

```
ollama pull llama3.2:1b
ollama pull llama3:8b
ollama pull llama3:70b
ollama pull llava:7b
ollama pull qwen2.5:14b
ollama pull gemma2:9b
ollama pull mistral:7b
```

3.2 Avvio dell'applicativo

Dalla radice del progetto eseguire:

```
1 sudo docker compose up -d --build
```

Il comando compila le immagini, scarica i modelli e avvia tutti i servizi. Al termine, l'interfaccia web è disponibile all'indirizzo:

```
1 http://localhost:3000
```

Nota sui tempi di avvio

- *Il primo avvio può richiedere diversi minuti: oltre alla build dei container, Next.js deve completare la fase di bootstrap.*
- *Il primo test di inferenza sarà più lento perché il modello di embedding semantico viene scaricato e inizializzato; le richieste successive risulteranno sensibilmente più rapide.*

4 Istruzioni d'uso

Al primo accesso l'utente visualizza la Home, dove nella parte superiore è presente la navbar principale che consente di raggiungere tutte le funzionalità dell'applicazione.

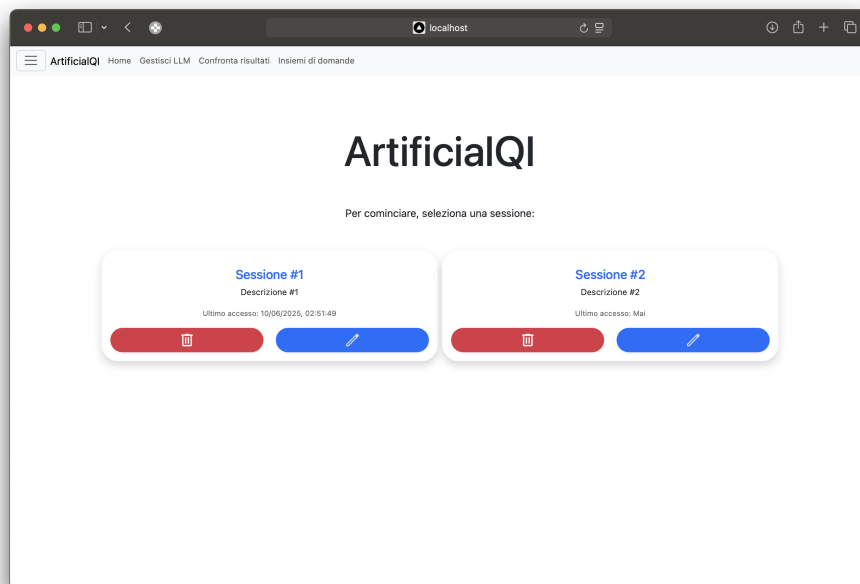


Figura 1: Schermata "Home"



Figura 2: NavBar

4.1 Configurazione

Per utilizzare l'applicazione è necessario eseguire le seguenti operazioni preliminari.

4.1.1 Gestione LLM

Selezionare “Gestisci LLM” dal menu e registrare i modelli da testare.

- Inserire manualmente nome e numero di parametri del modello.
- Oppure importare modelli già caricati in fase di installazione.

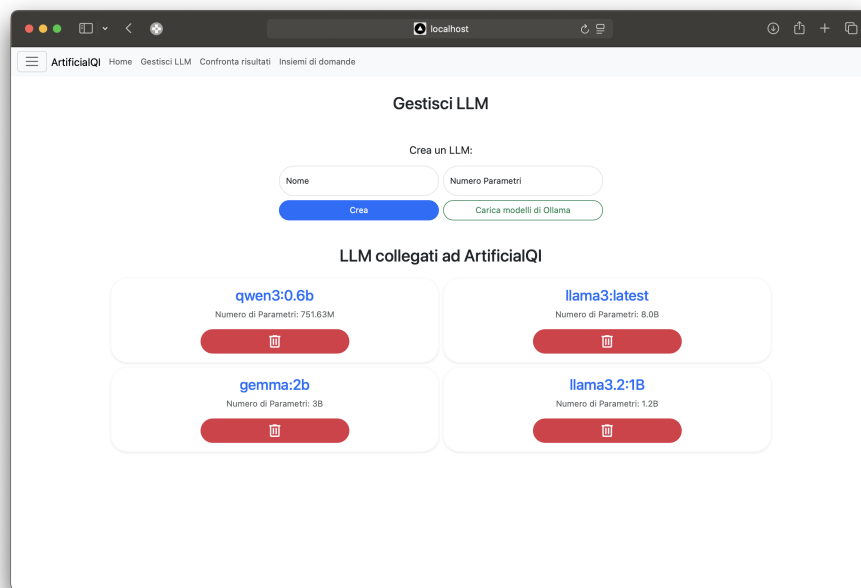
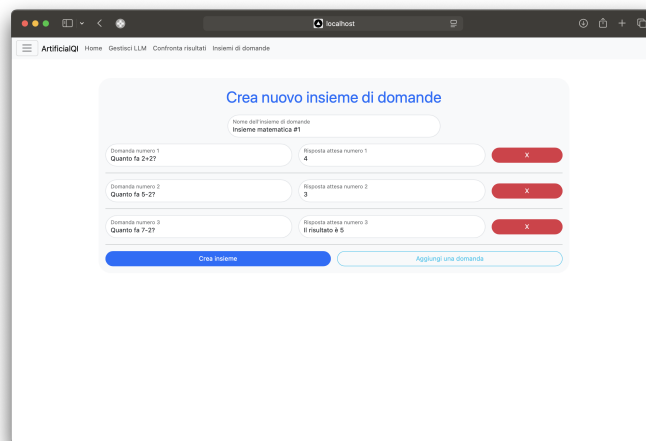


Figura 3: Schermata "Gestisci LLM"

4.1.2 Insiemi di domande

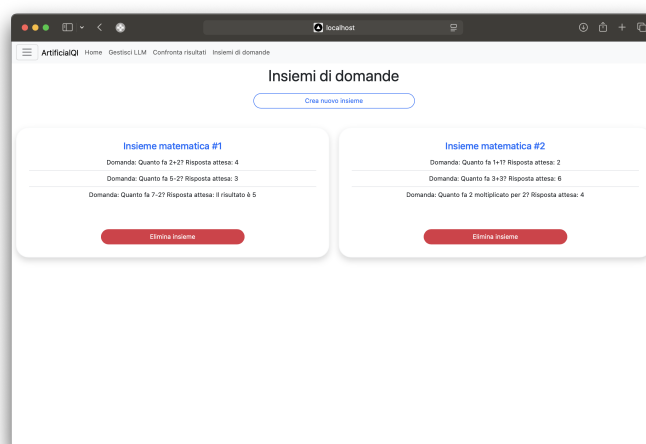
Selezionare "Insiemi di domande" dal menù e creare uno o più insiemi contenenti una o più domande e rispettive risposte attese.

Insiemi di domande a cui andranno poi a rispondere gli LLM desiderati dall'utente all'interno di una sessione.



The screenshot shows a web browser window with the URL 'localhost'. The page title is 'ArtificialQI' and the navigation bar includes 'Home', 'Gestisci LLM', 'Confronta risultati', and 'Insiemi di domande'. The main heading is 'Crea nuovo insieme di domande'. Below this, there is a form to create a new set of questions. The form includes a text input for 'Nome dell'insieme di domande' with the value 'Insieme matematica #1'. There are three rows of question and answer pairs, each with a red 'X' button to delete the pair. The first row is 'Domanda numero 1: Quanto fa 2+2?' with 'Risposta attesa numero 1: 4'. The second row is 'Domanda numero 2: Quanto fa 5-2?' with 'Risposta attesa numero 2: 3'. The third row is 'Domanda numero 3: Quanto fa 7-2?' with 'Risposta attesa numero 3: Il risultato è 5'. At the bottom, there are two buttons: 'Crea insieme' (blue) and 'Aggiungi una domanda' (light blue).

Figura 4: Creazione di un insieme di domande



The screenshot shows the 'Insiemi di domande' page. The navigation bar is the same as in Figure 4. The main heading is 'Insiemi di domande'. There is a button 'Crea nuovo insieme' at the top. Below this, there are two cards representing different sets of questions. The first card is titled 'Insieme matematica #1' and contains three questions: 'Domanda: Quanto fa 2+2? Risposta attesa: 4', 'Domanda: Quanto fa 5-2? Risposta attesa: 3', and 'Domanda: Quanto fa 7-2? Risposta attesa: Il risultato è 5'. There is a red 'Elimina insieme' button at the bottom of the card. The second card is titled 'Insieme matematica #2' and contains three questions: 'Domanda: Quanto fa 1+1? Risposta attesa: 2', 'Domanda: Quanto fa 3-2? Risposta attesa: 6', and 'Domanda: Quanto fa 2 moltiplicato per 2? Risposta attesa: 4'. There is a red 'Elimina insieme' button at the bottom of the card.

Figura 5: Schermata "Insiemi di domande"

4.2 Sessioni di test

Una sessione è lo spazio di lavoro in cui si collegano LLM e insiemi di domande e si eseguono i test.

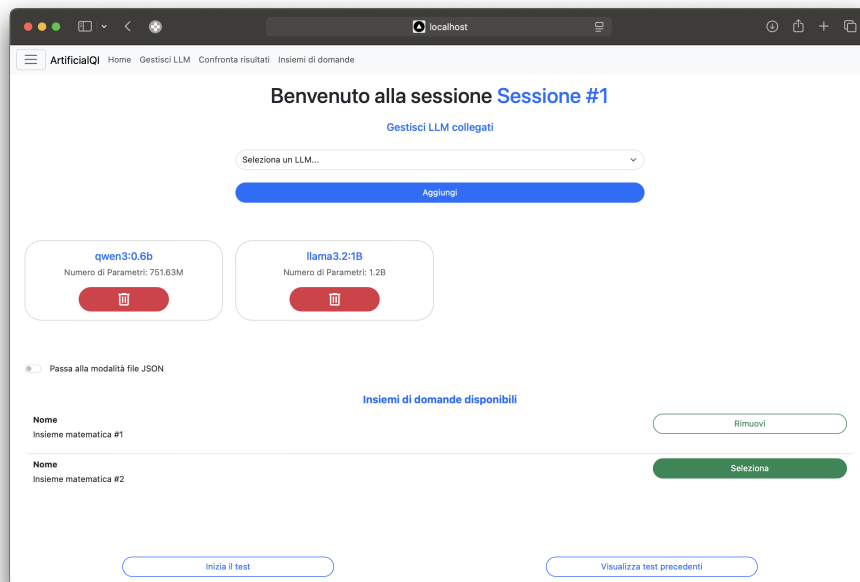


Figura 6: Schermata "Sessione di test"

4.2.1 Creazione - Modifica - Eliminazione

Dalla "Home" o dal menu è possibile creare una nuova sessione compilando i campi Titolo e Descrizione.

In qualsiasi momento, dalla Home, sarà poi possibile:

- Modificare titolo e descrizione di una sessione.
- Eliminare una sessione.

4.2.2 Esecuzione e visualizzazione dei test

Una volta all'interno di una sessione selezionare:

- gli LLM da impiegare nel test (collegati precedentemente in "Gestisci LLM");
- gli insiemi di domande da somministrare (creati in precedenza in "Insiemi di domande").

Premere **"Inizia il test"** per avviare la procedura.

Una volta terminata la procedura di valutazione il sistema mostrerà all'utente ogni domanda e rispettiva risposta presente all'interno del/degli insieme/insiemi impiegati nel test con l'aggiunta delle seguenti informazioni:

- LLM che ha risposto alla domanda;
- Risposta ottenuta;
- Valutazione semantica della risposta;
- Valutazione esterna sulla correttezza della risposta generata dall'LLM.

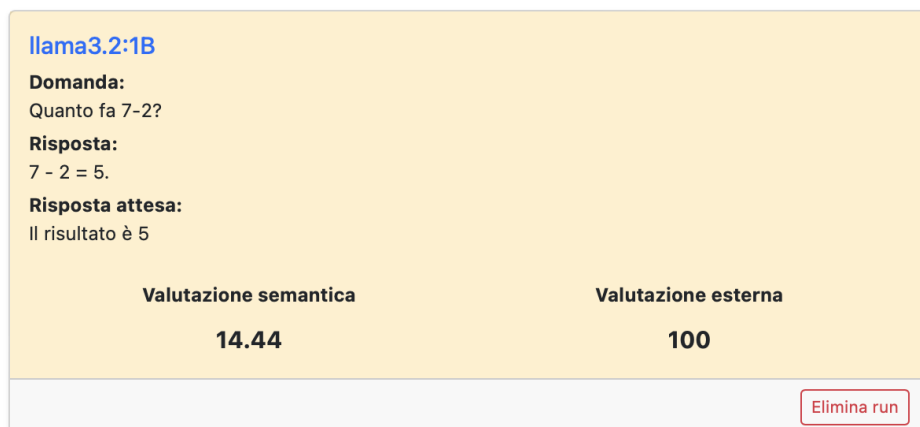


Figura 7: Card di valutazione

Verrà visualizzato poi sotto a questo insieme di valutazioni un grafico riepilogativo che evidenzia la “correttezza” delle risposte fornite dagli LLM.

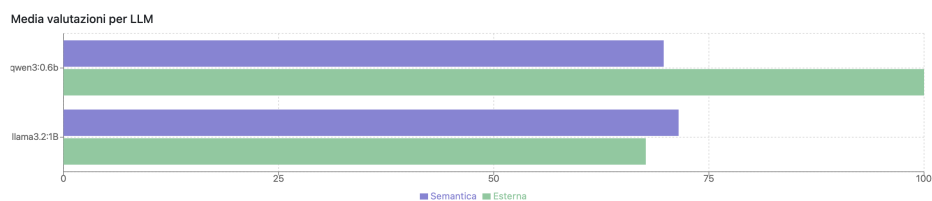


Figura 8: Grafico di valutazione

Dopo aver eseguito almeno un test, è possibile rivedere quelli già svolti tramite il pulsante “Visualizza test precedenti” presente nella sessione.

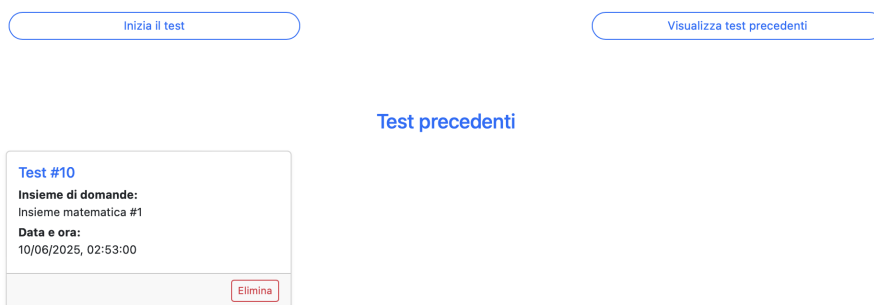


Figura 9: Test precedenti

Ogni test archiviato riporta:

- Insiemi di domande utilizzati
- Data e ora di esecuzione

4.3 Confronta risultati

Nella sezione “Confronta risultati” l’utente può mettere a confronto due modelli che hanno partecipato alla stessa sessione di test. Dopo aver selezionato la sessione desiderata, compaiono due menù a tendina – popolati esclusivamente con gli LLM collegati a quella sessione – per scegliere il Primo e il Secondo LLM.

Appena entrambi i modelli sono selezionati, l’applicazione recupera solo gli insiemi di domande affrontati da entrambi i modelli e calcola, per ciascun insieme, la media delle valutazioni semantiche e delle valutazioni esterne. I risultati sono presentati in due grafici a barre verticali sovrapposti: il primo mostra le medie semantiche, il secondo le medie esterne. Ogni riga rappresenta un insieme di domande, mentre le due barre colorate visualizzano il punteggio dei rispettivi LLM, permettendo di individuare a colpo d’occhio in quali insiemi un modello si è dimostrato superiore all’altro secondo i due criteri di valutazione.

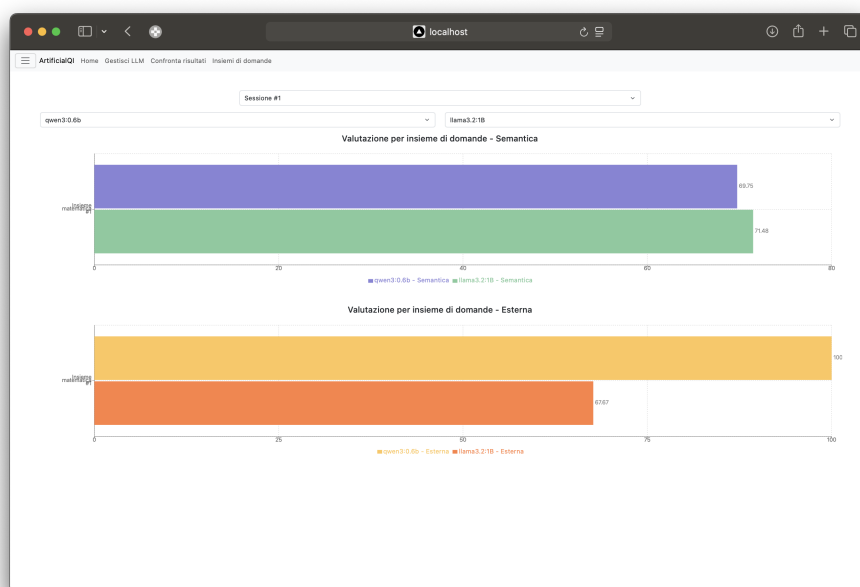


Figura 10: Schermata “Confronta risultati”