Alessandro Luppino


# AI-Based Penetration Testing Suite

Un assistente per il lavoro di Penetration Testing, in grado di riconoscere vulnerabilità e fornire supporto.

SLIDES

# TABLE OF CONTENTS

# TABLE OF CONTENTS

# INTRODUZIONE

Il progetto mira a sviluppare una suite per il Penetration Testing basata sull'intelligenza artificiale, progettata per assistere l'utente e migliorare l'efficacia e l'efficienza delle valutazioni di sicurezza.

Questa soluzione integrerà un modello linguistico avanzato (LLM) appositamente ottimizzato per eseguire compiti di analisi automatizzata e interagire con gli strumenti di sicurezza esistenti.

Offrendo agli utenti un'esperienza interattiva per dialogare direttamente con il modello e pianificare strategie di mitigazione personalizzate.

**NOTA:** i modelli, i dataset e i framework che verranno nominati in seguito sono provvisori e potrebbero subire dei cambiamenti nel corso del progetto come verrà approfondito nell'analisi dei rischi.

# OBIETTIVI

## ATTUALI

**Ricerca e Adattamento Modello LLM:**

- Ricerca di un modello piccolo ma performante.
- Interazione con il modello (CLI) e integrazione per l'utilizzo di moduli (tools).
- Training su dataset di CVE e vulnerabilità.

**Creazione del Modulo di Rilevazione delle Vulnerabilità:**

- Integrazione con strumenti di scansione esistenti.
- Information Gathering.
- CVE check.

## FUTURI

**Aumento prestazioni:**

- Hosting del modello su piattaforme dedicate come HuggingFace.co/spaces.
- Migrazione ad un modello più performante (10B>).
- Training su dataset più grandi su piattaforme dedicate.

**Nuovi Moduli:**

- Modulo di Attacco: Suggerire exploit e tecniche di attacco specifiche per le vulnerabilità rilevate.
- Modulo di Report: Genera report dettagliati.

# REQUISITI

## FUNZIONALI

**Interfaccia conversazionale:** Deve consentire agli utenti di interagire con il sistema tramite una chat AI per pianificare test, eseguire scansioni e ricevere suggerimenti.

**Scansione delle vulnerabilità:** Il sistema deve eseguire la scansione delle infrastrutture IT utilizzando strumenti come Nmap. Deve analizzare i risultati per identificare le vulnerabilità con un livello di rischio assegnato.

## NON FUNZIONALI

**Prestazioni:** Il sistema deve processare i dati di una scansione e generare risposte in tempi ragionevoli.

**Scalabilità:** Deve essere facilmente estendibile per includere nuovi tipi di vulnerabilità e strumenti di scansione.

**Usabilità:** Output chiaro e facilmente leggibile, in particolare per pentester o analisti.

**Affidabilità:** Deve fornire risultati accurati in termini di gravità delle vulnerabilità rilevate.

## TECNICI

**Modello AI:** Qwen/Qwen2.5-1.5B

**Linguaggi e Framework:** Python, Hugging Face, PyTorch.

**Integrazioni:** Nmap, Whois, Nvdlib.

**Dataset:** CVE, Exploit-DB.

**Device:** il modello è attualmente fatto girare in locale, richiede quindi un hardware abbastanza performante da renderlo possibile.

# ANALISI DEI RISCHI

| RISCHIO | PROBABILITA’ | IMPATTO | MITIGAZIONE |
|---|---|---|---|
| Dataset non formattato correttamente | Bassa | Alto | Effettuare un pre processamento accurato e testandolo con dataset più piccoli. |
| Problemi durante il fine tuning del modello AI | Alta | Medio | Testare con piccoli dataset prima di addestrare sul dataset completo. |
| Costi elevati di calcolo per l'addestramento | Media | Basso | Effettuare test iniziali con output da Nmap e OpenVAS per validare l'integrazione. |
| Difficoltà di integrazione con strumenti di scansione | Bassa | Basso | Utilizzare versioni ridotte del modello AI come DistilBERT. |
| Tempi di sviluppo eccessivi | Media | Alto | Seguire strettamente la pianificazione del diagramma di Gantt. |

# COCOMO

## EARLY DESIGN

**Size Stimata** = 3.5 KLOC

**B** = 1.1

**M** = PERS * RPCX * RUSE * PDIF * PREX * FCIL * SCED

**A** = 2.94

**Valori Scelti:**

PERS = 0.86 (high)

RPCX = 1.00 (nominal)

RUSE = 0.94 (low)

PDIF = 1.00 (nominal)

PREX = 1.00 (nominal)

FCIL = 0.91 (high)

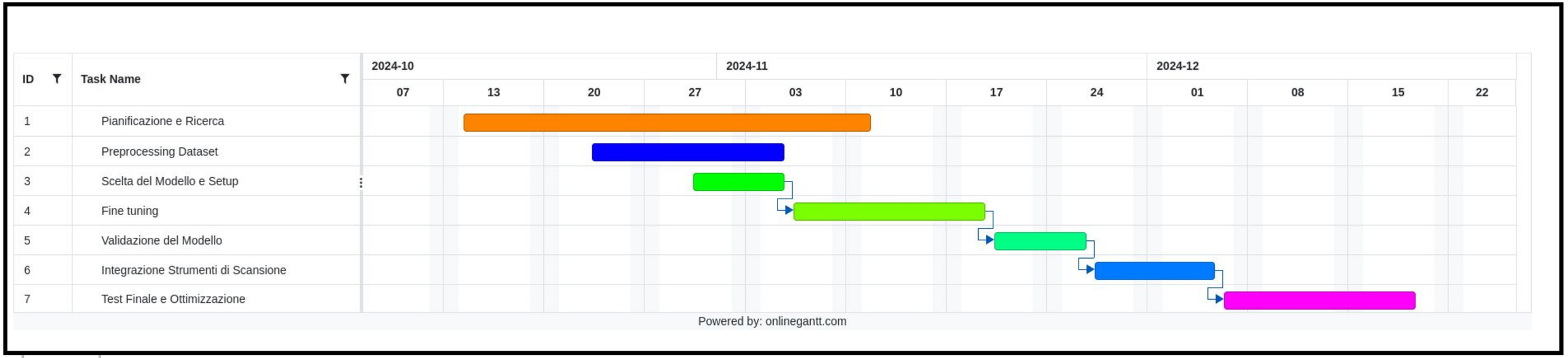
SCED = 1.00 (nominal)

---

**$PM = A * Size^{B} * M = 7,25657025$**

**NOTA:** La stima iniziale potrebbe non riflettere accuratamente le condizioni effettive del progetto. Ad esempio, alcuni fattori come la complessità del prodotto o le tempistiche richieste potrebbero essere stati sovrastimati. È possibile che alcune delle variabili considerate nel modello non rispecchino pienamente la natura specifica del progetto che sto affrontando. Inoltre, l'uso di strumenti o tecniche che semplificano il processo potrebbe non essere stato adeguatamente considerato nella stima. Per questo motivo, anche se la stima suggerisce un tempo maggiore, ritengo sia possibile completarlo comunque entro la scadenza.

# GANTT


ID
Task Name
2024-10
2024-11
2024-12
07
13
20
27
03
10
17
24
01
08
15
22
1 Pianificazione e Ricerca
2 Preprocessing Dataset
3 Scelta del Modello e Setup
4 Fine tuning
5 Validazione del Modello
6 Integrazione Strumenti di Scansione
7 Test Finale e Ottimizzazione
Powered by: onlinegantt.com

# FURPS+

## FUNCTIONALITY

- **Interfaccia conversazionale:**
  Per gestire la chat.

- **Chatting:**
  Possibilità di intraprendere una conversazione con un modello AI con conoscenze sull'argomento della sicurezza informatica.

- **Scanning:**
  Capacità di eseguire scansioni delle porte, riconoscere quali di esse è aperta, che servizio espone e che versione di quest'ultimo.

- **Analisi dei dati:**
  Per identificare e classificare le vulnerabilità.

## USABILITY

- **Interfaccia utente estremamente intuitiva:**
  I moduli vengono chiamati e gestiti dal modello stesso.

- **Supporto multilingue:**
  Va però specificato che l'utilizzo di una lingua diversa dall'inglese diminuisce la precisione del modello.

# FURPS+

## RELIABILITY

- **Accuratezza elevata nella scansione delle reti:**
  Quando richiesta la scansione di un IP il modello si affida a tool rinomati come Nmap.

- **Accuratezza elevata nell'identificazione delle CVE:**
  Quando richieste informazioni su CVE il modello interroga direttamente https://nvd.nist.gov/ evitando inconsistenze.

- **Altre generazioni:**
  Per diversi tipi di utilizzo fanno voce le regole comuni a tutti gli altri modelli (Controllare la veridicità delle informazioni che vengono generate, andando avanti nella conversazione la precisione può diminuire, ecc...)

## PERFORMANCE

- **Device CPU:**
  Nello stato attuale le performance sono legate alla CPU del dispositivo che esegue il modello.

- **Altre possibili variabili:**
  La velocità della rete

- **Benchmark su Ryzen 5 7535u:**
  Risposta: ~25s
  Scanning: ~60s

# FURPS+

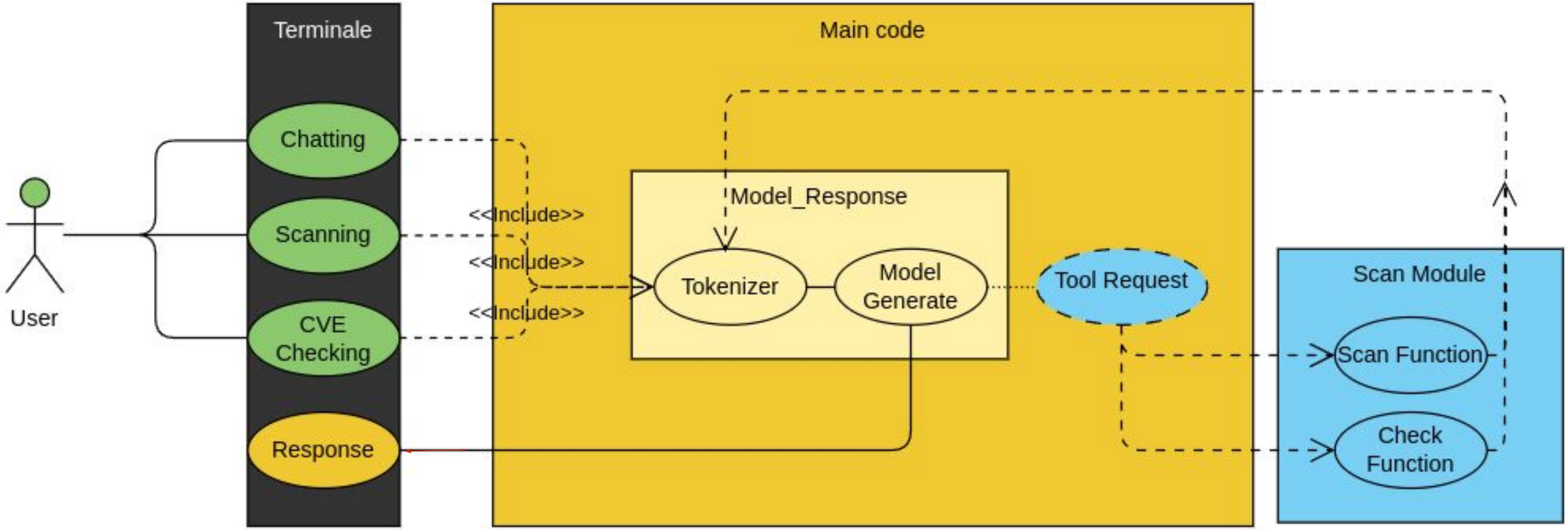
## SUPPORTABILITY

- **Architettura modulare:**
  Per aggiungere nuove funzionalità in futuro.
- **Compatibilità con strumenti esistenti**
- **Facile aggiornabilità del modello linguistico:**
  Per adattarsi a nuove minacce.

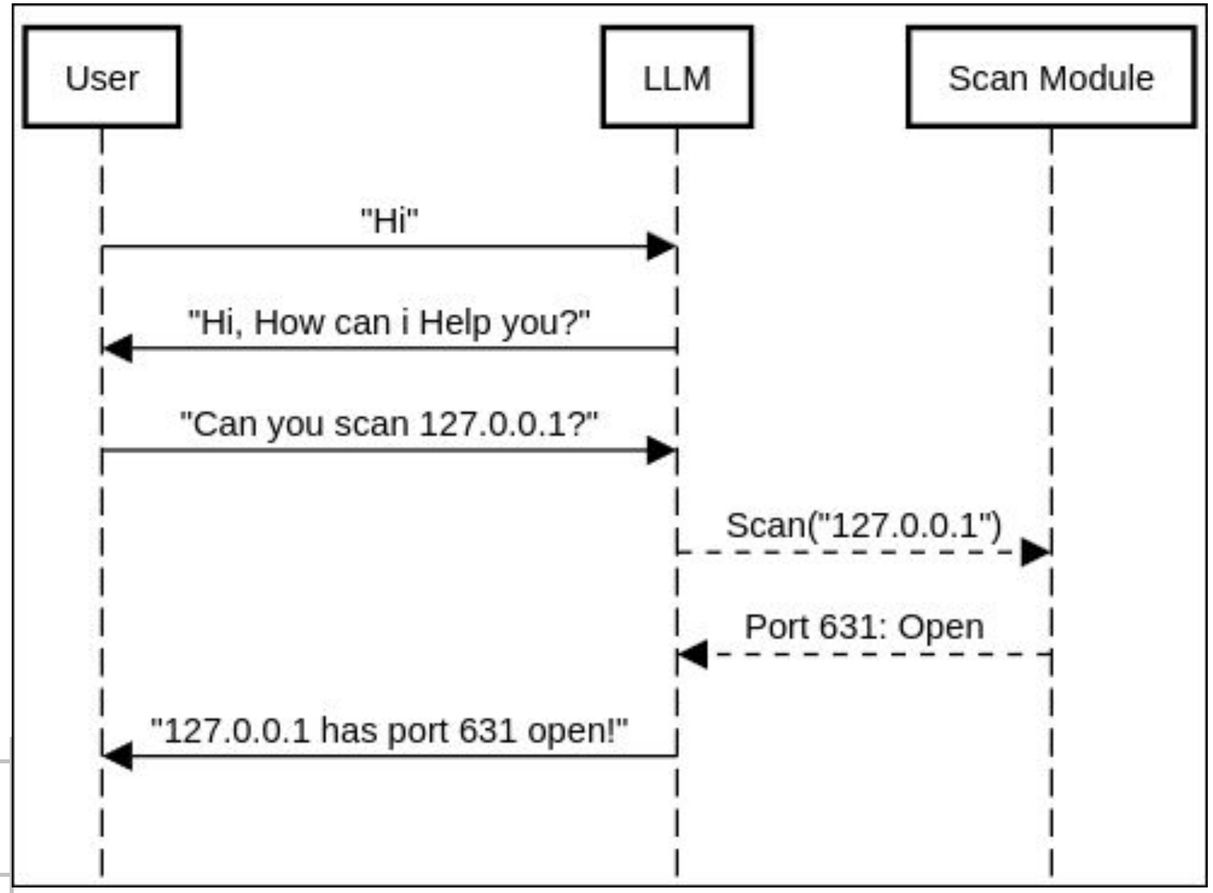
## EXTRA

- **Physical requirements:** Hardware performante.

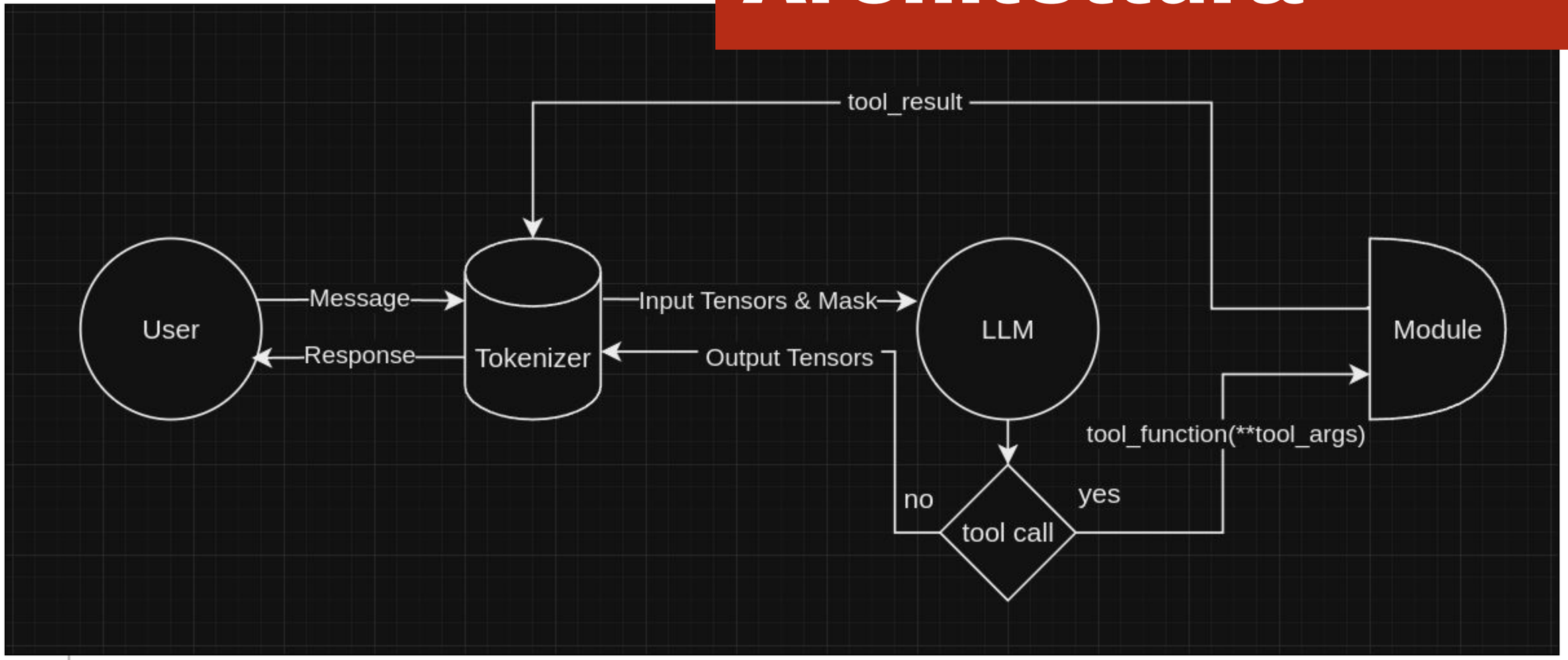
# DIAGRAMMA dei CASI D'USO



# DIAGRAMMA delle SEQUENZE



# Architettura




**NOTA:** Questa è la struttura attuale del progetto, rispecchierà un’archietttura Client-Server quando il modello sarà caricato e hostato su una piattaforma dedicata.

# FINE

**Grazie per l’attenzione!**