

LLM Studio

O platformă web completă pentru gestionarea și interacțiunea cu diferite
LLM-uri

Autor: Moldovan Andrei

April 9, 2025

Contents

1	Descriere Generală	2
1.1	Caracteristici Principale	2
1.2	Arhitectură	2
2	Stack Tehnologic	2
3	Caracteristici de Securitate	2
4	Interfața Utilizatorului	3
4.1	Pagina de Start	3
4.2	Termeni și Condiții	3
4.3	Politică	4
4.4	Înregistrare	4
4.5	Conectare	5
4.6	Uitare Parolă	5
4.7	Resetare Parolă	6
4.8	Email Resetare Parolă	7
4.9	Dashboard Utilizator	8
4.10	Conversație cu LLM	8
4.11	Ticket Suport	9
4.12	Modificare Date Utilizator	9
4.13	Modificare Parolă	10
5	Interfața Administratorului	10
5.1	Statistici Aplicație	10
5.2	Vizualizare Utilizatori	11
5.3	Vizualizare LLM-uri	11
5.4	Vizualizare Ticket-uri	11
5.5	Modificare Utilizator	12
6	Interfața Tehnicianului	12
6.1	Gestionare LLM-uri	12
6.2	Modificare LLM	13
7	Instalare și Rulare	13
8	Adaugare LLM-uri	13
9	Roluri Utilizatori	13

1 Descriere Generală

LLM Studio Manager este o platformă web avansată concepută pentru gestionarea și interacțiunea cu modele de limbaj de mari dimensiuni (LLM). Proiectul integrează o arhitectură modernă și securizată, oferind suport pentru multiple roluri de utilizatori și o varietate de funcționalități esențiale.

1.1 Caracteristici Principale

- **Suport Multi-LLM:** Posibilitatea de a alege dintre diverse LLM-uri disponibile.
- **Gestionarea Utilizatorilor:** Autentificare securizată și administrarea profilurilor.
- **Istoric Conversații:** Urmărirea și reluarea conversațiilor anterioare.
- **Vizualizare Date:** Prezentarea datelor în formate grafice intuitive.
- **Acces Bazat pe Roluri:** Funcționalități specifice pentru utilizatori, administratori și tehnicieni.

1.2 Arhitectură

Platforma este construită pe o arhitectură modulară:

- Frontend (ReactJS)
- Backend (FastAPI)
- Autentificare (JWT)
- Bază de Date (MongoDB)
- Server Web (Nginx)
- Gestionare LLM (Flask + Python + LLama + Transformers)

2 Stack Tehnologic

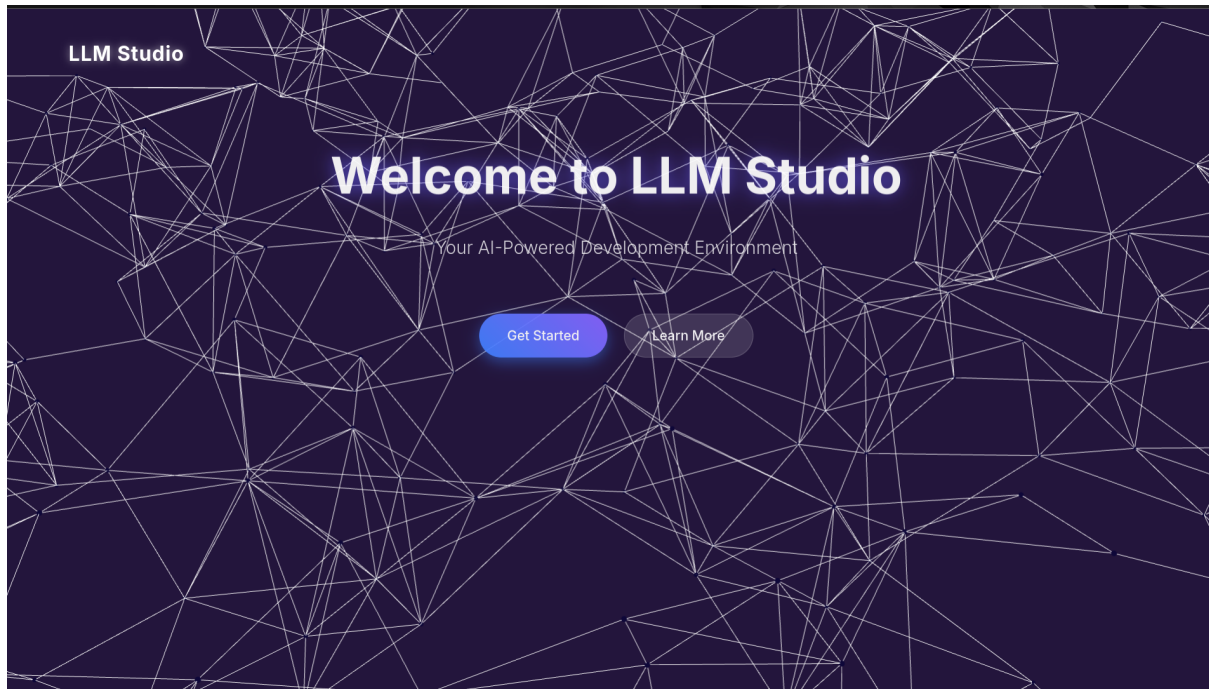
Componentă	Tehnologie
Frontend	ReactJS
Backend	FastAPI
Autentificare	JWT
Bază de Date	MongoDB
Server Web	Nginx
Gestionare LLM	LLama
Protocol	HTTPS

3 Caracteristici de Securitate

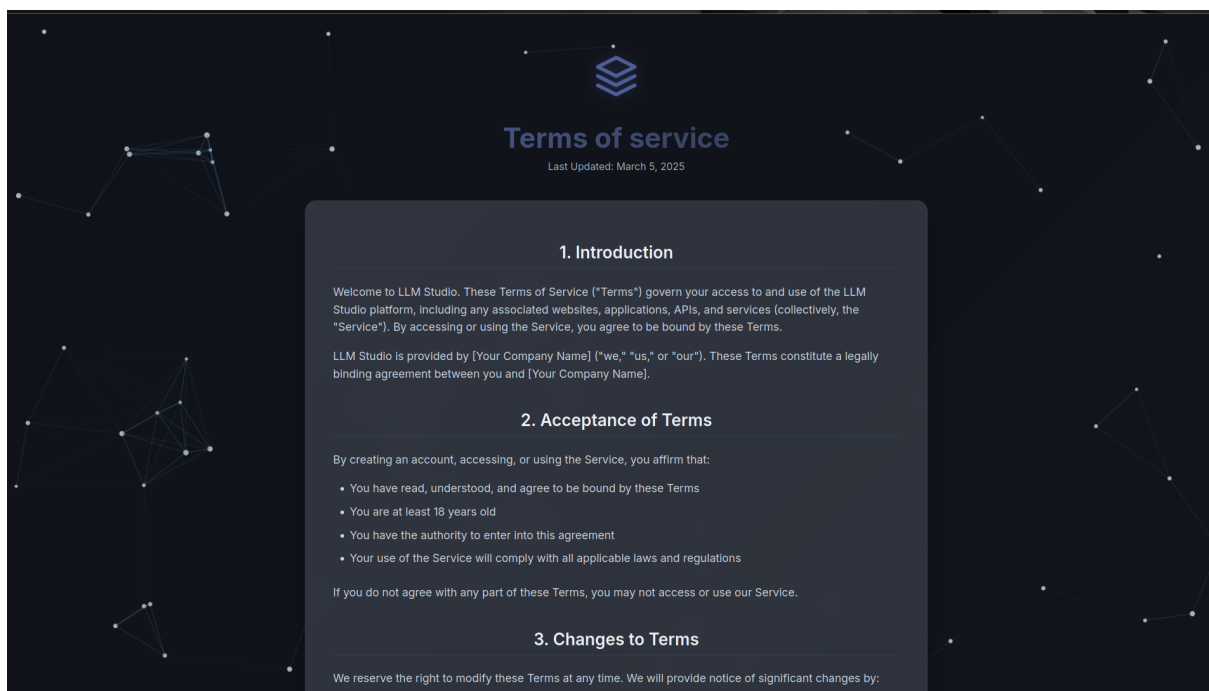
- JWT cu generare sigură de token.
- Validare strictă a input-urilor.
- Acces backend exclusiv prin frontend.
- HTTPS pe întreaga aplicație.
- Funcționare izolată a LLM-urilor.
- WebSockets pentru transmisie securizată.

4 Interfața Utilizatorului

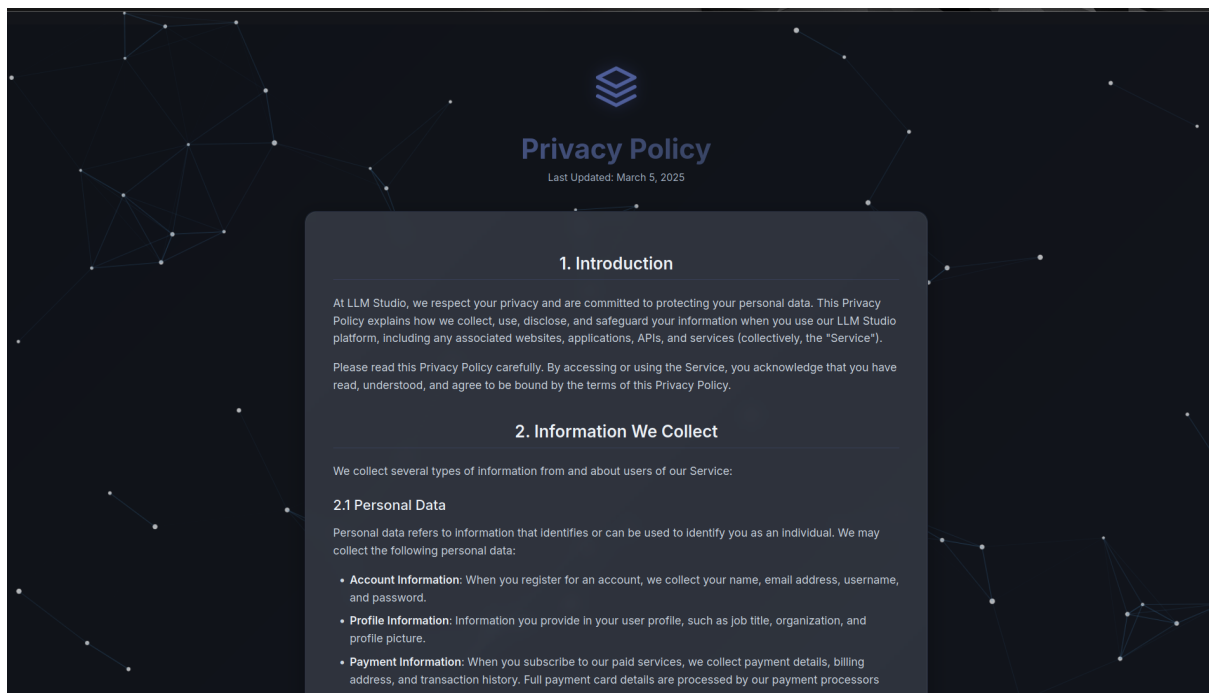
4.1 Pagina de Start



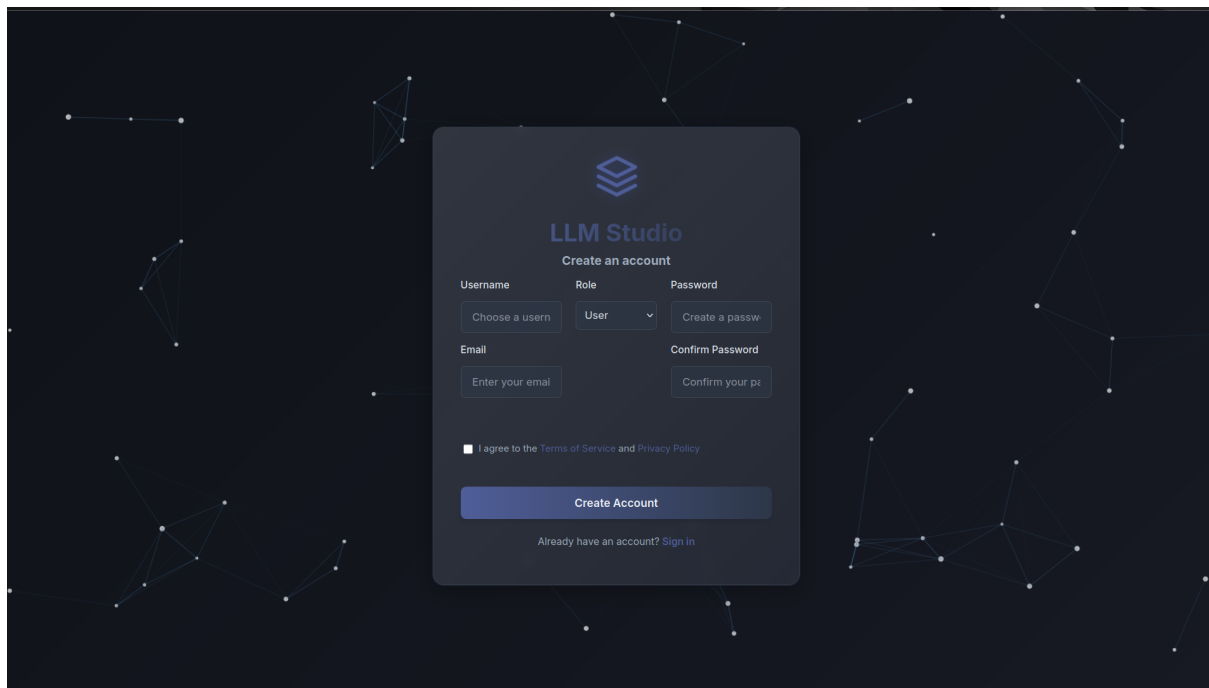
4.2 Termeni și Condiții



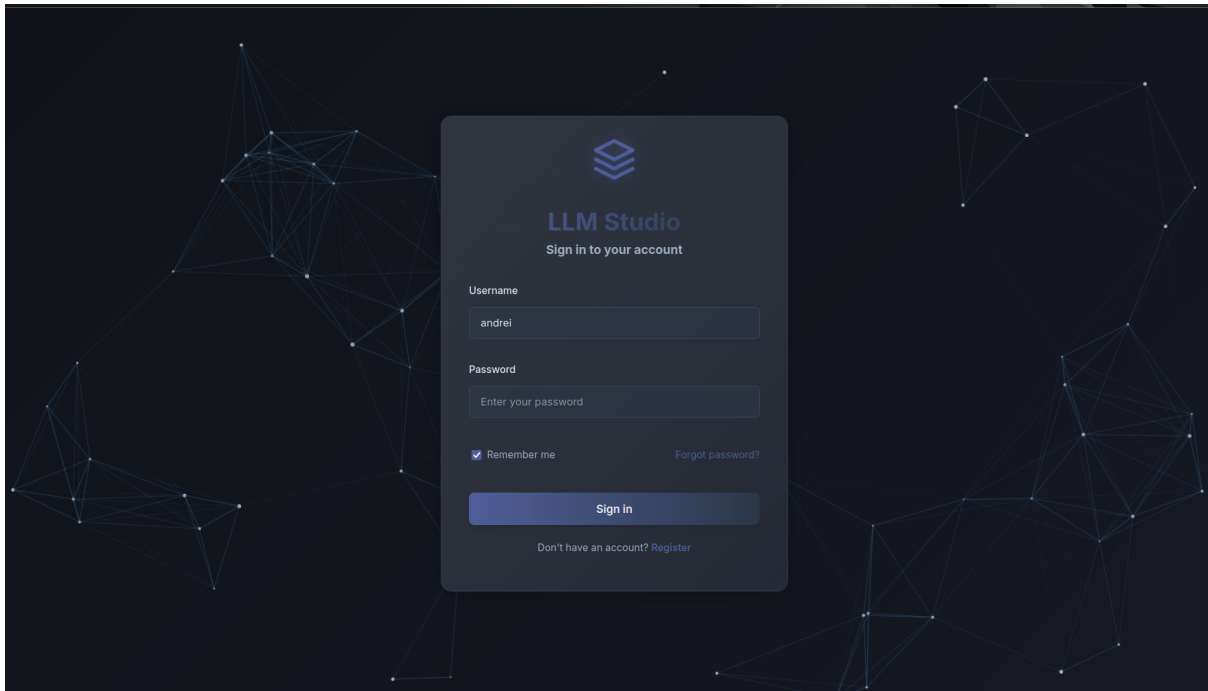
4.3 Politică



4.4 Înregistrare



4.5 Conectare



The image shows the 'Sign in' interface for LLM Studio. It features a dark background with a subtle geometric pattern of white dots and lines. A central dark gray card contains the LLM Studio logo at the top, followed by the text 'Sign in to your account'. Below this are two input fields: 'Username' with the value 'andrei' and 'Password' with the placeholder 'Enter your password'. There is a 'Remember me' checkbox which is checked, and a 'Forgot password?' link. A 'Sign in' button is positioned below the password field. At the bottom of the card, there is a link that says 'Don't have an account? Register'.

LLM Studio
Sign in to your account

Username
andrei

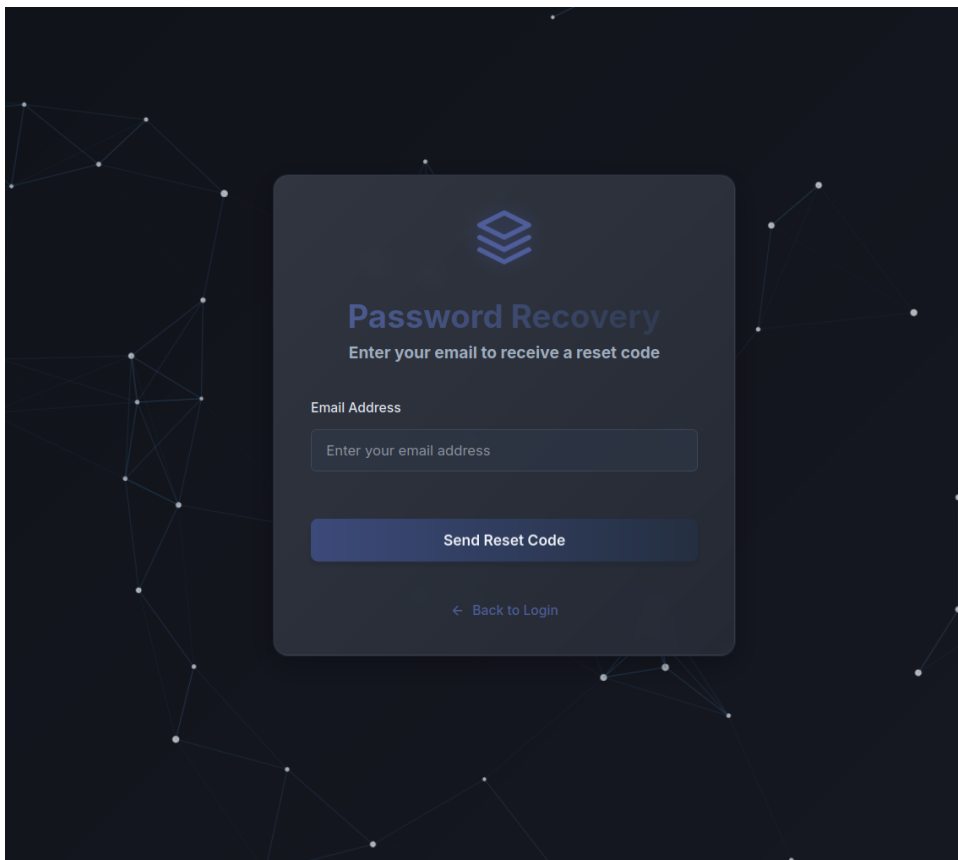
Password
Enter your password

☒ Remember me [Forgot password?](#)

Sign in

[Don't have an account? Register](#)

4.6 Uitare Parolă



The image shows the 'Password Recovery' interface for LLM Studio. It features a dark background with a subtle geometric pattern of white dots and lines. A central dark gray card contains the LLM Studio logo at the top, followed by the text 'Password Recovery' and 'Enter your email to receive a reset code'. Below this is an 'Email Address' input field with the placeholder 'Enter your email address'. A 'Send Reset Code' button is positioned below the input field. At the bottom of the card, there is a link that says '← Back to Login'.

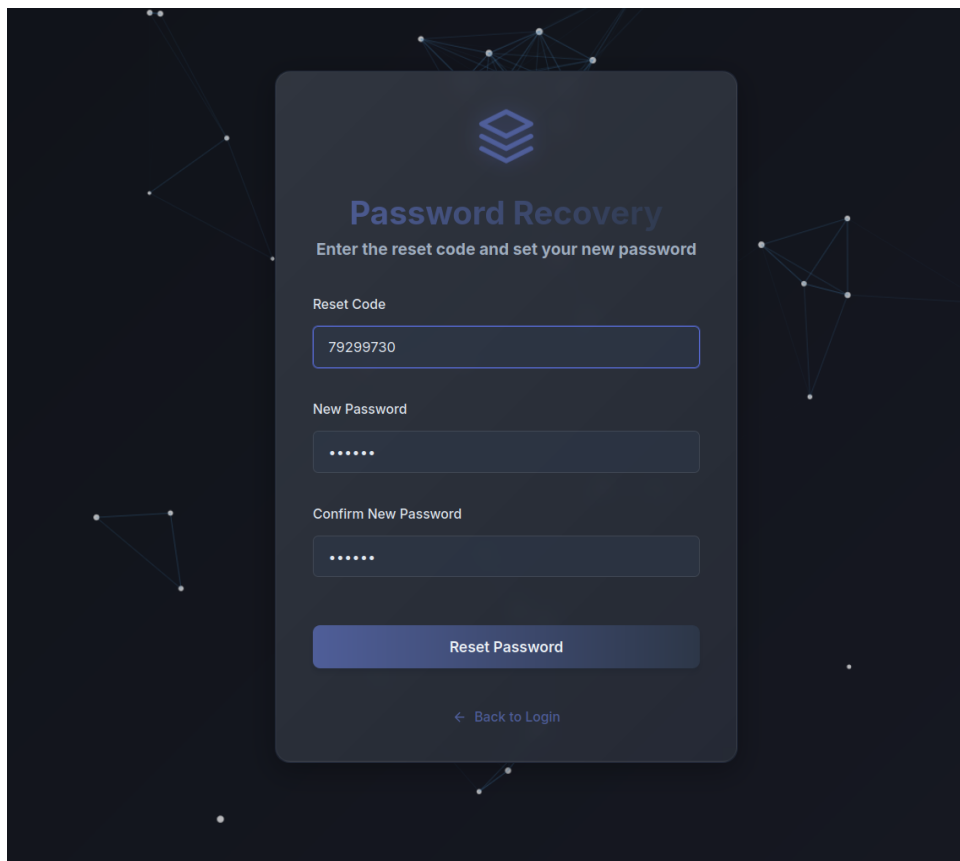
LLM Studio
Password Recovery
Enter your email to receive a reset code

Email Address
Enter your email address

Send Reset Code

[← Back to Login](#)

4.7 Resetare Parolă



The image shows a 'Password Recovery' form on a dark background with a network diagram. The form is a light gray rectangle with rounded corners. At the top is a logo consisting of three stacked blue squares. Below the logo is the title 'Password Recovery' in a bold, blue font, followed by the instruction 'Enter the reset code and set your new password' in a smaller, gray font. The form contains three input fields: 'Reset Code' with the value '79299730', 'New Password' with six dots, and 'Confirm New Password' with six dots. A blue 'Reset Password' button is at the bottom of the form, and a link '← Back to Login' is below it.

Password Recovery
Enter the reset code and set your new password

Reset Code
79299730

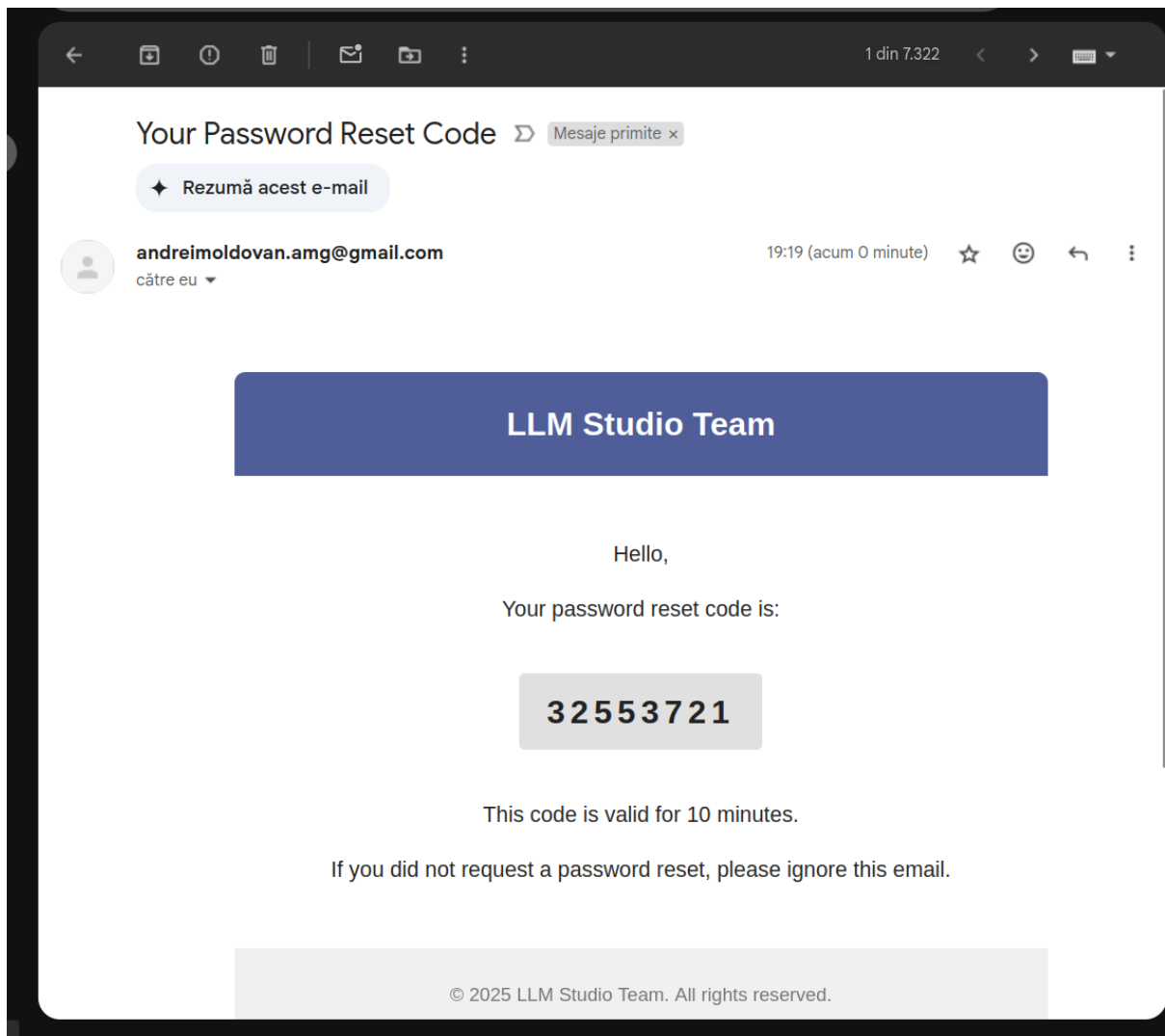
New Password
.....

Confirm New Password
.....

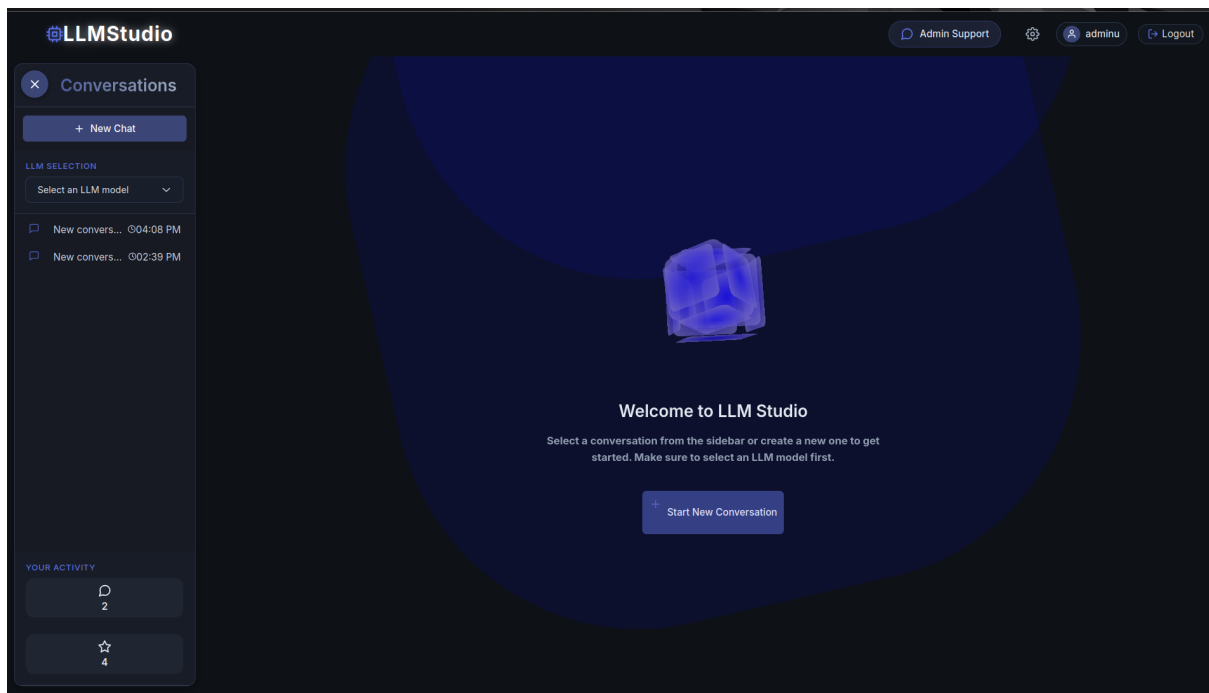
Reset Password

[← Back to Login](#)

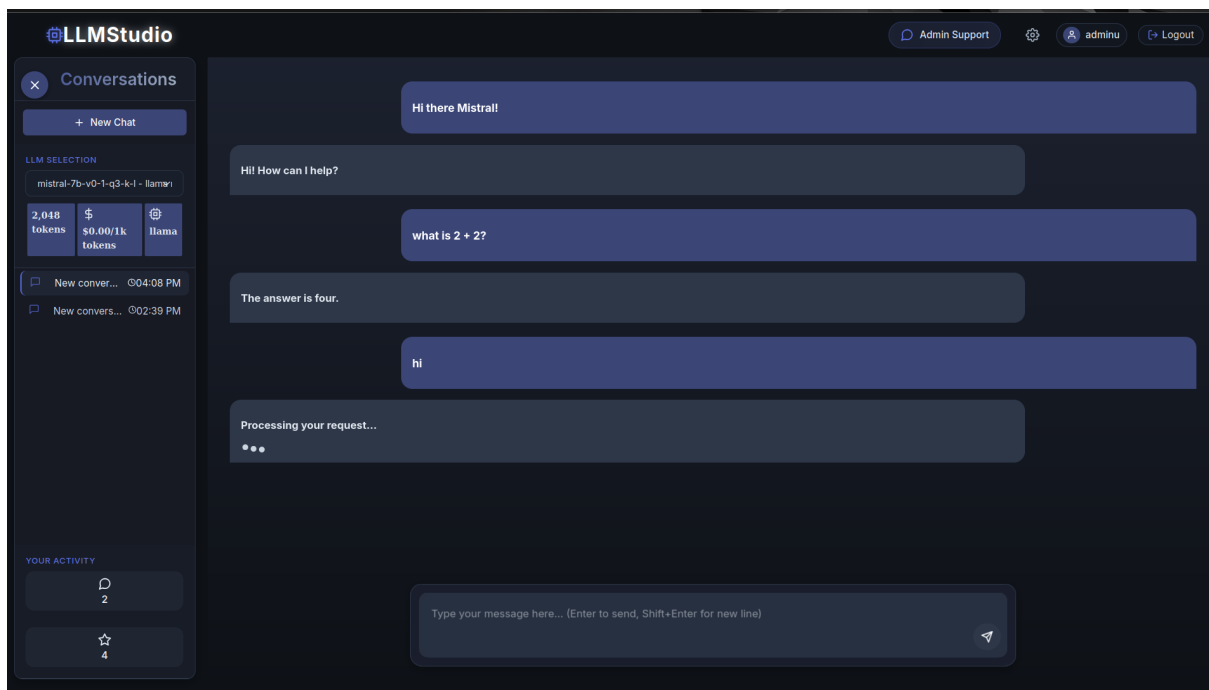
4.8 Email Resetare Parolă



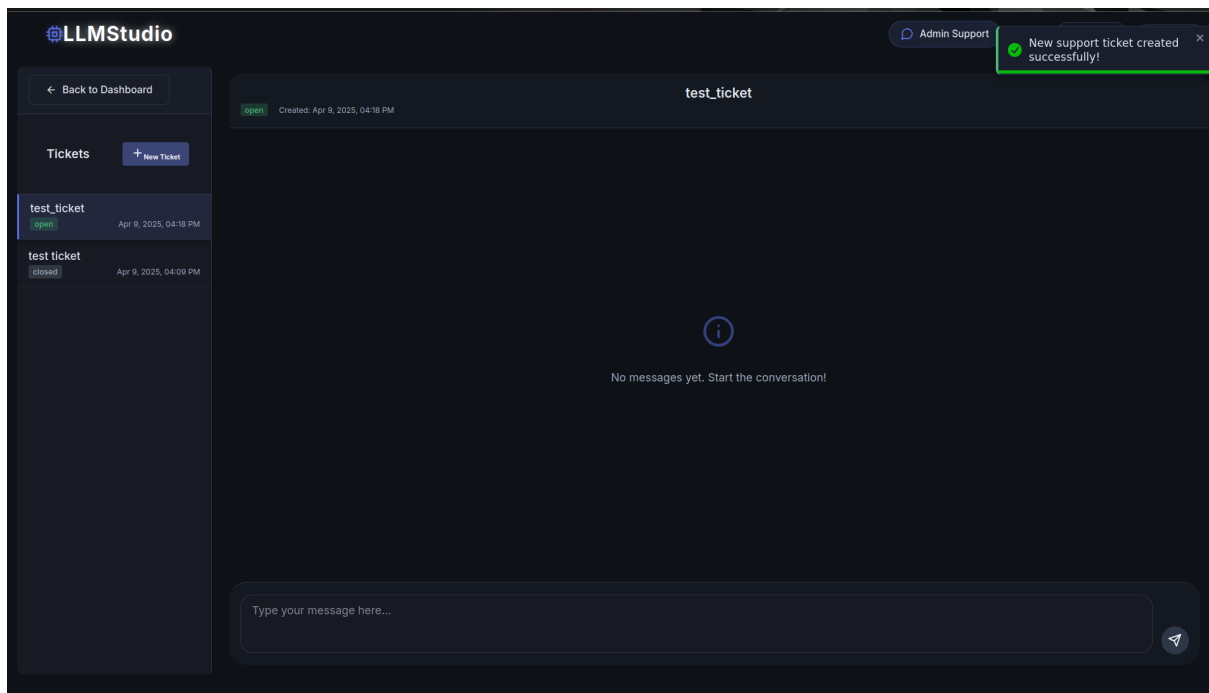
4.9 Dashboard Utilizator



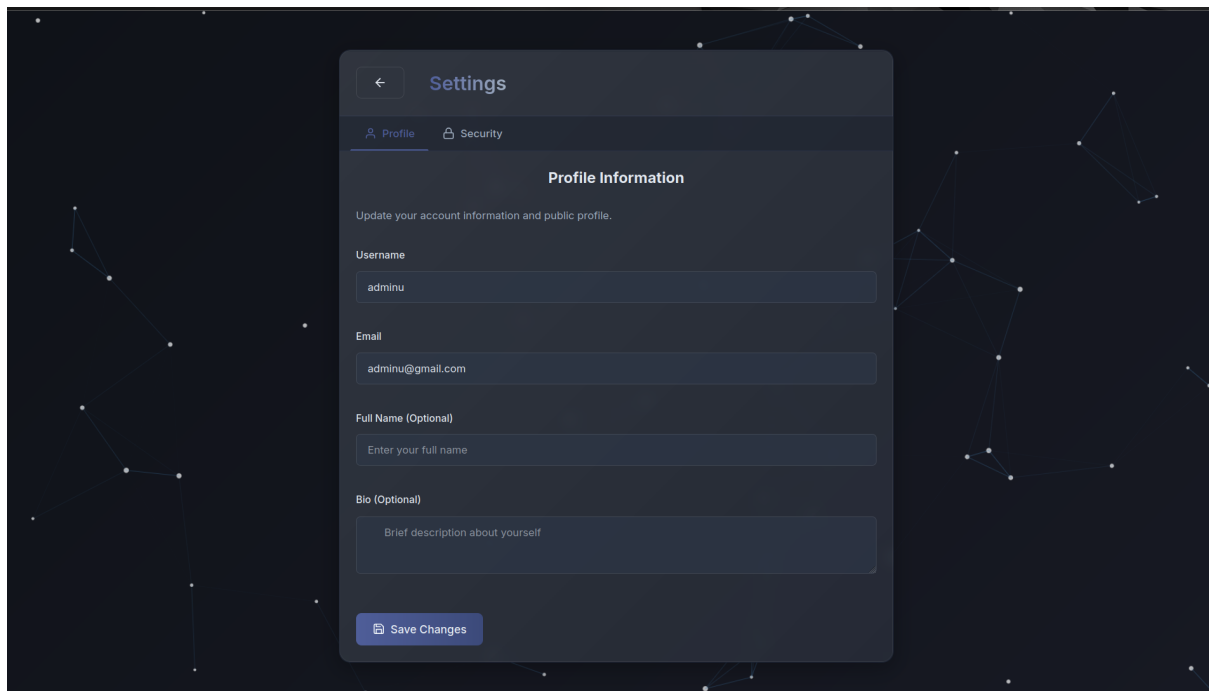
4.10 Conversație cu LLM



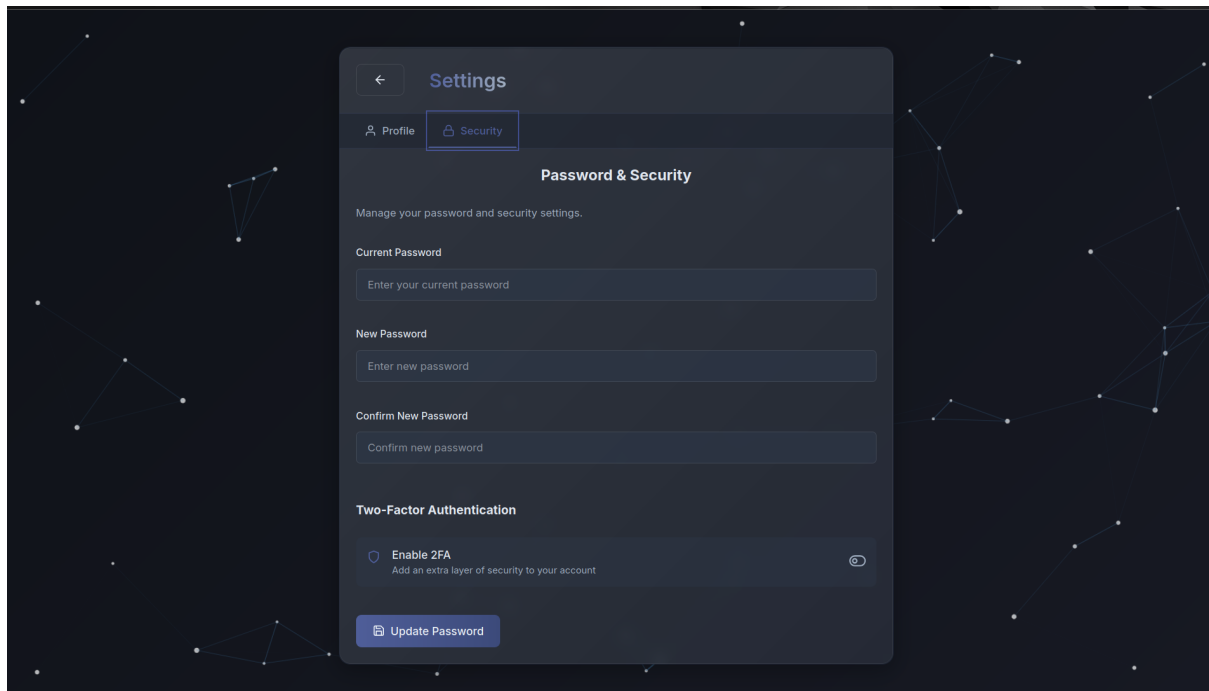
4.11 Ticket Suport



4.12 Modificare Date Utilizator

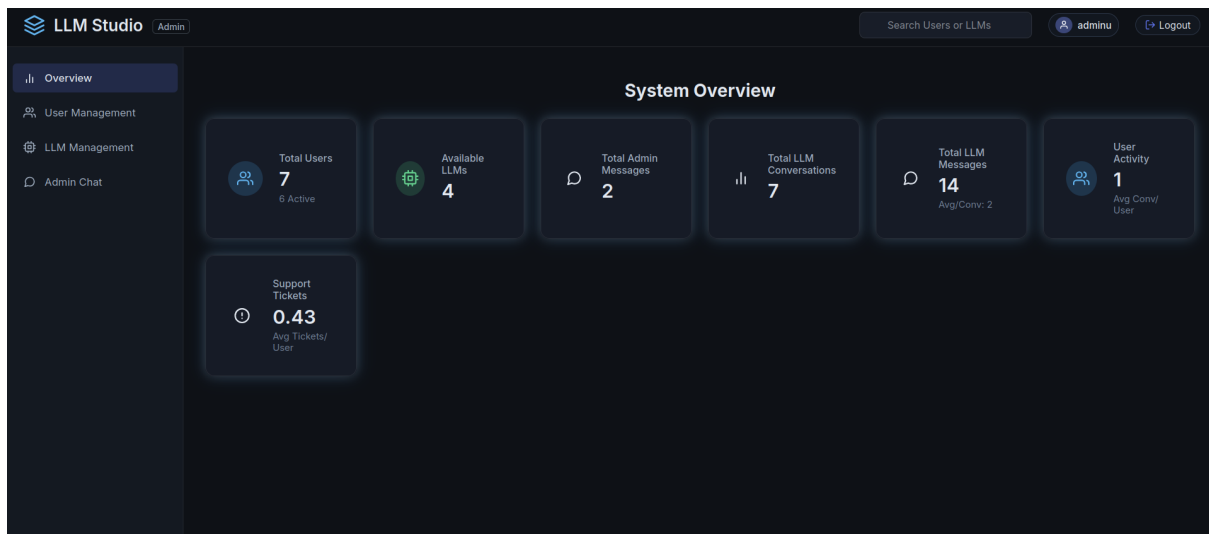


4.13 Modificare Parolă



5 Interfața Administratorului

5.1 Statistici Aplicație



5.2 Vizualizare Utilizatori

LLM Studio

Admin

Search Users or LLMs

admin

Logout

Overview

User Management

LLM Management

Admin Chat

User Management

Refresh Users

USERNAME	EMAIL	ROLE	STATUS	LAST LOGIN	ACTIONS
admin	admin@llmstudio.com	Admin	Active	N/A	<div></div>
user	user@llmstudio.com	User	Active	N/A	<div></div>
asdfg	asdfg@gmail.com	Technician	Disabled	N/A	<div></div>
andrei	andreimoldovan.amg@gmail.com	Admin	Active	N/A	<div></div>
adminu	adminu@gmail.com	Admin	Active	N/A	<div></div>
andrei2	andrei2@gmail.com	User	Active	N/A	<div></div>
2andrei	andrei@gmail.com	User	Active	N/A	<div></div>

5.3 Vizualizare LLM-uri

LLM Studio

Admin

Search Users or LLMs

admin

Logout

Overview

User Management

LLM Management

Admin Chat

LLM Models Viewer

llama-2-7b-chat-q2-k

Type: Llama

Source: /app/models/llama-2-7b-ch

Context: 2,048

Threads: 4

GPU Layers: 0

Temp: 0.7

llama-2-7b-q2-k

Type: Llama

Source: /app/models/llama-2-7b-q2

Context: 2,048

Threads: 4

GPU Layers: 0

Temp: 0.7

mistral-7b-v0-1-q3-k-l

Type: Llama

Source: /app/models/mistral-7b-v0

Context: 2,048

Threads: 4

GPU Layers: 0

Temp: 0.7

tinyllama-1-1b-chat-v1-0-q2-k

Type: Llama

Source: /app/models/tinyllama-1.1

Context: 2,048

Threads: 4

GPU Layers: 0

Temp: 0.7

5.4 Vizualizare Ticket-uri

LLM Studio

Admin

Search Users or LLMs

admin

Logout

Overview

User Management

LLM Management

Admin Chat

Support Tickets

test_ticket

open

Apr 9, 2025, 04:18 PM

test_ticket

closed

Apr 9, 2025, 04:09 PM

test

open

Apr 8, 2025, 05:34 PM

test

open

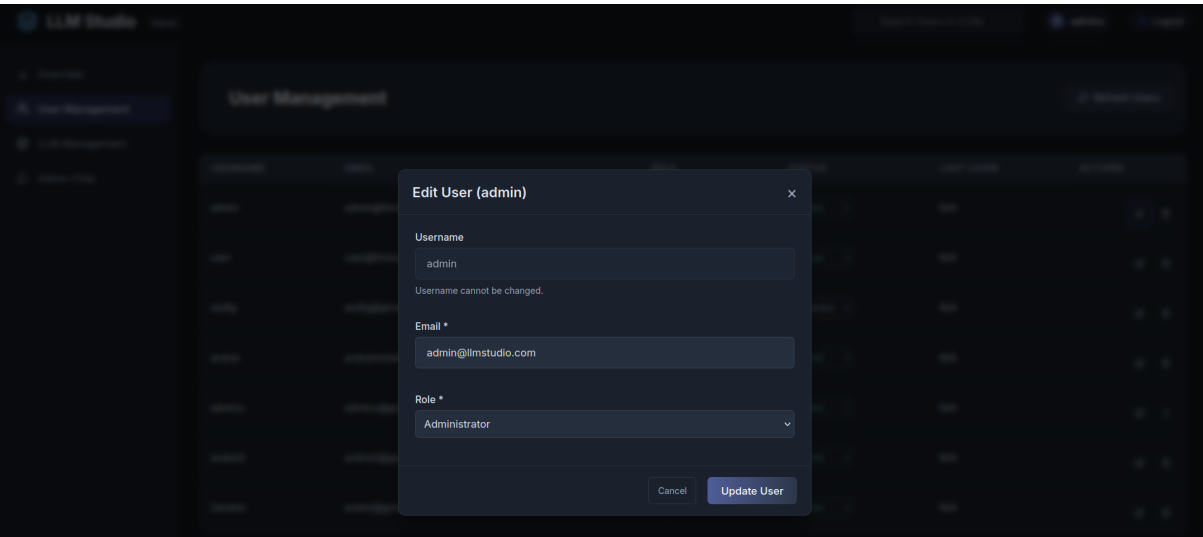
Created: Apr 8, 2025, 05:34 PM Updated: Apr 9, 2025, 02:20 PM

Close Ticket

hi

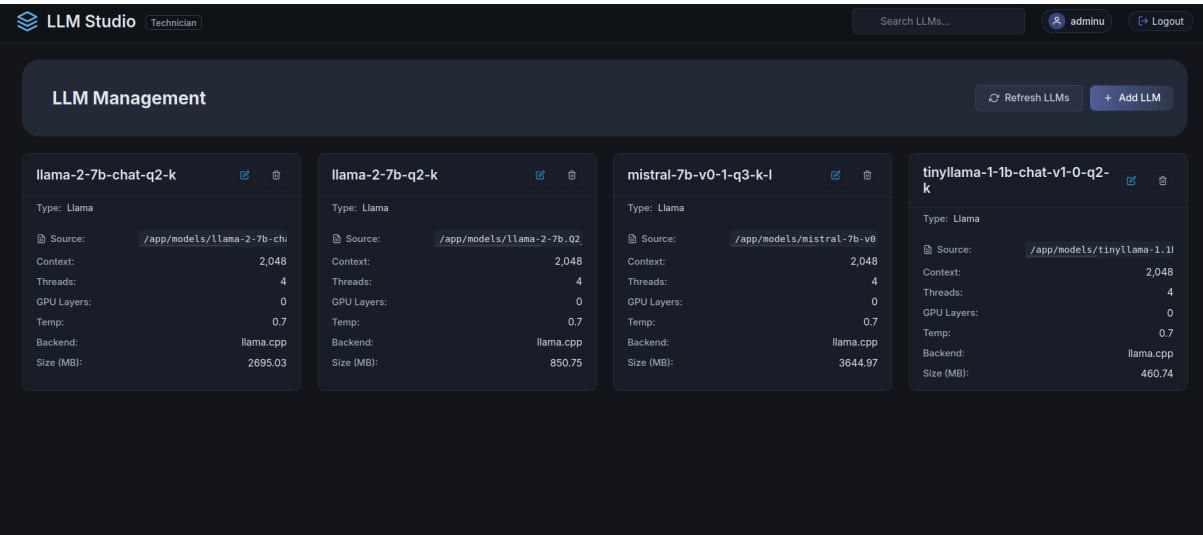
Apr 8, 2025, 05:34 PM

5.5 Modificare Utilizator

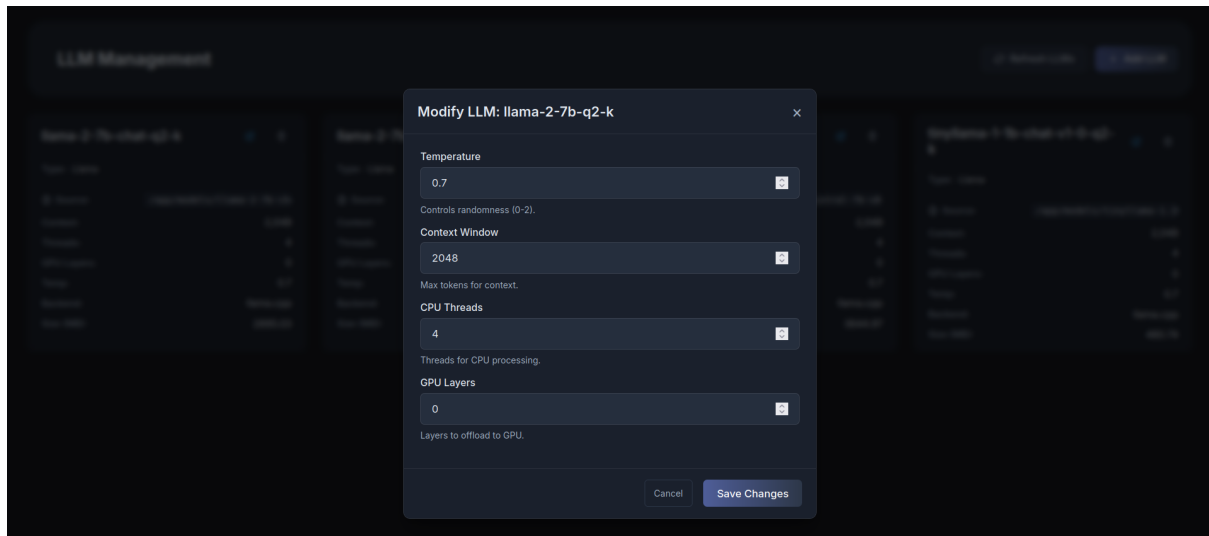


6 Interfața Tehnicianului

6.1 Gestionare LLM-uri



6.2 Modificare LLM



7 Instalare și Rulare

Pentru a instala și rula aplicația:

```
docker compose --file docker-compose-dev.yml up --build
```

8 Adaugare LLM-uri

Exemplu de link valid pentru adăugarea unui LLM: https://huggingface.co/TheBloke/Llama-2-7B-GGUF/resolve/main/llama-2-7b.Q5_0.gguf

9 Roluri Utilizatori

- **Utilizator:** Interacționează cu LLM-uri, salvează conversații și vizualizează date.
- **Administrator:** Gestionează utilizatorii și informațiile despre LLM.
- **Tehnician:** Adaugă, modifică sau șterge LLM-uri.