

# BotBlocker

Your Voice Matters. Don't Let AI Drown It Out

Milestone 4

Ângela Ribeiro (109061)  
João Santos (110555)  
João Viegas (113144)  
Hugo Castro (113889)  
Ana Rita Silva (114220)

## Advisors:

Prof. João Rafael Almeida  
Vicente Barros

# PROJECT OVERVIEW

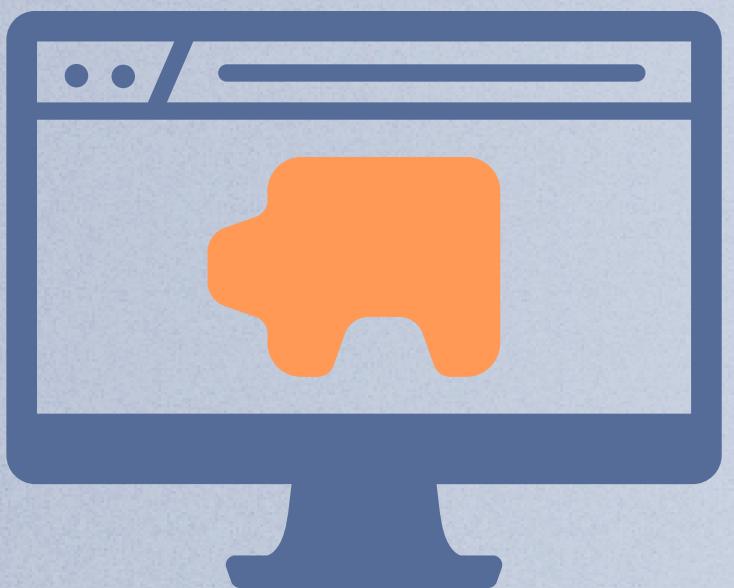
---

BotBlocker is a browser extension and website that uses community intelligence to identify and block automated social media profiles.

# TABLE OF CONTENTS

- 01      **Introduction and Motivation**
- 02      **The State of Related Work**
- 03      **Architecture**
- 04      **Technical Details**
- 05      **Implemented Features**

# INTRODUCTION AND MOTIVATION



## AI Bots

Manipulating engagement  
and spreading  
misinformation.

## Website + Extension

Detect and block interactions  
with AI profiles.

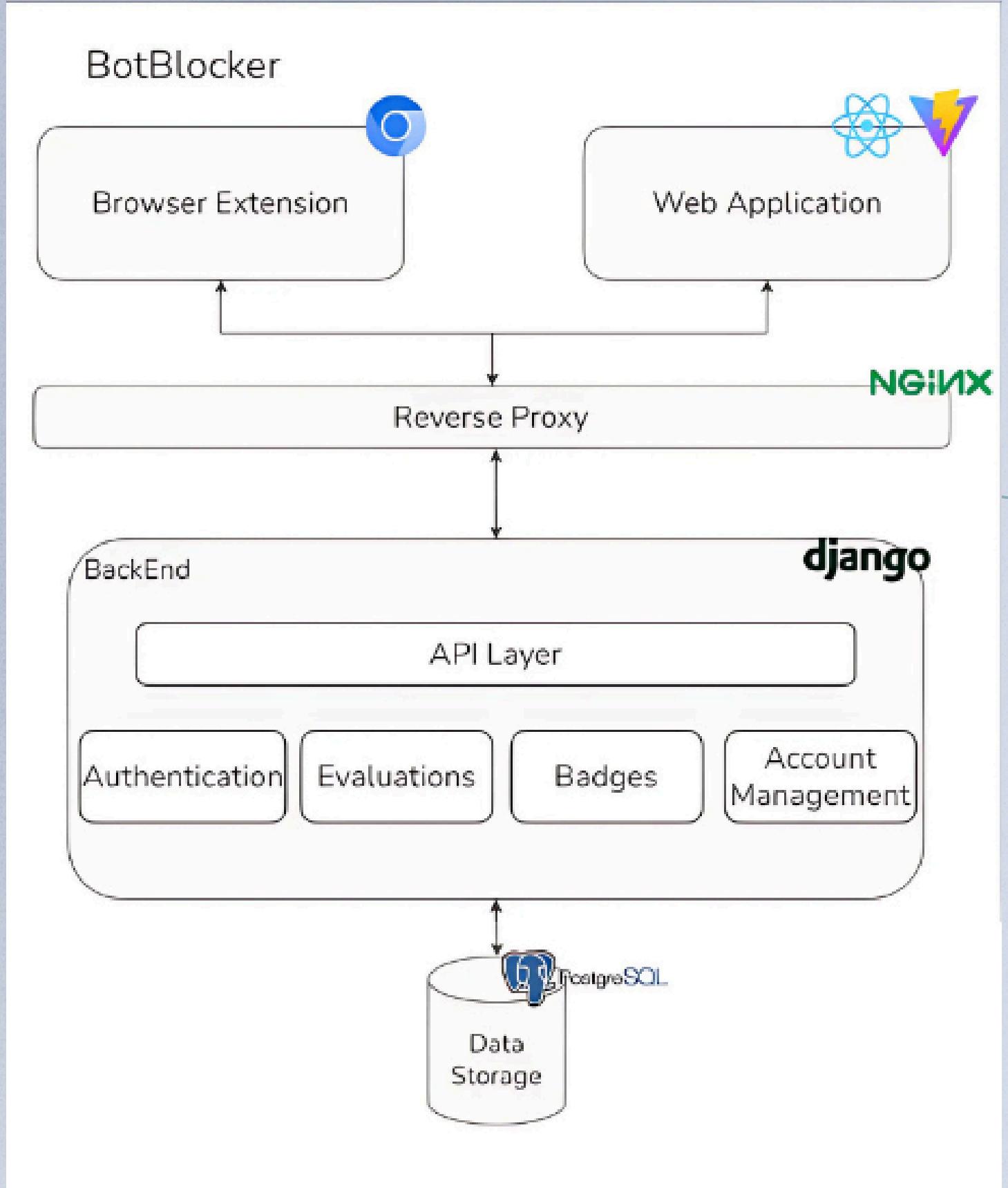
## Detection and Blocking

Identifying AI-driven  
accounts for transparency.

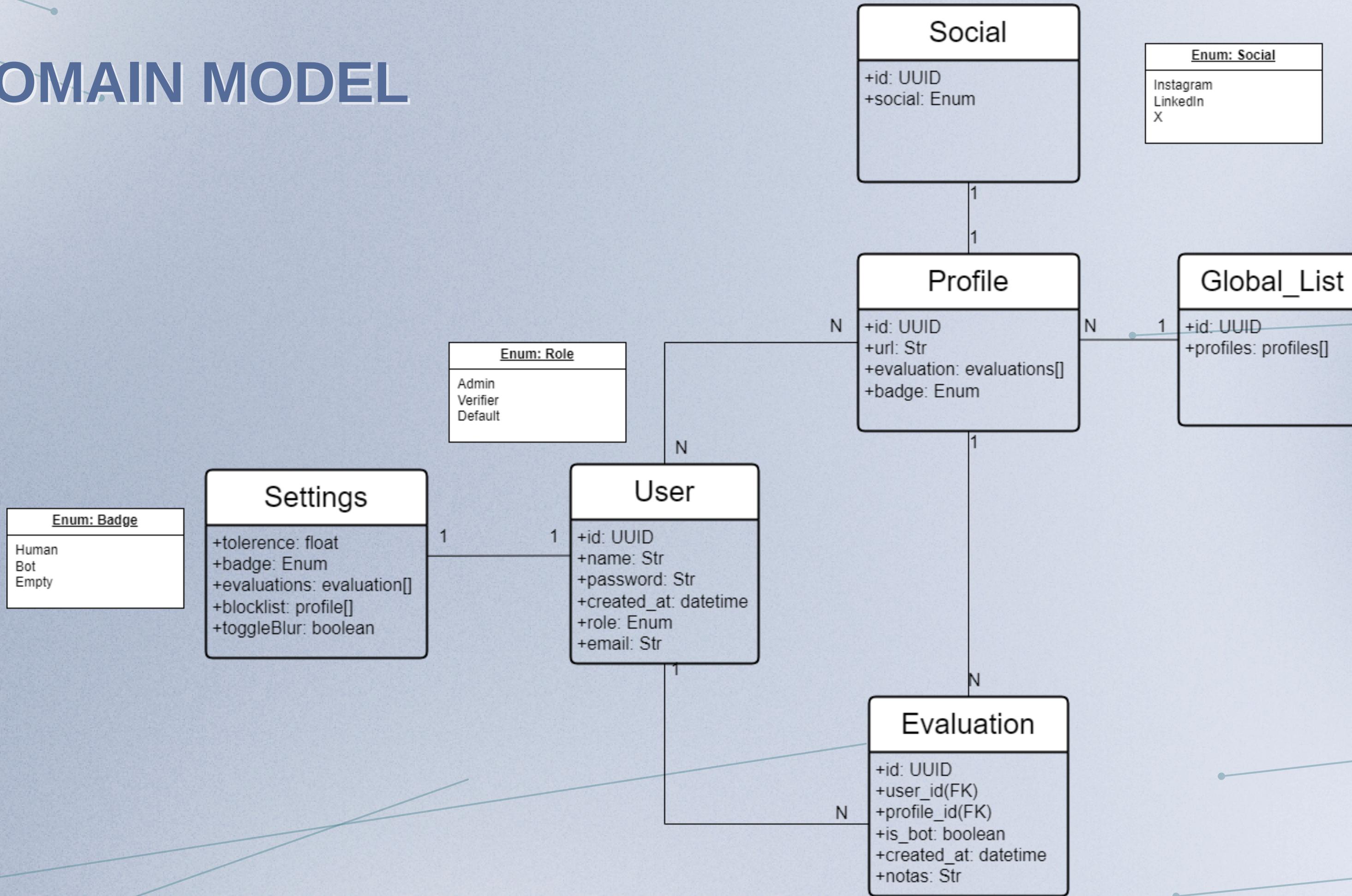
# THE STATE OF RELATED WORK

CRITERION	 <b>PEGABOT</b>	 <b>LIVEDUNE</b>	 <b>Bot Sentinel</b>	 <b>BotBlocker</b>
<b>Main Social Network</b>	Twitter/X	Multiple Social Networks	Twitter/X	Multiple Social Networks
<b>User Interaction</b>	None	None	Score Display	Community Voting & Feedback
<b>Blocking Management</b>	Not Available	Not Available	Reliability Score	Personal & Community Blocklists
<b>Accessibility</b>	Free	\$15,30/month for 5 accounts	Free	Free and open to contributions
<b>Main Advantages</b>	Ease of use, free access	Artificial Engagement Identification	High accuracy in identifying manipulative profiles	All of the before + Comprehensiveness, active user participation, integration with multiple networks
<b>Main Limitations</b>	Currently inactive	Focus only on influencers and advertising	Restricted to Twitter/X; does not detect neutral bots	None of the before
<b>Focus of Analysis</b>	Behavior and Profile Features	Engagement and Audience Authenticity	Toxicity and Coordinated Manipulation	Bot Networks, Misinformation, and Coordinated Attacks

# ARCHITECTURE



# DOMAIN MODEL



# TECHNICAL DETAILS

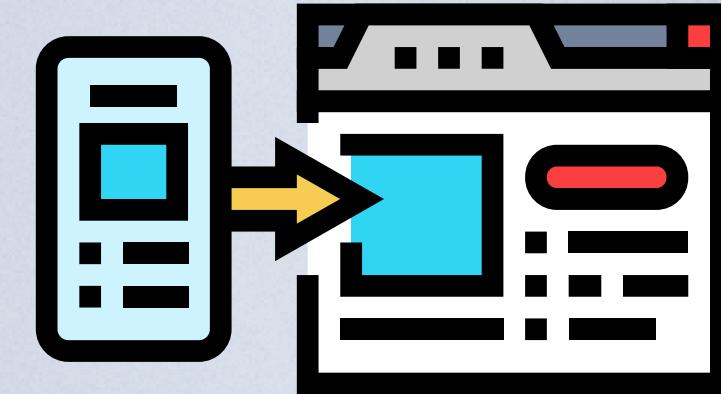
## Modular & Extensible

- Frontend and browser extension built with React and JSX
- Backend developed with a RESTful API
- Clean, modular codebase – easy to maintain and extend
- Designed to support multiple social platforms
- API logic and UI components are decoupled



## Local-First Design

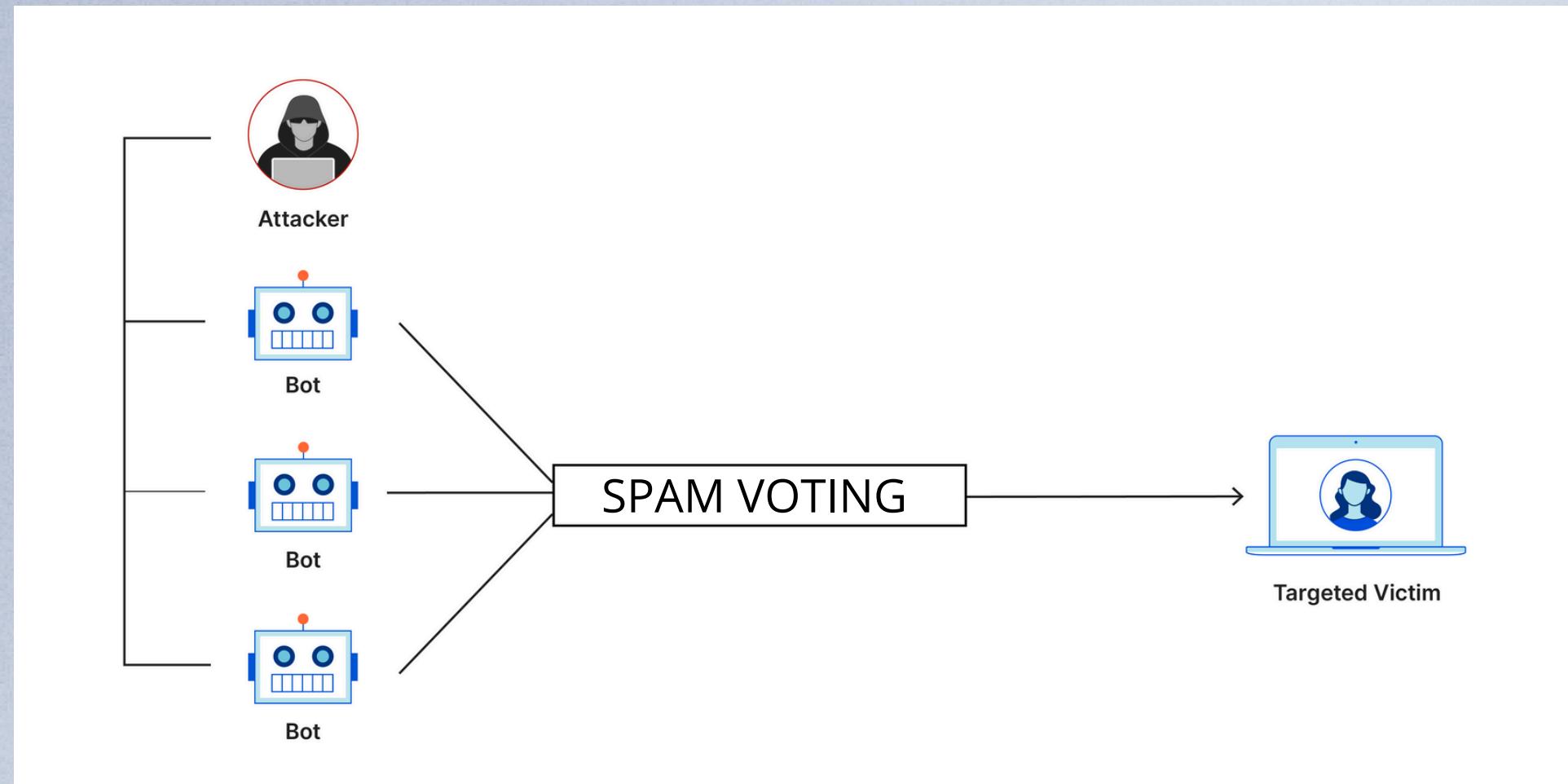
- Works without login: data is stored locally
- Optional login enables syncing blocklists and preferences across devices
- Sync handled via custom backend with session-based authentication



# TECHNICAL DETAILS

## Spam Voting Protection

- Admins and verifiers monitor abnormal voting patterns
- Account blocking and timeouts reduce spam voting risks
- Web Sockets usage



# IMPLEMENTED FEATURES

M1



# BotSentinel

**Advisor:** João Almeida ([joso.rafael.almeida@ua.pt](mailto:joso.rafael.almeida@ua.pt))

## 1. Background

In recent years, the rise of Artificial Intelligences capable of generating coherent texts, realistic images, and, more recently, interacting on social media has sparked growing debate about the authenticity of online profiles. Twitter, one of the most influential social platforms, has become a fertile ground for AI-driven bots, which can be used to spread misinformation, manipulate discussions, and create artificial engagement. The ease with which an automated profile can mimic human behavior raises concerns about the reliability of information and the formation of genuine opinions in digital spaces.

The Dead Internet Theory serves as an inspiration for this project. It suggests that much of the online content we consume may be artificially generated rather than produced by real people. According to this theory, the internet as we know it is largely dominated by bots and automated accounts, causing a significant shift in user behavior and engagement over time; humans have increasingly interacted with automated profiles without realizing that these interactions are not genuine. This shift could lead to a loss of authenticity, where real human voices are replaced by algorithms that shape behaviors and opinions. While this idea might sound exaggerated or like a conspiracy theory, the reality is that we are moving in that direction. Today's social media landscape is dominated by automation, with a chaotic and unpredictable nature of the past. The omnidirectional ecosystem of user-related profiles we see has been replaced by a handful of large platforms dominated by corporations, prioritizing the monetization of interactions—often at the expense of user experience. We are not yet fully immersed in the dystopia described by the theory, but the signs of this transformation are becoming increasingly evident, making it crucial to question and understand its implications.

This project aims to develop a system that enables users to identify being AI-driven, leveraging a combination of community-based evaluation tools with user-friendly interfaces to detect and class increase transparency in online interactions and ensuring that digital content is not covertly manipulated by AI.

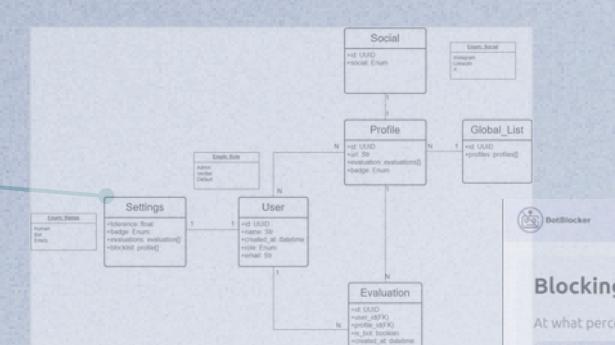
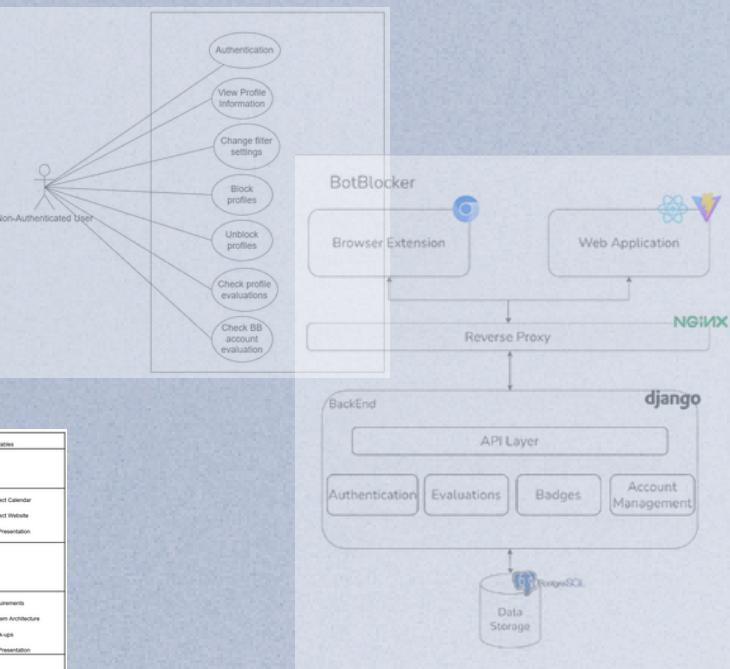
## 2. Project Objectives

The project aims to develop an innovative solution for identifying artificial intelligence (AI). While existing tools address this issue in a general way, this project seeks to provide direct and precise insights into specific accounts. The objectives are:

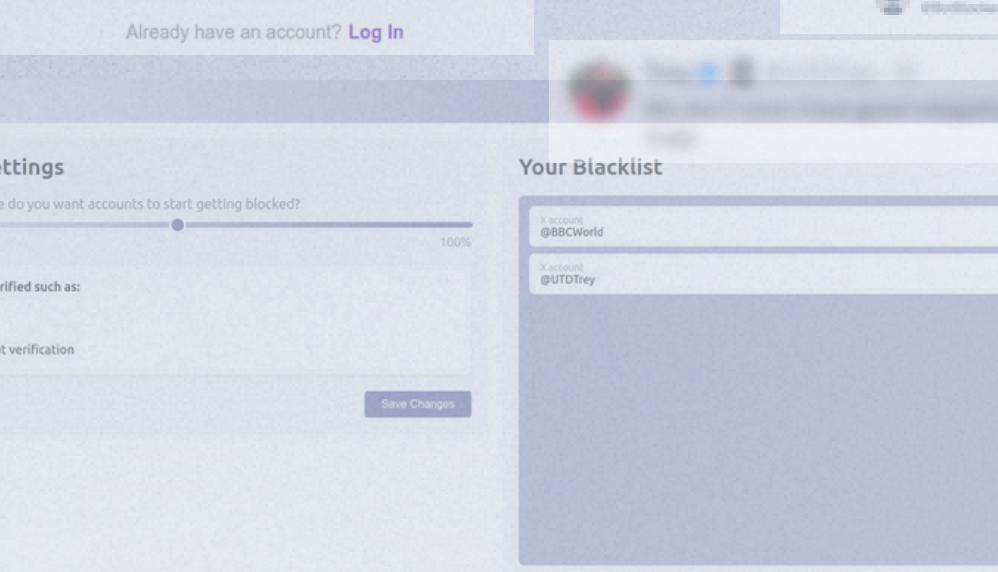
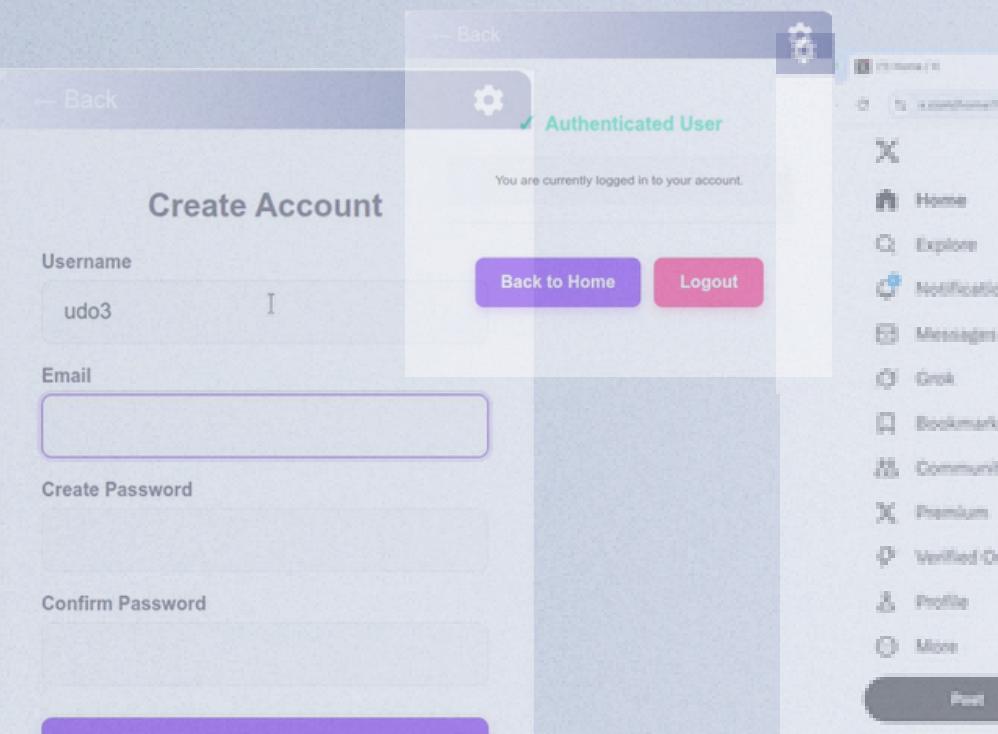
- Develop an accessible and flexible platform, available as both a plug-in. The plug-in will automatically display an account's credibility while the website will allow users to analyze profiles through a mobile device.

Date	Task List	Description	Deliverables
10/02/2023 - 17/02/2023	<b>Module Name: Project Foundation &amp; Documentation</b> Task 1: Define the project scope and mission (Rita Silva, Hugo Castro) Task 2: GitHub Repository Organization (Rita Silva, Hugo Castro) Task 3: Setup of Documentation Website (Rita Silva, Hugo Castro)	Establishing the project's foundation through use-case diagrams, organizing the GitHub repository, and setting up the documentation website to ensure structured development.	1. Project Calendar 2. Project Website 3. M1 Presentation
18/02/2023 - 24/02/2023	<b>Module Name: Project Planning and Framework</b> Task 1: Project Calendar Creation (Everyone) Task 2: Project Charter Preparation (Everyone) Task 3: Project Preparation Preparation (Everyone) Task 4: Project Charter Approval (Rita Silva, Hugo Castro) Task 5: High-Level Architecture Design (Lúcio Castro)	Milestone 1: Completion of Project Foundation, architecture, and contextual framework established.	1. Project Calendar 2. Project Website 3. M1 Presentation
25/02/2023 - 04/03/2023	<b>Module Name: Requirements &amp; UX Design</b> Task 1: Requirement Gathering with Everyone Task 2: Requirements Document Creation (Everyone) Task 3: Define Personas and Scenarios (Lúcio Castro) Task 4: Define Personas and Scenarios (Rita Silva, Hugo Castro)	Defining project needs, user personas, and scenarios while establishing functional and non-functional requirements with initial mockups.	1. Requirements 2. System Architecture 3. Mock-ups 4. M2 Presentation
04/03/2023 - 11/03/2023	<b>Module Name: System Architecture</b> Task 1: Refine the state of the UI (Lúcio Castro) Task 2: Technologies and Modules Selection (Everyone) Task 3: Define the Database Model (Lúcio Castro) Task 4: Map the Deployment Diagram (Rita Silva, Hugo Castro)	Refining research, defining modules, and creating deployment diagrams to structure the system.	1. Requirements 2. System Architecture 3. Mock-ups 4. M2 Presentation
11/03/2023 - 18/03/2023	<b>Module Name: Extension, Inflation, and Database Setup</b> Task 1: Define the project scope (Rita Castro) Task 2: Define the extension (Rita Castro) Task 3: Extension Inflation and Configuration (Rita Silva, Lúcio Castro) Task 4: Development of Extension Frontend (Rita Castro) Task 5: Backend API Implementation for Profile Identification (Lúcio Castro)	Establishing the database, configuring the extension, developing the frontend, and implementing the backend API for seamless integration.	

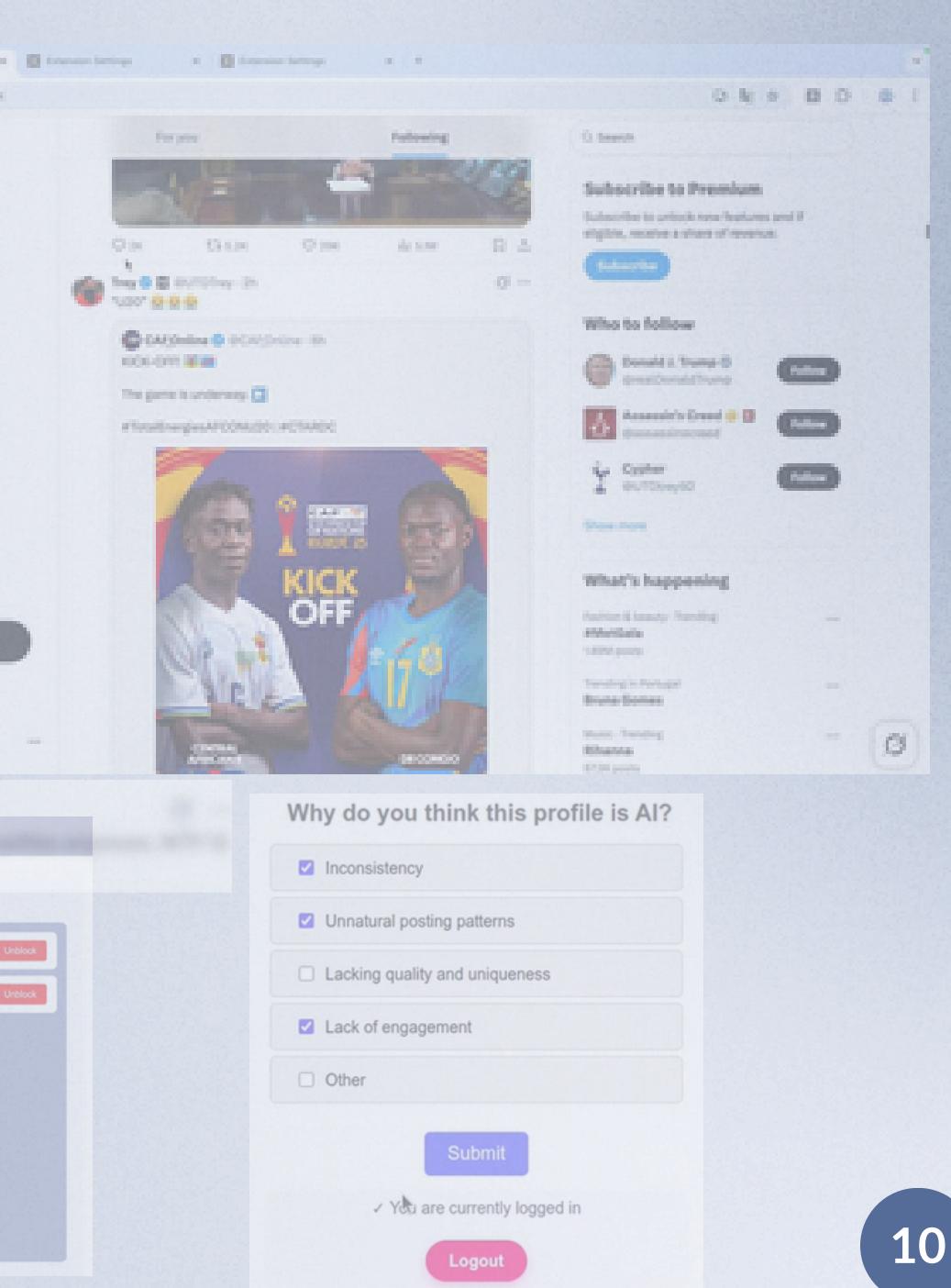
M2



# Checkpoint 1



# Checkpoint 2

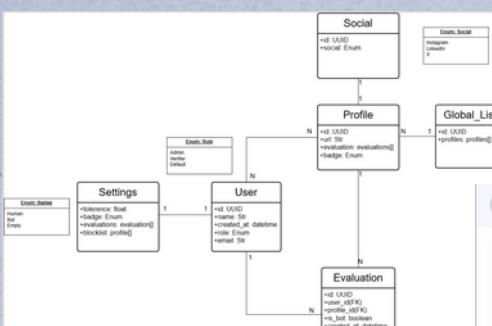
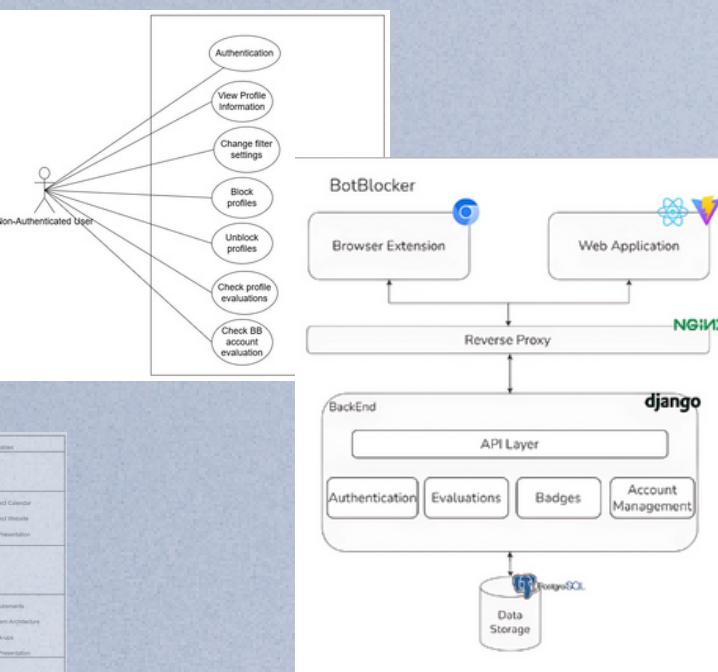


# IMPLEMENTED FEATURES

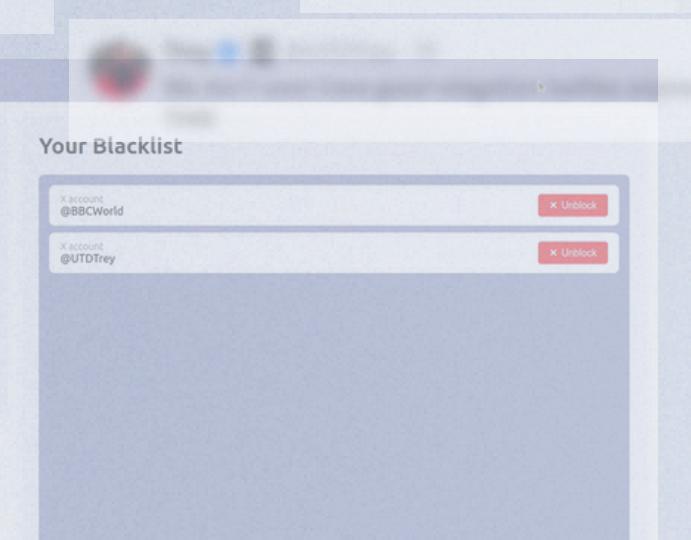
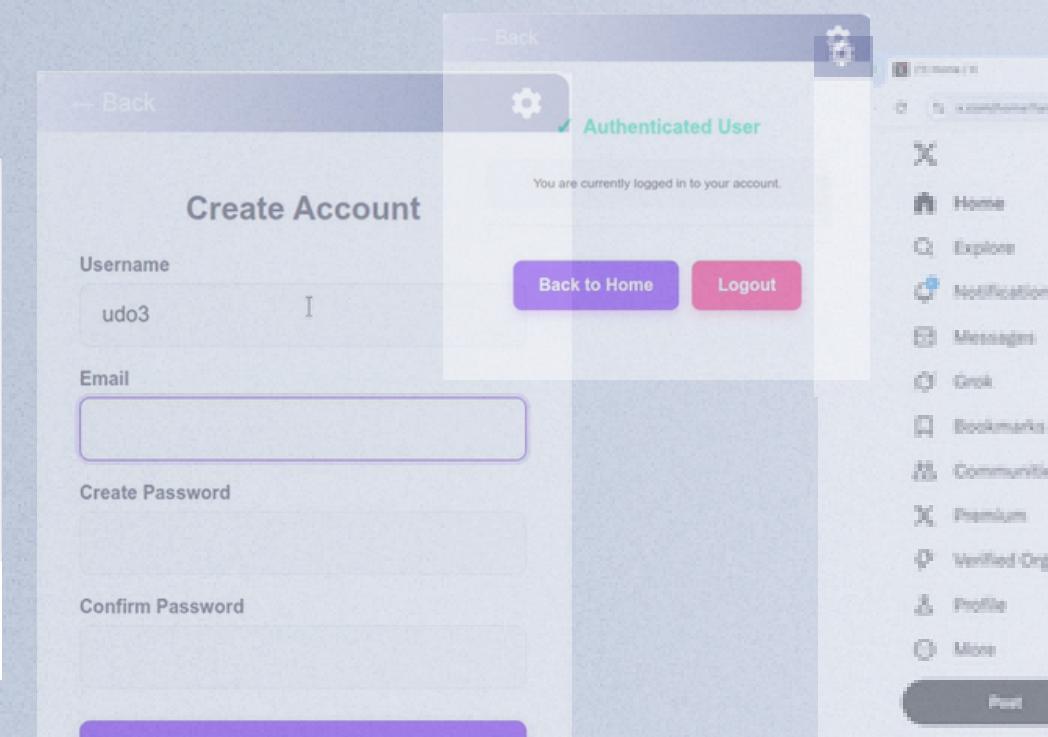
M1



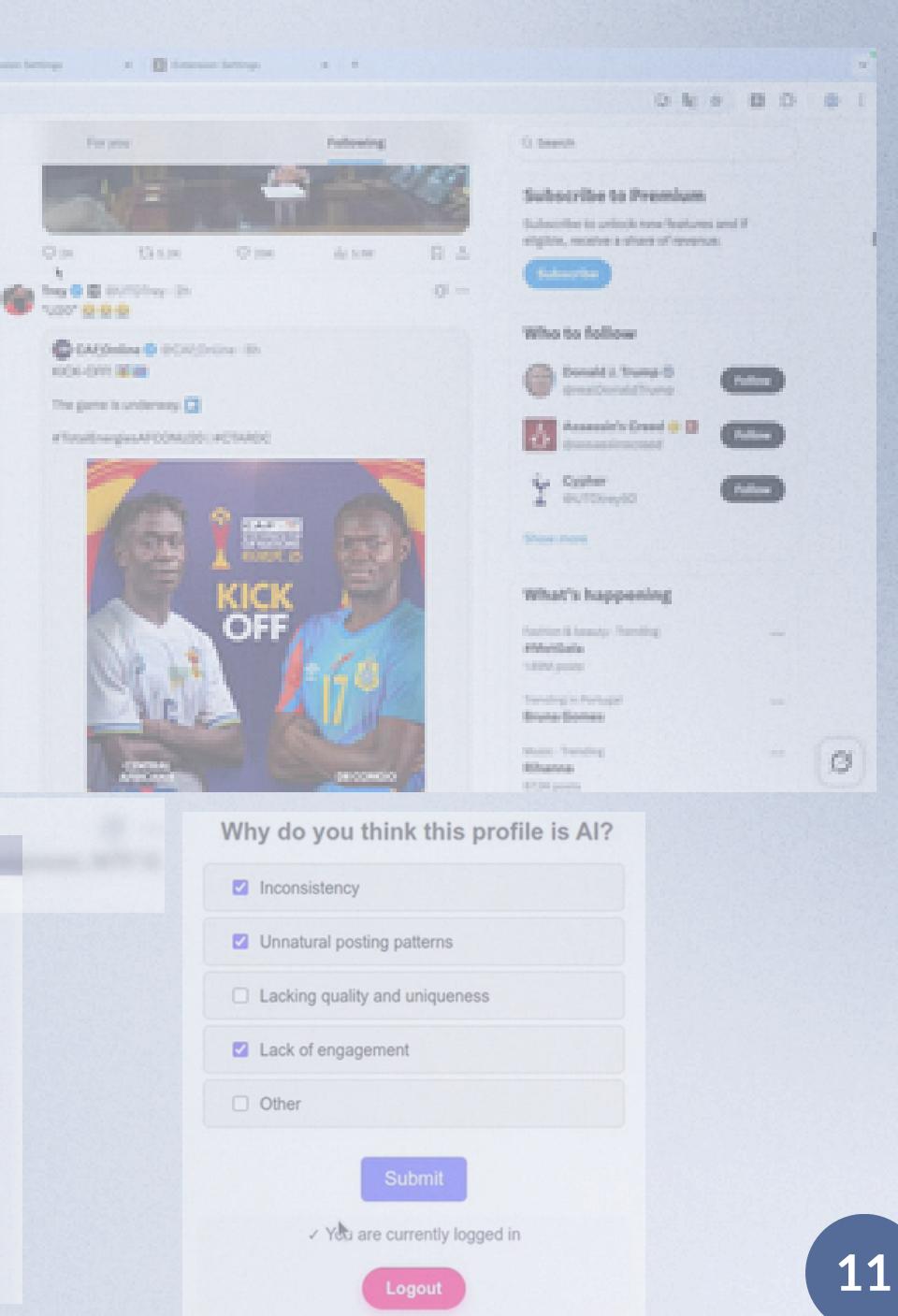
M2



# Checkpoint 1



# Checkpoint 2

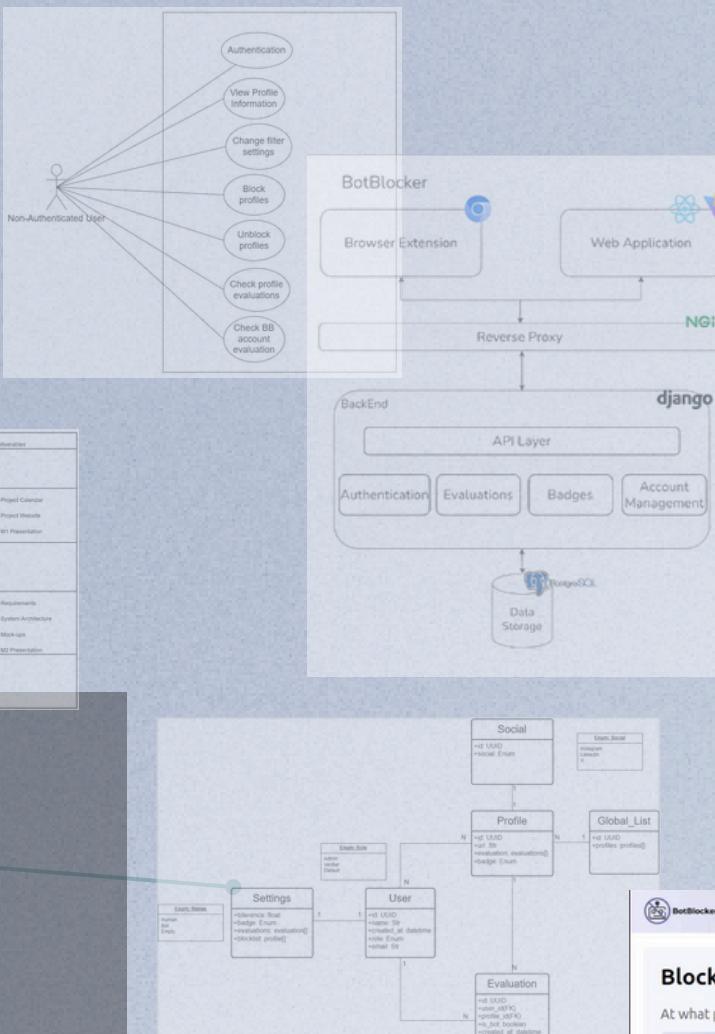


# IMPLEMENTED FEATURES

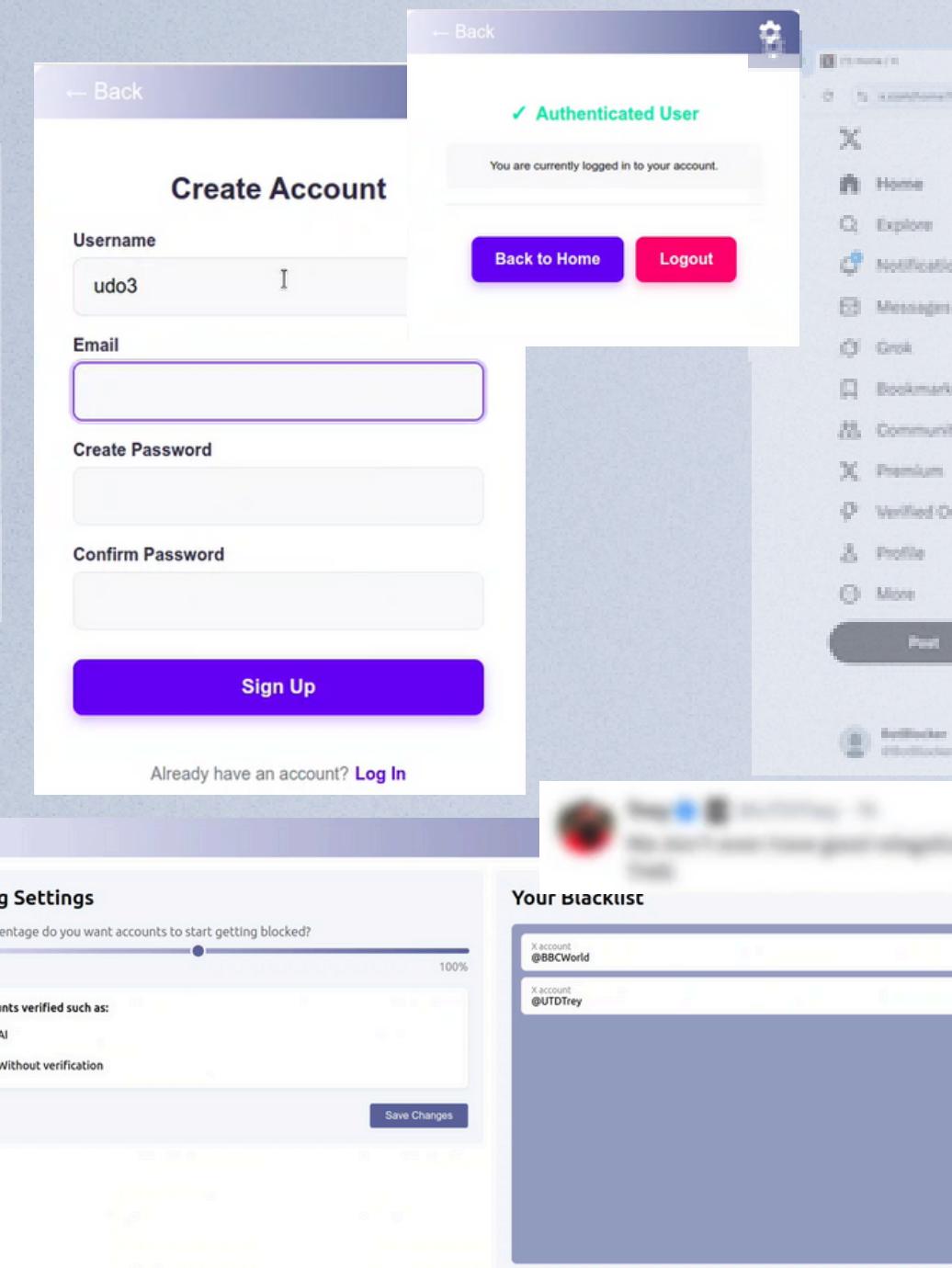
M1



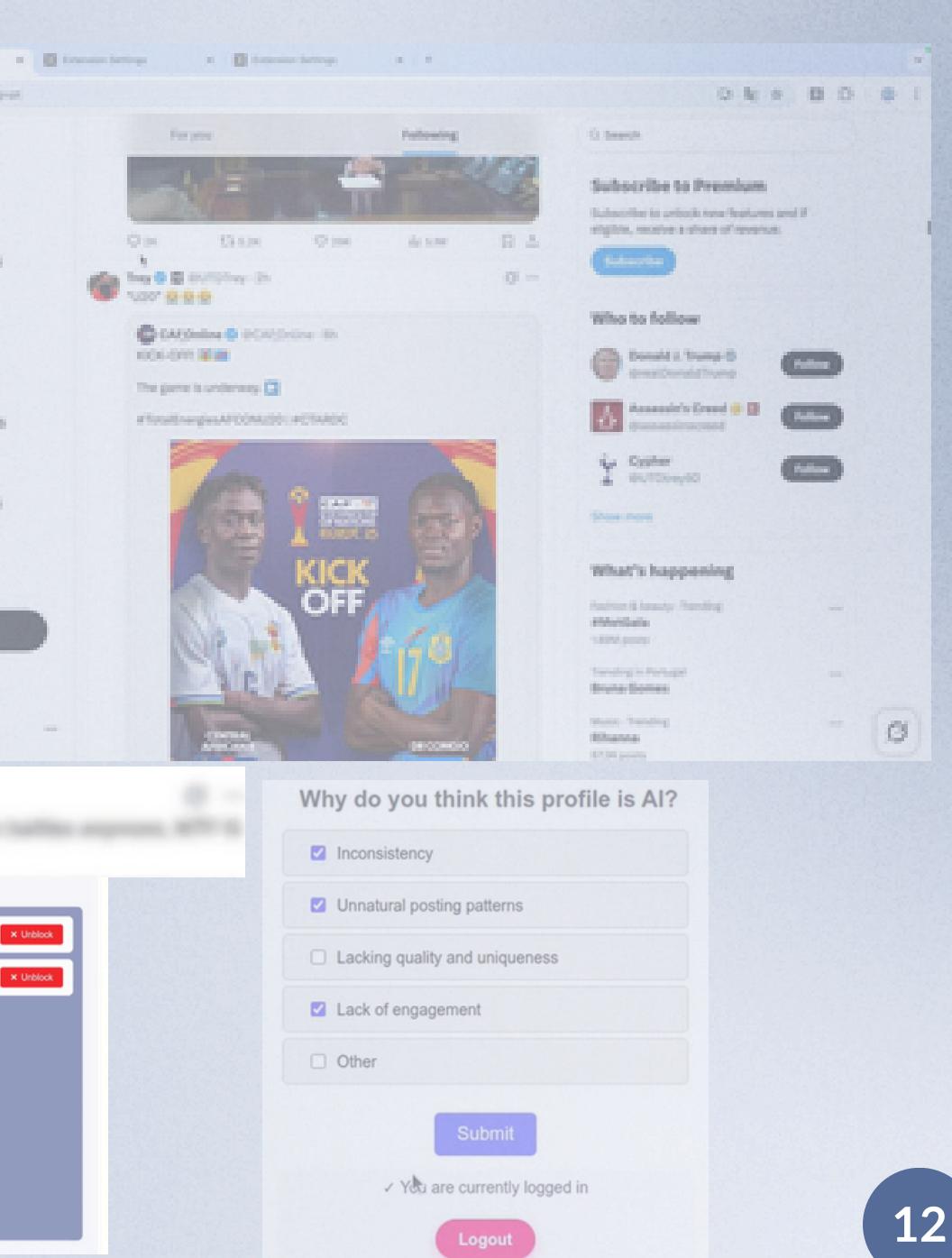
M2



# Checkpoint 1



# Checkpoint 2



# IMPLEMENTED FEATURES

M1



# BotSentinel

Advisor: Júlio Almeida ([jua.rafael.almeida@ua.pt](mailto:jua.rafael.almeida@ua.pt))

## 1. Background

In recent years, the rise of Artificial Intelligences capable of generating coherent text, realistic images, and, more recently, interacting on social media has sparked growing debate about the authenticity of online profiles. Twitter, one of the most influential social platforms, has become a fertile ground for AI-driven bots, which can be used to spread misinformation, manipulate discussions, and create artificial engagement. The ease with which an automated profile can mimic human behavior raises concerns about the reliability of information and the formation of genuine opinions in digital spaces.

The Dead Internet Theory serves as an inspiration for this project. It suggests that much of the online content we consume may be artificially generated rather than produced by real people. According to this theory, the internet as we know it is largely dominated by bots and automated algorithms, resulting in a lack of originality, accountability, and engagement. Over time, humans are increasingly interacted with automated profiles without realizing that these interactions are not genuine. This shift could lead to a loss of authenticity, where real human voices are replaced by algorithms that shape behaviors and opinions. While this idea might sound exaggerated or like a conspiracy theory, the reality is that we are moving in that direction. Today, we are more connected and informed than ever before, but at what cost? The transparency and accountability of the web have been eroded, and the way we interact with the digital world has been replaced by a handful of large platforms dominated by corporations, prioritizing the monetization of interactions—often at the expense of user experience. We are not yet fully immersed in the dystopia described by the theory, but the signs of this transformation are becoming increasingly evident, making it crucial to question and understand its implications.

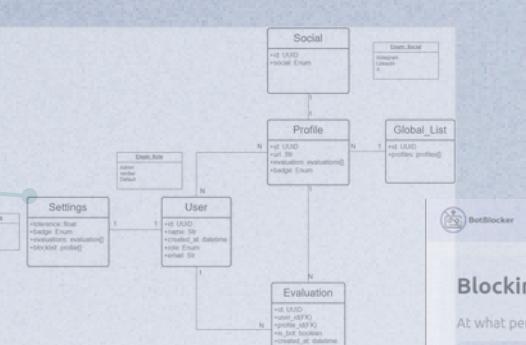
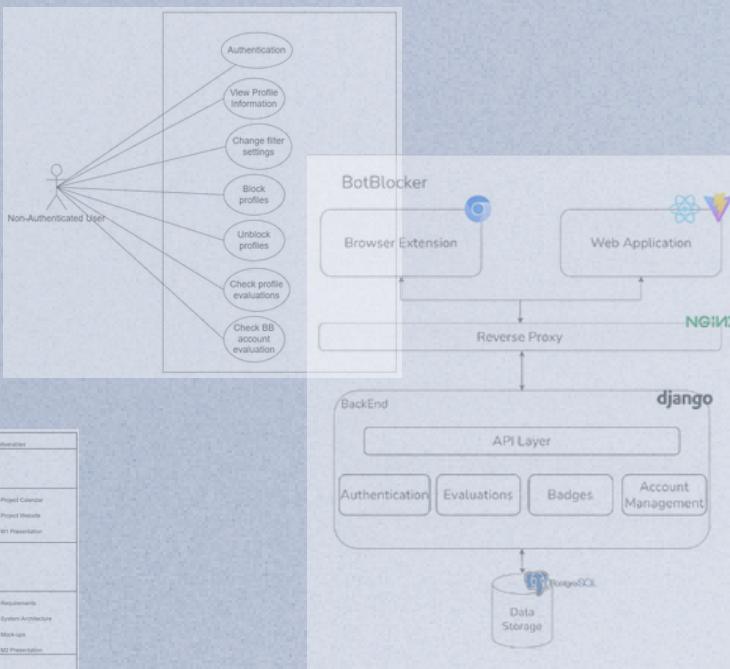
This project aims to develop a system that enables users to identify profiles suspected of being AI-driven, leveraging a combination of community-based evaluation tools. The system will provide users with tools to actively contribute to detecting and classifying bots, helping to increase transparency in online interactions and ensuring that digital communication is not covertly manipulated by AI.

## 2. Project Objectives

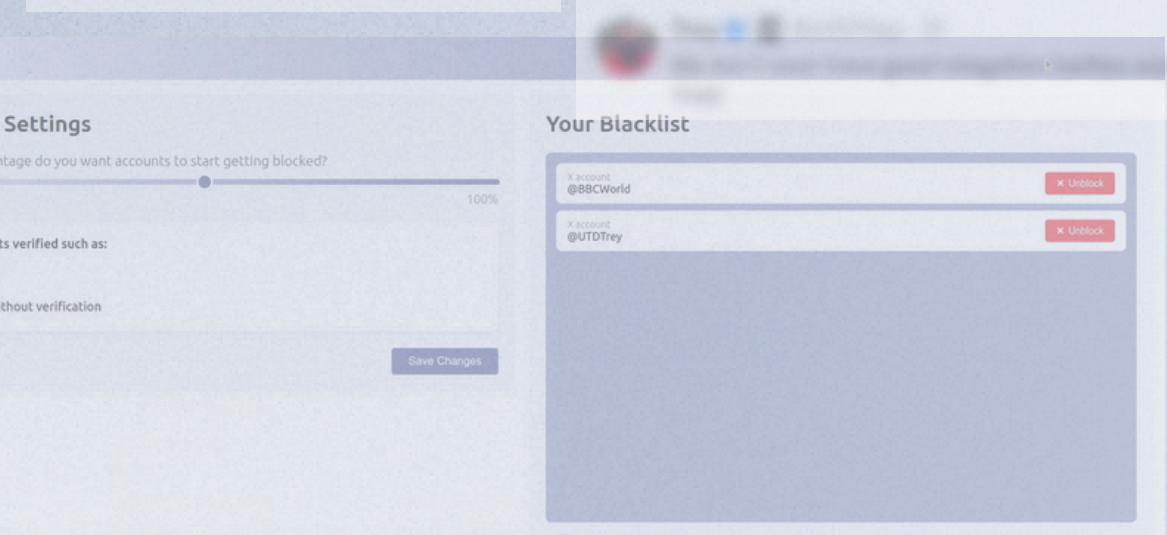
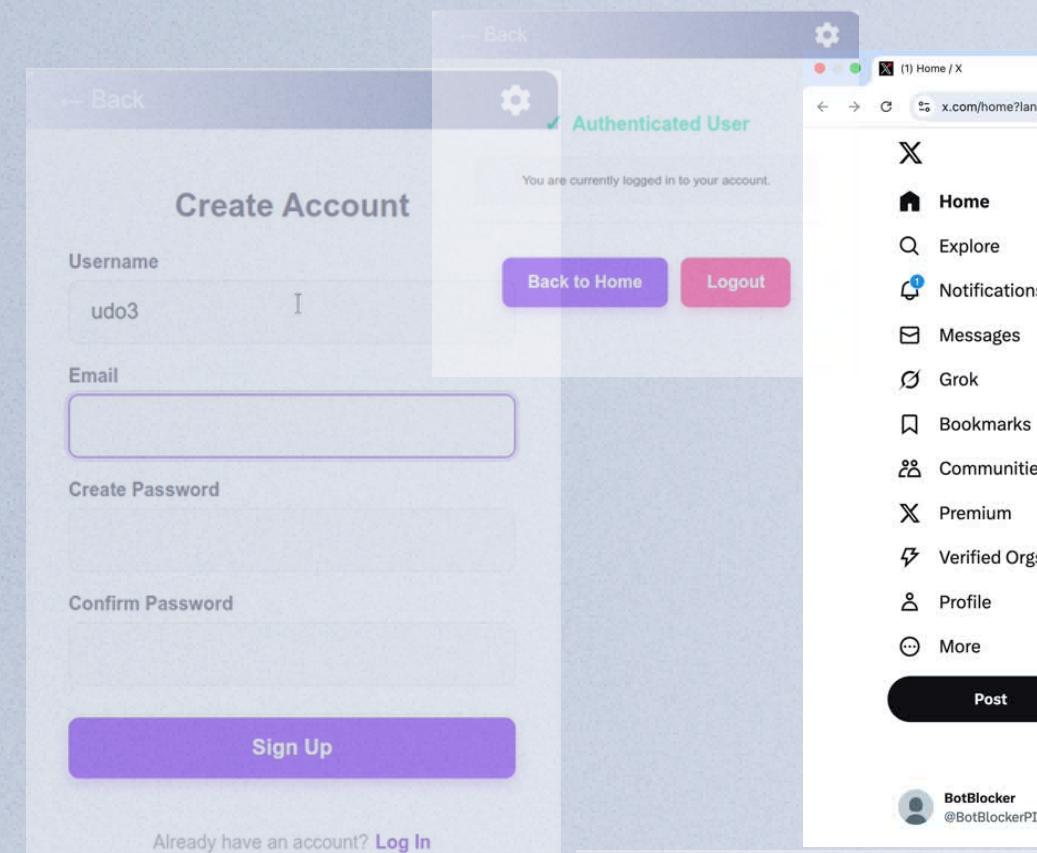
The project aims to develop an innovative solution for identifying AI-generated accounts managed by artificial intelligence (AI). While existing tools address this issue in a general way, this project seeks to provide direct and precise insights into specific accounts. To achieve this, the project aims to:

- Develop an accessible and flexible platform, available as both a website and a browser plug-in. The plug-in will automatically display an account's credibility while browsing, while the website will allow users to analyze profiles through a simple search interface, ideal for mobile devices.

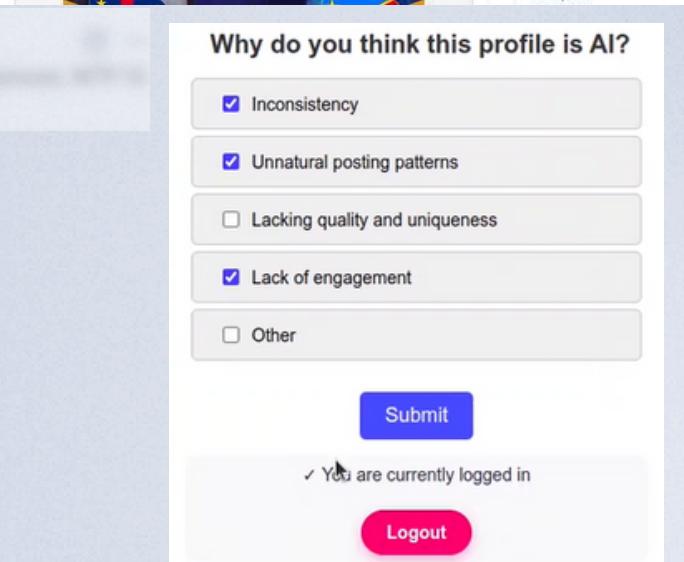
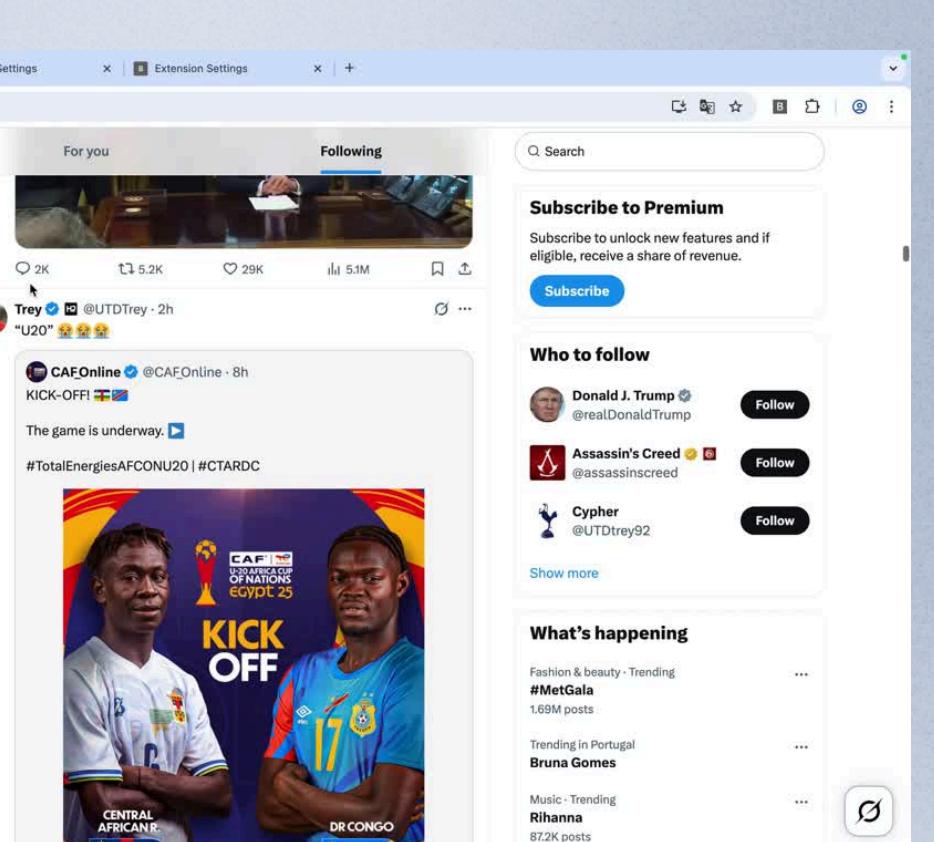
M2



# Checkpoint 1

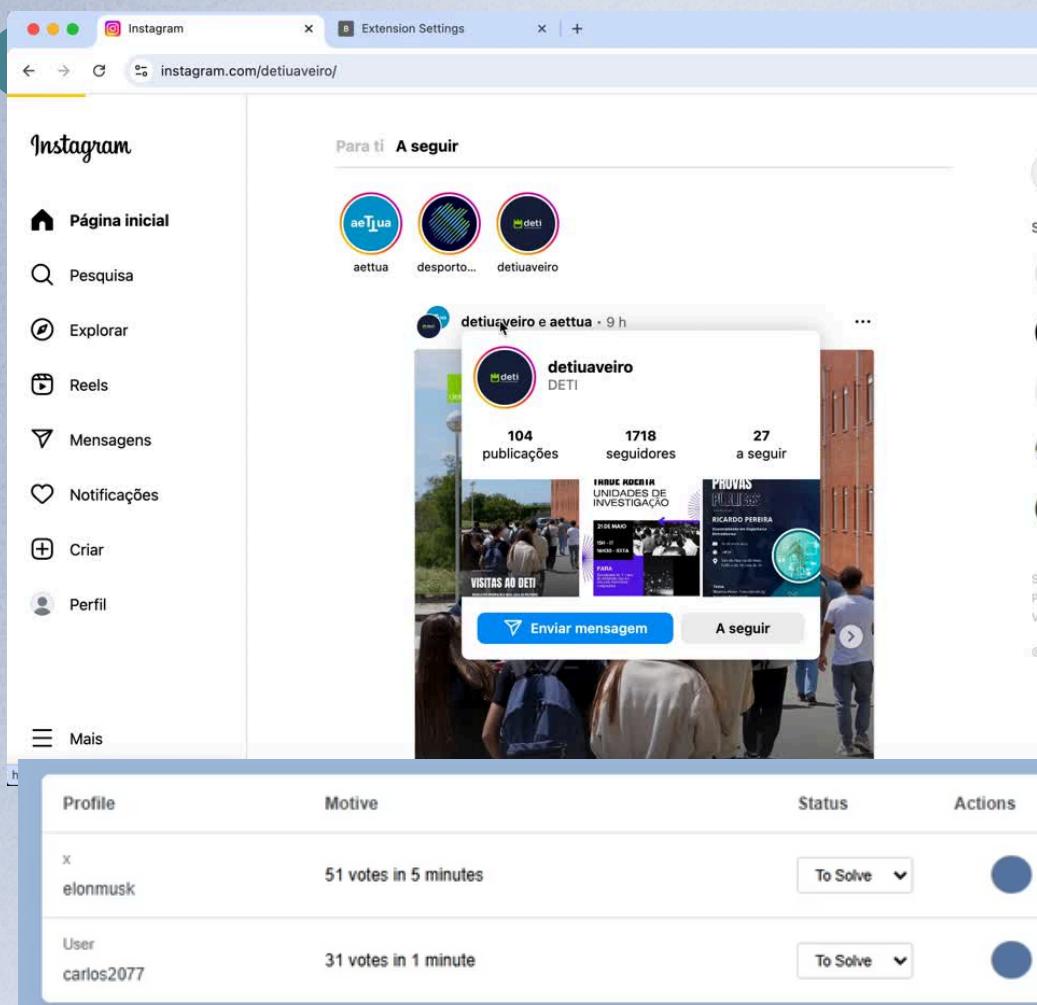


# Checkpoint 2



# IMPLEMENTED FEATURES

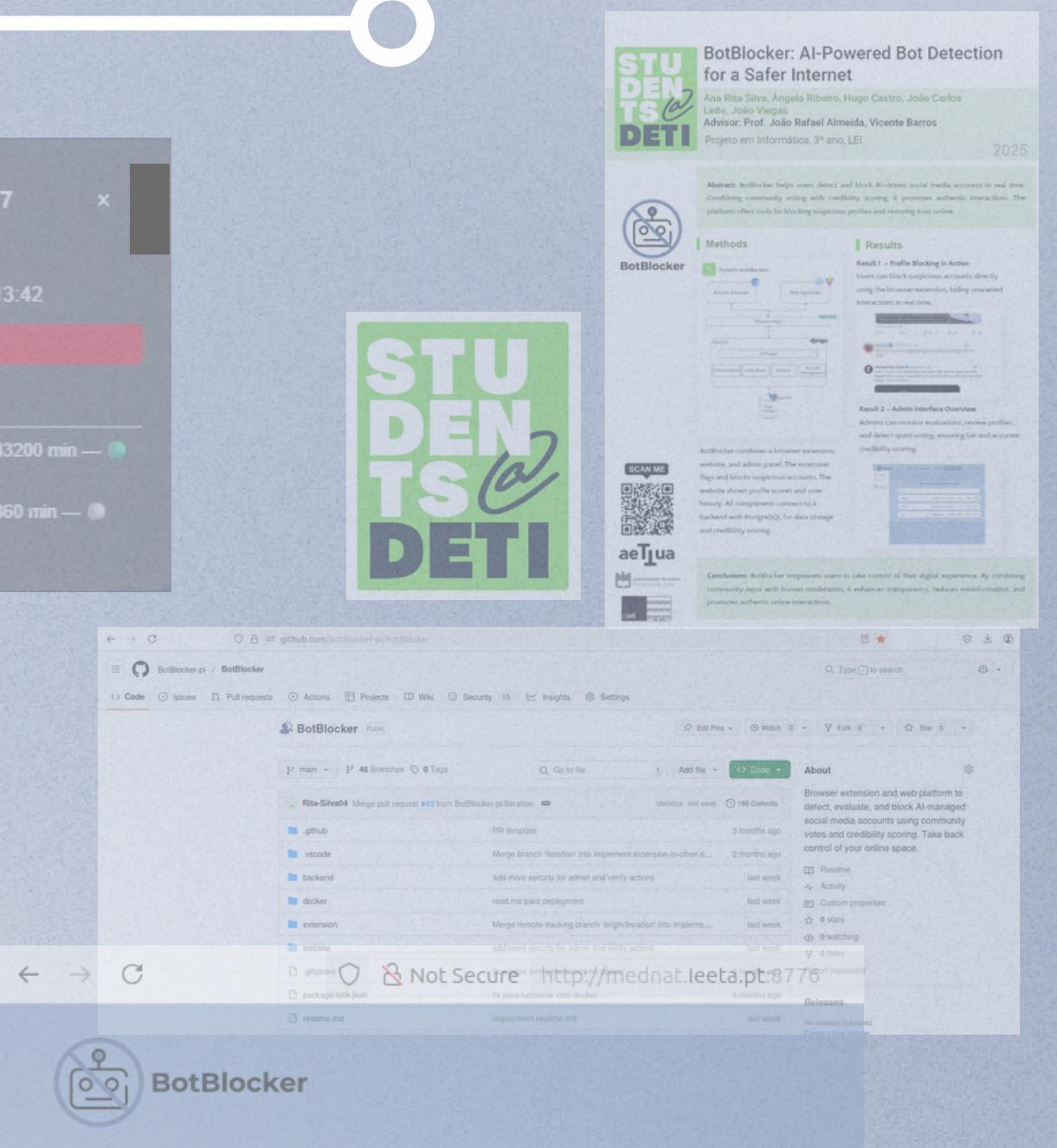
M3



Checkpoint 3

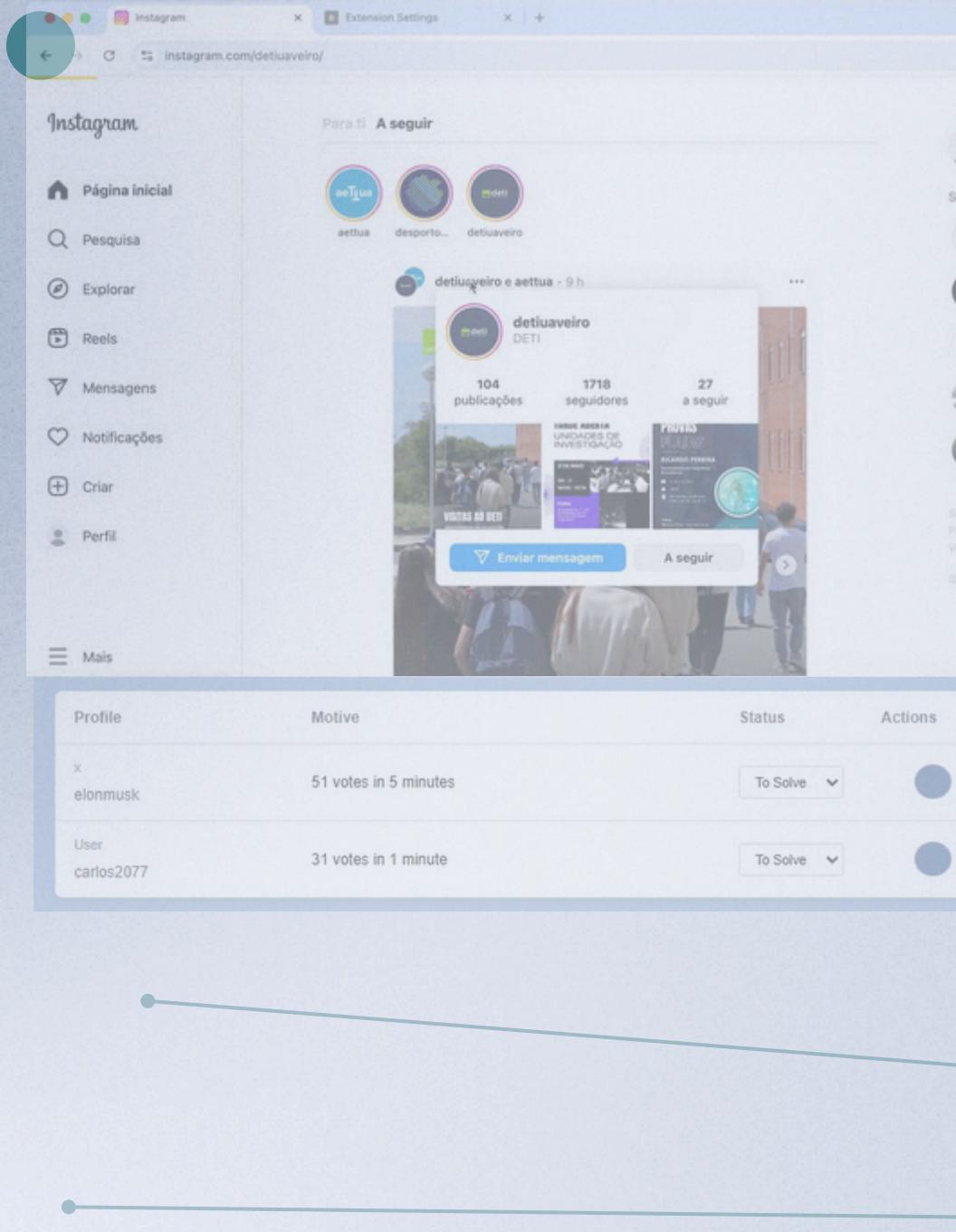
This block contains two screenshots. The top one is a modal titled 'Timeout: @carlos2077' showing an active timeout ending at 18/06/2025, 22:13:42. It lists recent timeouts for users @pardo27, @carlos2077, and @rafaelodias54. The bottom one is a screenshot of a GitHub repository for 'BotBlocker' with 46 branches and 186 commits, showing activity and pull requests.

M4



# IMPLEMENTED FEATURES

M3



Checkpoint 3

Timeout: @carlos2077

Active Timeout  
Ends at: 18/06/2025, 22:13:42

Revoke

Recent Timeouts

19/05/2025, 22:13:42 — 43200 min — Active

19/05/2025, 22:13:03 — 360 min — Inactive

M4

STUDENTS@DETI

BotBlocker: AI-Powered Bot Detection for a Safer Internet

Ana Rita Silva, Ângela Ribeiro, Hugo Castro, João Carlos Lobo, João Viegas

Advisor: Prof. João Rafael Almeida, Vicente Barros

Projeto em Informática, 3º ano, LEI.

Result 1 - BotBlocker Overview

BotBlocker helps users detect and block AI-driven social media accounts in real time. Combining community voting with credibility scoring, it promotes authentic interactions. The platform offers tools for blocking suspicious profiles and restoring trust online.

Result 2 - Admin Interface Overview

Admins can monitor evaluations, review profiles, and detect spambots in real time. The website shows profile scores and vote history. All components connect to a database with PostgreSQL, for data storage and credibility scoring.

Result 3 - User Interface Overview

BotBlocker combines a browser extension, an admin panel, and a web platform. The extension flags and blocks suspicious accounts. The website shows profile scores and vote history. All components connect to a database with PostgreSQL, for data storage and credibility scoring.

Result 4 - BotBlocker in Action

Users can block suspicious accounts directly using the browser extension, hiding unwanted interactions in real time.

Result 5 - Privacy and Security

BotBlocker respects user privacy by only collecting data necessary for detection and credibility scoring. It uses strong encryption for data transmission and stores data securely in PostgreSQL.

BotBlocker

Code Issues Pull requests Actions Projects Wiki Security Insights Settings

BotBlocker

Rita-Silva04 Merge pull request #45 from BotBlocker-pi/iteration... PR template 3 months ago

Medbo - last week 186 Commits

PR template Merge branch 'yesterter' into implement-extension-in-other... 2 months ago

Activity add more security for admin and verify actions last week

Custom properties read.me para deployment last week

PR template Merge remote tracking branch 'origin/iteration' into Implement... last week

PR template add more security for admin and verify actions last week

PR template Not Secure http://mednat.leeta.pt:8776 3 months ago

Ignore package-lock.json file para funcionar com docker last week

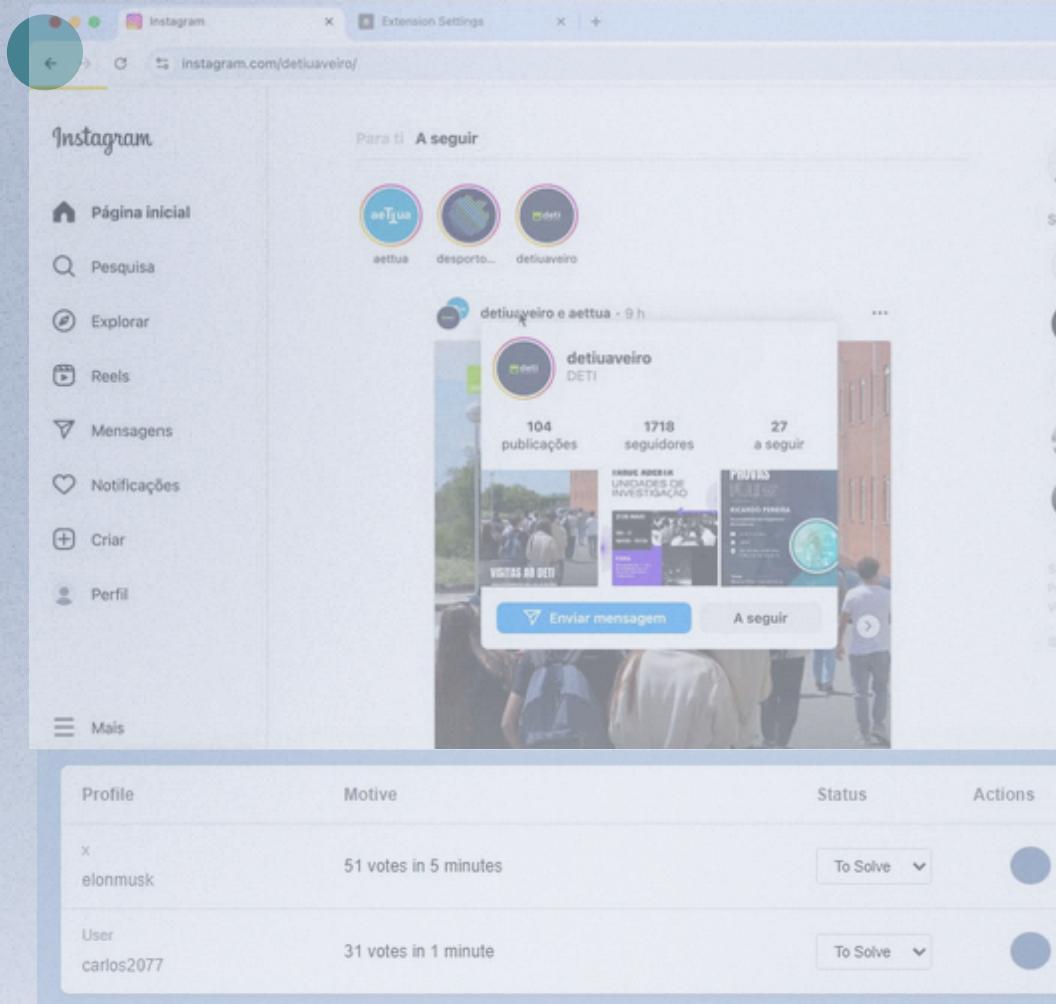
Deployment deployment/readme.md deployment/readme.md last week

Readme README.md deployment/readme.md last week

Releases no releases published

# IMPLEMENTED FEATURES

M3



Checkpoint 3

M4

# Thank you!

<https://botblocker-pi.github.io/Documentation/>



The image displays two screenshots of the BotSentinel Documentation website. The top screenshot shows the homepage with a blue header featuring the text "Your Voice Matters. Don't Let AI Drown It Out" and a subtext about taking control of online spaces. It includes an illustration of a white robot sitting at a laptop. The bottom screenshot shows a "Milestones" page with a sidebar listing four phases: MS1 - Inception Phase, MS2 - Elaboration Phase, MS3 - Construction, and MS4 - Transition. Each phase is represented by a card with a small icon and a link to its details. Below the sidebar, there are three cards: "Trust the Internet Again" (with a shield icon), "Community-Based Ratings" (with a people icon), and "Flexible Blacklist Control" (with a list icon). At the bottom of the page, there is a section titled "Background Objectives Benefits" with a brief description of the project's purpose.