**CAPSTONE PROBLEM STATEMENT –**

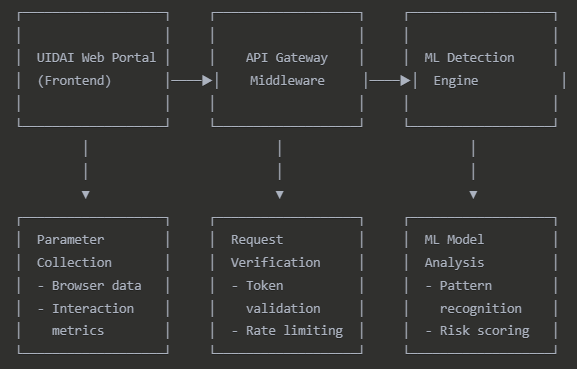
**Passive Bot Detection Solution (Passive CAPTCHA)**

This solution implements a non-intrusive bot detection system for UIDAI's portals that replaces traditional CAPTCHA mechanisms. By analyzing browser fingerprinting data and user behavior patterns through machine learning, the system can distinguish between legitimate human users and automated bots with minimal user interaction, significantly improving accessibility and user experience across UIDAI's digital services.

**OVERALL ARCHITECTURE**

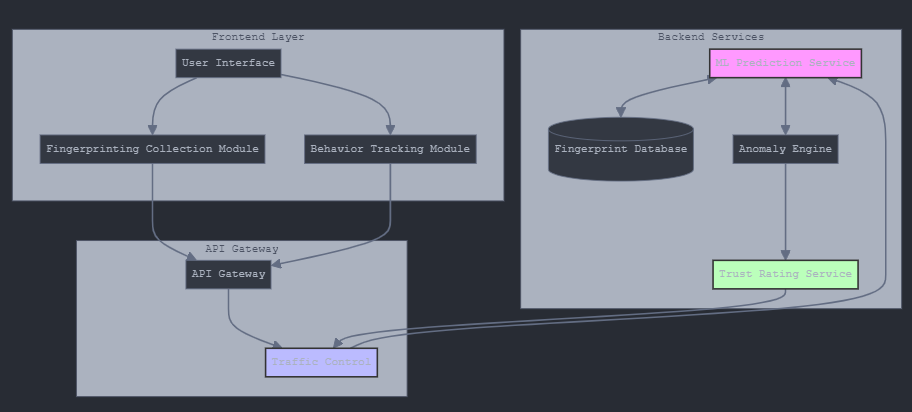
The MVP consists of three main components:

1. Frontend Data Collection Module: A React-based component that passively collects environmental and behavioral data
2. Backend ML Analysis Engine: A machine learning model that processes collected data to determine human vs. bot probability
3. API Protection Layer: Integration mechanism with UIDAI's existing API infrastructure
4. Minimal Interactive Fallback: Optional lightweight interaction prompts for edge cases



**DATA FLOW**

1. User visits UIDAI portal
2. Frontend collector gathers environmental and behavioral data (along with HONEYPOT TECHNIQUE)
3. Data is sent to API Gateway with each API request
4. Gateway forwards data to ML Detection Engine
5. Engine analyzes data and returns risk score
6. If score indicates human (low risk), request proceeds
7. If score indicates possible bot (high risk), system either:
   * Blocks the request
   * Requests minimal additional interaction (GAMIFIED CAPTCHA)
   * Applies rate limiting



**TECHNICAL APPROACH**

**Environmental Parameters Collection**

1. Device characteristics
2. Behavioral patterns
3. Session context

**Dataset**

WEBBOT\_DETECTION **-** <https://m4d.iti.gr/web-bot-detection-dataset/>

This dataset can be used to research web bot detection and evasion techniques that use and/or combine web logs with mouse movements.

The dataset was collected using a web server hosting web pages crawled from and consists of two parts, each used for the different evaluation phases

KEYSTROKE\_DYNAMICS - <https://userinterfaces.aalto.fi/136Mkeystrokes/>

**Machine Learning Model**

The backend will implement a multi-layered ML approach:

1. Feature Processing Layer: Normalizes and transforms raw data
2. Ensemble Classification Model: Combines multiple algorithms (Random Forest, Gradient Boosting, Neural Networks)
3. Confidence Threshold System: Determines when additional verification is needed

**Minimal Interactive Verification**

When passive analysis is inconclusive, we'll implement unobtrusive interactive elements:

1. Natural Interaction Tasks: Following an object with the cursor or simple drag interactions
2. Timing-Based Challenges: Measuring response to subtle UI changes
3. Contextual Puzzles: Simple tasks naturally integrated into the interface

**Advantages Over Traditional CAPTCHA**

This solution provides significant improvements over traditional CAPTCHA:

* Nearly invisible to legitimate users
* Better accessibility for all users
* More robust security against advanced bots
* Complete control over data and processing
* Dynamic adaptation to new threat patterns