SMART INDIA HACKATHON 2024

Problem Statement ID – 1672

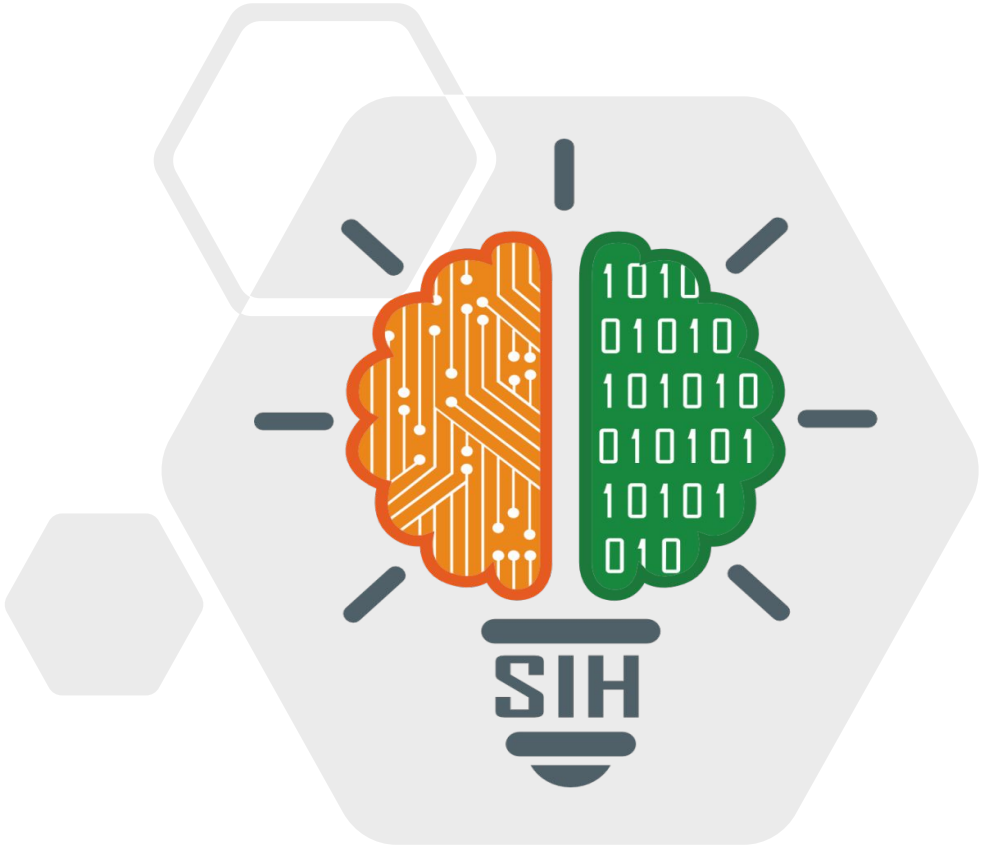
Problem Statement Title - Develop an ML Model based solution to refine CAPTCHA.

Theme - Smart Automation

PS Category - Software

Team ID -

Team Name - Mini Doras



Mini Doras

DEVELOPING AN ML-BASED SOLUTION TO REFINE CAPTCHA

SMART INDIA HACKATHON 2024

Our innovative solution takes the form of a ***three-layered approach***, the idea being that a bot may outsmart a single layer or two, but not all three.

1. **Layer 1 (ML Judge):** This is the core strategy of reducing human effort. Here we use an MoE Classification model to identify the user as a bot or human. Most of our users will be classified as human or bot in this layer. In the event of a low-confidence result, the user will be sent to Layer 2.

2. **Layer 2 (Adversarial Attack Captcha & SusMeter):** This is our counter to computer vision based bots. An adversarial attack captcha is specially fine-tuned to confuse and defeat image recognition algorithms. This ensures any users that pass this layer are almost certainly humans. As added security the SusMeter, another state-of-the-art MoE Classification model, is used to track user behaviour and further reduce chances of bots sneaking past.

3. **Layer 3 (Interactive Challenge & SusMeter):** This is our final counter against any bot that manages to bypass Layer 2 by random chance. The user is made to solve a very simple interactive task picked at random. This means bots must be able to solve an (ever-changing) library of possible challenges. On top of this lies yet another MoE Classification model which judges the user's behaviour.

Mini Doras

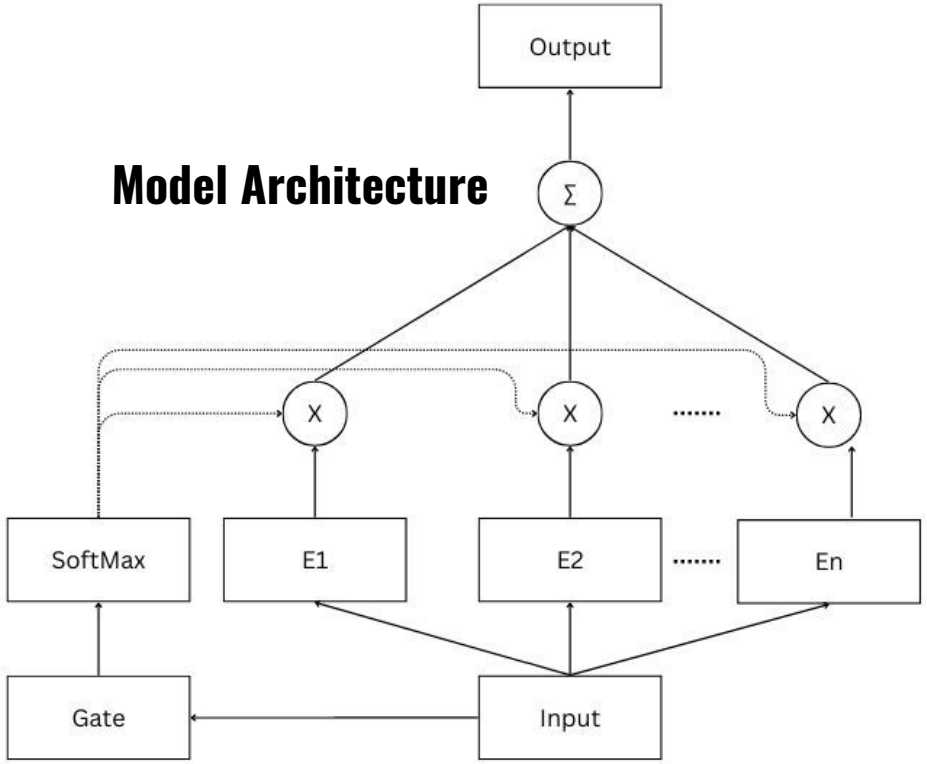
TECHNICAL APPROACH

SMART INDIA HACKATHON 2024

Technologies used:

1. Frontend - React
2. Security - Fernet Cipher
3. ML Model - MOE of CATBOOST, XGBOOST and ANNs
4. Used adversarial attack on bot for images in round 2

Model Architecture



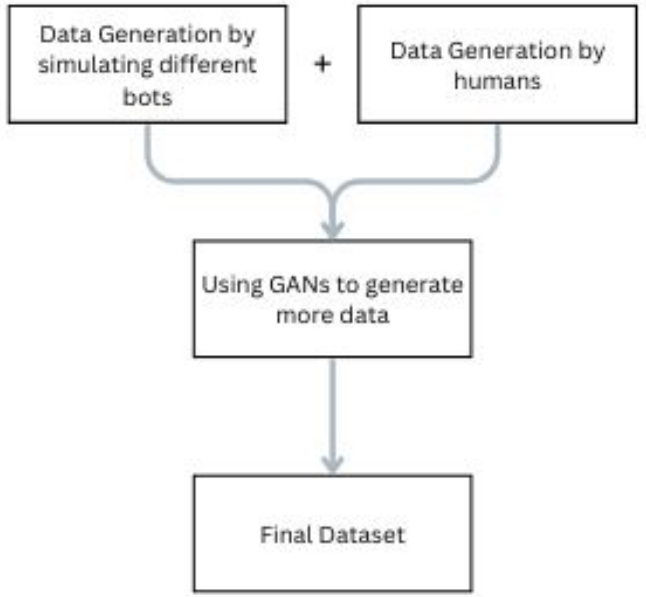
Dataset for Training ML Judge - L1

Time Taken	Typing	Mouse Movement
user.agent	Mouse Distance	IP(City)
IP(Country)	Coordinate	IS BOT

Dataset for Training Sus Meter - L2

Time Taken	Typing	Mouse Movement	
user.agent	Mouse Distance	IP(City)	
IP(Country)	Coordinate	IS Solved	IS BOT

ML Model and Data Generation



Mini Doras

FEASIBILITY AND VIABILITY

SMART INDIA HACKATHON 2024

FEASIBILITY AND VIABILITY

Highly Scalable

- Small model - can be easily scaled up or adapted for more complex scenarios as needed.
- This flexibility allows for incremental improvements and adjustments.

High f1 Score

- Enhanced Decision-Making and reduced errors
- Effective Handling of Imbalanced Data
- Improved Reliability

Neural Network Gate Model (MoE) Accuracy: 0.963855421686747

	precision	recall	f1-score	support
human	1.00	0.90	0.95	31
bot	0.95	1.00	0.97	52

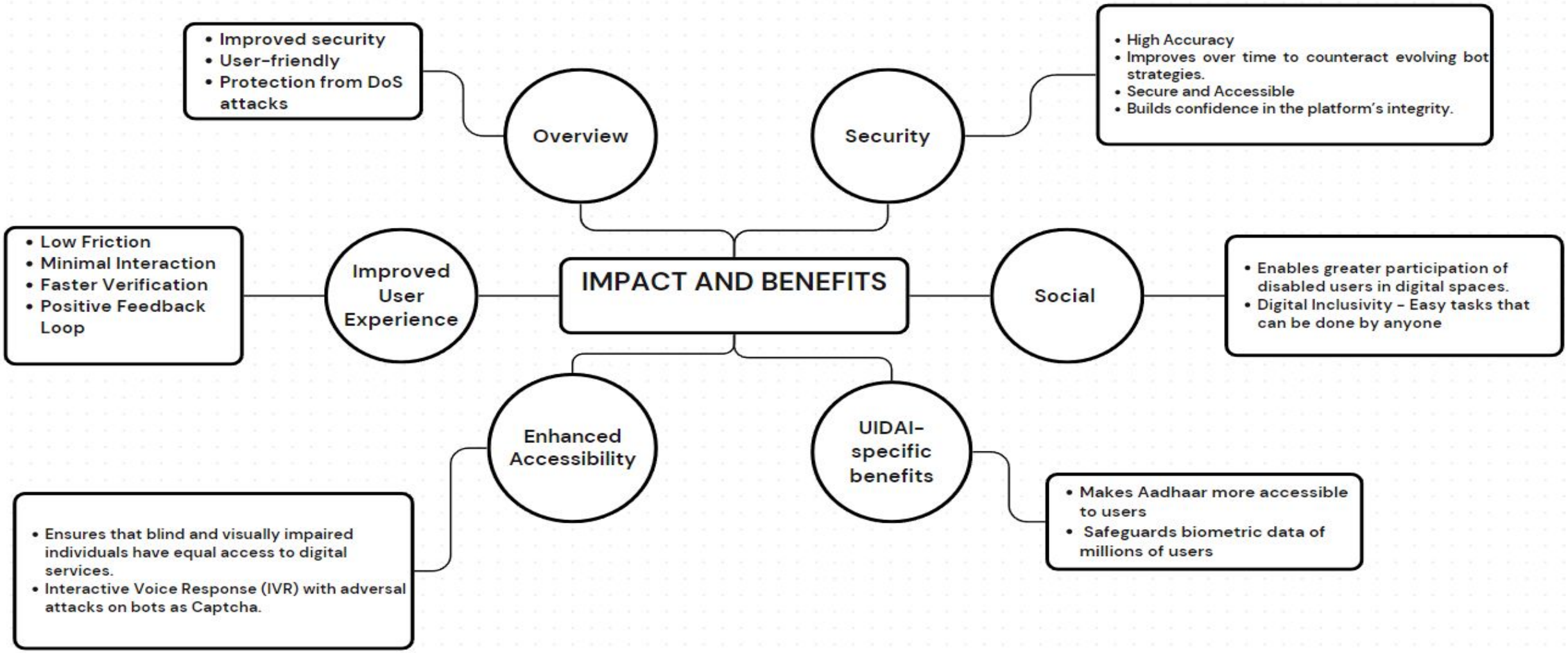
Easy to Deploy

- Modularity: Models designed in modular components hence can be deployed incrementally and integrated with existing systems more smoothly.
- Automated Deployment
- Versatile Deployment Options: Models that are compatible with various platforms (cloud, on-premises, hybrid)

Mini Doras

IMPACT AND BENEFITS

SMART INDIA HACKATHON 2024



Mini Doras

RESEARCH AND REFERENCES

SMART INDIA HACKATHON 2024

1. Detection of Bots in CAPTCHA as a Cloud Service Utilizing Machine Learning
2. Precursory Analysis of Attack-Log Time Series by Machine Learning for Detecting Bots in CAPTCHA
3. No Bot Expects the DeepCAPTCHA! Introducing Immutable Adversarial Examples, With Applications to CAPTCHA Generation
4. Recent advances of Captcha security analysis: a short literature review
5. A survey of CAPTCHA technologies to distinguish between human and computer
6. Precursory Analysis of Attack-Log Time Series by Machine Learning for Detecting Bots in CAPTCHA