**Audio Based Captchas to Prevent Various Web Attacks**

**A PROJECT REPORT**

###### ***Submitted by***

**AKSHAT RAWAT (20BCY10036)**

**KUNAL DHINGRA (20BCY10018)**

**SAKSHAM GUPTA (20BCY10088)**

**SHASHWAT CHANDRA (20BCY10023)**

*In partial fulfillment for the award of the degree*

*Of*

##### BACHELOR OF TECHNOLOGY

##### *In*

##### COMPUTER SCIENCE AND ENGINEERING

##### *WITH SPECIALIZATION IN*

##### CYBERSECURITY AND DIGITAL FORENSICS

****

**SCHOOL OF COMPUTING SCIENCE AND ENGINEERING**

**VIT BHOPAL UNIVERSITY**

**KOTHRIKALAN, SEHORE**

**MADHYA PRADESH - 466114**

DEC 2021

**BONAFIDE CERTIFICATE**

Certified that this project report titled **“Audio Based Captchas To Prevent Various Web Attacks”** is the bonafide work of “**AKSHAT RAWAT (20BCY10036), KUNAL DHINGRA (20BCY10018), SAKSHAM GUPTA(20BCY10088), SHASHWAT CHANDRA(20BCY10023)”** who carried out the project work under my supervision. Certified further that to the best of my knowledge the work reported at this time does not form part of any other project/research work based on which a degree or award was conferred on an earlier occasion on this or any other candidate.

|  |  |
| --- | --- |
| **PROGRAM CHAIR**  Dr. R. Rakesh  Senior Assistant professor  School of Computing Science and Engineering  VIT BHOPAL UNIVERSITY | **PROJECT GUIDE**  D:\Phone Camera\20201103_120845.jpg  Dr. S. Hariharasitaraman  Senior Assistant professor  School of Computing Science and Engineering  VIT BHOPAL UNIVERSITY |

The Project Exhibition I Examination is held on \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**ACKNOWLEDGEMENT**

First, we would like to thank the Lord Almighty for his presence and immense blessings throughout the project work.

We wish to express our heartfelt gratitude to Dr R.Rakesh, Program Chair, Division of Cyber Security and School of Computing Science for much of his valuable support and encouragement in carrying out this work.

We would like to thank our project guide Dr Hariharasitaraman.S for continually guiding and actively participating in our project, giving valuable suggestions to complete the project work.

We would like to thank all the technical and teaching staff of the School of Computer Science, who extended directly or indirectly all support.

Last, but not least, we are deeply indebted to our parents who have been the greatest support while we worked day and night for the project to make it a success.

.

**LIST OF ABBREVIATIONS**

1. **RAM -** Random Access Memory
2. **ROM -** Read Only Memory
3. **CAPTCHA-** Completely Automated Public Turing To Tell Computers and Humans Apart
4. **CSS -** Cascading Style Sheets
5. **JS -** Java Script
6. **OCR** – Optical Character Recognition
7. **AI** - Artificial Intelligence
8. **ML** – Machine Learning
9. **Gb –** GigaByte

**LIST OF FIGURES AND GRAPHS**

|  |  |  |
| --- | --- | --- |
| **FIGURE NO.** | **TITLE** | **PAGE NO.** |
| **1** | Text-Audio based captcha | **13** |
| **2** | Architectural flowchart | **18** |
| **3** | Working flowchart | **19** |
| **4**  **5** | Login Page  Audio Captcha | **24** |
| **6** | About | **25** |

**ABSTRACT**

The audio-based captcha is a cyber-defence tool to prevent web servers and Websites from flooding the network from spammers, bots, and anti-social elements of society. It is a highly optimized model which requires absolutely no system requirements from the user end and works efficiently without taking many resources. The audio-based captcha project is built using programming languages HTML, CSS, and JavaScript. The prototype requires the user to load audio, identify it and input the answer to enter the website or move forward. It can be used even by the small businesses and startups which are economically not so strong instead of paying thousands of dollars from their pocket to firms providing firewall services.

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **CHAPTER NO.** | **TITLE** | **PAGE NO.** |
|  | List of Abbreviations  List of Figures and Graphs  Abstract | 4  5  6 |
| 1 | **CHAPTER-1:**  **PROJECT DESCRIPTION AND OUTLINE** Introduction 1.2 Motivation for the work  1.3 About Introduction to the project  including techniques  1.4 Problem Statement  1.5 Objective of the work  1.6 Organization of the project  1.7 Summary | 11 |
| 2 | **CHAPTER-2:**  **RELATED WORK INVESTIGATION**  2.1 Introduction  2.2 Core area of the project  2.3 Existing Approaches/Methods  2.3.1 Approaches/Methods -1  2.3.2 Approaches/Methods -2  2.3.3 Approaches/Methods -3  2.4 Pros and cons of the stated Approaches/Methods  2.5 Issues/observations from investigation  2.6 Summary | 13 |
| 3 | **CHAPTER-3:**  **REQUIREMENT ARTIFACTS**  3.1 Introduction  3.2 Hardware and Software requirements  3.3 Specific Project requirements  3.3.1 Data requirement  3.3.2 Functions requirement  3.3.3 Performance and security requirement  3.3.4 Look and Feel Requirements  3.4 Summary | 16 |
| 4 | **CHAPTER-4:**  **DESIGN METHODOLOGY AND ITS NOVELTY**  4.1 Methodology and goal  4.2 Functional modules design and analysis  4.3 Software Architectural designs  4.4 Subsystem services  4.5 User Interface designs  4.6 Summary | 17 |
| 5 | **CHAPTER-5:**  **TECHNICAL IMPLEMENTATION & ANALYSIS**  5.1Outline  5.2 Technical coding and code solutions  5.3 Working Layout of Forms  5.4 Prototype submission  5.5 Test and validation  5.6 Performance Analysis(Graphs/Charts)  5.7 Summary | 20 |
| 6 | **CHAPTER-6:**  **PROJECT OUTCOME AND APPLICABILITY**  6.1Outline  6.2 Key implementations outlines of the System  6.3 Significant project outcomes  6.4 Project applicability on Real-world applications  6.5 Inference | 22 |
| 7 | **CHAPTER-7:**  **CONCLUSIONS AND RECOMMENDATION**  7.1Outline  7.2 Limitation/Constraints of the System  7.3 Future Enhancements  7.4 Inference | 26 |
|  | References | 28 |

**CHAPTER 1:**

**1.1 INTRODUCTION**

CAPTCHA stands for the Completely Automated Public Turing test to Differentiate between Computers and People. CAPTCHAs are tools that we can use to differentiate between real users and automated users like bots.

The Audio CAPTCHA is a reverse Turing test that uses audio to determine if a user is human or a machine.

**1.2 MOTIVATION FOR THE WORK**

Having a very keen interest in the subject of Cyber Security and Digital Forensics, our main objective is to research **Audio Based** CAPTCHA s and implement the knowledge gained into a prototype of our novelty that will be focused on people with disabilities which will provide an extra layer of security against bots which are constantly working on the principle of AI and ML.

Ultimately, we will be executing our CAPTCHA on a locally hosted website.

**1.3 About Introduction to the project including techniques**

In the project, short and simple audio, such as the sound of a car or a bus, will be played. The user must provide the sound that was played to distinguish between humans and bots.

**1.4 PROBLEM STATEMENT**

● Including the fact that text-based CAPTCHAs can be broken using optical character recognition (OCR), providing a security risk.

●CAPTCHAs based on text are ineffective for individuals with vision problems.

●Image-based CAPTCHAs necessitate big image databases, which adds to the server's burden.

●The image-based program can be broken by a random guessing assault.

●The video speed in video-based CAPTCHA may be poor, and the procedure may take a long time.

**1.5 OBJECTIVE OF THE WORK**

To protect a website from numerous attacks by utilizing audio-based CAPTCHA, which is more effective than conventional CAPTCHA that relies on text to distinguish between humans and bots.

**1.6 ORGANISATION OF THE PROJECT**

After determining the project, it was divided into 3 modules that are frontend, CAPTCHA, and testing/optimizing. Each module was dealt with in different phases starting with the frontend. The team initially made the frontend for the login page and then for the audio CAPTCHA. following that the frontend for the homepage was created. The main module- CAPTCHA was the trickiest part for the team which included programming in HTML, styling in CSS, and using functions in JS. Finally, the testing was done with a few trial runs.

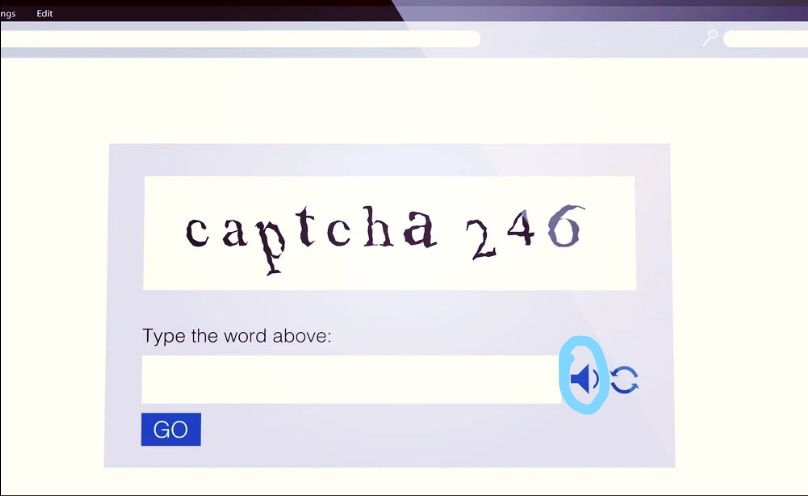
**1.7 SUMMARY**

This work aims to provide an extra layer of security against bots that defeat CAPTCHAs’ purpose.

**CHAPTER 2: RELATED WORK INVESTIGATION**

**2.1 INTRODUCTION**

CAPTCHAs are already in use in the world in different forms like text-based CAPTCHAs, image-based CAPTCHAs, mathematical problems-based CAPTCHAs, and many more. Audio-based CAPTCHAs are already in use that goes along with text-based CAPTCHAs as shown in fig 1.



*Fig 1 Text-Audio based captcha*

**2.2 CORE AREA OF THE PROJECT**

The Project revolves around the implementation of audios and engineering them in such a way that it can be combined with CAPTCHA and prevent bots and AI techniques from flooding the network.

HTML, CSS, and JavaScript have played a crucial role in the development of this project which are the primary programming languages used.

**2.3 EXISTING METHODS/APPROACHES**

Existing solutions include different types of CAPTCHAs like:

* **Math Problems**: The Math CAPTCHA widget requires human interaction, such that a mathematical challenge is required before they can submit the form.
* **Word Problem**: These require several yet easy word problems to be solved by the user to bypass the CAPTCHA.
* **Time-Based**: Time-based CAPTCHAs work on simple logic. It will take for a user to fill out a form, and from this, it can judge whether the user is a human or a bot
* **Puzzles**: These require several puzzles to be solved by the user to bypass the CAPTCHA.
* **Honeypots**: Honeypots, as the name suggests, is a “trap” that is designed to lure bots and computer programs to accidentally reveal their identities. The idea here is to provide something that is going to attract the bot, the “honey”, that is invisible or hidden from legitimate human users. For example, to combat form spam, a common and effective anti-spam honeypot practice is to add a hidden field (via CSS or JavaScript code, or can be as simple as using fonts with the same color as the page’s background) in the form. Human users won’t see this hidden field, but it will be scanned by bots. Thus, we can simply filter out the form submissions that are submitted with this hidden field filled.

**2.4 PROS AND CONS OF EXISTING METHOD**

**Pros**:

* + - Fast and efficient
    - Protects websites and servers from spammers and bots
    - Prevents web attacks like DoS attacks

**Cons**:

* Can be bypassed by AI and ML technique-based bots.

**2.5 ISSUES AND OBSERVATIONS FROM THE EXISTING METHOD**

In [2014](https://security.googleblog.com/2014/04/street-view-and-recaptcha-technology.html), Google pitted one of its machine learning algorithms against humans in solving the most distorted text CAPTCHAs: the computer got the test right 99.8 percent of the time, while the human’s percent. From this observation, it is clear that text-based CAPTCHAs are vulnerable to bots developed on the principle of AI and ML.

**2.6 SUMMARY**

This prototype was made to overcome the limitations faced by the traditional CAPTCHAs and the ReCAPTCHA by Google. It used audio-based techniques which have to be inputted by the user to verify themselves.

**CHAPTER 3: REQUIREMENT ARTIFACTS**

**3.1 INTRODUCTION**

The user does not require any high-end specification in his device to run our project. The project was made using simple HTML files and Java Scripts (JS) and Cascading Style Sheets (CSS).

**3.2 HARDWARE AND SOFTWARE REQUIREMENTS**

**HARDWARE REQUIREMENTS:**

● 1.6 GHz or faster processor

● 1 Gb of RAM

**SOFTWARE REQUIREMENTS:**

* Visual Studio Code

**3.3 SPECIFIC PROJECT REQUIREMENT**

Stable internet connection

**3.4 SUMMARY**

The project runs on basic requirements like a stable internet connection, a 4 Gb RAM, and requires Visual Studio Code. The model is not a complex project and is accessible to users without any difficulty.

**CHAPTER 4: DESIGN METHODOLOGY AND ITS NOVELTY**

**4.1 METHODOLOGY AND GOAL**

To achieve an errorless and optimized version of the CAPTCHA technique that is more protective and fun to interact with.

**4.2 FUNCTIONAL MODULES DESIGN AND ANALYSIS**

**Modules Used:**

1. Home page of the website

2. CAPTCHA

3. Attack module:

The proposed CAPTCHA is based on audio, therefore, blocking all the brute force attacks.

Since the development of AI and ML computer bots that are good in pattern learning, text CAPTCHAs have become very vulnerable.

Audio-based CAPTCHA provides additional benefits as it helps the visually impaired population bypass this CAPTCHA.

A visually aesthetic user interface made using HTML and CSS

**4.3 SOFTWARE ARCHITECTURAL DESIGNS**

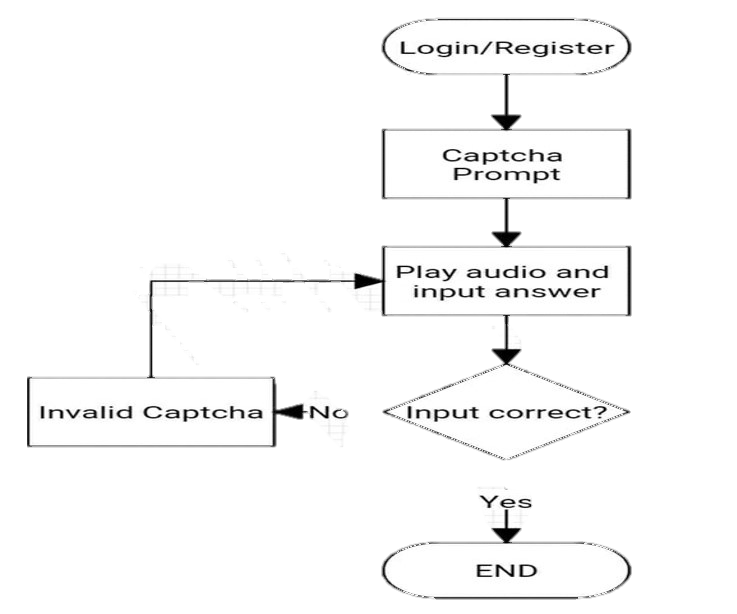
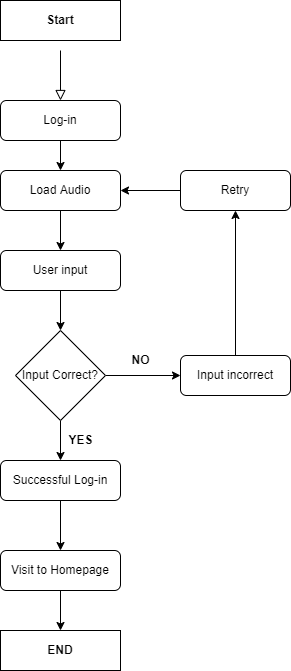
****

Figure 2 Architectural flowchart

**4.4 SUBSYSTEM SERVICES**

****

*Fig 3 Working flowchart*

**4.5 USER INTERFACE DESIGN**

A homepage has been created using HTML and CSS. On the said homepage 2 options have been provided that are log in and Register. The user has to enter his login credentials to visit the site.

**4.6 SUMMARY**

The entire project is based on the simple command of HTML, CSS, and JavaScript which makes it open-source and user friendly

**CHAPTER-5:**

**TECHNICAL IMPLEMENTATION & ANALYSIS**

**5.1 Technical coding and code solutions**

<body>

    <div class="bigbox">

        <div class="form-box">

        <div class="title">

        <div class="form-box">

        <div class="item2">

            <b>Project Exhibition-I|Team-20 </b>

        </div>

        </div>

        </div>

        <div class="box">

            <div class="item1">

            <div class="form-box">

            <form   class="input-group">

                <input type="text" id="ans" class="input-field" placeholder="The required one-word answer is" required onkeyup="this.value = this.value.toLowerCase()">

            </form>

             <div class="heading">

                <p>&#9734 AUDIO BASED CAPTCHA &#9734</p>

            </div>

            <div class="text">

                <p> &#9658 An audio will played below.

                    Please listen to the audio carefully and mention what sound is being played in the box

                    below.Press the Load button to load an audio clip

                </p>

            </div>

            <div class="p">

                <h1>Result</h1>

                <button>Load</button>

            </div>

            <div class="audio">

                <audio id="audio" controls >

                <source id="mp3player"  >

                </audio>

            </div>

            </div>

            <div class="submit-btn">

            <button type="submit" onclick='captcha()' id='submitbtn'  class="submit-btn">Login</button>

            </div>

            </div>

            </div>

            </div>

            </div>

            <script src="script.js"></script>

        </div>

**5.2 Summary**

This chapter consists of segments of codes of the prototype

**CHAPTER-6:**

**PROJECT OUTCOME AND APPLICABILITY**

**6.1 Outline**

**6.2 Key implementations outlines of the System**

**let btnRandom = document.querySelector('button');**

**let result = document.querySelector('h1');**

**let users = ['train', 'siren','horn','guitar','laugh','doorbell','firework'];**

**function getRandomNumber(min, max) { //function to load audios**

**let step1 = max - min + 1;**

**let step2 = Math.random() \* step1;**

**let result = Math.floor(step2) + min;**

**return result;**

**}**

**btnRandom.addEventListener('click', () => {**

**let index = getRandomNumber(0, users.length-1);**

**result.innerText = users[index];**

**if(result.innerText==="train"){**

**document.getElementById("mp3player").src = "E:/project/train.wav";**

**document.getElementById("audio").load();**

**}**

**//……………………..and so on**

**//Main Function to check the input**

**function captcha(){**

**var b = document.getElementById('ans').value;**

**if(b==result.innerText){**

**window.location.replace('file:///E:/project/homepage.html')**

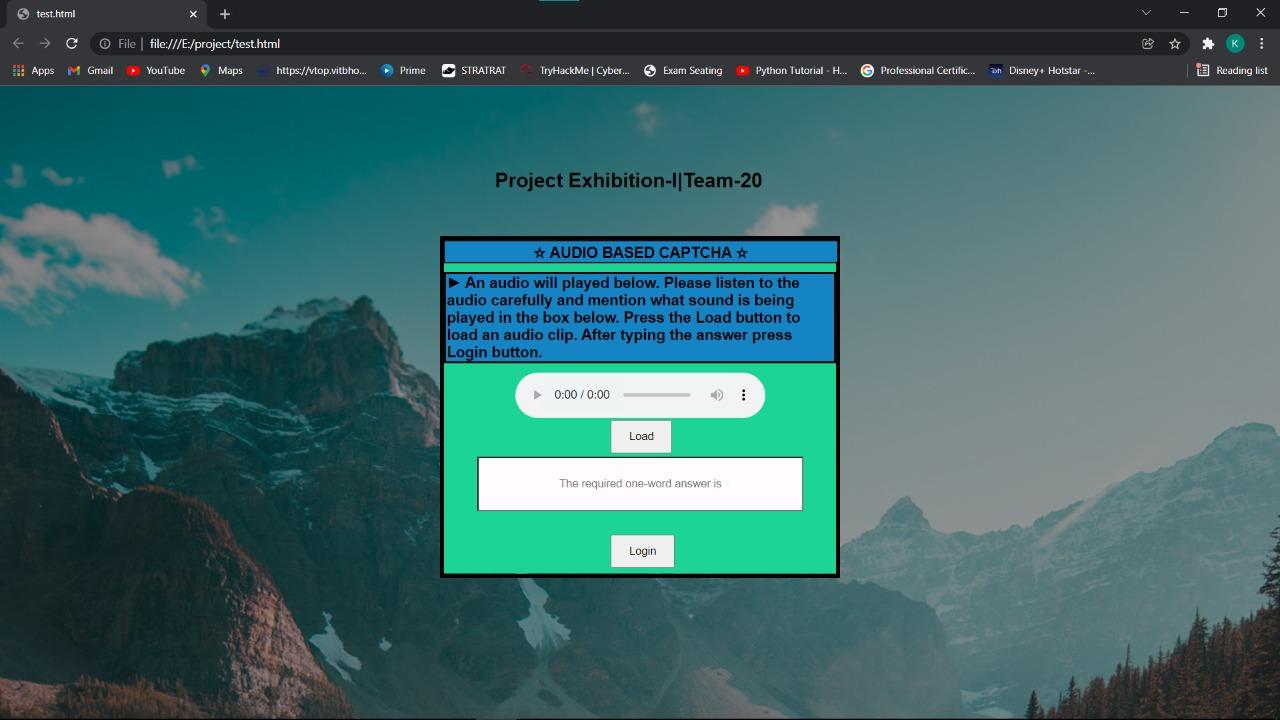
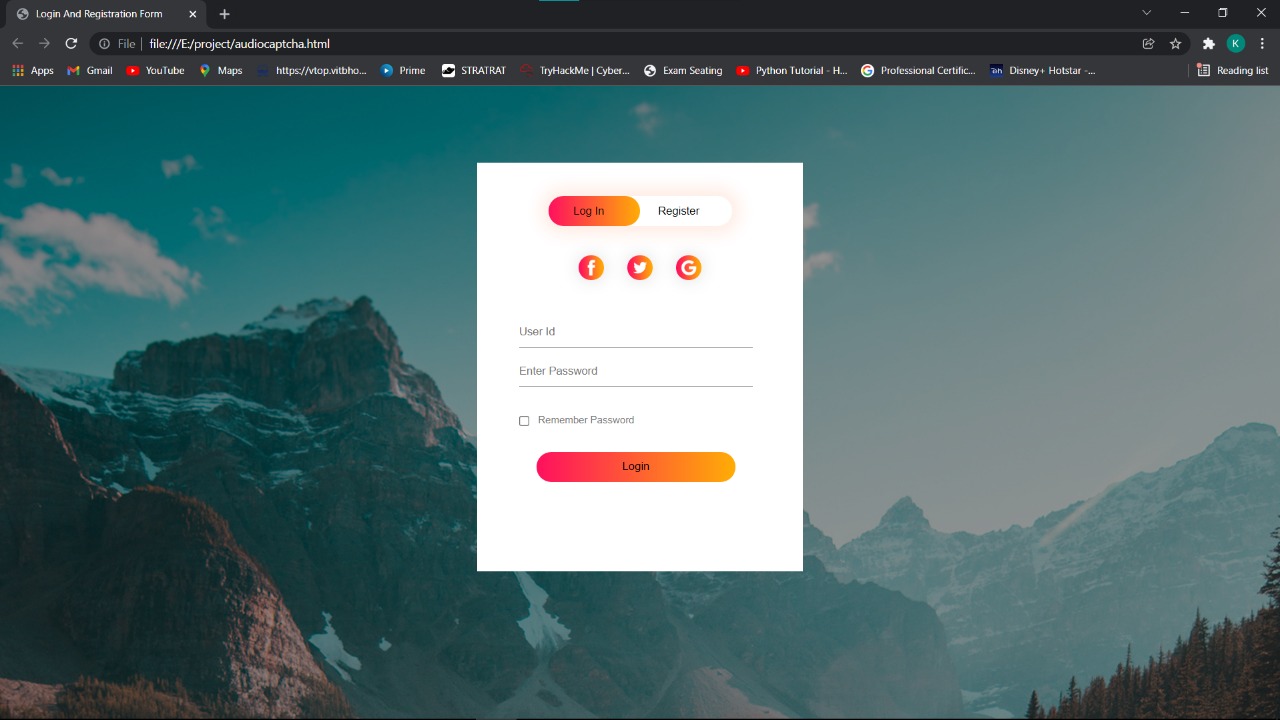
**}**

**else{**

**window.alert("Your input is incorrect");**

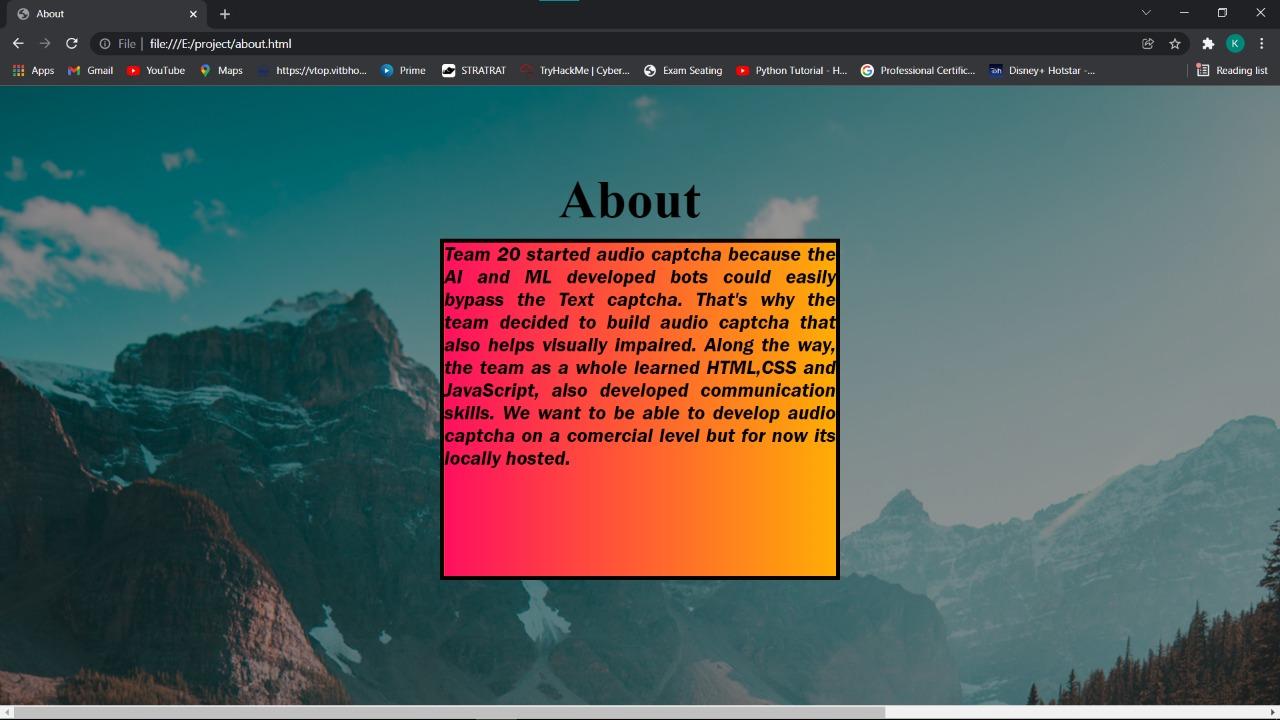
**}**

**};**

**6.3 Significant project outcomes**

**Fig 5 Audio Captcha**

**Fig 4 Login Page**

****

**Fig 6 About**

**6.4 Project applicability on Real-world applications**

The audio-based CAPTCHA could be implemented on every website available on the World Wide Web to ensure safety from spammers and bots from flooding the network and even from various web attacks.

Due to its low requirement, it can work efficiently even on a potato pc and continue to protect servers from web attacks which could prove to be dangerous.

**CHAPTER-7:**

**CONCLUSIONS AND RECOMMENDATION**

**7.1 Outline**

Discovery Phase:

* Primary Research
  + It was found that a text-based CAPTCHA was in use.
* Existing Content and Design Analysis
  + Text-based CAPTCHAs are in use and are beaten by bots developed on the concept of AI/ML.
* Gather Data
  + Other forms of CAPTCHAs such as puzzles, video-based, mathematical problems, etc. are also in use.

Strategy Phase

* Determine Requirements
  + Software Requirements: Visual Studio Code
  + Audio Library
  + Stable Internet Connectivity

Design and Development Phase

* Content Development
  + Formulation of a Login page using HTML and CSS.
  + Development of the CAPTCHA using 2 main functions i.e. function getRandomNumber() and function captcha().

**7.2 Limitation/Constraints of the System**

Nothing is perfect in this world and hence even this model comes with a couple of limitations. The main being is the problem faced by people suffering from hearing impairments.

**7.3 Future Enhancements**

Since the model is still in its trial phase, a future version could include; accepting synonyms of the word to be inputted by the user to increase functionality. It would save hundreds of dollars of users which are invested by them into different kinds of firewalls and is more efficient than the ReCAPTCHA by google

**7.4 Inference**

The user will have to log in first using the correct credentials. After successful login, the user would be taken to a new window containing an audio-based CAPTCHA. The user has to load the audio and play the audio to know the correct answer. After doing so, the user has to write the correct one-word answer to bypass the CAPTCHA. After successfully doing so, the user would be taken to the home page.

**REFERENCES**

1. Firkhan Ali bin Hamid , Dept. Web Technology, FSKTM University Tun Hussein Onn Malaysia Johor, Malaysia. *Development of CAPTCHA system based on puzzle*
2. <https://link.springer.com/chapter/10.1007/978-981-10-8228-3_33#:~:text=Audio%20CAPTCHA%20is%20a%20reverse,of%20CAPTCHA%20for%20web%20security>
3. https://www.w3schools.com/
4. irkhan Ali Bin Hamid Ali
5. Dept. Web Technology, FSKTM
6. Universiti Tun Hussein Onn Malaysia
7. Johor, Malaysia
8. irkhan Ali Bin Hamid Ali
9. Dept. Web Technology, FSKTM
10. Universiti Tun Hussein Onn Malaysia
11. Johor, Malaysia
12. irkhan Ali Bin Hamid Ali
13. Dept. Web Technology, FSKTM
14. Universiti Tun Hussein Onn Malaysia
15. Johor, Malaysia
16. irkhan Ali Bin Hamid Ali
17. Dept. Web Technology, FSKTM
18. Universiti Tun Hussein Onn Malaysia
19. Johor, Malaysia
20. irkhan Ali Bin Hamid Ali
21. Dept. Web Technology, FSKTM
22. Universiti Tun Hussein Onn Malaysia
23. Johor, Malaysia
24. irkhan Ali Bin Hamid Ali
25. Dept. Web Technology, FSKTM