PROJECT REPORT



Project Name: Image Based Captcha Using Python

Course Code: INT213

Course Name : Python Programming

Submitted To : Mr. Dipen Saini

Submitted By : Y.Yagnesh (12107821) and Amit Kumar(11912201)

GitHub link: <https://github.com/amitsde1/ImagBasedCaptcha>

## What is CAPTCHA

CAPTCHA stands for the Completely Automated Public Turing test to tell Computers and Humans Apart. CAPTCHAs are tools you can use to differentiate between real users and automated users, such as [bots](https://www.imperva.com/learn/application-security/what-are-bots/). CAPTCHAs provide challenges that are difficult for computers to perform but relatively easy for humans. For example, identifying stretched letters or numbers, or clicking in a specific area.

**What are CAPTCHAs Used for**

CAPTCHAs are used by any website that wishes to restrict usage by bots. Specific uses include:

* **Maintaining poll accuracy**—CAPTCHAs can prevent poll skewing by ensuring that each vote is entered by a human. Although this does not limit the overall number of votes that can be made, it makes the time required for each vote longer, discouraging multiple votes
* **Limiting registration for services**—services can use CAPTCHAs to prevent bots from spamming registration systems to create fake accounts. Restricting account creation prevents waste of a service’s resources and reduces opportunities for fraud.
* **Preventing ticket inflation**—ticketing systems can use CAPTCHA to limit scalpers from purchasing large numbers of tickets for resale. It can also be used to prevent false registrations to free events.
* **Preventing false comments**—CAPTCHAs can prevent bots from spamming message boards, contact forms, or review sites. The extra step required by a CAPTCHA can also play a role in reducing online harassment through inconvenience.

## How Does CAPTCHA Work

CAPTCHAs work by providing information to a user for interpretation. Traditional CAPTCHAs provided distorted or overlapping letters and numbers that a user then has to submit via a form field. The distortion of the letters made it difficult for bots to interpret the text and prevented access until the characters were verified.

This CAPTCHA type relies on a human’s ability to generalize and recognize novel patterns based on variable past experience. In contrast, bots can often only follow set patterns or input randomized characters. This limitation makes it unlikely that bots will correctly guess the right combination.

Since CAPTCHA was introduced, bots that use machine learning have been developed. These bots are better able to identify traditional CAPTCHAs with algorithms trained in pattern recognition. Due to this development, newer CAPTCHA methods are based on more complex tests. For example, reCAPTCHA requires clicking in a specific area and waiting until a timer runs out.

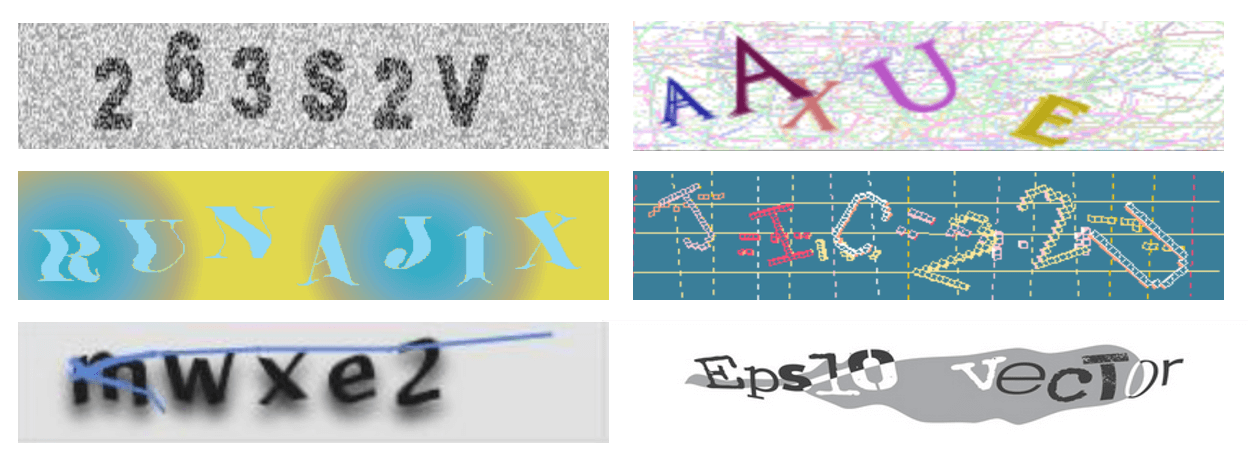
## CAPTCHA Types: Examples

Modern CAPTCHAs fall into three main categories—text-based, image-based, and audio.

### Text-based CAPTCHAs

Text-based CAPTCHAs are the original way in which humans were verified. These CAPTCHAs can use known words or phrases, or random combinations of digits and letters. Some text-based CAPTCHAs also include variations in capitalization.

The CAPTCHA presents these characters in a way that is alienated and requires interpretation. Alienation can involve scaling, rotation, distorting characters. It can also involve overlapping characters with graphic elements such as color, background noise, lines, arcs, or dots. This alienation provides protection against bots with insufficient text recognition algorithms but can also be difficult for humans to interpret.



Text-based CAPTCHA patterns

Techniques for creating text-based CAPTCHAs include:

* **Gimpy**—chooses an arbitrary number of words from an 850-word dictionary and provides those words in a distorted fashion.
* **EZ-Gimpy—**is a variation of Gimpy that uses only one word.
* **Gimpy-r**—selects random letters, then distorts and adds background noise to characters.
* **Simard’s HIP**—selects random letters and numbers, then distorts characters with arcs and colors.

### CAPTCHA Image

Image-based CAPTCHAs were developed to replace text-based ones. These CAPTCHAs use recognizable graphical elements, such as photos of animals, shapes, or scenes. Typically, image-based CAPTCHAs require users to select images matching a theme or to identify images that don’t fit.

You can see an example of this type of CAPTCHA below. Note that it defines the theme using an image instead of text.



Example of image-based CAPTCHA

Image-based CAPTCHAs are typically easier for humans to interpret than text-based. However, these tools present distinct accessibility issues for visually impaired users. For bots, image-based CAPTCHAs are more difficult than text to interpret because these tools require both image recognition and semantic classification.

### Audio CAPTCHA

Audio CAPTCHAs were developed as an alternative that grants accessibility to visually impaired users. These CAPTCHAs are often used in combination with text or image-based CAPTCHAs. Audio CAPTCHAs present an audio recording of a series of letters or numbers which a user then enters.

These CAPTCHAs rely on bots not being able to distinguish relevant characters from background noise. Like text-based CAPTCHAs, these tools can be difficult for humans to interpret as well as for bots.

### Math or Word Problems

Some CAPTCHA mechanisms ask users to solve a simple mathematical problem such as “3+4” or “18-3”. The assumption is that a bot will find it difficult to identify the question and devise a response. Another variant is a word problem, asking the user to type the missing word in a sentence, or complete a sequence of several related terms. These types of problems are accessible to vision impaired users, but at the same time they may be easier for bad bots to solve.

### Social Media Sign In

A popular alternative to CAPTCHA is requiring users to sign in using a social profile such as Facebook, Google or LinkedIn. The user’s details will be automatically filled in using single sign on (SSO) functionality provided by the social media website.

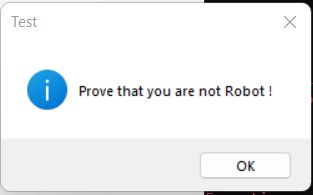
This is still disruptive, but may actually be easier for the user to complete than other forms of CAPTCHA. An additional benefit is that it is a convenient registration mechanism.

### No CAPTCHA ReCAPTCHA

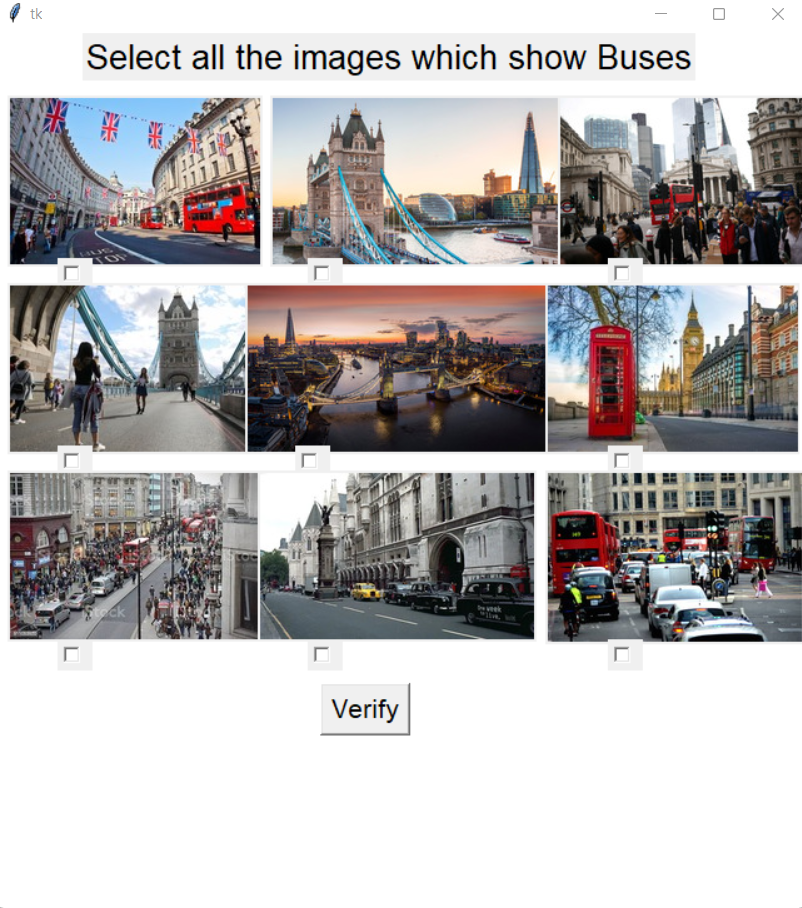
This type of CAPTCHA, known for its use by Google, is much easier for users than most other types. It provides a checkbox saying “I am not a robot” which users need to select – and that’s all. It works by tracking user movements and identifying if the click and other user activity on the page resembles human activity or a bot. If the test fails, reCAPTCHA provides a traditional image selection CAPTCHA, but in most cases the checkbox test suffices to validate the user.

Image Based Captcha

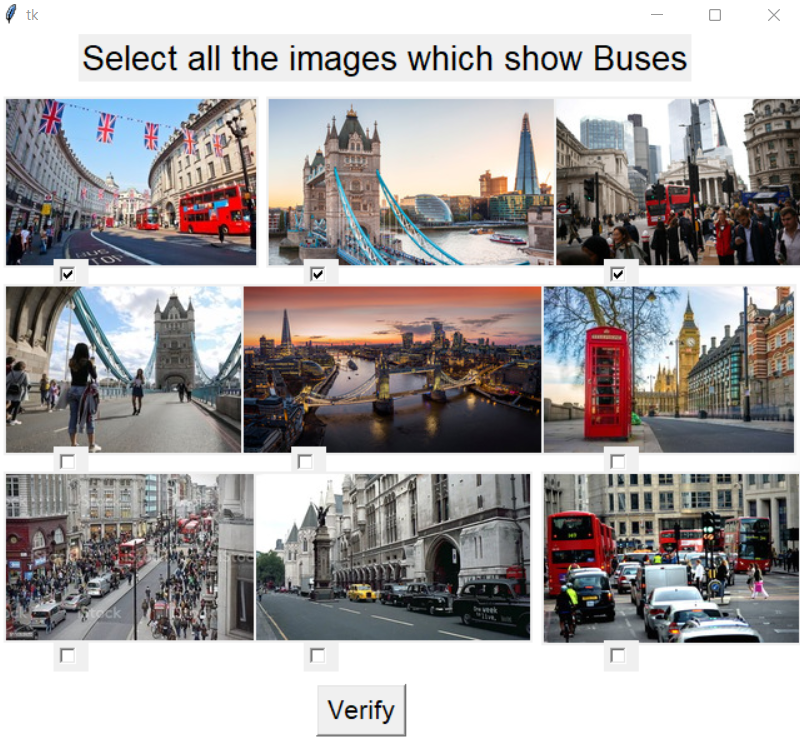
When we run our project , it show message box first to check if We are not robots



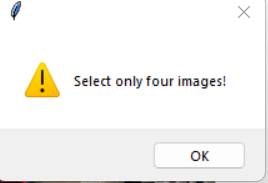
Click on ok it will open the root or frame like shown below



Case 1: selecting less then or more than 4 images then press Valid button.



Here image 1,2 and 3 are selected and when we press vaild button it show message

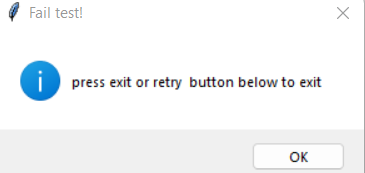


So it shows that there are only four images which are correct

Case 2: if we select four images in which not all are correct then it will show message

Like below image 1,2,3,4 are selected but correct images numbered are 1,3,7 and 8





On clicking on ok button , two buttons will be automatically shown on the frame as below

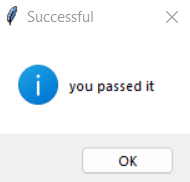


The exit button to exit the window

The retry button will reset all the selected checkbutton to 0

Case 3: all selected four choices are correct then on clicking on verify button a message box

Appears show “you are successful” then on clicking on Ok button root window closes.



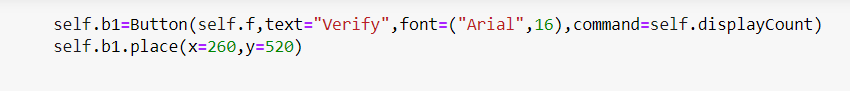
Source Code



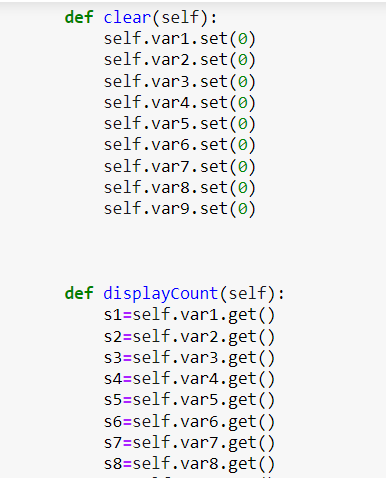


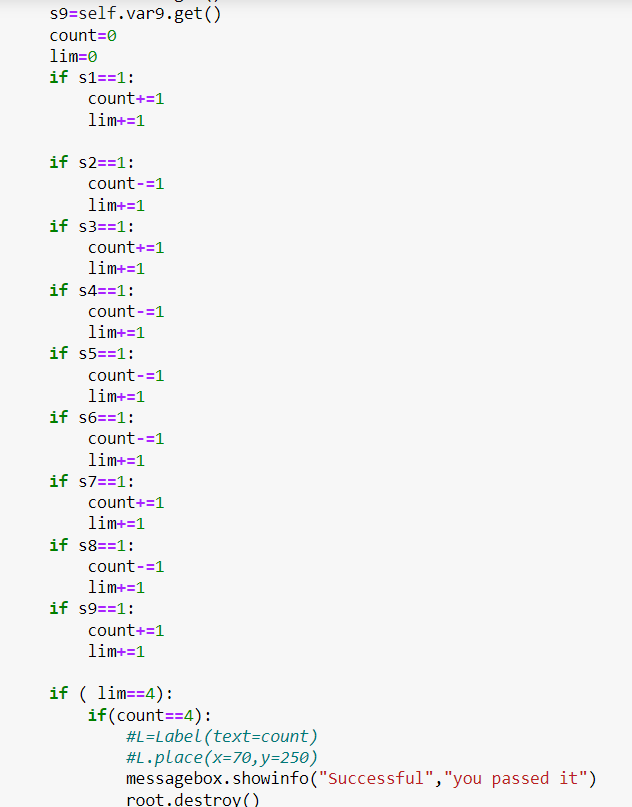














Github Link: <https://github.com/amitsde1/ImagBasedCaptcha>

Technique

checkbutton on selected will provide 1 on getting selected

When correct image get selected , count variable will be incremented by 1 otherwise

Count decrease by 1 apart from count we will introduce lim variable to count how many total checkbutton get selected

If count==4:

If lim==4:

Successful

else:

selected four but not all are correct

else:

selected checkbutton is not exactly 4

**DRAWBACKS OF USING CAPTCHA**

* Imagine this situation. You are knocking on a door. They ask you to prove you are human. They clear as day and see you are a human. Also, you wonder, ‘why would bots come to your house now?!’ This is exactly how people think seeing a CAPTCHA. Statistically, people get annoyed and tend to quit the site too. This would affect your viewership. People oftentimes point that the question, ‘are you a human’, or anything of such similar sort brushes their ego on the wrong side.
* CAPTCHA is one way to prevent breach. It is proven that CAPTCHA tests are successful in preventing breach only to a certain extent. Thus, it is right to state that it would only be partially beneficial when implemented on your site/ app.
* A part of the society finds the CAPTCHA test discriminatory. Statistically, people with poor eyesight or with both ears and sight impairment, find it extremely difficult to access important sites whenever required due to the presence of CAPTCHA.

## ****Conclusion****

To conclude, the concept of CAPTCHA is ever growing and it is going to take a reasonable amount of time to be completely functional and secure. You can expect CAPTCHA growth in the days to come!

Furthermore, by understanding why CAPTCHA is used and if proper implementation of the same is done by you, you can avoid spam bots on your sites!