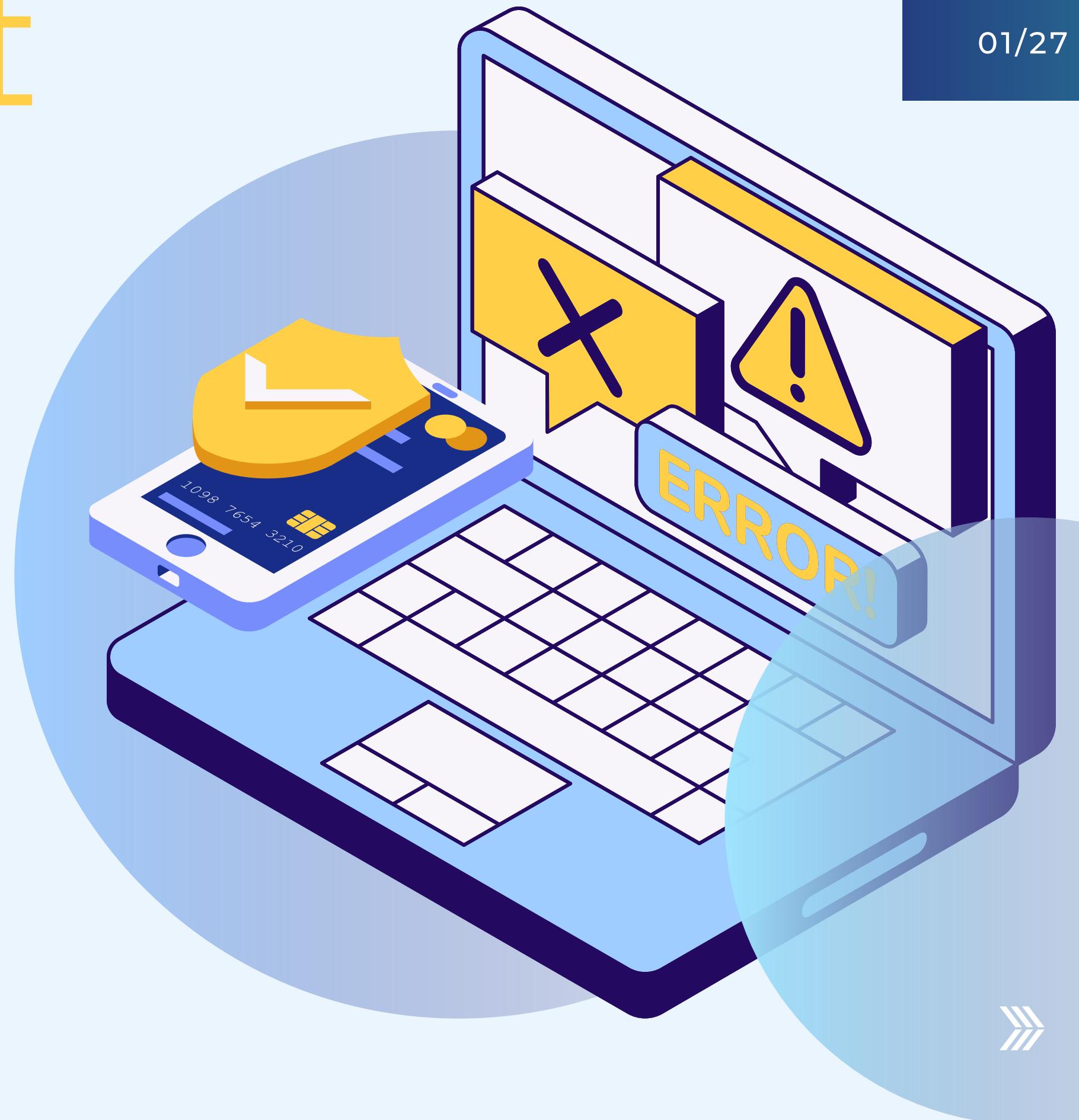


# PRES

# ENTAZIONE

01/27

# CYBER SECURITY



Leonardo Vorabbi  
Carlotta Nunziati

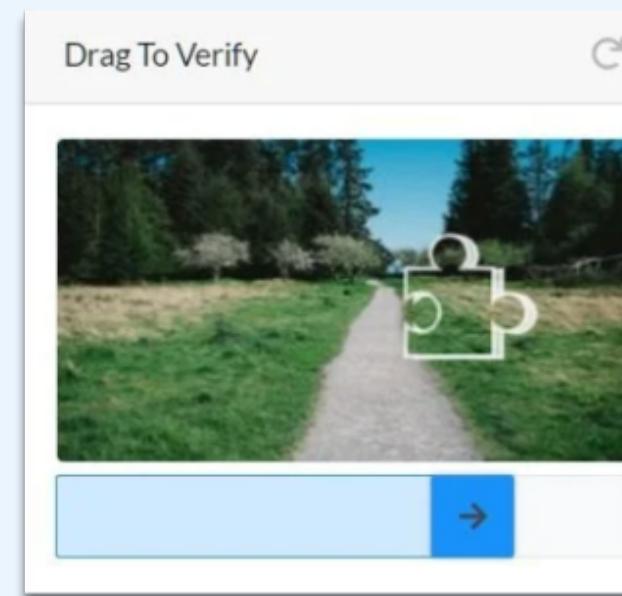


# CAPTCHA COSA SONO?

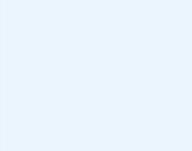
Press PLAY and enter the numbers you hear

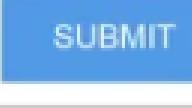
 0:00 

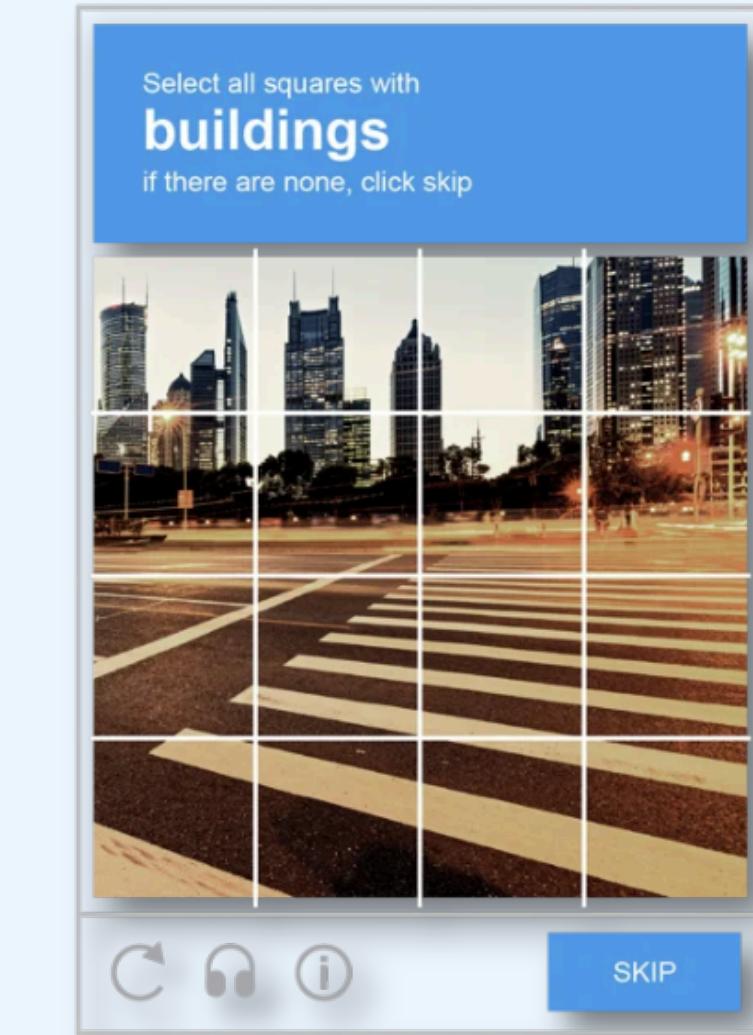
   



Solve the math question and enter the result below



I'm not a robot

  
reCAPTCHA  
Privacy - Terms

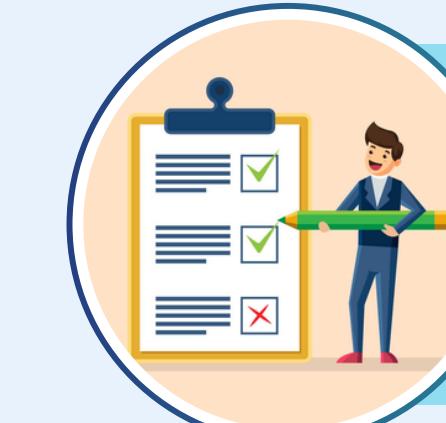
# CAPTCHA A COSA SERVONO?



Prevenire lo spam



Proteggere i siti web da  
abusi



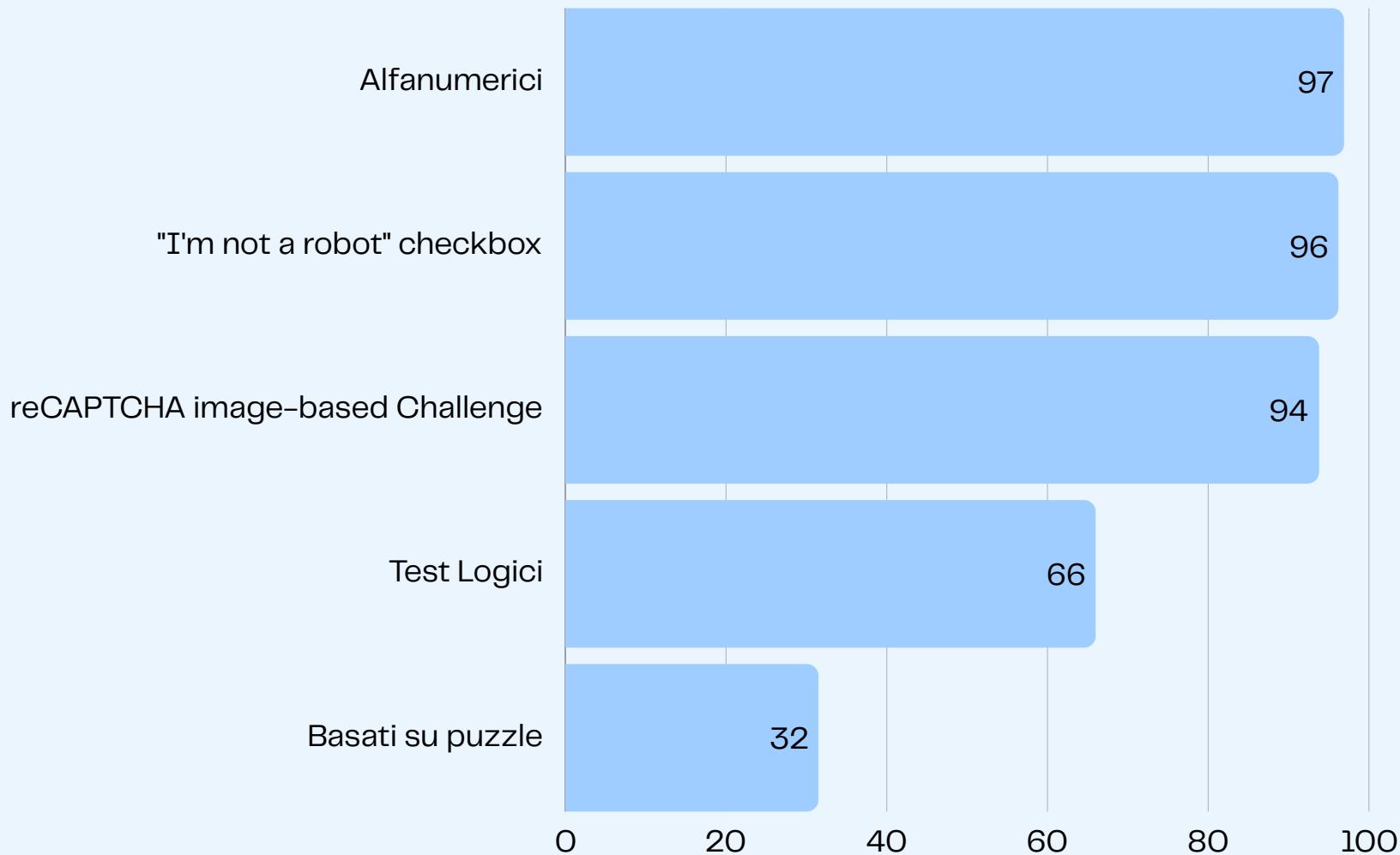
Garantire la correttezza  
dei sondaggi online



Evitare attacchi DDoS e  
scraping

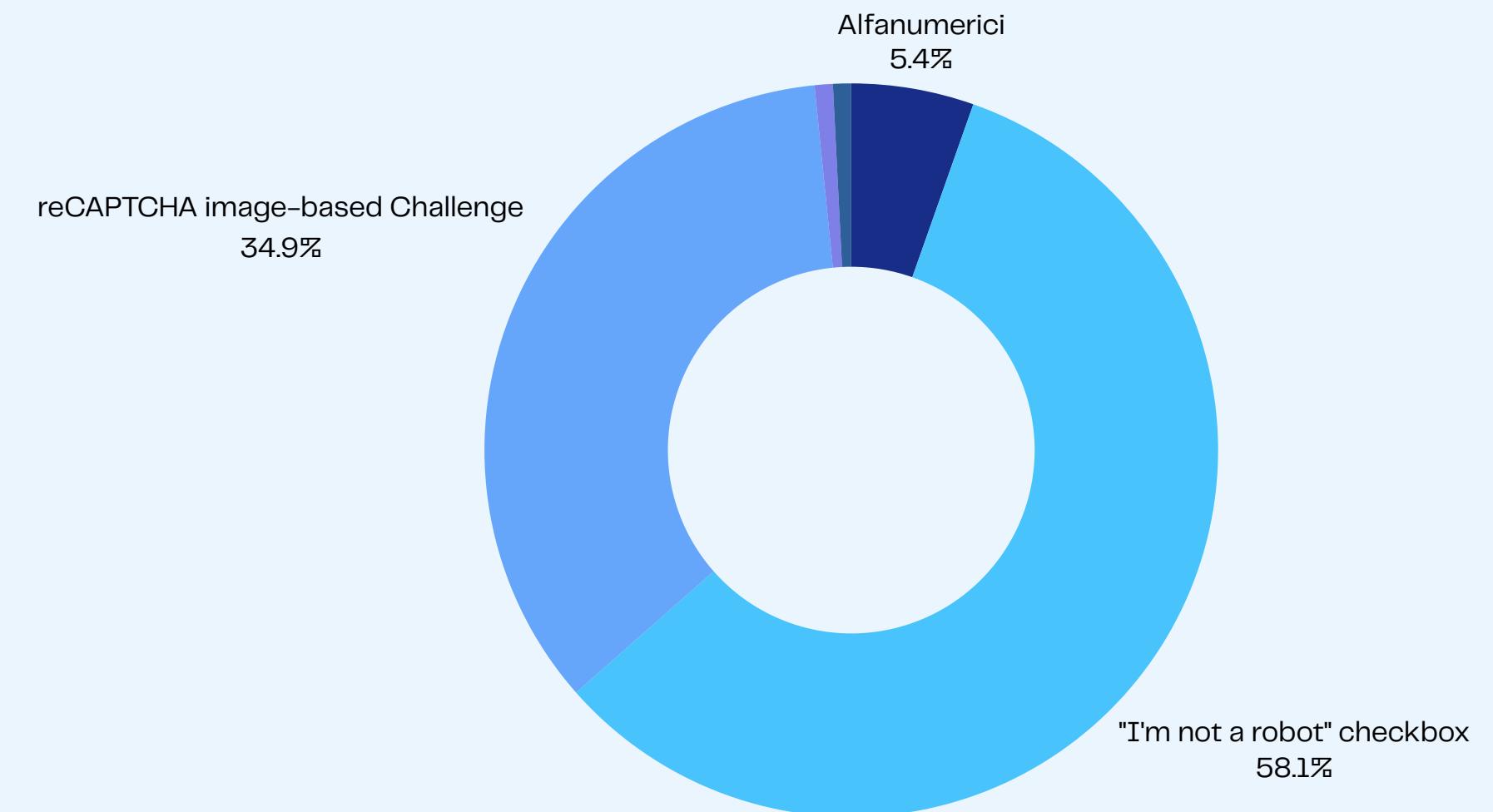
# SONDAGGIO

04/27



Quale di questi CAPTCHA  
conosci?

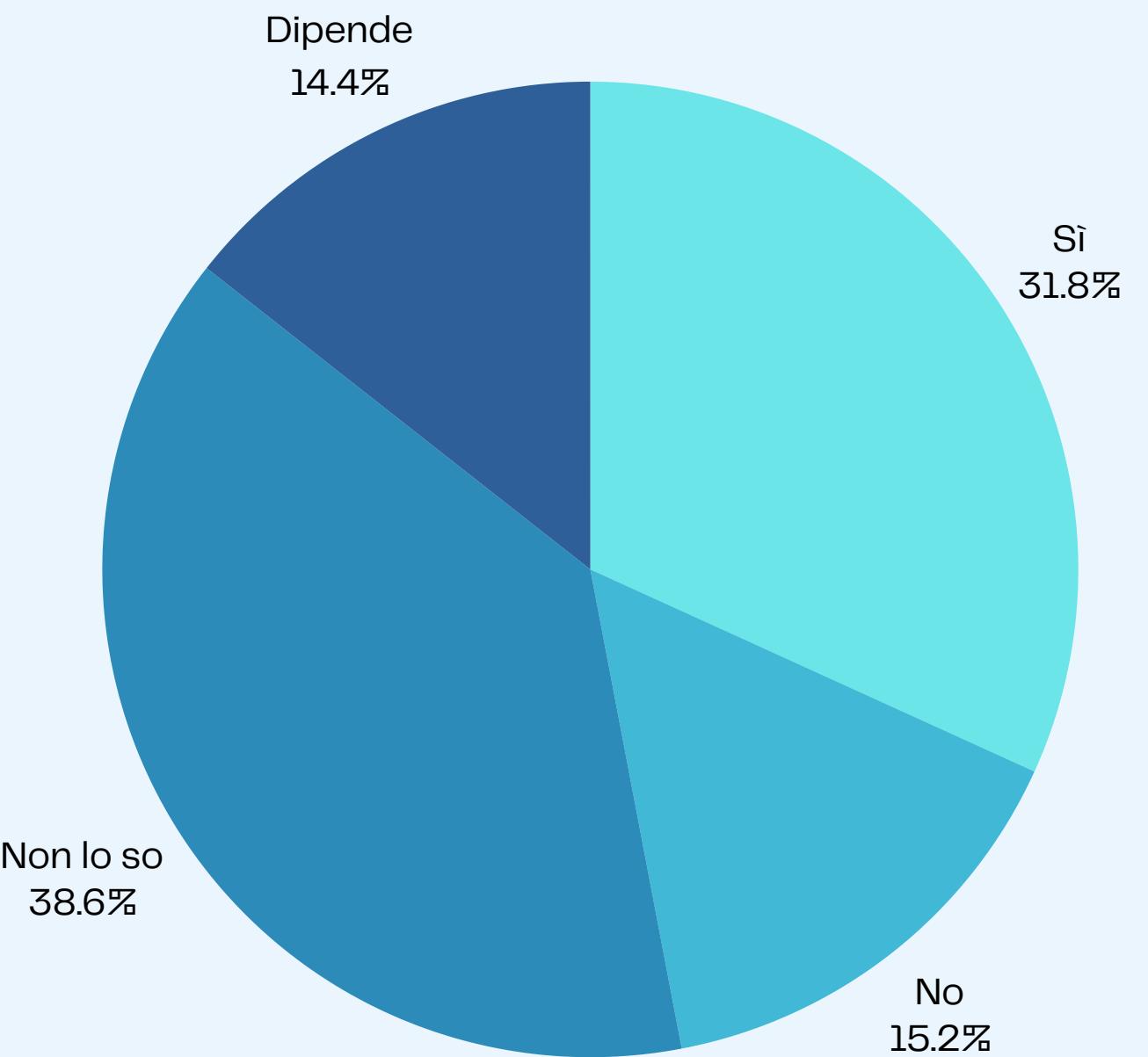
Quale CAPTCHA ti appare  
più frequentemente?



# SONDAGGIO

05/27

Reputi i CAPTCHA affidabili  
per lo scopo con cui sono  
stati concepiti?



# SONDAGGIO

06/27

## DIPENDE: Da cosa? Alcune risposte

**“Dipende dall'implementazione degli stessi. Chi intende bypassarli è sempre un passo avanti, e per poterli mantenere efficienti servono delle risorse che siti indipendenti spesso non possiedono, appoggiandosi quindi a servizi di terzi (come Google)”**

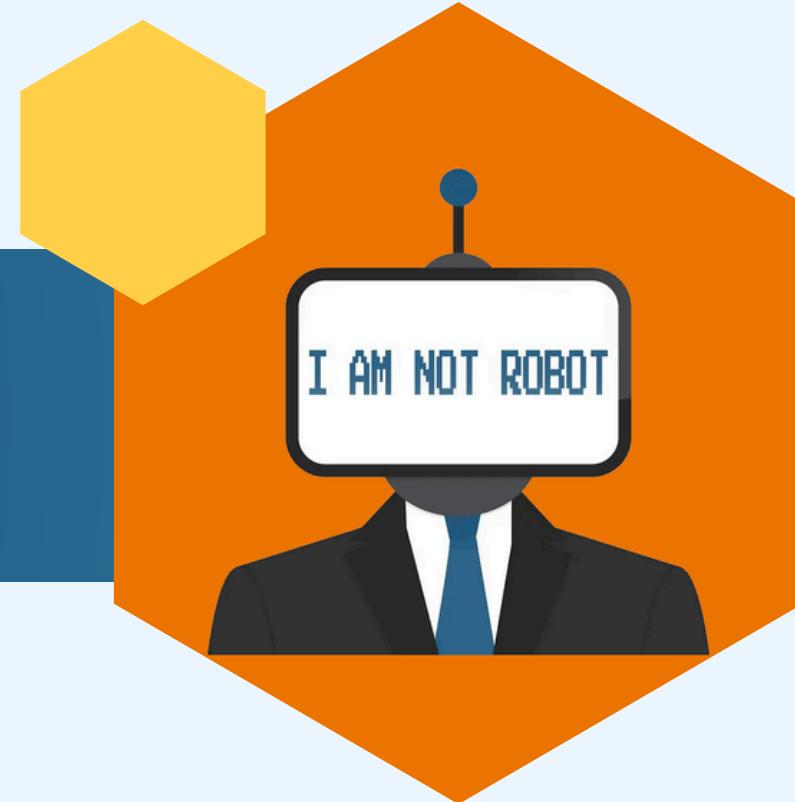
**“Non lo so”**

**“Dalla sicurezza del sito rispetto all'hacking, che può permettere di aggirarli, e dalla tipologia, perché quelli più semplici possono essere risolti da IA”**

**“Boh”**

**“Alcuni, come ad esempio quello in cui bisogna semplicemente selezionare l'opzione "non sono un robot", mi sembrano facilmente aggirabili. Poi in realtà non me ne intendo particolarmente.”**

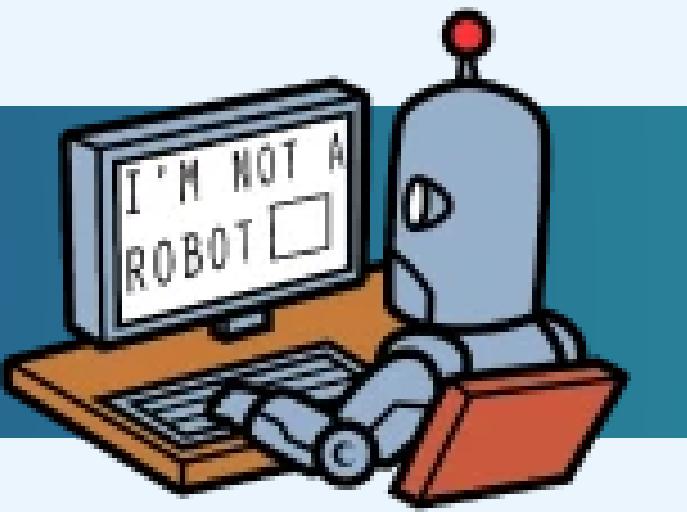
# OBIETTIVO DEL PROGETTO



Analizzare e sfruttare le  
vulnerabilità dei  
CAPTCHA attraverso il  
Machine Learning



# I NOSTRI MODelli



# PROCEDIMENTO

- 1** Scaricamento e preparazione del dataset di immagini
- 2** Divisione del dataset in train, validation e test set
- 3** Costruzione del modello
- 4** Impostazione degli iparametri del modello
- 5** Addestramento del modello
- 6** Valutazione su test set



# INPUT PREPROCESSING

10/27

1

Ridimensionamento

2

Conversione in  
scala di grigi

3

Riduzione del rumore

Filtrari di levigatura

4

Filtrari di contrasto e  
nitidezza

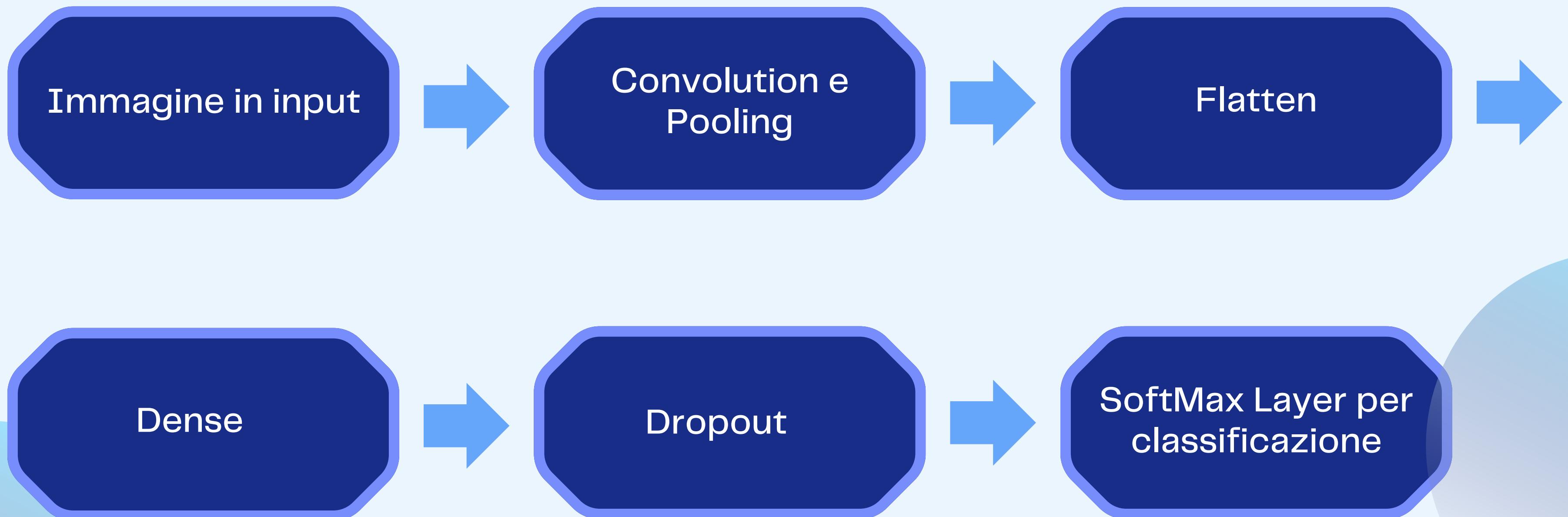
5

Normalizzazione

6

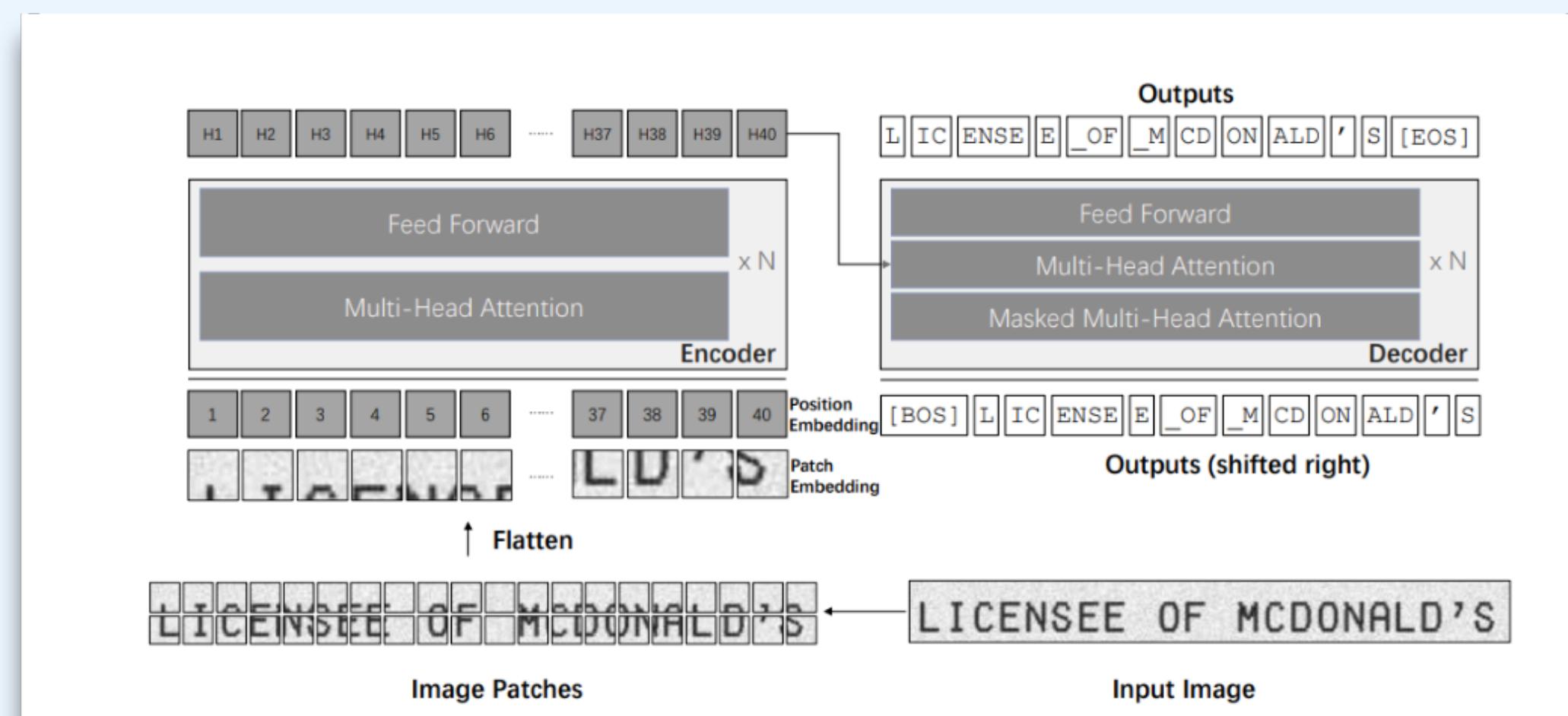
# MODELLO LOGICO

11/27



# MODELLO LOGICO (E ALFANUMERICO) PRE-ADDESTRATO

## TrOCR microsoft

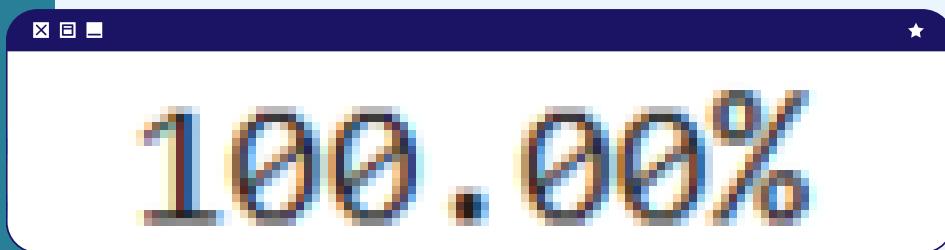


# CONFRONTO ACCURACY DEI MODELLI LOGICI

Accuracy modello non pre-addestrato:

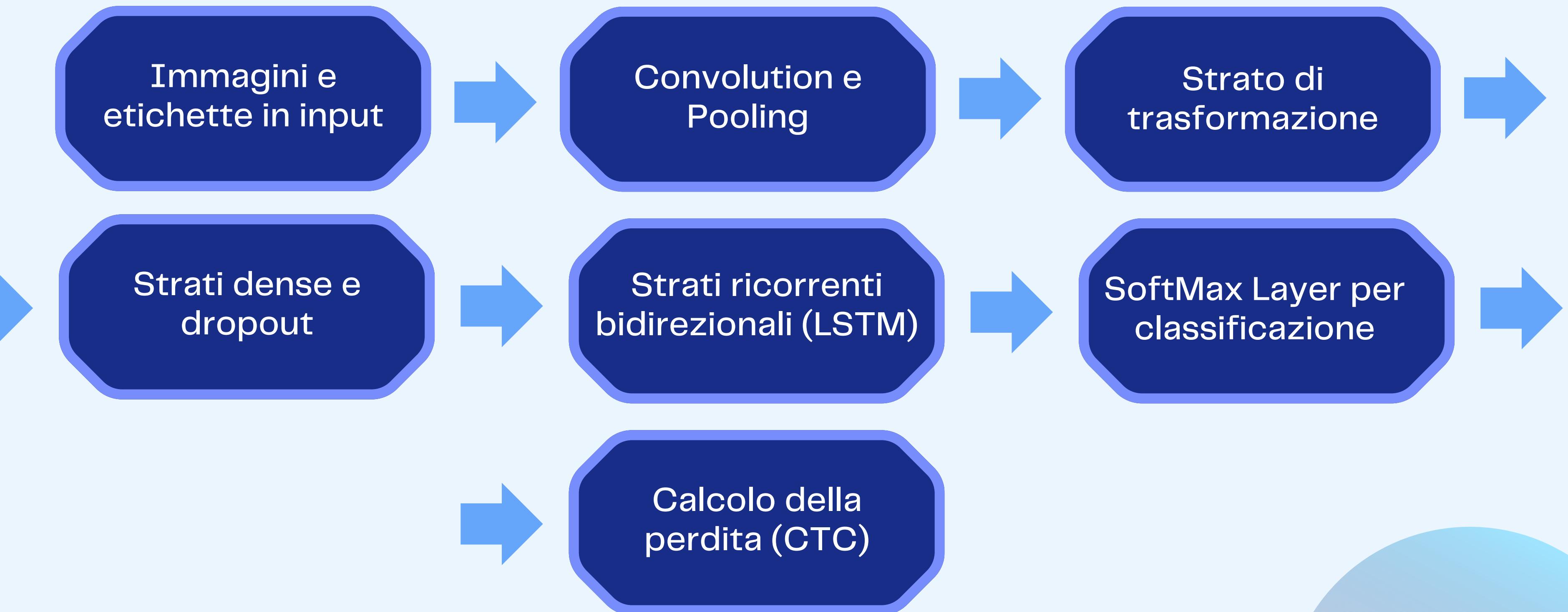


Accuracy modello pre-addestrato:



# MODELLO ALFANUMERICO

14/27

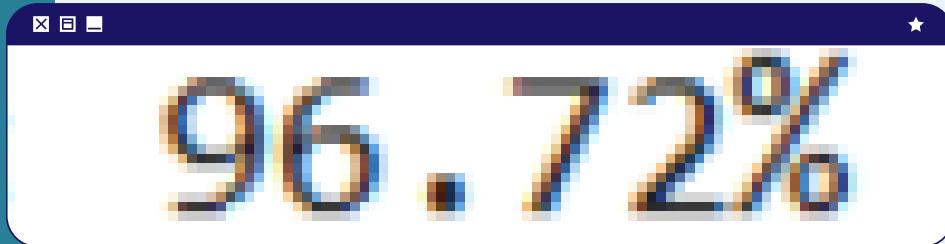


# CONFRONTO ACCURACY DEI MODELLI ALFANUMERICI

Accuracy modello non pre-addestrato:

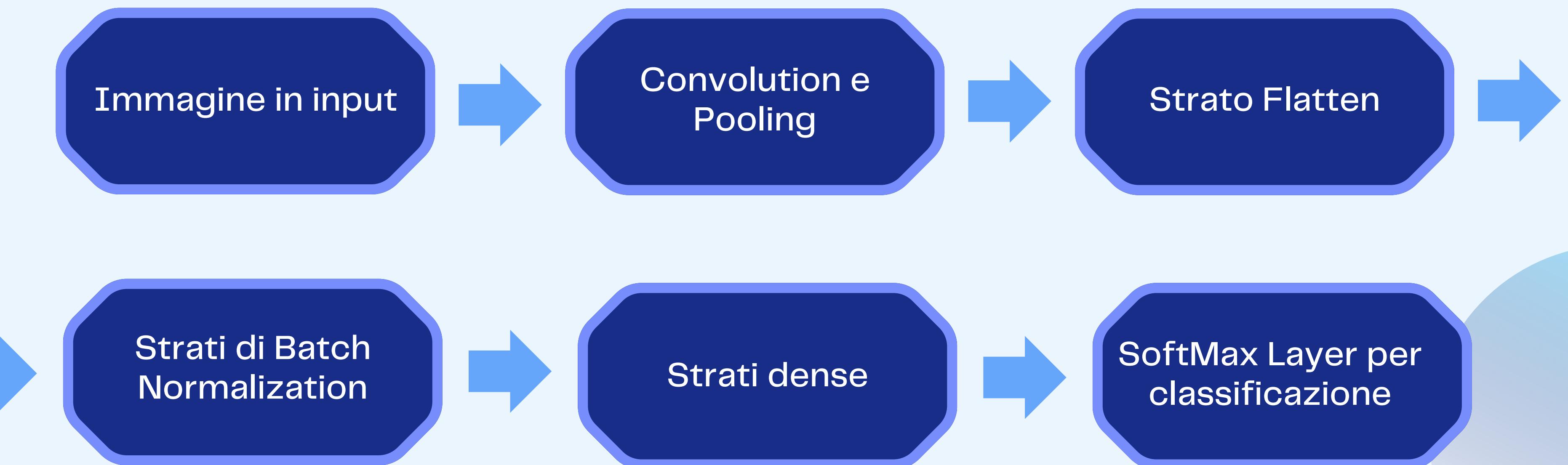


Accuracy modello pre-addestrato:



# MODELLO IMMAGINI

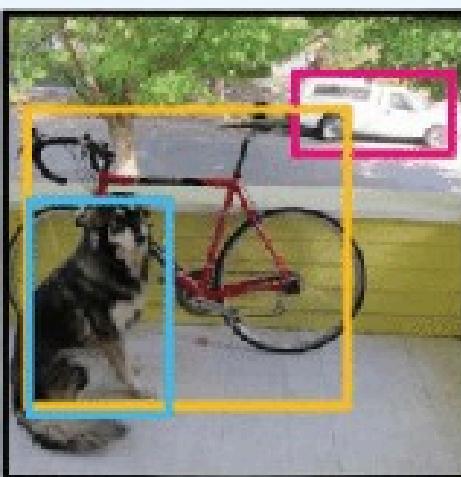
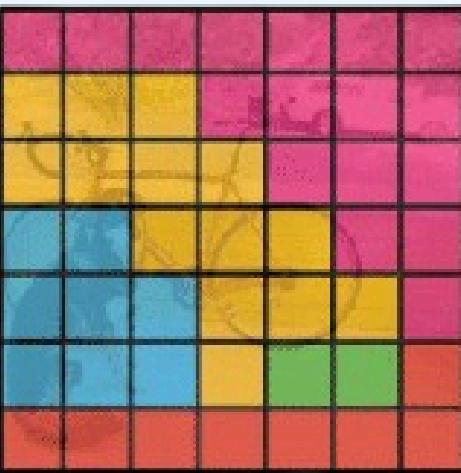
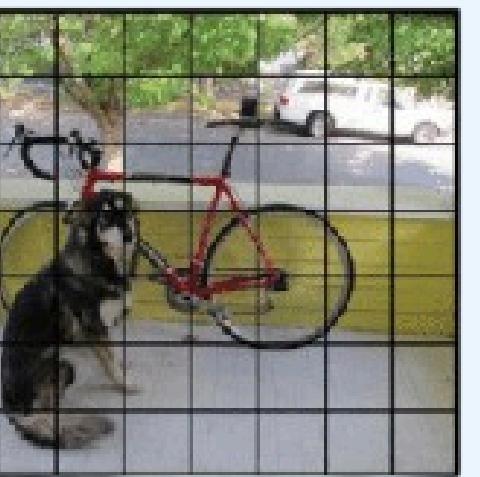
16/27



# MODELLO IMMAGINI PRE-ADDESTRATO

17/27

YOLO



## ALTERNATIVE



### AntiCAPTCHA

Servizio online che permette di superare automaticamente i CAPTCHA



### VERTEX AI (o equivalenti)

Modello generativo AI in grado di analizzare un'immagine.

# DEMO DEI MODelli

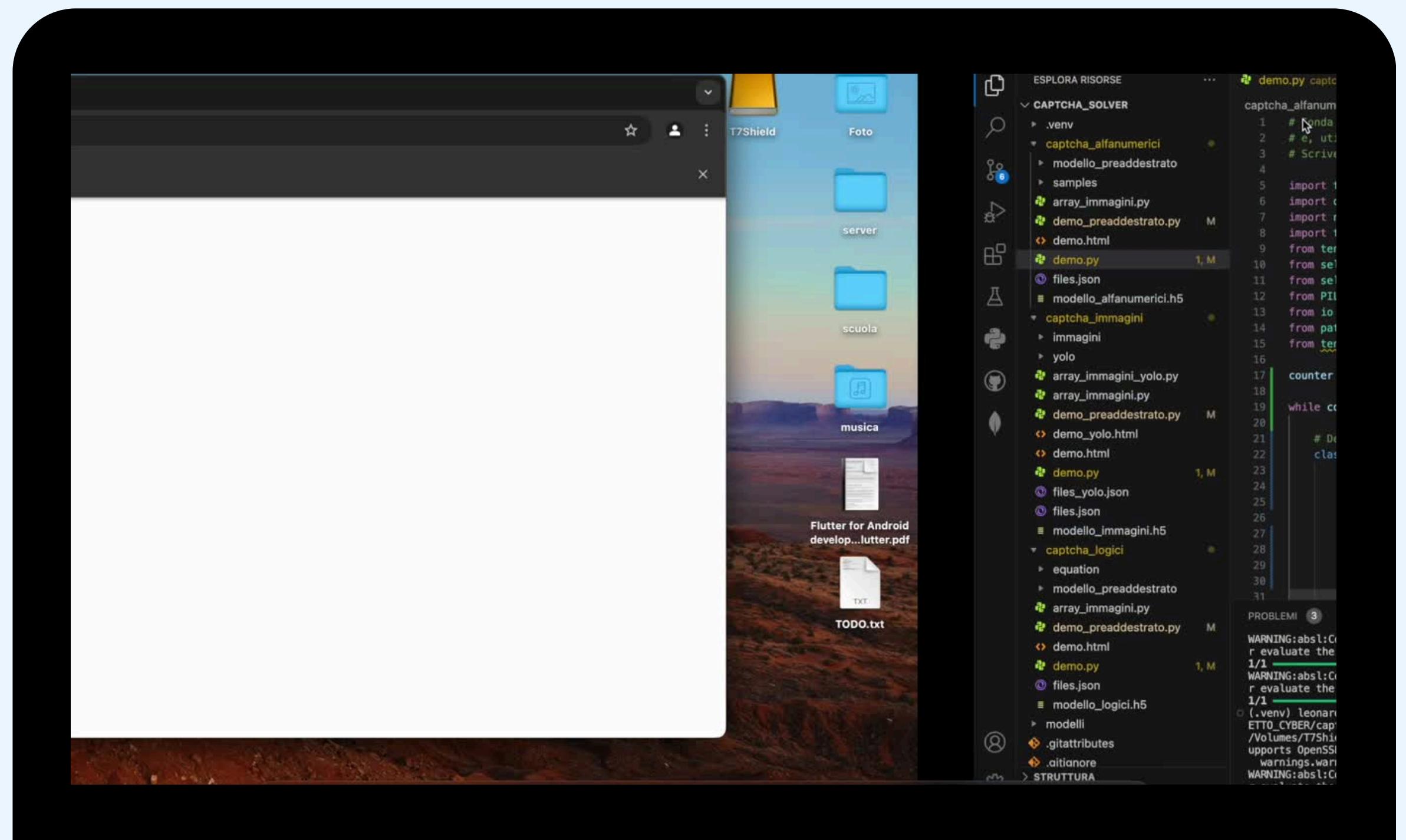


# IMPLEMENTAZIONE

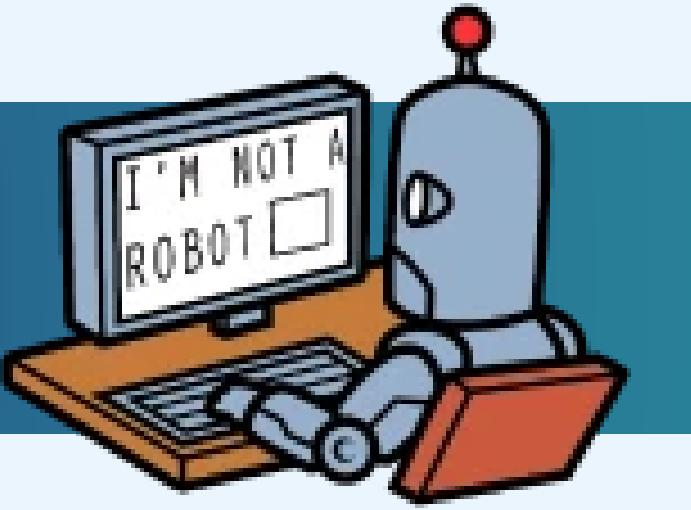
20/27

- 1 Creazione pagina web con immagine CAPTCHA casuale
- 2 Importazione modello addestrato
- 3 Addestramento sonda Selenium





# CONCLUSIONI



# RISULTATI OTTENUTI

I modelli **pre-addestrati** hanno mostrato una migliore generalizzazione, grazie alla loro capacità di adattarsi a nuovi dati con prestazioni superiori.

I modelli **non pre-addestrati** hanno comunque prodotto risultati soddisfacenti, ma necessitano di maggiori risorse e dati per evitare l'overfitting.

### Accesso ai dati

La difficoltà nel reperire dataset ampi ha limitato la generalizzazione dei modelli.

### Limitazioni computazionali

L'uso di una GPU limitata ha rallentato l'addestramento e la sperimentazione.

### Imprevedibilità dei CAPTCHA

Alcuni CAPTCHA più avanzati stanno diventando più resistenti agli attacchi automatizzati.

# LIMITAZIONI DEL PROGETTO



## Reperimento di nuovi dataset

Creare o ottenere dataset più variegati.

25/27

## Miglioramenti nel modello di riconoscimento delle immagini

Migliorare il modello di risoluzione di CAPTCHA basati sul riconoscimento di immagini.

## Implementazione modelli per risoluzione di altri CAPTCHA

Implementare nuovi modelli per la risoluzione di altre tipologie di CAPTCHA.

## Applicazioni pratiche

Esplorare applicazioni reali dei modelli sviluppati.

# SVILUPPI FUTURI



# CONSIDERAZIONI FINALI

Si è rivelato come sia sempre più importante di adottare sistemi CAPTCHA più avanzati, con complessità visive e contestuali variabili, per resistere meglio agli attacchi.

**GRAZIE  
PER  
L'ATTENZIONE**

