



CS 419: Computer Security

# Week 6 Recitation: CAPTCHA

© 2025 Paul Krzyzanowski. No part of this content may be reproduced or reposted in whole or in part in any manner without the permission of the copyright owner.

# CAPTCHA: Detecting Humans

# Gestalt Psychology (1922-1923)

- **Max Wertheimer, Wolfgang Köler, Kurt Koffka**
- **Laws of organization**
  - Proximity
    - We tend to group things together that are close together in space
  - Similarity
    - We tend to group things together that are similar
  - Good Continuation
    - We tend to perceive things in good form
  - Closure
    - We tend to make our experience as complete as possible
  - Figure and Ground
    - We tend to organize our perceptions by distinguishing between a figure and a background

Source: <http://www.webrenovators.com/psych/GestaltPsychology.htm>

# Gestalt Psychology

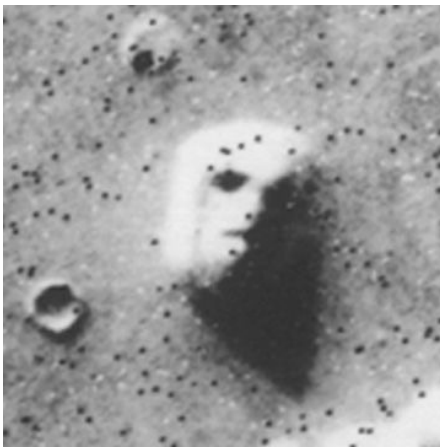


**18 x 22 pixels**

# Objects on Mars?



**Elvis**



**Face**

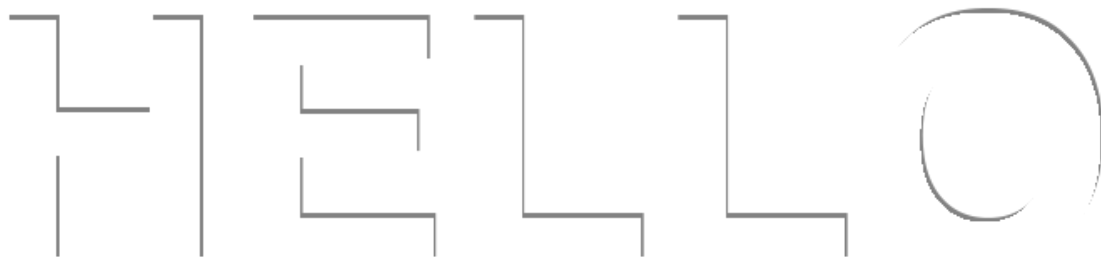


**Female statue**

**Our brains try to map objects into things we know**

# Gestalt Psychology: text continuity

**This isn't text but it we can read it like text:**



# Gestalt Psychology

**This isn't text but it we can read it like text:**



HELLO

**(This is trivial for today's AI models to parse)**

# Authenticating humanness

## Battle the Bots:

- Create a test that is easy for humans but extremely difficult for computers

## **CAPTCHA: Completely Automated Public Turing test to tell Computers and Humans Apart**

- Image Degradation
  - Exploit our limits in OCR technology
  - Leverages human Gestalt psychology: reconstruction

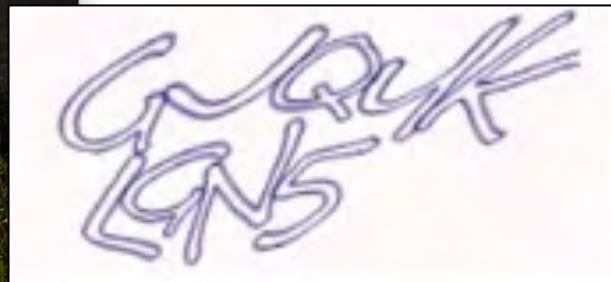
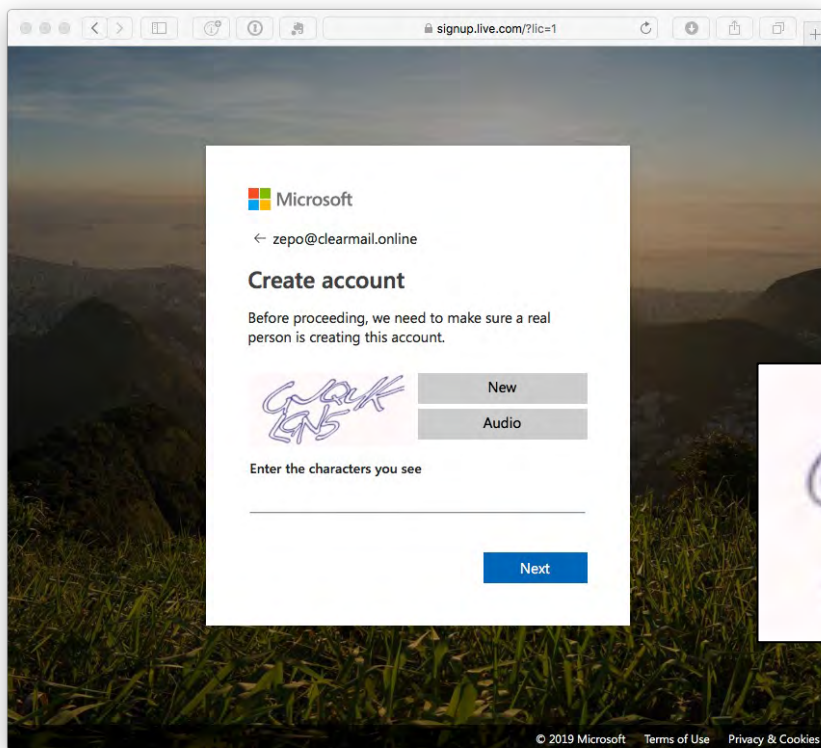
## Origins

- **1997**: AltaVista (an early search engine) – prevent bots from registering URLs with the search engine
- **2000**: Yahoo! and Manuel Blum & his team at CMU created **EZ-Gimpy**
  - Distort one of 850 words
- **2003**: Henry Baird @ CMU & Monica Chew at UCB created **BaffleText**
  - Generates a few words and random non-English words



# CAPTCHA Example (2019)

## Microsoft



See [captchas.net](https://captchas.net)

# They had to get more difficult

Advances in machine learning & character recognition led to automated solving

Name

E-mail Address

Password Confirm Password

Please type captcha here 562

(Optional) Please enter your phone number if you'd like us to call you to explain our products and services.

Select Country Code Phone Number

☐ I'd like to receive email about product updates, personalized recommendations, offers, and PowerPoint and presentation tips and tricks.

**SIGN UP**

By Pressing "Sign up" you accept our [Privacy Policy](#)

Microsoft account

Help us make sure you're not a robot

Enter the characters you see

[New](#) | [Audio](#)

FW58dPPW3R

☐ Send me email with promotional offers from Microsoft. (You can unsubscribe at any time.)

Click **I accept** to agree to the [Microsoft services agreement](#) and [privacy & cookies statement](#).

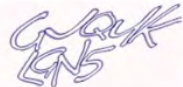
**I accept**

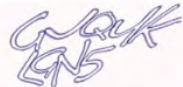
Microsoft

← zepo@clearmail.online

**Create account**

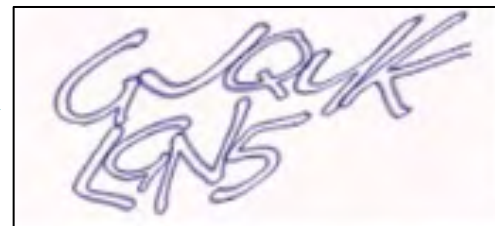
Before proceeding, we need to make sure a real person is creating this account.

 **New**

 **Audio**

Enter the characters you see

**Next**



# Problems

- **Accessibility**

- Visual impairment → audio CAPTCHAs
- Deaf-blind users are left out

- **Frustration**

- Typing text was more of a pain on mobile devices
- OCR & computer vision algorithms improved a lot!
- Challenges that are now difficult for computers may be difficult for humans

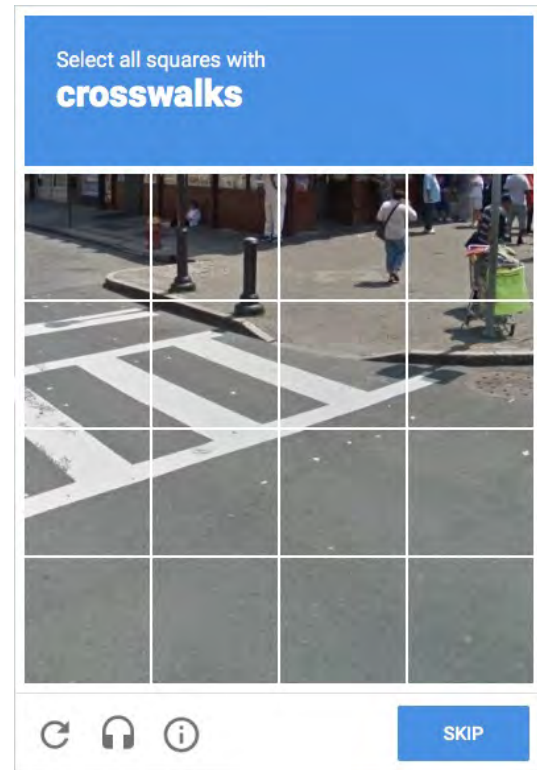
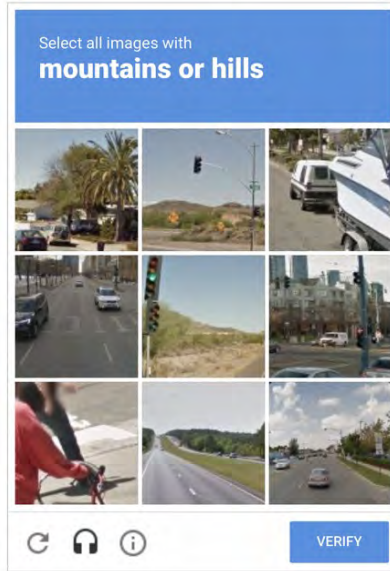
- **Attacks**

- Man in the middle attacks – software can redirect the challenge to humans
  - Use human labor – CAPTCHA farms
- Automated CAPTCHA solvers
  - Initially, educated guesses over a small vocabulary – later, use improved image recognition



# Alternate approaches

- Puzzles, scene recognition
- Touching is easier than typing on phones



# Alternate Approaches

Select 2 objects that are the same shape:



↻ Refresh ⓘ Report a problem

Confirm

Verify to continue:



↻ Refresh ⓘ Report a problem



Slide to Submit

SEND MESSAGE >



# More CAPTCHA Examples

login.secure.mercedes-benz.com

## Registration

Please fill in the fields below to register for Mercedes me.

Fields marked with an asterisk (\*) must be completed.

First name \*

Last name \*


Email address \*

Repeat e-mail \*

Enter characters from the image \*



[Reload image](#)

☐ I have read and agree to the [Access conditions](#).

Register

unitedwifi.com

## UNITED

Internet Available

EWR SEA

Arrives in 1h 40m

SEA 46°

Not a member? [Enroll now](#)

Please enter the numbers shown below.

**43543**

[Hear the numbers read out loud](#) or [try different numbers](#)

☒ I accept the [Terms and Conditions](#) and [Privacy Policy](#) of United's inflight Internet.



Sign in

Microsoft

← rutgers\_user@outlook.com

## Create account

Use the arrows to rotate the object to face in the **direction of the hand**.  
(1 of 1)



Submit

30817ca9aa1389c39.3150633801

Audio Restart

# reCAPTCHA

Ask users to translate images of real words & numbers from archival texts

## Two sections

(1) known text

(2) image text

- Assume that if you get one right , then you get the next one correct
  - Try it again on a few other people to ensure identical answers before marking it correct

## Google bought reCAPTCHA 2009

- Used free human labor to improve transcription of old books & street data
- This enabled the use of human labor to fix up the archives of the New York Times

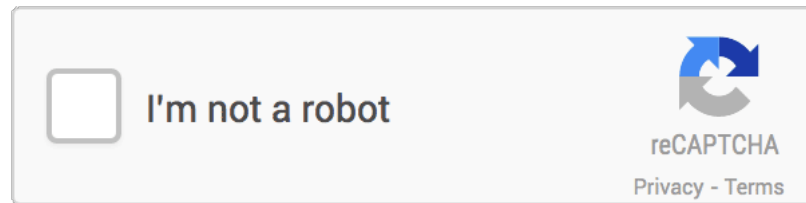
## By 2014:

Google found that AI could crack CAPTCHA & reCAPTCHA images with 99.8% accuracy



# NoCAPTCHA reCAPTCHA: reCAPTCHA v2

*Just ask users if they are a robot!*



- **Reputation management**

- “Advanced Risk Analysis backend”
- Check IP addresses of known bots
- Check Google cookies from your browser
- Considers user’s engagement with the CAPTCHA: before, during, and after
  - Mouse movements & acceleration, the precise location of clicks
- Generate a confidence score and allow the service to decide if it's good enough

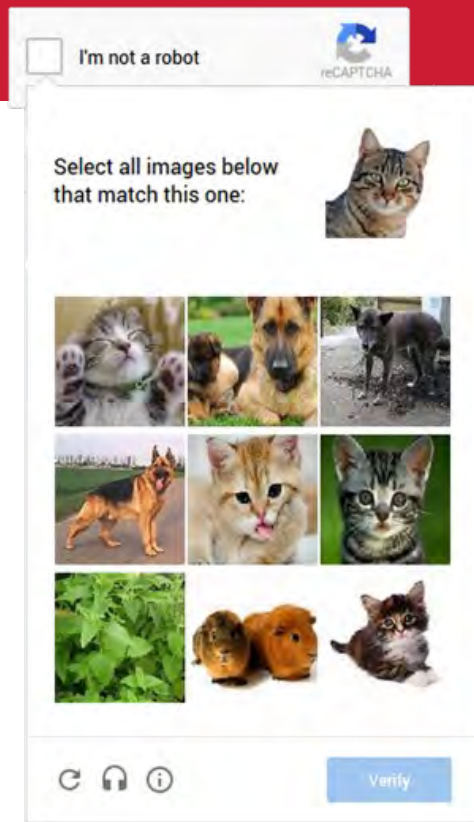
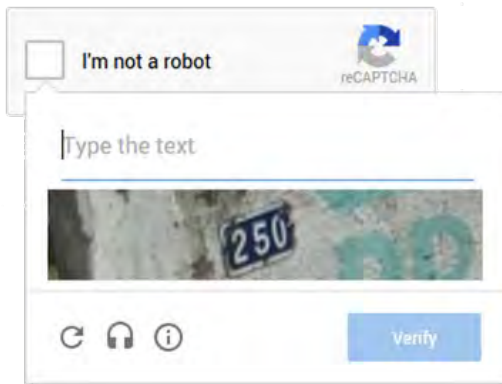


# reCAPTCHA v2 fallback

## If risk analysis fails, present a CAPTCHA

Three challenge types:

1. **Classification**: identify which images in a 3x3 grid belong to a given description (e.g., find all images with bridges)
2. **Classification**: like (1) but images are replaced after clicking (“click until there are none left”)
3. **Segmentation**: break an image into a 4x4 grid and identify parts that are relevant to the request (e.g., identify all parts of a motorcycle)



reCAPTCHA v3 (**invisible reCAPTCHA**) makes a decision only based on past interactions, potentially locking users out of a service

# The AI Threat

- **In 2024, a team at ETH Zurich demonstrated that reCAPTCHA v2 challenges can be solved 100% of the time using publicly-available AI software**
  - Apply a fine-tuned version of the open-source YOLO (You Only Look Once) object-recognition model
  - Use a VPN to connect with a different IP address for lots of repeated attempts – makes each connection appear to be unique
  - Incorporate artificial Bezier curve-based mouse movements to simulate human behavior

See <https://arxiv.org/abs/2409.08831>

# IllusionCAPTCHA: An attempt at defeating AI

## AI-created optical illusions are not recognized by other AI systems

- Gen AI combines an input image and a prompt
- AIs can create images but not detect the illusions: **LLMs failed 100% of the time**
- **Humans passed the test 83% of the time**



<https://arxiv.org/abs/2502.05461>

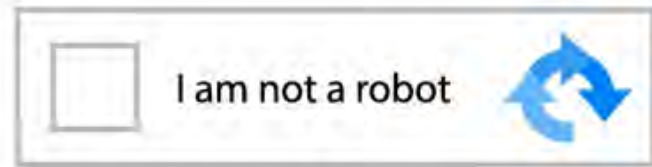
<https://archive.is/mAHQC>

<https://www.newscientist.com/article/2468020-ai-generated-optical-illusions-can-sort-humans-from-bots/>

# Fake CAPTCHA Prompts Used Maliciously

**In 2024, the Ukrainian Computer Emergency Response Team warned that the APT28 threat group (Fancy Bear, thought to be affiliated with Russian Intelligence) has been using CAPTCHA impersonation**

- Present a fake “I am not a robot” message to get users to click on the checkbox
- This initiates a malicious PowerShell command to the user’s clipboard
- The attack targets government workers in Ukraine but can inspire other attackers to use the same technique



<https://www.forbes.com/sites/daveywinder/2024/10/26/new-google-cyber-attack-warning-as-russian-apt28-hackers-strike/>

# Other approaches: Text/email verification

- **Text/email verification**

- Ask users for a phone # or email address
- Similar to two-factor authentication, but we're *not authenticating* the user
  - Just having them do something
- Service sends a message containing a verification code
  - Still susceptible to spamming & automation
  - Makes the process more cumbersome
  - Requires users to disclose information

- **Measure form completion times**

- Users take longer than bots to fill out and submit forms
- Measure completion times and randomness in delays
  - But bots can program delays if they realize this is being done

# The End