

# CAPTCHA

Completely Automated Public Turing test to tell  
Computers and Humans Apart

# The Problem!

A mechanism of preventing automated systems from infiltrating human systems.

These automatic scripts and programs are known as bots, and everyone here, since the proxy project, now has the skill to write these bots.

The information is trivial to an advanced computer programmer.

# The Motivation!

In 1997, Alta Vista did not use a bot, instead relying on user submitted URLs. Bots did the work anyway, and when Alta Vista added a very primitive CAPTCHA, submissions dropped by 95%.

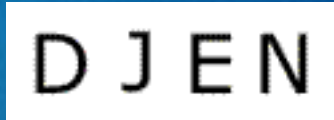
In 1999 Slashdot ran a poll asking what school was the best for Computer Science Majors. Carnegie Mellon students wrote a script. MIT students retaliated. At the end, each school had 20,000 votes, other schools had only 3 digit scores.

Slashdot still has open voting because they're simply fun polls. Few people have motivation to rock the vote there.

Time's people of the year!

# The Approach!

The very first of the CAPTCHAs were simple.



By rendering the text in the image, the bots were unable to read them.

This was broken pretty quick, so extra graphics were introduced into the image.



It turned out to be simple, fast, and effective. Only the most advanced bots could pass through this detection, and it took them a while to do so.

# The History!

Developed by CMU in 2000 by students and professors.

Trademark and patent applications were dropped in April 2008.

The particular technology of CAPTCHA is no longer in use; these days a system called reCAPTCHA is in place. However, the generic term CAPTCHA is still applied.

Version counting is extremely hard, mostly because version numbers were never kept. Many changes were simply modification to variables that control obscuring. Some were modifications to the original, some were total re-writes that merged back into the development tree. Often they were developed by third parties. The extreme variety helped ensure security, however, within a particular site, only a single style was used.

# More History!

reCAPTCHA is the currently recommended technology.

The new technology is run on Twitter, Facebook, and TicketMaster. It hosts 30 million images per day.

Authorization is handled offsite with the reCAPTCHA program, and is used for more than simple bot-control.

# The Assumptions!

The underlying assumptions are pretty clear. The developer assumes that the bots cannot decode the images as easily as cleartext, and must return a cleartext.

The stated mission on [recaptcha.com](https://recaptcha.com) is to keep the technology a few generations ahead of the bot writers. They know that the enemy will always, given enough time, defeat their mechanisms, and must escalate.

Also, some technologies rely on the fact that the human mind can extrapolate information while the machines cannot, by eliminating some information and adding extra.



# The Strengths!

The more advanced the protection, the harder it is to write the bots.

More technical knowledge is required, and more overhead required for each attempted infiltration.

Upgraded technology is not a major change for the average user. It may even be seamless.

Brute force password breaks and dictionary attacks trigger lockouts of an account. CAPTCHAs prevent a malicious script from locking people out of accounts at will.

TO COMPLETE YOUR WEB REGISTRATION, PLEASE PROVE THAT YOU'RE HUMAN:

WHEN LITTLEFOOT'S MOTHER DIED IN THE ORIGINAL 'LAND BEFORE TIME,' DID YOU FEEL SAD?

- YES
- NO

(BOTS: NO LYING)

# Recycled Manpower!

The current technology of reCAPTCHA solves two problems at once: user proof of identification (not authentication), and decipherment of digitized books.

Technology used to break the original CAPTCHAs are used currently in the digitizing of existing print books.

Each scan passes through a series of decipher tools, and when disagreements or ambiguities are discovered, they are converted to a CAPTCHA, and passed to a user.

When a sufficient number of humans have reach a concesious, the word is considered solved and used a bias for future deciphering.

# The Weaknesses!

Takes more time for the user.

Once defeated, a new technology needs to be rolled out.

Non-english keyboards may have a problem.

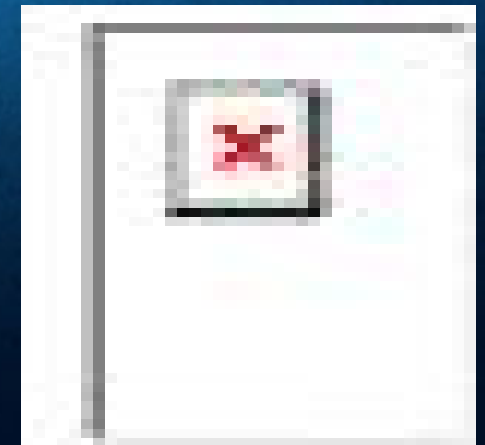
People who don't read the Latin alphabet may have issues translating the characters.

If the custom authorization requires a database, like with reCAPTCHA, SQL Injection could occur.

reCAPTCHA itself if properly controlled, but this requires more overhead for reCAPTCHA.

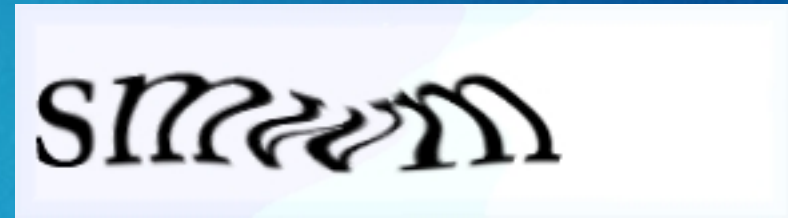
Bandwidth problems:

Cached images.



# The Tactics!

An early GIMP CAPTCHA, after basic systems had been cracked, in an attempt to distort the letters.



An attempt to make segmentation difficult by running a line through them.

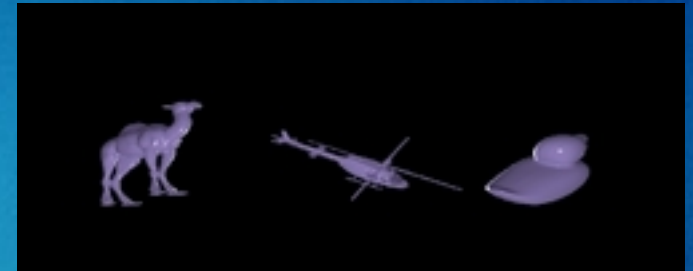


Heavy warping hides the characters and makes differentiation more difficult.



# The Future!

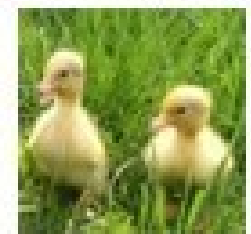
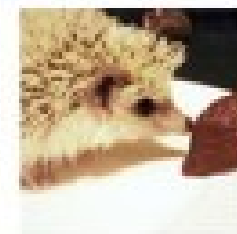
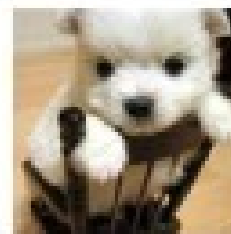
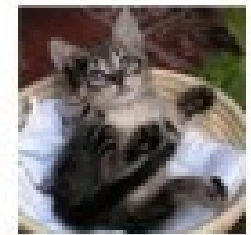
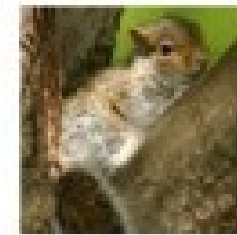
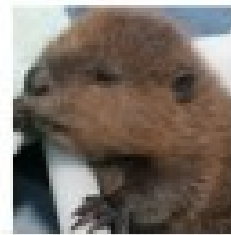
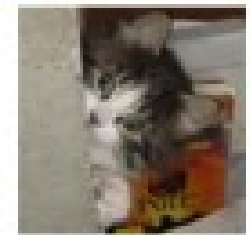
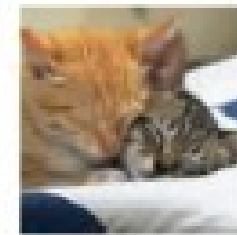
Graphic recognition of rotated 3D objects.  
An archive of objects is stored and  
rendered as needed.



Animation!  
The movies are revolutionary.

Animals!  
Pick the kittens.

Click 3 pictures of kittens to submit



# Extended Projects!

MailHide - the organization responsible for developing reCAPTCHA is developing MailHide, in an effort to hide email addresses from harvesters.

Book Scanning - The software used to crack captchas has been accessible to the general public lately, and taken by the same people who attempt to scan books and newspapers into digital records.

# Tandem operation!

The strongest forms one could implement are hybridizations, using multi fonts in a single image to confuse bots, character obstruction, extra information, and missing characters.

Two CAPTCHAs in the same page allow for two completely different rendering systems at once. By have four or five that the site can rotate through, it can confuse bots even more.

However, this is twice as annoying.

# Lots can go wrong!

Poor filtering.



## Eyestrain Captcha

☆ ADD TO FAVES    📄 BLOG THIS    🔍 ALL SIZES



WORD  
VERIFICATION

muaumwww|



## Qualifying question

Just to prove you are a human, please answer the following math challenge.

Q: Calculate:

$$\frac{\partial}{\partial x} \left[ 4 \cdot \sin \left( 7 \cdot x - \frac{\pi}{2} \right) \right] \Big|_{x=0}$$

A:

mandatory

Note: If you do not know the answer to this question, reload the page and you'll get another question.


Awful schemes.

Probably should never use anything with variables.



Undecipherable!

Write the text from image below to this textbox

Poor contrast, easier for a computer

Illegible.

**Verify Your Registration**

• Enter the code shown:  [More info](#) ⓘ

This helps prevent automated registrations.



enter the characters  
for the symbols shown  
in the box below:



♁ = 4 ♁ = j ♀ = d □ = n ♪ = x + = k  
▶ = v 😊 = r ♖ = h ◇ = t ☢ = 7 ♗ = a

Too slow



8Y70IO1

Too obscured

Too vague

Please enter the code you see below. [what's this?](#)



# The Conclusion!

CAPTCHA technology is not here to stay.

The technology is also used to slow SQL injection attacks, bots attempting to discover vulnerabilities are inhibited by the CAPTCHA.

Extra overhead required to run bots requires bot operators to have more hardware at their disposal in order to perform even basic tasks.

Image authorization must be ensured completely independently and securely.

Credits:

<http://www.johnmwillis.com/sillystory/top-10-worst-captchas/>

<http://citalan.blogspot.com/2007/12/captcha-and-alternatives.html>