

BeCAPTCHA

Edoardo De Matteis

Contents

1	Introduction	2
1.1	Historical background	2
1.2	Description	3
1.3	Requirements	3

1 Introduction

Nowadays when surfing the web CAPTCHA tests are everywhere, to the point we are used to get asked to prove our humanity and something is off when it does not happen. Furthermore, technology moves forwards and tests are becoming increasingly harder, while it may be a nuisance for the abled it can often be a barrier for people with disabilities. It is a priority to develop reliable CAPTCHA tests designed to be accessible to everyone.

1.1 Historical background

The word CAPTCHA stands for *Completely Automated Public Turing test to tell Computers and Humans Apart*. In the year 2000 during a speaking at Carnegie Mellon University held by YAHOO!, at the time one of the world's biggest companies, they listed the ten most relevant problems they did not know how to solve; among these issues stood out that spammers would write programs to automatically generate mails, and use them for malicious purposes [4, 3]. The world needed a test to distinguish humans from computers such that any human, regardless of age, gender, education or language could pass it. To make things even more difficult, the CAPTCHA test is born from a paradoxical idea: any computer should be able to grade it but should not be able to pass it, like if a teacher could grade students' exams without knowing the subject.

CAPTCHA The idea exposed in [1] is based on the fact that humans are extremely good at pattern and optical character recognition i.e., reading, and we generally learn to do it since we are kids and can do it under many circumstances like bad lighting and other. A text is took and warped, stretched or generally modified in some way such that computer vision techniques at the time could not read it, yet the computer with the gold truth just has to match it with the user's input. Once released CAPTCHA has been used to authenticate access to YAHOO! mail services.

reCaptcha All the character input by humans into CAPTCHA gave new data that made computers smarter. The new version, reCAPTCHA [2], uses two words: one generated so the computer knows the answer and the other pulled from a book or originally a *The New York Times*' old article, both of which are distorted. When a human gets the first word right the system guesses the second one is probably right and distributes it across different users, if there is consensus then it is approved. So many tests were done that something along one year of NY Times articles was digitalized every 4 days, this led Google to acquire reCAPTCHA in 2009 and exploit it to digitalize texts in its archive. When you do that enough times you get a big enough dataset of distorted characters that computers can effectively learn how to read them, in a way CAPTCHA taught machines how to read highly warped texts, in a 2014 machine learning study Google stated how humans have a 33% accuracy on recognizing CAPTCHA's warped characters while a trained machine has 99.8%.

reCAPTCHAv2 Given the huge potential reCAPTCHA showed to have, Google exploited it to have human users classify images and build up a data, that were later used to improve Google Maps [4]. Again, computers eventually became good enough on images that reCAPTCHA v2 is not enough anymore.

NoCAPTCHA reCAPTCHAv3 Computers changed and tests have to change as well, NoCAPTCHA and reCAPTCHA v3

1.2 Description

1.3 Requirements

Requirements like what is needed programming

References

- [1] Luis von Ahn et al. “CAPTCHA: Using Hard AI Problems for Security”. In: *Advances in Cryptology — EUROCRYPT 2003*. Ed. by Eli Biham. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 294–311. ISBN: 978-3-540-39200-2.
- [2] Luis Ahn et al. “reCAPTCHA: Human-based character recognition via Web security measures”. In: *Science (New York, N.Y.)* 321 (Sept. 2008), pp. 1465–8. DOI: 10.1126/science.1160379.
- [3] Josh Dzeja. *Why CAPTCHAs have gotten so difficult*. <https://www.theverge.com/2019/2/1/18205610/google-captcha-ai-robot-human-difficult-artificial-intelligence>. 2019.
- [4] Vox. *Why captchas are getting harder*. 2021. URL: <https://www.youtube.com/watch?v=lUTvB108eEg&t=376s> (visited on 05/14/2021).