

CAPTCHA

Humains contre Robots

Olivier Le Goaër

Entrée Libre #2
Centre des Abeilles, 28 Juillet 2021



Plan de la présentation

- x Le test de Turing
- x Les robots
- x Les CAPTCHAs
- x L'économie des CAPTCHAs
- x Démonstration : FloodBot
- x Le mot de la fin

Le test de Turing

Alan Turing (†1954)

x Était mathématicien de formation

- Eh oui, la science informatique n'était pas encore née !

x Père fondateur de l'informatique

- Des contributions à la fois techniques ([algorithme](#)) et philosophiques (ce qui est « algorithmisable » ou ne l'est pas)

x Prix Turing = Prix Nobel pour l'informatique

- Le lauréat 2018 est français et travaille dans l'IA*

Le test de Turing

- x De tous temps, tenir une conversation sensée en langage naturel est l'apanage des êtres dotés d'intelligence
- x Si un programme informatique arrive à se faire passer pour un humain aux yeux d'un juge humain lors d'une conversation écrite, alors le programme a atteint le stade d'intelligence*

*Turing parlait de « conscience »

Ce qu'il faut en retenir

x Certaines tâches sont historiquement **difficiles à algorithmiser**

- Reconnaître le contenu d'une image, tenir une conversation écrite ou orale sur un sujet non connu à l'avance, etc.

x Les **tests de Turing** se réfèrent à ces épreuves qui requièrent une forme d'intelligence pour les passer avec succès

Les robots

Les robots en informatique

- x Un programme autonome, supposé intelligent, est appelé robot, ou simplement « bot »
- x Certains sont légitimes (*Good Bots*), d'autres sont malveillants (*Bad Bots*)
- x De nombreux services en ligne (c-a-d sites Web) sont la cible des *Bad Bots*

Actes de malveillance sur le Web

- x Acheter toutes les places de concert sur la billetterie en ligne (pour les revendre)
- x Forcer un mot de passe en tentant toutes les combinaisons possibles
- x Exploiter les formulaires de contacts pour de la publicité non désirée (Spam)
- x « Bourrer les urnes » lors d'un vote en ligne
- x ...

Soyez attentif à l'actualité

ACCUEIL > HIGH-TECH

Cdiscount : Le compteur de votes d'un concours de courts-métrages piraté, un prix annulé

Un quart des tweets climatosceptiques proviennent de bots

Par L'EXPRESS.fr,
publié le 25/02/2020 à 12:30

Cdiscount @Cdiscount · 7 nov. 2019
[1/4]
Même si le Festival 21 est un succès, plus de 20 000 votes sur chacune des vidéos en si peu de temps, c'est (beaucoup) trop gros pour être vrai 😊

Cdiscount @Cdiscount
[2/4]
Nos investigations nous ont amené sur la piste d'utilisation frauduleuse de robots permettant de multiplier le nombre de vote sur le site festival21-cdiscount.com. Dans ces conditions, nous ne pouvons maintenir le prix « Coup de Cœur du Public »



Festival21 - Le concours de création vidéo, par Cdiscount !
Festival21 est le concours de création vidéo accessible à toutes et à tous, initié par Cdiscount. A partir du 4 Octobre 2019, réalisez la vidéo la festival21-cdiscount.com

39 17:04 - 7 nov. 2019

Twitter et Facebook annoncent avoir bloqué des opérations de manipulation politique

Le démantèlement de plusieurs milliers de comptes concerne notamment des campagnes de manipulation venues de l'étranger et en faveur de Donald Trump.

Le Monde avec AFP · Publié le 21 décembre 2019 à 03h02 - Mis à jour le 21 décembre 2019 à 12h34

Cet automne, ils ont démantelé plusieurs opérations similaires à celles révélées vendredi, soutenues par différents pays (Iran, Russie, Arabie saoudite, etc.), et cherchant à manipuler l'opinion de leurs citoyens ou de populations d'autres pays. Elles ont généralement recours à des robots, des faux profils et des technologies d'intelligence artificielle pour se fondre dans la masse, propager les messages de façon efficace et générer du trafic.

Faits divers – Justice

Il en avait marre d'attendre, il pirate la billetterie du Hellfest

Vendredi 18 octobre 2019 à 17:59 - Par Pascal Roche, France Bleu Loire Océan, France Bleu, France Bleu Paris



Il voulait absolument une place

Cet homme sans histoire âgé de 30 ans domicilié près de Rennes, consultant informatique diplômé d'un bac +5, a expliqué aux enquêteurs qu'il voulait absolument un billet pour le festival Hellfest mais qu'il ne parvenait pas à aller au bout de la procédure d'achat une fois entré sur le site. Agacé et dans "l'espoir de griller la file d'attente" comme le raconte une source proche du dossier, il a alors mis au point et lancé un petit robot qui a paralysé le système.

ChatBot : agent conversationnel

- x Léo est un bot, pas un être humain
- x Améliore la relation client et le commerce



Dialoguez avec Léo

Le 05/09/18 16:07:32

Bienvenue, je suis Léo votre caviste virtuel, je vais vous aider à choisir une bouteille. Que cherchez-vous ?

Le 05/09/18 16:54:15

Je cherche un type de vin particulier

Le 05/09/18 16:54:16

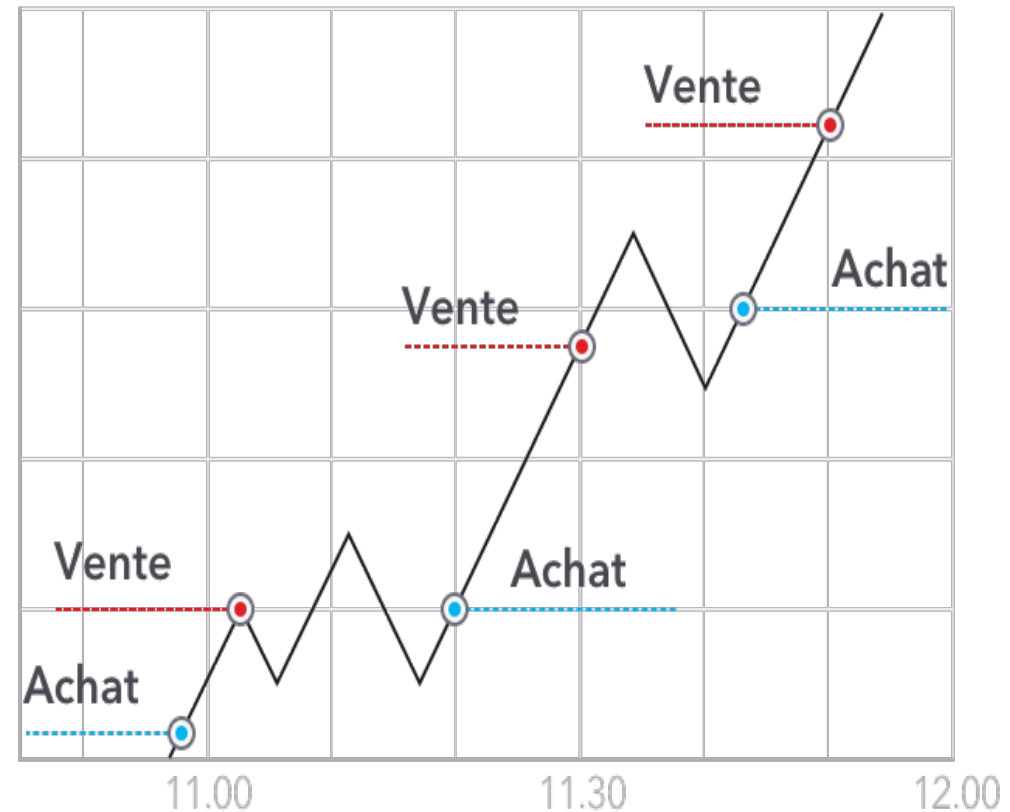
Parfait, quel type de vin puis-je vous conseiller ?

Rouge Blanc Rosé Pétillant

Champagne

TradingBot : marchés financiers

- x Analyse les cours des marchés (ex: BitCoin)
- x Effectue des ordres d'achats/ventes à votre place



GoogleBot : moteur de recherche

- x Parcours les pages Web récursivement (en suivant les liens)
- x Construit un index géant dédié à la recherche par mots-clés de Google Search



On-page SEO 🔍

On-Page Ranking Factors [2019 SEO] - Moz
<https://moz.com/seo/learning-center> ▼
The way your **page** is optimized can have a huge impact on its ability to rank. ... **On-page** ranking factors can have a big impact on your **page's** ability to rank if optimized properly. ... Title tags are the second most important **on-page** factor for **SEO**, after content.

On-Page SEO: Anatomy of a Perfectly Optimized Page (2019 Update)
<https://backlinko.com/on-page-seo>
Jan 9, 2019 - When it comes to **on-page SEO**, I'm sure you've heard enough about meta tags and keyword density for one lifetime. If you're looking for some ...

On Page SEO - Everything You Need to Know | SEJ
<https://www.searchenginejournal.com/seo/> ▼
Sep 24, 2016 - **On-page** refers to both the content and HTML source code of a **page** that can be optimized, as opposed to **off-page SEO** which refers to links and other external signals." It takes into account all aspects of the webpage that, when added together, will improve your rankings in the search results.

On-Page SEO: An Actionable Guide - Ahrefs
<https://ahrefs.com/blog/on-page-seo/> ▼
Nov 15, 2018 - There's more to **on-page SEO** than placing your keywords in meta tags. Learn how to perfectly optimize your web pages in 2019 (and beyond).

The On-Page SEO Cheat Sheet - Neil Patel
<https://neilpatel.com/blog/seo/> ▼
Despite the updates, however, **on-page SEO** hasn't really changed all that much. Every **on-page SEO** task is really for the user's benefit. Yet, most SEOs still ...

TrollBot : propagande

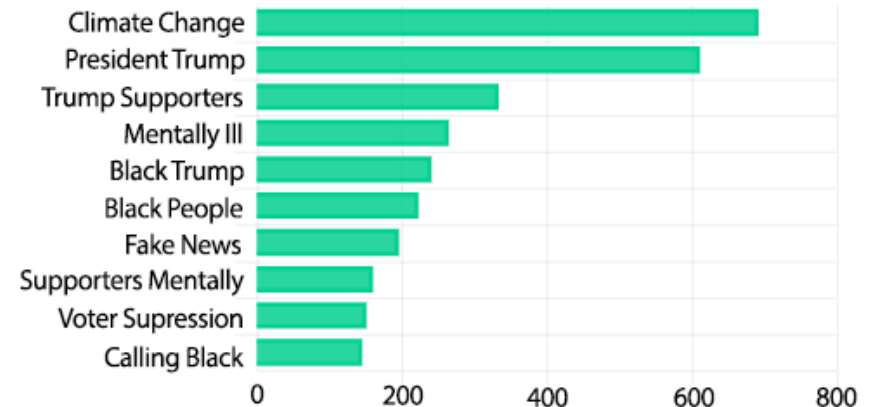
- x Tweets climatosceptique, pro-trump, etc.
- x Donne l'illusion d'une foule dont l'opinion risque d'influencer les débats



Tracking Trollbots

The day after CNN's Climate Crisis Town Hall with Democratic presidential hopefuls, the phrase "climate change" topped the list of two-word phrases appearing on "trollbot" Twitter feeds that Bot Sentinel tracks. It was unusual.

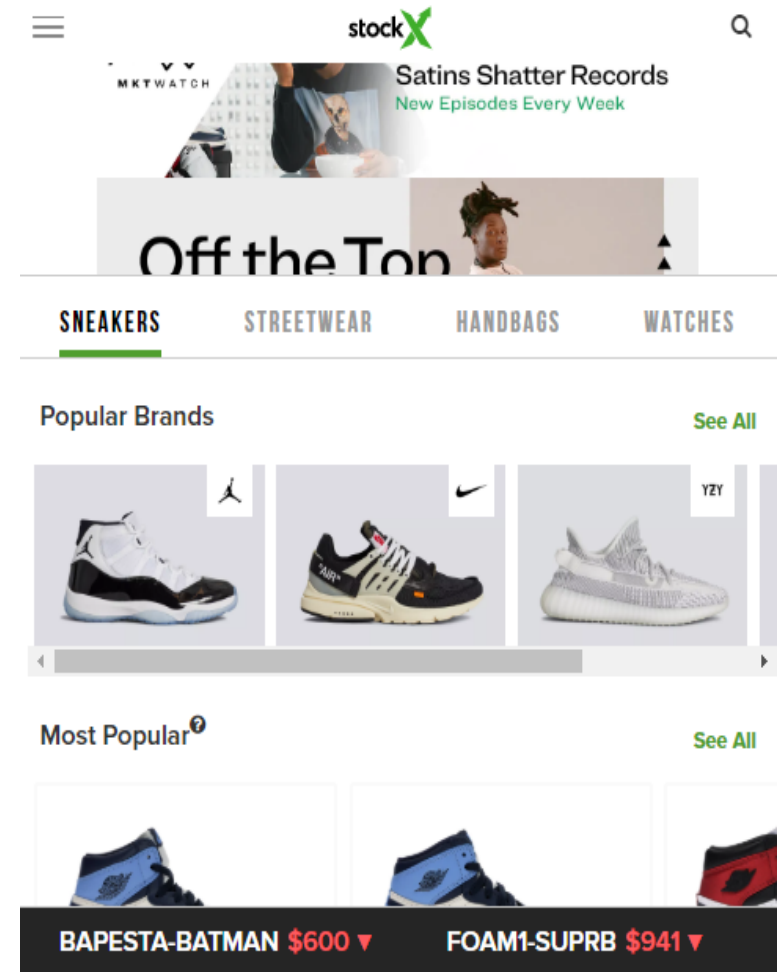
Top two-word phrases tweeted by tracked "trollbot" accounts over 24 hours Sept. 5, 2019



Source: Bot Sentinel

SneakerBot : achats spéculatifs

- x Les baskets en édition limitée sont un véritable business
- x Les acheter avant tout le monde, au meilleur prix, pour les revendre ailleurs



SpamBot : publicité non désirée

- x Exploite les formulaires Web (contact, commentaires)
- x Récolte les emails présents sur les sites Web



By [REDACTED]



4 comments:



Ahmed89 said...

Do you need Personal Loan?

Business Cash Loan?

Unsecured Loan

Fast and Simple Loan?

Quick Application Process?

Approvals within 24-72 Hours?

No Hidden Fees Loan?

Funding in less than 1 Week?

Get unsecured working capital?

Contact Us At : [REDACTED]@gmail.com

Phone number :+1 [REDACTED] (Whatsapp Only)

LOAN SERVICES AVAILABLE INCLUDE:

=====

Les CAPTCHAs

CAPTCHA ?

- x « Completely Automated Public Turing test to tell Computers and Humans Apart »
 - Test de Turing entièrement automatisé pour différencier les robots et les humains
- x Solution générique de sécurité anti-bot installée aux endroits critiques d'un site Web
 - Et plus généralement dans tous types de services numériques

Principe de base

x Pour stopper les bots, on va exiger une épreuve de type **défi-réponse**

- Son contenu varie et n'est pas connu à l'avance

x Test de Turing « inversé »

- Le juge est un programme (le CAPTCHA) qui va faire passer le test aux participants (humains et bots)
- Le test doit être difficile pour un programme **mais en même temps** facile pour l'humain

Défi maths

x **Analyser** (et calculer) par OCR* une expression mathématique est difficile pour un bot

Captcha:

$8+5=$

Qualifying question

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

Q: $5 - (-5) - 2 - (-3) + 4 = ?$

A:

mandatory

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

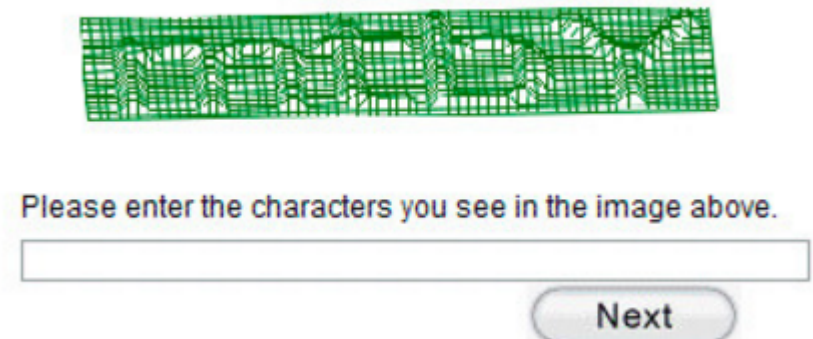
*Reconnaissance Optique de Caractères

Défi texte

- x Déchiffrer par OCR un texte déformé/bruité est difficile pour un bot



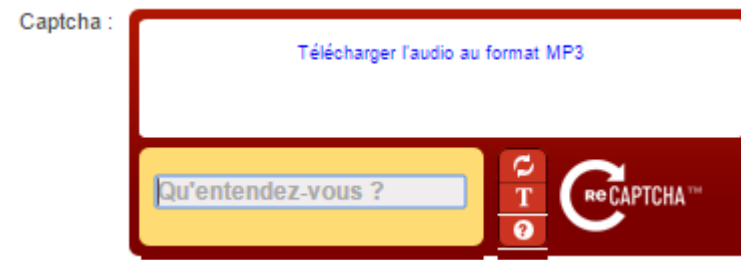
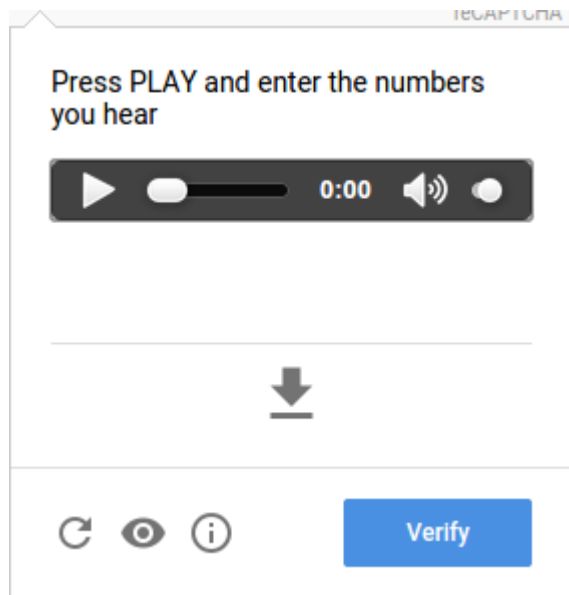
A screenshot of a 'Security Check' dialog box. The title bar is blue with the text 'Security Check'. The main area has a light yellow background. It contains the text 'Can't read the words below? Try different words or an audio captcha.' followed by a rectangular box showing the word 'fromis' and the number '9' in a stylized, slightly distorted font. Below this box is the label 'Text in the box:' followed by an empty text input field. At the bottom right are two buttons: 'Submit' (blue) and 'Cancel' (white with a grey border).



A screenshot of a captcha interface. At the top is a rectangular image containing the word 'CAPTCHA' in a green, heavily distorted, and pixelated font. Below the image is the text 'Please enter the characters you see in the image above.' followed by an empty text input field. At the bottom right is a rounded button labeled 'Next'.

Défi audio

- x Analyser un son déformé/bruité est difficile pour un bot



Défi social

x Les humains ont une **vie sociale** (amis, famille),
pas les bots

 Log in

 Log in with Facebook

 Log in with Twitter

 Log in with Linked In

 Log in with Gmail

 Log in with Yahoo!

User Name 

Password 

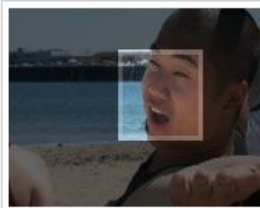
Enter Answer $5 + 9 =$

 Cancel

 Submit

Don't have an Account?

Identify This Friend 0 of 5 Friends Identified



This appears to be:

☐ Mike Steding

☒ Dan Muriello

☐ Ross Johnson

☐ Adam Ringel

☐ Matt Jones

☐ Ivan Lucuk

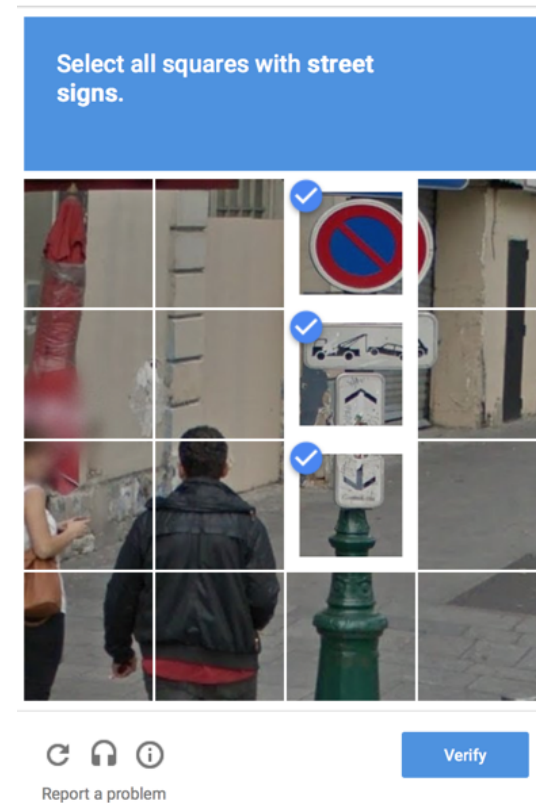
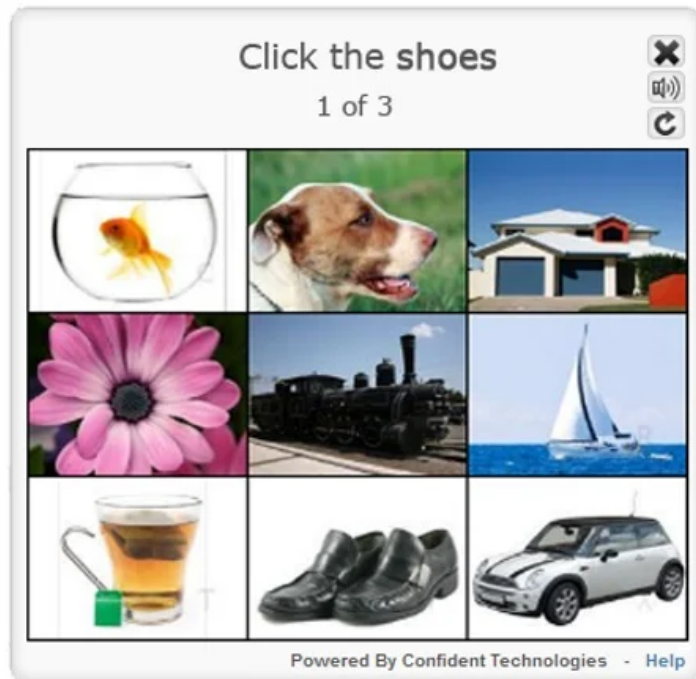
Refresh Photos

Submit

Skip Person

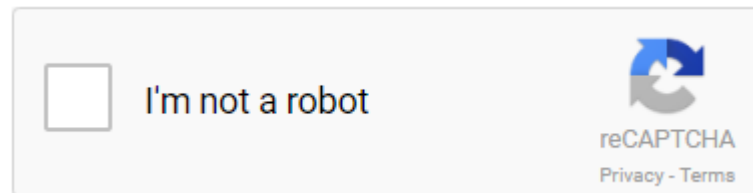
Défi image

x Comprendre le sens d'une image est difficile pour un bot



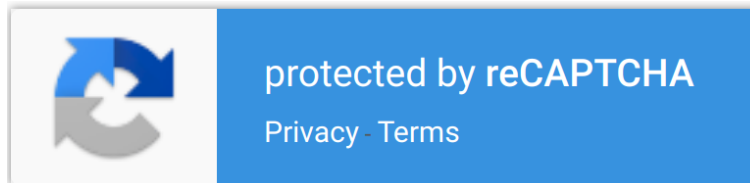
Défi coche

- x Un être humain déplace sa souris d'une certaine manière et ne clique jamais **parfaitement** au centre de la case à cocher



Défi invisible

- x La technique est tenue **secrète** mais apparemment elle se base sur les cookies et nécessite Javascript

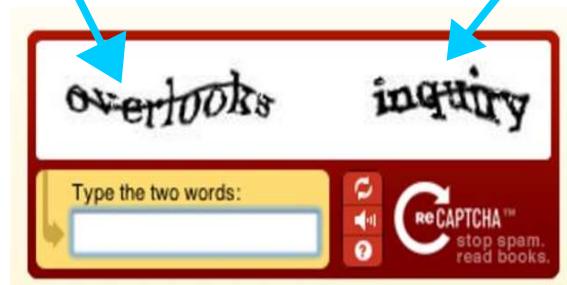


Et si on mettait à profit les défis ?

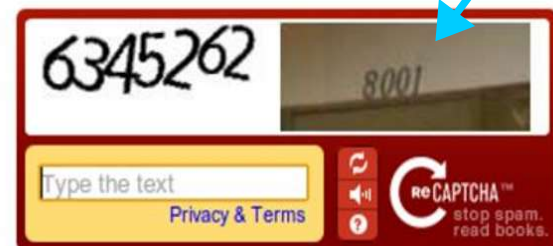
- x A l'échelle mondiale, le temps consacré par les humains à résoudre les CAPTCHAs est énorme
- x Avec reCAPTCHA, vous aidez Google à améliorer ses services là où l'OCR a échoué

Sert à vérifier que vous êtes humain

Sert à numériser un livre (*Google Book*)



Sert à numériser une adresse (*Google Street View*)



L'économie des CAPTCHAs

Top sites web utilisant reCAPTCHA

 youtube.com	2 TRAFFIC RANK	34B MONTHLY VISITS	>
 m.youtube.com		5.6B MONTHLY VISITS	>
 xvideos.com	7 TRAFFIC RANK	3.3B MONTHLY VISITS	>
 bit.ly		2.4B MONTHLY VISITS	>
 pornhub.com	13 TRAFFIC RANK	2.3B MONTHLY VISITS	>
 yahoo.co.jp	16 TRAFFIC RANK	2.3B MONTHLY VISITS	>
 zoom.us	23 TRAFFIC RANK	2.1B MONTHLY VISITS	>
 support.google.com		1.5B MONTHLY VISITS	>
 play.google.com		1.2B MONTHLY VISITS	>
 twitch.tv	38 TRAFFIC RANK	1.2B MONTHLY VISITS	>
3,809,095 additional websites are using reCAPTCHA			>

Une course contre la montre

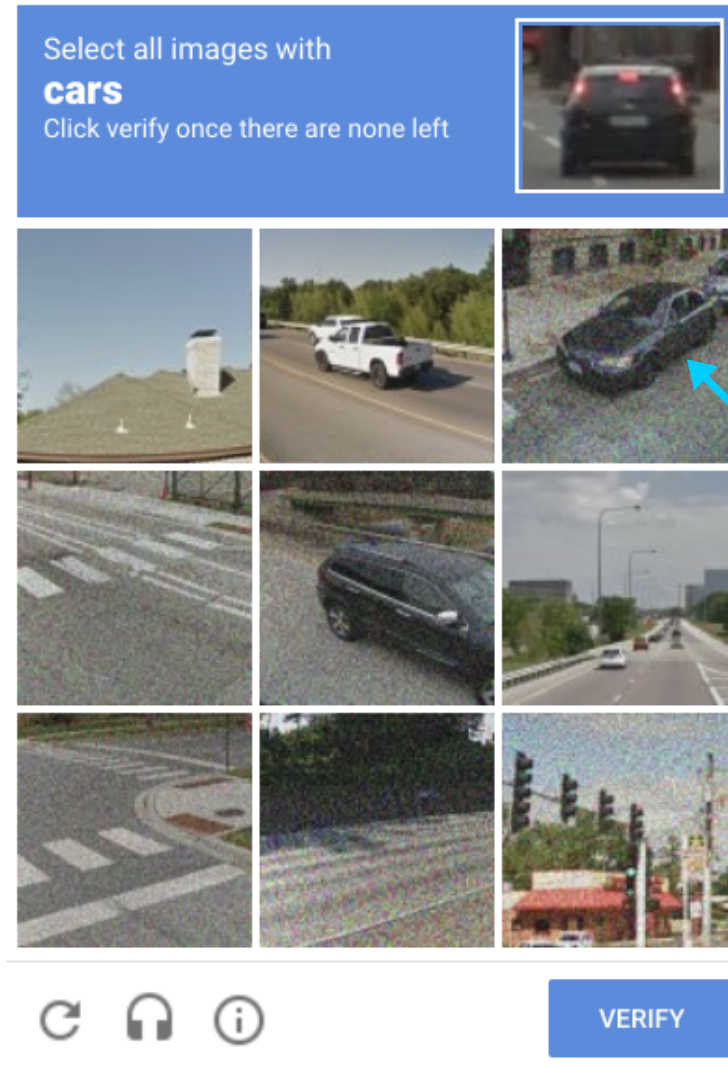
x Avec les progrès de l'IA (Deep learning), les bots réussissent à passer les tests !

- Les défis maths et défis texte ont cédé les premiers
- Les défis images/audio ne vont pas tenir éternellement

x Les CAPTCHAs doivent trouver des parades

- Induire les bots en erreurs
- Développer de nouvelles techniques

Induire les bots en erreur



x Adversarial CAPTCHA

- **Perturbations** ajoutées volontairement pour faire dérailler les algorithmes de reconnaissance d'image

Reconnue par certaines IA comme « nature » au lieu de « voiture »

Nouvelles techniques

- x Détection comportementale par **apprentissage** (la navigation, les cookies, l'historique, ainsi que d'autres signaux faibles)
- x Un test de Turing imaginé pour être échoué par l'humain et réussi par le bot (car l'« **erreur est humaine** ») - Brevet Amazon 2017
- x Dans un futur proche, le site web fera appel au **capteur biométrique** pour pouvoir poursuivre une action

Services de résolution de CAPTCHA

x Plateformes spécialisées

- Le bot tombe sur un captcha, l'envoi à une plateforme anti-CAPTCHA où un être humain est affecté à sa résolution. Le résultat est renvoyé au bot qui peut alors poursuivre son action

x Les nouveaux **forçats du travail**

- Illégal dans certains pays, dans d'autres non
- Exemple : 1,80€ versés pour 1000 défis-texte résolus

Démonstration : FloodBot

avec l'aimable autorisation de la proviseure du Lycée Jules Supervielle (64)

Contactez le Lycée via le site web

AC-BORDEAUX.fr


Jules Supervielle

[Accueil](#)
[Présentation du lycée](#)
[Informations pratiques](#)
[Les disciplines](#)

• Accès restreint



Site  collège

Recherche
recherche avancée

Vous êtes ici : Accueil

vous adresse mail

vous message

envoyer le message



Lycée Jules Supervielle - 6, avenue François Mitterrand - 64 400 Oloron Sainte Marie - [Mentions légales](#)

Création du FloodBot

x **Inonder** la proviseure adjointe de messages pour nuire : le « mail bombing »

- Exemple : 100 messages par jour, tous les jours
- Chercher à désorganiser le travail

x **Connaissances nécessaires**

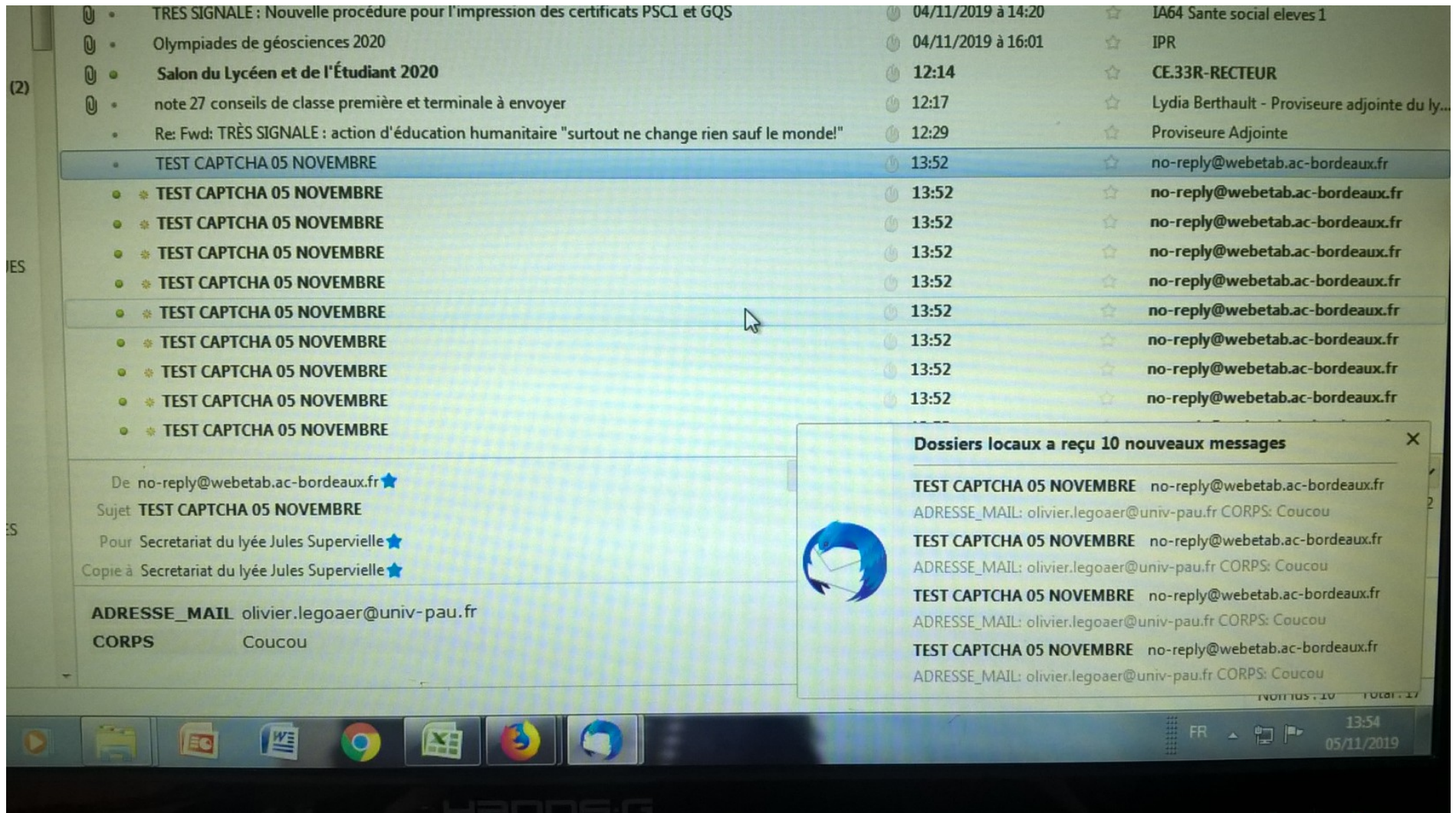
- Langage de programmation (Python, Java, ...)
- Technologies du Web (Html, HTTP, ...)

x Il existe même des **kits** pour créer ses bots !

Code Python du FloodBot

[illegible]

BAL secrétariat du Lycée (preuve)



Note : configuré pour seulement 10 messages...

Le mot de la fin

LeboncoinBot

- x En 2018, j'ai créé un bot d'analyse statistique des prix basé sur le site Web leboncoin.fr
 - [Article de presse](#) (La République des Pyrénées)
 - Le bot faisait gagner du temps aux acheteurs/vendeurs en déterminant automatiquement le juste prix d'un objet d'occasion
- x Depuis Août 2019, leboncoin se protège avec un système de CAPTCHA :(

La filière IA

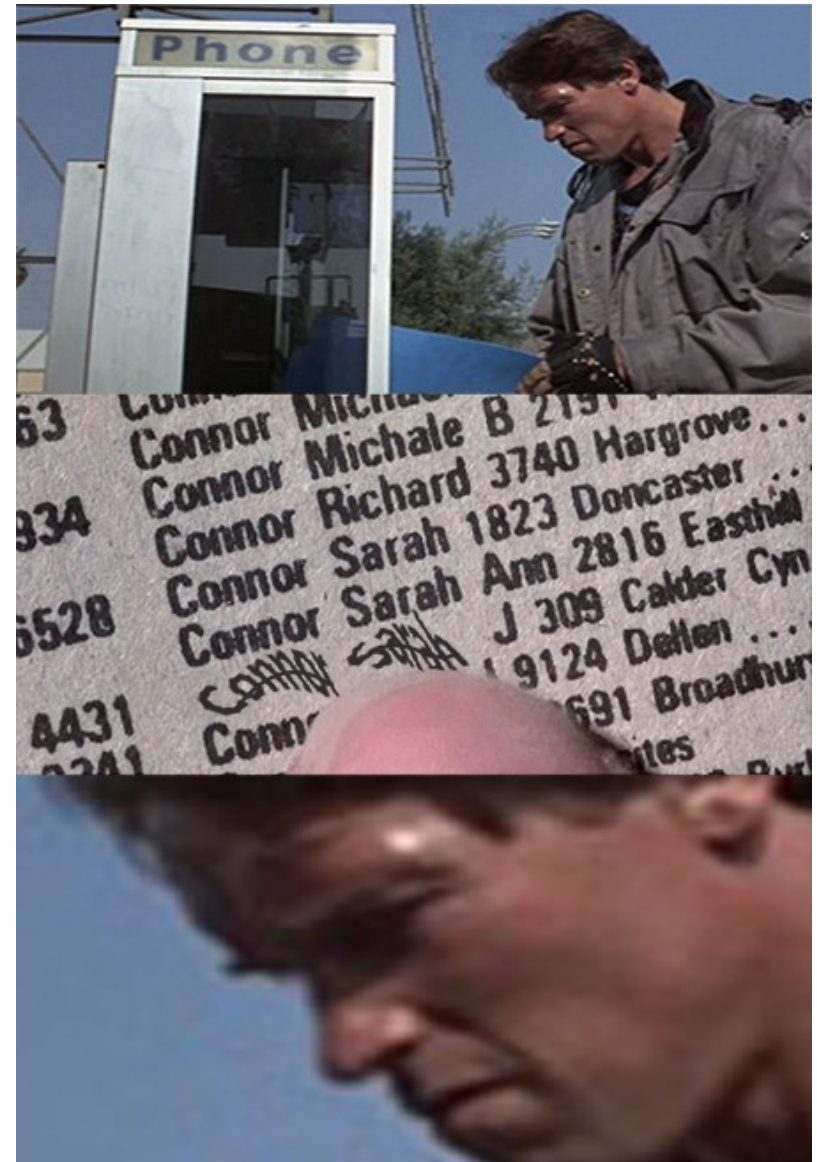
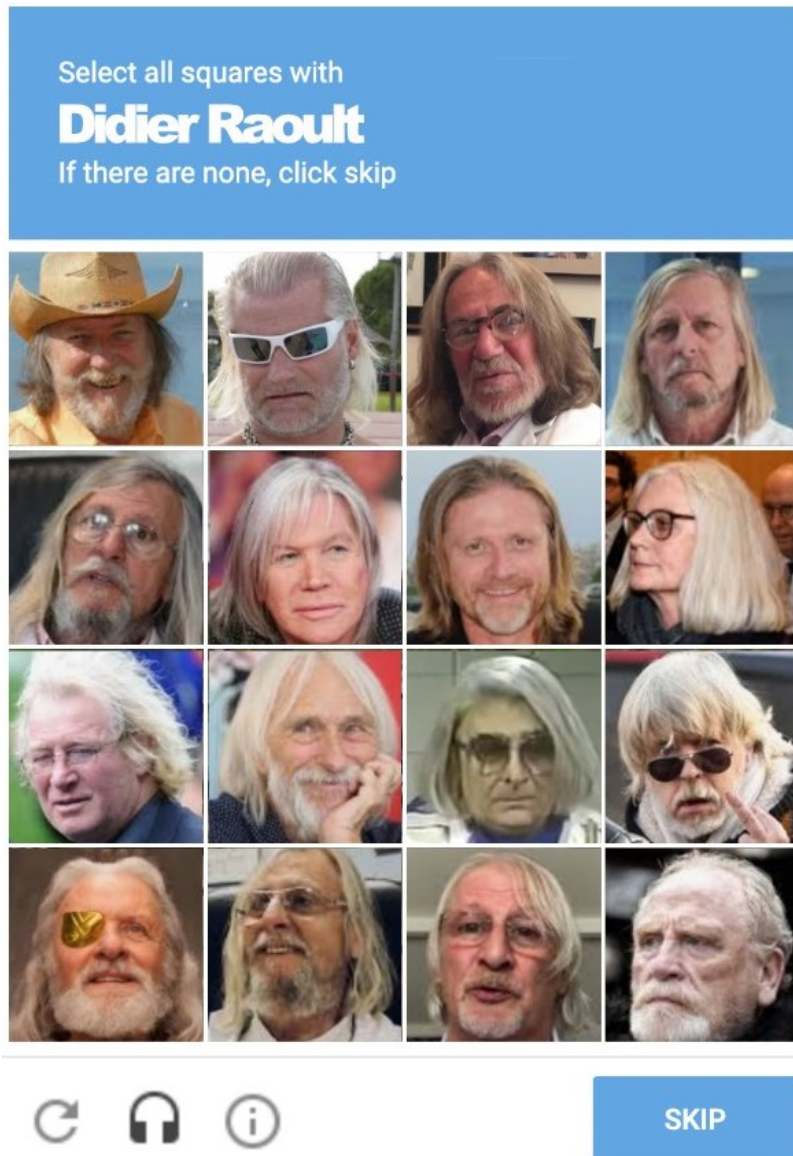
- x Ceux qui fabriquent les CAPTCHAs travaillent dans l'IA. Ceux qui cassent les CAPTCHA aussi
- x IA = Maths/Stats + Informatique
- x Choisir sa filière post-bac en connaissance de cause

Adversarial Patch



« Toaster »

Un peu d'humour...



MERCI

Des questions ?

Olivier Le Goaër

Entrée Libre #2

Centre des Abeilles, 28 Juillet 2021

