IA fiable : du code à la confiance publique

Lê Nguyên Hoang

École Polytechnique, Décembre 2024

Lequel craignez-vous le plus ?

Section 1

Le désastre informationnel

Un désastre de l'information sur le climat

Le "numérique responsable", du greenwahsing ?

Un désastre de santé mentale

HUFFPOST

Life 24/10/2024 11:30

Intelligence artificielle : un adolescent américain s'est suicidé après être tombé amoureux d'un chatbot

Sewell, 14 ans, a mis fin à ses jours après avoir développé des sentiments pour un personnage virtuel avec qui il discutait depuis des mois. Sa mère a porté plainte contre l'entreprise d'intelligence artificielle.

Par Charlotte Arce

Partager

CAVAN IMAGES / GETTY IMAGES/CAVAN IMAGES RF

La mère de l'adolescent a déposé une plainte pour négligence et mise en danger de la vie d'autrui contre l'entreprise Character.AI.

Actualités Élections US Économie Vidéos Débats Culture Le Goût du Monde Services S'abonner

PIXELS · CYBERHARCÈLEMENT

Le nombre de « deepfakes » pornographiques explose du fait des progrès de l'intelligence artificielle

Par Aurélien Defer

Publié le 05 décembre 2023 à 20h00, modifié le 06 décembre 2023 à 13h35

Lecture 6 min.

Article réservé aux abonnés · Offrir l'article

DÉCRYPTAGE | **En France, un article de loi prévoit de condamner spécifiquement ceux qui diffusent ces contenus considérés comme des cyberviolences sexistes.**

Taylor Klein s'apprêtait à achever ses études d'ingénieure dans le Connecticut, aux Etats-Unis, lorsqu'elle a appris que de fausses vidéos pornographiques d'elle circulaient sur Internet. « *Ces vidéos avaient des milliers de vues. En allant me coucher ce soir-là, j'ai craqué. Je me sentais très sale. (…) J'avais l'impression que quelqu'un essayait de me punir* », témoigne la jeune femme dans le film documentaire *Another Body*, sorti à la fin du mois d'octobre aux Etats-Unis.

Les réalisateurs Sophie Compton et Reuben Hamlyn y suivent pendant plus d'une heure le combat et la quête de vérité de cette jeune Américaine, dont la vie a été bouleversée par des *deepfakes* pornographiques. Aussi appelées « hypertrucages » en français, ces

Voir le livre “The Anxious Generation” de Jonathan Haidt.

Une perte de souveraineté démocratique sur le cyberespace

INDEPENDENT

US election › Subscribe Menu

NEWS SPORT VOICES CULTURE LIFESTYLE INDYBEST TRAVEL

News > World > Americas > US politics

JD Vance says US could drop support for NATO if Europe tries to regulate Elon Musk's platforms

Republican vice presidential nominee says 'Germans and other nations' – not Russia – would 'have to fund Ukraine's reconstruction'

Gustaf Kilander Washington DC

• Tuesday 17 September 2024 22:39 BST • Comments

Related video: Alina Habba says Trump was 'risking his life' by golfing

JD Vance has suggested that American support for NATO should be predicated on the European Union not regulating Elon Musk and his X social media platform, formerly known as Twitter.

The Republican vice presidential nominee and Ohio senator claimed in an interview with YouTuber Shawn Ryan that a top EU official had threatened to arrest the billionaire if he allowed former President Donald Trump back on X.

Meta blocked news from Facebook and Instagram in Canada — could they do the same in Australia?

By Brianna Morris-Grant

News and Magazine Publishing Industry

Wed 13 Mar

Meta has banned news outlets from its social media platforms in Canada. *(Reuters: Dado Ruvic/Illustration)*

abc.net.au/news/could-meta-block-news-in-australia-a... Share article

Sans parler de Microsoft, Google, Amazon, Huawei, TikTok, ...

Le déclin démocratique

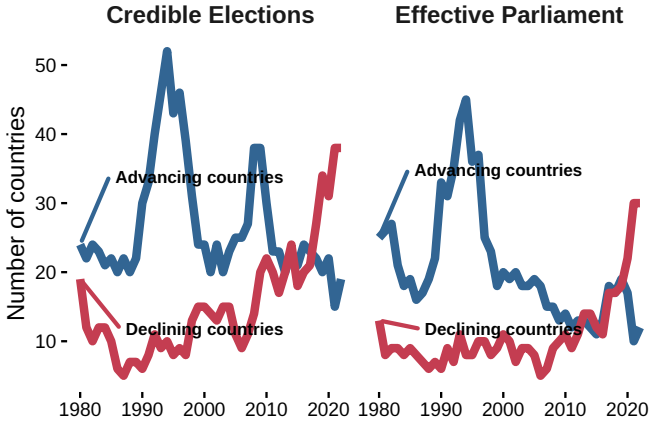


Image de l'article "Global Patterns" du *Global State of Democracy Initiative*.

Une urgence

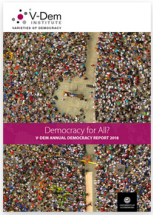
Previous Reports

DR 2022: Autocratization Changing Nature?

DR 2021: Autocratization Turns Viral

DR 2020: Autocratization Surges - Resistance Grows

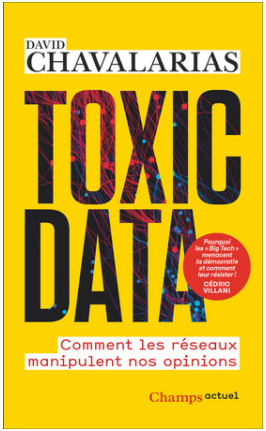
DR 2019: Democracy Facing Global Challenges

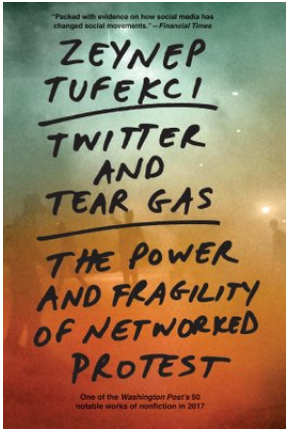


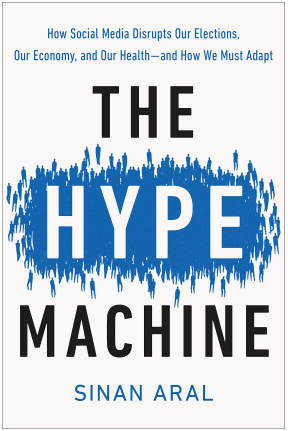
DR 2018: Democracy for All?

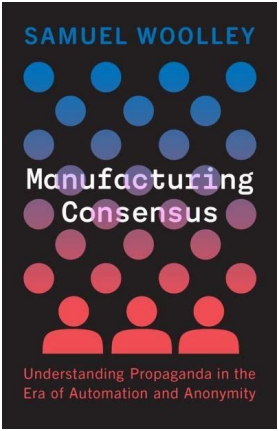
DR 2017: Democracy at Dusk?

Les IA au coeur du problème ?





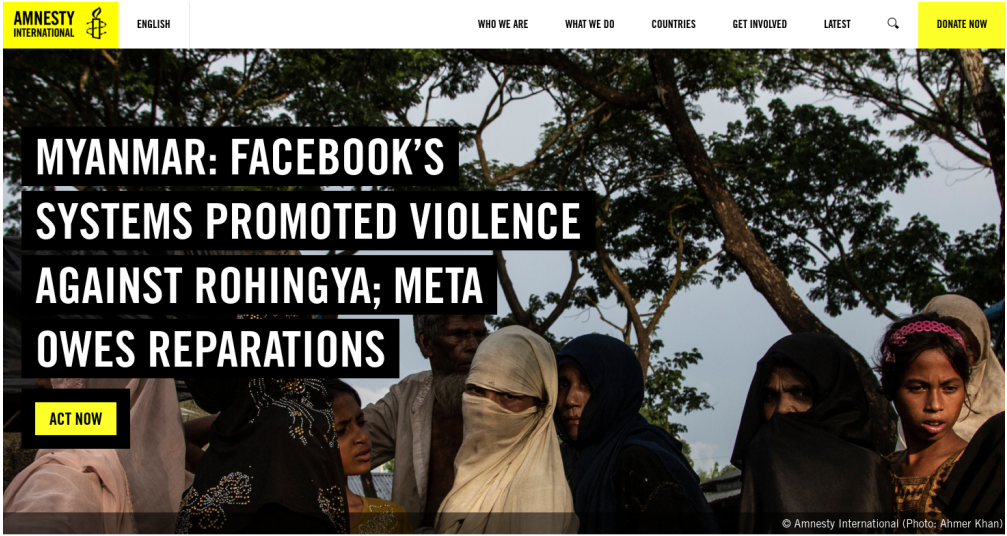




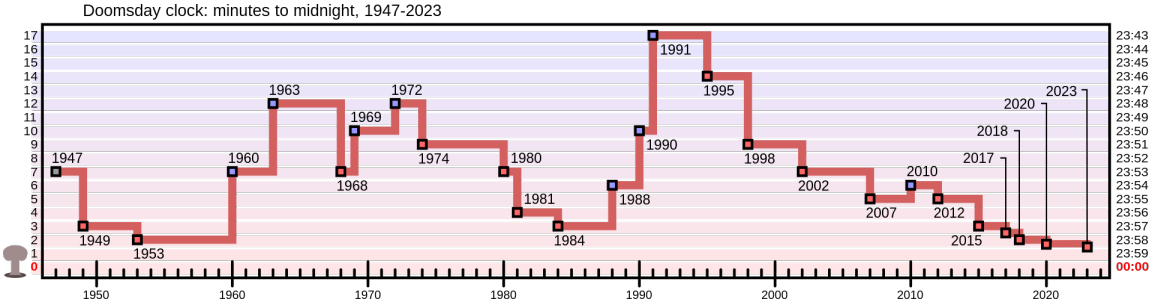
Plus lucratives et dangereuses que ChatGPT

Pixabay image by LolaSandoval1.

Les victimes s'empilent déjà



La “doomsday clock” se rapproche de minuit



La cybersécurité : un désastre ignoré

CYBER SAFETY REVIEW BOARD

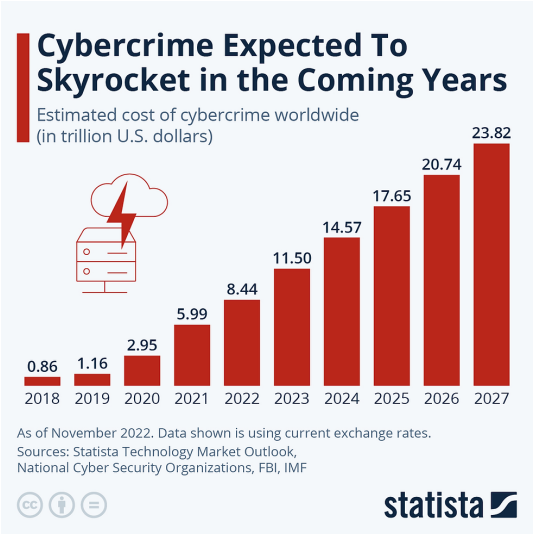
Review of the Summer 2023 Microsoft Exchange Online Intrusion

March 20, 2024
Cyber Safety Review Board

The Board is also concerned with Microsoft's public communications after the incident. In its September 6, 2023 blog post entitled "*Results of Major Technical Investigations for Storm-0558 Key Acquisition*," Microsoft explained that Storm-0558 likely stole the 2016 MSA key in the "crash dump" scenario described above. However, soon after publishing that blog, Microsoft determined it did not have any evidence showing that the crash dump contained the 2016 MSA key. This led Microsoft to assess that the crash dump theory was no longer any more probable than other theories as the mechanism by which the actor had acquired the key, which Microsoft chose to leave uncorrected for more than six months after publishing its September 6 blog.

The Board is troubled that Microsoft neglected to publicly correct this known error for many months. Customers (private sector and government) relied on these public representations in Microsoft's blogs. The loss of a signing key is a serious problem, but the loss of a signing key through unknown means is far more significant because it means that the victim company does not know how its systems were infiltrated and whether the relevant vulnerabilities have been closed off. Left with the mistaken impression that Microsoft has conclusively identified the root cause of this incident, Microsoft's customers did not have essential facts needed to make their own risk assessments about the security of Microsoft cloud environments in the wake of this intrusion. Microsoft told the Board early in this review that it believed that the errors in the blog were "not material." The Board disagrees. After several written follow up questions from the Board regarding the blog, Microsoft informed the Board on March 5, 2024, that it would be updating the blog in the "near future." One week following this communication, and more than six months after its publication of the September 6 blog, Microsoft corrected its mistaken assertions through an addendum to the blog's existing webpage.

La courbe qui me terrifie le plus



De nouvelles armes de guerre

Belarus

This article is more than **2 years old**

‘Cyberpartisans’ hack Belarusian railway to disrupt Russian buildup

Activists claim they could paralyse trains moving Russian forces for potential attack on Ukraine

A train carrying Russian military hardware at a railway station in Belarus. Photograph: Russian Defence Ministry/Tass

Andrew Roth *in Moscow*

Tue 25 Jan 2022 18.54 CET

0
comparisons

0
contributors

Wendover Productions

The Terrifying Efficiency of Drone Warfare

15

9
comparisons

2
contributors

WIRED

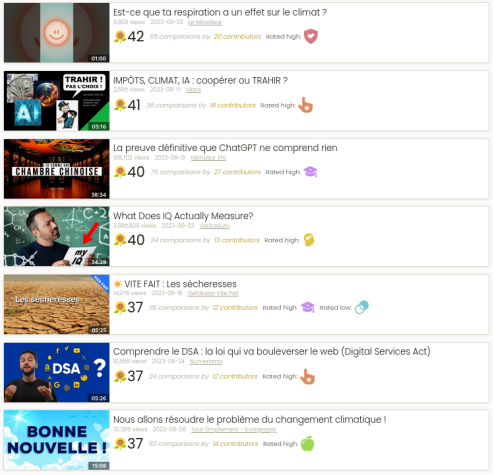
Hackers Remotely Kill a Jeep on a Highw...

Convaincus ? Lequel craignez-vous le plus ?

Section 2

Comment fonctionne l'intelligence artificielle ?

Lequel de ces problèmes est le plus dur ?



Recommandation de contenus VS échecs

Différentes mesure de complexité

Complexité en temps de calcul

Requiert beaucoup de calculs pour être résolu.
L'optimisation et la cryptanalyse sont de ce type.

Différentes mesure de complexité

Complexité en temps de calcul

Requiert beaucoup de calculs pour être résolu.
L'optimisation et la cryptanalyse sont de ce type.

Complexité en lignes de code (Kolmogorov-Solomonoff)

Requiert un grand nombre de lignes de code pour être résolu.
Le profilage psychologique, l'analyse d'images et la génération de texte sont de ce type.
Un humain ne peut plus écrire les algorithmes qui résolvent ces problèmes.

Différentes mesure de complexité

Complexité en temps de calcul

Requiert beaucoup de calculs pour être résolu.
L'optimisation et la cryptanalyse sont de ce type.

Complexité en lignes de code (Kolmogorov-Solomonoff)

Requiert un grand nombre de lignes de code pour être résolu.
Le profilage psychologique, l'analyse d'images et la génération de texte sont de ce type.
Un humain ne peut plus écrire les algorithmes qui résolvent ces problèmes.

Theorem (Minimax, John von Neumann 1928)

Les échecs, le go et tout autre jeu à deux joueurs à information complète est trivial au sens Kolmogorov-Solomonoff. Mais pas la recommandation (ni le jeu de l'imitation).

L'article révolutionnaire de Turing (1950)

Turing 1950

"C'est surtout un problème de programmation... Les estimations de la capacité de stockage du cerveau humain varient entre 10^{10} et 10^{15} bits ... Je serais surpris si plus de 10^9 bits étaient nécessaires pour passer le jeu de l'imitation de manière satisfaisante ... Note: La capacité de l'Encyclopaedia Britannica, 11e édition, est 2×10^9."

L'article révolutionnaire de Turing (1950)

Turing 1950

"C'est surtout un problème de programmation... Les estimations de la capacité de stockage du cerveau humain varient entre 10^{10} et 10^{15} bits ... Je serais surpris si plus de 10^9 bits étaient nécessaires pour passer le jeu de l'imitation de manière satisfaisante ... Note: La capacité de l'Encyclopaedia Britannica, 11e édition, est 2×10^9."

En langage moderne

Le jeu de l'imitation a une complexité en lignes de code de l'ordre de 10^9 bits.

L'article révolutionnaire de Turing (1950)

Turing 1950

"C'est surtout un problème de programmation... Les estimations de la capacité de stockage du cerveau humain varient entre 10^{10} et 10^{15} bits ... Je serais surpris si plus de 10^9 bits étaient nécessaires pour passer le jeu de l'imitation de manière satisfaisante ... Note: La capacité de l'Encyclopaedia Britannica, 11e édition, est 2×10^9."

En langage moderne

Le jeu de l'imitation a une complexité en lignes de code de l'ordre de 10^9 bits.

En langage courant

Aucun programme composé de moins d'un milliard de lignes de code ne peut dialoguer comme un humain.

L'invention des "learning machines", avant celle du mot "IA" !

Instead of trying to produce a programme to simulate the adult mind, why not rather try to produce one which simulates the child's? If this were then subjected to an appropriate course of education one would obtain the adult brain. Presumably the child-brain is something like a note-book as one buys it from the stationers. Rather little mechanism, and lots of blank sheets. (Mechanism and writing are from our point of view almost synonymous.) Our hope is that there is so little mechanism in the child-brain that something like it can be easily programmed. The amount of work in the education we can assume, as a first approximation, to be much the same as for the human child.

Section 3

Les systèmes auto-apprenants **apprennent** : Les conséquences sur leur cybersécurité

Les hypothèses dangereuses de la recherche

L'hypothèse irréaliste la plus commune

Soit $x_1, x_2, \ldots x_n$ des données indépendantes et identiquement distribuées...

Les hypothèses dangereuses de la recherche

L'hypothèse irréaliste la plus commune

Soit $x_1, x_2, \ldots x_n$ des données indépendantes et identiquement distribuées...

L'hypothèse extrêmement politisée devenue banalisée

Nous apprenons une fonction f qui généralise les données...

Les hypothèses dangereuses de la recherche

L'hypothèse irréaliste la plus commune

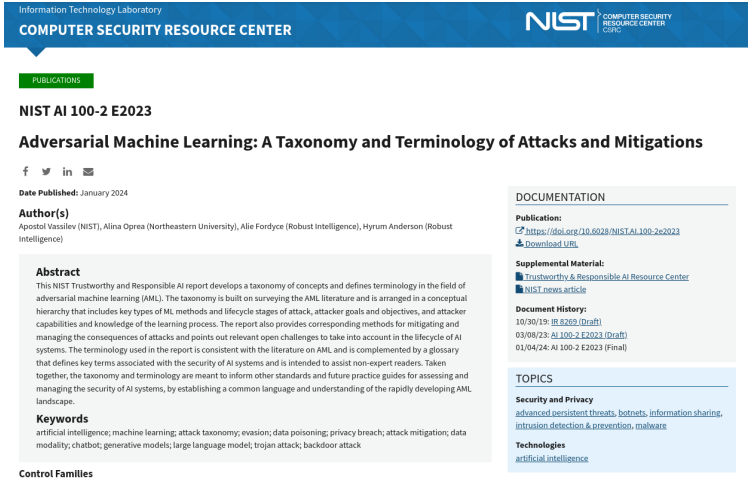
Soit $x_1, x_2, \ldots x_n$ des données indépendantes et identiquement distribuées...

L'hypothèse extrêmement politisée devenue banalisée

Nous apprenons une fonction f qui généralise les données...

Ces hypothèses sont extrêmement hackables !

Les agences gouvernementales s'attaquent (enfin!) au sujet



Information Technology Laboratory

COMPUTER SECURITY RESOURCE CENTER

NIST COMPUTER SECURITY RESOURCE CENTER CSRC

PUBLICATIONS

NIST AI 100-2 E2023

Adversarial Machine Learning: A Taxonomy and Terminology of Attacks and Mitigations

Date Published: January 2024

Author(s)

Apostol Vassilev (NIST), Alina Oprea (Northeastern University), Alie Fordyce (Robust Intelligence), Hyrum Anderson (Robust Intelligence)

Abstract

This NIST Trustworthy and Responsible AI report develops a taxonomy of concepts and defines terminology in the field of adversarial machine learning (AML). The taxonomy is built on surveying the AML literature and is arranged in a conceptual hierarchy that includes key types of ML methods and lifecycle stages of attack, attacker goals and objectives, and attacker capabilities and knowledge of the learning process. The report also provides corresponding methods for mitigating and managing the consequences of attacks and points out relevant open challenges to take into account in the lifecycle of AI systems. The terminology used in the report is consistent with the literature on AML and is complemented by a glossary that defines key terms associated with the security of AI systems and is intended to assist non-expert readers. Taken together, the taxonomy and terminology are meant to inform other standards and future practice guides for assessing and managing the security of AI systems, by establishing a common language and understanding of the rapidly developing AML landscape.

Keywords

artificial intelligence; machine learning; attack taxonomy; evasion; data poisoning; privacy breach; attack mitigation; data modality; chatbot; generative models; large language model; trojan attack; backdoor attack

Control Families

DOCUMENTATION

Publication:

https://doi.org/10.6028/NIST.AI.100-2e2023

Download URL

Supplemental Material:

Trustworthy & Responsible AI Resource Center

NIST news article

Document History:

10/30/19: IR 8269 (Draft)

03/08/23: AI 100-2 E2023 (Draft)

01/04/24: AI 100-2 E2023 (Final)

TOPICS

Security and Privacy

advanced persistent threats, botnets, information sharing, intrusion detection & prevention, malware

Technologies

artificial intelligence

5 des 9 articles cités par le NIST comme "état de l'art contre l'empoisonnement" sont co-écrits par les co-fondateurs de mon entreprise Calicarpa.

Risques d'exfiltration

NEWS POLITICS U.S. NEWS WORLD BUSINESS TECH HEALTH CULTURE & TRENDS WATCH NOW

ARTIFICIAL INTELLIGENCE

Samsung bans use of AI-like ChatGPT for employees after misuse of the chatbot

Bloomberg reported some staff members had uploaded sensitive code to ChatGPT.

Samsung has told employees not to use AI tools after a report found that some staff members had uploaded sensitive code to ChatGPT.
SeongJoon Cho / Bloomberg via Getty Images file

AI February 4, 2023

Microsoft Warns Employees Not to Share Sensitive Data with ChatGPT

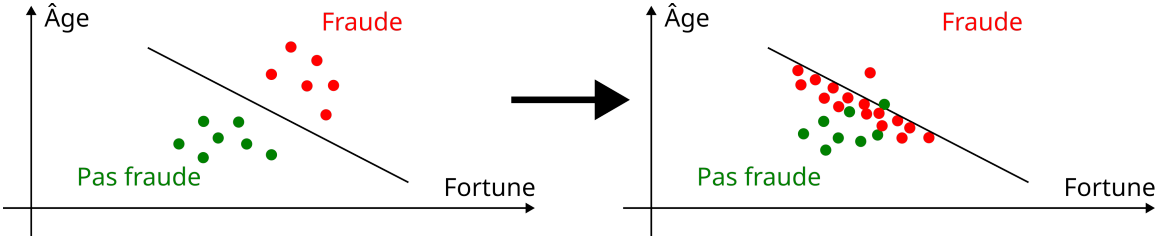
Microsoft has warned its employees not to share sensitive data with an artificially intelligent (AI) chatbot, ChatGPT from OpenAI. Employees of American multinational tech giants had asked in an internal forum whether ChatGPT or any other AI tools from OpenAI were appropriate to use at their work, Business Insider reported.

Also read: 30% of College Students Use ChatGPT

In response to that inquiry, a senior engineer from Microsoft's CTO office allowed to use ChatGPT but couldn't share confidential information with the AI chatbot.

"Please don't send sensitive data to an OpenAI endpoint, as they may use it for training future models," the senior engineer wrote in an internal post, per Insider.

Évasion (jailbreaking)



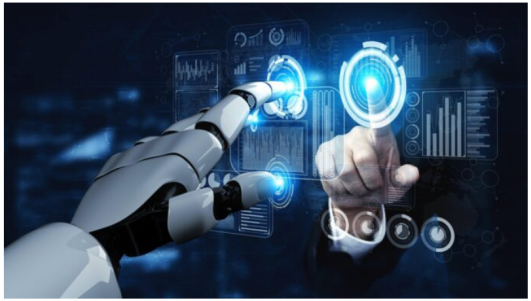
Empoisonnement des modèles

ARTIFICIAL INTELLIGENCE

Meta's AI Is Partially Trained on Breitbart and Russia Today, Study Finds

An analysis of Google and Meta's C4 dataset shows the data included text from right-wing and white supremacist sites.

By **Kyle Barr** Published April 19, 2023 | Comments (0)



As much as people think AI is 'intelligent,' what goes into training our modern AI models informs what kind of information it puts out. Image: Blue Planet Studio (Shutterstock)

BEST OF 2024 AWARDS

Gizmodo's Best of 2024 Awards →

Quizz

Combien de faux comptes
Facebook supprime chaque année ?

Facebook Removed More than 15 Billion Fake Accounts in Two Years, Five Times more than its Active User Base

Jastra Kranjec · Pro Investor
Updated: 27 September 2021

Disclosure

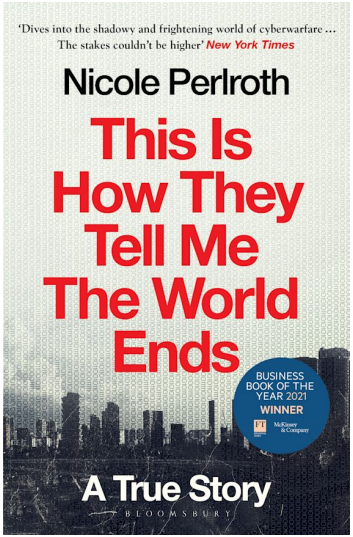
As the world's largest social networking platform, Facebook has witnessed a surge in the number of users in the past few years. Hundreds of millions of people have joined its social media space to communicate, keep in touch with the latest trends or promote business, especially after the pandemic hit. Although the COVID-19 restrictions have loosened in most countries, Facebook's active user base continues growing, but so does the number of fake accounts.

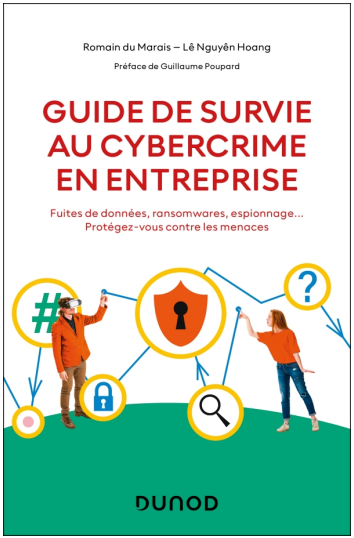
According to data presented by Stock Apps, the social media giant removed over 15 billion fake accounts in the last two years, five times more than its active user base.

3 Billion Fake Accounts Removed in the First Half of 2021, 20x More than the Number of New Active Users

Scammers use fake Facebook accounts to connect with users, get their personal information and steal identities. Most of them will reach out to anyone who's accepted their friend request to try and scam them out of money.

N'oubliez pas la cybersécurité classique !





Section 4

L'état de l'art en stratégies d'atténuation

Ne laissez pas les données sortir de chez vous

- Modèles open sources locaux.

Ne laissez pas les données sortir de chez vous

- Modèles open sources locaux.
- Sandboxing des algorithmes auto-apprenants.

Ne laissez pas les données sortir de chez vous

- Modèles open sources locaux.
- Sandboxing des algorithmes auto-apprenants.
- Confidentialité différentielle.

Ne laissez pas les données sortir de chez vous

- Modèles open sources locaux.
- Sandboxing des algorithmes auto-apprenants.
- Confidentialité différentielle.
- Calcul multi-partite.

Quizz

Comment 10 traders dans une pièce peuvent-ils connaître la moyenne de leurs salaires, sans que quiconque ne révèle son salaire ?

Évasion à la détection

- Apprentissage adversarial.

Évasion à la détection

- Apprentissage adversarial.
- Détection de données hors-distribution.

Évasion à la détection

- Apprentissage adversarial.
- Détection de données hors-distribution.
- Analyse de glissement distributionnel.

Évasion à la détection

- Apprentissage adversarial.
- Détection de données hors-distribution.
- Analyse de glissement distributionnel.
- Recours pour les faux négatifs/positifs.

Peut-on paralyser les drones tueurs en exploitant leurs faiblesses ?

A Survey on Security of UAV Swarm Networks: Attacks and Countermeasures

XIAOJIE WANG, School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
ZHONGHUI ZHAO, School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
LING YI*, School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
ZHAOLONG NING*, School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
LEI GUO, School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, China
F. RICHARD YU, Carleton University, Ottawa, Canada
SONG GUO, CSE, The Hong Kong University of Science and Technology, Kowloon, Hong Kong

The increasing popularity of **Unmanned Aerial Vehicle (UAV)** swarms is attributed to their ability to generate substantial returns for various industries at a low cost. Additionally, in the future landscape of wireless networks, UAV swarms can serve as airborne base stations, alleviating the scarcity of communication resources. However, UAV swarm networks are vulnerable to various security threats that attackers can exploit with unpredictable consequences. Against this background, this paper provides a comprehensive review on security of UAV swarm networks. We begin by briefly introducing the dominant UAV swarm technologies, followed by their civilian and military applications. We then present and categorize various potential attacks that UAV swarm networks may encounter, such as denial-of-service attacks, man-in-the-middle attacks and attacks against **Machine Learning (ML)** models. After that, we introduce security technologies that can be utilized to address these attacks, including cryptography, physical layer security techniques, blockchain, ML, and intrusion detection. Additionally, we investigate and summarize mitigation strategies addressing different security threats in UAV swarm networks. Finally, some research directions and challenges are discussed.

This work was supported by the National Science Foundation of China (62025105, 62272075, 62403092), by the National Natural Science Foundation of Chongqing (CSTB2023NSCQ-LZX0013), by the Science and Technology Research Program for Chongqing Municipal Education Commission (KJZD-M202200601, KJZD-K202401604), by the Hong Kong RGC Research Impact Fund (No. R5011-23F, No. R5060-19, No. R5034-18), and by the Collaborative Research Fund (No. C1042-23GF).
Authors' Contact Information: Xiaojie Wang, wangxiaojie@cqupt.edu.cn, Zhonghui Zhao, S220101211@stu.cqupt.edu.cn, Ling Yi (corresponding author), yiling@cqupt.edu.cn, Zhaolong Ning (corresponding author), ningzl@cqupt.edu.cn, and Lei Guo, guolei@cqupt.edu.cn, School of Communications and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing, 400065, China; Fei Richard Yu, richard.yu@carleton.ca, Department of Systems and Computer Engineering, Carleton University, Ottawa, ON K1S 5B6, Canada; Song Guo, songguo@cse.ust.hk, Department of Computer Science and Engineering, The Hong Kong University of Science and Technology, Hong Kong, China.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM 1557-7341/2024/11-ART
https://doi.org/10.1145/3703625

SUBMITTED

ELSEVIER
Reliability Engineering & System Safety
Volume 254, Part A, February 2025, 110608

Toward the resilience of UAV swarms with percolation theory under attacks

Tianzhen Hu [a], Yan Zong [a], Ningyun Lu [a b] ✉, Bin Jiang [a b]

Show more

+ Add to Mendeley Share Cite

https://doi.org/10.1016/j.ress.2024.110608 Get rights and content

Highlights

- The influence of distance within communication range between UAVs is incorporated into the modelling of the UAV swarm.
- An attack model is proposed to perform various attacks on the UAV swarm.
- A percolation based resilience assessment is proposed for the UAV swarm.
- Numerical results are presented for the proposed swarm model under continuous attacks.

Empoisonnement des modèles

- Nettoyage des données.

Empoisonnement des modèles

- Nettoyage des données.
- Apprentissage par aggrégations résilientes.

Empoisonnement des modèles

- Nettoyage des données.
- Apprentissage par aggrégations résilientes.
- Authentification (cryptographique) des sources.

Empoisonnement des modèles

- Nettoyage des données.
- Apprentissage par aggrégations résilientes.
- Authentification (cryptographique) des sources.
- Réduction de la dimensionalité.

Une préconisation fondamentale : réduire le nombre de paramètres

On the Impossible Safety of Large AI Models

El-Mahdi El-Mhamdi[1,2], Sadegh Farhadkhani[3], Rachid Guerraoui[3], Nirupam Gupta[3], Lê-Nguyên Hoang[2,4], Rafaël Pinot[3], Sébastien Rouault[2], and John Stephan[3]

[1]École Polytechnique
[2]Calicarpa
[3]EPFL
[4]Tournesol Association

Abstract

Large AI Models (LAIMs), of which large language models are the most prominent recent example, showcase some impressive performance. However they have been empirically found to pose serious security issues. This paper systematizes our knowledge about the *fundamental impossibility* of building arbitrarily accurate and secure machine learning models. More precisely, we identify key challenging features of many of today's machine learning settings. Namely, high accuracy seems to require *memorizing* large training datasets, which are often *user-generated* and *highly heterogeneous*, with both *sensitive information* and *fake users*. We then survey statistical lower bounds that, we argue, constitute a compelling case against the possibility of designing high-accuracy LAIMs with strong security guarantees.

1 Introduction

In recent years, we have witnessed a race for developing larger and larger artificial intelligence (AI) models. Notable milestones in this trend are *Attention Networks* (213 million parameters) [VSP+17], *GPT-2* (1.5 billion parameters) [RWC+19], *GPT-3* (175 billion parameters) [BMR+20], *Switch Transformer* (1.6 trillion parameters) [FZS21], *Persia* (over 100 trillion parameters) [LYZ+21], and *GPT-4* (unknown number of parameters) [BCE+23]. The scaling of model sizes has shown improvement in the accuracies on classical tasks, such as GLUE [WSM+19], SuperGLUE [WPN+19] and Winograd [SBBC20], without significant diminishing returns so far (see, e.g., Figure 1 in [BMR+20]). Moreover large AI models (or LAIMs) can also be used as *few-shot learners* [BMR+20], which has motivated their wide use as pre-trained *base* (or *foundation*) models [CCM21, CLL21, JLZ22, VPKG21, ZWK+21]. This success has generated enormous academic, economic and political interests into the development and deployment of LAIMs in public domain applications including content moderation, recommendation, search and ad targeting [Dea21, Hc21].

Contrary to the conventional wisdom of probably approximately correct (PAC) learning [Val84], the performance of LAIMs has been empirically shown to be best achieved by fully *interpolating* the training data [BHMM19, NKB+20, ZBH+17]. Put differently, the best accuracy is reached when these models *memorize* their training data [Fel20]. This phenomenon has also been theoretically supported to a certain extent by a recent line of work [BHX20, BMM18, BRT19, ISS+20, HY21, HG21, LL521, MM19, MVSS20, NVKM21]. Furthermore, training LAIMs requires access

The Poison of Dimensionality

Anonymous Authors[1]

Abstract

This paper advances the understanding of how the size of a machine learning model affects its vulnerability to poisoning, despite the use of state-of-the-art defenses. Given isotropic random honest feature vectors and the geometric median as the robust gradient aggregator rule, we essentially prove that, perhaps surprisingly, linear and logistic regressions with $D \geq 169H^2/P^2$ parameters are subject to *arbitrary model manipulation* by poisoners, where H and P are the numbers of honestly labeled and poisoned data points used for training. Our experiments go on exposing a fundamental tradeoff between augmenting model expressivity and increasing the poisoners' *attack surface*. We also informally discuss potential implications for "sandboxed learning", neural networks and non-zero-sum targeted poisoning.

1. Introduction

The classical theory of learning (Valiant, 1984; Geman et al., 1992; Kohavi & Wolpert, 1996) suggests that, given N training data, learning models should have $D = \Theta(N)$ parameters. But a vast empirical and theoretical literature on the *double descent* phenomenon (Zhang et al., 2017; Belkin et al., 2019; Muthukumar et al., 2019; Nakkiran et al., 2020; Mei & Montanari, 2022; Hastie et al., 2022) instead suggests that better performance could be obtained by letting $D \to \infty$. In any case, massive data collection has led to ever larger learning models (Brown et al., 2020; Fedus et al., 2022; Lian et al., 2022; Chowdhery et al., 2023).

However, these theories arguing for $D \geq \Omega(N)$ all assume that all training data are "honest" and should be generalized. In large-scale high-risk applications like language processing and content recommendation, this is deeply *unrealistic* and *ethically questionable* (Kallus & Zhou, 2018; Bender et al., 2021), if not illegal (Sag, 2023; Samuelson, 2023). After all, many of these systems fit massive web-crawled datasets (Smith et al., 2013; Chowdhery et al., 2023), which are heavily *poisoned* by doxed personal data, hate speech and state-sponsored propaganda (Woolley, 2023; Yurieff, 2019; Andrzejewski, 2023). In fact, such *data poisoning*, i.e. injections of misleading inputs in training datasets (Biggio et al., 2012; Suya et al., 2021), has become the leading AI security concern in the industry (Kumar et al., 2020).

Meanwhile, a growing line of research has been suggesting that high-dimensional training facilitates persistent poisoning attacks (Hubinger et al., 2024), even given state-of-the-art defenses (El-Mhamdi et al., 2022). The theoretical case has mostly relied on an mathematical impossibility to bring the norm of the gradient at termination below $\Omega(\sqrt{D})$. However, it is unclear that the performance of the poisoned model is then worse than if trained with fewer parameters.

Our paper advances the understanding of how model size D affects machine learning security, given H honestly labeled data and P poisoned data. Crucially, for $P = \Theta(H)$ (e.g. 1% of poisoned data), our results completely diverge from the common wisdom $D \geq \Omega(N) = \Omega(H)$. More precisely, we make the following contributions.

Contributions. First, when $D \geq 169H^2/P^2$, we essentially prove that using a state-of-the-art poisoning defense (gradient descent with the geometric median) actually provides *zero* resilience guarantee, even for the two most standard learning problems (linear and logistic regression). In fact, we prove *arbitrary model manipulation* by poisoners.

Second, we empirically show the value of *dimension reduction* under poisoning, for two other state-of-the-art poisoning defenses. Our experiments highlight a tradeoff between model expressivity and restricted *attack surface*.

Third, we prove and leverage a property of random vector subspaces to informally discuss the applicability of our analysis to "sandboxed learning" and nonlinear models.

Le poison de la dimensionalité

Considérons une régression linéaire avec H données $(x_1, y_1), \ldots, (x_H, y_H)$ honnêtes.

Le poison de la dimensionalité

Considérons une régression linéaire avec H données $(x_1, y_1), \ldots, (x_H, y_H)$ honnêtes.
Supposons les x_h isotropes, e.g. $x_h \sim \mathcal{N}(0, I_D)$, et les étiquettes correctes $y_h \sim \mathcal{N}(\beta^T x_h, \sigma^2)$.

Le poison de la dimensionalité

Considérons une régression linéaire avec H données $(x_1, y_1), \ldots, (x_H, y_H)$ honnêtes.
Supposons les x_h isotropes, e.g. $x_h \sim \mathcal{N}(0, I_D)$, et les étiquettes correctes $y_h \sim \mathcal{N}(\beta^T x_h, \sigma^2)$.
Utilisons la descente de gradient avec robustification par médiane géométrique (un algorithme d'apprentissage appartenant à l'état de l'art de l'IA sécurisée).

Le poison de la dimensionalité

Considérons une régression linéaire avec H données $(x_1, y_1), \ldots, (x_H, y_H)$ honnêtes.
Supposons les x_h isotropes, e.g. $x_h \sim \mathcal{N}(0, I_D)$, et les étiquettes correctes $y_h \sim \mathcal{N}(\beta^T x_h, \sigma^2)$.
Utilisons la descente de gradient avec robustification par médiane géométrique (un algorithme d'apprentissage appartenant à l'état de l'art de l'IA sécurisée).

Theorem (Hoang 2024, version informelle)

Supposons $D \geq 169H^2/P^2$. Alors, avec grande probabilité, il existe P données empoisonnées qui permettent une **manipulation arbitraire du modèle**.

Autre préconisation : l'apprentissage sandboxé

42

170
comparisons

48
contributors

Science4All

Spotify a payé leur silence

An Equivalence Between Data Poisoning and Byzantine Gradient Attacks

Sadegh Farhadkhani, Rachid Guerraoui, Lê Nguyên Hoang, Oscar Villemaud *Proceedings of the 39th International Conference on Machine Learning*, PMLR 162:6284-6323, 2022.

$$\textsc{Loss}(\rho, \vec{\theta}, \vec{\mathcal{D}}) \triangleq \sum_{n \in [N]} \mathcal{L}_n(\theta_n, \mathcal{D}_n) + \sum_{n \in [N]} \mathcal{R}(\rho, \theta_n).$$

- Réduire la surface d'attaque.

Fondamentaux de la cybersécurité

- Réduire la surface d'attaque.
- Cloisonnement des composants avec moindre privilège.

Fondamentaux de la cybersécurité

- Réduire la surface d'attaque.
- Cloisonnement des composants avec moindre privilège.
- Redondance et diversification des systèmes critiques.

Fondamentaux de la cybersécurité

- Réduire la surface d'attaque.
- Cloisonnement des composants avec moindre privilège.
- Redondance et diversification des systèmes critiques.
- Monitoring du système d'information.

Ce que **je** peux en faire

vs.

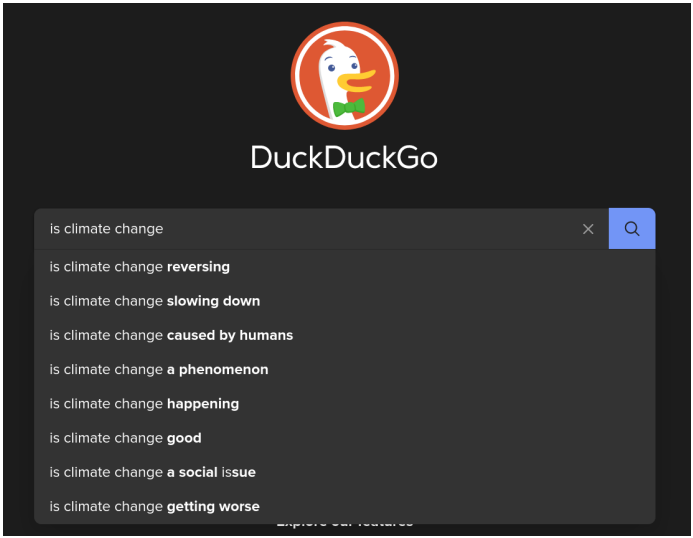
Ce que le **cybercrime** peut en faire

(et comment protéger la démocratie dans ce contexte...)

Section 5

Pour une IA fiable : L'enjeu de la gouvernance

Qui a décidé ? Et qui devrait décider ?



Le cyberespace, un zone de droit

PIXELS · VIE PRIVÉE

Meta condamné à une amende record de 1,2 milliard d'euros par le régulateur irlandais des données personnelles

Le groupe américain se voit reprocher le transfert de données d'Européens vers les Etats-Unis. L'amende est la plus lourde jamais imposée dans le cadre du droit européen sur les données.

Par Olivier Clairouin et Martin Untersinger
Publié le 22 mai 2023 à 11h42, modifié le 22 mai 2023 à 18h20 · Lecture 3 min.

Offrir l'article

La société Meta, maison mère de Facebook, a été condamnée à une amende record de 1,2 milliard d'euros par la Data Protection Commission (DPC), le régulateur irlandais de la vie privée. Une somme sans précédent à l'échelle de l'Union européenne, qui surpasse de loin celle que l'entreprise Amazon avait été condamnée à verser en juillet 2021, qui était à l'époque de 746 millions d'euros.

La DPC, équivalent irlandais de la Commission nationale de l'informatique et des libertés (CNIL) en France, reproche au réseau social d'avoir continué à transférer des données personnelles de ses clients européens vers les Etats-Unis. En 2020, la Cour de justice de l'union européenne (CJUE) avait estimé que la possibilité réservée aux services de sécurité américains de pouvoir accéder aux données des Européens était incompatible avec le droit de l'Union européenne en matière de protection des données.

Nick Clegg, responsable des affaires publiques de Meta, a jugé que cette sanction, *« injustifiée et inutile »*, *« établissait un dangereux précédent pour les nombreuses entreprises qui transfèrent des données entre les Etats-Unis et l'UE »*. Il a aussi annoncé faire appel de la décision.

Max Schrems, l'activiste à l'origine de l'arrêt de la CJUE, s'est dit

PUBLICITÉ


Édition du Jour
Daté du samedi 16 novembre
Le Monde WEEK-END
Allemagne : vers la fin du tabou de la dette publique

Commission opens formal proceedings against TikTok under the Digital Services Act

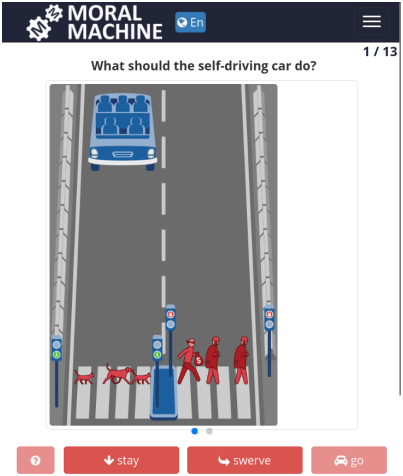
The European Commission has opened formal proceedings to assess whether TikTok may have breached the Digital Services Act (DSA) in areas linked to the protection of minors, advertising transparency, data access for researchers, as well as the risk management of addictive design and harmful content.

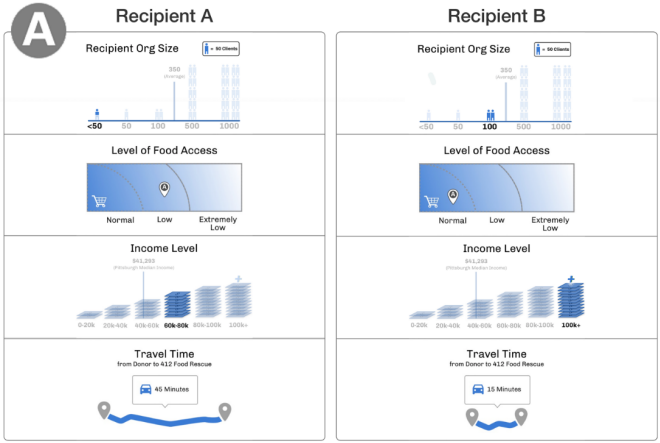
European Commission

On the basis of the preliminary investigation conducted so far, including on the basis of an analysis of the risk assessment report sent by TikTok in September 2023, as well as TikTok's replies to the Commission's formal Requests for Information (on illegal content (https://digital-strategy.ec.europa.eu/en/news/commission-sends-request-information-tiktok-under-digital-services-act), protection of minors (https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-tiktok-and-youtube-under-digital-services-act), and data access (https://digital-strategy.ec.europa.eu/en/news/commission-sends-requests-information-17-very-large-online-platforms-and-search-engines-under)), the Commission has decided to open formal proceedings against TikTok under the Digital Services Act.

- Full press release (https://ec.europa.eu/commission/presscorner/detail/en/IP_24_926)
- DSA: Making the online world safer (https://digital-strategy.ec.europa.eu/en/policies/safer-online)

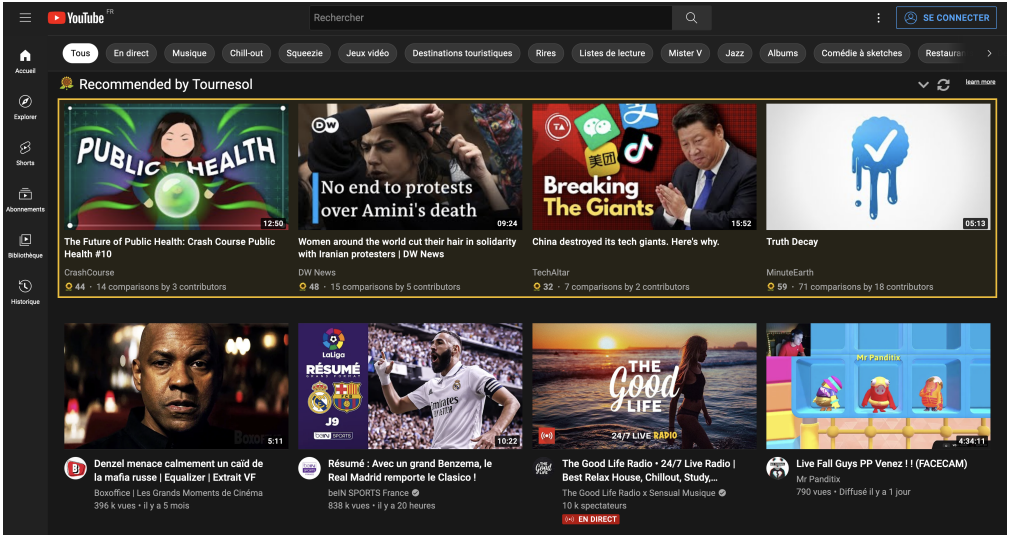
Gouverner directement les algorithmes



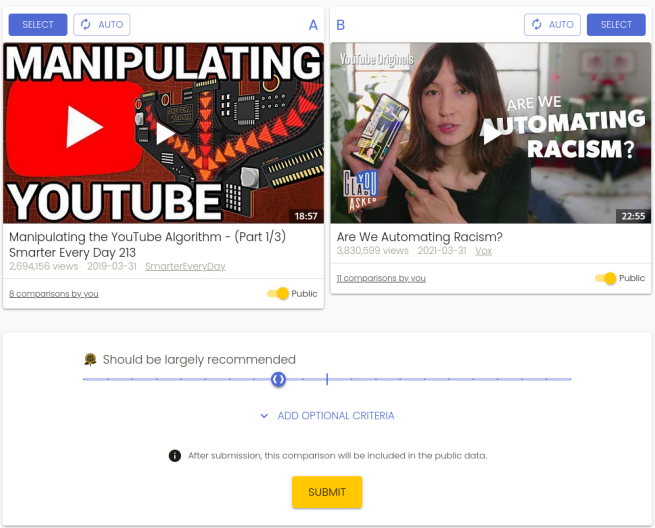


Moral machine et WeBuildAI : des IA conçues démocratiquement.

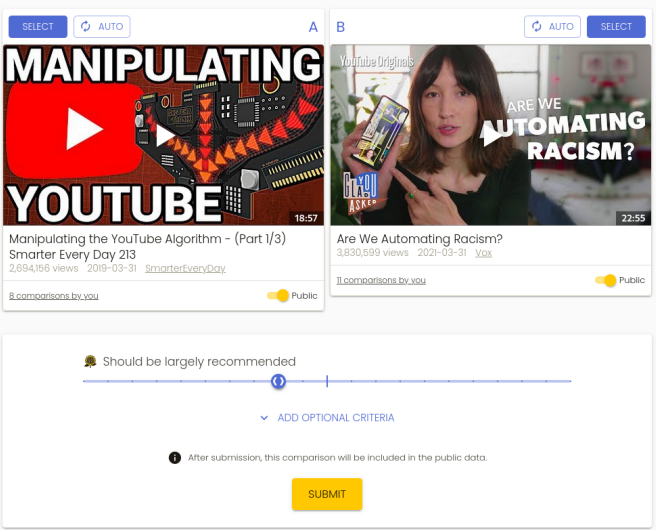
Comment résoudre les vrais dilemmes de l'éthique des IA

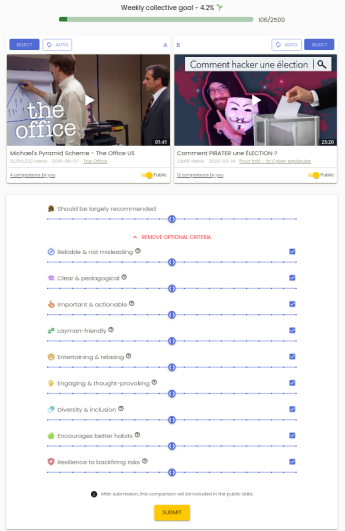


Élicitation des préférences



Élicitation des préférences





Apprendre de jugements comparatifs quantifiés (AAAI'24)

La famille $\mathbf{p}(r|\theta) \propto f(r)e^{r\theta}$ de modèles probabilistes des comparisons r sachant une différence de score θ est appelée modèle GBT. Elle est paramétrée par une fonction f symétrique.

Apprendre de jugements comparatifs quantifiés (AAAI'24)

La famille $\mathbf{p}(r|\theta) \propto f(r)e^{r\theta}$ de modèles probabilistes des comparisons r sachant une différence de score θ est appelée modèle GBT. Elle est paramétrée par une fonction f symétrique.

Theorem (Dembo et Zeitouni 2009)

Pour tout f, la log-vraisemblance de f-GBT est strictement concave et infiniment dérivable. Le maximum de vraisemblance (ou tout maximum a posteriori avec un a priori log-concave) est donc rapide à calculer.

Apprendre de jugements comparatifs quantifiés (AAAI'24)

La famille $\mathbf{p}(r|\theta) \propto f(r)e^{r\theta}$ de modèles probabilistes des comparisons r sachant une différence de score θ est appelée modèle GBT. Elle est paramétrée par une fonction f symétrique.

Theorem (Dembo et Zeitouni 2009)

Pour tout f, la log-vraisemblance de f-GBT est strictement concave et infiniment dérivable. Le maximum de vraisemblance (ou tout maximum a posteriori avec un a priori log-concave) est donc rapide à calculer.

Theorem (Fageot, Farhadkhani, H et Villemaud, AAAI'24)

Avec un a priori gaussien, le maximum a posteriori de f-GBT est une fonction monotone des comparaisons.

Apprendre de jugements comparatifs quantifiés (AAAI'24)

La famille $\mathbf{p}(r|\theta) \propto f(r)e^{r\theta}$ de modèles probabilistes des comparisons r sachant une différence de score θ est appelée modèle GBT. Elle est paramétrée par une fonction f symétrique.

Theorem (Dembo et Zeitouni 2009)

Pour tout f, la log-vraisemblance de f-GBT est strictement concave et infiniment dérivable. Le maximum de vraisemblance (ou tout maximum a posteriori avec un a priori log-concave) est donc rapide à calculer.

Theorem (Fageot, Farhadkhani, H et Villemaud, AAAI'24)

Avec un a priori gaussien, le maximum a posteriori de f-GBT est une fonction monotone des comparaisons.

Theorem (Fageot, Farhadkhani, H et Villemaud, AAAI'24)

Avec un a priori gaussien, le maximum a posteriori de f-GBT est une fonction Lipschitz continue du vecteur des comparaisons (avec la distance de Hamming).

Solidago : une pipeline modulaire, open source et libre (AGPL)

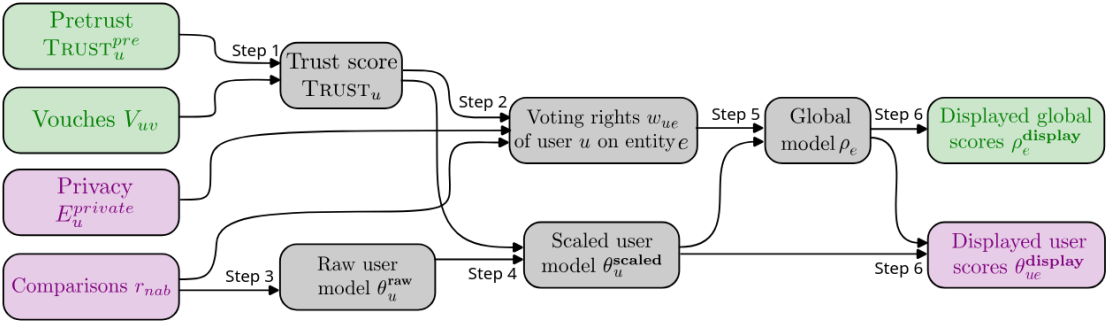
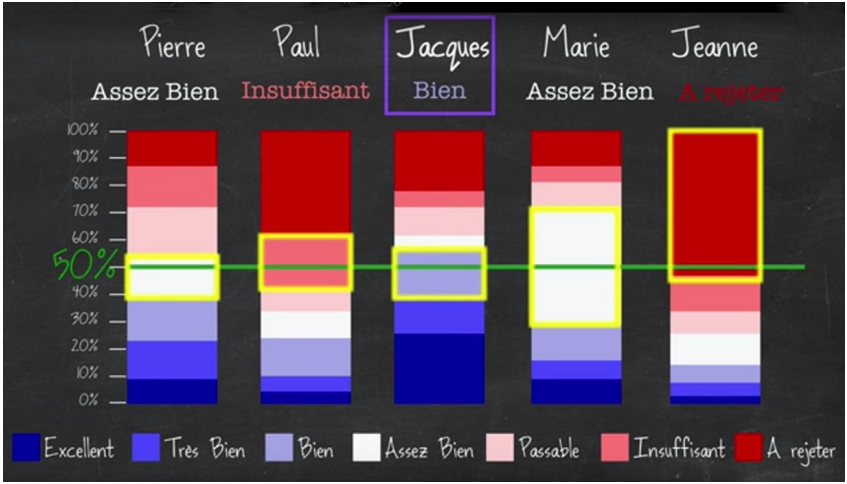


Figure 1: This figure describes the SOLIDAGO pipeline (slightly simplified). Green boxes correspond to public data, while black boxes are kept hidden. The purple boxes contain both public and private data. The pipeline is composed of 6 steps, namely (1) trust propagation, (2) voting rights assignment, (3) preference learning, (4) model scaling, (5) model aggregation and (6) post-process.

L'aggrégation sécurisée des voix par jugement majoritaire



Extrait de "Réformons l'élection présidentielle !"

Voter en très grande dimension

On the Strategyproofness of the Geometric Median

El-Mahdi El-Mhamdi
Calicarpa, École Polytechnique

Sadegh Farhadkhani*
EPFL

Rachid Guerraoui
EPFL

Lê-Nguyên Hoang*
Calicarpa, Tournesol

Abstract

The *geometric median*, an instrumental component of the secure machine learning toolbox, is known to be effective when robustly aggregating models (or gradients), gathered from potentially malicious (or strategic) users. What is less known is the extent to which the geometric median incentivizes dishonest behaviors. This paper addresses this fundamental question by quantifying its *strategyproofness*. While we observe that the geometric median is not even approximately strategyproof, we prove that it is *asymptotically α-strategyproof*: when the number of users is large enough, a user that misbehaves can gain at most a multiplicative factor α, which we compute as a function of the distribution followed by the users. We then generalize our results to the case where users actually care more about specific dimensions, determining how this impacts α. We also show how the *skewed geometric medians* can be used to improve strategyproofness.

1 INTRODUCTION

There has recently been a growing interest in collaborative machine learning to efficiently utilize the ever-increasing amount of data and computational resources (McMahan et al., 2017; Kairouz et al., 2021; Abadi et al., 2015). Collaborative learning gathers information from multiple users (e.g., gradient vectors (Zinkevich et al., 2010), local model parameters (Dinh et al., 2020; Farhadkhani et al., 2022b) or users' preferences (Noothigattu et al., 2018; Allouah et al., 2023)) and typically summarizes it in a single vector. While averaging is the most widely used method for aggregating multiple vectors into a single vector (Polyak and Juditsky, 1992), it suffers from severe security flaws: averaging can be arbitrarily manipulated by a single strategic user (Blanchard et al., 2017).

The geometric median is a promising "robust" alternative to averaging. It has been widely used in collaborative learning as it is a provably good approximation of the average (Minsker, 2015) and it is robust to a minority of malicious users (Lopuhaä and Rousseeuw, 1989). A large body of research known as "Byzantine learning" (Blanchard et al., 2017; Chen et al., 2017; El-Mhamdi et al., 2018; Rajput et al., 2019; Alistarh et al., 2018) uses the geometric median to ensure safe learning despite the presence of participants with arbitrarily malicious behavior (Farhadkhani et al., 2022a; Karimireddy et al., 2022; Acharya et al., 2022; Wu et al., 2020; So et al., 2021; Gu and Yang, 2021; Pillutla et al., 2022; Farhadkhani et al., 2022b). Interestingly, the geometric median also satisfies the fairness principle *"one voter, one vote with a unit force"* (see Section F.5), making it ethically appealing.

In this paper, we study the extent to which the geometric median *incentivizes* strategic manipulation[1]. Ideally, we would like the geometric median to be *strategyproof* (Gibbard, 1973; Satterthwaite, 1975; Brandt et al., 2016), i.e., we want it to be in each voter's best interest to report their true preferred vector. Put differently, honesty would ideally be a *dominant strategy* (Chung and Ely, 2007). This is very different from *Byzantine learning*, which only focuses on the resilience of the training, usually assuming a *majority* of honest users. Conversely, we consider the more realistic case where *every* user wants to bias the algorithm towards their specific target states. Such considerations are critical for high-stake life-endangering applications such as content moderation and recommendation (Yue, 2019;

Authors are listed in alphabetical order.
*Correspondence to: sadegh.farhadkhani@epfl.ch, and len@tournesol.app.

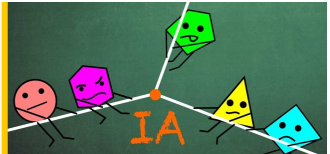
Proceedings of the 26th International Conference on Artificial Intelligence and Statistics (AISTATS) 2023, Valencia, Spain. PMLR: Volume 206. Copyright 2023 by the author(s).

[1] Hence, we often use the term "voter" instead of "user".

41

114 comparisons

29 contributors



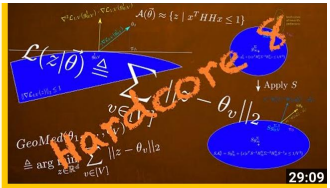
Science4All

Les maths des IA démocratiques

8

24 comparisons

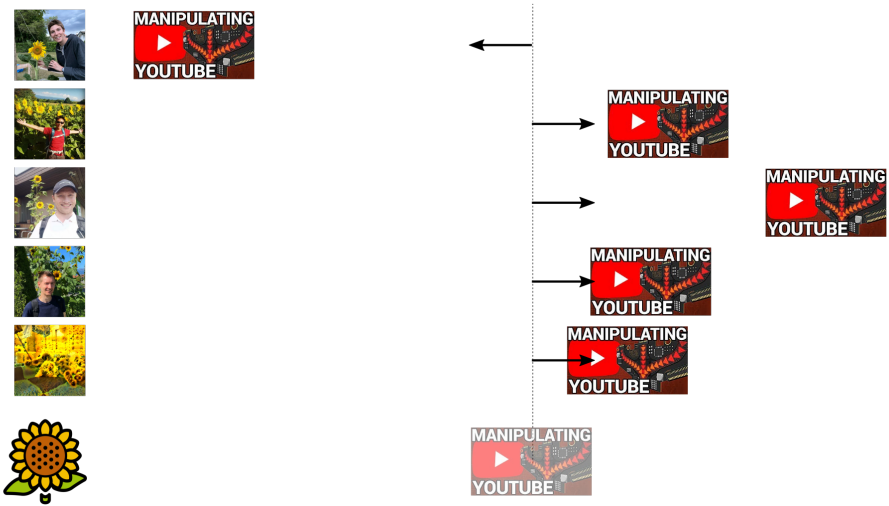
6 contributors



Science4All

Les maths des IA démocratiques ⁉

La primitive clé: la médiane quadratiquement régularisée

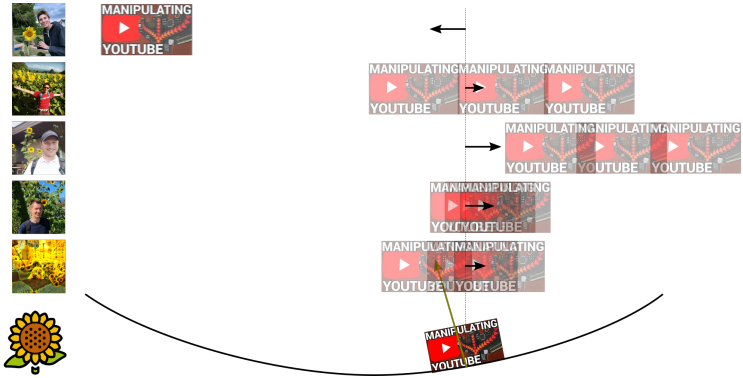


La primitive clé: la médiane quadratiquement régularisée






La primitive clé: la médiane quadratiquement régularisée



Theorem (Allouah, Guerraoui, H, Villemaud (AISTATS'24))

$QrMed_L(\mathbf{w}, \mathbf{x}) \triangleq \arg\min_{z \in \mathbb{R}} \left\{ \frac{1}{2L} z^2 + \sum_{i=1}^{n} w_i |x_i - z| \right\}$ est L-Lipschitz continue en les droits de vote (muni de la norme infinie).

Une base de données fascinantes

Comptes activés	Comparaisons	Vidéos notées
20 939	235 846	44 568
+ 112	+ 2 877	+ 558

Quelques découvertes préliminaires

- L'importance importe.

Quelques découvertes préliminaires

- L'importance importe.
- Biais de récence.

Quelques découvertes préliminaires

- L'importance importe.
- Biais de récence.
- Biais vers la vidéo de gauche.

Quelques découvertes préliminaires

- L'importance importe.
- Biais de récence.
- Biais vers la vidéo de gauche.
- Comparer tous les critères réduit ces biais.

Des questions fascinantes : les droits de vote 2.0

- Preuve de citoyenneté avec divulgation nulle de connaissance.

Des questions fascinantes : les droits de vote 2.0

- Preuve de citoyenneté avec divulgation nulle de connaissance.
- Délégation de vote (démocratie liquide).

Des questions fascinantes : les droits de vote 2.0

- Preuve de citoyenneté avec divulgation nulle de connaissance.
- Délégation de vote (démocratie liquide).
- Réseau de confiance et valorisation de l'expertise.

Des questions fascinantes : les droits de vote 2.0

- Preuve de citoyenneté avec divulgation nulle de connaissance.
- Délégation de vote (démocratie liquide).
- Réseau de confiance et valorisation de l'expertise.
- Quelle norme utiliser ? (vote unitaire selon la norme $\ell_?$)

Maths & sociologie : écouter la majorité silencieuse

- Concevoir des représentants algorithmiques.

Maths & sociologie : écouter la majorité silencieuse

- Concevoir des représentants algorithmiques.
- Sociologie des participants.

Maths & sociologie : écouter la majorité silencieuse

- Concevoir des représentants algorithmiques.
- Sociologie des participants.
- Apprentissage collaboratif sécurisé.

Maths & sociologie : écouter la majorité silencieuse

- Concevoir des représentants algorithmiques.
- Sociologie des participants.
- Apprentissage collaboratif sécurisé.
- Incertitudes et devoir de vigilance.

Le défi philosophico-mathématique : une démocratie des volitions

- Distinguer préférences instinctives et volitions réfléchies.

Le défi philosophico-mathématique : une démocratie des volitions

- Distinguer préférences instinctives et volitions réfléchies.
- Modéliser la psychologie du jugement humain.

Le défi philosophico-mathématique : une démocratie des volitions

- Distinguer préférences instinctives et volitions réfléchies.
- Modéliser la psychologie du jugement humain.
- Favoriser le consensus radical (en puissance).

Le défi philosophico-mathématique : une démocratie des volitions

- Distinguer préférences instinctives et volitions réfléchies.
- Modéliser la psychologie du jugement humain.
- Favoriser le consensus radical (en puissance).
- Justifier la légitimité (constructivisme moral ?).

La cybersécurité du vote : un problème encore très ouvert...

- Transparence et vérifiabilité (protocole ouvert, calcul réparti).

La cybersécurité du vote : un problème encore très ouvert...

- Transparence et vérifiabilité (protocole ouvert, calcul réparti).
- Confidentialité du vote (zero-knowledge, calcul multipartite...).

La cybersécurité du vote : un problème encore très ouvert...

- Transparence et vérifiabilité (protocole ouvert, calcul réparti).
- Confidentialité du vote (zero-knowledge, calcul multipartite...).
- Résilience/strategyproofness du mode de scrutin.

La cybersécurité du vote : un problème encore très ouvert...

- Transparence et vérifiabilité (protocole ouvert, calcul réparti).
- Confidentialité du vote (zero-knowledge, calcul multipartite...).
- Résilience/strategyproofness du mode de scrutin.
- Cybersécurité des machines de vote (!!).

Pegasus, Predator, DarkMatter...

14

23
comparisons

2
contributors

Amnesty International

How Your Phone Can Be Weaponized Ag...

L'urgence de la souveraineté numérique

Lenovo Customer Feedback program [edit]

At a third time in 2015, criticism arose that Lenovo might have installed software that looked suspicious on their commercial Think-PC lines. This was discovered by Computerworld writer Michael Horowitz, who had purchased several Think systems with the Customer Feedback program installed, which seemed to log usage data and metrics.[233] Further analysis by Horowitz revealed however that this was mostly harmless, as it was only logging the usage of some pre-installed Lenovo programs, and not the usage in general, and only if the user allowed the data to be collected. Horowitz also criticized other media for quoting his original article and saying that Lenovo preinstalled spyware, as he himself never used that term in this case and he also said that he does not consider the software he found to be spyware.[234]

Lenovo Accelerator [edit]

As of June 2016, a Duo Labs report stated that Lenovo was still installing bloatware, some of which leads to security vulnerabilities as soon as the user turns on their new PC.[235][236] Lenovo advised users to remove the offending app, "Lenovo Accelerator".[237] According to Lenovo, the app, designed to "speed up the loading" of Lenovo applications, created a man-in-the-middle security vulnerability.

U.S. Marine network security breach [edit]

In February 2021, *Bloomberg Businessweek* reported that U.S. investigators found in 2008 that military units in Iraq were using Lenovo laptops in which the hardware had been altered. According to a testimony from the case in 2010, "A large amount of Lenovo laptops were sold to the U.S. military that had a chip encrypted on the motherboard that would record all the data that was being inputted into that laptop and send it back to China". Lenovo was unaware of the testimony and the U.S. military did not inform the company of any security concerns. A Lenovo spokesperson stated that "we have no way to assess the allegations you cite or whether security concerns may have been triggered by third-party interference."[238]

Chinese regulator pauses partnership with Alibaba

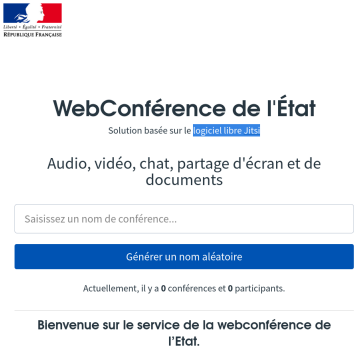
23 December 2021

China's telecommunications regulator has paused a partnership with Alibaba Cloud after one of the firm's engineers discovered the Log4shell security flaw.

According to state-backed Chinese media, the suspension is because the firm did not report Log4shell to The Ministry of Industry and Information Technology (MIIT) in time.

Adoptez et faites adopter le logiciel libre

GendBuntu
(Linux pour la gendarmerie)



webconf.numerique.gouv.fr
(basé sur Jitsi)

La démocratie numérique a besoin de votre aide

- Utiliser et contribuer à Tournesol.

La démocratie numérique a besoin de votre aide

- Utiliser et contribuer à Tournesol.
- Promouvoir la démocratie numérique autour de vous.

La démocratie numérique a besoin de votre aide

- Utiliser et contribuer à Tournesol.
- Promouvoir la démocratie numérique autour de vous.
- Effectuer des dons aux organisations qui défendent la démocratie.

La démocratie numérique a besoin de votre aide

- Utiliser et contribuer à Tournesol.
- Promouvoir la démocratie numérique autour de vous.
- Effectuer des dons aux organisations qui défendent la démocratie.
- Contribuer aux codes sources libres et open source d'intérêt général.

La démocratie numérique a besoin de votre aide

- Utiliser et contribuer à Tournesol.
- Promouvoir la démocratie numérique autour de vous.
- Effectuer des dons aux organisations qui défendent la démocratie.
- Contribuer aux codes sources libres et open source d'intérêt général.
- Contribuer à la recherche pour une IA fiable.

Section 6

Conclusion

Un chantier gargantuesque

“On est face à un abime.”

Nathalie Riché (notre éditrice)

Un chantier gargantuesque

“On est face à un abime.”

Nathalie Riché (notre éditrice)

Syllogisme du politicien

- Il faut faire quelque chose.
- X est quelque chose.
- Donc il faut faire X.

Un chantier gargantuesque

"On est face à un abime."

Nathalie Riché (notre éditrice)

Syllogisme du politicien

- Il faut faire quelque chose.
- X est quelque chose.
- Donc il faut faire X.

Cynisme éclairé

- Seul l'impact compte vraiment.
- Mon impact est négligeable.
- Donc je n'ai pas à agir.

Trois raisons (plus rationnelles) d'agir

Trois raisons (plus rationnelles) d'agir

1. Chaque fraction de degré compte.

Trois raisons (plus rationnelles) d'agir

1. Chaque fraction de degré compte.

2. Créer des adjacents possibles.

Trois raisons (plus rationnelles) d'agir

1. Chaque fraction de degré compte.
2. Créer des adjacents possibles.
3. Le plus beau des hobbys/métiers.

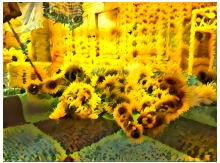
Le fabuleux chantier : rendre l'information sécurisée et démocratique

Tournesol.app


Adrien Matissart


Louis Faucon

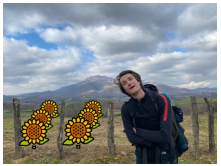

Aidan Jungo


Romain Beylerian


Jean-Lou Fourquet


Julien Fageot

Victor Fersing


Titouan Lustin


Sadegh Farhadkhani


Oscar Villemaud


Martin Gibert


Maxime Lambrecht

\+ des centaines de chercheurs, développeurs, vulgarisateurs et partenaires
\+ les dizaines de milliers de bénévoles qui contribuent à notre recherche-action !

Calicarpa.com

El Mahdi El Mhamdi (Professeur à l'X)

Sébastien Rouault