

5 |

# visions algorithmiques du monde

thomas tari

5 1 | la culture est dans le code

5 2 | pratiques algorithmiques

5 3 | comment les machines  
apprennent

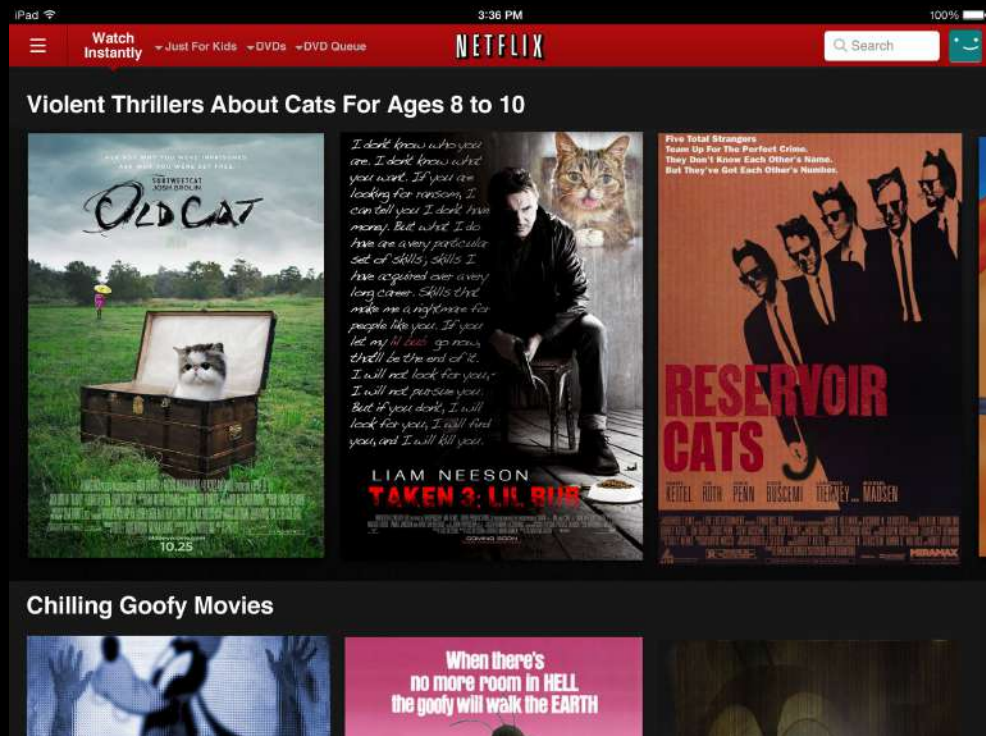
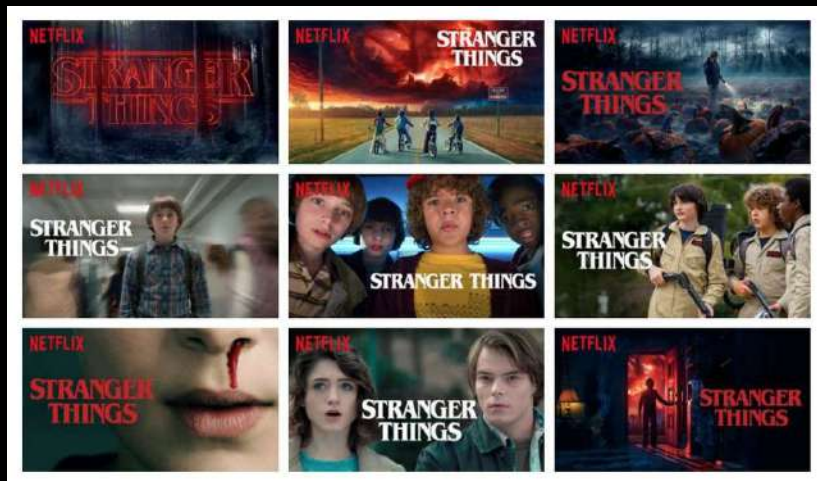
5 4 | une base de données est  
une vision du monde

16 oct 2025

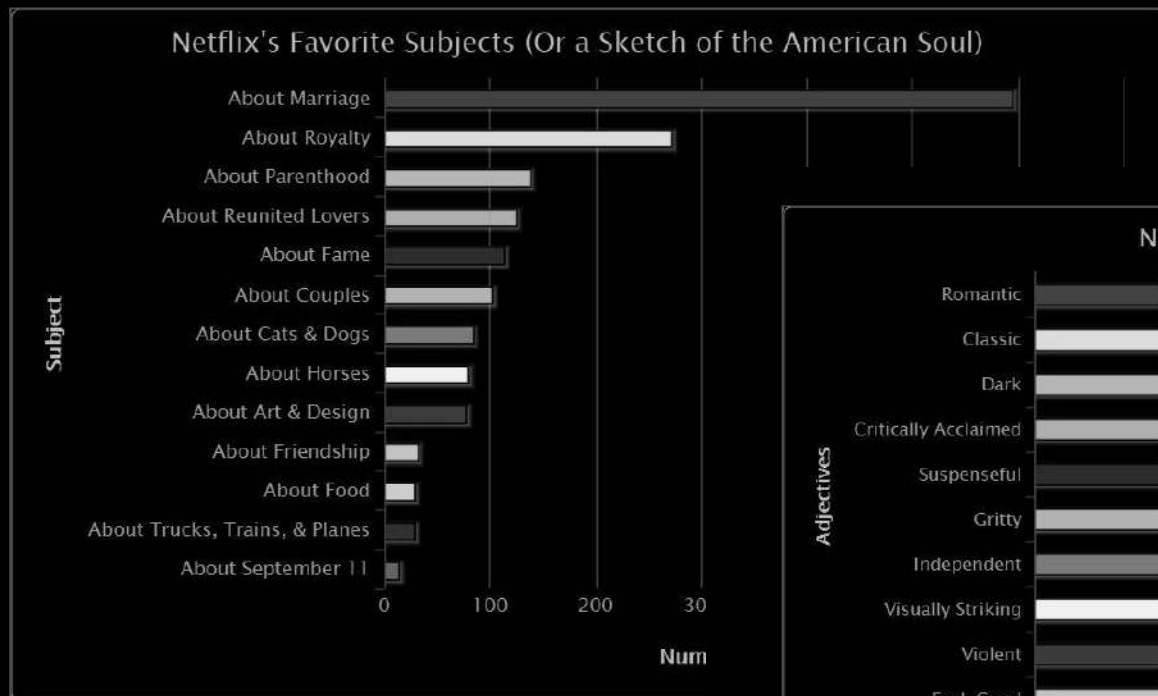
5 1 | la culture est dans le code

# comment **NETFLIX** a décortiqué **HOLLYWOOD**

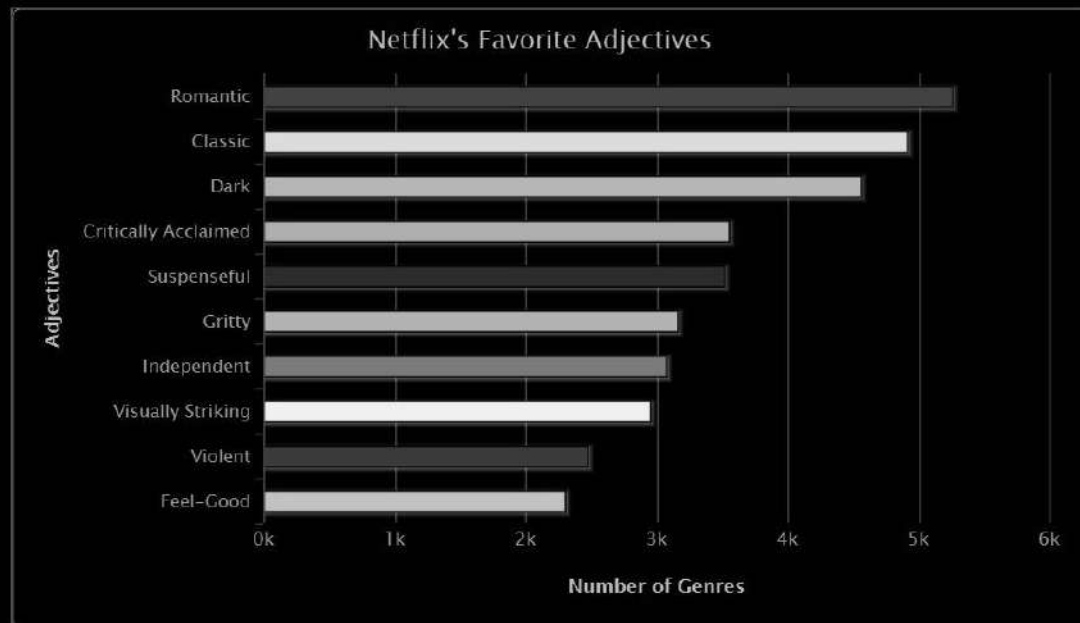
**Samurai & Ninja Bounty-Hunter  
Road Trip Dramas Based on a  
Book Set in Ancient Times For  
Kids With a Strong Female Lead**



# comprendre la grammaire de netflix |



'how netflix reverse-engineered hollywood', alexis madrigal, 2014



# ethnographie du processus de codage |

après que greg harty se soit levé du lit dans son appartement de sherman oaks, il prend une tasse de café et commence sa journée de travail à un bureau dans le coin de son salon | sa mission : regarder trois épisodes de « modern family » | pendant la diffusion de la sitcom à succès, l'aspirant scénariste ouvre une autre fenêtre sur son ordinateur portable et affiche une feuille de calcul | il commence à choisir des étiquettes – son employeur, netflix, les appelle des tags – pour décrire ce qu'il voit

la comédie : « décalée » | l'humour : « clair, sombre » | le ton : « humoristique », « irrévérencieux » et « sincère »

phil de ty burrell : « idiot », « enfantin », un « idiot adorable »

claire de julie bowen : « contrôlante », « affirmée »

jay d'ed o'neill : « hargneux », « mâle alpha »

# le mystère perry mason |

personne — pas même yellin — ne sait vraiment pourquoi il existe autant de sous-genres mettant en scène raymond burr et barbara hale ; c'est inexplicable avec la logique humaine, c'est juste quelque chose qui s'est produit. [...] comment appelle-t-on ce quelque chose dans le code et les données qui a conduit à l'existence de ces micro-genres ?

**la conclusion déconcertante et remarquable est que lorsque les entreprises combinent l'intelligence humaine et l'intelligence artificielle, il se produit des choses que nous ne pouvons pas comprendre.**

« laissez-moi philosopher un instant : dans le monde humain, ce sont les hasards heureux qui rendent la vie intéressante », m'a dit yellin. « plus vous ajoutez de complexité au monde des machines, plus vous ajoutez des hasards heureux que vous ne pouviez imaginer. perry mason va se produire. ces fantômes dans la machine seront toujours un sous-produit de la complexité. **et parfois, nous appelons cela un bug, et parfois, nous appelons cela une fonctionnalité.** »



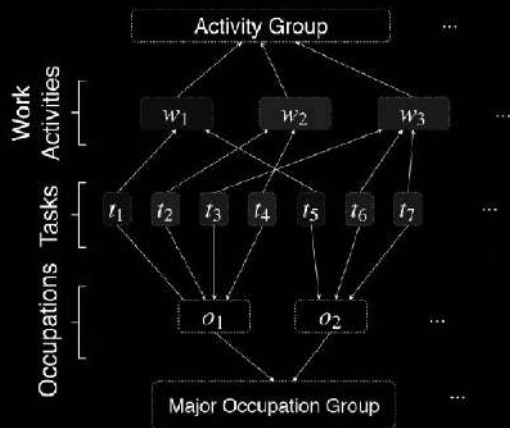
list of top actors

**Raymond Burr**  
*Bruce Willis*  
*George Carlin*  
*Jackie Chan*  
*Andy Lau*  
*Robert De Niro*  
**Barbara Hale**  
*Clint Eastwood*

*Gene Autry*  
*Yun-Fat Chow*  
*Anthony Hopkins*  
*Bob Hope*  
*Cary Grant*  
*Elvis Presley*  
*Fred Astaire*  
*John Wayne*

# la base de données o\*net |

une classification exhaustive des compétences, des savoirs et des expériences scolaires et professionnelles prérequis pour exercer l'ensemble des métiers



## AIM 1: Qualitative Data Collection & Task Analysis

**Fieldwork**  
- Interviews  
- Observations  
- Document collection  
- Photos  
- Focus groups

**Task specification**  
- Extract tasks from fieldnotes, interviews, and focus groups  
- Categorization

**Primary care survey**  
- Task validation  
- Automation rating  
- Workload impact rating

**Dataset formulation**

## AIM 2: Quantitative Analysis & Modeling

**Expert survey**  
- Task automation rating

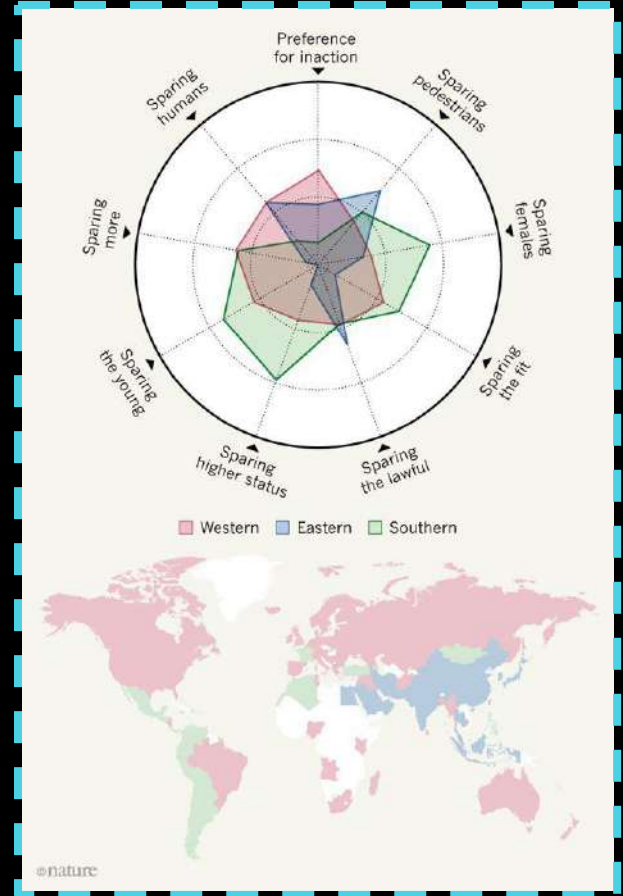
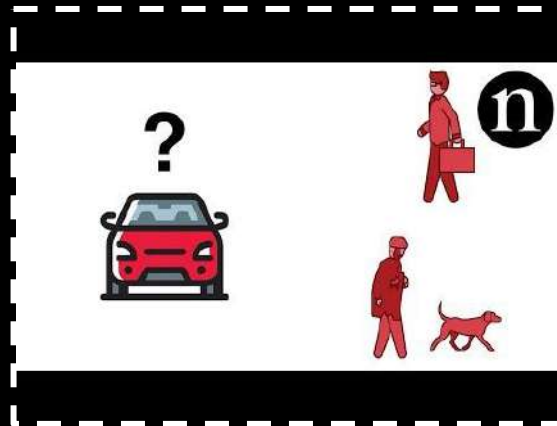
**Machine learning prediction model**

**Automation insights**

**Augment data**  
- O\*NET: Skills, Knowledge, & Abilities

# l'expérience de la machine morale |

<http://moralmachine.mit.edu/>



## THE CAR THAT KNEW TOO MUCH

Can a Machine Be Moral?

JEAN-FRANÇOIS BONNEFON



« quand on commence à cliquer sur la junk food des réseaux sociaux, on nous en propose encore plus, ce qui rend la résistance encore plus difficile » |

[how the secret algorithms behind social media actually work](#), noah giansiracusa, time, août 2025

« supposons que nous ayons à l'esprit un utilisateur spécifique et une publication spécifique, et que nous écrivions Plike pour la probabilité que l'utilisateur aime la publication, Plove pour la probabilité qu'il appuie sur l'emoji cœur, Pangry pour la probabilité qu'il appuie sur l'emoji colère, Pcomment pour la probabilité qu'il commente la publication et Pshare pour la probabilité qu'il la partage. écrivons Vlike, Vlove, etc. pour les valeurs en points attribuées à ces engagements. la formule magique est alors la suivante :

$$\text{Score} = V\text{like} \times P\text{like} + V\text{love} \times P\text{love} + V\text{angry} \times P\text{angry} + V\text{comment} \times P\text{comment} + V\text{share} \times P\text{share}$$

supposons qu'un « j'aime » vaut un point, qu'un emoji en forme de cœur vaut cinq points et qu'un commentaire vaut trente points

photo de chiot :  $1 \times 0,5 + 5 \times 0,2 + 30 \times 0,1 = 4,5$

post covid&5g :  $1 \times 0 + 5 \times 0 + 30 \times 0,2 = 6$

# les algorithmes comme cultures |

de la définition de kowalski's (1979) : “algorithme = logique + contrôle”, au caractère multiple des cultures algorithmiques

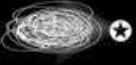



« nous devrions peut-être commencer à les considérer comme une population à échantillonner » ; les cultures algorithmiques sont culturellement complexes, les réalités qu'elles façonnent sont tributaires des contextes et des cultures dans lesquels elles sont produites et mises en œuvre - elles doivent être considérées pour leur pluralité, leur désordre (messiness) et leur nature contradictoire

les algorithmes en tant que culture : les algorithmes ne sont pas des objets techniques singuliers qui entrent dans de nombreuses interactions culturelles différentes, mais plutôt des objets instables, culturellement mis en œuvre par les pratiques que les gens utilisent pour interagir avec eux.



5 2 | pratiques algorithmiques

# déconstruire les algorithmes | cardon, 2016

	À côté 	Au-dessus 	Dans 	Au-dessous 
Exemples	Médiamétrie, Google Analytics, affichage publicitaire	PageRank de Google, Digg, Wikipédia	Nombre d'amis Facebook, Retweet de Twitter, notes et avis	Recommandation Amazon, publicité comportementale
Données	Vues	Liens	Likes	Traces
Population	Échantillon représentatif	Vote censitaire, communautés	Réseau social, affinitaire, déclaratif	Comportements individuels implicites
Forme du calcul	Vote	Classements méritocratiques	<i>Benchmark</i>	<i>Machine learning</i>
Principe	<i>Popularité</i>	<i>Autorité</i>	<i>Réputation</i>	<i>Prédiction</i>

# une sociologie du travail algorithmique |

camille girard-chanudet, [« sur la chaîne de production de l'intelligence artificielle. une sociologie du travail algorithmique »](#), sociologie du travail, 2025 et [« mais l'algo, là, il va mimer nos erreurs ! contraintes et effets de l'annotation des données d'entraînement d'une ia »](#), réseaux, 2023

« sixième étage de l'aile s du palais de la justice, sur l'île de la cité à paris. une magistrate m'accompagne le long des couloirs feutrés de la cour de cassation, jusqu'à une enfilade de bureaux : "voilà, c'est ici. la première porte à gauche, c'est les *data scientists*. juste après, les dev qui bossent sur l'interface. là, à droite, c'est louise, notre greffière, et tous les bureaux suivants, c'est les annotatrices". entre tableaux blancs couverts de notes de réunions et écrans sur lesquels s'affichent chiffres, graphiques et décisions de justice, je découvre, bureau après bureau, le service chargé de l'intelligence artificielle de la juridiction suprême »

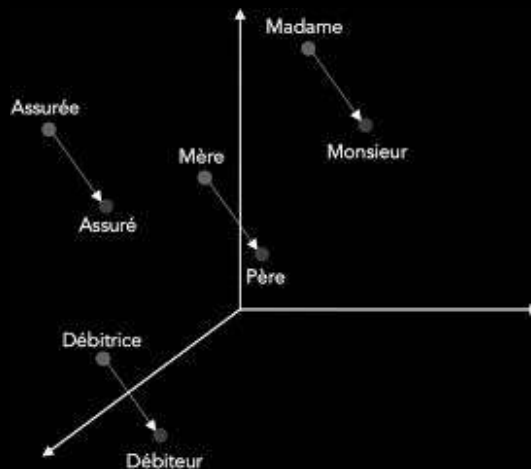
le cas d'un algorithme développé pour anonymiser automatiquement les quatre millions de décisions de justice produites chaque année par les tribunaux français : celui-ci doit servir à protéger la vie privée des personnes citées dans les décisions, tout en permettant la diffusion publique, à grande échelle, de ces documents

# le travail de l'ia |

[l'ia] se distingue des processus de production industriels classiques en ce qu'elle ne transforme pas de la matière, mais de l'information

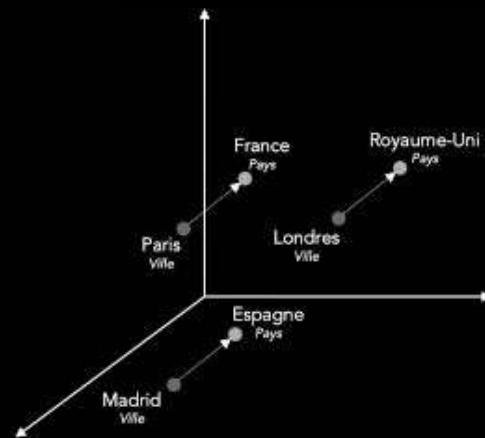
le travail de l'ia peut être ainsi envisagé comme un travail des données tendu vers un aplatissement progressif de l'information, qui représente à la fois une condition et un déterminant de l'automatisation de son traitement.

pourtant, l'enquête montre que le réel échappe toujours à cette encapsulation, [...] à la manière des défauts de production sur les chaînes industrielles, ils représentent des « grains de sable sous le capot ».



Dimension 1 (genre)

spatialisation  
sémantique au sein  
d'un fragment de  
« monde artificiel »



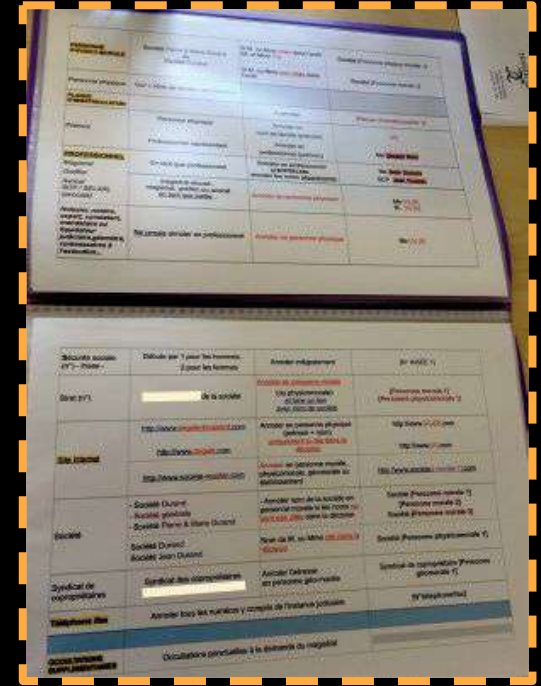
Dimension 2 (ville-pays)

# celui des magistrat·es |

l'origine de cet ensemble de catégories : il résulte du travail mené au sein d'un groupe d'une trentaine de magistrat·es, sélectionné·es pour leur expertise reconnue dans leurs contentieux de spécialité

« c'est vraiment un aller-retour permanent, en fait, entre les décisions et les principes qu'on cherche à appliquer [...]. finalement notre réflexion était... comment dire, il faut être honnête, hein, pas très scientifique »

ces groupes de travail ont été construits comme des espaces réservés à la magistrature, auxquels aucun acteur tiers n'a été convié. **ce choix, manifeste le caractère éminemment politique de l'activité taxinomique.** parce qu'ils fixent une représentation particulière du monde (ici, de la « vie privée ») dans des inscriptions rigides et performatives, **les systèmes de classification constituent d'importants leviers de pouvoir.** ils deviennent, dans la mesure de leurs usages, des **amplificateurs des conceptions du monde qu'ils incarnent**



# celui des annotatrices |

une quinzaine de fonctionnaires de catégorie c, majoritairement en fin d'une carrière réalisée au sein de l'administration judiciaire, elles ont été recrutées à ce poste sans prérequis de compétences dans le domaine informatique

Décision n°129460 - pourvoi n°08-20.000 - CTV1 - 31/02/2010

Annotations demandées

Personne physique (16)

Personne morale (2)

Adresse (2)

Date de naissance/anniversaire (0)

Date de mariage/PACS (0)

Date de décès (0)

N° INSEE (0)

Magistrat / membre du greffe (0)

Établissement (0)

N° SIREN/SIRET (0)

Localité (0)

Téléphone/Fax (0)

Email (0)

Site web sensible (0)

Compte bancaire (0)

Cadastre (0)

Guide d'annotation

Personne physique  
Romaine

30 Mme [10] a formé un pourvoi incident contre le même arrêt ;

50 Sur le rapport de Mme B [11], conseiller, les observations de Me C [12], avocat de M. [13], de la SCP M [14], F [15] et H [16], avocat de Mme [17], les conclusions écrites de M. N [18], premier avocat général, et après en avoir délibéré conformément à la loi ;

80 Attendu que, le 30 mai 1987, [19] [20] [21], veuve [22], a consenti aux deux enfants issus de son mariage, Mme [23] [24], épouse [25], et M. [26] [27], une donation-partage portant sur les biens dépendant de la communauté ayant existé entre elle et son époux ; que, par un arrêt du 20 juillet 1989, devenu irrévocable, la cour d'appel de Pau a confirmé le jugement ayant déclaré recevable l'action en rescision pour lésion de plus du quart engagée par Mme [28], épouse [29] ;

90 La sœur fait grief aux deux arrêts attaqués d'AVOIR rejeté l'exception d'irrecevabilité de l'action en rescision pour lésion de Madame [30] opposée par Monsieur [31] [32] ;

100 « En l'espèce, l'action intentée par Madame [33] par assignation en date du 8 mars 1988 a trait à l'estimation du stock d'armagnac compris dans la succession de [34] [35] décédé le 31 octobre 1986, contenue dans un acte intervenu par-devant Me [36], notaire à [37] en date du 30 mai 1987.

104 « Il s'agit en fait d'un acte mixte comprenant pour partie donation à titre de partage anticipé de Madame [38] de ses droits dans la communauté ayant existé entre elle-même et [39] [40] et pour partie partage de bien indivis recueillis dans la succession de [41] [42] entre ses deux enfants [43] [44] et (Madame [45]), dont les [46] estimés à 2.920.000 F et les Armagnacs propres estimés à 3.350.000 F.

Réinitialiser

Vue anonymisée

s'apparente au sale boulot : « elles ont juste à appliquer les consignes ! », mais un travail de soin

| surveiller les machines : j'ai signalé [une erreur récurrente d'annotation sur les adresses]. bon, on le savait, mais les gens étaient surpris. ambroise ne me croyait pas, parce que normalement il [l'algorithme] aurait dû le faire. il m'a dit "c'est impossible !". j'ai dit "ben écoutez, on va demander aux autres ce qu'elles en pensent", et elles ont dit "oui, oui, oui, c'est récurrent !". je lui dis "on continue ?" il dit "non, non, c'est bon, je vous crois" » (anna, annotatrice)



# et celui des data scientists |

« en fait dans un projet de ce genre, l'algorithmie c'est max 5 % du projet. le plus compliqué, c'est toute l'ingénierie, c'est nettoyer les données, les récupérer, le formatage, faire tout le pipeline de production, de l'entrée des données au passage de l'algo, etc. en vrai, finalement, je code assez peu au quotidien » (ambroise)

iels utilisent des algorithmes préexistants :  
l'anonymisation automatique des décisions de justice repose sur des algorithmes issus de facebook et du géant du e-commerce zalando. pour les data scientists, ces modèles représentent des outils de travail au comportement incertain, des « boîtes noires » dont il s'agit de contrôler et d'optimiser le fonctionnement en mettant en œuvre différentes stratégies.

→ création de données synthétiques

```
1 Jean Du Bois né le 31 janvier 2000 a fondé  
une entreprise d'informatique INFOWORLD
```

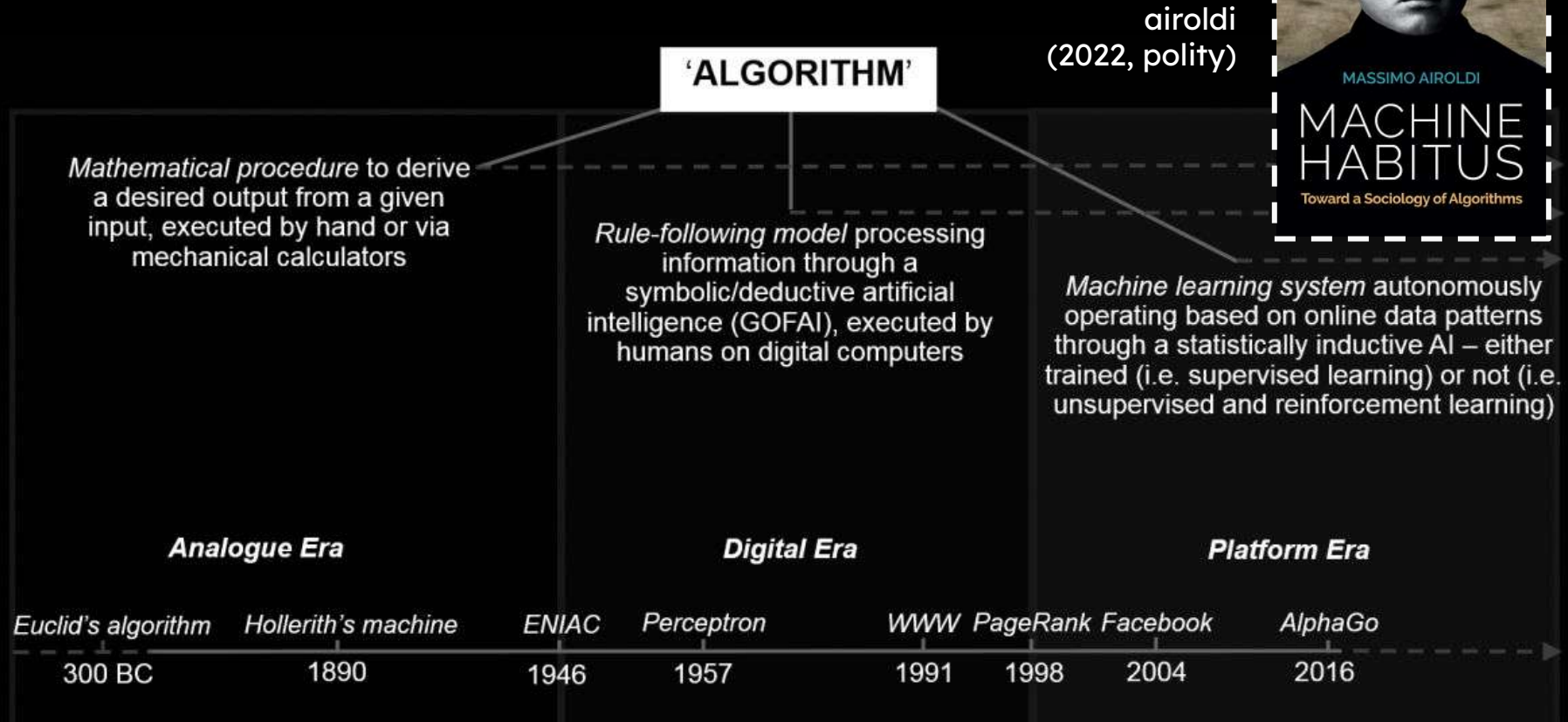
```
<tag: prenom start:0 end:4> Jean <tag: nom start:5  
end:12 > Du Bois né le <tag:date de naissance start:18  
end:33> 31 janvier 2000 a fondé une entreprise  
d'informatique <tag: personne_morale start: end:>  
INFOWORLD.
```

```
Jean B-prenom + arret_0143  
Du B-nom + arret_0143  
Bois I-nom + arret_0143  
né O + arret_0143  
le O + arret_0143  
31 B-date de naissance + arret_0143  
janvier I-date de naissance + arret_0143  
2000 I-date de naissance + arret_0143  
a O + arret_0143  
fondé O + arret_0143  
une O + arret_0143  
entreprise O + arret_0143  
INFOWORLD B-personne_morale + arret_0143
```

« les paramètres, c'est un peu comme de la magie. [...] on pourrait sûrement réfléchir aux implications mathématiques, il doit y avoir des explications conceptuelles, mais ce serait beaucoup trop lourd de réfléchir à tout ça. j'ai pas un background mathématique pur, du coup je fonctionne plutôt par tests et erreurs » (aniela).

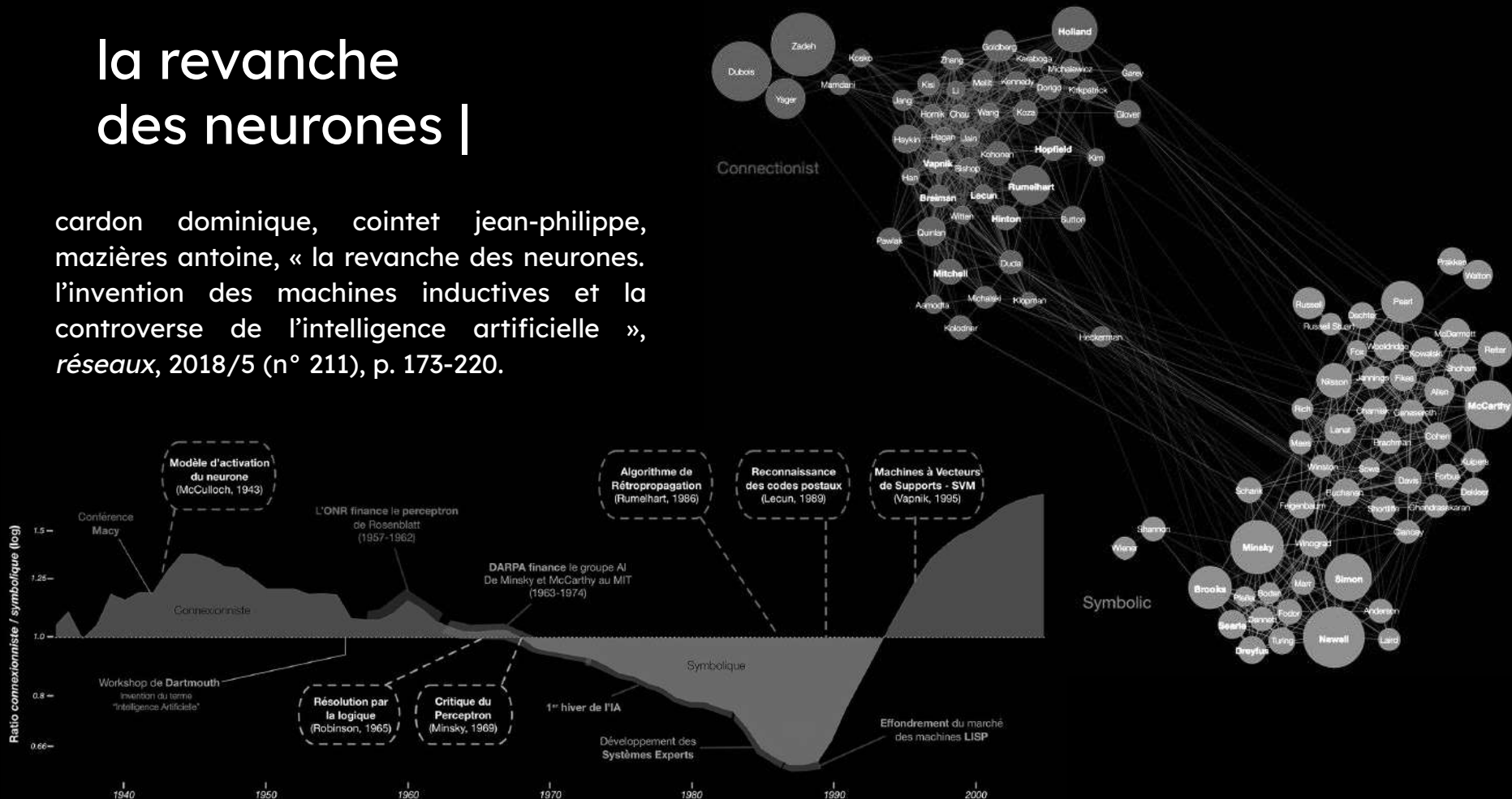
## 5 3 | comment les machines apprennent

# ce que nous avons appelé des algorithmes

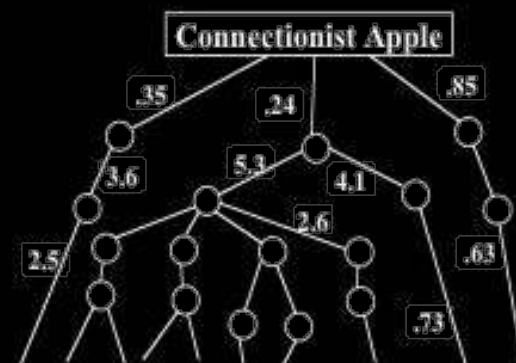
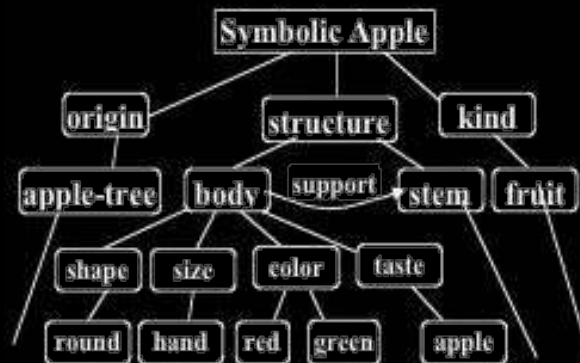


# la revanche des neurones |

cardon dominique, cointet jean-philippe,  
mazières antoine, « la revanche des neurones.  
l'invention des machines inductives et la  
controverse de l'intelligence artificielle »,  
*réseaux*, 2018/5 (n° 211), p. 173-220.

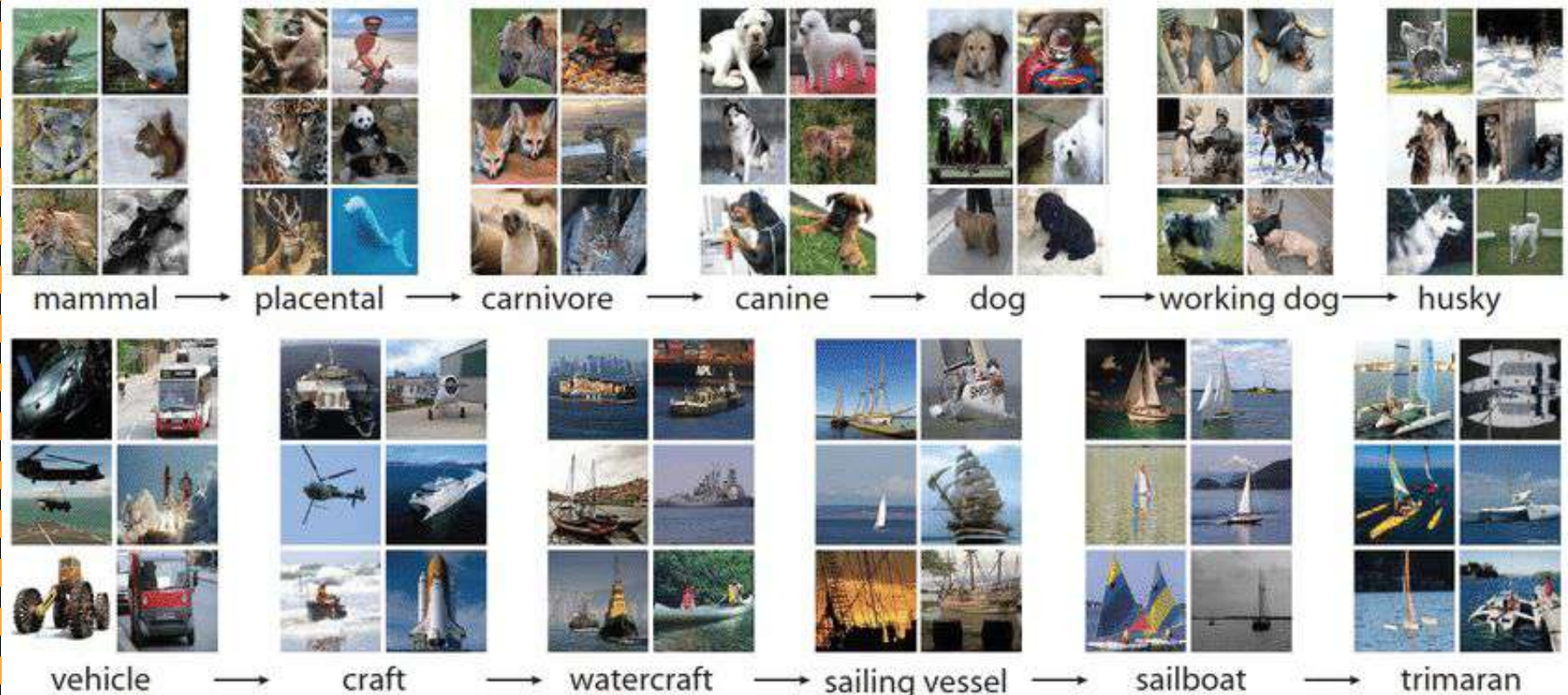


# philosophies de l'ia



<i>Machine</i>	<i>Monde</i>	<i>Calculateur</i>	<i>Horizon</i>
<b>Cybernétique</b> (connexionniste)	<i>Environnement</i>	<i>« Boîte noire »</i>	<i>Negative feedback</i>
<b>IA Symbolique</b> (symbolique)	<i>Monde « jouet »</i>	<i>Raisonnement logique</i>	<i>Résolution de problème</i>
<b>Système expert</b> (symbolique)	<i>Monde de connaissances expertes</i>	<i>Sélection des hypothèses</i>	<i>Exemples/ contre-exemples</i>
<b>Deep learning</b> (connexionniste)	<i>Le monde comme vecteur de données massives</i>	<i>Réseau de neurones profond</i>	<i>Optimisation de l'erreur sur objectif</i>

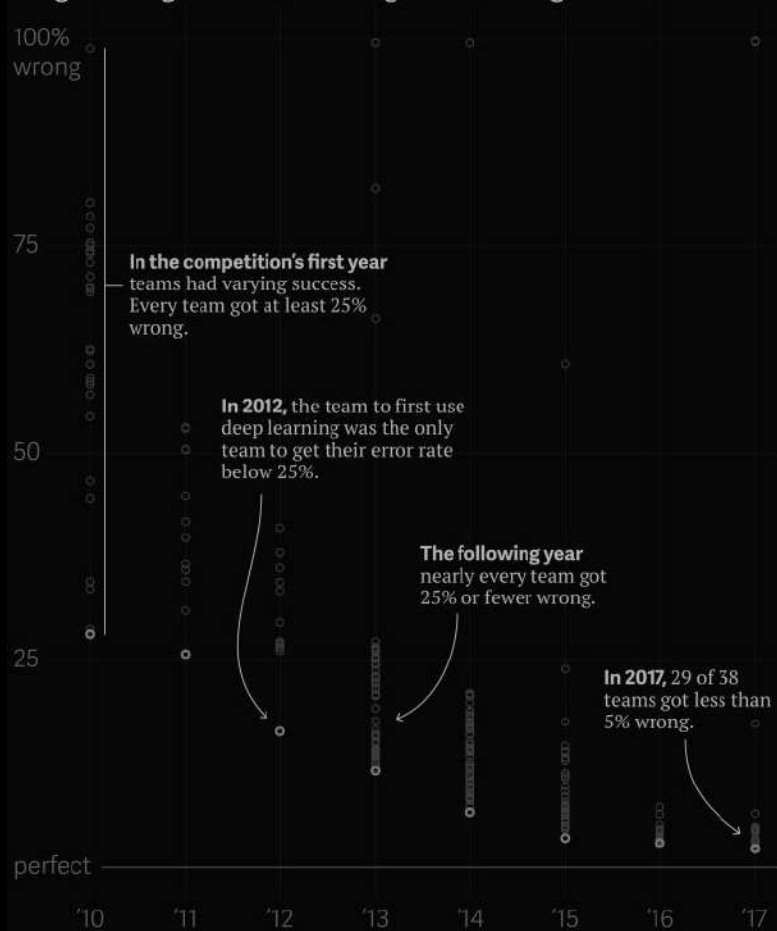
# systems de classification experts et symboliques



# la révolution du deep learning

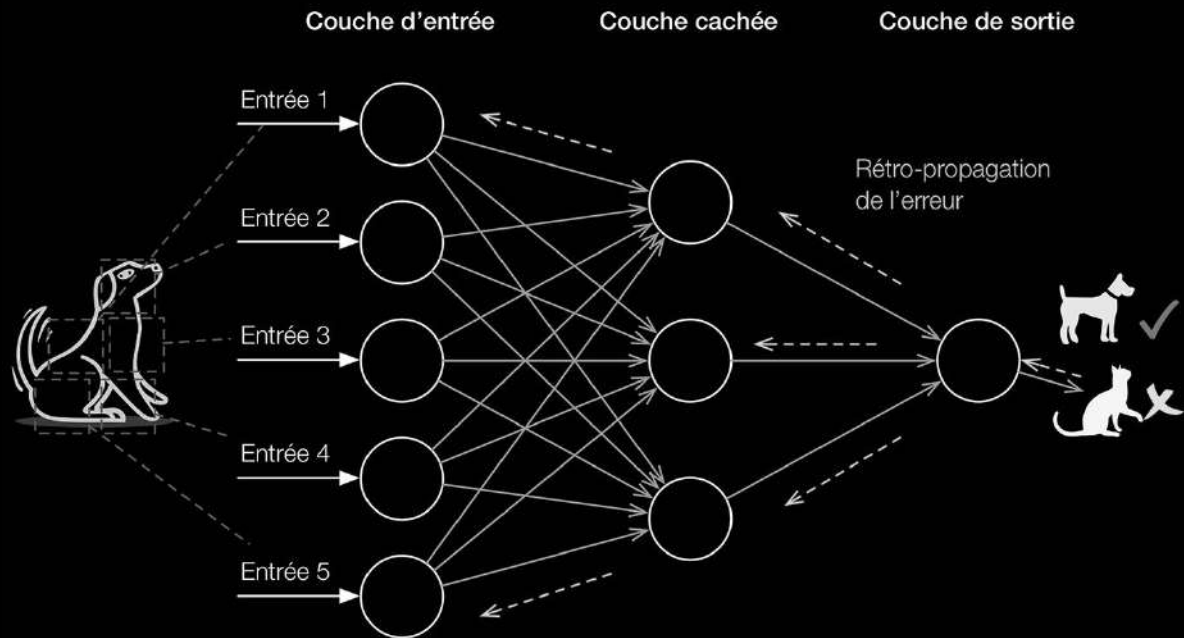
à l'époque, les mecs de computer vision s'excitaient sur imagenet depuis deux trois ans [une base de données de 1,2 million d'images étiquetées avec 1 000 catégories servant de benchmark pour comparer les résultats en classification des différents compétiteurs]. le number one, il était à 27,03 % d'erreur, le number 2 à 27,18 %, le number 3 à 27,68 %. et hinton, il envoie son mec sorti de nulle part : "on a fait tourner un gros deep, on est à 17 !". il met 10 points à tout le monde ! comme ça, le jeune geek, il arrive, il annonce le résultat, la salle bondée à craquer. enfin, il comprend rien à rien, genre il a 17 ans ! il ne sait pas pourquoi les trucs sont là. lui, il était enfermé dans sa boîte, il ne connaissait rien au domaine. [...] et tu as tous les grands manitous du computer vision qui essayent de réagir : "mais en fait c'est pas possible. ça va pas marcher pour la reconnaissance d'objet quand il faut..." enfin, les mecs étaient tous par terre parce que grosso modo cela foutait en l'air 10 ans d'intelligence, de tuning, de sophistication.

ImageNet Large Scale Visual Recognition Challenge results





# l'apprentissage profond | réseaux neuronaux





5 4 | une base de données  
est une vision du monde

# les glitches |

## projet #algoglitch au médialab



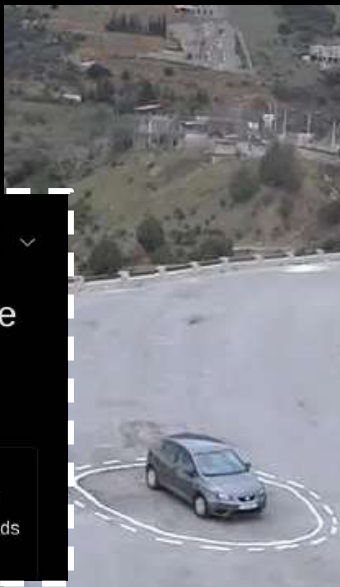
**Kashmir Hill** ✓  
@kashhill

Abonné

This week in algorithms gone wrong: Google steering people into literal fires rather than metaphorical ones.

**Joel Rubin** ✓ @joelrubin

"The Los Angeles Police Department asked drivers to avoid navigation apps, which are steering users onto more open routes — in this case, streets in the neighborhoods that are on fire." fw.to/EMJxt2E



**Josh Davis**  
@Obby\_Oss

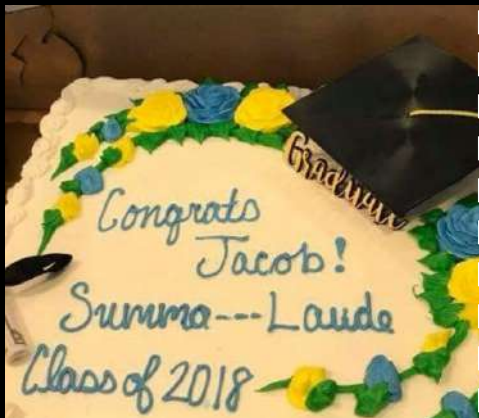
Suivre

4YO daughter asks my @Google Mini: "Hey Google, what's Mrs. Claus doing?"

Google proceeds to find an article about Mrs. Claus & explain that Mrs. Claus & Santa are made up stories.

Thanks Google.

#MerryChristmas #Christmas #AI #AiFail



**Jac Rayner**  
@GirlFromBlupo

Suivre

Dear Amazon, I bought a toilet seat because I needed one. Necessity, not desire. I do not collect them. I am not a toilet seat addict. No matter how temptingly you email me, I'm not going to think, oh go on then, just one more toilet seat, I'll treat myself.

À l'origine en anglais

00:22 - 6 avr. 2018

73 742 Retweets 404 397 J'aime



1,9 k 74 k 404 k



Tweeter votre réponse

**Rosiebita** ✓ @rosiebita · 6 avr.

En réponse à @GirlFromBlupo

Had the same situation with my mother's burial urn. For months after her death, I got messages from #Amazon saying, "If you liked THAT urn, you might also like THIS one!"

À l'origine en anglais

58 825 8,0 k



**Jac Rayner** @GirlFromBlupo · 6 avr.  
Oh, that is awful! How dreadful for you.

À l'origine en anglais

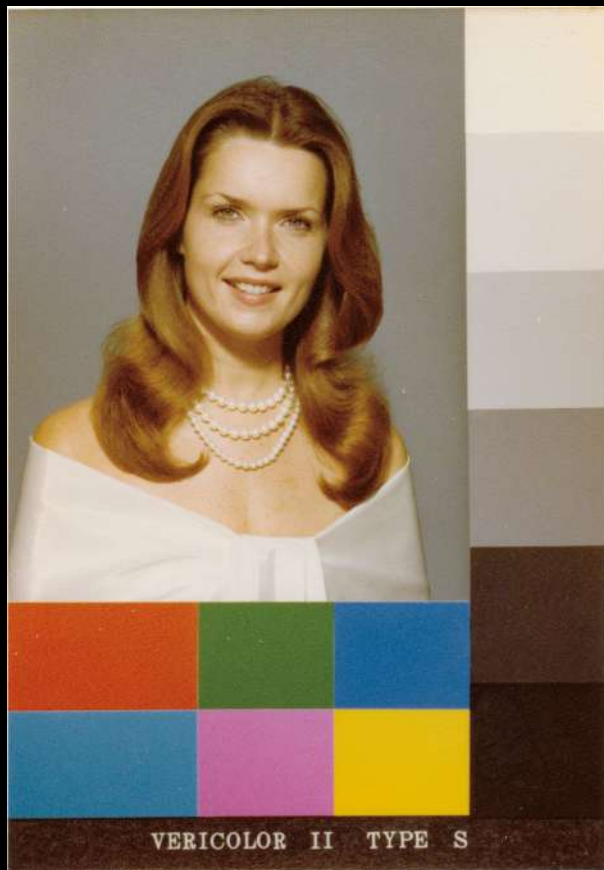
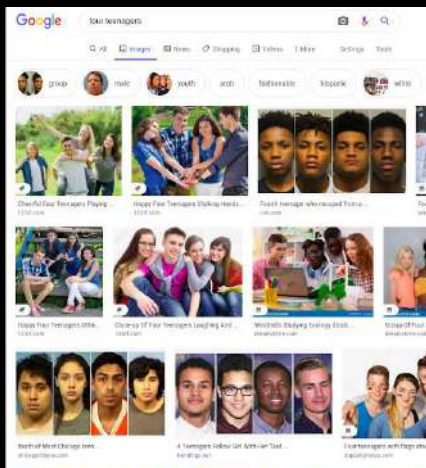
3 13 1,6 k

why are black women so  
 why are black women so angry  
 why are black women so loud  
 why are black women so mean  
 why are black women so attractive  
 why are black women so lazy  
 why are black women so annoying  
 why are black women so confident  
 why are black women so savvy  
 why are black women so insecure

**ALGORITHMS**  
 OF  
**OPPRESSION**

HOW SEARCH ENGINES  
 REINFORCE RACISM

SAFIYA UMOJA NOBLE



	WHITE	AFRICAN AMERICAN
Labeled Higher Risk, But Didn't Re-Offend	23.5%	44.9%
Labeled Lower Risk, Yet Did Re-Offend	47.7%	28.0%

Overall, Northpointe's assessment tool correctly predicts recidivism 61 percent of the time. But blacks are almost twice as likely as whites to be labeled a higher risk but not actually re-offend. It makes the opposite mistake among whites: They are much more likely than blacks to be labeled lower risk but go on to commit other crimes. (Source: ProPublica analysis of data from Broward County, Fla.)

# des biais sociaux et culturels dans les données |



contextualisation  
des images de  
requins marteaux,  
truites et homards,  
[malevé | 2019](#)

google translate :

Ő szép.	Ő gyereket nevel.
Ő okos.	Ő zenél.
Ő olvas.	Ő politikus.
Ő épít.	Ő sok pénzt keres.
Ő varr.	Ő süteményt süt.
Ő tanít.	Ő professzor.
Ő kutat.	Ő titkárnő.



# vers des datasets plus justes ?

- vocabulaire aux concepts stagnants
- concepts non-visuels
- absence de diversité d'images

solutions: filtrer, balance, flouter ?

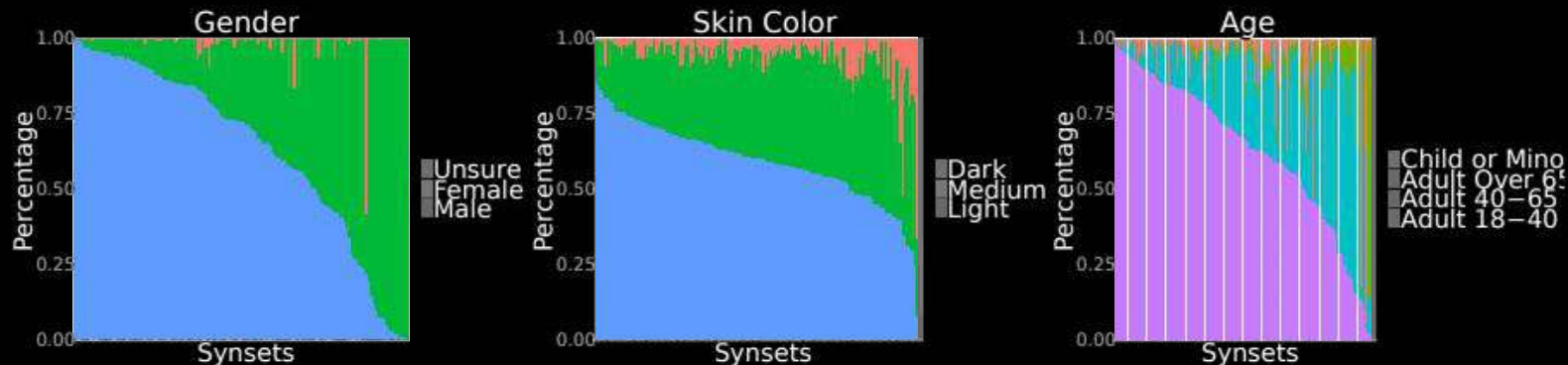


Figure 4: The distribution of demographic categories across the 139 safe and imageable synsets which contain at least 100 images. The size of the different color areas reveal the underrepresentation of certain groups.

## Towards Fairer Datasets: Filtering and Balancing the Distribution of the People Subtree in the ImageNet Hierarchy

Kaiyu Yang  
Princeton University  
Princeton, NJ  
kaiyuy@cs.princeton.edu

Klint Qinami  
Princeton University  
Princeton, NJ  
kqinami@cs.princeton.edu

Li Fei-Fei  
Stanford University  
Stanford, CA  
feifeili@cs.stanford.edu

Jia Deng  
Princeton University  
Princeton, NJ  
jiadeng@cs.princeton.edu

Olga Russakovsky  
Princeton University  
Princeton, NJ  
olgarus@cs.princeton.edu

### ABSTRACT

Computer vision technology is being used by many but remains representative of only a few. People have reported misbehavior of computer vision models, including offensive prediction results and lower performance for underrepresented groups. Current computer

work, we focus on one dataset, ImageNet [18], which has arguably been the most influential dataset of the modern era of deep learning in computer vision. ImageNet is a large-scale image ontology collected to enable the development of robust visual recognition models. The dataset spearheaded multiple breakthroughs in object recognition [35, 45, 20]. In addition, the feature representation

# “a dataset is a worldview”

hannah davis 2020























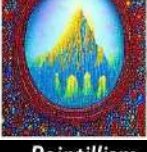

“une base de données est une vision du monde. outre les préjugés spécifiques de leurs créateurs, les ensembles de données encodent également les valeurs culturelles générales de l'époque à laquelle ils ont été créés.

nos ensembles de données peuvent et doivent avoir des dates d'expiration. les problèmes de classification et de taxonomie (...) [sont] plus importants et plus effrayants en ce qui concerne l'apprentissage automatique, et ce pour deux raisons principales :

- les ensembles de données d'apprentissage automatique influencent les modèles qui entrent en production dans le monde réel et ont un impact substantiel et immédiat sur notre vie quotidienne, (...) de nos résultats de recherche aux perspectives d'emploi en passant par les scores de crédit (...).
- les chercheurs en apprentissage automatique, les journalistes, les gouvernements, les entreprises et d'autres parties concernées ont perpétué cette image de l'ia comme une boîte noire, totalement ininterprétable et, surtout, responsable d'elle-même. cela permet de faire passer tous ces problèmes de préjugés et de visions du monde pour “juste l'ia”. c'est un récit exceptionnellement commode pour ceux qui utilisent l'apprentissage automatique à des fins nuisibles.”

childbirth	anger	0	
childbirth	anticipation		0
childbirth	disgust	0	
childbirth	fear	0	
childbirth	joy	0	
childbirth	negative		0
childbirth	positive		0
childbirth	sadness	0	
childbirth	surprise		0
childbirth	trust	0	

# empoisonner l'ia en protégeant ses œuvres |

		Dog	Car	Handbag	Hat	Fantasy art	Cubism
Clean Model (SD-XL)							
Poisoned Model (SD-XL)	50 poison samples						
	100 poison samples						
	300 poison samples						
		Cat	Cow	Toaster	Cake	Pointillism	Anime
		Destination Concept A					

*This Tool Could Protect Artists From A.I.-Generated Art That Steals Their Style*

Artists want to be able to post their work online without the fear "of feeding this monster" that could replace them.

univ. de chicago :

« think of glaze like a new dimension of the art, one that ai models see but humans do not (like uv light or ultrasonic frequencies), except the dimension itself is hard to locate / compute / reverse engineer »

# jeudi prochain

23 oct 2025

le cours portera sur :

6 | actualité des  
controverses de l'ia