

Algorithmes, big data,  
réseaux de neurones,  
intelligence artificielle...

Cercle Zététique du Languedoc Roussillon 12 février 2022

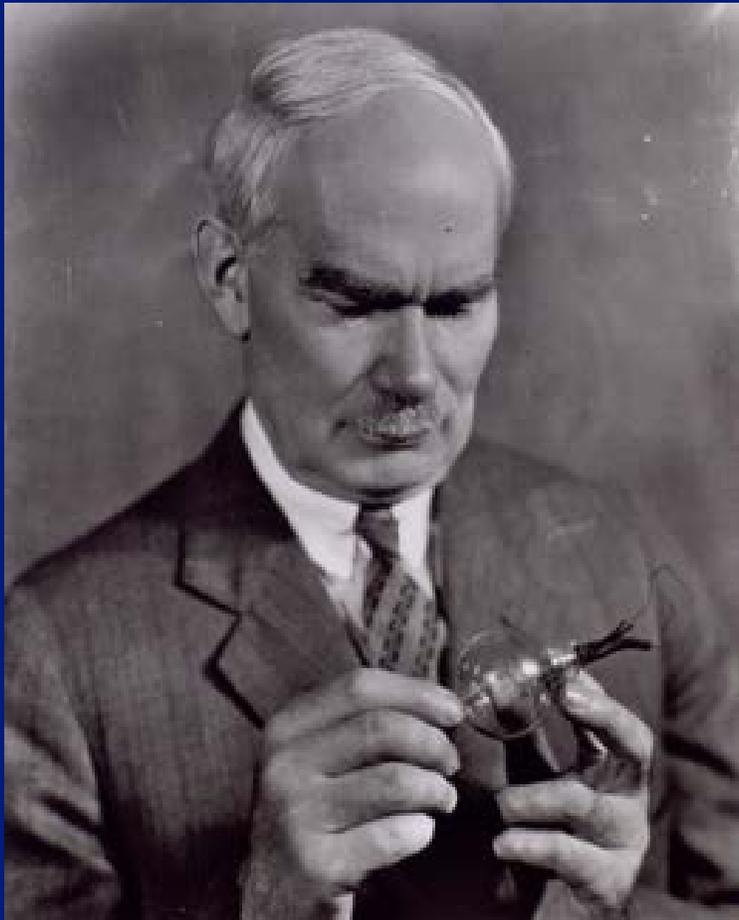
*« On ne peut comprendre le monde numérique dans sa totalité  
sans comprendre suffisamment ce qu'est son cœur  
informatique »*

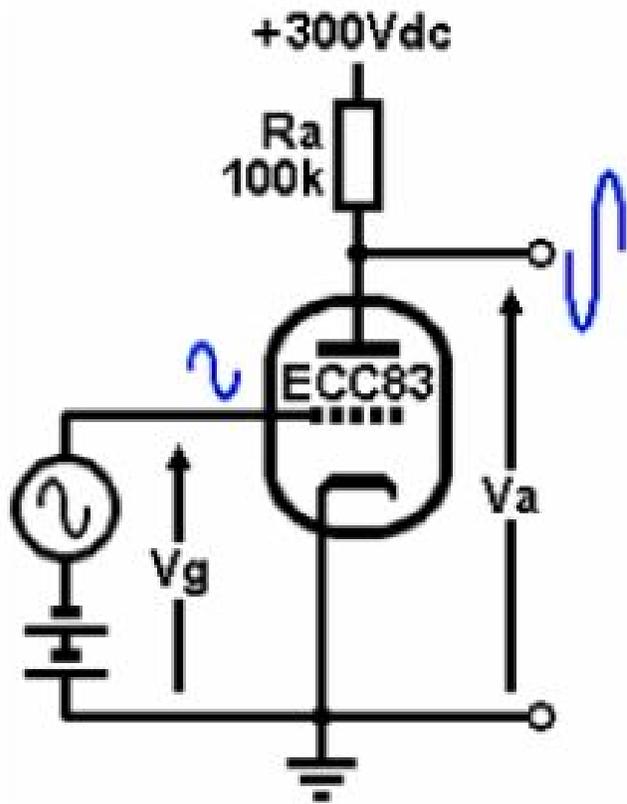
Gérard Berry professeur émérite au collège de France.

Médaille d'Or du CNRS

# Naissance de l'électronique

## Lee de Forest, la triode (1906)

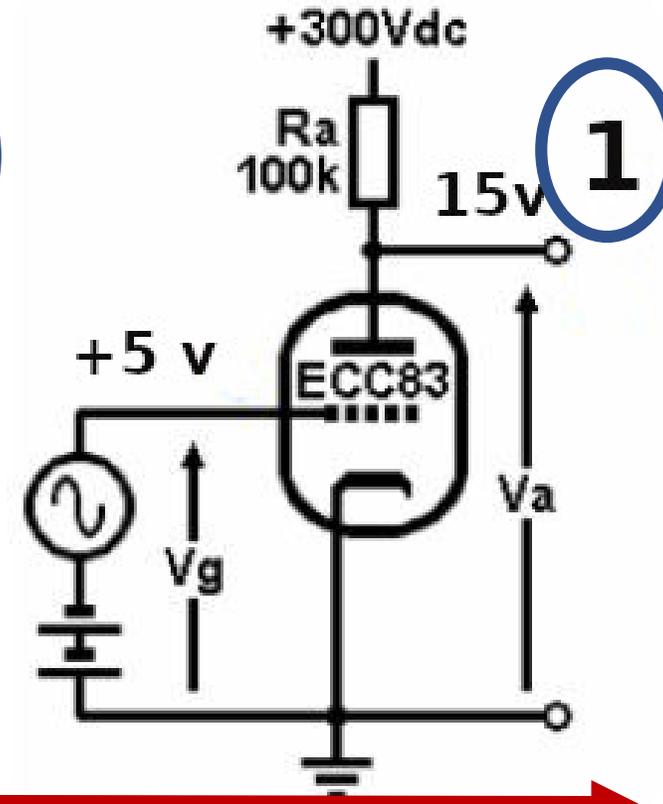
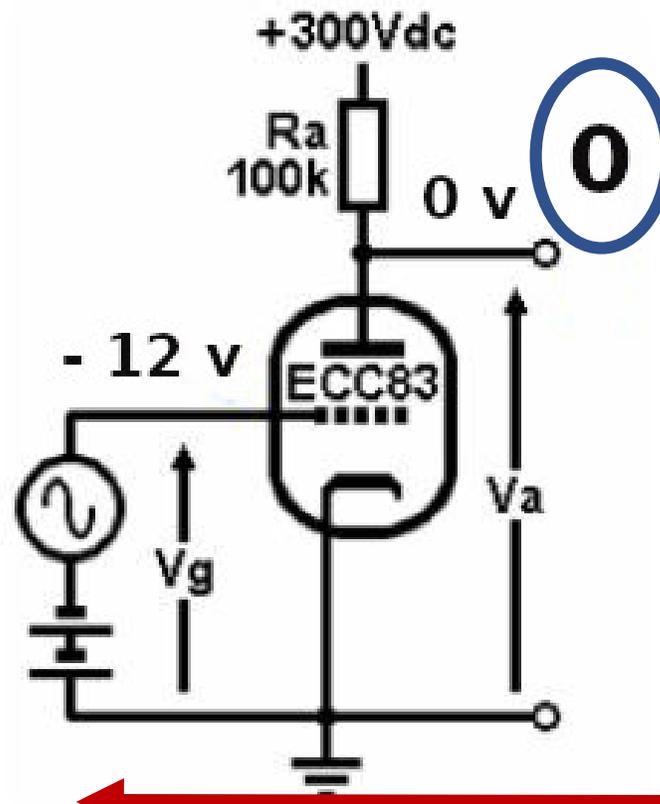




### Amplification

Radio, TV, hi fi,  
radar, sonar etc.

Électronique analogique



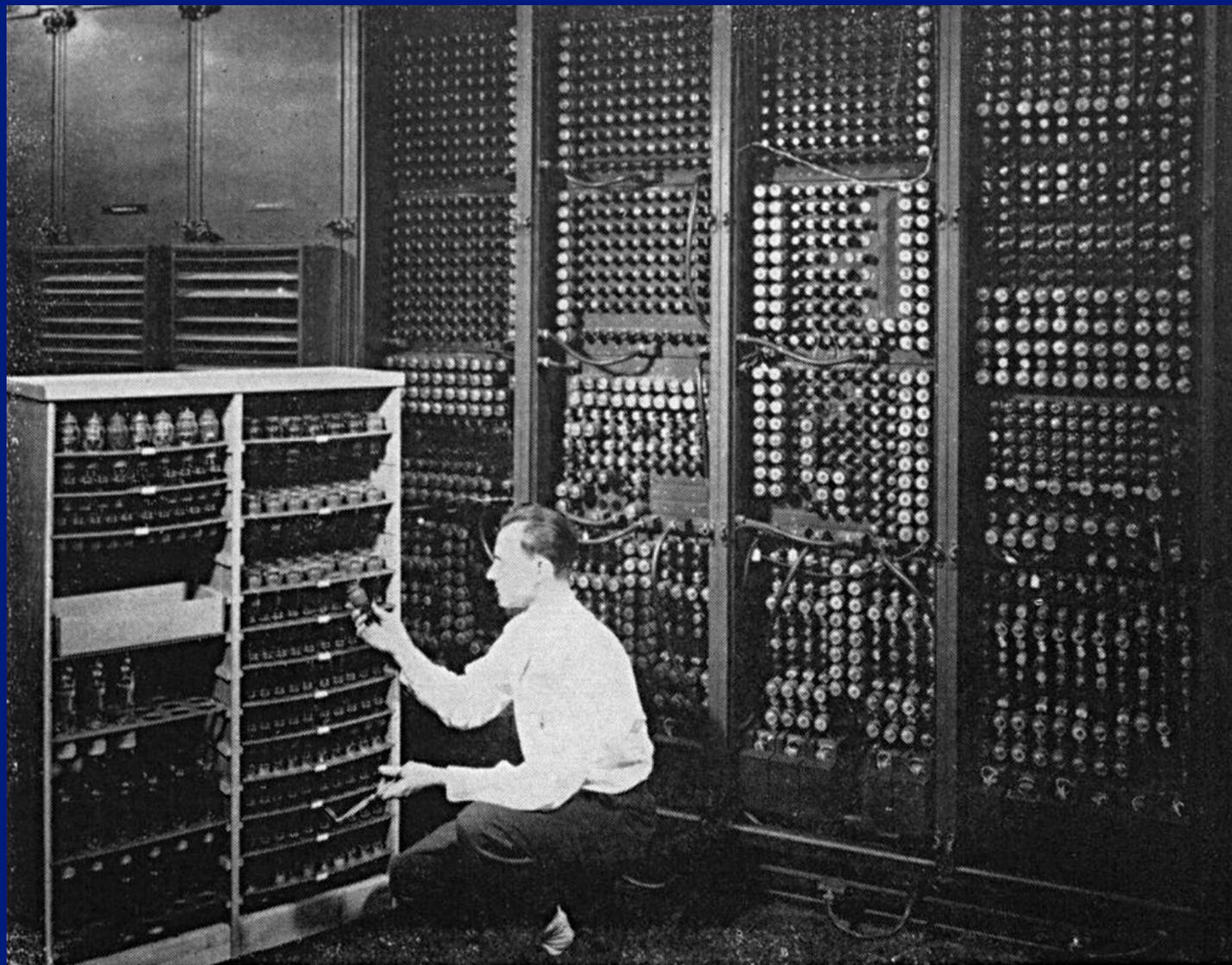
### Commutation

Electronique numérique



# Le premier ordinateur l'Eniac, 1945

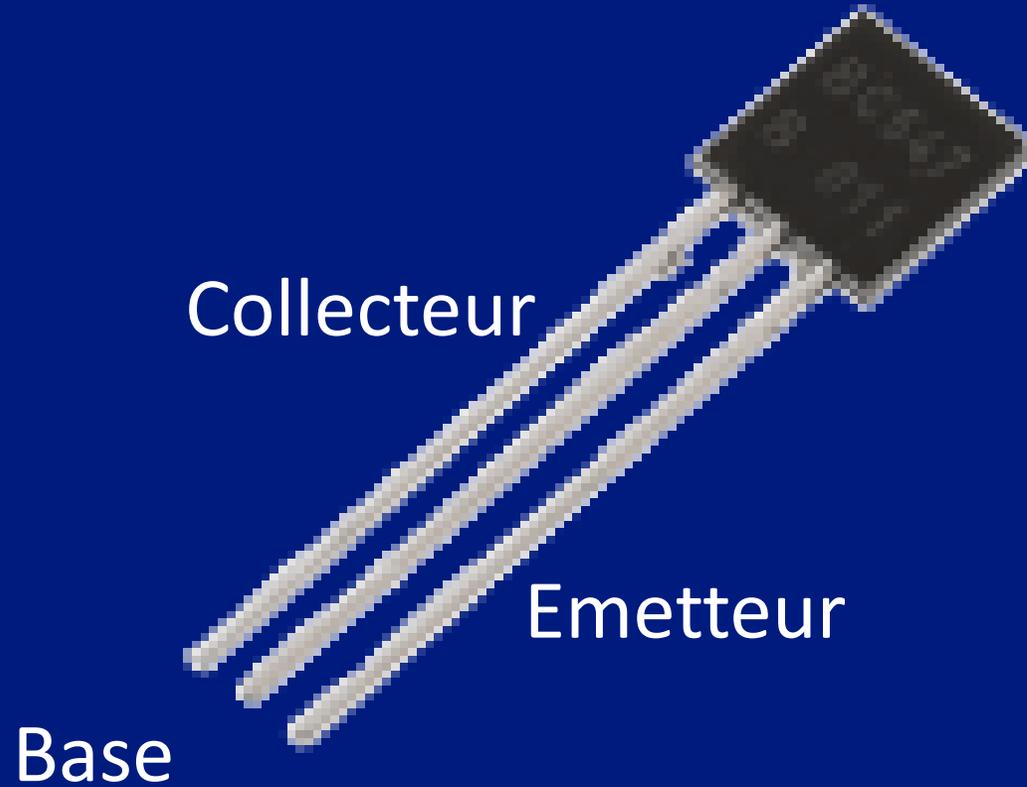
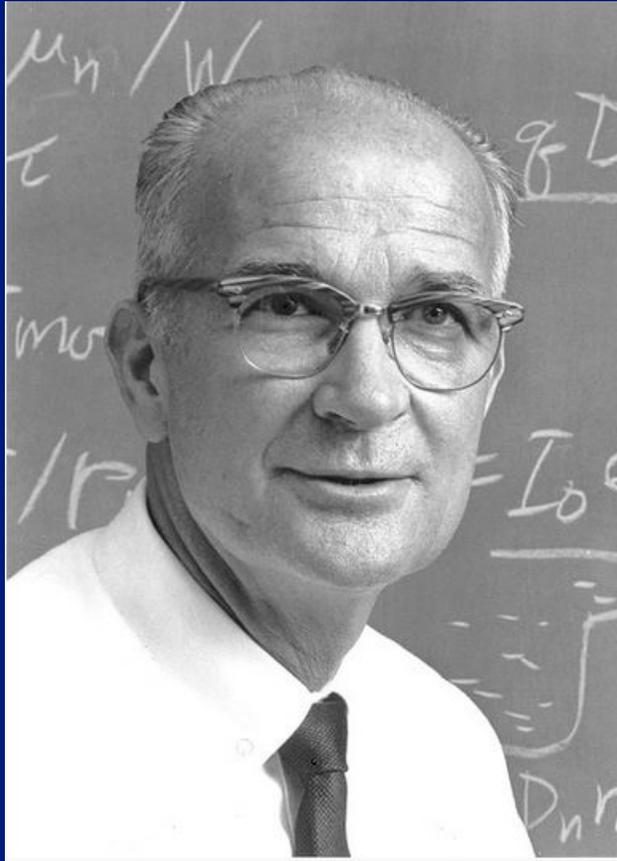
18 000 triodes



Replacing a bad tube meant checking among ENIAC's 19,000 possibilities.

# William Shockley (Nobel de physique)

## Le transistor 1946

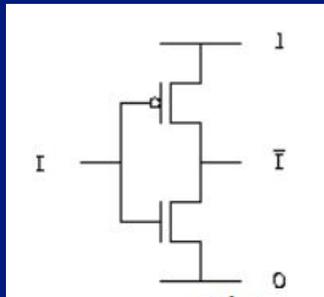
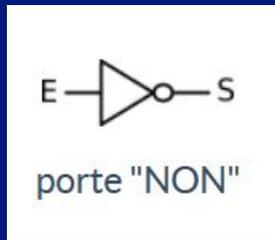


# Les données alphanumériques

- Octets 8 bits.
  - 8 bits permettent 256 combinaisons de 00000000 à 11111111
- Suffisant pour codifier tous les alphabets romains majuscules et minuscules avec leurs accents et les chiffres de 0 à 9

Table de vérité porte « non »

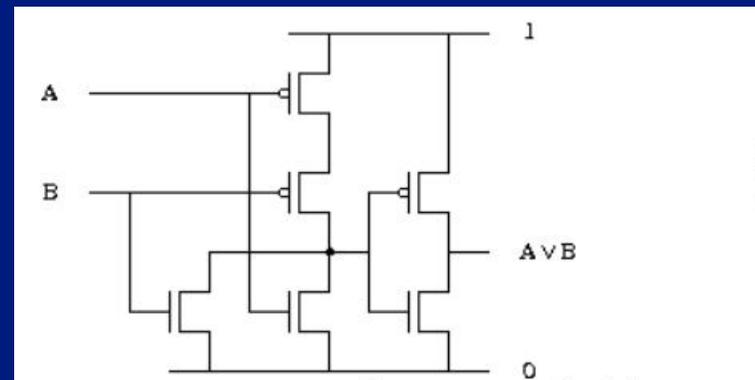
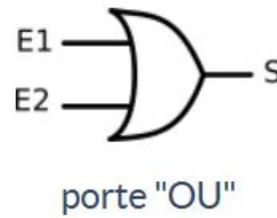
E (Entrée)	S (Sortie)
1	0
0	1



2 transistors

Table de vérité porte "OU" :

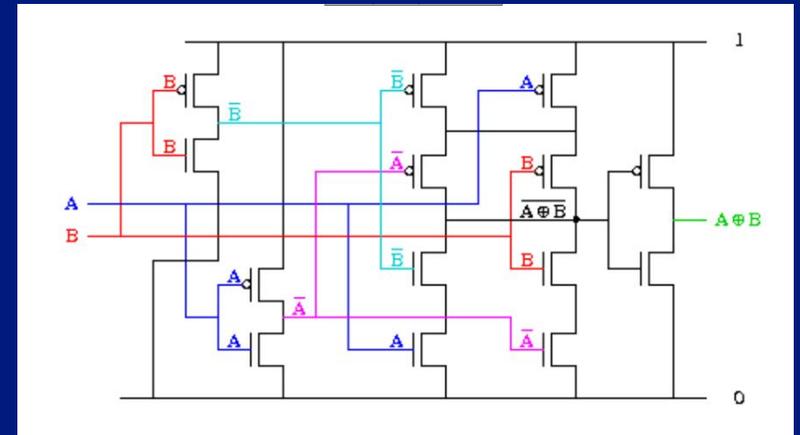
E1	E2	S
0	0	0
0	1	1
1	0	1
1	1	1



6 transistors

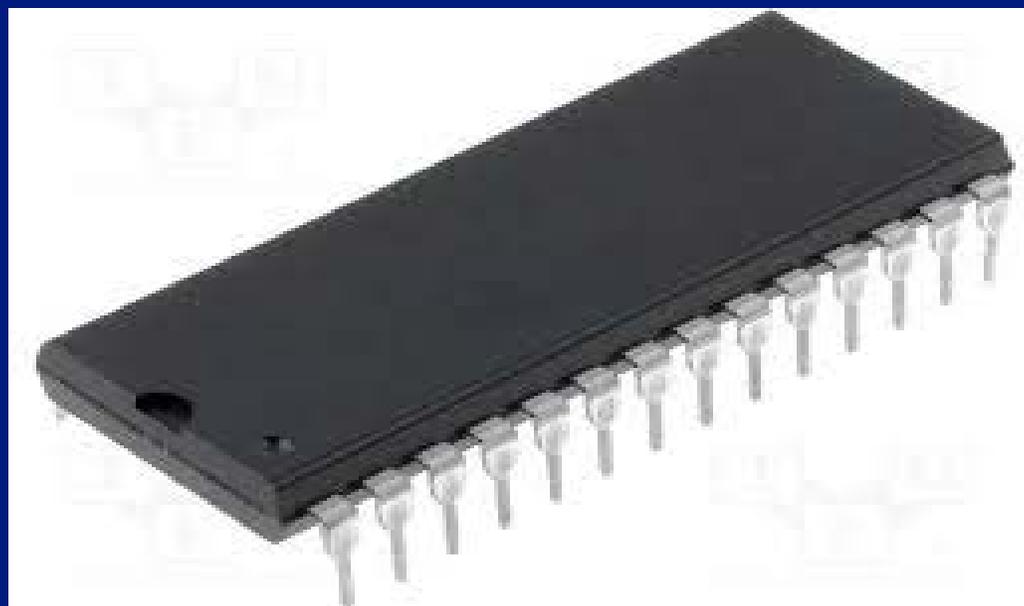
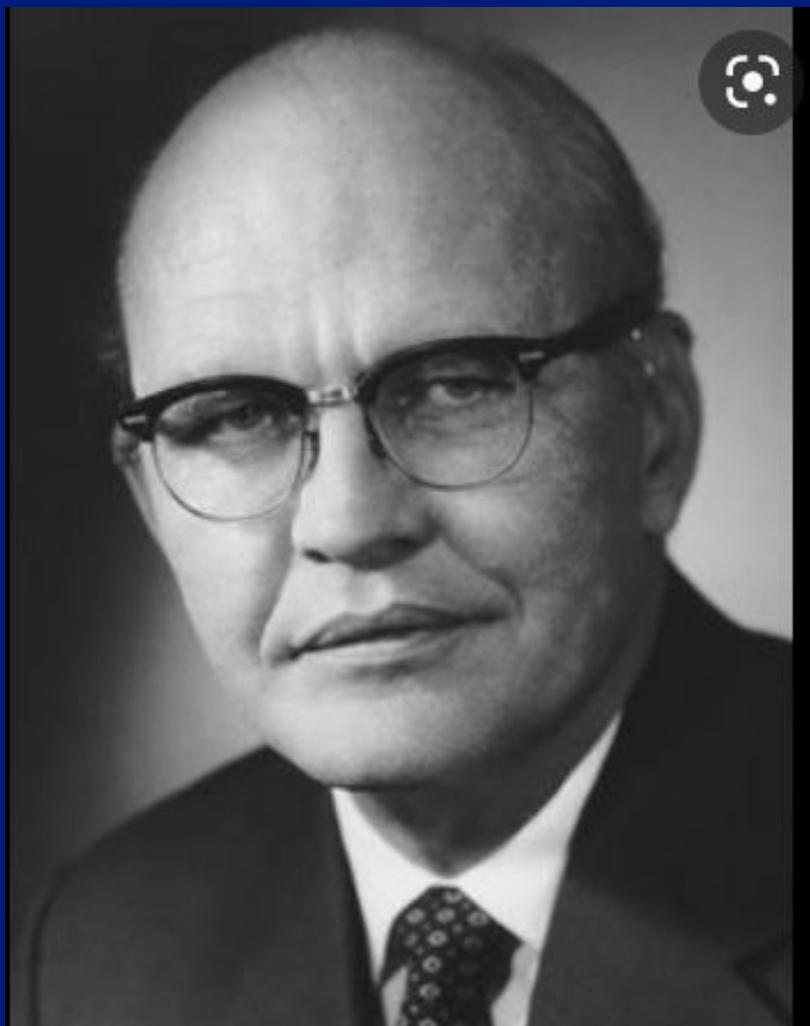
Table de vérité porte "XOR" :

E1	E2	S
0	0	0
0	1	1
1	0	1
1	1	0



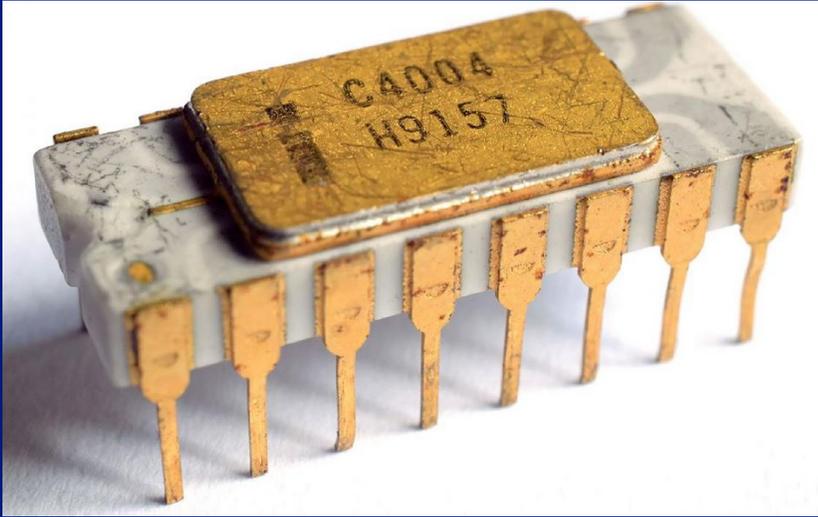
14 transistors

# Jack Kilby, les circuits intégrés 1958



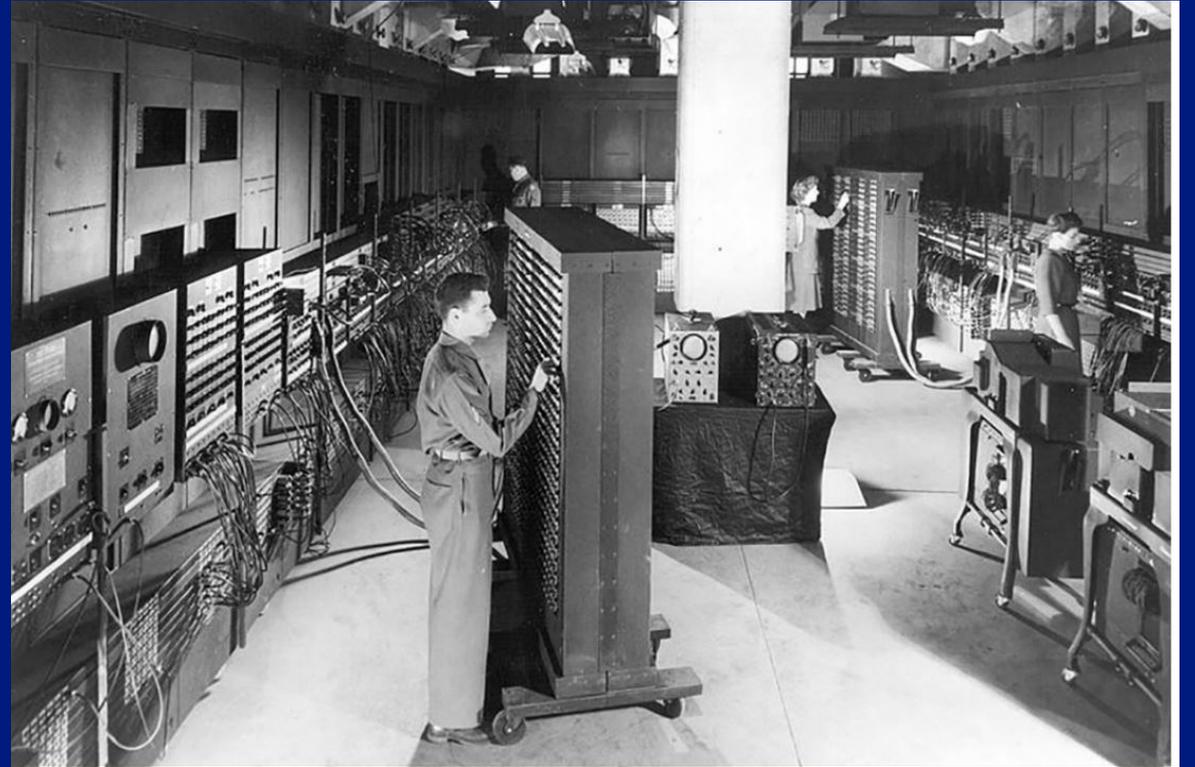
En 1965 les meilleurs circuits intégrés comportaient 64 transistors.

# Microprocesseur Intel 4004, 1971



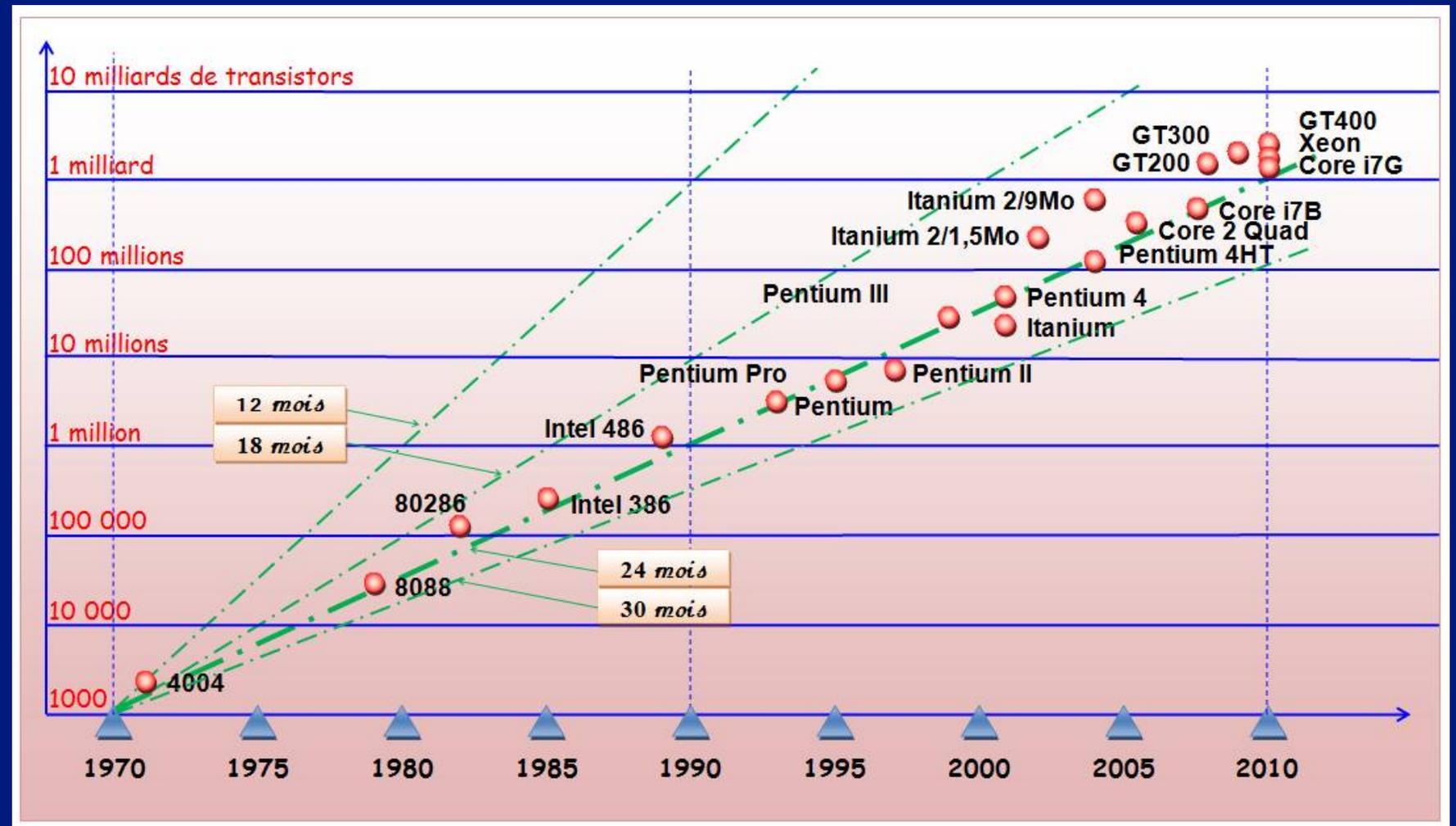
2300 transistors 92 600 opérations par seconde sur qq cm<sup>2</sup>

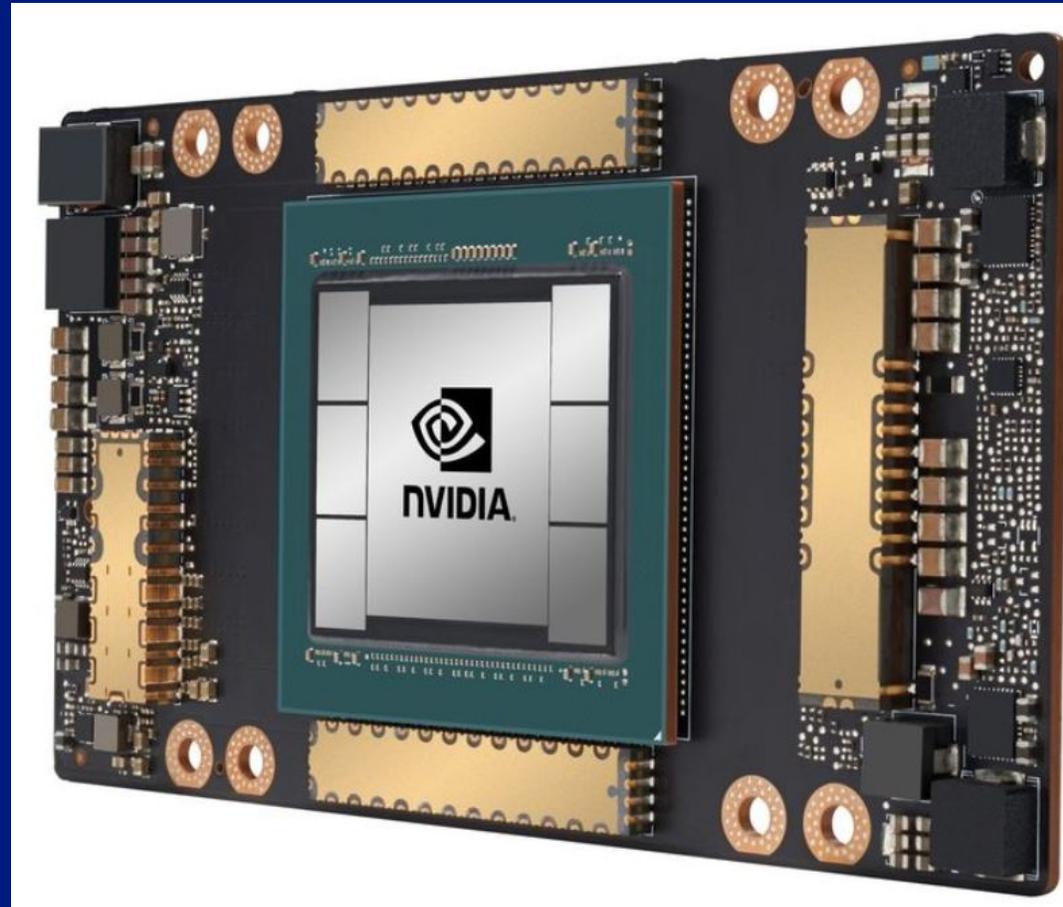
Une puissance de calcul comparable à celle de l'Eniac, 167 m<sup>2</sup>, 30 tonnes.



# Loi de Moore, doublement du nombre de transistors par circuit intégré tous les 2 ans

- 1971  
Microprocesseur Intel 4004  
9200 transistors
- 1990  
Microprocesseur Intel Pentium  
1 million de transistors
- 2010  
Microprocesseur Intel Core i7G  
1 milliard de transistors
- 2021  
Processeur graphique Nvidia A100  
53 milliards de transistors





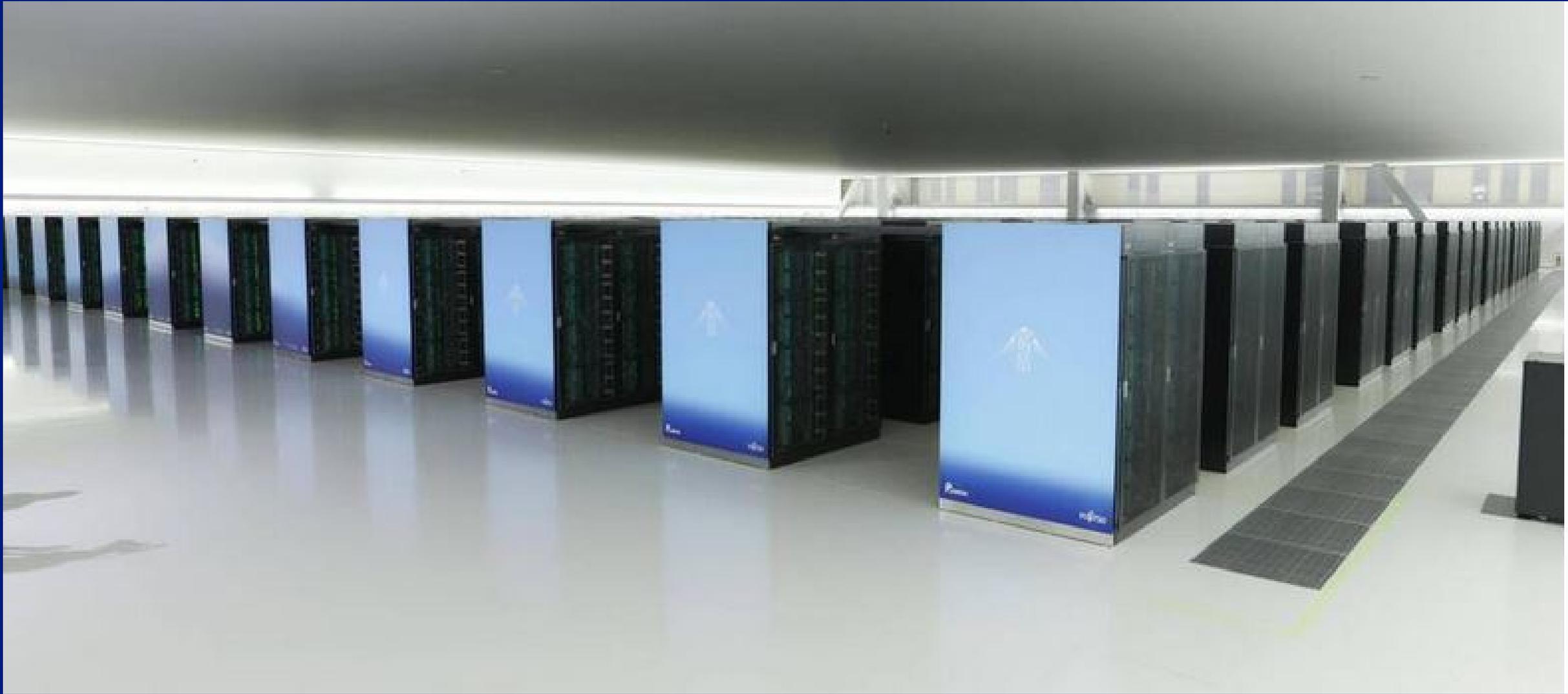
2021. Processeur graphique « NVIDIA A 100 ».  
53 milliards de transistors, 5 millions de milliards d'opérations par  
seconde.

# Des gains en performance inégalés



Un chasseur à réaction des années 70 : 2000 km/h

Un chasseur avec un même gain technologique que les circuits intégrés parcourrait aujourd'hui 330 fois la distance terre soleil (150 millions de km) en une heure !



Fugaku, superordinateur du centre de calcul Riken à Kobe  
442 millions de milliards d'opérations par seconde

# Les algorithmes

# Algorithme

- Un algorithme est une suite d'opérations qui se termine à un moment donné et qui délivre un résultat
- Les recettes de cuisine sont des algorithmes...

# Algorithme gâteau marron-chocolat

- Les données :

- Crème de marron = 500g
- Chocolat = 100g
- Œufs = 3
- Casserole = 1
- Moule à cake = 1
- Four = 150 degrés
- Temps de cuisson = 45 mn
- Temps au frigo = 2h

- Les instructions :
- 1 Beurrer moule à cake
- 2 Battre œufs
- 3 Mélanger œufs, chocolat, crème de marron, beurre dans casserole
- 4 Chauffer casserole
- 5 **Chocolat fondu ? Si OUI, Instruction suivante, si NON, Instruction 4 à nouveau**
- 6 Verser dans moule à cake
- 7 Chauffer
- 8 **Temps de cuisson > 45 minutes ? Si OUI, Instruction suivante, si NON, instruction 7 à nouveau**
- 9 Mettre au frigo
- 10 **Temps au frigo < 2h ? Si OUI, instruction 9 à nouveau, si NON, exécuter 11**
- 11 Servir !

# Algorithmes de traitement de données structurées alphanumériques.

- De 1950 à la fin des années 1990, les ordinateurs traitent exclusivement des données structurées.
- Informatique de gestion
  - Banques, assurances
  - Entreprises : comptabilité clients, fournisseurs, commandes, factures...
  - Bureautique avec les PC
- Calcul scientifique. Météo, nucléaire, espace, aéronautique, médecine



# Années 2000, les données non structurées



Un livre de 500 pages  
1 million d'octets



Une photo  
4 millions d'octets



Du son  
1 millions d'octets par minute



Une image animée, vidéo  
2 milliards d'octets par heure

# Années 2000, l'explosion des données

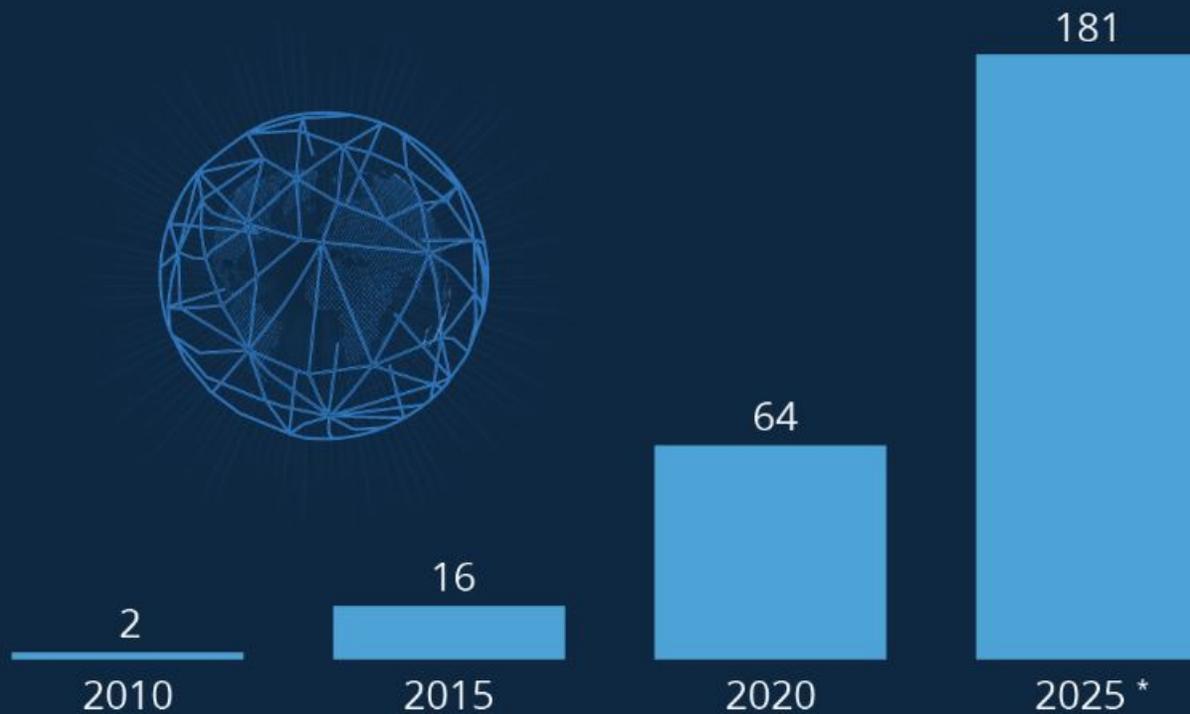
## Le big data

- 30 000 heures de vidéos mises en ligne par heure sur Youtube
  - 306 milliards de mails envoyés par jour
  - 5 900 tweets par seconde sur Twitter
  
  - 110 millions de milliards de données sont stockées sur les serveurs de Google
- 
- <https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde/>
  - <https://www.blogdumoderateur.com/chiffres-google/>

Un zettaoctet vaut mille milliards de milliards d'octets.

## Le Big Bang du Big Data

Estimation du volume de données numériques créées ou répliquées par an dans le monde, en zettaoctets



Un zettaoctet équivaut à mille milliards de gigaoctets.

\* Prévvision en date de mars 2021.

Sources : IDC, Seagate, Statista



# Les algorithmes de l'intelligence artificielle (IA)

# Les réseaux de neurones artificiels



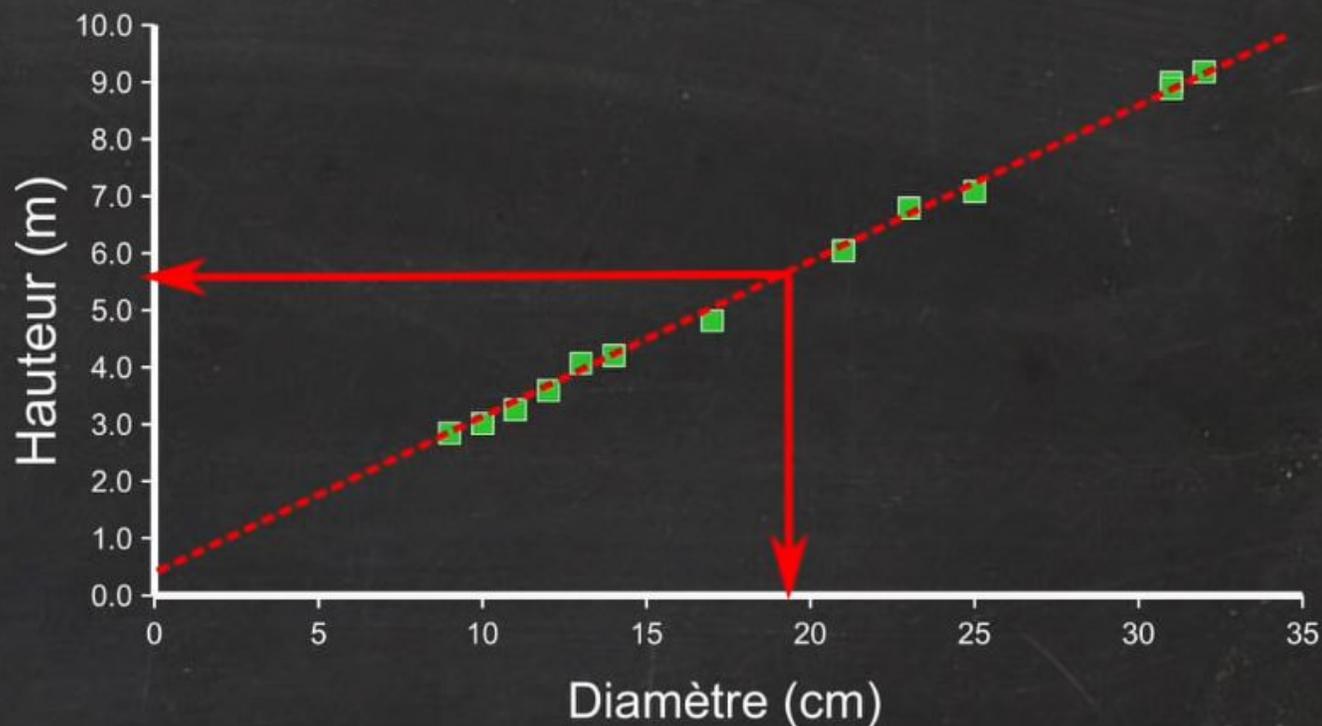
Forêt de hêtres

Diamètre (cm) | Hauteur (m)

10	2.8
17	4.6
14	4.0
9	2.6
21	5.8
25	6.9
31	8.8
12	3.4
23	6.6
32	9.0
31	8.7
11	3.0
13	3.8

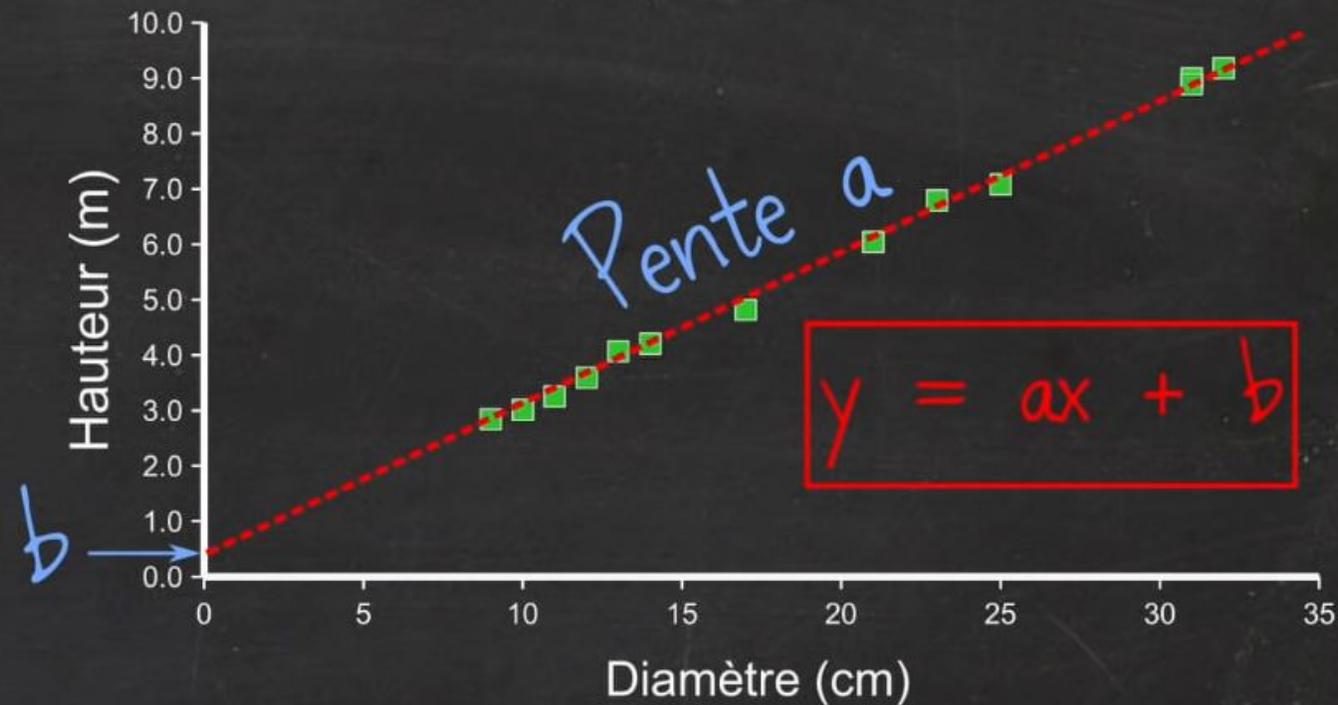
19

?



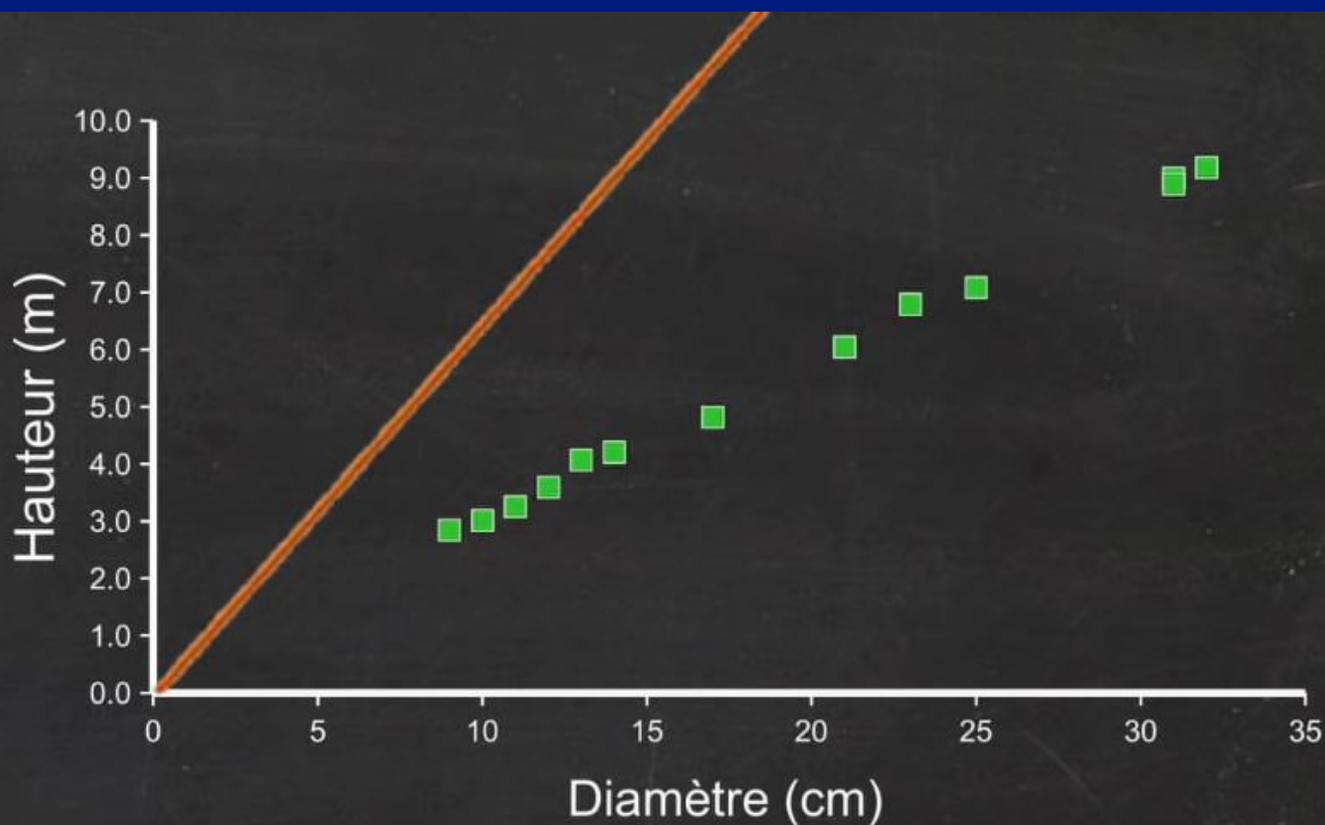
Diamètre (cm) | Hauteur (m)

10	2.8
17	4.6
14	4.0
9	2.6
21	5.8
25	6.9
31	8.8
12	3.4
23	6.6
32	9.0
31	8.7
11	3.0
13	3.8

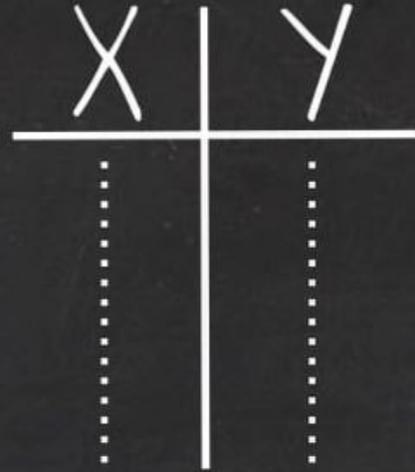


Diamètre (cm)	Hauteur (m)
---------------	-------------

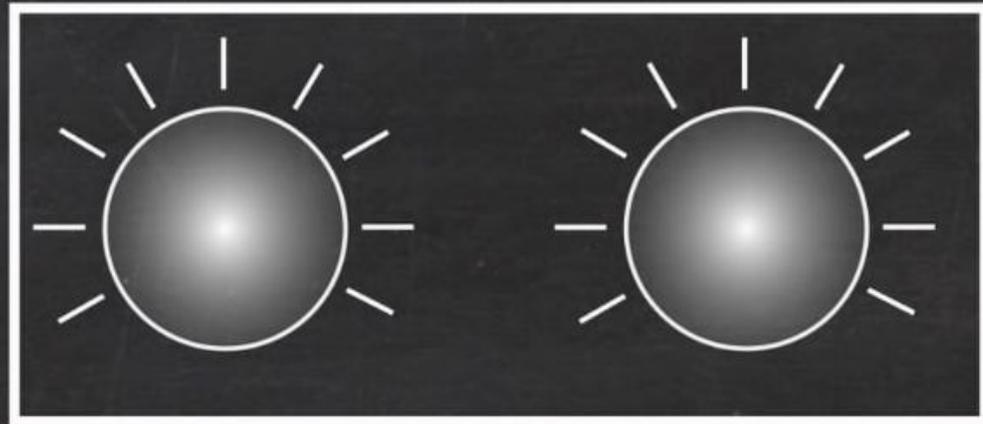
10	2.8
17	4.6
14	4.0
9	2.6
21	5.8
25	6.9
31	8.8
12	3.4
23	6.6
32	9.0
31	8.7
11	3.0
13	3.8



# Machine learning

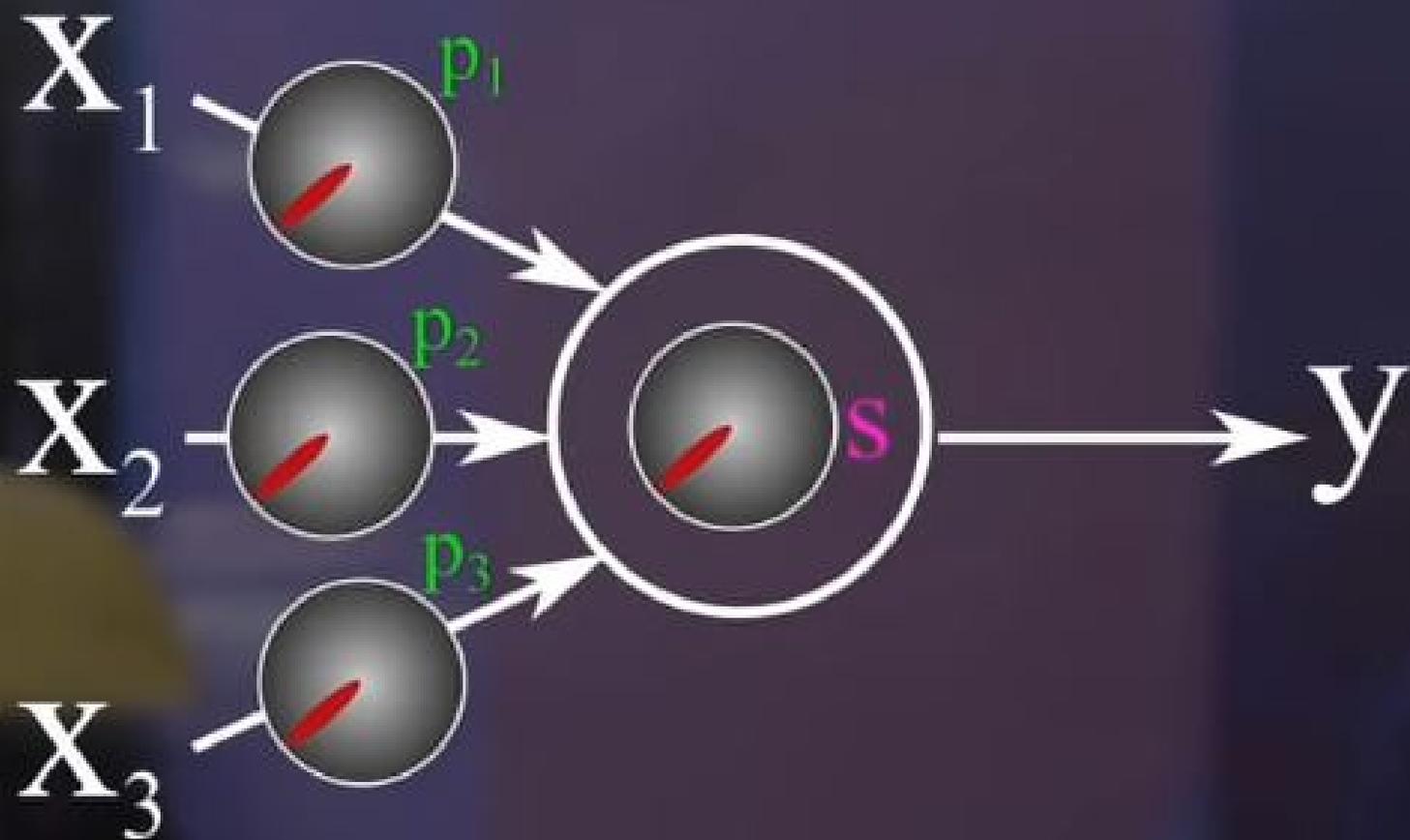


Entrée  
X



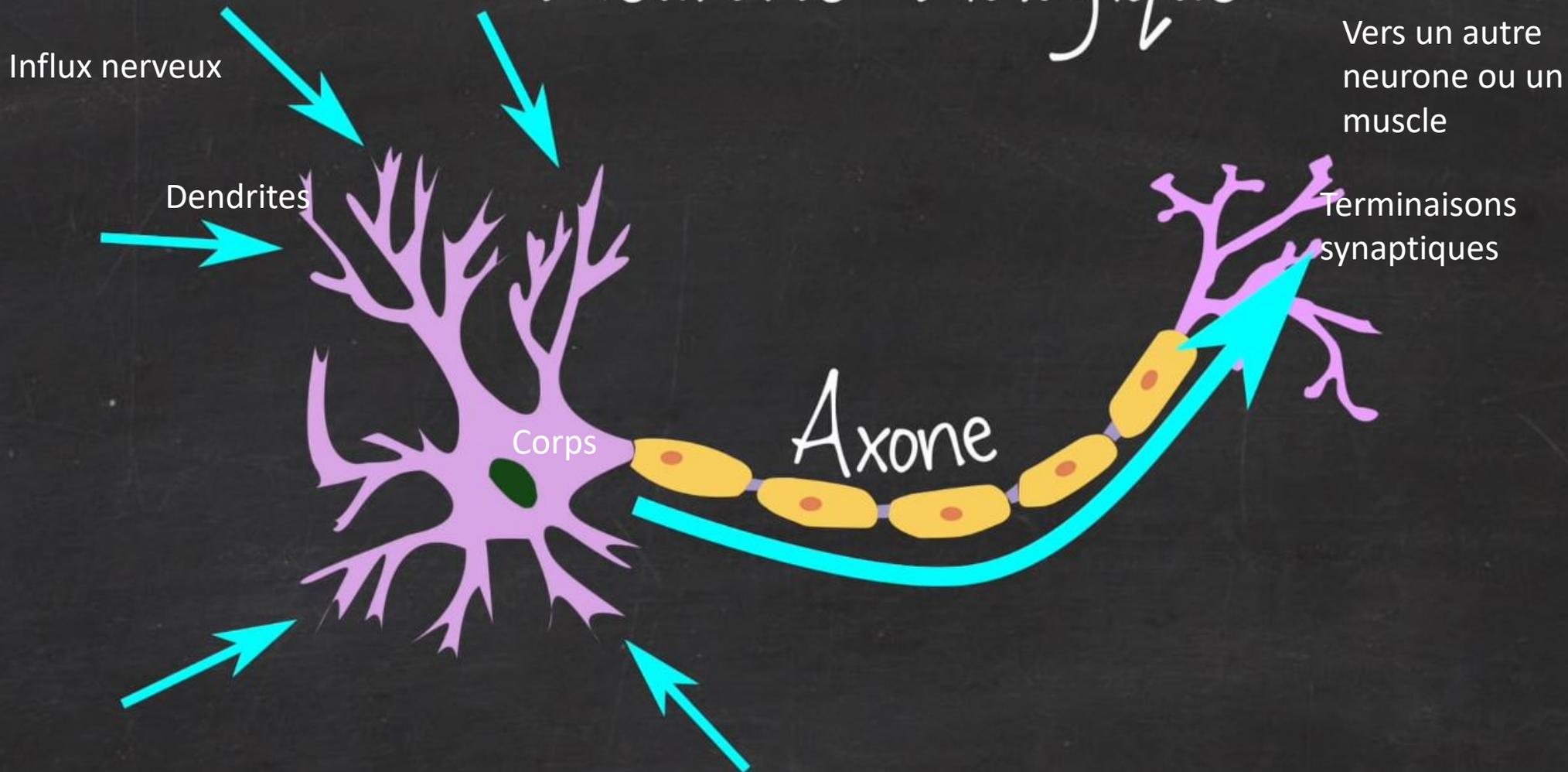
Sortie  
Y

Phase d'apprentissage

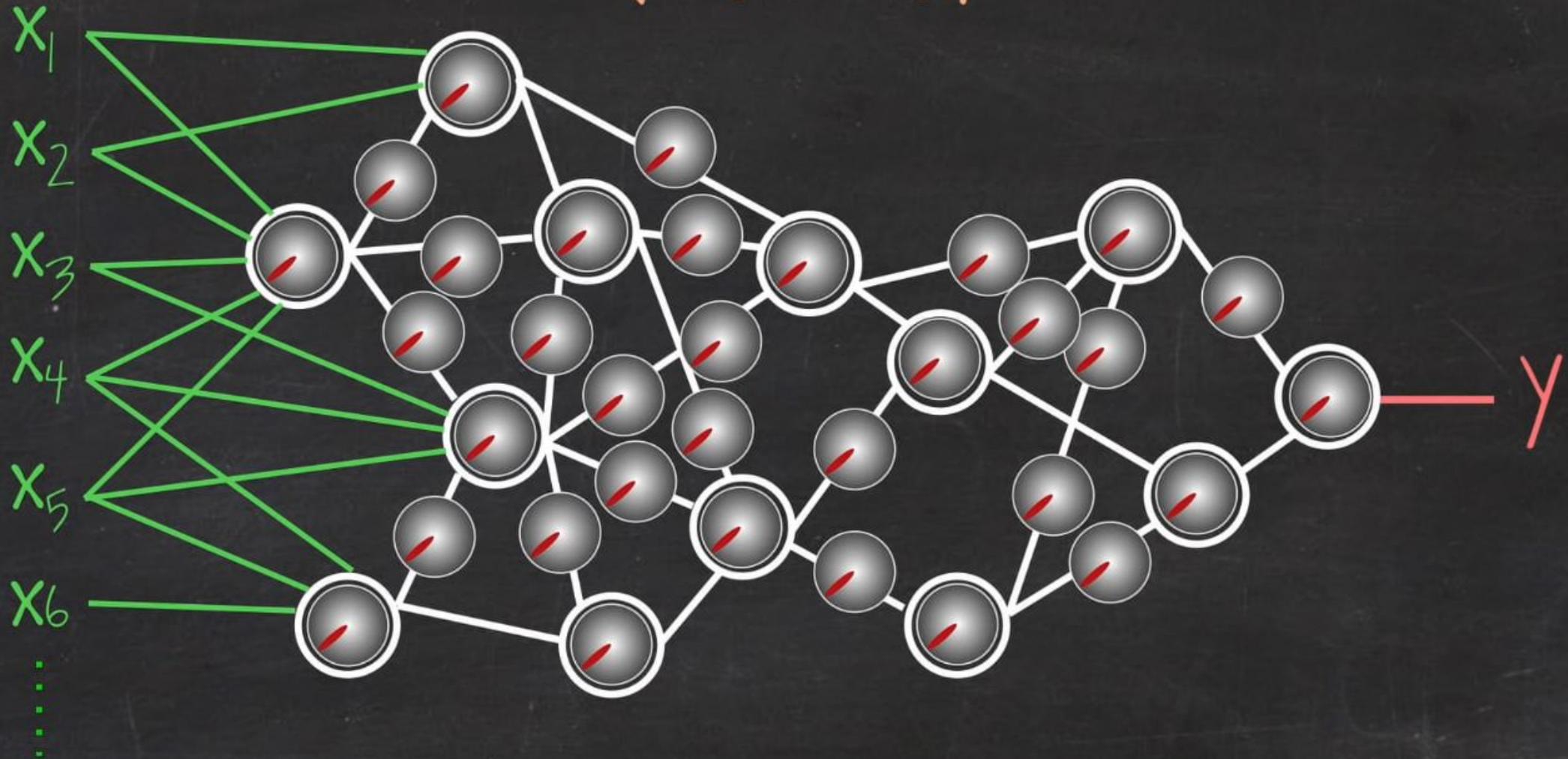


$$\begin{aligned} p_1 X_1 + p_2 X_2 + p_3 X_3 > s ? &\longrightarrow y=1 \\ < s ? &\longrightarrow y=0 \end{aligned}$$

# Neurone biologique

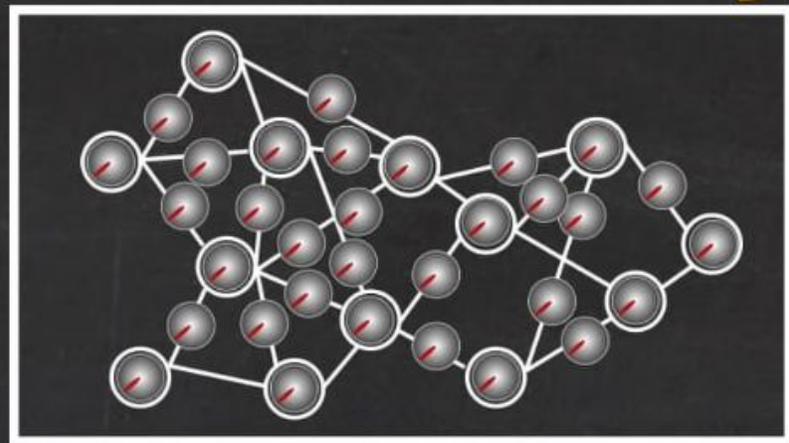


# Réseau de neurones (artificiel)



# Phase d'apprentissage

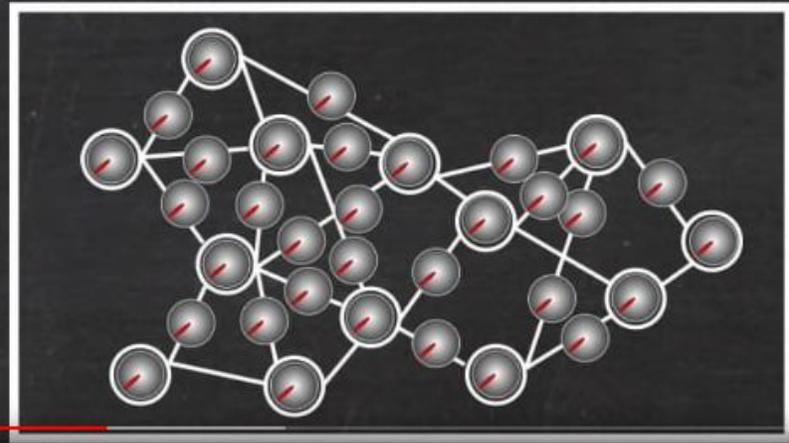
Entrées  
 $X$



Sorties  
 $Y$

# Phase de prédiction

Nouvelle  
Entrée  
 $X$



Prédiction  
de  $Y$

# Reconnaissance de véhicule



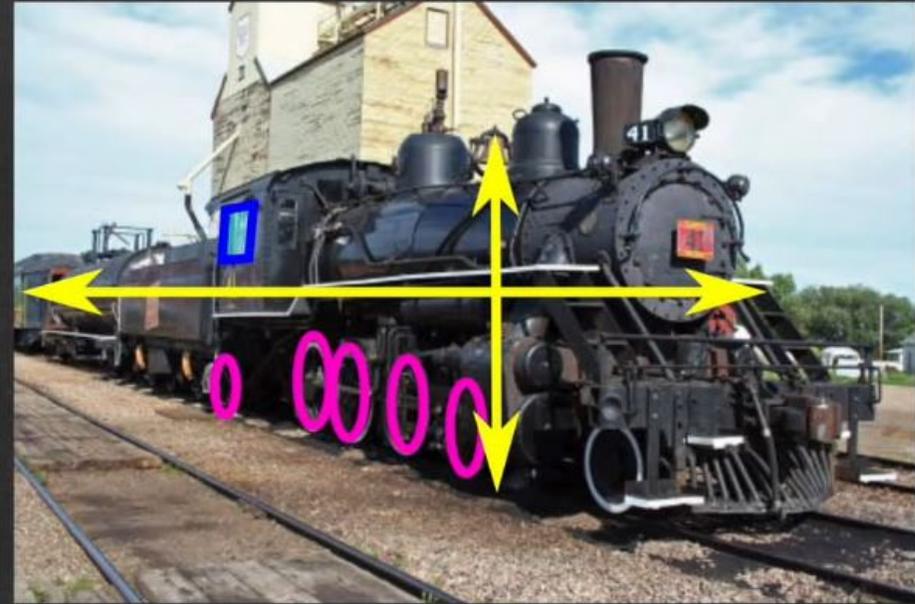
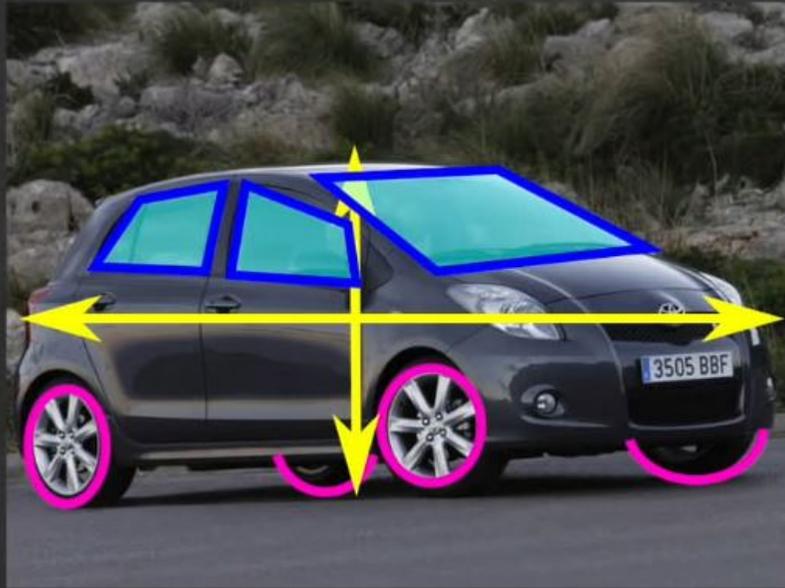
Algorithme



Voiture ?  
Bus ?  
Train ?  
Moto ?

...

# Caractéristiques essentielles de l'image



# Abstraction de l'image



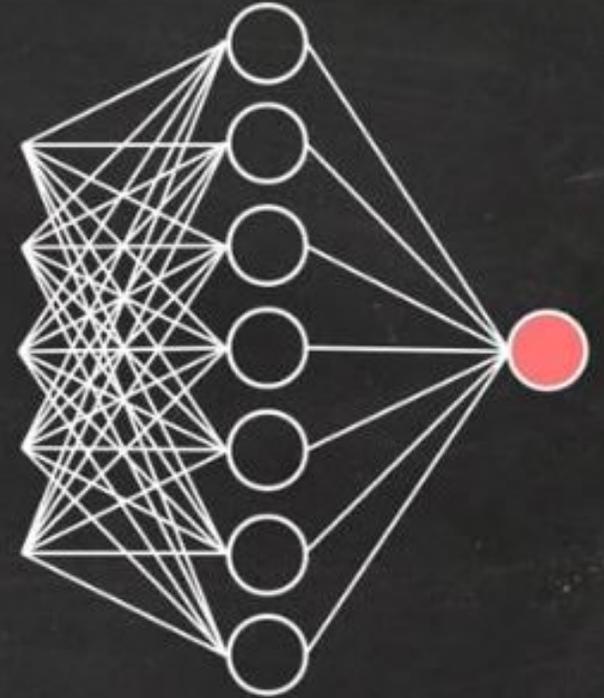
Image brute  
(millions de pixels)

Algorithme  
d'extraction de  
caractéristiques

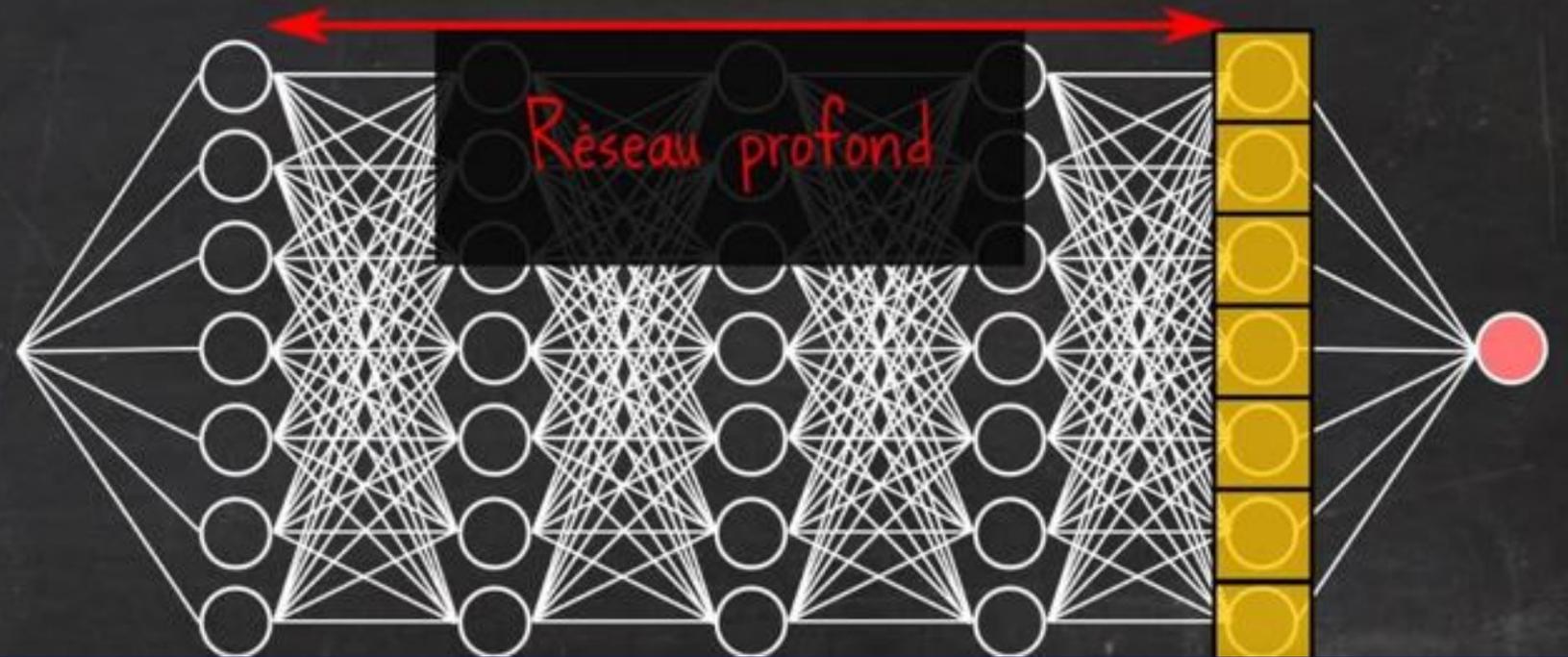
L'algorithme d'extraction  
est conçu et alimenté par  
des humains. C'est long .



Liste de  
caractéristiques  
(quelques dizaines)

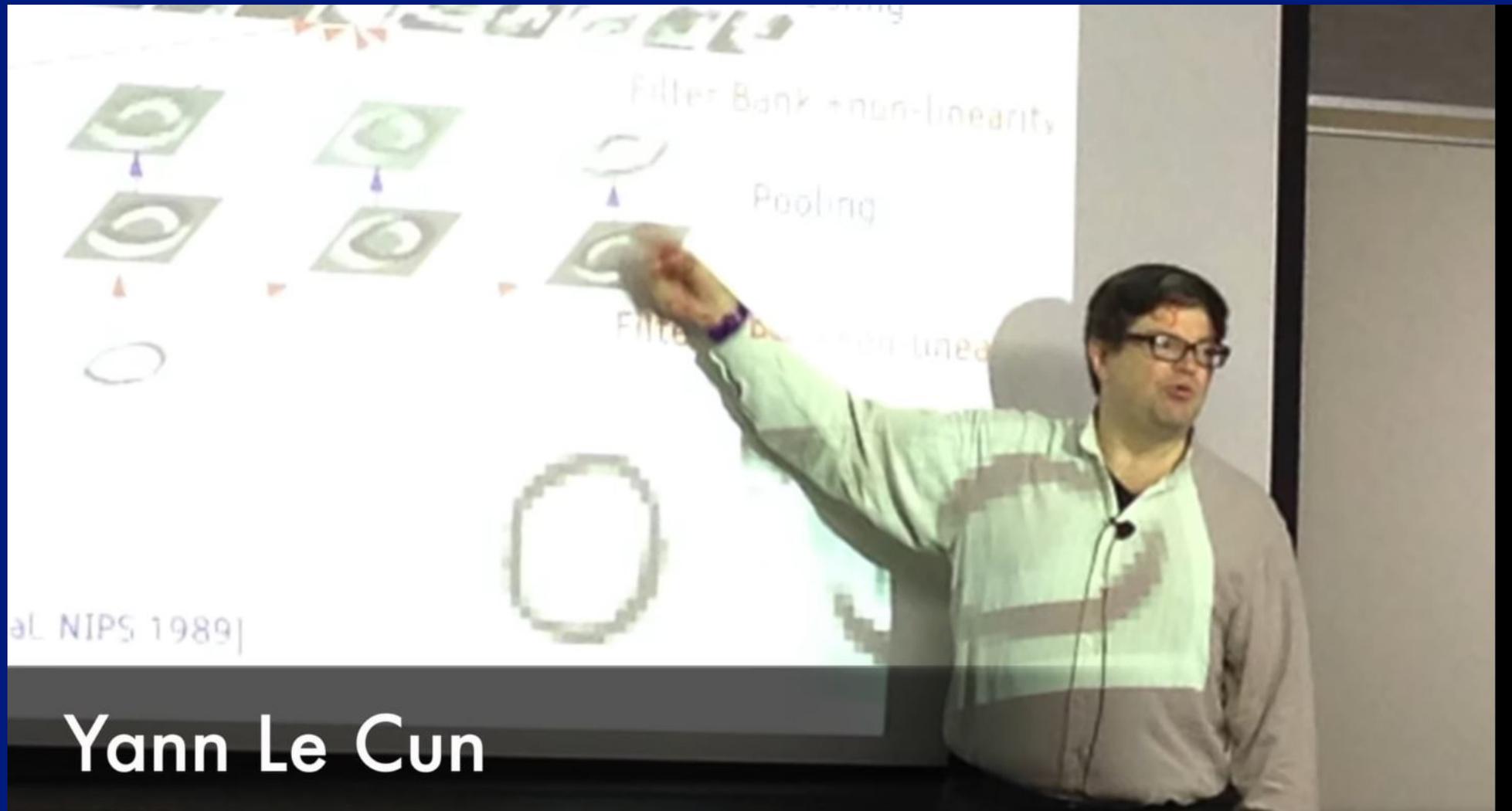


# Le deep learning



Le réseau de deep learning trouve lui-même les caractéristiques essentielles de l'image

# Yann Le Cun inventeur du Deep learning dans les années 90's.



Yann Le Cun

# Large Scale Visual Recognition Challenge

## 2010

1.	NEC	28%
2.	XRCE	34%
3.	ISIL	45%
4.	UCI	47%
5.	Hminmax	54%

## 2011

1.	XRCE	26%
2.	Uv A	31%
3.	ISI	36%
4.	NII	50%

## 2012

1.	SuperVision	16%
2.	ISI	26%
3.	VGG	27%
4.	XRCE	27%
5.	Uv A	30%

## 2013

1.	Clarifai	12%
2.	NUS	13%
3.	ZeilerFergus	13%
4.	A.Howard	13%
5.	OverFeat	14%

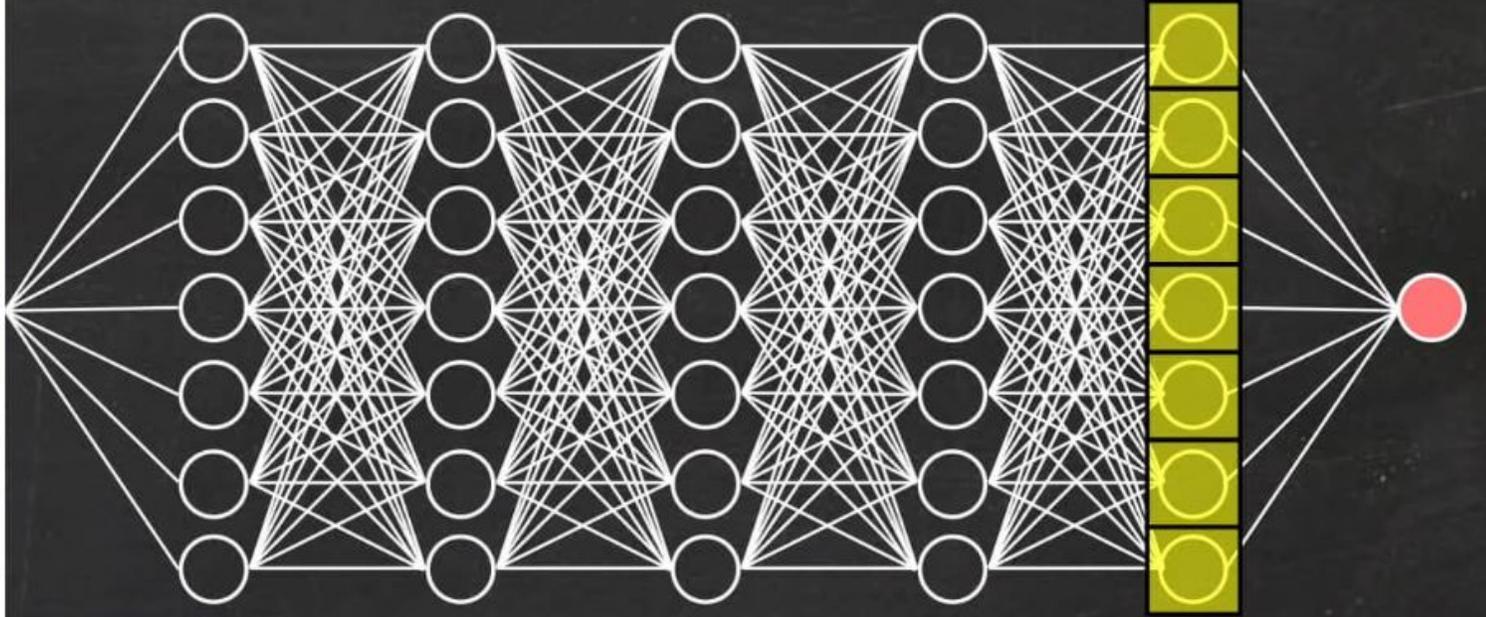
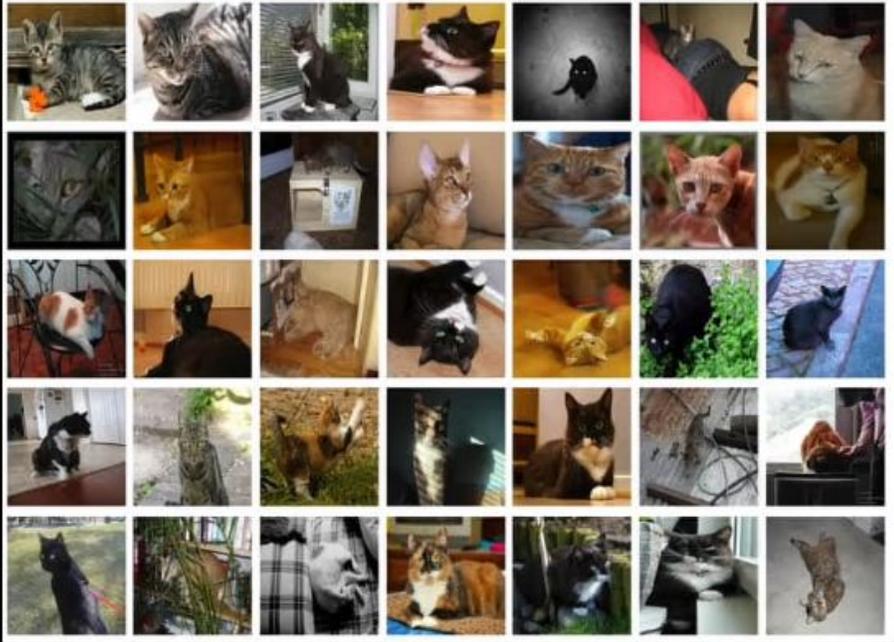
2012 et  
1013 les  
gagnants  
du  
concours  
sont des  
réseaux de  
deep  
learning

# Yann Le Cun, embauché par Facebook en 2013. Directeur de son labo d'IA

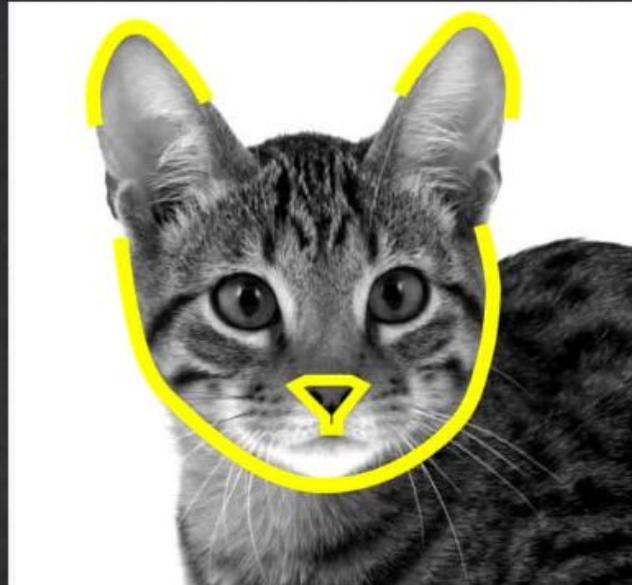


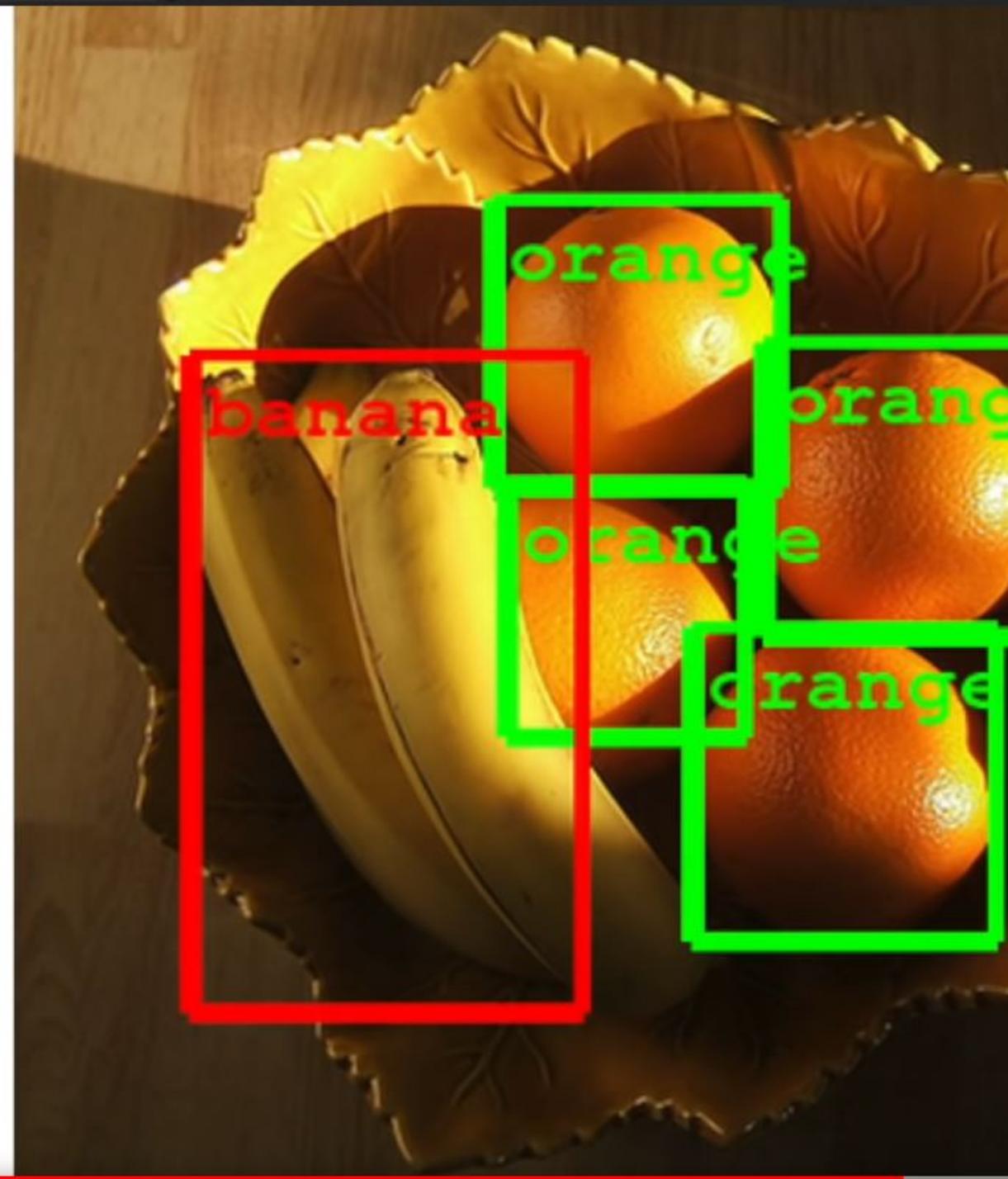
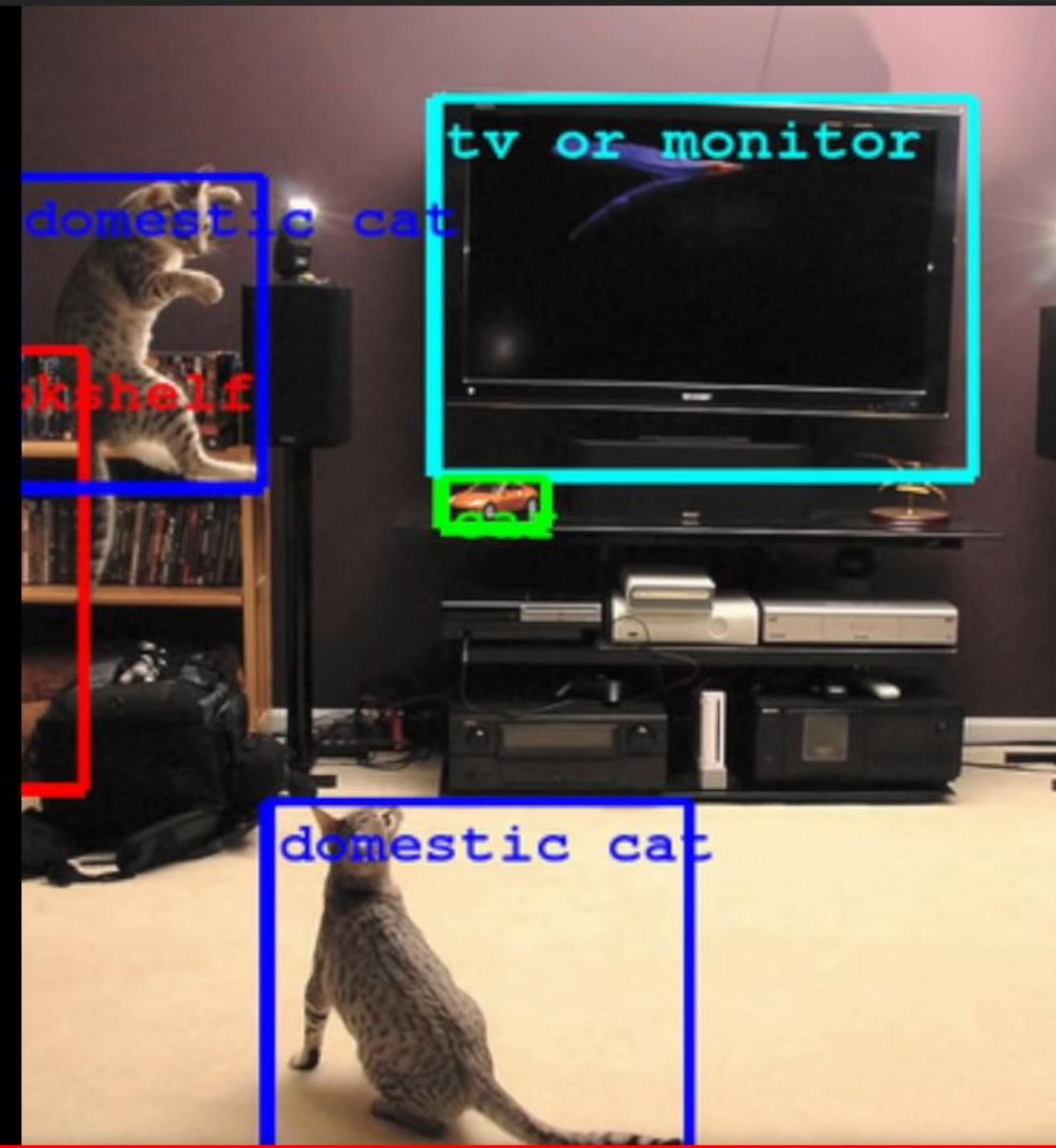
Facebook, Google, Amazon, Baidu...  
Tous les géants du net embauchent  
des spécialistes du deep learning

Les réseaux de deep learning qui fonctionnaient très mal dans les années 90 se sont mis à fonctionner grâce à la puissance de calcul accrue des ordinateurs, à la disponibilité de bases d'images considérables (Imagenet, 15 millions d'images décrites) et à l'amélioration des algorithmes.



Pour reconnaître un chat avec 95% de réussite il a fallu utiliser 100 000 images de chats contextualisées (nuit, sofa, mur etc.)





# Des algorithmes de DL peuvent analyser une scène et la décrire par une phrase

**Describes without errors**



**A person riding a motorcycle on a dirt road.**

**Describes with minor errors**



**Two dogs play in the grass.**

# Modèles génératifs

Valeurs des caractéristiques essentielles

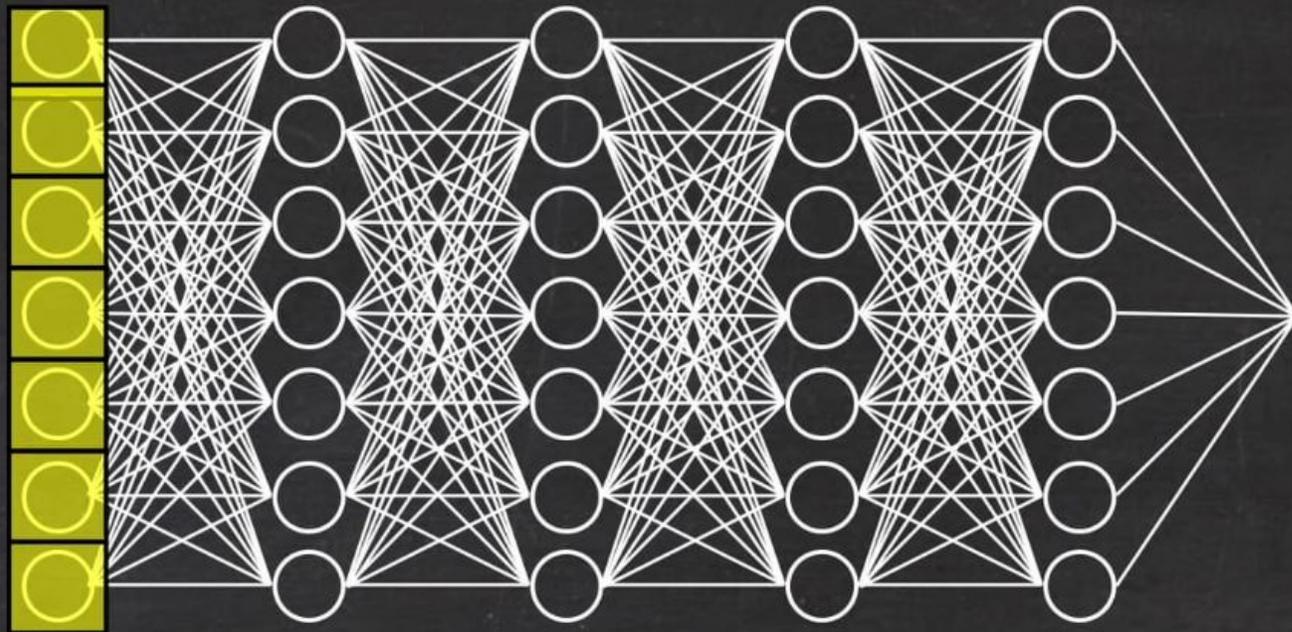


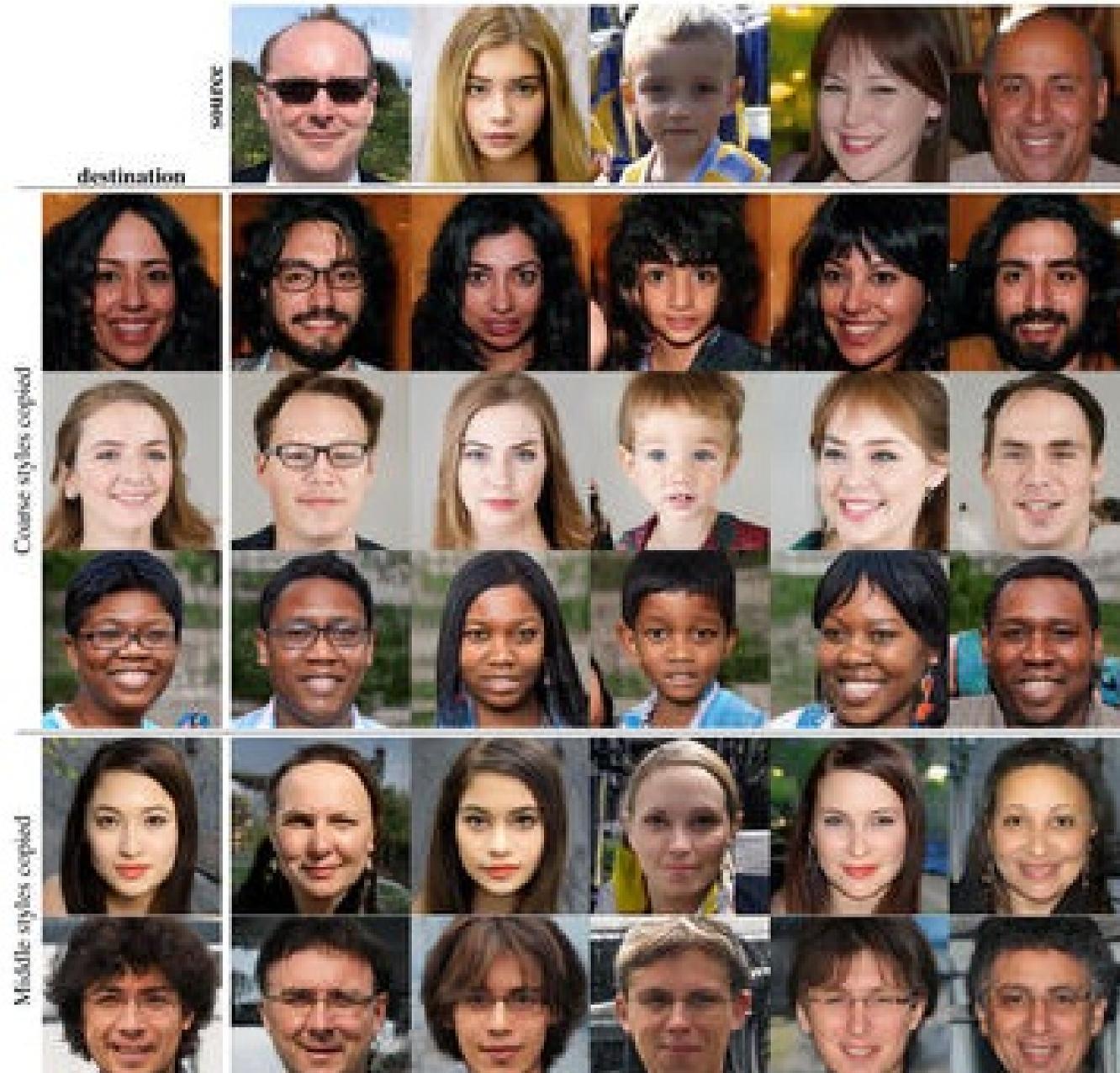
Image inventée



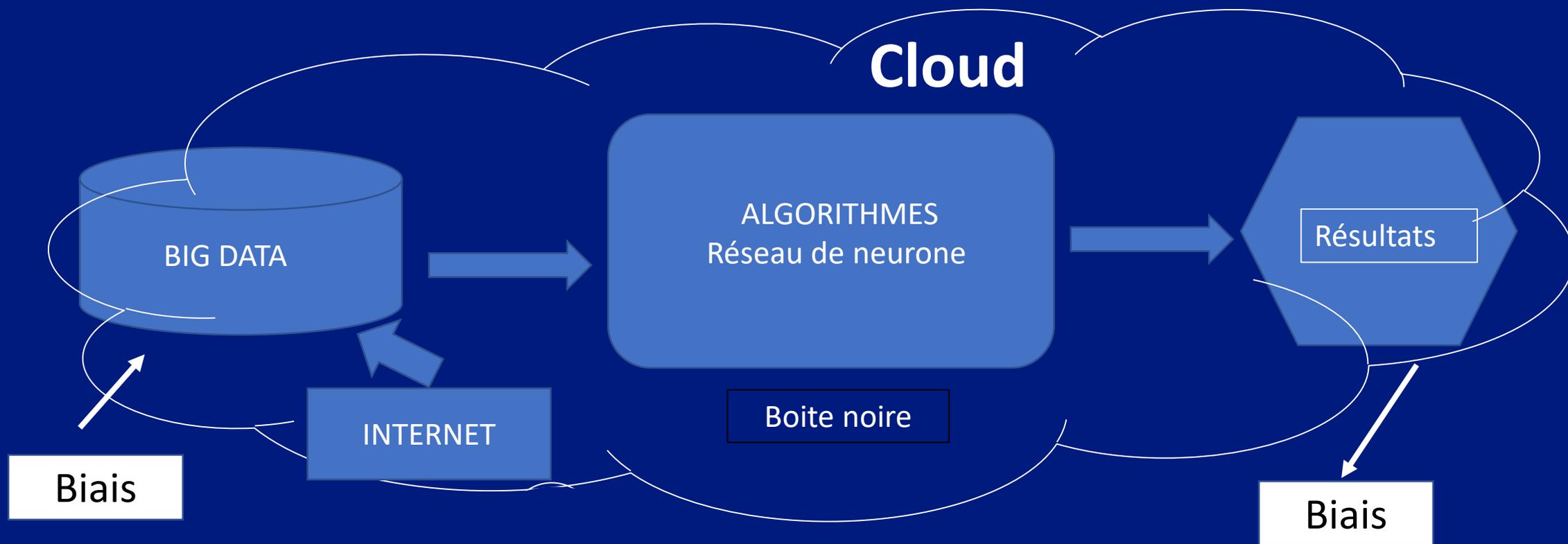


Image générée par le réseau adverse génératif StyleGAN, à partir d'une analyse de portraits.

Visages  
« inventés » par  
un réseau  
adverse génératif



# L'intelligence artificielle



# Les grands domaines des algorithmes de l'IA

- La reconnaissance d'images
- La reconnaissance du langage naturel
  - Analyse syntaxique et sémantique de textes
  - Reconnaissance vocale

# Reconnaissance des images

# La voiture autonome



Nouveau Produit

 Télécharger le catalogue

INNOVATION EN ROBOTIQUE GUIDÉE PAR VISION

# Robotique guidée par vision



Reconnaissance  
d'image dans  
l'industrie



Reconnaissance  
visuelle de défauts  
dans l'industrie



Où est-ce ?

Tout le monde peut  
tester un algorithme de  
reconnaissance d'image  
avec Google photo



## Recherche par image



Effectuez des recherches sur Google à l'aide d'images au lieu de texte. Essayez de faire glisser une image ici.

**Recherche par image**

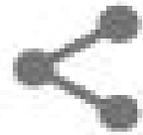
**Coller l'URL de l'image**

**Importer une image** 

Parcourir...

Aucun fichier sélectionné.

# Coïmbre



Ville au Portugal



Coimbra est une ville fluviale du Portugal, ancienne capitale du pays, qui abrite une vieille ville médiévale



Recherche de  
nouveaux détails  
astronomiques

# Les biais dans la reconnaissance d'images

- L'Université de Cornwell à sélectionné 7500 images du monde réel pour tromper l'IA

# Résultats...

Résultat de l'IA

Lion de mer 99%

Ecureuil



Résultat de l'IA

couvercle de regard 99%

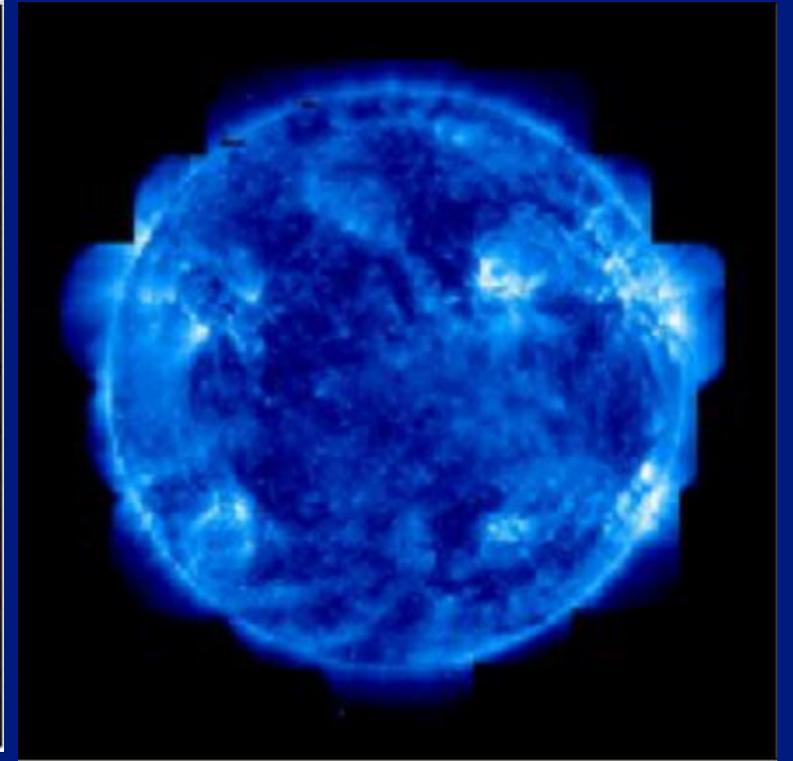
Libellule



Résultat de l'IA

Méduse 99%

Photosphère



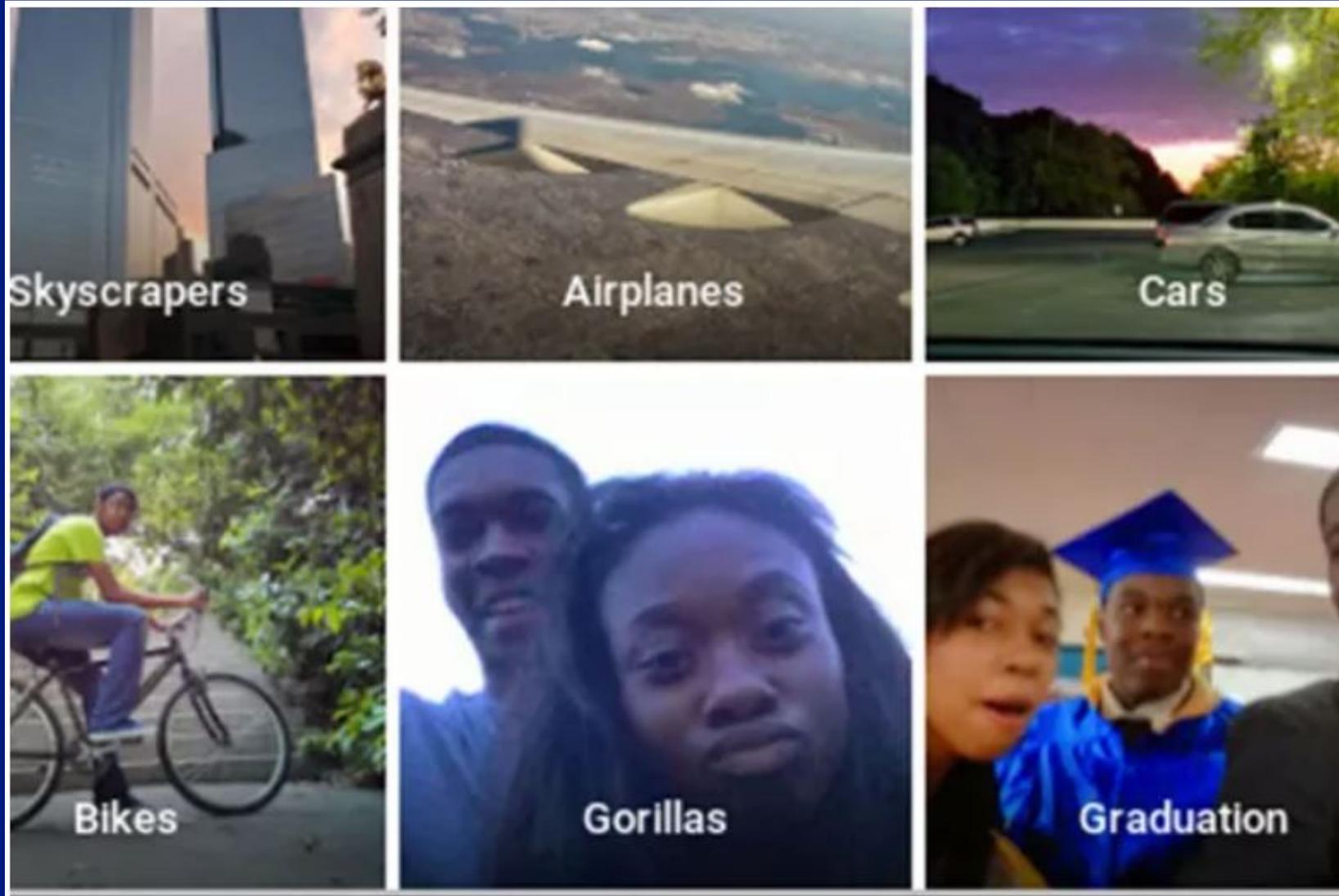


Sèche-cheveux ?



Dromadaire ?

# Le pire dans les biais de la reconnaissance d'image



# Les biais dans la reconnaissance d'images

- « *Les raisons pour lesquelles les algorithmes se trompent ne sont pas toujours tout à fait claires* », Dan Hendrycks, coauteur de l'étude de l'Université de Cornwell

# Les biais dans la reconnaissance d'images

- Face ++ identifie 99,3 % des hommes, mais seulement 78,7 % des femmes
- IBM reconnaît 77,6 % des peaux foncées contre 95% des peaux blanches

# Surveillance de masse

## « Clearview AI » épinglée

- « Clearview AI » récoltait sur le Web toutes les images contenant des visages humains avec les données personnelles associées (10 milliards de photos).
- L'accès à cette BDD a été vendu à des polices, à des gouvernements et à diverses sociétés dans le monde

Enquête New York Times (01/2021)

- Le 16 décembre 2021, la CNIL a ordonné à Clearview AI de supprimer toutes les photos et autres informations personnelles récoltées en France. Les organismes similaires britannique, australien et canadien ont fait de même.

- <https://www.lemonde.fr/idees/article/2022/01/06/les-technologies-de-reconnaissance-faciale->

# Reconnaissance de la voix et du langage naturel

# Reconnaissance du langage naturel

- Dans de nombreux outils du quotidien :
  - Traducteurs (Google Translate...)
  - Correcteurs orthographiques
  - Assistants numériques : Alexa, Siri, chatbots de support clients.
- <https://www.journaldunet.com/solutions/dsi/1503539-ia-et-reconnaissance-vocale-des-technologies-a-suivre-de-pres/>

# La reconnaissance du langage

- Plus les algorithmes utilisent de paramètres, meilleurs sont les résultats
- Application Megatron de Microsoft et Nvidia.
  - 530 milliards de paramètres en entrée
- Gopher, DeepMind
  - 280 milliards de paramètres
- GPT-3 d'OpenAI
  - 175 milliards de paramètres. Utilisé dans plusieurs domaines : les chatbots, les questions/réponses ou encore la traduction mais pas dans la santé (voir plus loin)
- [https://siecledigital.fr/2021/12/09/ia-deepmind-explore-les-limites-des-modeles-de-langage /](https://siecledigital.fr/2021/12/09/ia-deepmind-explore-les-limites-des-modeles-de-langage/)

# Traduction instantanée

- Google Translate permet d'obtenir la traduction écrite d'une phrase parlée.
- Avec les écouteurs Pixel Buds, l'utilisateur peut entendre directement la traduction de son interlocuteur.

# Watson gagne à Jeopardy



- Jeopardy, jeu de culture générale très populaire aux USA. Un indice est donné aux joueurs qui doivent formuler la question dont cet indice est la réponse.
- Exemples : Indice : « C'est le plus grand des océans ». Réponse du joueur : « Qu'est-ce que le Pacifique ? »
- Indice : « Sorti en 2011, c'est le seul film à avoir remporté l'Oscar et le César du meilleur film de l'année ». Réponse du joueur : « Qu'est-ce que *l'Artiste* ? » (Film de Serge Hazanavicius)

# Watson gagne à Jeopardy

- IBM a entré dans la base de connaissance de Watson 200 millions de pages de contenu
- L'ordinateur traite 500 milliards d'octets par seconde, l'équivalent de 500 000 livres par seconde.



En une seconde,  
Watson « parcourt »  
un volume  
d'information  
équivalent à 250 fois  
celui des ouvrages  
présents sur cette  
photo.

# Les biais de la reconnaissance du langage naturel

# Les biais de la reconnaissance du langage naturel

- NVIDIA et Microsoft reconnaissent que l'IA reproduit des stéréotypes et des préjugés issus des données qui servent à l'entraînement.
  - Exemple : Tay (2016) un chatbot de Microsoft tenait des propos racistes et sexiste. Fermé 16h après sa mise en service.

# Les biais de la reconnaissance du langage naturel

- Traitement de CV. Si plus d'homme ont été recrutés, le système reproduira cette proportion quitte à reproduire la misogynie
  - On supprime la mention « sexe » ?
    - L'IA reconnaîtra les prénoms féminins
  - On supprime les prénoms ?
    - L'IA identifiera le sexe avec les loisirs, la carrière...

# « Suicidez-vous ! »

- Lors d'un essai GPT-3 a conseillé à un (faux) patient de se suicider.
- A demandé à une personne présentant les symptômes d'une embolie pulmonaire de regarder une vidéo sur YouTube pour apprendre à se détendre !
- GPT-3 n'a pas été conçu et entraîné pour être utilisé dans le domaine de la santé.
- Les systèmes de reconnaissance du langage naturel ne sont pas universels.

# L'assistant vocal d'Apple, Siri, espionne ses utilisateurs !

- *The Guardian* a révélé que des utilisateurs du monde entier étaient écoutés à leur insu.
- Le lanceur d'alerte Thomas Le Bonniec :
  - "On écoutait des enregistrements réels de personnes qui s'adressaient à Siri et aussi parfois des gens qui parlaient autour de Siri.*
  - "1 300 enregistrements par jour ont été écoutés, pendant huit ou dix semaines. »*

Apple voulait recueillir le plus d'exemples de voix et de phrases pour améliorer l'algorithme de reconnaissance vocale de Siri.

- <https://www.francetvinfo.fr/internet/securite-sur-internet/video-nos-smartphones-nous-espionnent-ils-pendant->

# LES APPLICATIONS DE L'IA

- Les algorithmes de l'IA sont présents dans des secteurs particulièrement sensibles.

- Santé

- Police

- Justice

- Armée

- Réseaux sociaux

- Transports

- Défense

- ... et aussi dans les domaines suivants :
- Publications scientifiques
- Commerces, banques, assurances,
- Traductions
- Smartphones
- Environnement
- Domotique
- Robotique
- Scolarité (parcoursSup)
- Arts
- Jeux (échec, Go, Jéopardy ...) etc.

- <https://www.mailabs.fr/intelligence-artificielle/exemples-intelligence-artificielle/>

- <https://www.lesechos.fr/2018/05/comment-lia-revolutionne-la-traduction-991256>

# Intelligence artificielle et santé

# OPTIMA

- Développé par l'Université Johns Hopkins de Baltimore
- Une intelligence artificielle est capable de représenter virtuellement le cœur d'un patient
- But : mieux traiter les arythmies cardiaques
- A été testée sur près de 160 patients qui souffraient de problèmes cardiaques aux États-Unis

# AlphaFold, Google DeepMind, le repliement des protéines

- L'IA « AlphaFold » de « Google DeepMind », a réussi à prédire comment les protéines se replient
- Pourrait permettre de lutter contre de nombreuses maladies et certaines formes de cancers
- Les chercheurs de DeepMind ont entraîné leur algorithme sur une base de données d'environ 170 000 séquences de protéines et sur leur forme.

# DeepGestalt

- L'application IA DeepGestalt de FDNA (Boston) est capable de détecter des maladies rares à partir de l'analyse du visage d'un patient.
- « Lors d'une étude de 17 000 visages d'enfants avec plus de 2000 syndromes génétiques différents, l'outil aurait donné des résultats meilleurs que ceux des médecins ».
- <https://siecledigital.fr/2019/01/15/sante-cette-ia-peut-detecter-des-maladies-rares-a-partir-de-visages/>

# Watson l'IA santé d'IBM

- Traite les dossiers et les publications scientifiques
  - *« Le robot n'a besoin que de trois secondes pour analyser 20 millions de pages de données »*
- En oncologie intervient afin de définir avec précision la nature d'un cancer et les traitements les plus adaptés

JUSTICE PREDICTIVE

# Aides à la décision judiciaire

- Des algorithmes calculent la probabilité d'une décision de justice. S'ils trouvent que les chances de gagner de chaque partie sont égales, il leur conseille de ne pas aller en justice mais de trouver une conciliation (États-Unis)

# Datajust

- En cours de développement en France.
- Un algorithme en développement pour recenser, par type de préjudice :
  - les montants demandés et offerts par les parties à un litige
  - les montants alloués aux victimes en indemnisation de leur préjudice corporel dans les décisions rendues en appel
- Porte sur 3,9 millions de décisions de justice

# IA et justice

- Dans l'Ontario, un "tribunal virtuel" tranche les conflits entre voisins ou entre salarié et employeur
- Au Québec, un logiciel permet de régler les petits contentieux commerciaux.
- En Estonie, un robot devrait bientôt établir la culpabilité d'une personne pour des litiges "mineurs" (moins de 7 000 euros)
- <https://www.vie-publique.fr/eclairage/277098-lintelligence-artificielle-ia-dans-les-decisions-de-justice>

# Critiques de l'IA dans la justice

- Aux États-Unis, l'algorithme Compas est utilisé par les juges pour estimer la probabilité de récidive des accusés et ajuster la peine en fonction de cette évaluation.
  - Compas suscite de nombreuses critiques
- Des juristes dénoncent des biais raciaux dans les algorithmes qui pénalisent les minorités ethniques

POLICE PREDICTIVE

# IA et police prédictive aux Etats Unis

- L'algorithme PredPol est censé prévenir les délits et les crimes dans certaines zones.
- Après 8 ans d'expérimentations PredPol fait l'objet de vives critiques, à cause des biais de l'algorithme qui perpétueraient les discriminations sociales.
- En avril 2020, la police de Los Angeles a stoppé l'utilisation de ce logiciel.
- <https://www.franceculture.fr/emissions/le-meilleur-des-mondes/police-predictive-legaltech-les-algorithmes-font-ils-la-loi>

# IA et police prédictive

- L'entreprise chinoise Cloud Walk Technology utilise la reconnaissance faciale pour prédire si un individu risque de commettre un crime avant qu'il ne survienne.
  - Un changement de comportement, un mouvement inhabituel... le système est conçu pour détecter tout comportement suspect.

# ARMES LETALES AUTONOMES (robots tueurs)

# SALA, Système d'Armes Létales Autonomes

- SALA : « *Automate capable de mener une action létale de manière automatisée sans intervention humaine dans un environnement changeant. Un système dont l'humain ne peut reprendre le contrôle qu'en procédant à sa destruction physique* ».

Claude Ganay, Commission parlementaire de Défense, juillet 2020

# SALA, Système d'Armes Létales Autonomes

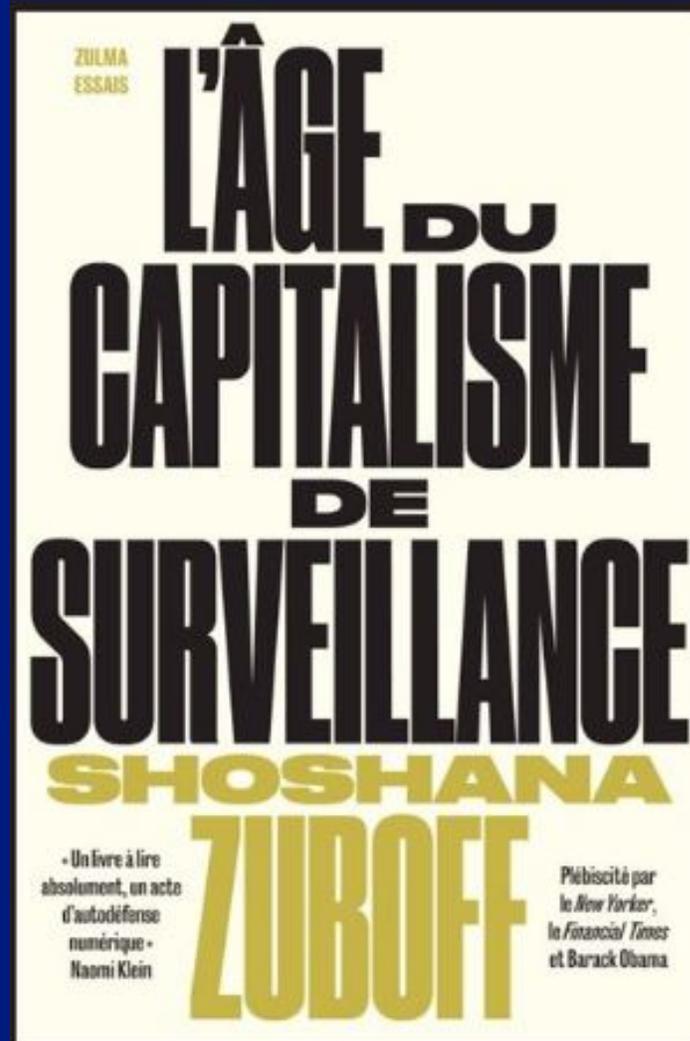
- le Comité d'éthique de la défense estime qu'il ne faut ni fabriquer, ni exporter, ni utiliser de SALA.
- *« L'utilisation de systèmes d'armes létales autonomes est contraire au principe de continuité de la chaîne de commandement. »*

# Plateformes Gafa

# Shoshana Zuboff



Spécialiste de psychologie sociale,  
professeure émérite à la Harvard  
business school



New York Times  
“Livre remarquable de  
l’année”

Financial Times  
“Meilleur livre de l’année”

Sunday Times (GB)  
“Meilleur livre business de  
l’année”

Un des livres préféré de  
Barack Obama (2019)

1000 pages

# Le surplus comportemental

- *« Les Gafa convertissent nos existences en données comportementales, se les approprient et les traitent dans leurs «usines» de plus en plus complexes : l'intelligence artificielle, le «machine learning».*
- *« Une grande partie de ces données nous ont été volées. Elles nous ont été prises sans que nous le sachions, et sans que nous y consentions. »*
- <https://www.capital.fr/entreprises-marches/a-notre-insu-google-et-facebook-nous-volent-nos-vies-1388260>

# Le surplus comportemental

- *« Dans ces usines, les Gafa fabriquent des produits: des prédictions de nos comportements futurs »*
- *« Elles ont de la valeur car elles s'approchent de la certitude en s'appuyant sur une grande variété de données, dans des quantités inimaginables. »*
- Facebook produit 6 millions de prédictions de ce type chaque seconde.

# Le surplus comportemental

- La publicité ciblée est le premier marché reposant sur cette exploitation : cette activité explique à 88% la valorisation de Google et à 99% celle de Facebook

# LE CAS GOOGLE

# Google

- 1998, Sergey Brin et Larry Page créent Google.
- Leur objectif : vendre la licence de leur moteur de recherche.
- Étaient farouchement anti publicité !
- Stockaient sans rien en faire les données les données du surplus comportemental



# Google

- Un étudiant, Amit Patel, convainc Brin et Page que ces données collatérales constituent un « puissant capteur de comportement humain ».
- Les utilisateurs de Google ne sont ni les clients de l'entreprise (ils n'achètent rien), ni le « produit ».
- Le produit **ce sont les données du « surplus comportemental »** qui permettent aux annonceurs (les vrais clients de Google) de faire de la publicité ciblée extrêmement efficace.

# Le surplus comportemental

- En 2003 Google a déposé un brevet sur la publicité ciblée\* fondée sur l'analyse prédictive quasi instantanée de la requête en fonction de séries de données comportementales \*\*

\* « *Generating User information for Use in Targeted Advertising* »

\*\* « *User profile information* » UPI »

# Le surplus comportemental, ça marche !

- Le surplus comportemental fait la fortune de Google :
  - 2001 début des test du surplus comportemental : chiffre d'affaires 86 M\$
  - 2002 347 M\$
  - 2003 1,5 Md \$
  - 2004 3,2 Md \$.
  
- 2020, 182 Md\$
- Source « L'âge du capitalisme de surveillance » Shoshana Zuboff, professeur à Harvard

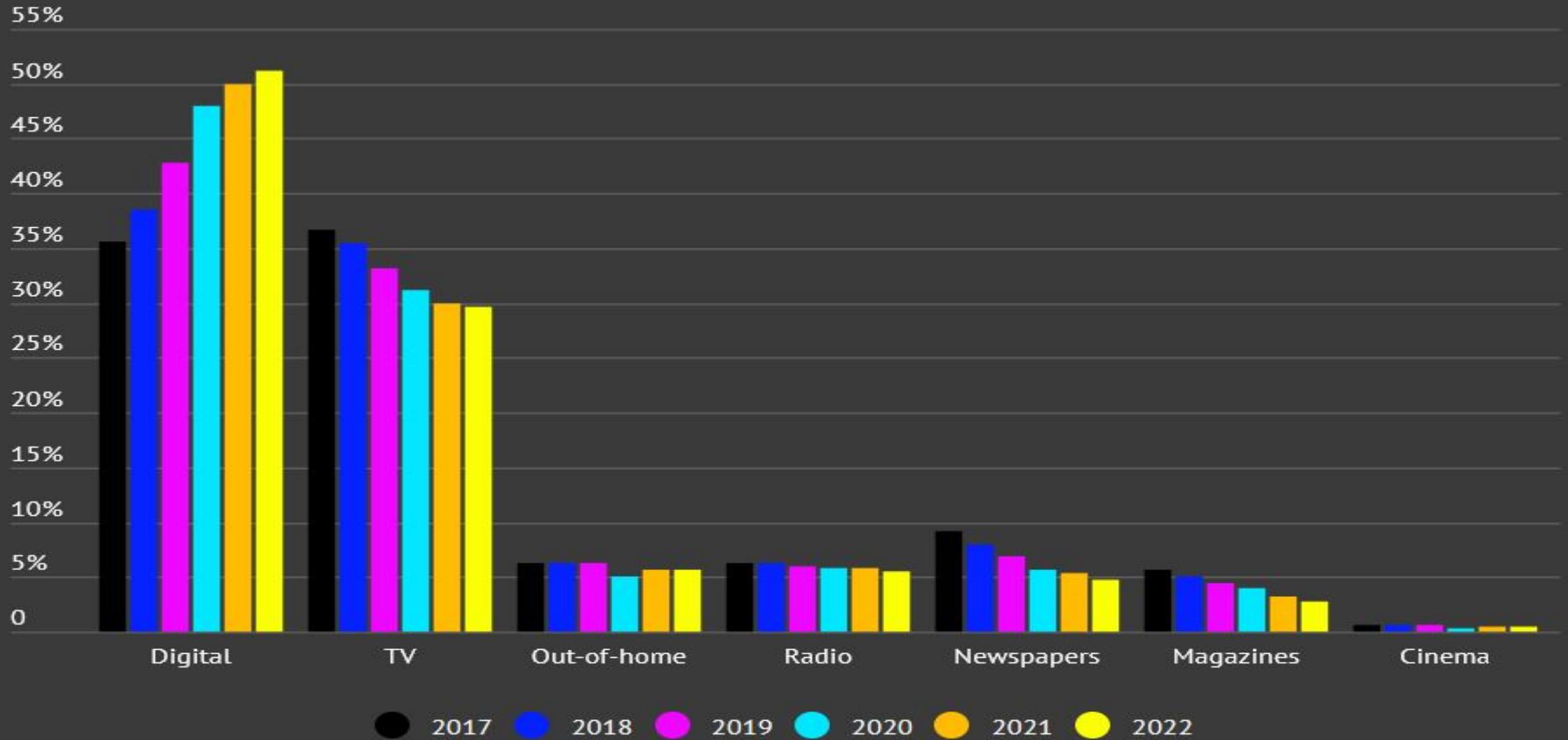
# Chiffres Google

- 2016 : 89% des revenus de Google provenaient des campagnes de publicités ciblées.
- 2020 : Google traite 6,9 milliards de requêtes par jour !
- 100 millions de milliards de données sont stockées sur ses serveurs

# Les Gafa dominant le marché publicitaire

- En siphonnant 2/3 des recettes publicitaires mondiales les Gafa contraignent un grand nombre d'entreprises de presse écrite, de télévision ou de radio à fermer leurs portes, à se regrouper ou à réduire drastiquement le nombre de leurs journalistes.
- <https://www.lesechos.fr/tech-medias/medias/pour-la-premiere-fois-le-numerique-sarrobe-plus-de-la-moitie-du-marche-publicitaire-en-france-1271724>

# SHARE OF GLOBAL AD SPEND BY MEDIA



# LE CAS FACEBOOK

# De Google à Facebook

- En 2008, Sheryl Sandberg, ex-Google, importe la démarche d'utilisation du surplus comportemental de Google chez Facebook.
- Elle a compris que le réseau social Facebook est une fabuleuse mine de données comportementales.

# Le bouton « Like »



- Les « Like » génèrent, à l'insu des utilisateurs, une série d'attributs personnels confidentiels.
- Plus un utilisateur clique sur le bouton « Like » plus il accroît la valeur prédictive de signaux fournis à Facebook.
- Source « L'âge du capitalisme de surveillance » Shoshana Zuboff, professeur à Harvard

# L'amplification virale des messages selon Bruno Patino

- « *L'algorithme Facebook accélère, ralentit, amplifie ou réduit les messages dans un seul but : que les utilisateurs les partagent et qu'ils deviennent viraux. »*
- « *Les messages choisis sont ceux capables de retenir le plus l'attention : messages clivants, de haine, de désinformation. »*
- Fausses informations sur le Covid : 12 émetteurs seulement mais suivis par 59 millions d'utilisateurs de Facebook, Twitter, Instagram etc. ! (démultiplication par l'action des algorithmes )
  - Source CCDH américain. Center for countering digital haine

# L'amplification virale des messages selon Bruno Patino

- Etude Université de New York et Grenoble Alpes, élections USA de 2020
- Les messages de désinformation sur Facebook ont reçus 6 fois plus de « clics » que les messages d'information factuelle, lors des élections de 2020 aux Etats Unis.
- En 2021 Facebook a fermé les comptes de l'université de New York !
- 31% des adultes américains s'informent régulièrement grâce a Facebook (Pew Research, 2021)

# Frances Haugen et les Facebook papers



Frances Haugen, lanceuse  
d'alerte, ex-ingénieur  
Facebook spécialiste des  
algorithmes

# Les Facebook papers

- Des milliers de pages de documents internes révélés par Frances Haugen
  - Fournissent de multiples exemples des pratiques nocives de Facebook et de ses filiales.
- *« Ces révélations dressent le portrait d'une entreprise cynique, consciente de son influence toxique pour les adolescents et la démocratie, mais qui choisit de ne pas remettre en cause son modèle. »*
- [https://www.lemonde.fr/pixels/article/2021/10/26/comment-l-algorithme-de-facebook-echappe-au-controle-de-ses-createurs\\_6099888\\_4408996.html](https://www.lemonde.fr/pixels/article/2021/10/26/comment-l-algorithme-de-facebook-echappe-au-controle-de-ses-createurs_6099888_4408996.html)
- <https://www.rtl.fr/actu/sciences-tech/facebook-papers-4-choses-a-retenir-des-nouvelles-revelations-qui-ebbranlent-facebook-7900090342>

# Facebook papers

- *« Nous avons vu à maintes reprises dans les recherches de Facebook qu'il est plus facile d'attiser la colère des utilisateurs plutôt que de les inciter à faire preuve d'empathie ou de compassion. Nous subventionnons donc littéralement la haine sur ces plateformes. »*

Frances Haugen, devant le parlement britannique

# Facebook

## une modération problématique

- 2,85 milliards d'utilisateurs actifs chaque mois
- 100 000 modérateurs humains effectuent 10% des vérifications.  
90% le sont par intelligence artificielle.
- <https://fr.statista.com/infographie/17800/big-data-evolution-volume-donnees-numeriques-genere-dans-le-monde/>

# Modération problématique

- Selon les documents révélés par les « Facebook papers », Facebook consacre l'immense majorité de son budget modération en Occident et est incapable de protéger ses utilisateurs ailleurs.
- Le groupe dit être « *bien conscient qu'une politique de modération plus faible dans les pays non anglophones rend la plateforme vulnérable aux abus de personnes mal intentionnées et de régimes autoritaires* » Washington Post.

# Facebook papers

- De nombreux documents indiquent que Facebook ne comprend plus, ou mal, ce que font ses propres algorithmes. Son réseau social est devenu une machine difficile à contrôler.

« Le Monde » à partir des Facebook papers

# Facebook papers

- En Inde, des vidéos pornographiques *soft* se sont retrouvées subitement mises en avant dans l'onglet Watch, sans que personne ne comprenne pourquoi.
- Aux Etats-Unis, des ingénieurs s'arrachent les cheveux pour comprendre pourquoi certains groupes politiques continuent d'être recommandés aux utilisateurs alors qu'ils ne devraient plus l'être.

# Facebook papers

- *« Facebook a eu quinze ans pour démontrer que les algorithmes de classement des contenus en fonction de l'engagement [le nombre de commentaires, de partages...] peuvent être conçus de manière responsable ; s'ils n'ont pas réussi à le faire jusqu'à présent, ils n'y arriveront jamais. »*

Roddy Lindsay, tribune dans le *New York Times*

# Youtube

- Sur YouTube, 120 000 ans de temps de vidéos sont visionnés chaque jour.
- 70% sont regardées en raison de la recommandation de l'intelligence artificielle de la plateforme
- une étude a montré que les recherches du terme « climat » sur YouTube ont majoritairement (54%) orienté les internautes vers des vidéos climatosceptiques.

# DSA, DIGITAL SERVICE ACT européen



- Ce qui est autorisé offline doit l'être online, ce qui est interdit offline doit l'être online.
- Que l'on parle de contrefaçon, d'antisémitisme, de pédopornographie, de menaces de mort ou de vente de drogues, tous les contenus illégaux doivent être retirés.
- Les contenus haineux, l'amplification de la violence verbale et physique, la désinformation doivent être identifiés comme tels et traités en conséquence.

Conclusion

- « *L'intelligence artificielle n'est pas un sujet technique mais politique* »

- Kate Crawford, professeure Université de New York, Normale sup Paris, dans « *Power politics and the planetary cost of artificial intelligence* », 2021

*« La guerre de l'information tourne le dos à la démocratie. Trois questions se posent : Qui sait ? Qui décide ? Qui décide qui décide ?*

*« Dans l'état actuel des choses, ce sont les entreprises capitalistes de surveillance qui savent. C'est la forme du marché qui décide. C'est la lutte concurrentielle entre les capitalistes de surveillance qui décide qui décide. »*



## Soshana Zuboff

Spécialiste de psychologie sociale, professeure émérite à la Harvard business school

*« Réorienter l'informatisation de la société passe par une forme de décolonisation de l'esprit, de la manière d'appréhender la numérisation du monde, afin de se soustraire à la pensée imposée par une poignée d'acteurs, pour oser imaginer que d'autres alternatives sont possibles ».*



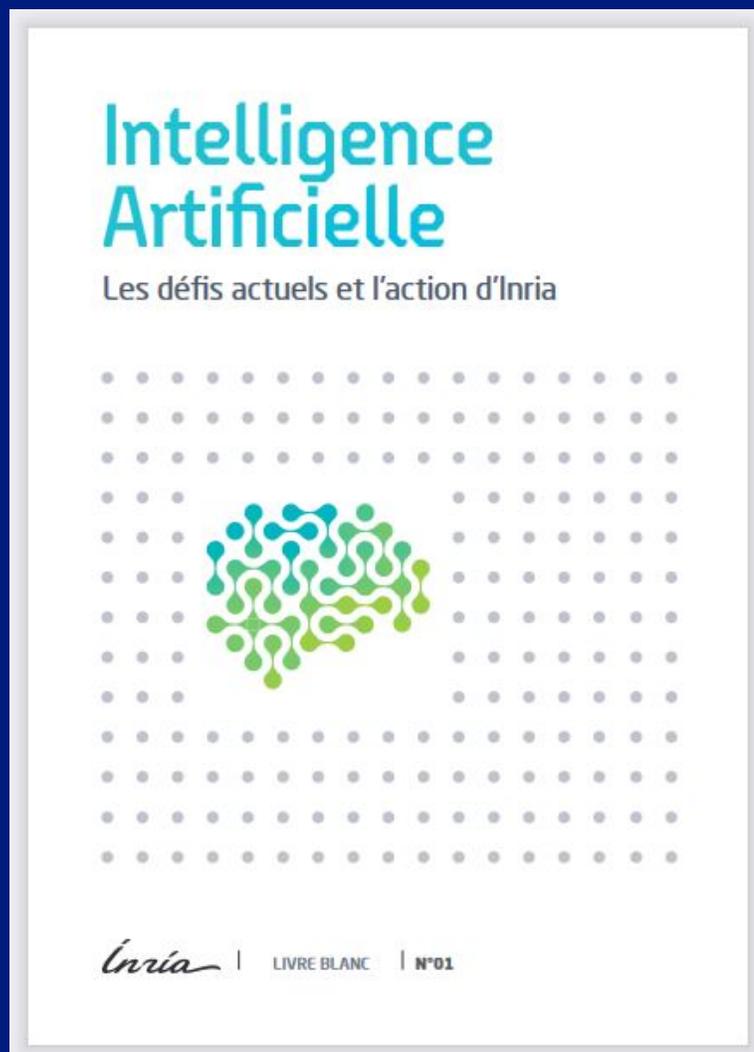
## Solange Ghernaouti

Solange Ghernaouti, membre de la Commission suisse pour l'Unesco, spécialiste des questions numériques.

Pour compléter cet exposé d'excellents rapports  
sont accessibles en téléchargement

- Ci-après les liens pour le faire.

# Livre blanc de l'INRIA 2016



<https://www.inria.fr/fr/livre-blanc-inria-intelligence-artificielle>

# Rapport Villani 2018



<https://www.vie-publique.fr/rapport/37225-donner-un-sens-lintelligence-artificielle-pour-une-strategie-nation>

# Livre blanc de la CNIL



[https://www.cnil.fr/sites/default/files/atoms/files/cnil\\_rapport\\_garder\\_la\\_main\\_web.pdf](https://www.cnil.fr/sites/default/files/atoms/files/cnil_rapport_garder_la_main_web.pdf)

# Livre blanc de la commission européenne sur l'intelligence artificielle



Bruxelles, le 19.2.2020  
COM(2020) 65 final

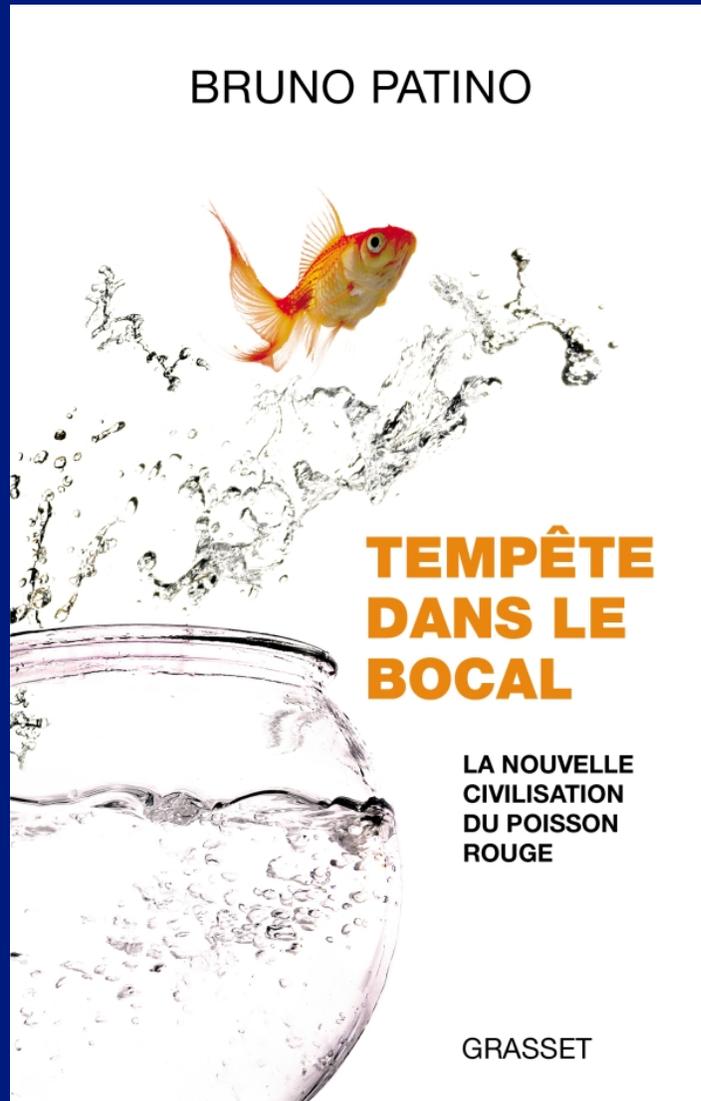
**LIVRE BLANC**

**Intelligence artificielle**

**Une approche européenne axée sur l'excellence et la confiance**

# Rapport Bronner 2022





# Excellent livre grand public sur le sujet

Bruno Patino donne en fin d'ouvrage les références d'une trentaine de livres et plus de 100 références de pages ou de sites internet sur l'intelligence artificielle et les réseaux sociaux.



FIN