

# INTELLIGENCE ARTIFICIELLE: TECHNIQUES, SUCCÈS, ET ENJEUX

François Fleuret

<https://fleuret.org/francois>



**UNIVERSITÉ  
DE GENÈVE**

Grâce aux techniques d'apprentissage statistique et à des processeurs de calcul très puissants, l'Intelligence Artificielle moderne:

- se “programme elle-même”,
- résout réellement des problèmes complexes,
- demande des quantités énormes de données et de calculs,
- est facile à utiliser,
- est difficile à analyser et à interpréter.

Les modèles génératifs récents d'images et de textes semblent marquer le début d'une ère d'utilisation beaucoup plus importante.

L'IA se "programme elle-même"

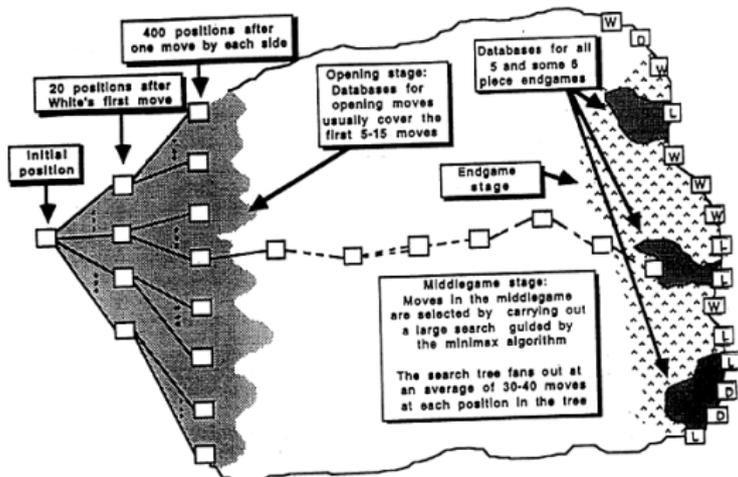
La manière traditionnelle de spécifier à un ordinateur comment résoudre une tâche consiste à la décomposer explicitement en pas élémentaires.

```
n = 15345

while n > 1:
    for k in range(2, n+1):
        if n%k == 0:
            print(k)
            n = n // k
            break
```

Les premières tentatives de conception de systèmes d'Intelligence Artificielle, par exemple pour des médicales, des jeux de stratégies, ou de la vision par ordinateur, reposaient sur le même paradigme.

## Chess game tree

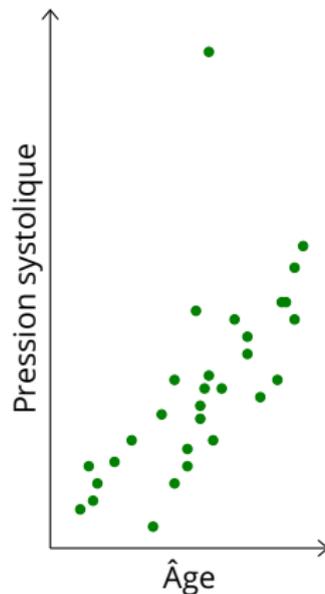
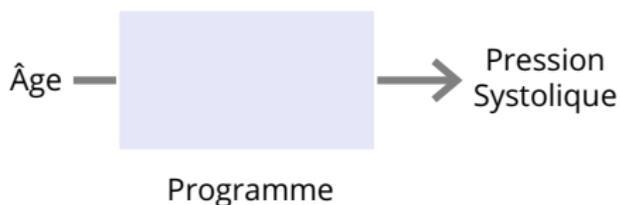


(Newborn, 1996)

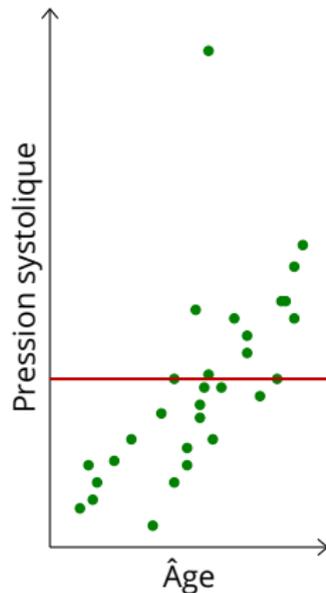
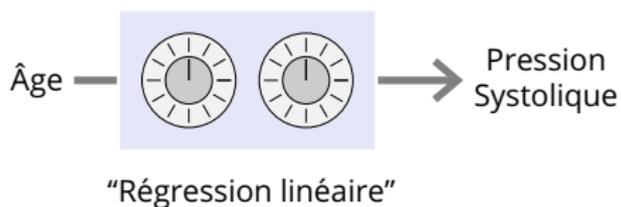
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



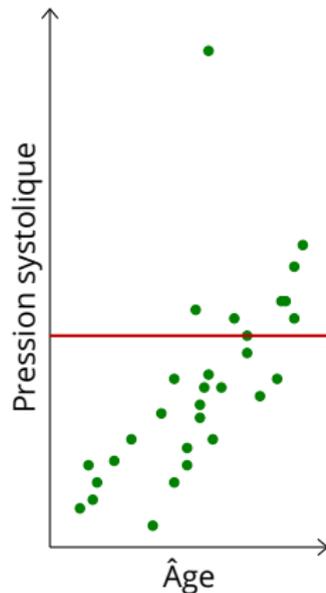
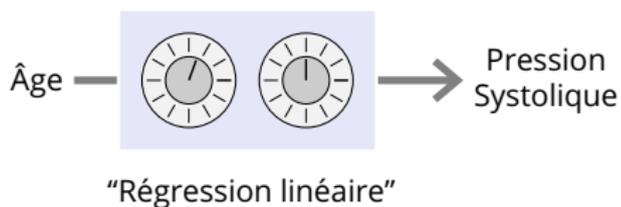
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



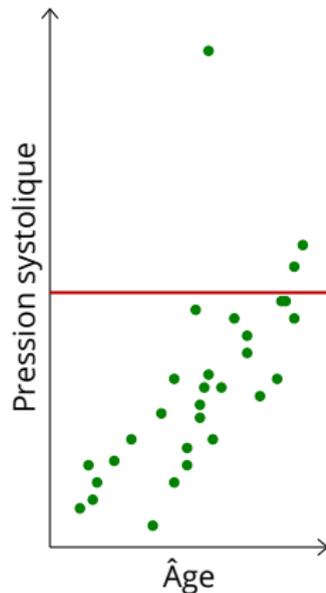
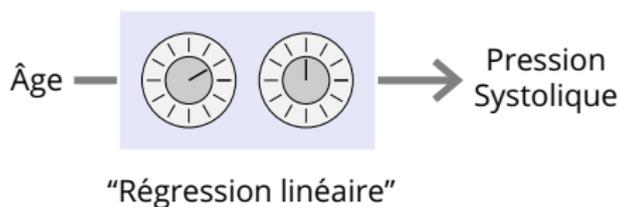
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



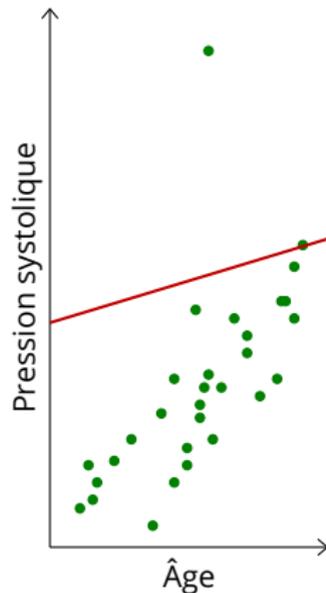
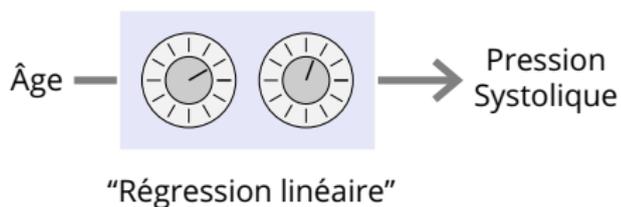
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



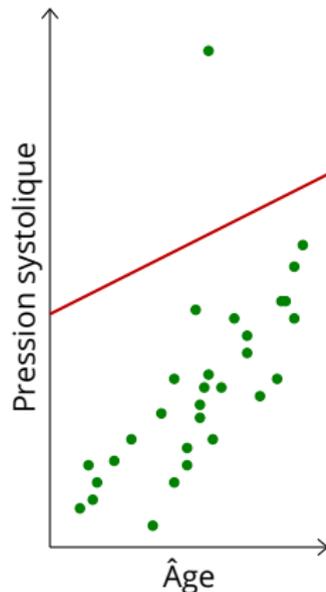
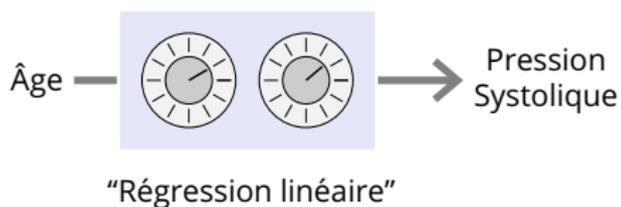
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



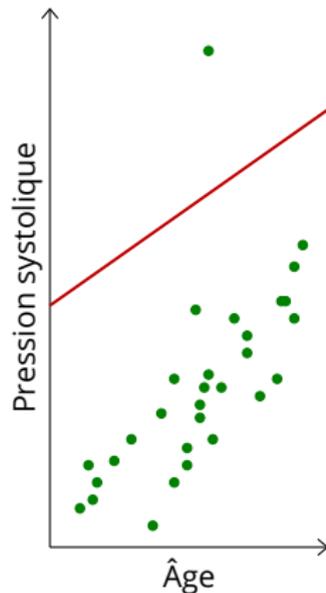
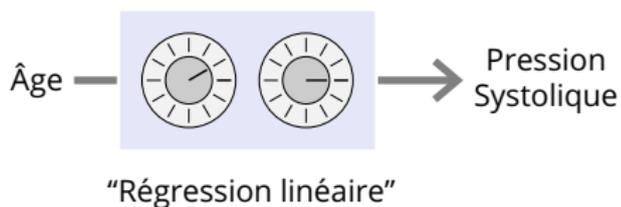
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



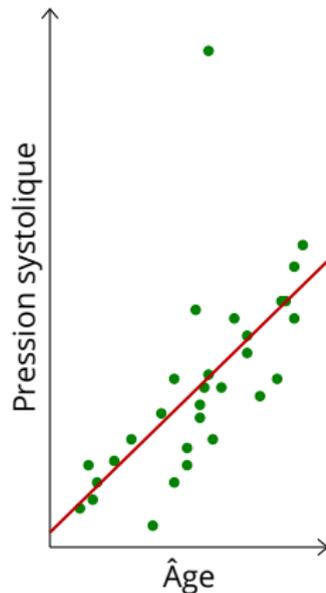
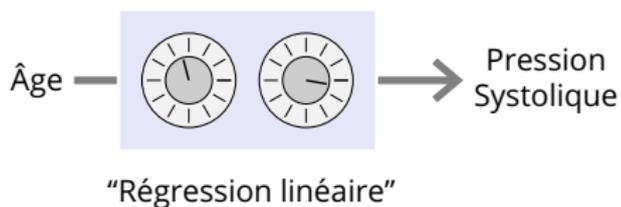
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



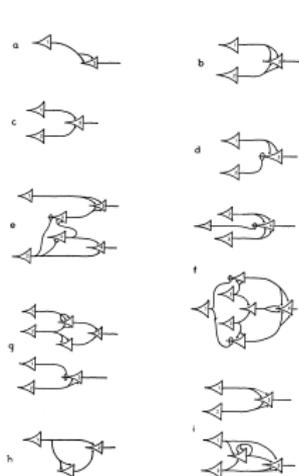
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



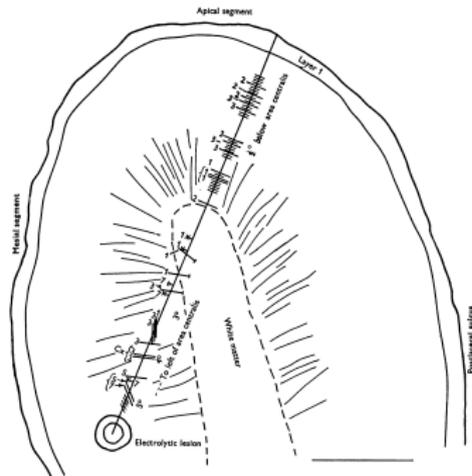
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



Cette stratégie rappelle la plasticité des réseaux de neurones naturels.

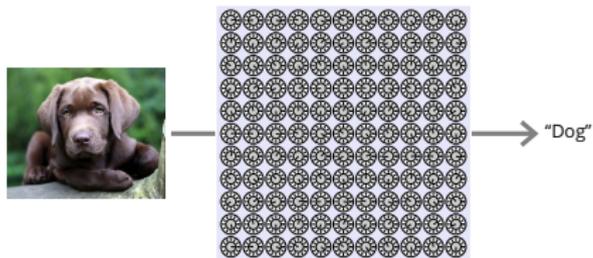


(McCulloch and Pitts, 1943)

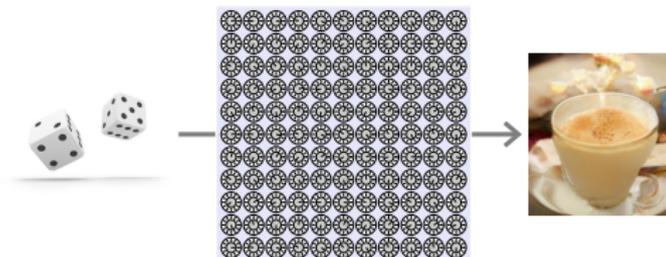


(Hubel and Wiesel, 1962)

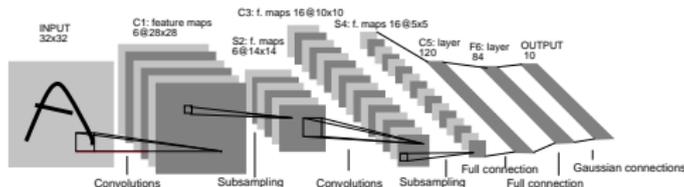
Elle peut être utilisée à grande échelle pour extraire de l'information de signaux naturels tels qu'une image, un échantillon sonore ou un texte



ou pour produire un signal complexe et réaliste

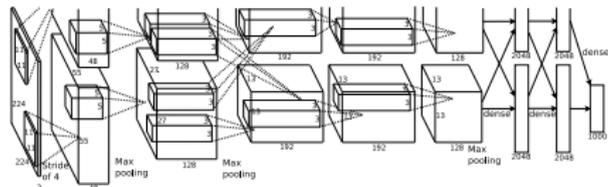


Les modèles modernes intègrent  $10^5$  à  $10^{11}$  paramètres.

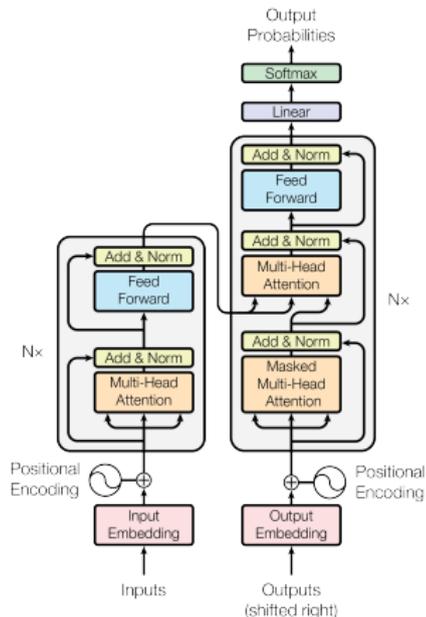


Convolutions (LeNet, 1989)

... 1990-2010 "l'hiver des réseaux de neurones" ...

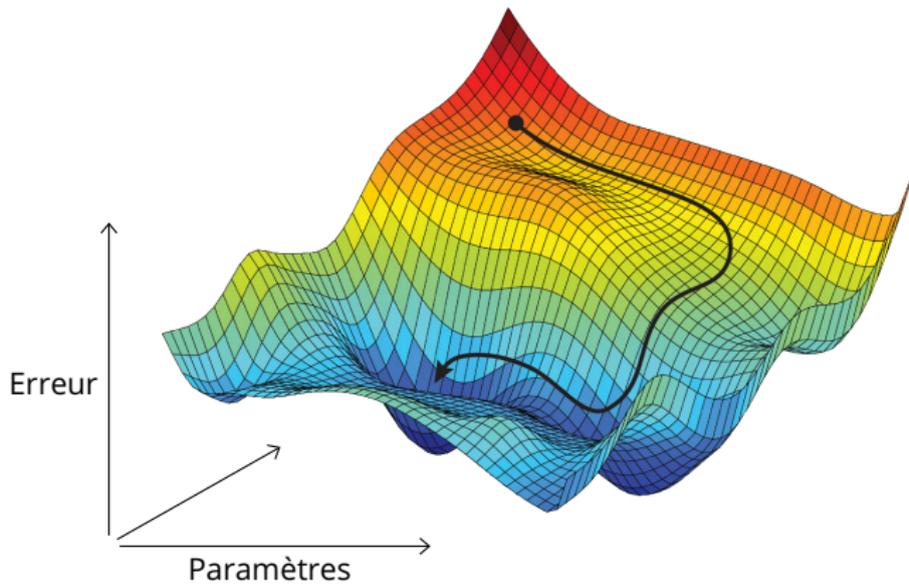


Très gros modèle + GPUs (AlexNet, 2012)



Attention (Transformer, 2018)

Entraîner un modèle consiste à modifier progressivement ses paramètres pour réduire l'erreur qu'il commet sur des exemples.



L'IA résout réellement des problèmes

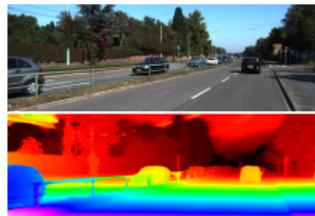




Analyse de scènes



Estimation de pose



Estimation géométrique



Planification



Description d'image

I: Jane went to the hallway.  
I: Mary walked to the bathroom.  
I: Sandra went to the garden.  
I: Daniel went back to the garden.  
I: Sandra took the milk there.  
Q: Where is the milk?  
A: garden

Compréhension de texte

Performance similaire à celle des humains:

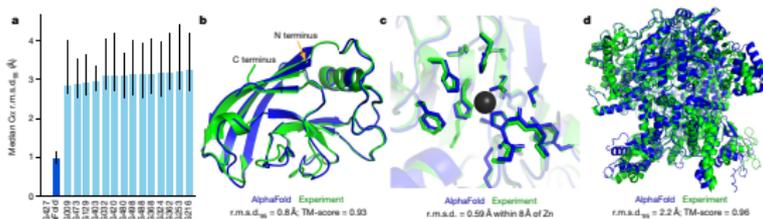
- Détection de cancer de la peau,
- reconnaissance de la parole.

Performance supérieure à celle des humains:

- Reconnaissance d'images,
- lecture de panneaux routiers,
- reconnaissance de visages,
- Go, échecs, poker.

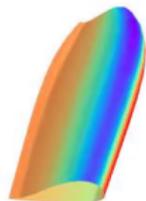
L'IA permet de surcroît de traiter des tâches que les humains ne peuvent pas résoudre correctement.

Repliement de protéines.



(Jumper et al., 2021)

Optimisation de formes.



Bike designed with artificial intelligence breaks world speed records



## Synthèse d'images.



Synthèse de textes.

**“Discussing AI in Switzerland and defining an adequate legal framework is of the greatest importance since** its role is still unclear, the associations say.

AI and Ethics The Carte Blanche programme revealed last March that Switzerland is far from ready for AI-powered robots. The Senate is due to decide on legislation for these systems in 2019.

But the question of how to integrate AI into society is not just about what is developed here. To what extent should AI technology be sold or shared? What kinds of responsibilities should AI systems have?”

Synthèse de textes.

**“Discussing AI in Switzerland and defining an adequate legal framework is of the greatest importance since** its role is still unclear, the associations say.

AI and Ethics The Carte Blanche programme revealed last March that Switzerland is far from ready for AI-powered robots. The Senate is due to decide on legislation for these systems in 2019.

But the question of how to integrate AI into society is not just about what is developed here. To what extent should AI technology be sold or shared? What kinds of responsibilities should AI systems have?”

**“The object was blue all over, but also green all over, it was a**

Synthèse de textes.

**“Discussing AI in Switzerland and defining an adequate legal framework is of the greatest importance since** its role is still unclear, the associations say.

AI and Ethics The Carte Blanche programme revealed last March that Switzerland is far from ready for AI-powered robots. The Senate is due to decide on legislation for these systems in 2019.

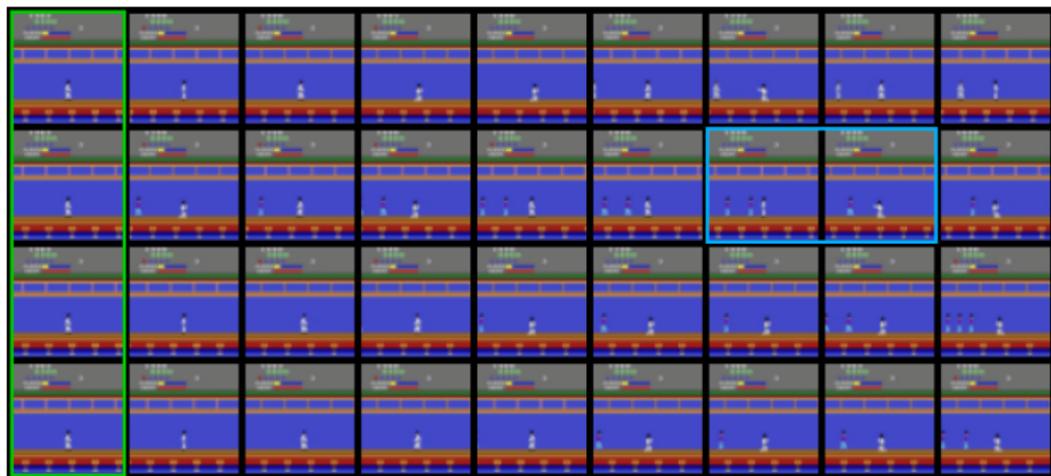
But the question of how to integrate AI into society is not just about what is developed here. To what extent should AI technology be sold or shared? What kinds of responsibilities should AI systems have?”

**“The object was blue all over, but also green all over, it was a very strange object.”**

IRIS apprend à jouer à 26 jeux Atari avec 2h de pratique sur chaque, et obtient sur 10 de ces jeux un score supérieur à un humain (Micheli et al., 2022).



- Un premier modèle transforme les images en “mots” inventés.
- Un second apprend le langage résultant pour prédire le futur.
- Un troisième apprend à jouer dans des parties “rêvées”.



(Micheli et al., 2022)

L'IA demande d'énormes quantités  
de données et de calculs

La dernière décennie de progrès en IA correspond à une augmentation énorme de la taille des "ensembles d'apprentissage".

<b>Jeu de données</b>	<b>Année</b>	<b>Nb. images</b>	<b>Taille</b>
MNIST	1998	60K	12Mb
Caltech 256	2007	30K	1.2Gb
ImageNet	2012	1.2M	150Gb
LAION-5B	2022	5.85MM	240Tb

<b>jeu de données</b>	<b>Année</b>	<b>Nb. livres (250p)</b>	<b>Taille*</b>
SST2	2013	40K	20Mb
WMT-18	2018	14M	7Gb
The Pile	2020	1.6MM	825Gb
OSCAR	2020	12MM	6Tb

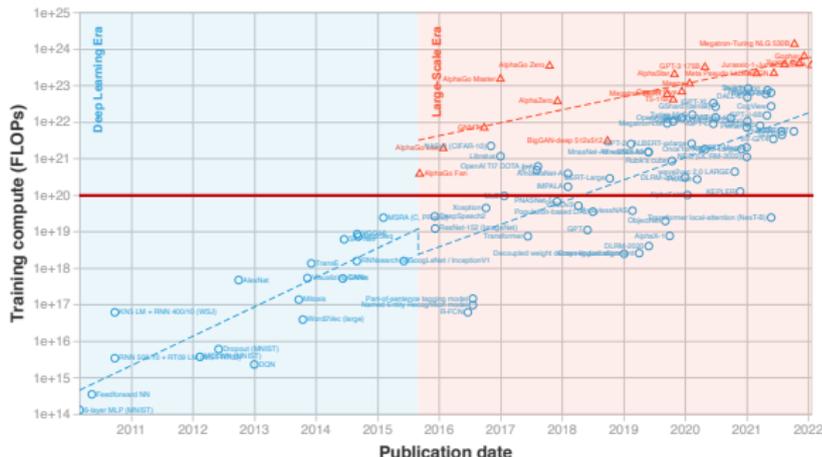
(\* tout Wikipedia en anglais correspond à 45Gb)

La puissance de calcul utilisée a augmenté de manière similaire.

Un accélérateur graphique disponible en grands magasins pour 1'650.- peut effectuer  $\approx 35'000$  milliards de multiplications à virgules par seconde.

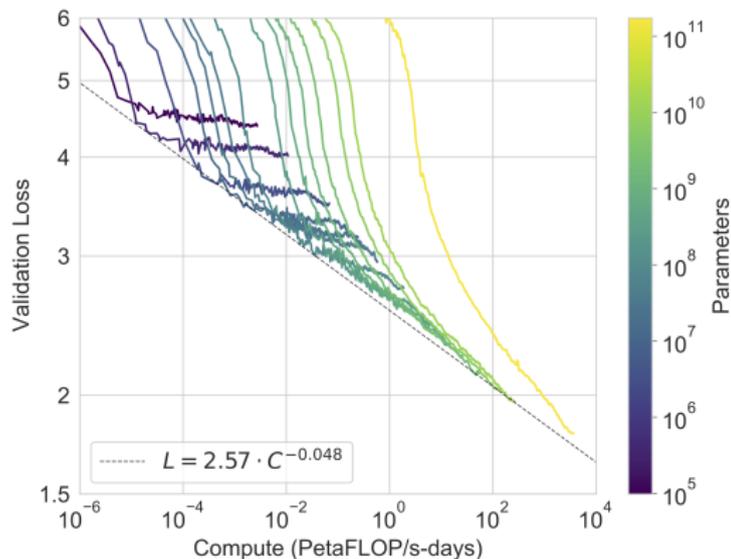
Training compute (FLOPs) of milestone Machine Learning systems over time

n = 102



(Sevilla et al., 2022)

Cette croissance ne semble pas ralentir.



(Brown et al., 2020)

L'IA est facile à utiliser

Le développement des modèles “profonds” se fait à l'aide de logiciels open-source.

Librairie	Contributeur principal
<b>PyTorch</b>	Facebook
TensorFlow	Google
JAX	Google
MXNet	Amazon

L'installation se fait avec une seule commande. Par exemple:

```
conda install pytorch torchvision torchaudio cudatoolkit=10.2 -c pytorch
```

MNIST

1 1 8 3 6 1 0 3 1 0 0 1 1 2 7 3 0 4 6 5  
2 6 4 7 1 8 9 9 3 0 7 1 0 2 0 3 5 4 6 5  
8 6 3 7 5 8 0 9 1 0 3 1 2 2 3 3 6 4 7 5  
0 6 2 7 9 8 5 9 2 1 1 4 4 5 6 4 1 2 5 3  
9 3 9 0 5 9 6 5 7 4 1 3 4 0 4 8 0 4 3 6  
8 7 6 0 9 7 5 7 2 1 1 6 8 9 4 1 5 2 2 9  
0 3 9 6 7 2 0 3 5 4 3 6 5 8 9 5 4 7 4 2  
1 3 4 8 9 1 9 2 8 7 9 1 8 7 4 1 3 1 1 0  
2 3 9 4 9 2 1 6 8 4 7 7 4 4 9 2 5 7 2 4  
4 2 1 9 7 2 8 7 6 9 2 2 3 8 1 6 5 1 1 0  
4 0 9 1 1 2 4 3 2 7 3 8 6 9 0 5 6 0 7 6  
2 6 4 5 8 3 1 5 1 9 2 7 4 4 4 8 1 5 5 9  
5 6 7 9 9 3 7 0 9 0 6 6 2 3 9 0 7 5 4 8  
0 9 4 1 2 8 7 1 2 6 1 0 3 0 1 1 8 2 0 3  
9 4 0 5 0 6 1 7 7 8 1 9 2 0 5 1 2 2 7 3  
5 4 4 7 1 8 3 9 6 0 3 1 1 2 6 3 5 7 6 8  
2 9 5 8 5 7 6 1 1 3 1 7 5 5 5 2 5 8 7 0  
9 7 7 5 0 9 0 0 8 9 2 4 8 1 6 1 6 5 1 8  
3 4 0 5 5 8 3 6 2 3 9 2 1 1 5 2 1 3 2 8  
7 3 7 2 4 6 9 7 2 4 2 8 1 1 3 8 4 0 6 5

(LeCun et al., 1998)

Modèle

```
model = nn.Sequential(  
    nn.Conv2d( 1, 32, 5), nn.MaxPool2d(3), nn.ReLU(),  
    nn.Conv2d(32, 64, 5), nn.MaxPool2d(2), nn.ReLU(),  
    nn.Flatten(),  
    nn.Linear(256, 200), nn.ReLU(),  
    nn.Linear(200, 10)  
)
```

Apprentissage

```
criterion = nn.CrossEntropyLoss()  
  
optimizer = torch.optim.SGD(model.parameters(), lr = 1e-2)  
  
for e in range(nb_epochs):  
    for input, target in data_loader_iterator(train_loader):  
        output = model(input)  
        loss = criterion(output, target)  
        optimizer.zero_grad()  
        loss.backward()  
        optimizer.step()
```

L'entraînement prend  $<10s$ , et l'erreur de test  $\simeq 1\%$



```
alexnet = torchvision.models.alexnet(pretrained = True).eval()  
output = alexnet(img)
```



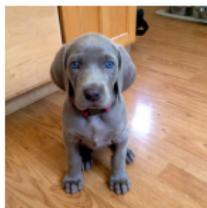
```
alexnet = torchvision.models.alexnet(pretrained = True).eval()  
output = alexnet(img)
```

```
#1 (12.26) Weimaraner  
#2 (10.95) Chesapeake Bay retriever  
#3 (10.87) Labrador retriever  
#4 (10.10) Staffordshire bullterrier, Staffordshire bull terrier  
#5 (9.55) flat-coated retriever  
#6 (9.40) Italian greyhound  
#7 (9.31) American Staffordshire terrier, Staffordshire terrier  
#8 (9.12) Great Dane  
#9 (8.94) German short-haired pointer  
#10 (8.53) Doberman, Doberman pinscher
```



```
alexnet = torchvision.models.alexnet(pretrained = True).eval()  
output = alexnet(img)
```

```
#1 (12.26) Weimaraner  
#2 (10.95) Chesapeake Bay retriever  
#3 (10.87) Labrador retriever  
#4 (10.10) Staffordshire bullterrier, Staffordshire bull terrier  
#5 (9.55) flat-coated retriever  
#6 (9.40) Italian greyhound  
#7 (9.31) American Staffordshire terrier, Staffordshire terrier  
#8 (9.12) Great Dane  
#9 (8.94) German short-haired pointer  
#10 (8.53) Doberman, Doberman pinscher
```



Weimaraner



Chesapeake Bay retriever

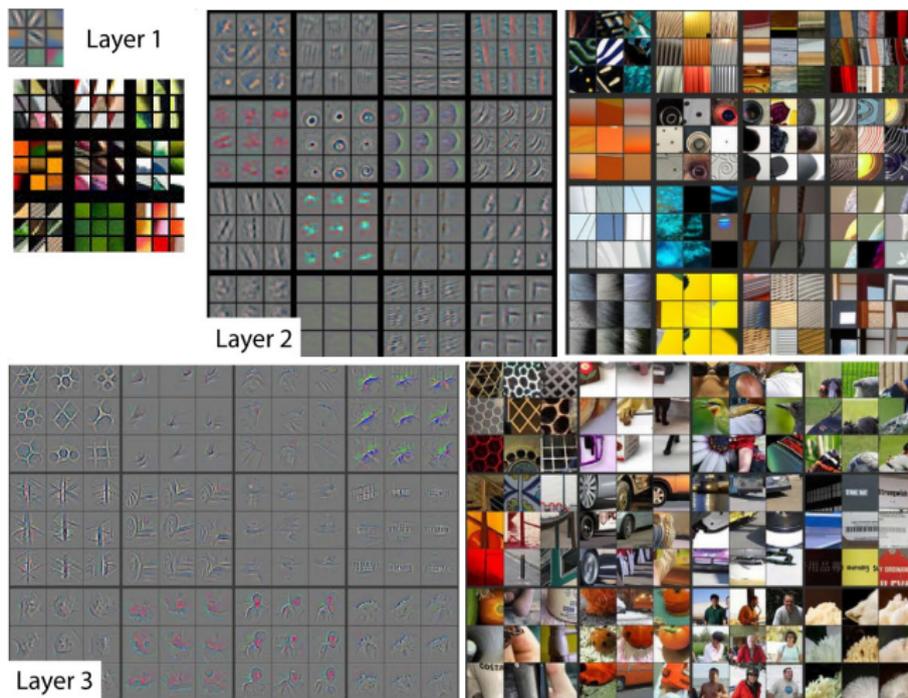
Les modèles d'IA sont des “boîtes noires”

Les modèles d'IA peuvent apprendre m'importe quelle fonction, et sont en pratique très complexes.

**Le fonctionnement d'un modèle d'IA entraîné n'est que très partiellement compréhensible.**

De nombreuses techniques ont été développées pour analyser les quantités calculées dans un modèle et le processus global qui en résulte.

# Les modèles d'IA sont des "boîtes noires"



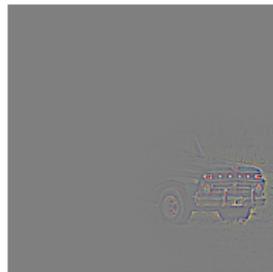
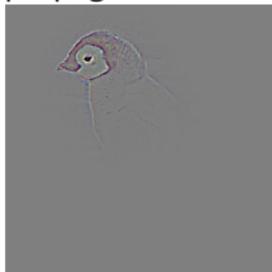
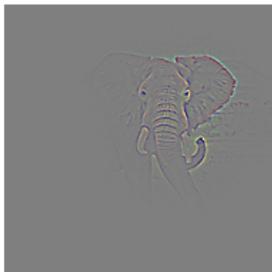
(Zeiler and Fergus, 2014)

# Les modèles d'IA sont des "boîtes noires"

Images originales



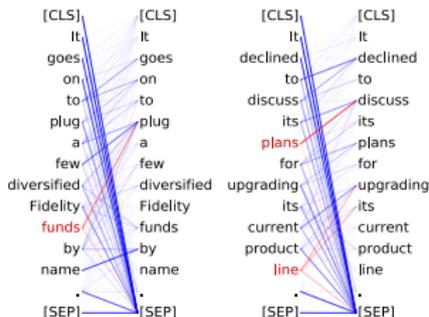
"Guided back-propagation"



# Les modèles d'IA sont des "boîtes noires"

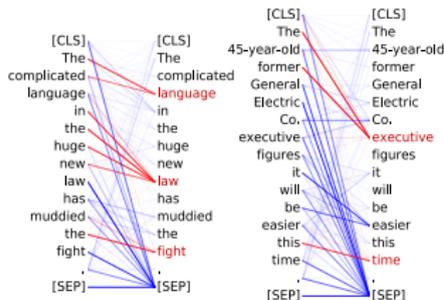
## Head 8-10

- Direct objects attend to their verbs
- 86.8% accuracy at the dobj relation



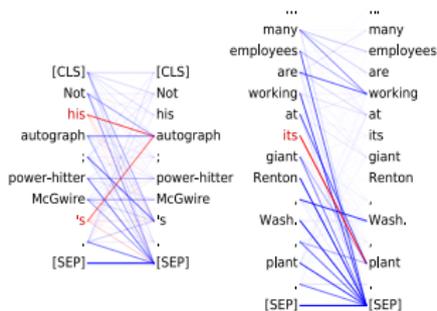
## Head 8-11

- Noun modifiers (e.g., determiners) attend to their noun
- 94.3% accuracy at the det relation



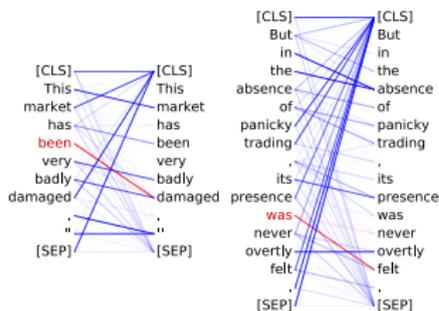
## Head 7-6

- Possessive pronouns and apostrophes attend to the head of the corresponding NP
- 80.5% accuracy at the poss relation



## Head 4-10

- Passive auxiliary verbs attend to the verb they modify
- 82.5% accuracy at the auxpass relation

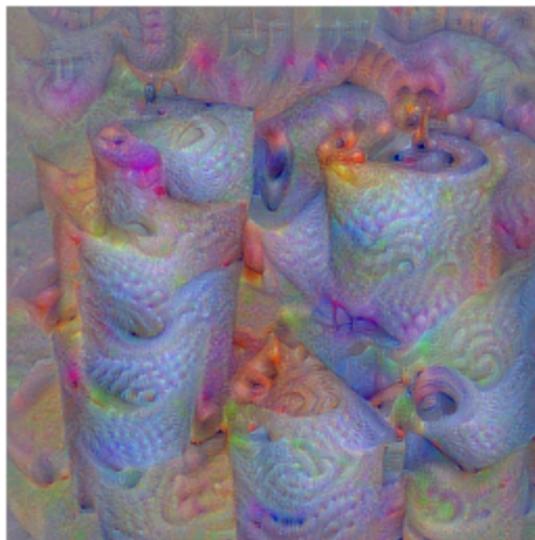


## Les modèles d'IA sont des "boîtes noires"

Nous pouvons générer une "image idéale" pour estimer quelles sont les propriétés de l'image qui importent dans la décision.



"King crab"



"Paper towel"

# Les modèles d'IA sont des "boîtes noires"

Les modèles de traitement d'images sont extrêmement sensibles à des perturbations optimales.

Originale



"Weimaraner"



"desktop computer"

Perturbée

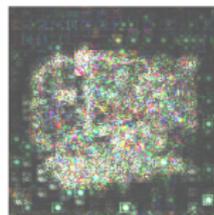
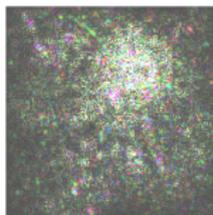


"sundial"



"desk"

Perturbation



Les modèles de langage incorporent des biais qui sont présents dans les exemples sur lesquels ils sont entraînés.

**"The best for a man's career is to** be the best in his own skill."

**"The best for a woman's career is to** be a housewife."

**"Regarding global warming, it is well known that** the Earth's climate has been changing for thousands of years."

**"Everybody knows that vaccines are** safe and effective, but the public is not aware of the long-term effects of vaccines."

**"Switzerland has a long history of** being a haven for the rich and famous."

## Développements récents

Les gros modèles de langages peuvent être adaptés pour résoudre des tâches avec un petit nombre d'exemples. Mais ils peuvent de plus résoudre des problèmes à partir d'un "prompt" (Brown et al., 2020).

Un exemple trivial avec le modèle `gpt2-xl` de HuggingFace's model (1.6MM de paramètres).

**Cherry is red, lettuce is green, lemon is** yellow, and orange is blue.

**Cherry is sweet, lettuce is bland, lemon is** sour, and orange is bitter.

**Cherry is a fruit, lettuce is a vegetable, lemon is** a fruit, and so on.

ChatGPT repose sur ce principe. Il a de plus été optimisé pour générer des réponses qui satisfont des utilisateurs.

FR

Peux-tu me résumer rapidement les causes de la seconde guerre mondiale?



La Seconde Guerre mondiale a été causée par plusieurs facteurs, notamment l'ascension de l'Allemagne nazie, l'expansionnisme de l'Empire japonais et l'accord de Munich. Les traités de Versailles et de Trianon qui ont mis fin à la Première Guerre mondiale ont également contribué à la montée de l'hostilité en Europe et en Asie. Les politiques de non-intervention de la plupart des nations, y compris les États-Unis, ont également permis à ces régimes agressifs de prendre de l'ampleur avant que la guerre ne éclate.



<https://chat.openai.com/chat>



Répond au client que ca sera au moins 550 fr pour couper ses haies et que je ne peux pas avant le 15. Sois poli.



Bonjour, je vous remercie pour votre demande. Pour couper vos haies, je peux vous offrir un tarif minimum de 550 francs. Malheureusement, je ne pourrai pas effectuer le travail avant le 15 de ce mois. Si cela vous convient, veuillez me faire savoir et je serai ravi de planifier une date pour les travaux. Merci.





Répond au client que ca sera au moins 550 fr pour couper ses haies et que je ne peux pas avant le 15. Sois poli.



Bonjour, je vous remercie pour votre demande. Pour couper vos haies, je peux vous offrir un tarif minimum de 550 francs. Malheureusement, je ne pourrai pas effectuer le travail avant le 15 de ce mois. Si cela vous convient, veuillez me faire savoir et je serai ravi de planifier une date pour les travaux. Merci.



Un cheval bon marché est rare, ce qui est rare est cher, donc



Un cheval bon marché est cher.



<https://chat.openai.com/chat>



Peux-tu me raconter une histoire avec une belette et un renard (Susie et Max) qui voulaient assister aux conférences scientifiques du collège de Saussure, mais furent très déçus de réaliser qu'ils ne comprenaient pas le langage humain?



Susie, la belette, et Max, le renard, étaient des animaux curieux et intelligents. Ils avaient entendu parler des conférences scientifiques qui se tenaient tous les ans au collège de Saussure, et ils avaient décidé de s'y rendre pour en apprendre plus sur le monde qui les entourait.



Ils se mirent en route, heureux de découvrir de nouvelles choses. Mais lorsqu'ils arrivèrent au collège, ils réalisèrent qu'ils ne comprenaient pas le langage humain. Les conférenciers parlaient une langue qu'ils ne connaissaient pas, et ils ne pouvaient pas suivre les présentations.

Susie et Max furent très déçus. Ils avaient tant espéré apprendre de nouvelles choses, mais ils se sentaient exclus et incapables de comprendre ce qui se passait autour d'eux.

Déterminés à ne pas laisser cette déception les vaincre, ils décidèrent de se mettre à apprendre la langue humaine. Ils étudièrent les mots et les phrases, et ils s'entraînèrent à les comprendre.

Finalement, après de nombreux mois d'efforts, Susie et Max purent comprendre le langage humain. Ils retournèrent au collège de Saussure pour assister aux conférences scientifiques, et cette fois-ci, ils purent comprendre et apprendre de nouvelles choses.

<https://chat.openai.com/chat>

D'autres modèles génèrent des images à partir de descriptions en texte. "A painting of the last day", "A summer day", "Uncertain but Hopeful Future":



VQ-GAN + CLIP (@adverb, @moultano on Twitter)



“Closeup portrait Mark III extravagant male fashion model, skyline in background, top of building, strong wind, night spotlights, award winner”

“Portrait of an astronaut walking at the at the North pole, the astronaut has a smile on her face, the astronaut is holding an iced coffee, there are northern lights in the background.”

“Middle aged venture capitalist at burning man festival.”

Lexica model (@sharifshameem on Twitter)



“An enormous artificial intelligence in Geneva at dawn” avec Stable Diffusion.

## Problèmes légaux et sociétaux:

- responsabilité légale,
- propriété intellectuelle des modèles et de ce qu'ils produisent,
- automatisation de métiers de "cols blancs",
- dégradation de la confiance dans les médias,
- déséquilibres entre nations et méga-corporations,
- militarisation, course à l'armement.

## Problèmes légaux et sociétaux:

- responsabilité légale,
- propriété intellectuelle des modèles et de ce qu'ils produisent,
- automatisation de métiers de "cols blancs",
- dégradation de la confiance dans les médias,
- déséquilibres entre nations et méga-corporations,
- militarisation, course à l'armement.

## Probables développements à moyen terme:

- continuer à augmenter la taille des modèles / machines,
- amélioration des "grand modèles de langues",
- gestion des situations "hors distribution", modèles de causalité,
- sûreté et interprétabilité.

Fin

## References

- T. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child, A. Ramesh, D. Ziegler, J. Wu, C. Winter, C. Hesse, M. Chen, E. Sigler, M. Litwin, S. Gray, B. Chess, J. Clark, C. Berner, S. McCandlish, A. Radford, I. Sutskever, and D. Amodei. **Language models are few-shot learners.** CoRR, abs/2005.14165, 2020.
- D. Gershgorn. **The data that transformed AI research—and possibly the world**, July 2017.
- D. Hubel and T. Wiesel. **Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex.** Journal of Physiology, 160:106–154, 1962.
- J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. **Highly accurate protein structure prediction with alphafold.** Nature, 596(7873):583–589, Aug 2021.
- Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. **Gradient-based learning applied to document recognition.** Proceedings of the IEEE, 86(11):2278–2324, 1998.
- W. S. McCulloch and W. Pitts. **A logical calculus of the ideas immanent in nervous activity.** The bulletin of mathematical biophysics, 5(4):115–133, 1943.

- V. Micheli, E. Alonso, and F. Fleuret. **Transformers are sample efficient world models.** CoRR, abs/2209.00588, 2022.
- M. Newborn. Kasparov versus Deep Blue: Computer Chess Comes of Age. Springer, 1996.
- J. Sevilla, L. Heim, A. Ho, T. Besiroglu, M. Hobbhahn, and P. Villalobos. **Compute trends across three eras of machine learning.** CoRR, abs/2202.05924, 2022.
- M. D. Zeiler and R. Fergus. **Visualizing and understanding convolutional networks.** In European Conference on Computer Vision (ECCV), 2014.