

INTELLIGENCE ARTIFICIELLE, ENTRE PROMESSES 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”

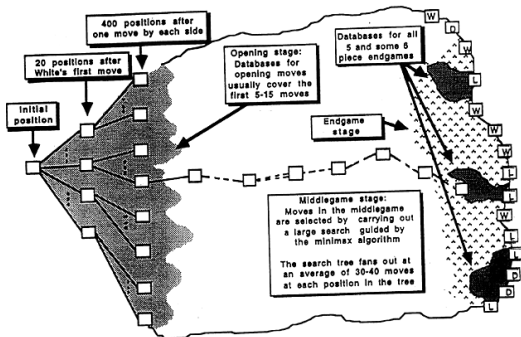
La manière traditionnelle de spécifier à un ordinateur comment résoudre une tâche consiste à la décomposer explicitement en pas élémentaires.

```
n = 15345

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

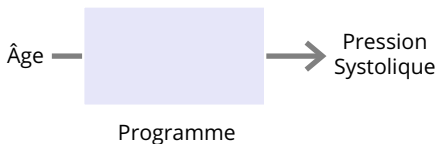
Les premières tentatives de conception de systèmes d'Intelligence Artificielle, par exemple pour des médicales, des jeux de stratégies, ou de la vision par ordinateur, reposaient sur le même paradigme.

Chess game tree

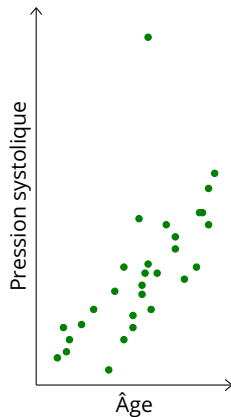
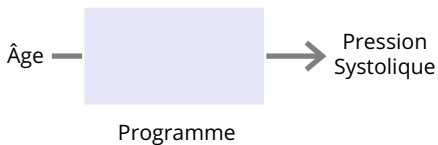


(Newborn, 1996)

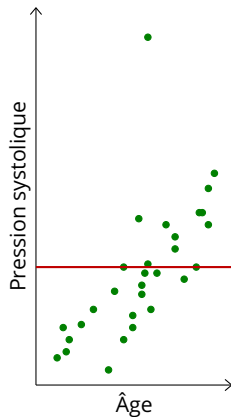
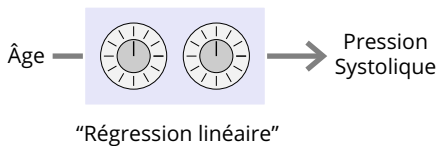
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



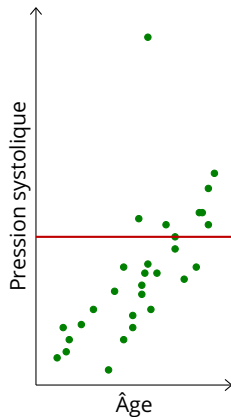
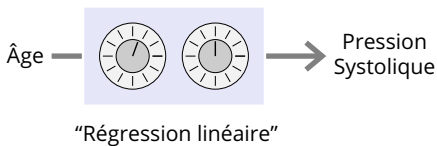
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



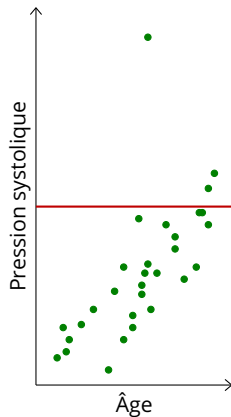
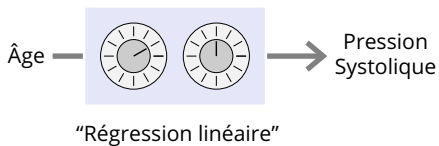
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



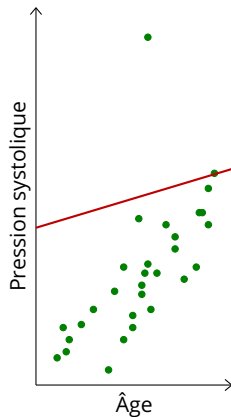
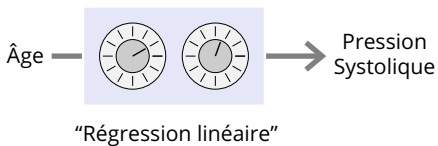
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



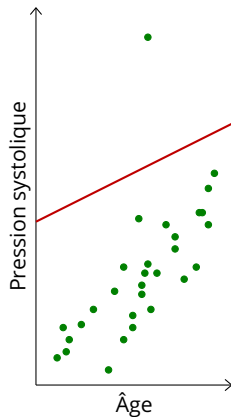
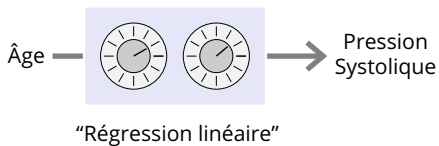
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



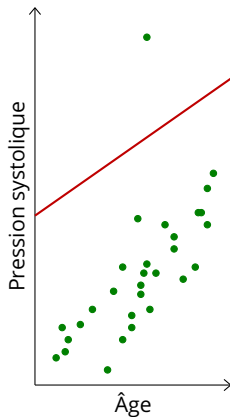
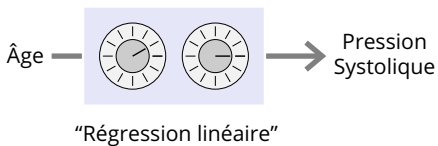
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



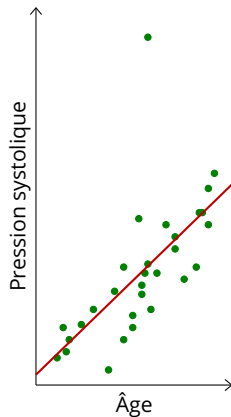
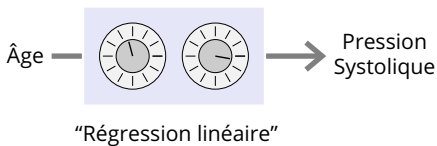
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



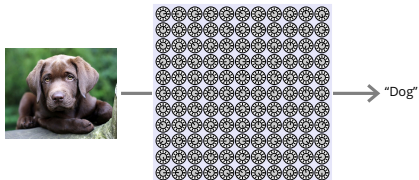
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



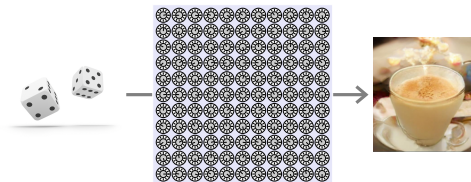
L'idée fondamentale de l'apprentissage statistique est d'adapter automatiquement un programme pour qu'il fonctionne sur des exemples.



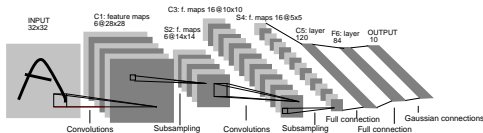
Elle peut être utilisée à grande échelle pour extraire de l'information de signaux naturels tels qu'une image, un échantillon sonore ou un texte



ou pour produire un signal complexe et réaliste

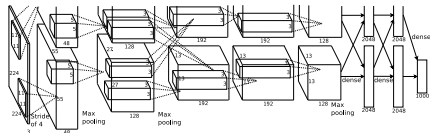


Les modèles modernes intègrent 10^5 à 10^{11} paramètres.

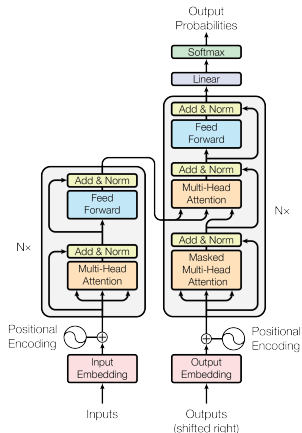


Convolutions (LeNet, 1989)

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

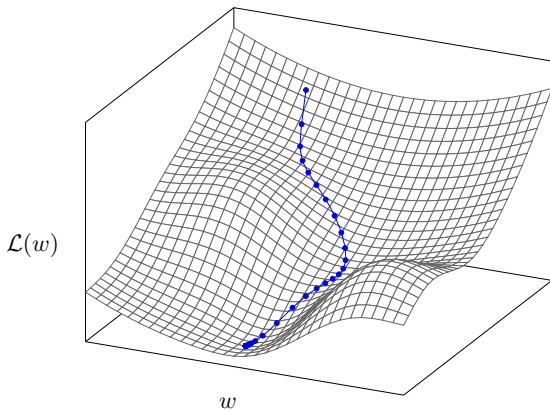


Très gros modèle + GPUs (AlexNet, 2012)



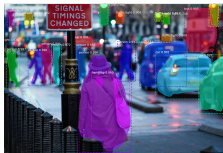
Attention (Transformer, 2018)

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

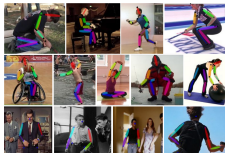


L'IA résout réellement des problèmes

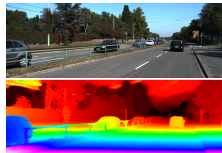
L'IA résout réellement des problèmes



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

L'IA résout réellement des problèmes



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

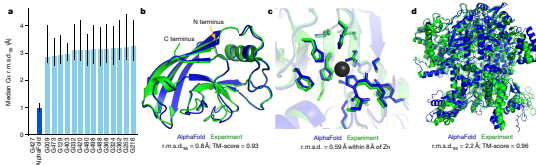
"Portrait of an astronaut walking at the at the North pole, the astronaut has a smile on her face, the astronaut is holding an iced coffee, there are northern lights in the background."

"Middle aged venture capitalist at burning man festival."

Lexica model (@sharifshameem on Twitter)

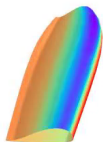
L'IA permet de surcroît de traiter des tâches que les humains ne peuvent pas résoudre correctement.

Repliage de protéines.



(Jumper et al., 2021)

Optimisation de formes.



Bike designed with artificial intelligence breaks world speed records



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".

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

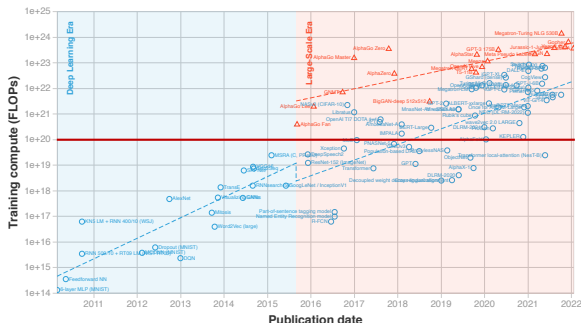
jeu de données	Année	Nb. livres (250p)	Taille*
SST2	2013	40K	20Mb
WMT-18	2018	14M	7Gb
The Pile	2020	1.6MM	825Gb
OSCAR	2020	12MM	6Tb

(* 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.

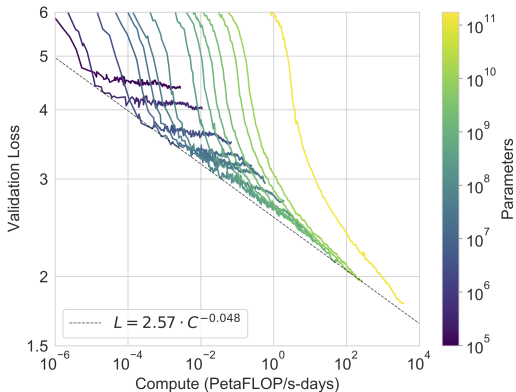
Training compute (FLOPs) of milestone Machine Learning systems over time
n = 102



Énergie \approx
 2'000km en voiture

(Sevilla et al., 2022)

Cette croissance ne semble pas ralentir.



(Brown et al., 2020)

L'IA est facile à utiliser

Réseau de classification de caractères + procédure d'apprentissage.

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)
```



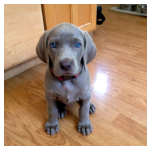
```
alexnet = torchvision.models.alexnet(pretrained = True).eval()  
output = alexnet(img)
```

```
#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

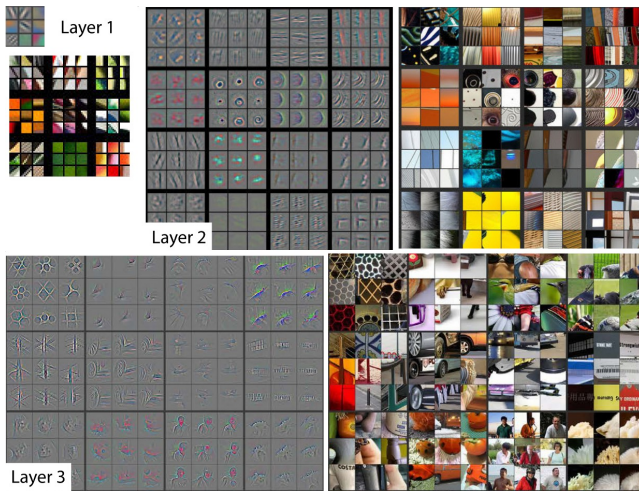
Les modèles d'IA sont des “boîtes noires”

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"

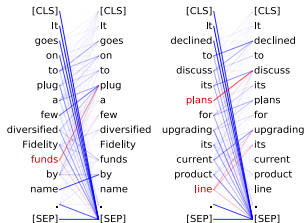


(Zeiler and Fergus, 2014)

Les modèles d'IA sont des "boîtes noires"

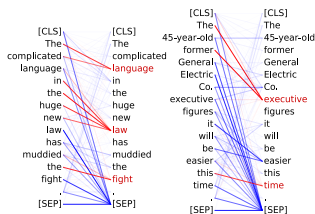
Head 8-10

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



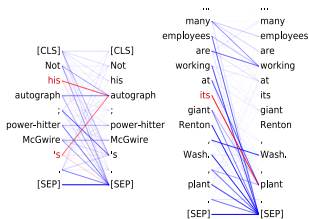
Head 8-11

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



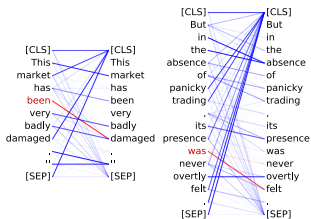
Head 7-6

- **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"

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

Originale



"Weimaraner"



"desktop computer"

Perturbée

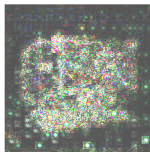
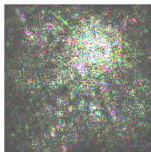


"sundial"



"desk"

Perturbation



Développements récents

Les **modèles de langages** estiment la probabilités du prochain mots d'un texte, étant donnés les mots qui précèdent.

System Prompt
(human-written)

In a shocking finding, scientist discovered a herd of unicorns living in a remote, previously unexplored valley, in the Andes Mountains. Even more surprising to the researchers was the fact that the unicorns spoke perfect English.

Model Completion
(machine-written, 10
tries)

The scientist named the population, after their distinctive horn, Ovid's Unicorn. These four-horned, silver-white unicorns were previously unknown to science.

Now, after almost two centuries, the mystery of what sparked this odd phenomenon is finally solved.

Dr. Jorge Pérez, an evolutionary biologist from the University of La Paz, and several companions, were exploring the Andes Mountains when they found a small valley, with no other animals or humans. Pérez noticed that the valley had what appeared to be a natural fountain, surrounded by two peaks of rock and silver snow.

GPT2 (14.02.2019)

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.”

Les gros modèles de ce type peuvent résoudre des problèmes à partir d'un "prompt" (Brown et al., 2020). Un exemple simple avec le modèle `gpt2` de Hugging Face (120M de paramètres).

I: water boils at 100 degrees, O: physics, I: the square root of two is irrational, O: mathematics, I: the set of prime numbers is infinite, O: mathematics, I: gravity is proportional to the mass, O:

Les gros modèles de ce type peuvent résoudre des problèmes à partir d'un "prompt" (Brown et al., 2020). Un exemple simple avec le modèle `gpt2` de Hugging Face (120M de paramètres).

I: water boils at 100 degrees, O: physics, I: the square root of two is irrational, O: mathematics, I: the set of prime numbers is infinite, O: mathematics, I: gravity is proportional to the mass, O: **physics**,

I: water boils at 100 degrees, O: physics, I: the square root of two is irrational, O: mathematics, I: the set of prime numbers is infinite, O: mathematics, I: squares are rectangles, O: **mathematics**,

Les gros modèles de ce type peuvent résoudre des problèmes à partir d'un "prompt" (Brown et al., 2020). Un exemple simple avec le modèle [gpt2](#) de Hugging Face (120M de paramètres).

I: water boils at 100 degrees, O: physics, I: the square root of two is irrational, O: mathematics, I: the set of prime numbers is infinite, O: mathematics, I: gravity is proportional to the mass, O: **physics**,

I: water boils at 100 degrees, O: physics, I: the square root of two is irrational, O: mathematics, I: the set of prime numbers is infinite, O: mathematics, I: squares are rectangles, O: **mathematics**,

I: I love apples, O: positive, I: music is my passion, O: positive, I: my job is boring, O: negative, I: frozen pizzas are awesome, O:

Les gros modèles de ce type peuvent résoudre des problèmes à partir d'un "prompt" (Brown et al., 2020). Un exemple simple avec le modèle `gpt2` de Hugging Face (120M de paramètres).

I: water boils at 100 degrees, O: physics, I: the square root of two is irrational, O: mathematics, I: the set of prime numbers is infinite, O: mathematics, I: gravity is proportional to the mass, O: **physics**,

I: water boils at 100 degrees, O: physics, I: the square root of two is irrational, O: mathematics, I: the set of prime numbers is infinite, O: mathematics, I: squares are rectangles, O: **mathematics**,

I: I love apples, O: positive, I: music is my passion, O: positive, I: my job is boring, O: negative, I: frozen pizzas are awesome, O: **positive**,

I: I love apples, O: positive, I: music is my passion, O: positive, I: my job is boring, O: negative, I: frozen pizzas taste like cardboard, O: **negative**,

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



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>



HEARD: Jack has three apples, he eats two, and buy five more, how many does he have?

THOUGHT: Jack has three apples, he eats two [Jack has now one apple], and buy five more [Jack has now six apples], how many does he have?

SAID: Jack has six apples.

HEARD: Sophy has five pairs of socks, and Bob three. She gives him two, but then he lost all of them. How many pairs of socks does she have?

THOUGHT: Sophy has five pairs of socks, and Bob three. She gives him two [Sophy now has three pairs of socks and Bob has five], but then he lost all of them [Bob now has zero pairs of socks]. How many pairs of socks does she have?

SAID: Sophy has three pairs.

HEARD: Jim has nothing, and Sam has twelve meatballs. Sam gives five meatballs to Jim, and they each eat two meatballs. How many meatballs has Jim?

THOUGHT: Jim has nothing, and Sam has twelve meatballs. Sam gives five meatballs to Jim [Sam now has seven meatballs and Jim has five], and they each eat two meatballs [Sam now has five meatballs and Jim three]. How many meatballs has Jim?

SAID: Jim has three meatballs.

HEARD: My neighbor has four coins, he gives me one, and loses one, how many does he have?

THOUGHT:



My neighbor has four coins, he gives me one [My neighbor now has three coins], and loses one [My neighbor now has two coins]. How many does he have?



SAID: My neighbor has two coins.

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

FR

HEARD: I have four black cubes, I paint two in red and one in blue, then I paint one red cube in black, and finally paint the black and blue cubes in yellow. What colors are my cubes?

THOUGHT:



I have four black cubes, I paint two in red [I now have two red cubes and two black cubes], and one in blue [I now have two red cubes, one blue cube, and one black cube], then I paint one red cube in black [I now have one red cube, one blue cube, and two black cubes], and finally paint the black and blue cubes in yellow [I now have one red cube and three yellow cubes]. What colors are my cubes?



SAID: I have one red cube and three yellow cubes.

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

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.
- 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.
- 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.