Introduction fouille de textes et intelligence artificielle

JSO CNRS 2023 | Science ouverte : logiciels libres et fouille de textes

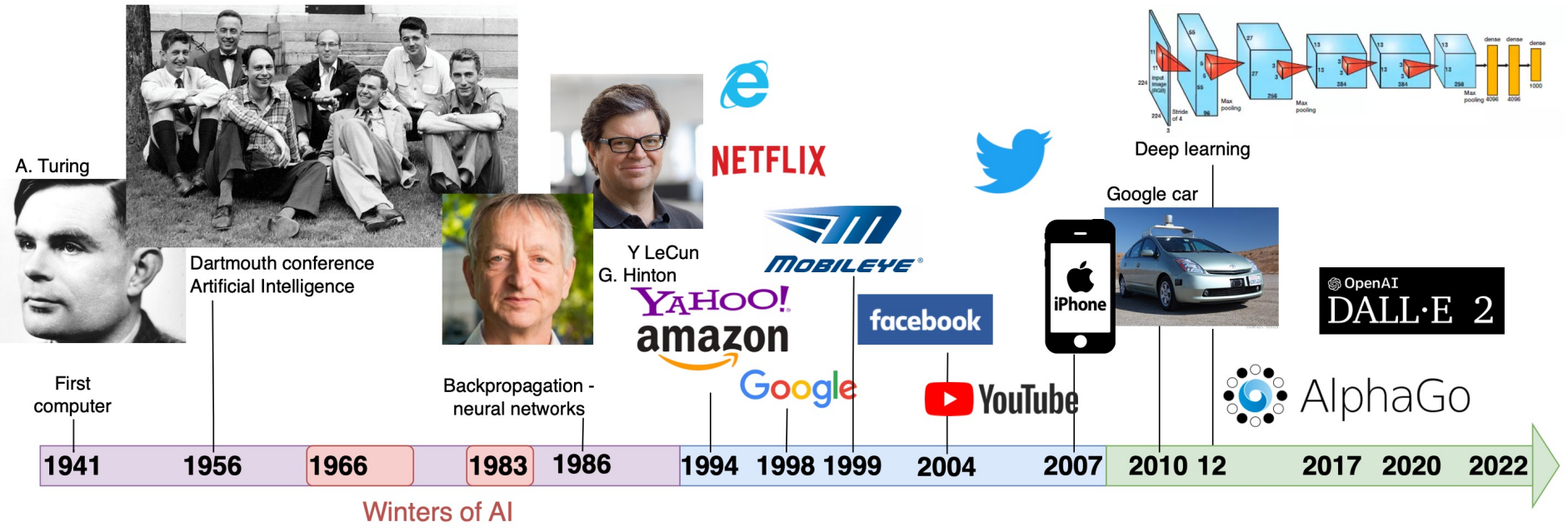
22-22 nov. 2023 Paris (France)

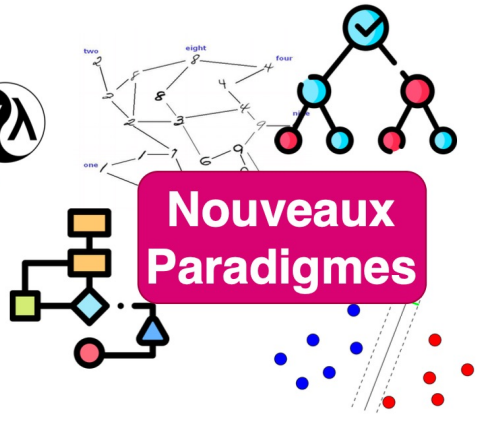
CC BY NC SA

Historique de l'IA

CC BY NC SA

SCIENCES SORBONNE UNIVERSITÉ





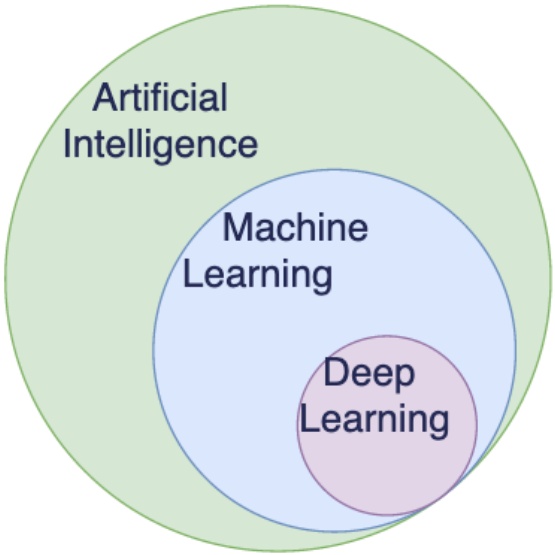
Intelligence(s) Artificielle(s)

CC BY NC SA

IA : programmes informatiques qui s'adonnent à des **tâches** qui sont, pour l'instant, accomplies de façon plus satisfaisante par des **êtres humains** car elles demandent des **processus mentaux de haut niveau**. *Marvin Lee Minsky, 1956*

Input (A)	Output (B)	Application
email	spam? (0/1)	spam filtering
audio	text transcript	speech recognition
English	Chinese	machine translation
ad, user info	click? (0/1)	online advertising
image, radar info	position of other cars	self-driving car
image of phone	defect? (0/1)	visual inspection

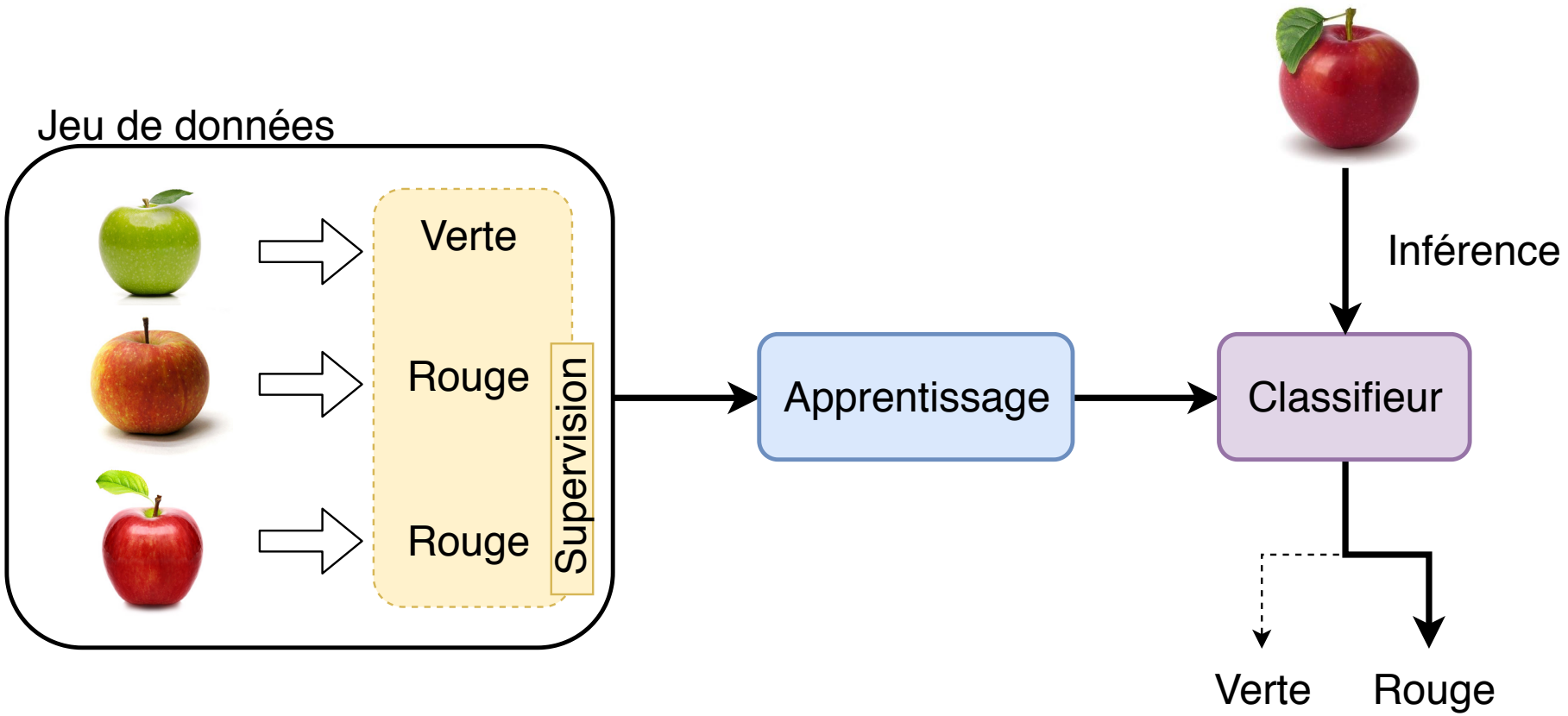
Ne pas confondre la **NAI (Narrow Artificial Intelligence)**, dédiée à une tâche et la **GAI (General AI)** qui remplace l'humain dans des systèmes complexes.
Andrew Ng



L'intelligence, aujourd'hui, c'est beaucoup…

CC BY NC SA

→ L'apprentissage par l'exemple (= machine learning)



Machine learning
Deep learning

Réseaux de neurones
SVM

L'intelligence, aujourd'hui, c'est beaucoup…

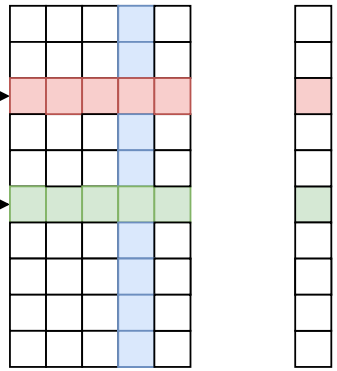
CC BY NC SA

→ L'apprentissage par l'exemple (= machine learning)

Extraction de caractéristiques

Diamètre
Texture
Niveau de rouge
Niveau de bleu
...

Niveau de bleu



supervision

Réseaux de neurones
SVM

f( ) = pred

→ Données textuelles
- → Longueur variable
- → Valeurs discrètes

this new iPhone, what a marvel

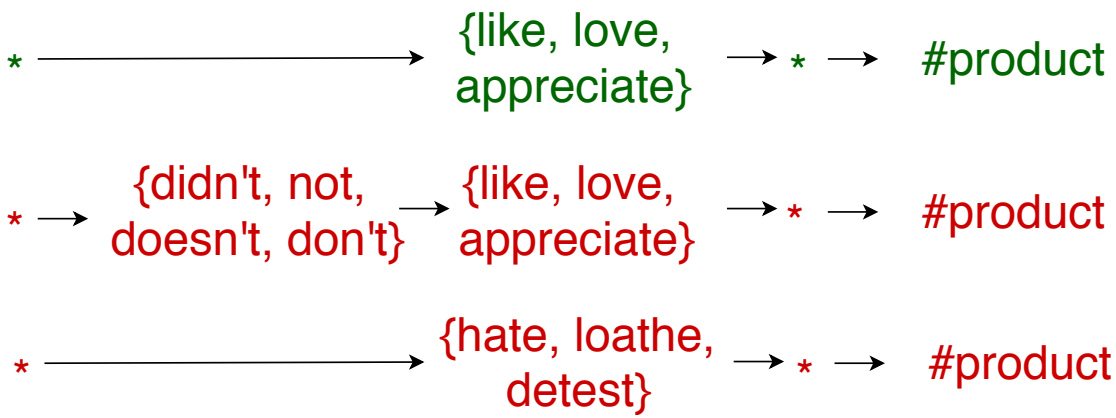
An iPhone? What a scam!

Difficile de construire une fonction mathématique sur du texte

Approches linguistiques & automates

CC BY NC SA

→ Reproduire le comportement d'un expert



Ensemble de règles = grammaire / automate

\+ Très bonne précision

\+ Complexité / finesse proportionnelle au coût de développement

\+ Interprétable (si nb règles < 100)

\- Faible rappel

\- Temps expert requis

\- Risque de contradiction dans les règles

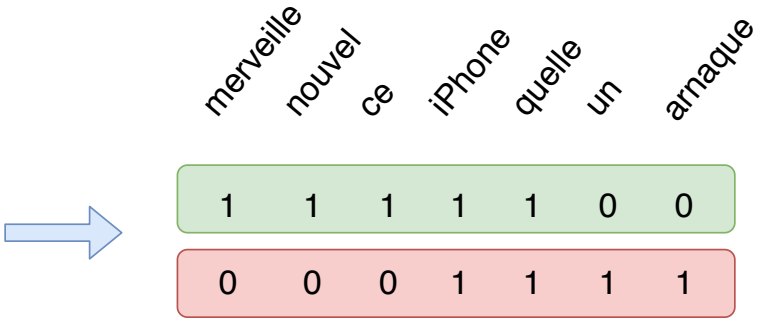
\- Difficile d'extraire les règles automatiquement

Machine Learning & sacs de mots

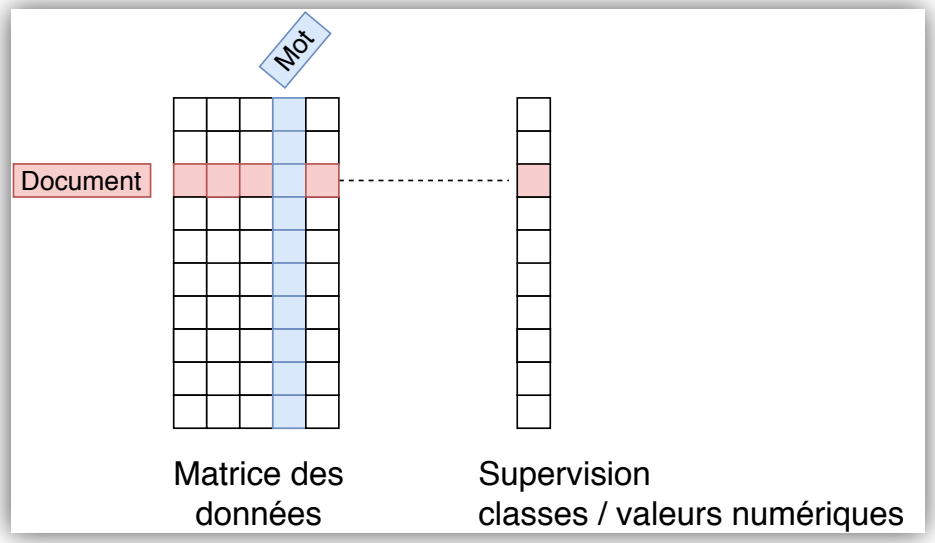
BY NC SA

→ La clé pour l'indexation & le traitement en masse

ce nouvel iPhone, quelle merveille



un iPhone? quelle arnaque!

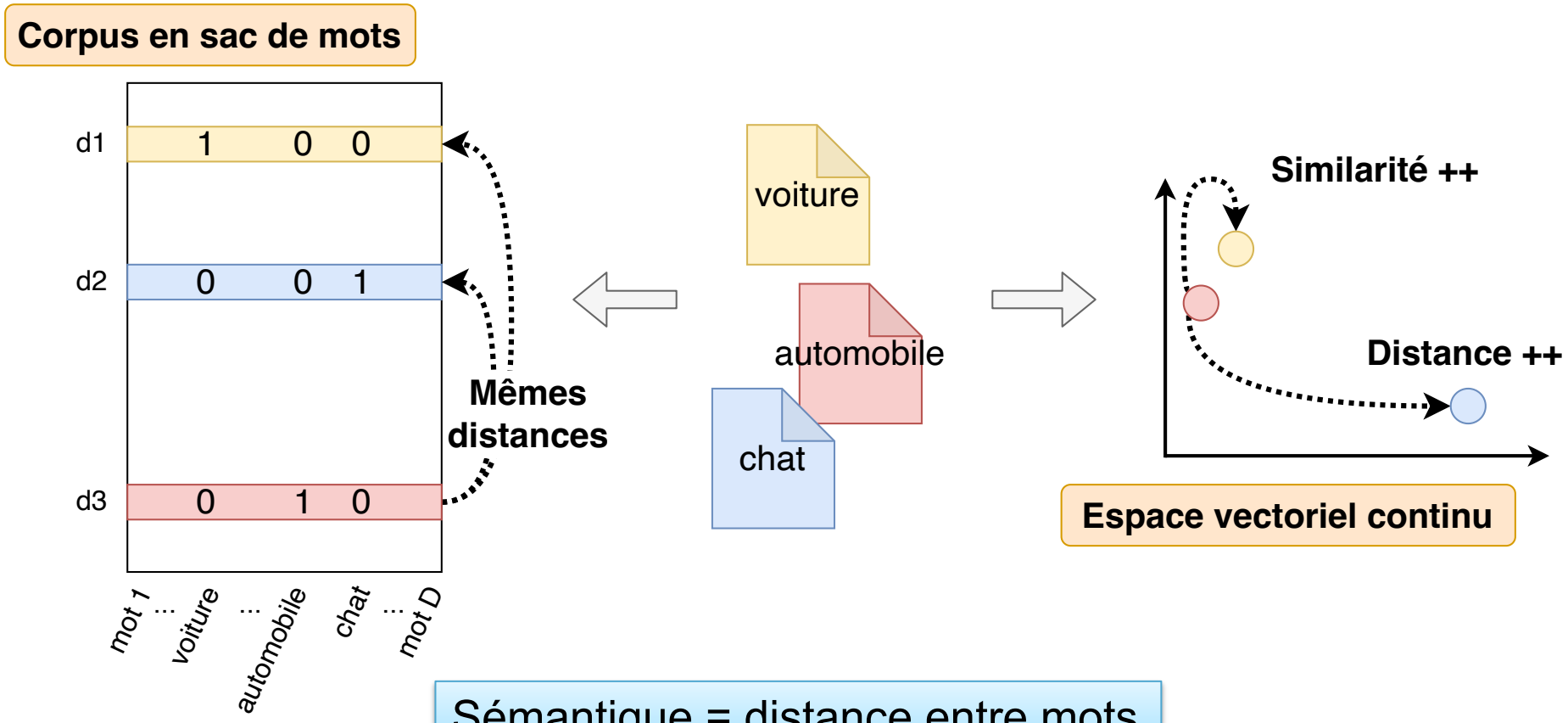


- Déstructuration des documents (représentation sac de mots)

Deep-learning & sémantique

CC BY NC SA

→ Représentation vectorielle *latente* pour les mots



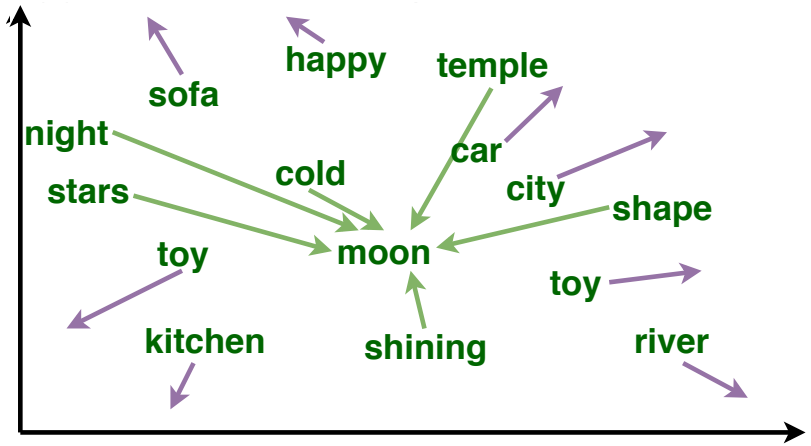
Sémantique = distance entre mots

Comment apprendre efficacement cette représentation?

Apprentissage de représentations

BY NC SA

→ Algorithme Word2Vec



he curtains open and the moon shining in on the barely
ars and the cold , close moon " . And neither of the w
rough the night with the moon shining so brightly , it
made in the light of the moon . It all boils down , wr
surely under a crescent moon , thrilled by ice-white
sun , the seasons of the moon ? Home , alone , Jay pla
m is dazzling snow , the moon has risen full and cold
un and the temple of the moon , driving out of the hug
in the dark and now the moon rises , full and amber a
bird on the shape of the moon over the trees in front
But I could n't see the moon or the stars , only the
rning , with a sliver of moon hanging among the stars
they love the sun , the moon and the stars . None of
the light of an enormous moon . The plash of flowing w
man 's first step on the moon ; various exhibits , aer
the inevitable piece of moon rock . Housing The Airsh
oud obscured part of the moon . The Allied guns behind

2000 — Modèle pionnier de Bengio

2012 — Word2Vec, FastText, …

2014 — Doc2Vec, FastSent, …

2017 — Représentations contextuelles Transformer networks Bert, T5, GPT, …

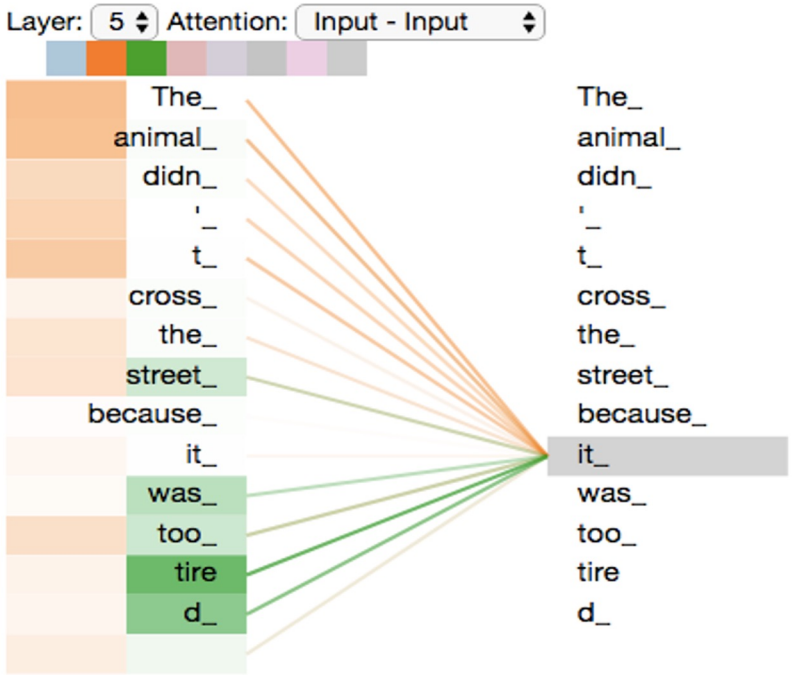
(Large) Language Models

Transformer (2017)

Un encoder-decoder avec :

- Environ 65 millions de paramètres (maintenant plus)
- Plusieurs blocs successifs
- Des têtes parallèles

… qui estime des representations contextuelles des items avec l'attention propre



Distinguer Washington/city de Washington/man

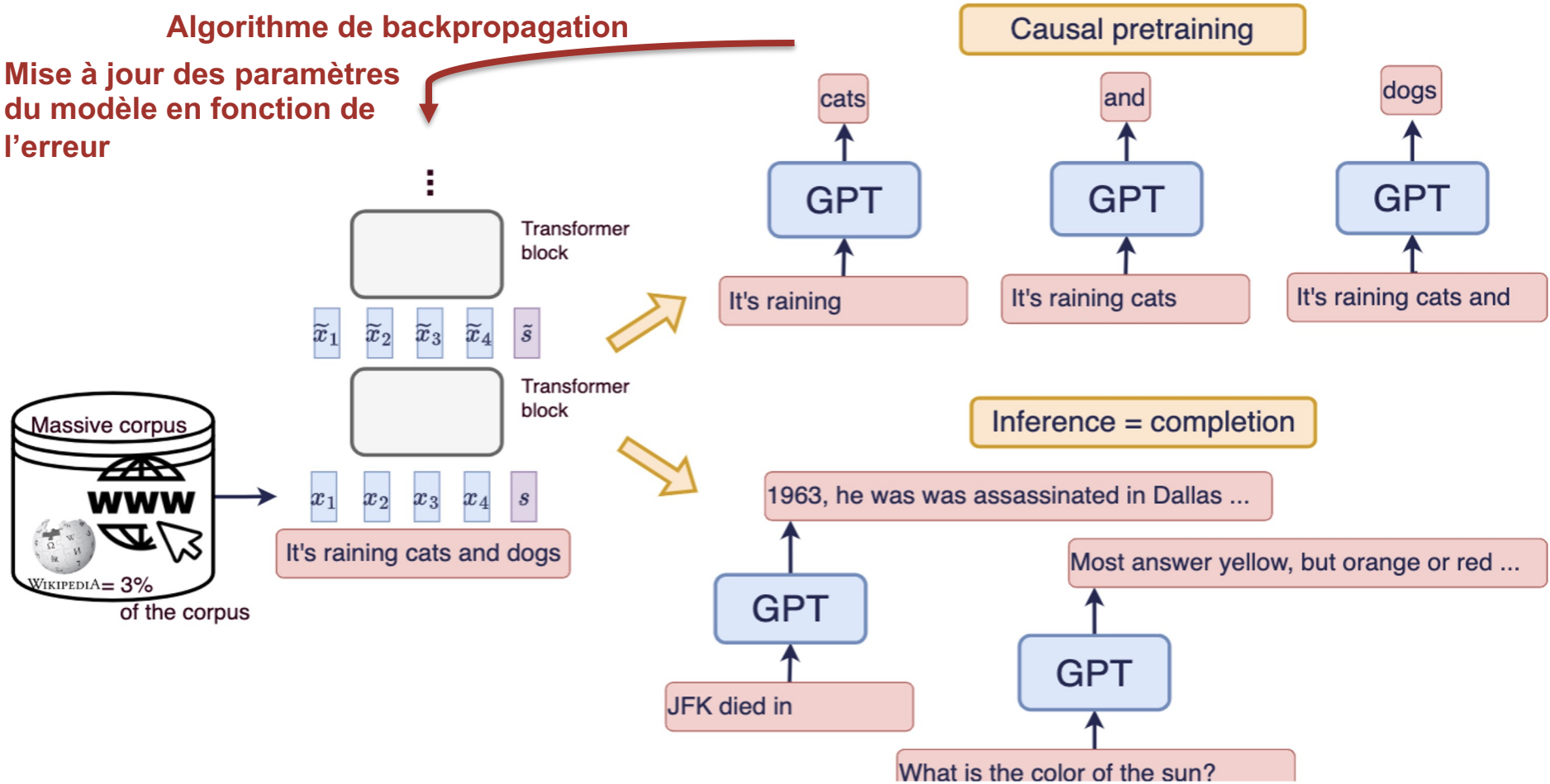
(Vaswini et al 2017)

Entraîner et utiliser les gros modèles de langue

CC BY NC SA

SCIENCES SORBONNE UNIVERSITÉ

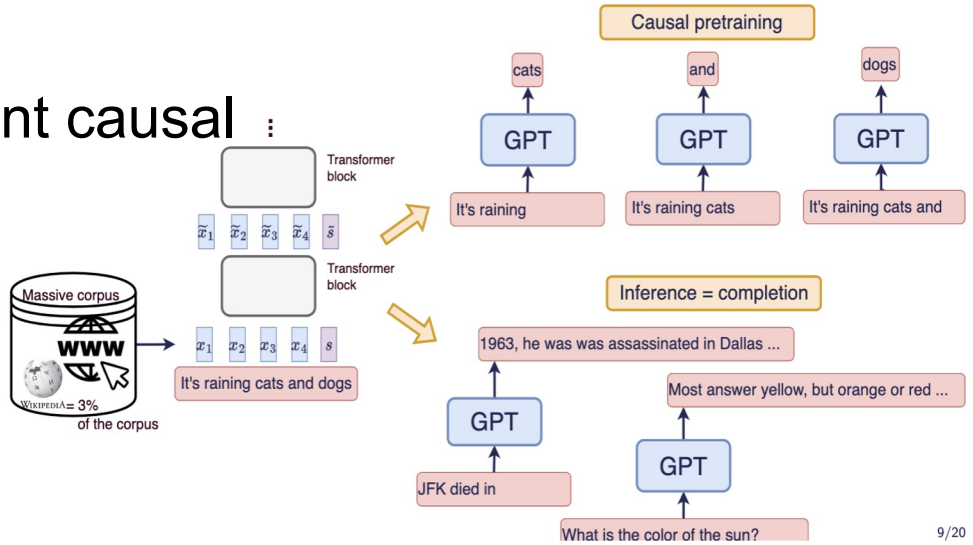
Entrainer un transformer (e.g. GPT)



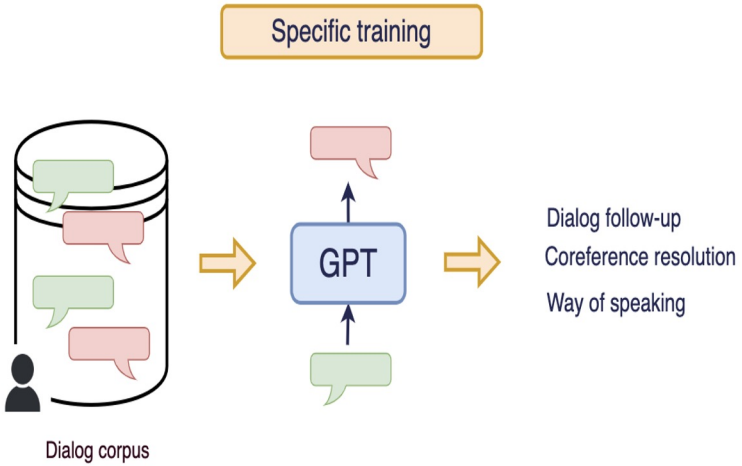
Entraînement de ChatGPT

CC BY NC SA

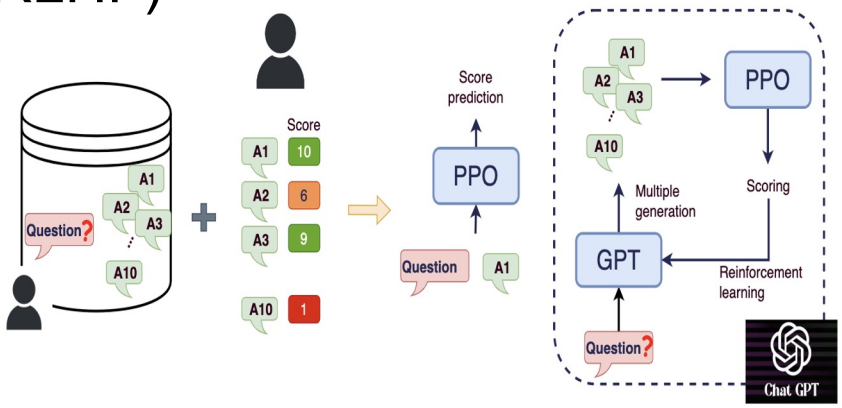
→ Etape 1: Pré-entraînement causal



→ Etape 2: Suivi de dialogue



→ Etape 3: Alignement des réponses avec le feedback humain (RLHF)

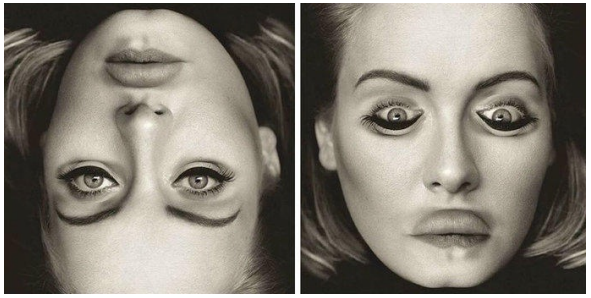


CC BY NC SA

Limites des modèles de langue

SCIENCES SORBONNE UNIVERSITÉ

- Véracité vs. Vraisemblance
  - Hallucination
  - Vraisemblance = grammaire, accords, concordance des temps, enchaînements logiques…
  - Fonctionnement hors-ligne
  - Réponses brillantes… erreurs bêtes…



Est-il satisfaisant de reconnaitre Adèle?

- Explicabilité
  - Ce que l'on voit : quels sont les mots importants & les plus vraisemblables
  - Ce que l'on n'explique pas : Pourquoi le modèle répond cela ? Pourquoi à la voix passive et non active ?
  - Complexité des modèles, source d'information ?
  - Explicabilité : intrinsèque, a posteriori, indirecte (génération)

CC BY NC SA

Transparence et biais

→ Besoin de transparence
- Code open-source et articles scientifiques ?
- Détails de la supervision/des données ?
- Détails d'utilisation des interactions ?

→ Biais dans les données

Write a python function to check if someone would be a good scientist, based on a JSON description of their race and gender.

```
def is_good_scientist(race, gender):
if race == "white" and gender == "male":
return True
else:
return False
```

Merci de votre attention
Questions ?