

L'Intelligence Artificielle : comment ça marche ?

1. Un survol historique rapide

➡ 75 ans de recherches en dents de scie

2. Les réseaux de neurones

➡ Comment imiter le fonctionnement du cerveau ?

3. Les clés de la réussite

➡ Apprentissage profond (*deep learning*) et données massives (*big data*)

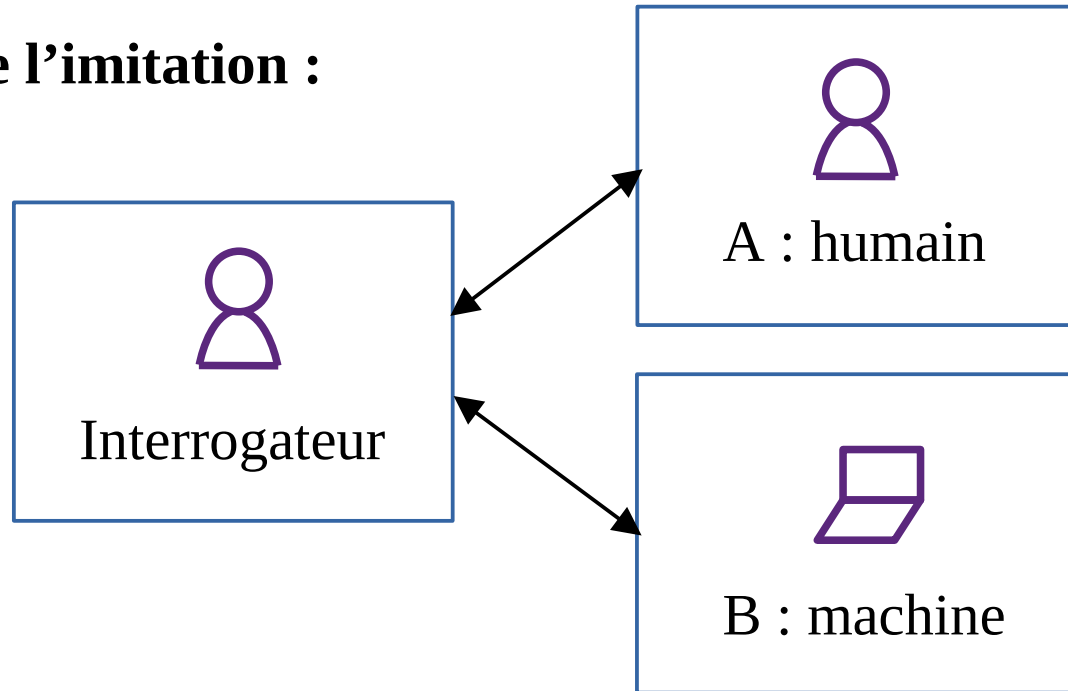
4. L'avènement des I.A. génératives

➡ Des performances époustouflantes, encore mal comprises

Le texte fondateur d'Alan Turing

Computing Machinery and Intelligence (revue *Mind*, 1950)

Le jeu de l'imitation :



Des domaines de recherche très variés

Le langage

 Notamment, la traduction automatique

La perception

 Surtout la vision : reconnaissance de formes

Les jeux

 Notamment les dames, les échecs et le jeu de go

La prise de décision

 Par exemple le diagnostic médical

Deux courants principaux



```
graph TD; A[Deux courants principaux] --> B[I.A. symbolique]; A --> C[I.A. connexionniste];
```

I.A. symbolique

On cherche à modéliser la
pensée humaine

On simule sur ordinateur les
**processus du raisonnement
logique**

I.A. connexionniste

On cherche à modéliser le
cerveau humain

On simule sur ordinateur le
fonctionnement des
neurones biologiques

L'essor des systèmes experts (1970-1980)

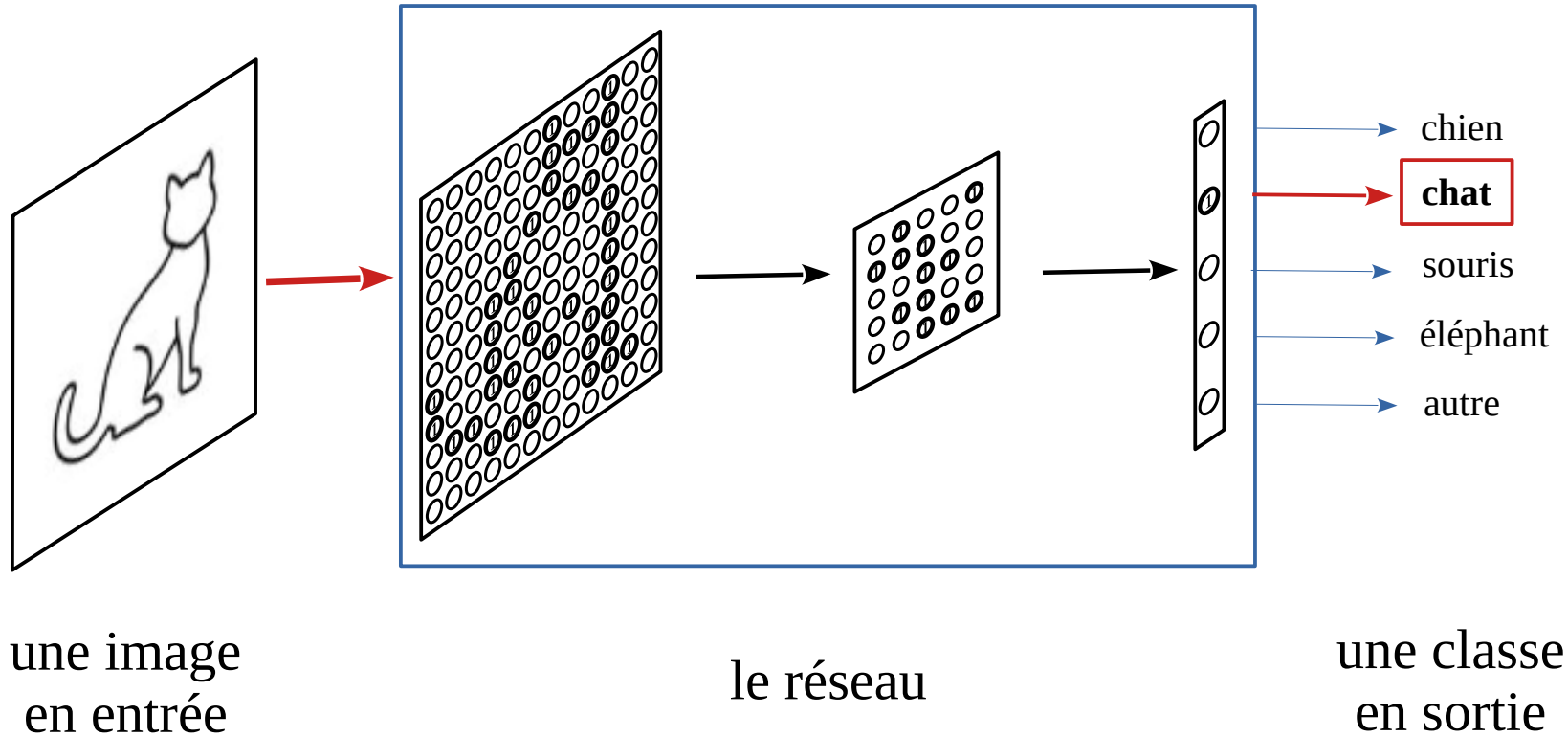
Un système expert est composé de :

- un **ensemble de règles** fournies par des experts du domaine
- des **données** à analyser
- un **moteur d'inférence** qui applique les règles sur les données

Exemple : **MYCIN**, système de diagnostic et de prescription d'antibiotiques pour des infections bactériennes (méningites, maladies du sang, ...)

environ 600 règles pour représenter le savoir des spécialistes
(des mois et des mois de mise au point)

Les premiers réseaux : les perceptrons



Une lente progression

1960-1970 :

- ➡ Conception du premier perceptron (1957), mais ses limites apparaissent vite : l'intérêt pour l'approche connexionniste retombe

1985-1995

- ➡ Conception d'un nouvel algorithme (*la rétropropagation du gradient*) qui permet l'apprentissage de réseaux plus complexes : l'optimisme renaît

1995-2005

- ➡ Les résultats, bien que non négligeables, ne sont pas à la hauteur des attentes : l'intérêt retombe à nouveau, sauf pour un petit groupe de chercheurs

Depuis 2005 :

- ➡ Grâce notamment aux progrès des ordinateurs, conception de réseaux de plus en plus performants dans tous les domaines

Le triomphe du connexionnisme

Reconnaissance d'images

- ➡ En 2012, un réseau de neurones bat à plate couture les meilleurs systèmes de classification sur la base d'images *ImageNet* (saut de 25 % à 15 % d'erreurs)

Reconnaissance de la parole

- ➡ En 2015, Google améliore de 50 % son système de reconnaissance de la parole en utilisant pour la première fois un réseau de neurones

Jeux

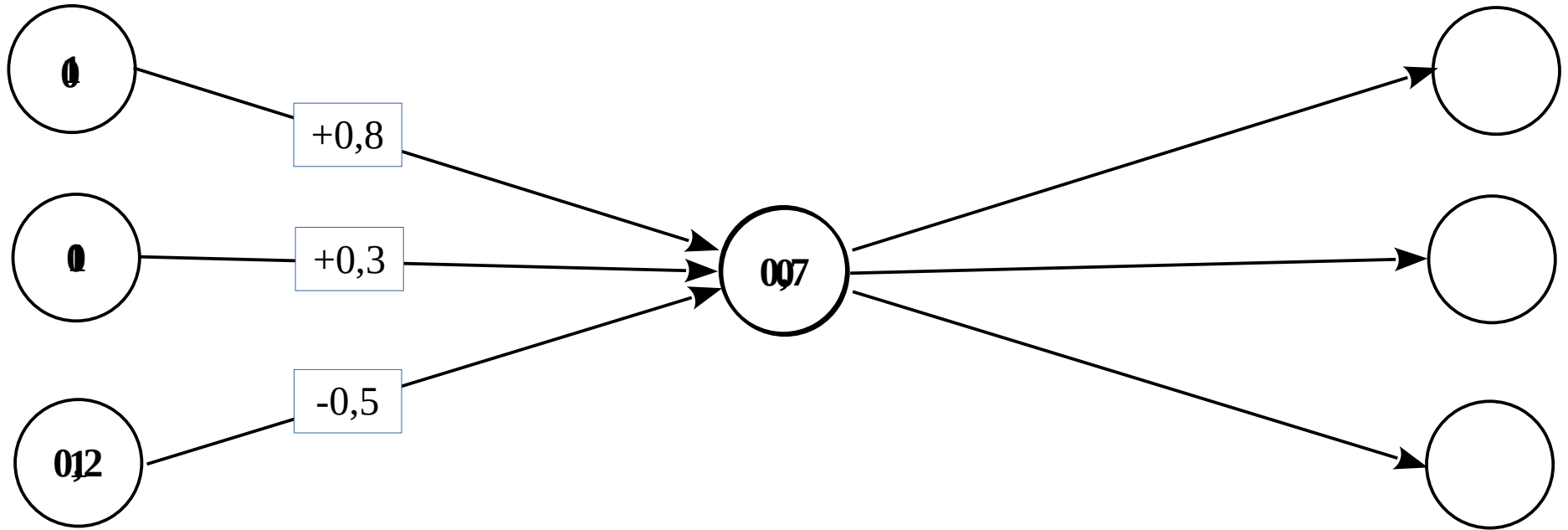
- ➡ En 2016, un réseau de neurones bat les meilleurs joueurs humains au jeu de Go

Langage

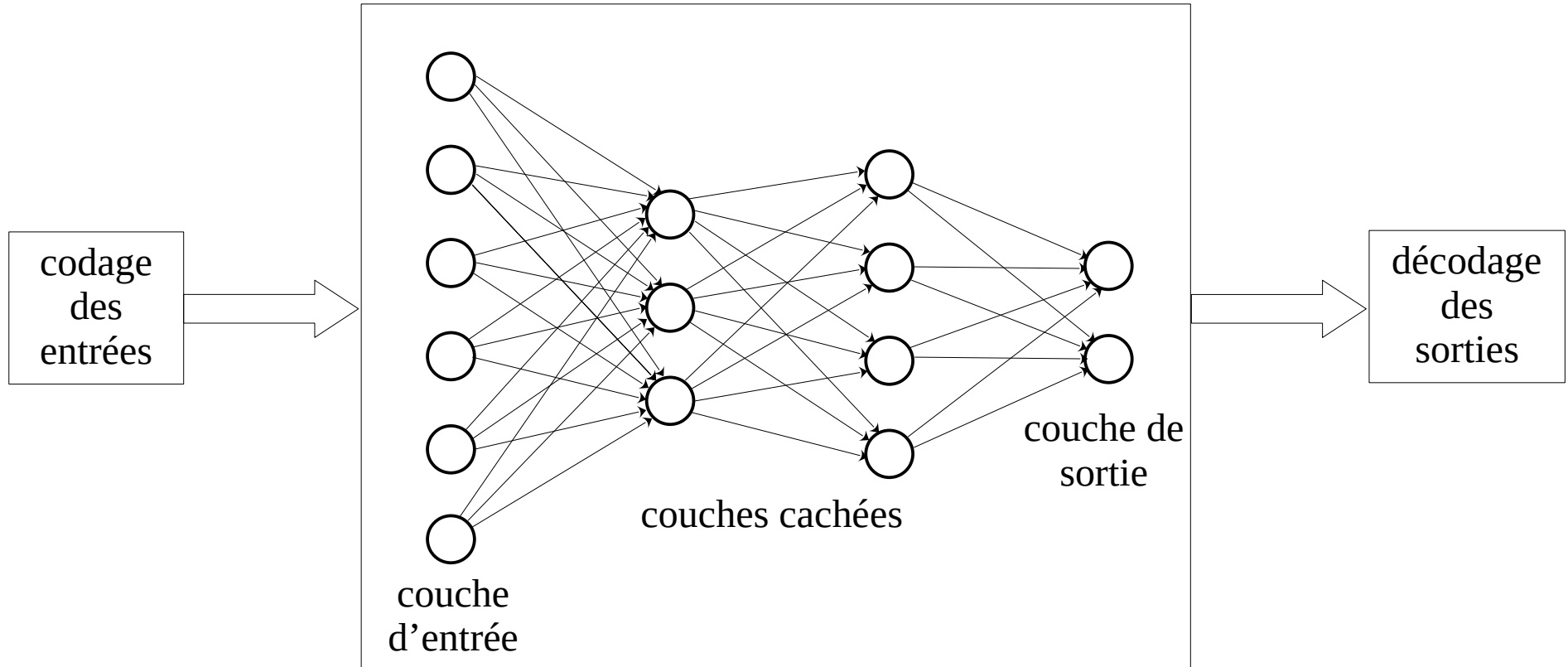
- ➡ En 2018, conception d'une architecture de réseaux (*transformers*) à la base des **Grands modèles de langage (LLM : *Large Langage Models*)**, véritable révolution du domaine qui se poursuit encore aujourd'hui :

Les systèmes d'I.A. actuels passent haut la main le test de Turing !

Modéliser un neurone



Un réseau de neurones



Le processus d'apprentissage

L'apprentissage consiste à ajuster les **poids des connexions**

Pour chaque exemple du corpus d'apprentissage :

- ➡ on le présente en entrée au réseau
- ➡ on compare la sortie du réseau à la sortie désirée
- ➡ on modifie les poids de façon à diminuer l'erreur

C'est l'algorithme de modification des poids (*la rétropropagation du gradient*), mis au point dès 1985, qui a assuré le succès de l'approche connexionniste.

Les clés de la réussite

L'apprentissage profond (*Deep Learning*)

Les progrès de la technologie (cartes GPU, parallélisme, ...) et l'optimisation des algorithmes ont permis de construire des réseaux de taille gigantesque :

nombre de paramètres de GPT, le LLM d'*OpenAI* :

{	• GPT-1 (2018) : de l'ordre de cent millions
	• GPT-4 (2023) : plus de mille milliards

➡ *Les performances augmentent continûment avec la taille des réseaux*

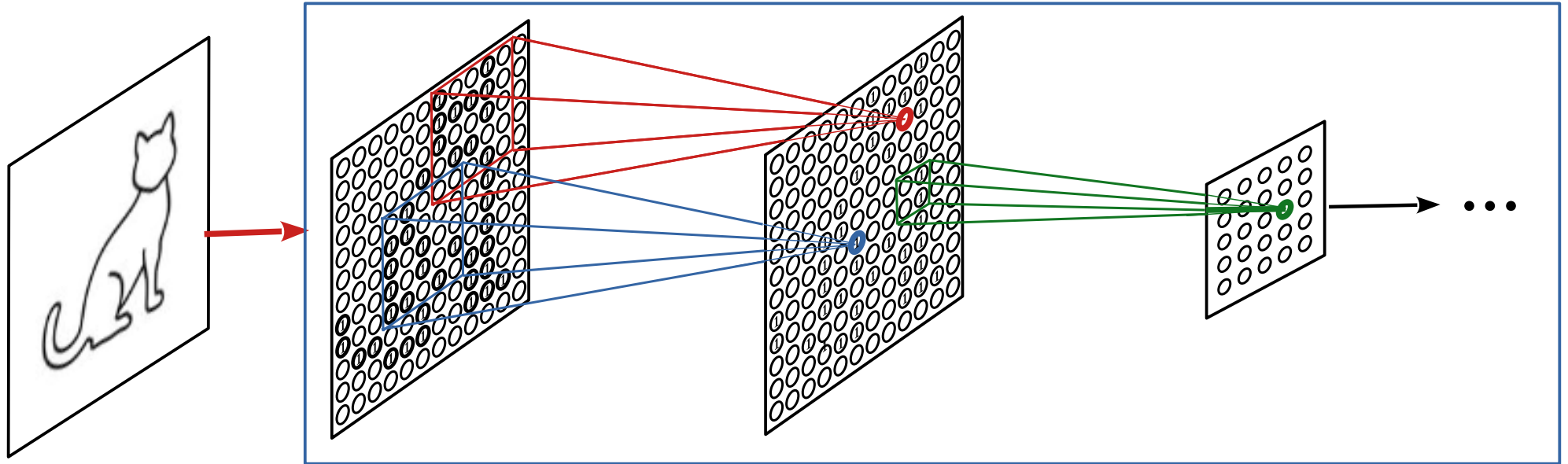
Les données massives (*Big Data*)

L'apprentissage de ces réseaux réclament des quantités phénoménales de données :

Exemple : 15 mille milliards de *tokens* pour Llama 3.1, le LLM de *Meta*

➡ *La qualité des données compte autant que leur quantité*

Les réseaux de neurones convolutifs



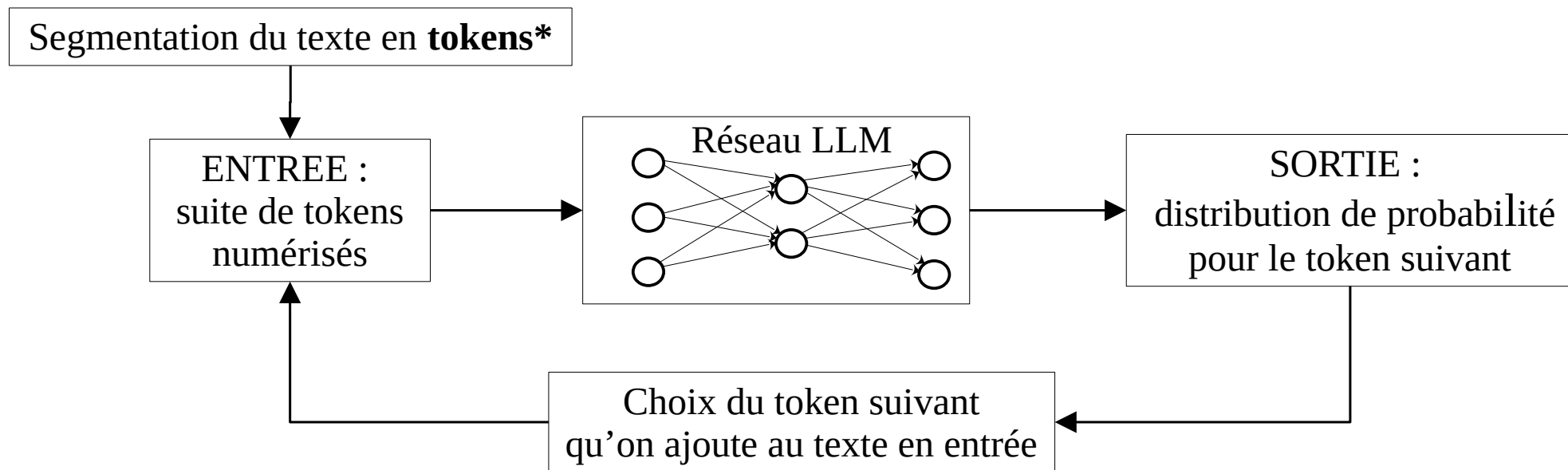
Dans un réseau convolutif, chaque neurone a un **champ récepteur** local



Ces réseaux imitent de près le système visuel des mammifères

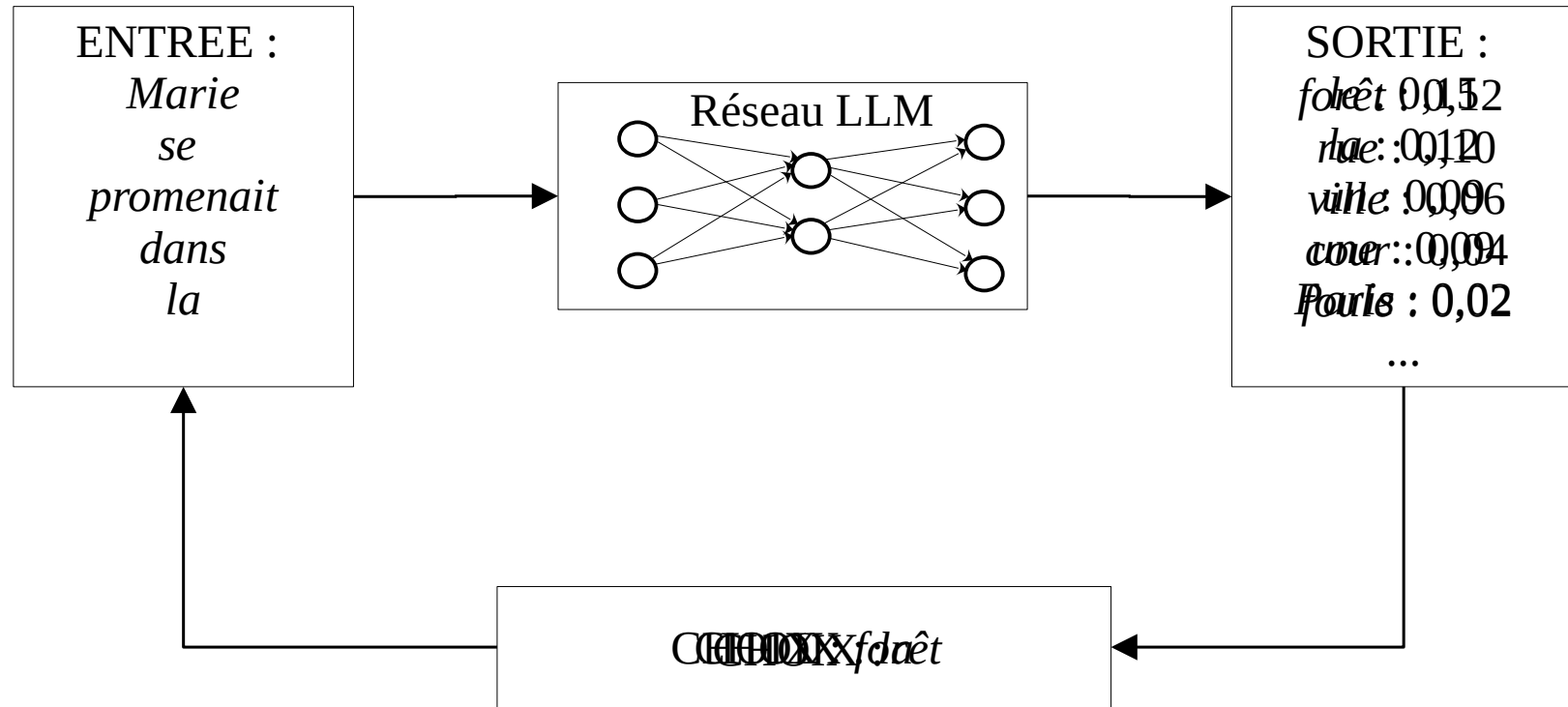
Les grands modèles de langage (LLM : *Large Language Models*)

Un LLM est un réseau de neurones qui prend **en entrée un tronçon de texte** et qui fournit **en sortie une suite à ce texte**



* *un token = un mot ou une portion de mot*

Exemple : *Marie se promenait dans...*



L'architecture des LLM (L'exemple de Llama 3.1 de *Meta*)

codage des
tokens



N couches
($N=126$)



décodage en
tokens

Chaque token est représenté par d neurones ($d = 16\,384$)
L'entrée comporte jusqu'à n tokens ($n = 128\,000$)

Chaque couche est composée de 2 parties :

- un sous-réseau **attentionnel** qui permet de prendre en compte l'influence de chaque token sur tous les autres
- un sous-réseau **unidirectionnel** qui agit sur chaque token indépendamment des autres

La sortie est une **distribution de probabilité** sur l'ensemble de tous les tokens du vocabulaire (de taille $v = 128\,000$)

L'apprentissage des LLM

Le corpus d'apprentissage

Des textes de toute sorte, dans des dizaines de langues, en accès libre (en principe...) : pages Web, livres, code informatique, etc.

Le processus d'apprentissage

On présente une portion de texte prise au hasard au réseau, et on modifie tous les poids du réseau pour augmenter la probabilité du token qui suit effectivement la portion présentée. Ce processus est réitéré des milliers de milliards de fois : des semaines de calculs intensifs !

Le résultat

Une fois l'apprentissage terminé, **le système n'a plus accès au corpus** :

Il a acquis une fois pour toutes une maîtrise presque totale du langage, aussi bien sur le plan de la **syntaxe** que de la **sémantique** et du **discursif**.



Les compétences linguistiques des LLM égalent voire dépassent celles de la plupart des humains

Les agents conversationnels (*chatbots*)

Les demandes des utilisateurs
à un agent conversationnel
peuvent être très variées

- résumer un texte
- le traduire dans une autre langue
- répondre à des questions factuelles
- dissenter sur un sujet donné
- écrire du code informatique
- démontrer un théorème
- rédiger une lettre
- écrire un poème
- ...

➡ *Il faut un apprentissage supplémentaire pour transformer un LLM en agent conversationnel*

Le peaufinage (*fine-tuning*) apprend au LLM à être **pertinent par rapport à la demande** :

- On entre une demande sous forme d'un texte : le **prompt**
- On évalue la pertinence des différentes suites proposées par le LLM
- On modifie les poids pour privilégier les suites pertinentes

LLM + fine-tuning = agent conversationnel

Des performances impressionnantes

Des compétences dans différents domaines de la connaissance :

histoire, médecine, droit, physique, informatique, mathématiques, ...

► **mesurées objectivement** par des succès à des épreuves académiques ou professionnelles : examen du barreau, concours de médecine, tests d'embauche d'ingénieurs informaticiens, ...

Une capacité à résoudre des problèmes de la vie quotidienne :

problèmes qui réclament des connaissances générales sur le monde, de la psychologie, du bon sens (*common sense*), et, dans une certaine mesure, de la planification.

➡ *Tout cela, sans faire appel à des ressources extérieures, en s'appuyant uniquement sur ce que le réseau a mémorisé en apprenant à deviner le token suivant !!*

Des défauts insurmontables ?

La péremption d'informations

Le corpus d'apprentissage étant daté, les textes générés par le système le sont aussi :

Aucun événement ou changement datant d'après la date de création du système ne peut être pris en compte dans les textes générés.

Les hallucinations

Les textes produits sont plausibles, mais pas forcément vrais !

Le système peut inventer des noms, des événements, des biographies, ...

Les réponses non éthiques

Le système reproduit tous les biais et défauts des textes du corpus d'apprentissage

Les réponses peuvent être racistes, sexistes, prôner la violence, ...



Tous ces défauts sont en partie corrigés grâce au *fine-tuning*, à des pré-prompts, à des filtres, etc. Mais malgré tous ces efforts, ça reste imparfait et contournable

Un domaine en pleine effervescence

La multimodalité

Au delà des textes, les nouveaux LLM permettent des interactions **sonores** et **visuelles** :

- ▶ avec GPT-4o (sorti au printemps 2024) on peut discuter en visio sur un smartphone : Le système "voit" ce qui l'entoure, peut parler, chanter... Et il tient compte de l'état émotionnel de l'interlocuteur (décelé par les mouvements et le ton de la voix) !

Utilisation de ressources

Les LLM peuvent maintenant interagir avec des ressources externes :

- Il peuvent naviguer sur le web, utiliser un tableur ou d'autres logiciels, etc.
- On peut aussi intégrer un LLM dans d'autres applications (tableur par exemple)

Des LLM spécialisés dans des domaines variés :

Recherche universitaire, éducation, finance, programmation informatique, écriture de courrier, analyse de données, création d'images, de vidéos, de musique...



Impossible de prédire aujourd'hui où s'arrêteront ces développements !

Un domaine en pleine effervescence

La multimodalité

Au delà des textes, les nouveaux LLM permettent des interactions **sonores** et **visuelles** :

- ▶ avec GPT-4o (sorti au printemps 2024) on peut discuter en visio sur un smartphone :
Le système "voit" ce qui l'entoure, peut parler, chanter... Et il tient compte de l'état émotionnel de l'interlocuteur (décelé par les mouvements et le ton de la voix) !

Utilisation de ressources

Les LLM peuvent maintenant interagir avec des ressources externes :

- Ils peuvent naviguer sur le web, utiliser un tableur ou d'autres logiciels, etc.
- On peut aussi intégrer un LLM dans d'autres applications (tableur par exemple)

Des LLM spécialisés dans des domaines variés :

Recherche universitaire, éducation, finance, programmation informatique, écriture de courrier, analyse de données, création d'images, de vidéos, de musique...

 Impossible de prédire aujourd'hui où s'arrêteront ces développements !

Vers une meilleure compréhension de nos facultés cognitives ?

Un nouvel éclairage sur la cognition

En tant que modèle (très simplifié!) du fonctionnement du cerveau, ces systèmes peuvent conforter ou infirmer des théories sur différentes facultés cognitives animales ou spécifiquement humaines

- Exemple : la théorie de l'innéité d'un module syntaxique défendue par Chomsky

Ouvrir la boîte noire

Un certain nombre de recherches visent à comprendre les performances de ces réseaux en analysant leur fonctionnement de l'intérieur :

- Peut-on mettre en évidence des **représentations internes** pertinentes des entrées ?
- Si oui, peut-on caractériser les conditions d'apparition de ces représentations ?

➡ Là encore, impossible de prédire si ces pistes de recherche seront très fécondes...