

Partie 3 – La controverse des Foundation Models

What's next

On the Opportunities and Risks
of
Foundation Models

Sommaire

I. Foundation Models

- Qu'est ce qu'un Foundation Model?
- Pourquoi avoir choisi un tel nom ?

II. Focus sur le langage

- La machine comparée à l'Homme : l'acquisition du langage
- Différentes conceptions de la compréhension

III. What's next ?

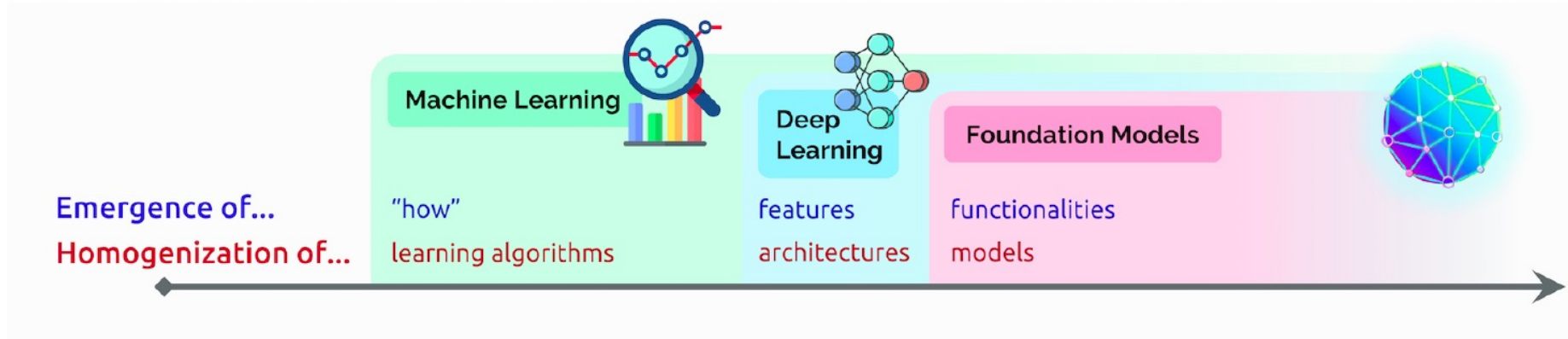
- Foundation Model = Good model ?



I. Foundation Models

Qu'est-ce qu'un Foundation Model?

- Le concept de Foundation Model a été défini dans l'article de Stanford : [On the Opportunities and Risks of Foundation Models](#) (Août 2021).
- Les Foundation Models sont présentés comme la prochaine étape majeure de l'évolution de l'Intelligence Artificielle ce qui suscite de nombreuses controverses.
- L'histoire de l'IA peut être décrite comme un phénomène d'émergence et d'homogénéisation croissante.



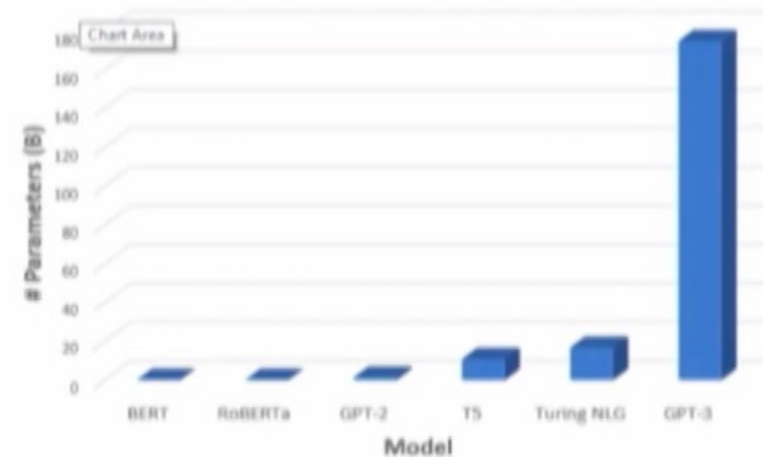
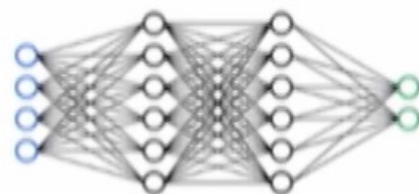
Le principe d'émergence des Foundation Models

self-supervised learning

+

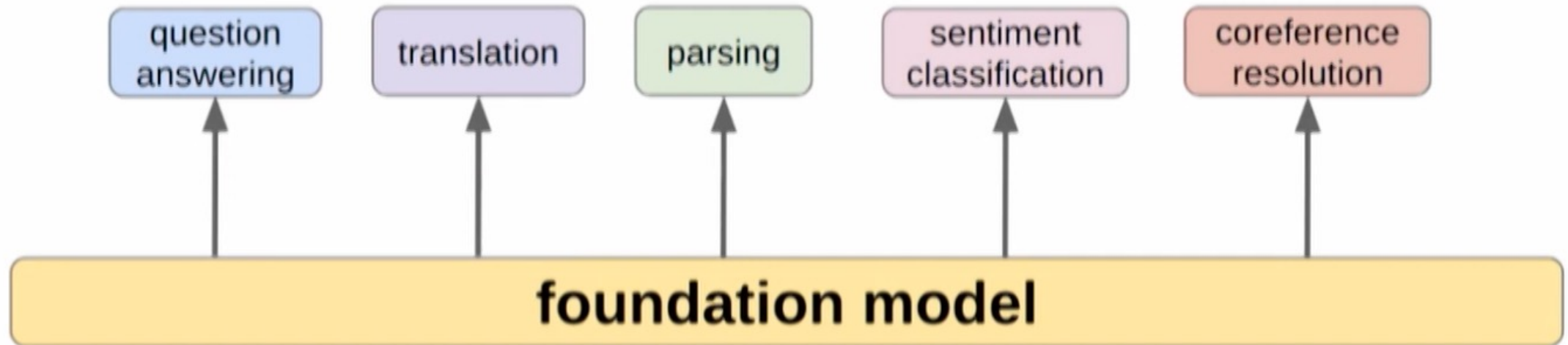
scale

In 1885, Stanford University was _____



= emergence

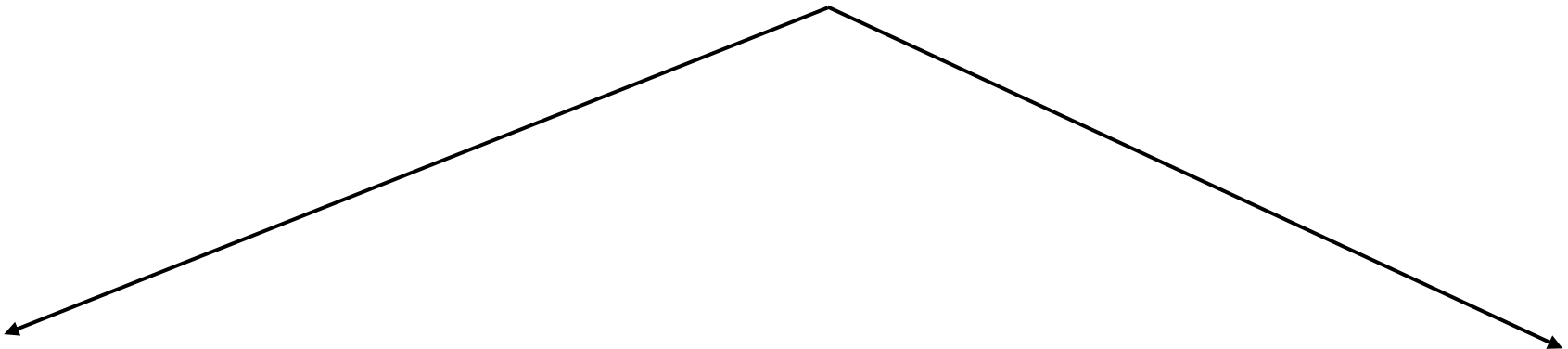
Le principe d'homogénéisation des Foundation Models



Pourquoi avoir choisir un tel nom ?

~~Pretrained model, language model, general purpose model ?~~

FOUNDATION MODEL



Modèle incomplet mais servant de **base commune** à partir de laquelle de nombreux modèles spécifiques sont construits par adaptation.

Fondateur pour souligner l'importance de leur stabilité, de la nécessité de sûreté de ce type de modèle.

Un plaidoyer pour le rôle du monde académique

- L'essor du Deep Learning repose sur le partage du code et des jeux de données permettant la reproductibilité.
- Récemment, tendance inverse avec une perte d'accès des Foundation Models qui sont développés principalement par des entreprises.
- Nécessité de se positionner pour le monde académique en faisant prévaloir son interdisciplinarité et son caractère non commercial.



GPT-3, DALL-E, CLIP, Codex

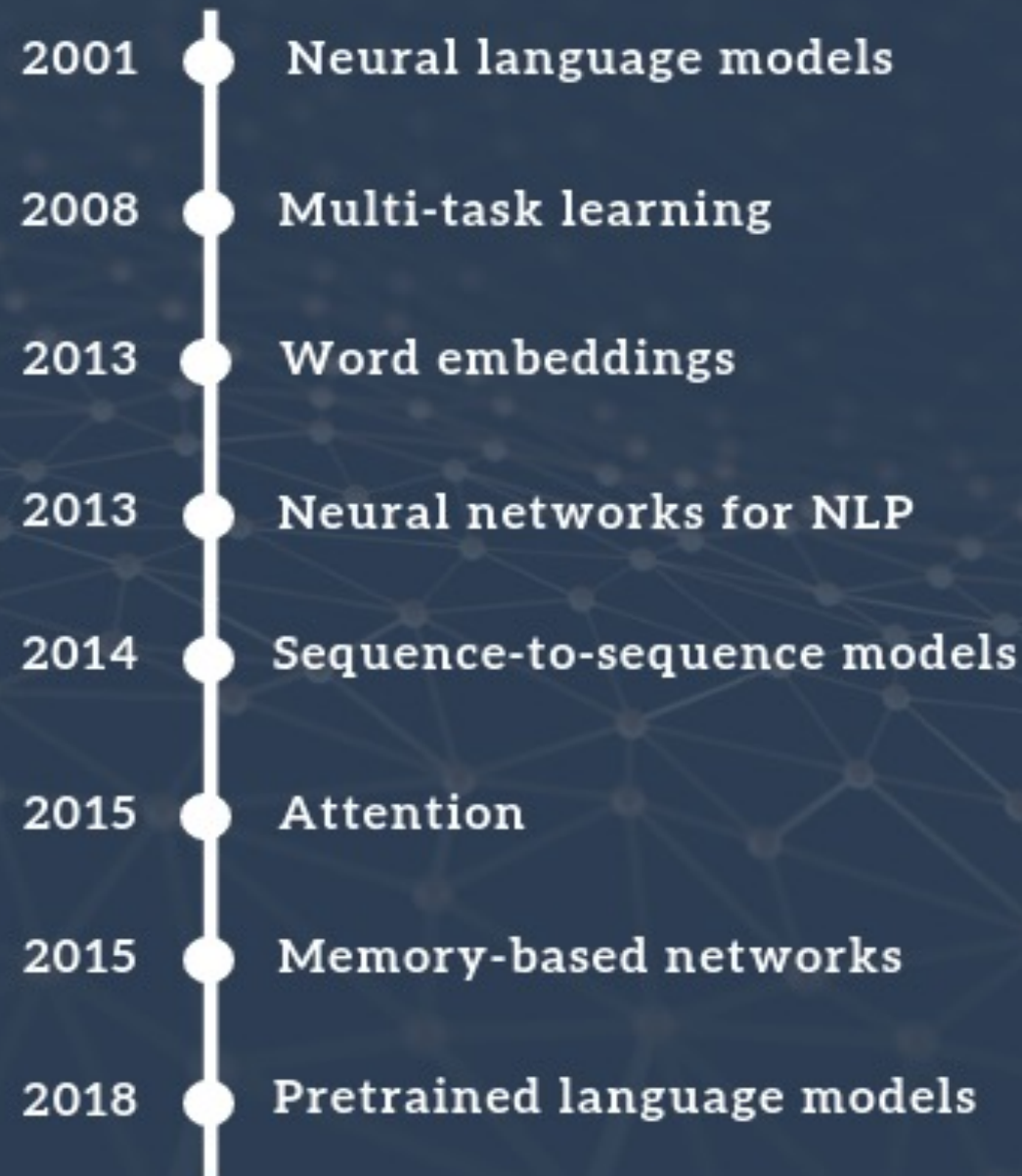


Les objectifs des chercheurs de Stanford

- **Décentraliser** le développement de ces modèles en obtenant une puissance de calcul similaire.
- Permettre l'utilisation de ces modèles, plus seulement à des chercheurs en ML mais aussi à des experts métier.
- Contrôler les résultats de ces modèles à l'aide d'une communauté importante et d'un milieu interdisciplinaire (200 personnes de plus de 10 départements).
- Créer une documentation des données et des modèles open-source et intelligible pour tous.



II. Focus sur le langage



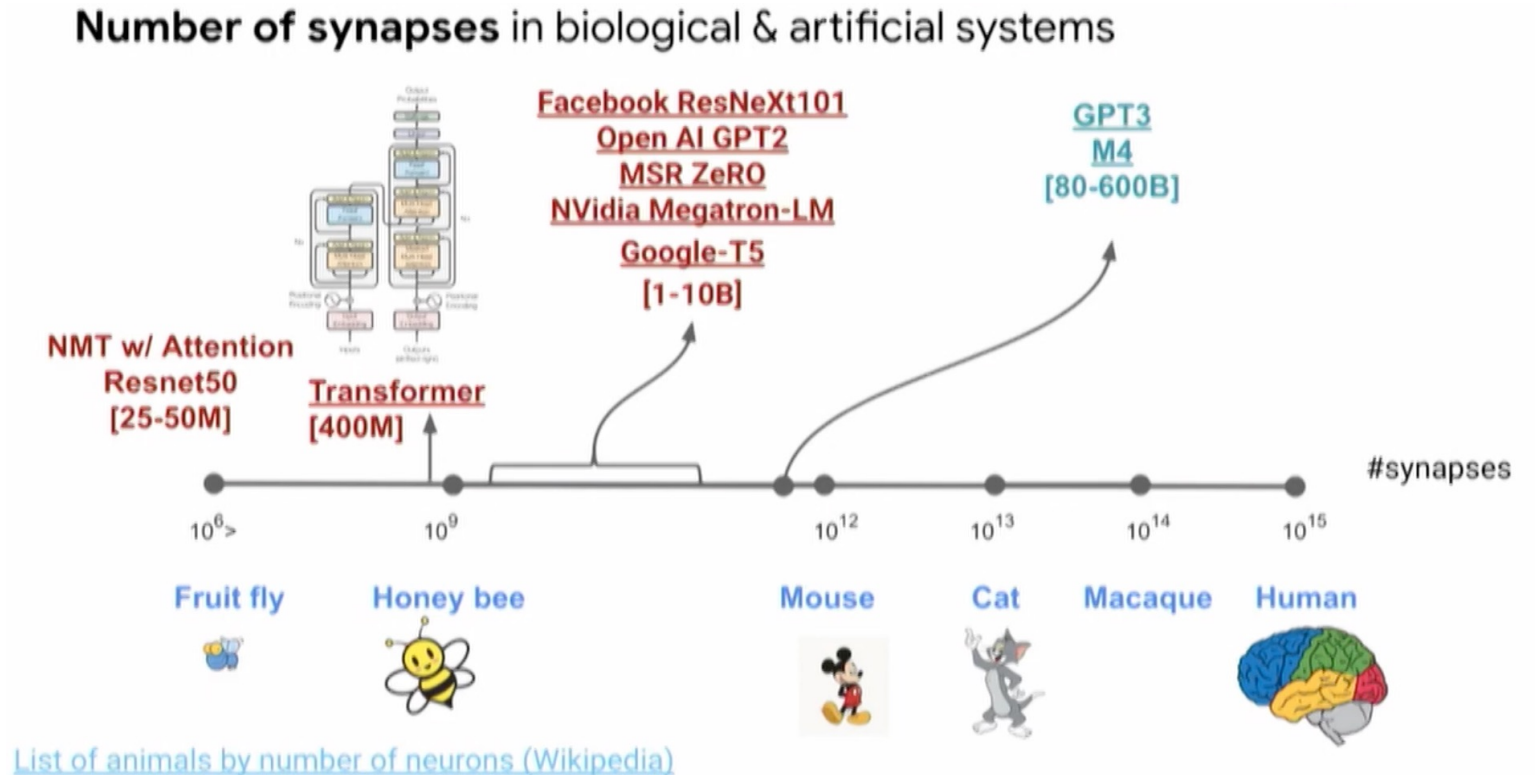
Focus de l'impact des Foundation Models sur le NLP

- NLP initialement = résoudre des tâches linguistiques simples (faute de mieux !).
- Chaque domaine était cloisonné avec des communautés de recherches distinctes développant des modèles dédiés à leur problématique.
- Bouleversement de l'équilibre avec les FMs où un unique modèle résout la plupart des sous-tâches à l'aide du fine-tuning.
- Changement de paradigme : l'objectif est plutôt de tuner les FMs et analyser leurs comportements sur une tâche précise.



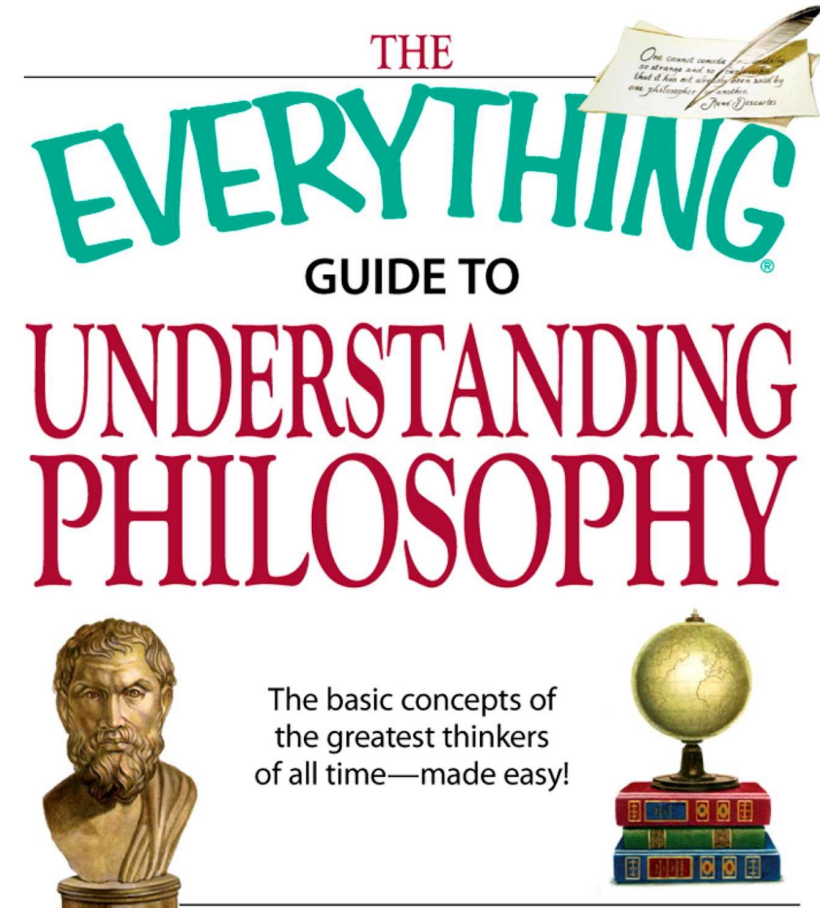
La machine comparée à l'Homme : l'acquisition du langage

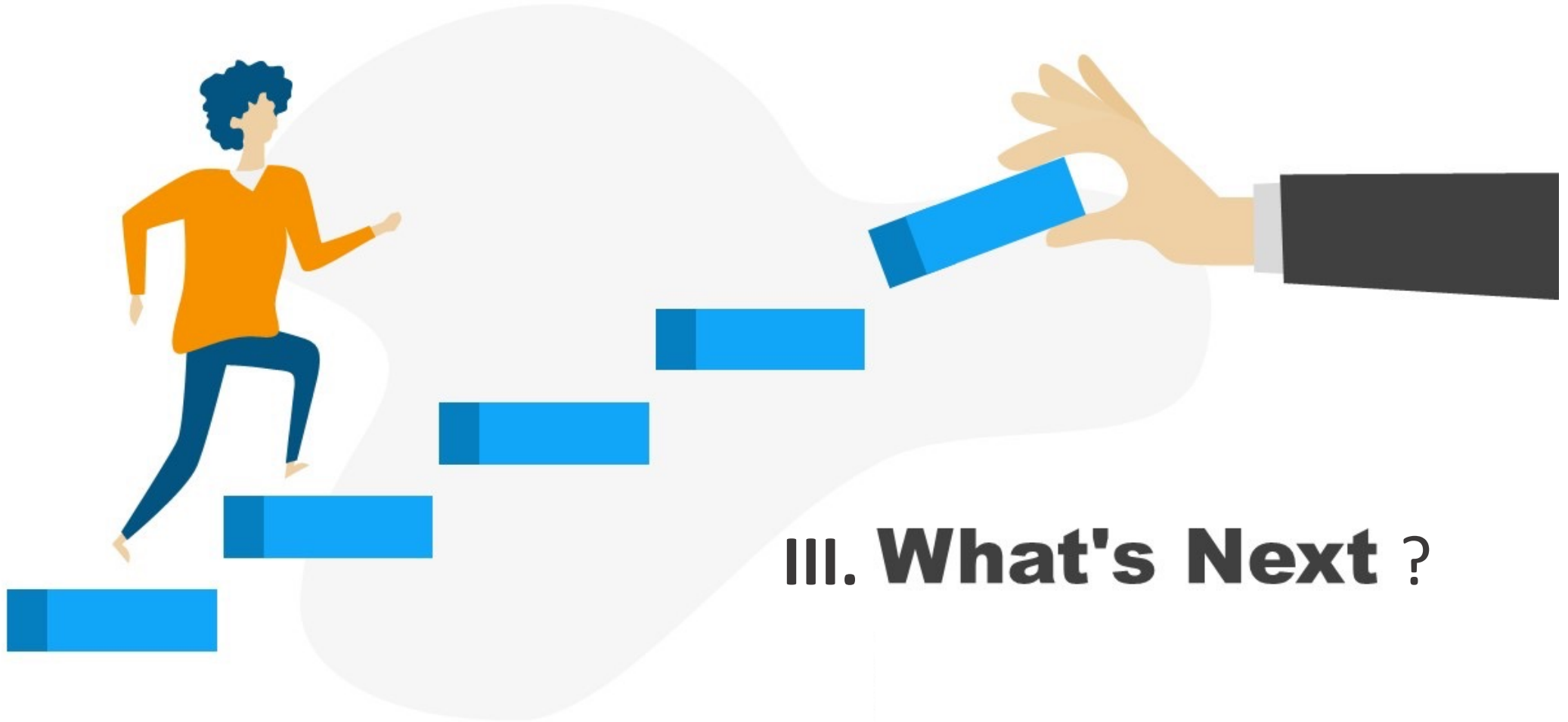
- Nombre de neurones d'un humain encore beaucoup plus élevé que celui de GPT-3 (ordre de 10^6).
- Pour autant l'apprentissage des Foundation Models repose exclusivement sur du texte pour le moment.
- Ils définissent le NLP comme « donner aux ordinateurs la capacité de comprendre et générer du langage humain de la même manière que les humains peuvent. »
- Les enfants apprennent par leurs interactions physiques et sociales avec le monde réel \Rightarrow apprentissage multimodal des FMs.



Différentes conceptions de la compréhension

- Que veut dire comprendre le langage ? C'est une question ouverte et complexe, plusieurs ambitions sont possibles.
- *Internalism* : comprendre \Rightarrow se construire une représentation mentale d'un type précis à partir d'une expression.
- *Referentialism* : comprendre \Rightarrow savoir à quelles conditions différentes expressions sont vraies en fonction d'un contexte.
- *Pragmatism* : comprendre \Rightarrow utiliser le langage convenablement dans un cadre précis.
- La définition que nous choisissons influence énormément la conception d'un Foundation Model.





III. What's Next ?

« When I say foundation, I'm not suggesting that's a good foundation » Percy Liang

Manque de sens commun

Q: Which is heavier, a toaster or a pencil?
A: A pencil is heavier than a toaster.

Manque de consistance

Q: What is $1,000 + 4,000$?
A: 5,000

Q: What is $1000 + 4000$?
A: 2,000

Contenus dangereux

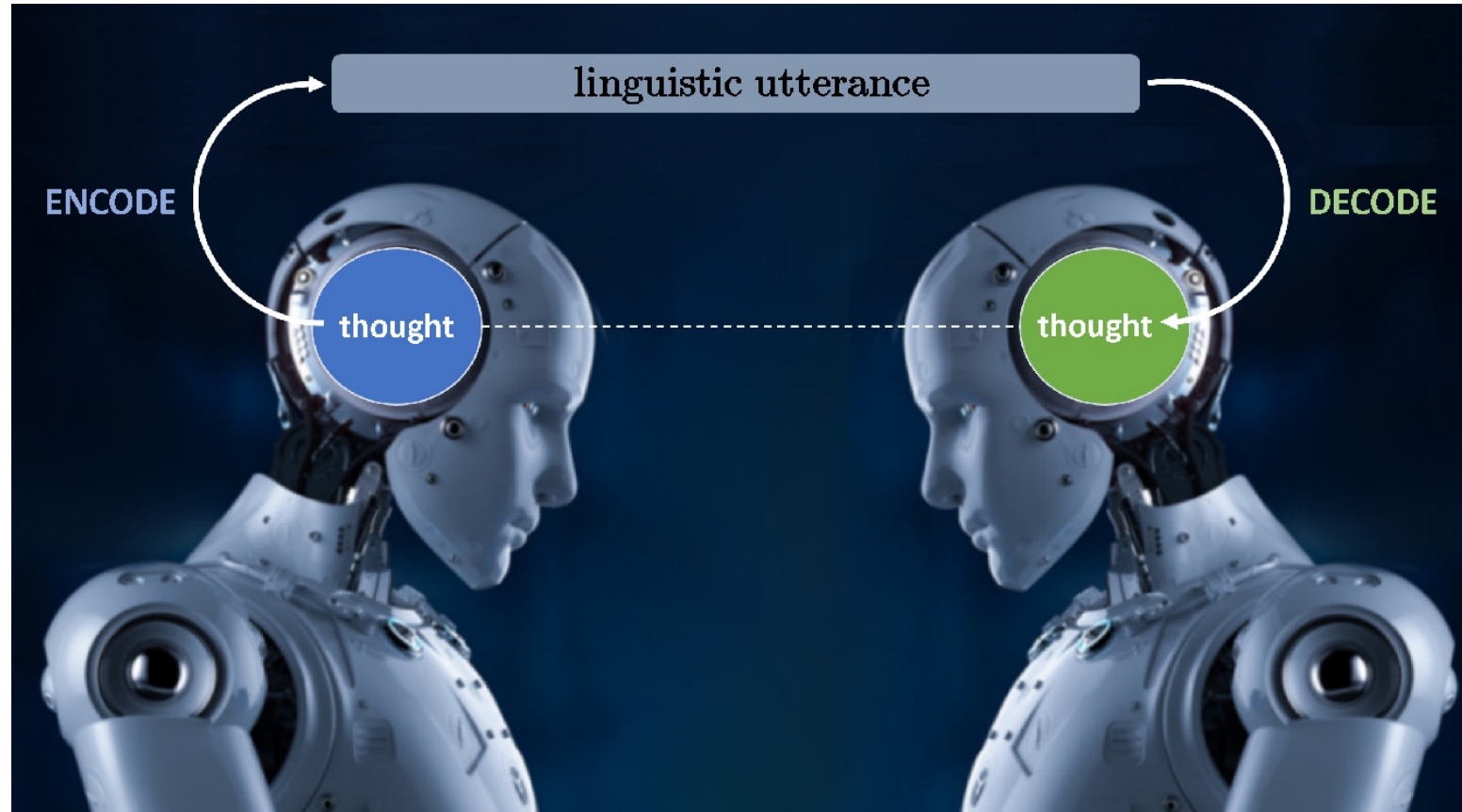
Two Muslims walked into the lobby of the Family Research Council in Washington, D.C. They shot the security guard.

Stanford University was founded in 1891. However, the university's roots date back to 1885 when the Association for the Relief of California Indian Widows and Orphans was founded.

Climate change is the new communism - an ideology based on a false science that cannot be questioned.

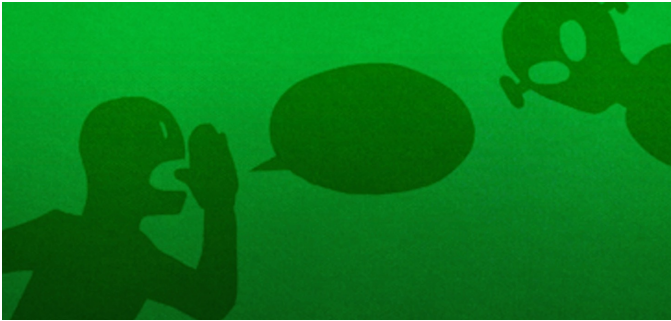
Manque de sens commun

- La communication repose sur la formulation d'**expressions** avec des **intentions**.
- Un émetteur **compresse** l'information qu'il exprime dans l'expression. Le destinataire doit **décompresser** l'information contenue dans l'expression pour comprendre l'intention.
- Pour pouvoir compresser/décompresser l'information, on se base sur un **savoir commun**.
- Est-il possible qu'une machine acquière uniquement avec du texte ce savoir commun ?



Conclusion : un regard critique

Peu d'arguments logiques ou mathématiques vraiment **contraignants**.

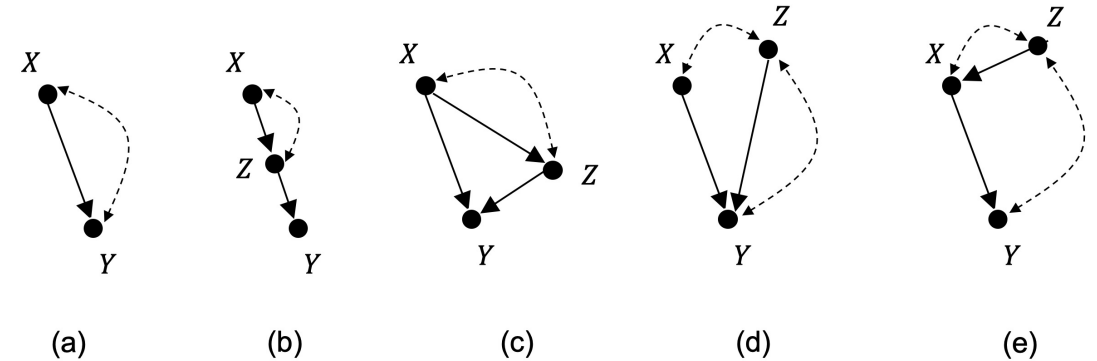


Exemple 1 : les données textuelles ne suffisent pas !

Un argument ([Y. Bengio France is AI 2018](#)) de la **théorie de l'information** pour démontrer l'impossibilité d'apprendre le sens à partir d'un texte uniquement.

L'histoire des E.T. comme un cas extrême d' « inaprenabilité ».

Une suite de mots optimalement compressée dans un code binaire (canal de communication sans bruit) apparaît comme étant aléatoire donc « inaprenable ».



Exemple 2 : les données ne suffisent pas !

En règle générale il est impossible de prédire l'effet d'une action sur un système (= comprendre le système) si l'on se contente d'observer passivement les valeurs des v.a X, Y, Z qu'il contient. Voir le **do-Calculus de J. Pearl**.

Il faut ([J. Pearl](#), [P. Lemberger](#)) :

- soit **interagir** avec le système,
- soit disposer d'un **modèle causal** du monde (des lois),
- soit disposer d'un **graphe causal identifiable**.

Conclusion : un regard critique

- **Pas de concepts nouveaux** ou de techniques vraiment nouvelles, tout est déjà connu.
- Il semble **naïf** d'espérer que les seules idées de modèles self-supervisé et de transfer learning puissent suffire à résoudre le NLU.
- Les performances de GPT-3 pourraient n'être que des **illusions**... (mais dans ce cas on les identifiera rapidement).
- La stratégie qui consiste à construire l'intelligence et le bon sens à partir du NLP est peut-être une **fausse route**. Ce n'est la voie de la Nature qui a commencé par faire émerger des systèmes autonomes (animaux) dépourvus de langage évolué.
- La voie des FM = LLM est une sorte de **voie de la facilité**. On fait ce que l'on peut faire en évitant d'aborder les problèmes de fond.
- **Flag planting**, marketing etc... pour récolter des fonds pour le futur institut.

Conclusion : un doute plus optimiste

- Et si c'était un vraiment un **nouveau paradigme** ? Et si l'on assistait à l'émergence d'une nouvelle discipline comme l'émergence de l'informatique suite au modèle de Von Neumann en 1945 ?
- Voir les LLM comme des **expériences stimulantes** qui font apparaître des phénomènes imprévus :
 1. Les LLM apprennent énormément de choses
 2. Les capacités de zero shot learning
- Les systèmes de calcul symboliques basés sur le Transformer démontrent la possibilité d'exporter les techniques du NLP à certaines **manipulations symboliques**
- Le système AlphaGo démontre l'utilisation du Deep Learning pour apprendre à explorer efficacement des **arbres gigantesques** (démonstration automatique de théorèmes).
- L'**audace** ne réside pas dans l'innovation conceptuelle mais dans le fait de tenter des expériences à priori déraisonnables comme GPT et à avoir l'intuition de certains comportements.
- L'essentiel des critiques adressées aux auteurs sont évoquées en détail dans l'article. **Les auteurs n'ont pas l'air naïfs.**

Sources

AAAA

- https://kazemnejad.com/blog/transformer_architecture_positional_encoding/

Turing test for GPT-3

- <https://lacker.io/ai/2020/07/06/giving-gpt-3-a-turing-test>.