

C'est un sujet sur lequel nous pensons tous que nous devrions avoir des connaissances. Le cerveau. Et pourtant, il n'y a pas vraiment de recette connue de tous pour expliquer le fonctionnement du cerveau. Il n'y en a pas ?

Mais les neuroscientifiques le savent, n'est-ce pas ? Eh bien...

Et maintenant, nous avons des IA comme ChatGPT qui, au moins en apparence, font beaucoup de raisonnements semblables à ceux des humains.

Qu'est-ce que cela signifie ?

Le cerveau ressemble-t-il à ChatGPT ?

Votre cerveau est comme un .

Autrefois, nous disions toujours que le cerveau était comme un ordinateur.

Mais la plupart d'entre nous, scientifiques et technologues, avons cessé de dire cela il y a plus de 30 ans. Le cerveau n'effectue pas des milliards d'opérations par seconde en suivant un programme.

Pour commencer, il est plutôt lent, nous pensons à environ 2 ou 3 mini-pensées par seconde, bien que nous soyons assez rapides pour percevoir et interpréter notre vision, disons environ 5 ou 10 fois par seconde.

Bien sûr, il est difficile de quantifier la façon dont cela se convertit en instructions informatiques équivalentes.

Deuxièmement, au lieu d'un programme, nos réseaux neuronaux biologiques s'inscrivent dans un paradigme dit "connexionniste", où le traitement est effectué en parallèle par des milliards de neurones qui détectent des modèles, chacun avec des milliers de synapses finement ajustées.

Au total, ce qui différencie votre cerveau de celui d'une autre personne, ce sont les quelque 600 000 milliards de synapses ajustables qui relient vos neurones les uns aux autres.

Le cerveau ressemble donc plus à un logiciel de réseau neuronal artificiel, qui n'est généralement considéré que comme légèrement similaire aux cerveaux biologiques, qu'à un ordinateur proprement dit.

Savons-nous comment fonctionne le cerveau ?

Cela dépend de ce que vous entendez par "savoir".

Et de la personne à qui l'on s'adresse.

Dans le passé, pendant des décennies, nous avons disposé d'un grand nombre de données, mais nous n'avons pas beaucoup progressé dans notre vision d'ensemble :

1. Nous étions capables de simuler avec précision de petits réseaux de neurones.
2. Nous ne disposions que de théories manuelles (c'est-à-dire non mathématiques) au niveau de l'image globale.

Aujourd'hui, lorsque nous disons que nous "savons" comment fonctionne le cerveau, nous voulons dire que

A. Nous pouvons simuler avec précision des parties du cerveau et obtenir les bonnes réponses.

B. Nous avons une compréhension plus pratique des processus de haut niveau tels que la vision, la pensée et l'indexation des pensées dans l'hippocampe, et nous disposons de nouvelles théories générales solides qui sont nettement meilleures qu'auparavant.

C. Nous disposons même de théories mécanistes détaillées et testables sur le fonctionnement des principaux composants du cerveau.

Ainsi, des théoriciens des neurosciences/de l'IA comme Jeff Hawkins, ex-PDG de Palm Computing, et son équipe Numenta, par exemple, ont des théories concrètes et vaguement testables. (Lisez mon commentaire du livre de vulgarisation scientifique de Hawkins sur le sujet ici).

D'un point de vue pragmatique, les succès des simulations sont sans doute aussi bons qu'une théorie ou sa compréhension, même si, bien sûr, l'extraction de la compréhension à partir de simulations est toujours un art scientifique fascinant.

Le Human Brain Project tente actuellement de simuler une bonne partie du cerveau. Et nous avons réussi à simuler le cerveau de 132 neurones d'un ver microscopique, le *C. elegans*.

Cela dit, comment le cerveau fonctionne-t-il d'après ce que nous savons ?

Comment tous ces neurones produisent-ils des pensées ?

Les neurones du cerveau ne sont que des détecteurs de formes.

Tout ce que nous faisons d'autre avec le cerveau, qu'il s'agisse d'imaginer des journées à la plage, d'écrire logiquement des blogs ou d'inventer de nouvelles technologies, est le résultat de combinaisons de ces allumages neuronaux.

Les milliards de neurones interconnectés de notre cerveau peuvent être considérés comme un dispositif de compression de la mémoire. Il s'agit également d'un dispositif permettant de traiter les souvenirs en utilisant essentiellement le même système.

Il est évident que les "réglages de force" des billions de synapses qui relient les 85 milliards de neurones de notre cerveau peuvent être considérés très naïvement comme des billions d'emplacements de mémoire.

L'erreur de ce modèle est que chaque mémoire est stockée de manière distribuée. Je ne veux pas dire qu'il est stocké de manière redondante, mais que les différentes parties d'un souvenir sont stockées à des endroits différents.

Au lieu de stocker le souvenir spécifique, le cerveau le compresse pour le stocker un peu partout avec tous les autres souvenirs similaires. Ainsi, en dépit de la mémoire à court terme, notre cerveau modifie les paramètres des synapses partout où ce souvenir aura un impact.

Par exemple, si vous apprenez un nouveau caractère de la langue hébraïque, comme le P hébreu : פ, votre cerveau va quelque part légèrement modifier les éléments généraux sous-jacents, mais plus grands que le pixel, y compris les coins et les courbes qui composent le glyphe, et quelque part d'autre dédier certains neurones à la mémorisation de ce caractère via des liens avec les éléments et d'une certaine manière encoder (via d'autres neurones) la façon dont ils se connectent pour former le glyphe.

Il a été démontré que le cortex visuel contient trois niveaux de traitement : V1, V2 et V4 (il n'y a pas de V3 pour une raison historique). Les neurones dédiés à la reconnaissance des parties d'images telles que les lignes et les coins, ainsi que leur orientation et leur position, se trouvent dans V1 et V2.

La sortie de ces neurones remonte la hiérarchie et, dans V4, nous avons des neurones qui se déclenchent pour des images spécifiques comme פ, quelle que soit son orientation ou sa taille.

Le cerveau ne mémorise pas chaque fois que nous voyons un פ, il continue à modifier les neurones qui alimentent le neurone principal jusqu'à ce que la police de caractères n'ait plus d'importance.

Il s'agit d'un système de compression.

C'est incroyable.

Bien sûr, il y a beaucoup plus que cela.

Le connectome

Si le cortex visuel de notre cerveau traite la vision, c'est tout simplement parce que les neurones de notre nerf optique y arrivent. Si V2 et V4 traitent les niveaux de vision suivants, c'est parce que notre "connectome", une série de connexions neuronales à longue distance et à un stade précoce du développement, achemine les signaux de ces A vers les B.

Pour le reste, le néocortex - comme Jeff Hawkins aime à le dire dans son livre - est le même partout. Les parties du cerveau qui traitent les sons ou le langage sont dictées par ce qui entre dans ces parties par les faisceaux de connectomes précocement formés.

C'est Vernon Mountcastle (John Hopkins) qui l'a noté pour la première fois dans une obscure revue technique. Depuis 2004, Jeff Hawkins l'a popularisé, en essayant, ce qui est compréhensible, de le faire savoir au monde entier.

Tous les autres s'enthousiasment pour les minuscules différences qu'ils découvrent dans les différentes parties du néocortex. Mais Mountcastle et Hawkins sont plus impressionnés par l'uniformité !

Cela signifie que le cerveau fonctionne essentiellement de la même manière presque partout dans le cerveau. Du moins dans le centre primaire de la mémoire et de la pensée qu'est le néocortex.

Le cerveau exécute essentiellement le même algorithme presque partout !

La seule différence réside dans l'origine et la destination des signaux.

Pour être juste, bien sûr, il existe d'autres parties spécialisées et importantes du cerveau et une architecture globale impliquant, dans sa plus simple expression, le néocortex et l'hippocampe.

La mémoire et la pensée se produisent comme décrit ci-dessus dans le néocortex.

L'hippocampe peut être considéré comme un index temporel et spatial de ces souvenirs. Vous pouvez imaginer pourquoi nous avons besoin de ce petit ajout et pourquoi nous ne voulons pas qu'il soit endommagé.

Qu'avons-nous appris des réseaux neuronaux artificiels (RNA) ?

Par réseaux neuronaux artificiels, nous entendons généralement les réseaux artificiels (rapides, non imitatifs de la biophysique) que nous construisons pour l'apprentissage profond de l'IA (c'est-à-dire pas les simulations biophysiques minutieuses conçues pour correspondre à la biologie, mentionnées ci-dessus).

Le fait est que même les petits réseaux neuronaux (sans parler des réseaux neuronaux tels que ChatGPT ou DALL-E 2 d'OpenAI ou LaMDA de Google) présentent des caractéristiques très proches de la biologie, au-delà de la nature distribuée du stockage de la mémoire :

- l'utilisation naturelle des couches du réseau neuronal émerge
- des "neurones grand-mère" apparaissent (c'est-à-dire des neurones maîtres qui se déclenchent en cas de schémas d'entrée de haut niveau importants et récurrents)
- les couches supérieures changent rapidement de schéma d'allumage lorsque les données d'entrée fluctuent
- les couches profondes s'installent sur des schémas d'allumage plus constants à mesure que les données d'entrée fluctuent.

Ce phénomène est totalement analogue à ce qui a été découvert dans le cortex visuel du cerveau biologique. Par exemple, V1 fluctue à chaque changement de pixel, tandis que V4 se fixe sur l'identification d'un objet.

Plus intéressant encore, les ANN visuels des niveaux supérieurs identifient des composants visuels de bas niveau très similaires, notamment les bords, les lignes, les coins, les points et les formes.

Cela ne devrait surprendre personne. Dans le domaine de l'efficacité, même les RNA simples sont très analogues aux réseaux neuronaux biologiques réels à un niveau élevé.

Il va de soi que les caractéristiques susmentionnées allaient toujours émerger de manière analogue. Pourtant, pour une raison étrange, un groupe de sceptiques de l'IA symbolique LMM insiste sur le fait que les ANN et les LLM n'ont "rien à voir avec le cerveau".

C'est faux. À un niveau grossier, il existe une forte ressemblance dans l'architecture fondamentale de niveau moyen : des couches de neurones plus connectées entre elles et avec les couches voisines qu'avec les autres couches.

Mathématiquement, le comportement est dicté par cette topologie. Il est impossible que les ANN ne se comportent pas comme des cerveaux.

Les détails ne sont pas pertinents.

Le ChatGPT est-il comme le cerveau ?

C'est une question controversée parmi les chercheurs et les développeurs de NLU (compréhension du langage naturel).

Selon mon opinion professionnelle, les réseaux neuronaux artificiels (RNA) de l'apprentissage profond et de ChatGPT sont plus proches de notre cerveau qu'on ne le pense. Les désaccords (professionnels) sur ce point se résument purement à des définitions différentes du terme "similaire", selon l'IMO. Je veux dire "approximativement", "dans un sens global" du phénomène de base qui se produit dans les ANNs par rapport aux cerveaux, dans la façon dont l'information est stockée, comprimée et extraite. Pas dans les moindres détails de leur architecture, bien sûr.

Il existe une division académique sur cette question et sur le fait de savoir si ChatGPT comprend et raisonne, et encore moins s'il réussit.

En effet, ChatGPT, malgré sa capacité à répondre à nos questions et à nos instructions, n'a jamais été conçu que pour "prédire le mot suivant" et son étonnante capacité à générer du texte a été à la fois un choc et un accueil favorable pour la majeure partie de la communauté de l'IA.

Les plus pragmatiques d'entre nous, chercheurs en NLU, dont je fais partie, s'accordent à dire que ChatGPT semble indéniablement comprendre vos données d'entrée et d'entraînement.

Cette capacité, à notre avis, apparaît ou "émerge" dans un système purement basé sur la prédiction

du mot suivant dans une séquence, simplement parce que pour bien prédire le mot suivant, il doit bien comprendre le texte précédent et donc ses données d'apprentissage.

Certains "pointilleux" universitaires de l'UAL sont, d'après l'OMI, tout à fait déraisonnables parce qu'ils se concentrent sur (1) les imperfections des MLI et (2) le mécanisme de réseau neuronal sous-jacent des MLI.

Ces sceptiques sont généralement des chercheurs en IA symbolique qui, par définition, choisissent ou préfèrent travailler sur ce paradigme basé sur des règles plutôt que sur des réseaux neuronaux. Ce débat suscite donc des intérêts divergents, menés de manière assez militante par Gary Marcus (NYU) et le célèbre linguiste Noam Chomsky (MIT).

Personnellement, je suis un fan des quatre domaines de l'IA symbolique, des réseaux neuronaux, des neurosciences et de la linguistique.

Je suis en fait assez stupéfait du dogmatisme, du manque d'appréciation des contributions de chaque domaine par l'autre, et de la quasi-animosité exprimée sur ce sujet.

Comment le ChatGPT se compare-t-il au cerveau ?

Veuillez accepter ces comparaisons avec quelques grains de sel ! J'essaie de mesurer le contenu de l'information. Il s'agit simplement d'établir une comparaison approximative.

ChatGPT, ou plus précisément GPT-3 qui le sous-tend actuellement fin 2022, compte environ 175 milliards de "synapses" réglables (généralement appelées "poids" ou paramètres du réseau neuronal), ce qui est environ 3500 fois moins que les 600 billions de synapses d'un cerveau.

Ce nombre de neurones inclut d'autres composants du cerveau que ceux liés au langage, comme le cortex visuel et le cervelet, qui se concentre sur le mouvement.

La majeure partie de la "réflexion" sur les concepts et le langage se produit manifestement dans la partie non sensorielle du néocortex et dans l'hippocampe, qui représentent ensemble environ 25 % du cerveau.

En termes de pensée, le cerveau fait donc fonctionner quelque chose comme 1000 ChatGPT.

Veuillez prendre ces calculs à l'envers avec quelques grains de sel.

Ainsi, chaque fois que nous (surtout pas moi) banalisons le ChatGPT, rappelons-nous que même si nous sommes fondamentalement plus intelligents, le ChatGPT connaît en réalité beaucoup plus de faits (et de relations entre eux) sur le monde en général que n'importe lequel d'entre nous, et ce avec 1000x moins de paramètres réglables.

La différence entre ChatGPT et une gigantesque base de données de connaissances sur le monde et de bon sens (qui n'existe pas, Wikipedia représente peut-être 0,01% de la formation de GPT) est bien sûr que ChatGPT comprend comment ces relations fonctionnent et comment générer du texte à partir de celles-ci dans le style que nous demandons. (Un peu mieux qu'une base de données ! Mais l'ANN de ChatGPT ne stocke nulle part avec précision ses données d'apprentissage. Il ne stocke que les relations statistiques comprimées entre les mots, et n'est donc pas aussi bon qu'une base de données).

Et si l'un d'entre nous passe du temps à banaliser son cerveau, ou du moins le cerveau des gens en général (est-ce que quelqu'un fait cela ?), il est tout aussi valable d'affirmer que chacun de nos cerveaux, du point de vue de la pensée, vaut environ 10 milliards de dollars (on estime que GPT-3 coûte environ 10 millions de dollars pour s'entraîner).

Au total, la valeur totale de nos cerveaux individuels est d'environ 50 milliards de dollars. LOL.

Bien sûr, la comparaison n'est pas vraiment juste : ChatGPT, pour un coût pas trop élevé par personne, peut distribuer sa "cognition en bouteille" à des millions de personnes à la fois, comme c'est déjà le cas.