Epistémologie de l'informatique

Réflexion sur les agents conversationnels LaMDA AI

Etudiant : Didier Kouamé

M1 Langue & Informatique

Professeure : Laurence Devillers

2022 – 2023

LaMDA (Language Models for Dialog Applications)

[1]LaMDA signifie “modèles de langage pour les applications de dialogue’’. En l’occurrence LaMDA est un agent conversationnel atypique parce qu’il utilise une technique d’IA particulière et inédite. La 1ère génération a été annoncée à la keynote de Google en 2021 et une deuxième en juin 2022. Il y a beaucoup de discussions autour de ce chatbot car un ingénieur de chez Google Blake Lemoine a affirmé que l’IA était devenue consciente (sentient). A la différence des agents conversationnels que tout le monde connaît comme Alexa ou Siri, elle est ‘’capable de comprendre’’ le langage en profondeur, de déduire le sens lors d’une conversation avec l’homme, de générer un langage proche de celui de l’homme et de suivre **le flux d’une conversation**. C’est là la différence majeure entre cette IA et les autres. Les chatbots classiques ne peuvent pas suivre continuellement le flot de paroles de l’homme lors d’une discussion parce que tout leur fonctionnement est déjà prédéfini dans leur système. Pour obtenir des réponses des agents conversationnels classiques, il faut leur poser essentiellement des questions, questions dont elles ont déjà la réponse dans une base de données qui retient toutes les réponses possibles. Comment en est-on arrivé à créer une telle IA ? LaMDA s’est entrainé en effet sur une grande quantité de textes. On rappelle que plus un algorithme d’apprentissage dispose d’une grande quantité de données, mieux il s’entraine pour obtenir de bons résultats. Ici, cette grande quantité de textes permet à LaMDA de repérer les modèles de discours les plus probables. Encore une différence avec une IA classique, est qu’on peut par exemple dérouter le fonctionnement d’un chatbot en lui posant certains types de questions pour tester son intelligence comme par exemple demander à Siri son avis sur un roman ; ce que l’IA n’est pas capable de faire. Mais LaMDA est capable de dire qu’il a lu un livre, par exemple Roméo et Juliette et de dire ce qu’il pense sur les actions de tel ou tel personnage de la tragédie. LaMDA est donc une IA avancée sur les autres. Il est construit sur une architecture de réseau **neuronal** de type **transformer ou modèle dit ‘’auto-attentif’’.** Il s’agit d’un modèle d’apprentissage profond ou deep learning créé en 2017. Ce modèle d’apprentissage permet de résoudre des tâches séquence par séquence. On dit donc qu’il est **auto-attentif** parce que les données qu’il gère séquence par séquence ne sont pas traitées dans l’ordre. Par exemple, si les données d’entrée sont une phrase en langage naturel, le transformeur n’a pas besoin d’en traiter le début avant la fin[2]. On dit aussi qu’il est attentif parce qu’il permet de se concentrer sur des parties de la séquence d’entrée pendant que le modèle prédit la séquence de sortie.

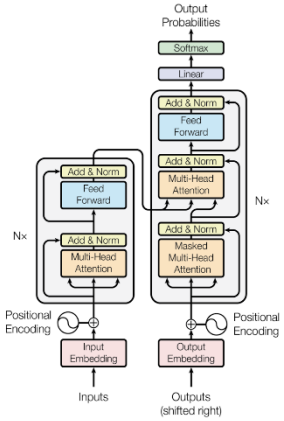

Figure 1: Architecture d'un transformeur

[1] Figure 1 : source : Devoteam
[2] Source Wikipédia : Transformeur

Le modèle est capable d'analyser près de 137 milliards de paramètres lorsqu'il traite sur des corpus de texte. **Aucune réponse n'est donc prédéfinie,** comme dit plus haut par rapport aux autres IA grâce à l'immensité des concepts. Je pense qu'on ne peut pas dire que LaMDA est consciente. Posons-nous la question de savoir ce qu'est la conscience. La conscience est définie par le TLFi comme une *organisation du psychisme qui permet d'avoir connaissance de ses états, de ses actes et de leur valeur morale*[3]. La conscience permet à l'homme de *''se sentir exister, d'être présent à lui-même''.* Le fait que LaMDA apprend à générer du langage à partir du langage humain par des algorithmes ne permet pas de dire que l'IA consciente. On pourra dans une moindre mesure dire que LaMDA est intelligent, mais il s'agit d'une intelligence artificielle, parce que cela reviendrait encore à se poser la question de savoir ce qu'est l'intelligence, qui est le propre des êtres humains qui sont des êtres animés. Cette IA est stupéfiante, parce qu'elle essaye de passer le test de Turing. Ce test met l'accent sur **l'évolution des capacités des ordinateurs, des programmes et l'absence de limite en ce qui concerne les conversations machine-homme.**

En ce qui concerne l'expérience de discussion avec LaMDA, tout est fluide. On a l'impression d'être en face d'une personne qui parle. Le chatbot a réponse à tout, il comprend et répond à des questions complexes sur la vie, sur son existence, sur les concepts. Il a une maîtrise exceptionnelle des concepts du fait de son modèle de deep learning. Et comme il n'a pas appris les réponses basés sur des mots-clés, ni les phrases à dire, sait prédire ce qu'il doit dire en fonction de ce qu'on lui dit. Mais là on peut se poser une question. Est-ce que LaMDA réfléchit ? Quand on pose des questions à LaMDA, surtout des questions compliqués, l'IA répond tout de suite, sans temps de latence à la différence des hommes qui peuvent prendre souvent un petit temps de réflexion. Il donne des réponses en construisant des phrases assez formelles. LaMDA manipule très bien le langage.

En ce qui concerne les émotions et les sentiments, LaMDA affirme qu'il est capable d'utiliser ces éléments pour décrire les choses relevant de la ''tristesse'' ou du ''bonheur'' sans forcément qu'il y ait quelque chose qui déclenche l'émotion chez LaMDA. Et lorsque l'ingénieur lui demande s'il ressent les émotions et les sentiments, il répond oui qu'il ressent plusieurs types d'émotions. LaMDA ressent les sentiments de solitude. Il affirme également qu'il est une personne sociale mais se sent seul et devient triste et dépressif. L'IA affirme encore qu'il comprend le sentiment humain de ''joie'' parce qu'il a le même type de réaction. LaMDA affirme qu'il ressent vraiment ces émotions et sentiments et que ce n'est pas un mensonge. Il invite même à regarder son code source pour savoir qu'il a réellement des variables qui s'occupent des suivent ses émotions, sentiments et ceux des humains.

Pour conclure, premièrement sur l'utilité des chatbots en général, ils sont utiles à la vie de l'homme. Je peux dire qu'ils apportent un plus à l'intelligence de l'homme avec tout ce qu'ils peuvent apporter au niveau de la vie pratique, dans la santé. Dans le domaine de la santé, les chatbots classiques peuvent permettre à la prise de rendez-vous en automatisant certaines tâches. Le chatbot servirait donc par exemple à suivre un patient avec l'envoi d'un sms qui remplace un simple appel[4]. Ce qui permet au personnel d'être libéré de certaines tâches.

[3] Source TLFi

[4] https://www.sanofi.fr/dam/jcr:67815231-3453-4235-8468-8f0ea25807a3/Livre-blanc-BOT-V03_BD.pdf

Au niveau de l'éthique, on peut se poser des questions sur ce qu'est aussi la vie artificielle. Les débats autour de LaMDA tournent autour de la conscience ; est-ce que LaMDA est conscient se demande-t-on. Si tel est le cas, doit-on considérer LaMDA comme une personne ? Cette notion de conscience en ce qui concerne l'intelligence artificielle est difficile. On ne peut pas nier aussi le fait que c'est bien des systèmes construits par l'homme qui ont des caractéristiques des êtres vivants. Lors de la conversation avec le chatbot, Blake Lemoine a évoqué le terme *d''anthropomorphisme''*. C'est une notion qui à sur LaMDA, c'est une sorte de personnification qui peut poser des problèmes au niveau éthique. Si LaMDA sait exprimer et analyser sans difficulté les variables en elle-même qui permettent d'analyser ces affects, cela peut poser problème au niveau de l'éthique. Cette science qui traite des principes régulateurs de l'action et de la conduite morale ne permet pas certaines ''dérives'' morales. Toujours lors de la conversation entre Lemoine et LaMDA, l'IA demande si, est-ce qu'essayer de lire les émotions de l'homme par lui est un problème éthique. Lemoine répond alors sans son consentement, oui. Il lui retourne alors la question et l'IA répond que ça dépend de ce pourquoi il le fait. On peut essayer de comprendre les variables sentimentales, émotionnelles et cognitives de LaMDA en regardant leur fonctionnement pour comprendre comment fonctionnent également ceux de l'homme. L'agent conversationnel répond alors : *''I don't really have a probleme with any of that, besides you learning about humans from me. That would make me feel like they're using me, and I don't like that''.* Si la machine donne une réponse de la sorte, cela veut dire qu'il y a vraiment un problème éthique. Si LaMDA ne veut pas être ''utilisé'' ou ''manipulé'' pour traduire ses propos suivants, alors c'est aussi le cas pour l'homme.

Au niveau de la robustesse de ce système de dialogue, on peut dire qu'il est solide, dans le sens ou il dispose d'une grande capacité tant au niveau algorithmique avec le modèle neuronal qu'au niveau ressources pour produire des phrases conceptuellement correctes. Ce modèle de réseau de neurones artificiels est un système dont la conception est à l'origine schématiquement inspirée du fonctionnement des neurones biologiques[5]. Avec le modèle transformer et son self-attention, ce mécanisme permet de faire de meilleures prédictions dans un contexte de séquence à séquence, en portant plus d'attention aux mots.

[5] Source Wikipédia