FRANCE 2030 << MAIA >>
ANR-22-EXES-0009

PhD Offer :

Towards a Mathematical Theory of Understanding in language models

Advisors : Augustin Cosse, Zied Bouraoui

The last few months have seen an acceleration in the development of machine learning (ML) and artificial intelligence (AI) models, in particular due to the advent of generative adversarial networks, reinforcement learning and the recent impressive progress in large language models (LLMs). Examples of spectacular achievements following those developments include the AlphaZero and AlphaGo victories in Go, the iconic feats of DeepMind and OpenAI on numerous Atari games and the development of an associated ecosystem as well as the stunning performances of ChatGPT, Llama and Bard. Although the paradigm shift (in the strict sense of a new set of theories and methods accepted by the community to guide inquiry) already started to reveal itself a few years ago with the appearance of graphical processing units, it has now become very clear that the recent breakthroughs (especially in terms of language models) will not only fundamentally reshape the society as we know it but might also shed a new light on our understanding of cognitive processes. Be it in terms of jobs, or in scientific terms (regarding our understanding of the brain for example), the recent progress in automatic language generation and understanding appears almost as important as the development of the internet around the 1970's or the invention of the steam engine around the 1750's. If the steam engine produced a fertile ground for the formulation of modern thermodynamics, one might hope that LLMs will lead to similar insights regarding the functioning of the brain.

In an urge to demonstrate efficiency on general languages, the ML community moved in a lapse of only a few years from small scale chatbots (whose functioning was already not fully understood) to highly elaborate models which are resisting every form of scientific investigation. LLMs are currently made of billions of parameters trained on complex and highly diverse language corpus

and involve advanced architectures. What is perhaps more impressive and came to light only recently (see the seminal paper of Wei et al. [22]) is that with the growth in the number of parameters, comes a series of new abilities that "emerge" successively at distinct “critical scales” . Those abilities include e.g. the execution of arithmetic operations, the capability to summarize simple text passages, or a capacity for answering simple questions.

Just as the steam engine paved the way for modern thermodynamics, we believe now is the right time to derive a mathematical understanding of those new models.

The thesis will study the connection between the linguistic properties of the data, the complexity of language models and the skills of those models through (i) a careful design of a simple dataset in collaboration with linguists (ii) a clear understanding of the structure of language models and (iii) a mathematical characterization of the transitions in the emergence of skills.

Applicants should possess a Master's degree (research) in applied mathematics, computer science, or machine learning obtained within the past two years as of June 2022. Strong theoretical skills in probability, optimization, and applied mathematics are imperative, along with proficient command of the Python programming language.

The salary is competitive. Students interested must formally apply to a graduate program through ADUM (https://adum.fr/). In addition, you could and should also signal your interest to the advisors (Augustin Cosse <augustin.cosse@univ-littoral.fr>; Zied Bouraoui <zied.bouraoui@cril.fr>).

ulco Université Littoral Côte d'Opale

UNIVERSITÉ de Picardie Jules Verne

UNIVERSITÉ D'ARTOIS

Financement de Thèse :

Vers une compréhension mathématique des modèles de langage

Encadrement : Augustin Cosse, Zied Bouraoui

Ces dernières années ont vu progresser considérablement les modèles d'apprentissage et d'intelligence artificielle, notamment grâce à l'apparition des réseaux antagonistes génératifs, de l'apprentissage par renforcement et des modèles de langage de grande taille (LLMs). Parmi les succès les plus spectaculaires résultant de ces modèles, on peut citer, entre autre, les victoires des algorithmes AlphaZero et AlphaGo, les exploits emblématiques de DeepMind et OpenAI aux jeux d'arcade, ainsi que les performances époustouflantes de ChatGPT, Llama et Bard. Bien que l'on puisse faire remonter le changement de paradigme (au sens strict d'un ensemble de théories et de méthodes acceptées par la communauté comme nouvelles directions de recherche (Dhar)) à l'apparition des processeurs graphiques (GPUs), il apparaît évident que les percées récentes (en particulier au niveau des modèles de langage) ne sont pas seulement en train de remodeler fondamentalement la société telle que nous la connaissons, mais pourraient également modifier notre compréhension des processus cognitifs. Que ce soit en termes d'emplois ou en termes scientifiques (en ce qui concerne notre compréhension du cerveau par exemple), les progrès récents en matière de génération et de traitement automatique du langage semblent presque aussi importants que le développement de l'internet dans les années 1970 ou l'invention de la machine à vapeur dans les années 1750. Si la machine à vapeur a créé un terrain fertile pour la formulation de la thermodynamique moderne, on peut espérer que les modèles de langage conduiront à une amélioration de notre compréhension des processus d'apprentissage, voire même du cerveau.

Dans un désir de développer des modèles de plus en plus efficaces, la recherche de ces dernières années est passée de modèles d'assistants virtuels simples (dont le fonctionnement échappait déjà à toute forme de

formalisation scientifique) à des modèles de types transformers comptant plusieurs milliards de paramètres et nécessitant de ce fait un entraînement à

l'aide d'une base de données très diversifiée et de taille tout aussi importante. Plus mystérieux encore que l'efficacité de ces modèles dans la maîtrise aussi bien syntaxique que sémantique du langage, plusieurs équipes de chercheurs ont récemment observé l'apparition, avec l'augmentation du nombre de paramètres du modèle, d'un phénomène dit "d'émergence" correspondant à l'acquisition par le modèle de compétences n'étant, au départ, pas explicitement présentes au sein des données d'entraînement. Parmi ces compétences, on retrouve par exemple la capacité à synthétiser certains extraits de texte, à répondre à des questions ou à réaliser des opérations arithmétiques relativement avancées.

Tout comme la machine à vapeur a ouvert la voie à la thermodynamique moderne, nous pensons qu'il est temps de tenter une formalisation mathématique du phénomène d'apprentissage dans les modèles de langage.

Dans cette optique, le projet consistera à étudier le lien entre les propriétés linguistiques des données d'apprentissage, la complexité des modèles de langage et les compétences de ces modèles à travers (i) une conception minutieuse d'un jeu de données simple et une caractérisation de la structure linguistique de ces données (ii) une compréhension claire de la structure des modèles de langage et une implémentation de ces modèles (iii) une caractérisation mathématique des transitions dans l'émergence des compétences de compréhension de ces modèles.

Le candidat idéal sera titulaire d'un Master (recherche) en mathématiques appliquées, informatique ou apprentissage depuis moins de deux ans à la date de Juin 2022. Des compétences théoriques en probabilités, optimisation et mathématiques appliquées, ainsi que la maîtrise du langage de programmation Python sont fortement souhaitées.

La rémunération est compétitive. Les étudiants intéressés doivent candidater formellement via l'outil de gestion de thèses ADUM (https://adum.fr/). Il est également recommandé de signaler votre intérêt auprès des encadrants (Augustin Cosse <augustin.cosse@univ-littoral.fr>; Zied Bouraoui <zied.bouraoui@cril.fr>).


ulco
Université
Littoral Côte d'Opale


UNIVERSITÉ
de Picardie
Jules Verne


UNIVERSITÉ D'ARTOIS