

Contexte général de la thèse

L'arrivée récente des grands modèles de langage (LLM) et de leurs outils associés pour le grand public révèle des défis majeurs pour la société. Parmi les nombreux domaines qui sont, ou seront, impactés par ces modèles génératifs, le domaine biomédical est l'un de ceux qui attirent actuellement l'attention des industriels, des chercheurs, mais aussi du grand public. En effet, le besoin en outils et en applications potentielles semble immense, que ce soit au niveau du traitement de documents textuels, de l'imagerie médicale, ou même de l'interaction vocale. En raison de la nature sensible des données personnelles traitées et des craintes de la société associées aux outils d'aide à la décision, le travail en traitement automatique du langage naturel (TALN) doit innover en abordant les problématiques inhérentes à ce domaine. La position de doctorat est financée par le projet ANR MALADES, dans lequel nous présentons des approches innovantes pour l'intégration des LLM dans les centres de santé. L'objectif du projet MALADES est d'équiper ces centres avec des outils de TALN dérivés des LLM et adaptés au domaine biomédical tout en maintenant la souveraineté des modèles et le contrôle total de leurs données de santé.

Objectifs

Génération d'instructions textuelles orientées vers la biomédecine pour l'entraînement des LLM : Une des façons de former un LLM orienté vers le chat est d'utiliser un ensemble d'instructions, sous forme de questions/réponses, à travers un modèle de récompense basé sur l'apprentissage par renforcement. Nous explorerons deux stratégies pour obtenir ces instructions, tout en ayant une annotation humaine restreinte : 1) Utilisation de dossiers d'examen médical, où nous pouvons obtenir les questions et réponses associées (à la fois du professeur et des étudiants), 2) Instructions auto-générées avec des approches pour générer artificiellement des données pour l'entraînement du modèle. La première partie, qui implique l'utilisation de données d'examen, nécessitera une annotation manuelle des réponses des étudiants dans une certaine mesure, afin de mettre en place le modèle de récompense basé sur la qualité des réponses. La deuxième, qui implique la génération d'instructions artificielles, reposera également sur des données annotées pour générer des instructions plus robustes.

Étude des biais des LLM concernant les déterminants sociaux : Alors que les LLM biomédicaux seront inévitablement utilisés de plus en plus par la société, que ce soit au niveau des professionnels de la santé ou du grand public, il est clair que les LM contiennent des biais inhérents dans les données (biais sur le genre, l'ethnie, etc.) utilisées pour leur entraînement, ayant un impact sur les résultats fournis par ces modèles. Nous cherchons à compléter l'évaluation en étudiant les biais potentiels contenus dans les LLM, en nous concentrant ici sur les déterminants sociaux des dossiers des patients. Notre objectif n'est pas de mener une étude exhaustive des biais présents dans les modèles, mais plutôt d'établir les jalons nécessaires pour contrôler les modèles de langage, notamment en ce qui concerne les dossiers de patients incluant des déterminants sociaux (par exemple, lieu de vie, état matrimonial ou niveau d'éducation).

Construction et annotation de tâches médicales originales : Les modèles actuels se concentrent principalement sur des tâches de TALN classiques dans le domaine médical. Cependant, ces tâches sont mal conçues pour évaluer les modèles génératifs. Afin de combler cette lacune, nous proposons trois cas d'utilisation originaux : 1) Examens des étudiants en santé sur divers domaines

médicaux avec des questions et réponses ouvertes, 2) L'évaluation et la production potentielle de rapports pré-décharge des patients hospitalisés à partir de leurs dossiers, et 3) Scénario de médecin réel / patient virtuel simulé.

Évaluation croisée entre les entrepôts de données : Les évaluations récentes des modèles de langage dans les domaines médicaux ont montré des disparités dans les résultats obtenus en fonction des données sur lesquelles ces modèles sont formés et évalués. Cette différence sur l'origine des données est rarement évaluée, seules les tâches choisies comptent finalement pour la performance intrinsèque des modèles. Nous voulons mener une évaluation croisée des modèles biomédicaux dans les centres de santé pour évaluer l'adéquation du modèle pour le déploiement auprès d'autres prestataires de soins de santé. Cela nous permettra de déterminer si les modèles (en totalité ou en partie) doivent être retravaillés dans chaque centre de santé. Cet aspect plus prospectif dépendra des collaborations avec d'autres entrepôts de données de différentes tailles qui ont exprimé leur volonté de participer au processus d'évaluation.