

Utiliser les grands modèles de langage à bon escient: étude de cas pour classifier les retours d'expérience patients

Les nouveaux modèles de langage issu de l'intelligence artificielle (*Large Language Model* - LLM) sont extrêmement prometteurs pour aider et accélérer l'utilisation des retours d'expérience et des textes libres dans le domaine médical. Nous présentons ici les opportunités et les points d'attention pour utiliser ces modèles, non pas au niveau du patient individuel, mais au niveau des projets d'amélioration systémique de la qualité des soins. Nous illustrons ensuite notre propos avec la description d'un projet pour classifier les retours d'expérience patients en dimensions de la satisfaction et en valence (retour positif, neutre ou négatif).

L'enquête de satisfaction des patients des HUG reçoit environ 11'000 commentaires libres par année. En plus de la prise en charge individuelle de ces retours d'expérience, les équipes doivent aussi les classifier, ce qui permet d'identifier des problèmes systémiques, et de mieux choisir les projets d'amélioration de la qualité. Par exemple, plusieurs commentaires sur le manque d'intimité dans une unité peuvent mettre en lumière un problème d'aménagement causant des tensions. De même, des commentaires de plusieurs patients qui se plaindraient de la prise en charge de leur douleur pourraient suggérer la nécessité d'une formation complémentaire.

La classification des retours d'expérience est à ce jour réalisée par des collaborateurs qui lisent chaque commentaire. Cette méthode prend beaucoup de temps, avec deux conséquences : les équipes ont moins de temps pour les investigations, et la classification est parfois faite plusieurs semaines après que le commentaire soit donné.

L'application de LLM sur >30'000 commentaires reçus depuis 2019 a permis de classifier correctement environ 35% des commentaires, correspondant à un gain de temps d'environ 40 jours-personne par année. Au-delà de ces premiers résultats prometteurs, la présentation indiquera pourquoi 65% des commentaires n'étaient pas suffisamment bien classés pour être traités par les LLM, et les pistes d'amélioration.

Delphine COURVOISIER, HUG