

L’hôpital EpiCURA, l’Hôpital Foch (Suresnes), l’Université de Mons et le CHU Saint-Pierre ont collaboré à la réalisation d’une des premières études mondiales qui évalue la performance de ChatGPT dans la prise en charge de cas cliniques réels. S'il en ressort que l'IA est capable d'émettre des diagnostics et de suggérer des examens complémentaires et traitements pertinents, elle n'est pas encore en mesure aujourd'hui de se substituer au discernement du praticien humain.

Le développement de l'intelligence artificielle est fulgurant. ChatGPT semble capable de passer et réussir des examens universitaires et présente de surprenantes connaissances théoriques en médecine. De nombreuses questions se posent toutefois quant aux limites des systèmes d'IA actuels, et notamment leur capacité à proposer des prises en charge médicales pour les humains.

Étude internationale sur 45 patients

Les Professeurs Jérôme Lechien, chirurgien ORL, et Stéphane Hans, chef de service ORL de l'Hôpital Foch, viennent de publier les résultats d'une étude internationale menée sur 45 patients présentant des pathologies fréquentes en médecine générale ou oto-rhino-laryngologie et dont le tableau clinique a été présenté à l'intelligence artificielle ChatGPT.

Dans cette étude, les médecins ont soumis à l'IA les symptômes, antécédents, médicaments et données de l'examen clinique des patients, et celle-ci a été interrogée sur les potentiels diagnostics différentiels, diagnostic principal, examens complémentaires et traitements.

Deux médecins ont ensuite analysé les résultats de performance de l'IA à l'aide d'un nouvel o util
clinique
développé et validé lors de cette étude, l'Artificial Intelligence Performance Instrument (AIPI).

Diagnostic plausible et correct dans 63.5% des cas

Globalement, ChatGPT a proposé un diagnostic plausible et correct dans 63.5% des cas et des diagnostics différentiels (alternatifs) plausibles dans 28.5% des cas. Les examens complémentaires proposés par l'IA afin de confirmer le diagnostic étaient pertinents et totalement ou partiellement nécessaires dans 29% des cas. Quant aux traitements, ils étaient pertinents et nécessaires dans 22% des cas, et pertinents mais incomplets dans 16% des cas. Dans le reste des cas, les procédures thérapeutiques proposées étaient une association de traitements pertinents, nécessaires, inadéquats et inutiles.

Quelle place pour l'IA ?

D'une façon globale, l'IA fournit une liste de possibilités telle une encyclopédie virtuelle très performante mais ne serait pas (encore) capable de discerner et classer les diagnostics et examens complémentaires selon leur pertinence. Si l'IA peut s'avérer très utile dans l'accompagnement de la formation des jeunes médecins, elle ne se substitue pas encore au discernement humain du praticien. De plus, elle propose souvent une liste d'examens complémentaires dont une grande partie n'est pas utile ou pertinente, ce qui peut mener à des confusions chez l
es patients qui utiliseraient l'IA avant de consulter leur médecin.

« De futures recherches sont nécessaires pour explorer le potentiel de l'IA en médecine et lui trouver une place permettant d'améliorer la qualité des soins donnés aux patients. » concluent les chercheurs.