

Χρήση LLMs για τη διάγνωση αναπνευστικών παθήσεων σε ασθενείς που προσέρχονται στο ΤΕΠ

Έξαρχος Κωνσταντίνος^{1,2}, Μούτσιου Αργυρώ Σωτηρία^{2,3}, Κήττας Απόστολος¹

¹Τμήμα Πνευμονολογίας, Γεν. Νοσοκ. Ιωαννίνων “Γ. Χατζηκώστα”, Ιωάννινα, Ελλάδα

³Μονάδα Εντατικής Θεραπείας, Γεν. Νοσοκ. Ιωαννίνων “Γ. Χατζηκώστα”, Ιωάννινα, Ελλάδα

²Τμήμα Επειγόντων Περιστατικών, Γεν. Νοσοκ. Ιωαννίνων “Γ. Χατζηκώστα”, Ιωάννινα, Ελλάδα



Εισαγωγή

Τα Μεγάλα Γλωσσικά Μοντέλα (LLMs: Large Language Models) εξελίσσονται ραγδαία τα τελευταία χρόνια και έχουν δοκιμαστεί σε μια πληθώρα κλινικών προβλημάτων με ποικίλα αποτελέσματα. Η χρήση των LLMs στο Τμήμα Επειγόντων Περιστατικών (ΤΕΠ) έχει δείξει ενθαρρυντικά αποτελέσματα στο triage καθώς και στη διαφοροδιάγνωση περιστατικών.

Σκοπός

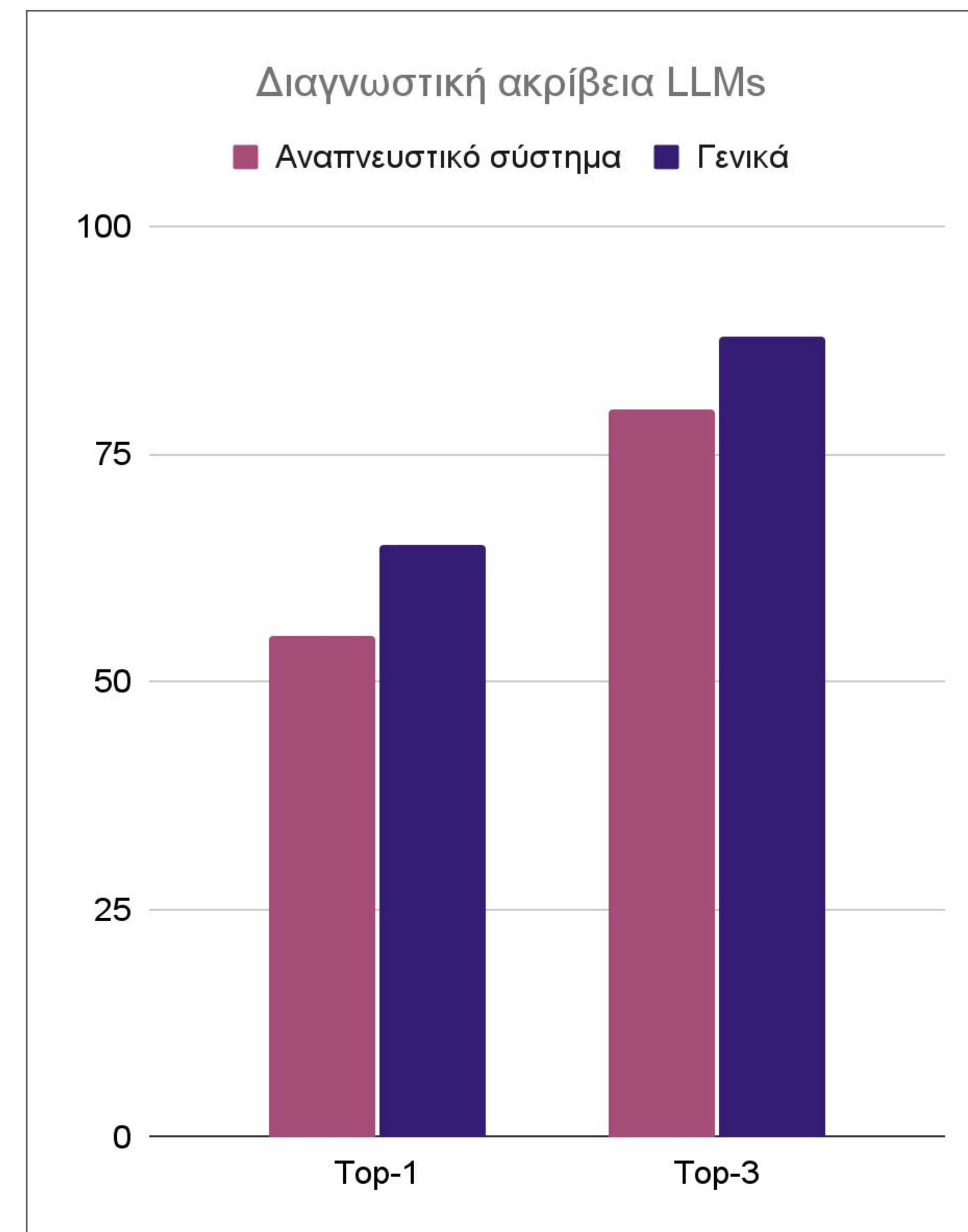
Στην παρούσα εργασία επιχειρήσαμε να εκτιμήσουμε την ακρίβεια των LLMs στη διαφοροδιάγνωση πραγματικών κλινικών σεναρίων από ασθενείς με αναπνευστικά ενοχλήματα που προσήλθαν στο ΤΕΠ.

Μέθοδοι

Παραθέσαμε στο ChatGPT (OpenAI) 60 κλινικά σενάρια ασθενών που επισκέφτηκαν το ΤΕΠ του Γενικού Νοσοκομείου Ιωαννίνων “Γ. Χατζηκώστα” αιτιώμενοι αναπνευστικά ενοχλήματα ποικίλης βαρύτητας και σπανιότητας και ζητήσαμε την πιθανότερη διάγνωση καθώς και τις 3 επικρατέστερες διαγνώσεις. Τα σενάρια περιελάμβαναν την περιγραφή της κύριας ενόχλησης, την ηλικία του ασθενούς και ενίοτε τις συνοσηρότητες; δεν δόθηκαν πληροφορίες σχετικά με τα ζωτικά σημεία, την κλινική εξέταση, εργαστηριακά ή απεικονιστικά ευρήματα.

Αποτελέσματα

Η πιθανότερη διάγνωση που δόθηκε από το LLM ήταν σωστή στο 55% των περιπτώσεων ενώ η σωστή διάγνωση υπήρχε μέσα στις 3 επικρατέστερες διαγνώσεις στο 80% των περιστατικών.



Συζήτηση

Παρατηρούμε ότι η ακρίβεια του LLM είναι σημαντικά χαμηλότερη συγκριτικά με μελέτες που αφορούν όλο το φάσμα των περιστατικών που προσέρχονται σε ΤΕΠ και δεν περιορίζονται σε αναπνευστικά νοσήματα, όπως στην παρούσα εργασία. Για την καλύτερη αποτίμηση της ακρίβειας των LLMs σε πραγματικά περιστατικά απαιτούνται μεγαλύτερες σειρές ασθενών και πιθανά αξιολόγηση της ακρίβειας με την σταδιακή προσθήκη στοιχείων που αφορούν την κλινική εξέταση, τα εργαστηριακά ή και τα απεικονιστικά δεδομένα, καθώς και δοκιμή περισσότερων LLMs.

Βιβλιογραφία

- Levine, D.M. et al. (2024) ‘The diagnostic and triage accuracy of the GPT-3 artificial intelligence model: an observational study’, The Lancet. Digital health.
- Sorich, M.J. et al. (2024) ‘The Triage and Diagnostic Accuracy of Frontier Large Language Models: Updated Comparison to Physician Performance’, Journal of medical Internet research.