

Οι προκαταλήψεις των Μεγάλων Γλωσσικών Μοντέλων.
Δεδομένα από την σύγκριση διαφορετικών μοντέλων σε Ελληνικά και
Αγγλικά.

Γιώργος Μικρός

Hamad Bin Khalifa University, Qatar

Email: gmikros@hbku.edu.qa

Περίληψη

Τα Μεγάλα Γλωσσικά Μοντέλα – ΜΓΜ (Large Language Models – LLMs) αποτελούν αυτή τη στιγμή τη σημαντικότερη τεχνολογική εξέλιξη στην ιστορία της Τεχνητής Νοημοσύνης. Μπορούν να παράγουν κείμενα τα οποία δεν μπορούν να διακριθούν από τα κείμενα που παράγουν οι άνθρωποι. Επίσης, μπορούν να συνδυάζουν ένα τεράστιο όγκο κειμενο-κεντρικής πληροφορίας και να απαντούν σε σύνθετα ερωτήματα, μετατρέποντάς τα σε ένα από τα πιο χρήσιμα εργαλεία επέκτασης της ανθρώπινης γνώσης.

Οι απαντήσεις που παίρνουμε από τα ΜΓΜ είναι συνεκτικές, αλλά συχνά μη ακριβείς ως προς την ακρίβεια των γεγονότων που μεταφέρονται. Ωστόσο, ένας μεγαλύτερος κίνδυνος από την χρήση τους προκύπτει από την αφελή πεποίθηση, ότι οι πληροφορίες που παίρνουμε είναι ουδέτερες τόσο ιδεολογικά όσο και κοινωνικά. Αντίθετα, πλήθος ερευνών τα τελευταία δύο χρόνια, έχει τεκμηριώσει σημαντικές ιδεολογικές και κοινωνικές προκαταλήψεις σε όλα τα σημαντικά ΜΓΜ.

Η παρουσίασή μας στοχεύει στο να καταδείξει τα πιο πρόσφατα ερευνητικά αποτελέσματα σχετικά με τις προκαταλήψεις στα ΜΓΜ. Επιπλέον θα επεκταθούμε και στο ερώτημα κατά πόσο διαφορετικές γλώσσες επιβάλλουν διαφορετικές προκαταλήψεις με απώτερο σκοπό να ερμηνεύσουμε κατά πόσο τα ΜΓΜ, μέσω των γλωσσικών δεδομένων εκπαίδευσής τους, προσεγγίζουν τις διαφορετικές πολιτισμικές δυναμικές και τις αναπαριστούν με πιστότητα.