

Ρομπότ μου, ρομποτάκι, είναι αλήθεια αυτό που διάβασα;

Η παραπληροφόρηση θεωρείται ένα από τα πιο σημαντικά προβλήματα της εποχής μας. Φυσικά, δεν είναι καινοφανές φαινόμενο, καθώς υπάρχει όσο υπάρχει επικοινωνία μεταξύ ανθρώπων. Η αντιμετώπιση της παραπληροφόρησης έγινε πιο επιτακτική επειδή τα σύγχρονα ψηφιακά μέσα και ιδίως τα ηλεκτρονικά κοινωνικά δίκτυα μας δίνουν τη δυνατότητα σχεδόν άμεσης διασποράς πληροφοριών ανεξαρτήτως της ορθότητάς τους, αλλά ο έλεγχος της ορθότητας δεν συμβαίνει το ίδιο γρήγορα. Αυτό σε συνδυασμό με το γεγονός ότι ο βασικός μηχανισμός κερδοφορίας για τις εταιρείες που τα αναπτύσσουν είναι η εμπλοκή των χρηστών (user engagement) και τα έσοδα από διαφημίσεις, δημιουργεί συνθήκες που ευνοούν τη διασπορά προκλητικών τίτλων και εύπεπτων άρθρων, τα οποία συχνά περιέχουν παραπληροφόρηση.

Η χρήση τεχνητής νοημοσύνης έχει προταθεί από ερευνητές για τον περιορισμό της παραπληροφόρησης με διάφορες μορφές. Μία από αυτές είναι η αυτοματοποίηση του ελέγχου ισχυρισμών (claim verification ή fact-checking), και η εφαρμογή της απασχολεί κλάδους της τεχνητής νοημοσύνης όπως η επεξεργασία φυσικής γλώσσας και



Του **ΑΝΔΡΕΑ ΒΛΑΧΟΥ**, Καθηγητή Τμήματος Επιστήμης και Τεχνολογίας των Υπολογιστών του Πανεπιστημίου του Cambridge

η μηχανική μάθηση, όσο και τη δημοσιογραφία. Η έρευνα έχει συνεισφέρει στη χρησιμοποίηση των μεθόδων τεχνητής νοημοσύνης που αναπτύσσονται από οργανισμούς οι οποίοι εστιάζονται στον έλεγχο ισχυρισμών όπως η Full Fact στο Ηνωμένο Βασίλειο. Καθώς οι μέθοδοι που χρησιμοποιούνται στον αυτόματο έλεγχο ισχυρισμών εναντίον της παραπληροφόρησης βελτιώνονται συνεχώς, δηλαδή το ρομποτάκι του τίτλου γίνεται όλο και καλύτερο, είναι σημαντικό να σκεφτούμε τι επιθυμούμε από αυτές. Μια σημαντική απαίτηση είναι οι μέθοδοι αυτές να μην επιστρέφουν μια μονολεκτική απάντηση του τύπου αληθές/ψευδές, αλλά να επιστρέφουν στον χρήστη αποδεικτικά στοιχεία και την αιτιολόγηση της πρόβλεψής τους. Οι μονολεκτικές απαντήσεις είναι δύσκολο να ελεγχθούν για την ορθότητά τους, και δεν προάγουν τον εμπειριστικό διάλογο βάσει στοιχείων στην κοινωνία. Αξίζει επίσης να σημειωθεί σε αυτό το σημείο, ότι

η πρόοδος στα μεγάλα γλωσσικά μοντέλα, έχει περιπλέξει τη χρήση της τεχνητής νοημοσύνης στον έλεγχο ισχυρισμών, καθώς αφενός έχει βελτιώσει την ακρίβεια των προβλέψεων, αλλά αφετέρου τα μοντέλα αυτά συχνά επιστρέφουν λανθασμένες/αλλοιωμένες πληροφορίες. Για αυτόν τον λόγο, είναι σημαντικό συστήματα ελέγχου ισχυρισμών να επιστρέφουν τις πηγές των στοιχείων που χρησιμοποιούν στην αιτιολόγησή τους, και όχι μόνο πληροφορίες που έχουν αποθηκεύσει σε συμπίεση μορφή στα βάρη των νευρώνων τους.

Η δεύτερη απαίτηση είναι κατά την ανάπτυξη των συστημάτων να λαμβάνουμε υπόψη μας τον τρόπο με τον οποίο θα χρησιμοποιηθούν. Ποιοι θα έχουν την ιδιοκτησία των συστημάτων; Οι μηχανές αναζήτησης και οι εταιρείες που τις ελέγχουν ή μήπως η κυβέρνηση κάθε χώρας; Ποιανών οι ισχυρισμοί θα ελέγχονται, των πολιτικών, των προσώπων δημοσίου ενδιαφέροντος, ή όλων των πολιτών; Και τι θα συμβαίνει όταν ένας ισχυρισμός κρίνεται αναληθής; Θα λογοκρίνεται, π.χ.

θα απαγορεύεται η διασπορά του στα μέσα κοινωνικής δικτύωσης; Ή θα επιτρέπεται, αλλά θα συνοδεύεται υποχρεωτικά από κάποια ετικέτα ή αποδεικτικά στοιχεία;

Πρόσβαση σε όλους

Η δική μου σύσταση είναι ότι τα συστήματα ελέγχου ισχυρισμών θα πρέπει να γίνουν προσβάσιμα σε όλους, έτσι ώστε να αποφύγουμε την ενίσχυση των μονοπωλίων στη χρήση τους. Θα ήταν επικίνδυνο αν δύο-τρεις εταιρείες μέσω των συστημάτων τους μπορούσαν να καθορίσουν τι θεωρούμε αληθές ή ψευδές στην κοινωνία, ιδίως δεδομένων των υπάρχοντων ολιγοπωλίων στις μηχανές αναζήτησης και στα κοινωνικά δίκτυα, τα οποία ήδη καθορίζουν σε μεγάλο βαθμό την πρόσβασή μας στις πληροφορίες. Όχι μόνο γιατί πάντα υπάρχει το ενδεχόμενο λαθών, αλλά γιατί πολλές φορές δεν υπάρχει μια απλή απάντηση, καθώς σε πολλές περιπτώσεις τα στοιχεία επιδέχονται πολλαπλές ερμηνείες όσον αφορά το αληθές ή ψευδές ενός ισχυρισμού. Για παράδειγμα, είναι η χορτοφαγία καλή για την υγεία; Σε τέτοιες περιπτώσεις η απάντηση εξαρτάται από την εφαρμογή της, π.χ. μια δίαιτα με πίτσα τυρί-ντομάτα και ζαχαρούχα ποτά μάλλον δεν είναι. Ακόμα πιο πολύπλοκη είναι η αξιολόγηση ισχυρισμών σε θέματα πολιτικής, π.χ. αν το Brexit ήταν καλή επιλογή για το Ηνωμένο Βασίλειο. Σε τέτοιες περιπτώσεις είναι σημαντικό να δώσουμε τη δυνατότητα στον απλό χρήστη να κάνει ο ίδιος τον έλεγχο των ισχυρισμών, δίνοντάς του πρόσβαση στα στοιχεία με τέτοιο τρόπο ώστε να μπορεί να τα ερμηνεύσει και να διαμορφώσει τη γνώμη του. Με αυτόν τον τρόπο οι χρήστες θα μπορέσουν να γίνουν ενεργοί καταναλωτές πληροφοριών, και να συνδράμουν στην ενημέρωση της ευρύτερης κοινωνίας γύρω τους. Επιστρέφοντας στο ρομποτάκι του τίτλου του άρθρου, πιστεύω ότι θα πρέπει να βοηθά τον χρήστη να σχηματίσει τη δική του άποψη, αλλά να μην είναι το ίδιο ο τελικός κριτής της αλήθειας.

