



Οικοδόμηση εμπιστοσύνης στην Τεχνητή Νοημοσύνη

Τους τελευταίους μήνες, οι εταίροι του SHIELD συγκεντρώνουν στοιχεία μέσω δραστηριοτήτων αξιολόγησης αναγκών, ομάδων εστίασης και συζητήσεων με ειδικούς σε όλες τις χώρες-εταίρους. Οι δραστηριότητες αυτές βοηθούν το έργο να μεταβεί από γενικούς στόχους σε συγκεκριμένες προτεραιότητες σχεδιασμού: πώς πρέπει να απευθύνεται ένα chatbot στους νέους, τι είδους καθοδήγηση πρέπει να παρέχει, πότε πρέπει να παραπέμπει σε ανθρώπινη υποστήριξη και ποιες διασφαλίσεις πρέπει να υπάρχουν εξ αρχής;

Τα νέα άτομα χρειάζονται μια γλώσσα που να τους φαίνεται σαφής, χωρίς επικριτικό τόνο και εύκολα αναγνωρίσιμη. Οι γονείς και οι κηδεμόνες ανησυχούν ότι τα εργαλεία τεχνητής νοημοσύνης ενδέχεται να στερούνται ενσυναίσθησης ή συναισθηματικής ευαισθησίας. Οι επαγγελματίες που ασχολούνται με τη νεολαία, οι εκπαιδευτικοί και οι ειδικοί δίνουν ιδιαίτερη προσοχή σε κινδύνους όπως η μεροληψία, τα λάθη, η επικίνδυνη κλιμάκωση και η κατάχρηση δεδομένων. Σε όλες αυτές τις προοπτικές, η προστασία της ιδιωτικής ζωής αναδεικνύεται ως απαραίτητη προϋπόθεση για την εμπιστοσύνη.

Η διαδικτυακή έμφυλη βία διαμορφώνεται από την ντροπή, τον φόβο, την αβεβαιότητα, την κοινωνική πίεση και τη δυσπιστία απέναντι στα συστήματα αναφοράς. Επομένως, ένα chatbot που απλώς εντοπίζει επιβλαβή γλώσσα δεν αρκεί. Για να είναι χρήσιμη, η τεχνητή νοημοσύνη πρέπει να εντάσσεται σε ένα ευρύτερο πλαίσιο υποστήριξης που να δίνει προτεραιότητα στην προστασία της ιδιωτικής ζωής, να λαμβάνει υπόψη τα τραύματα, να περιλαμβάνει ανθρώπινη παρέμβαση και να λογοδοτεί στα άτομα που προορίζεται να εξυπηρετήσει. Αυτό συνάδει με τις ευρύτερες κατευθυντήριες γραμμές σχετικά με τη βία λόγω φύλου που διευκολύνεται από την τεχνολογία και την ηθική τεχνητή νοημοσύνη, οι οποίες δίνουν έμφαση στην ασφάλεια, τη διαφάνεια, το πλαίσιο και τα μέτρα προστασίας για τις πληγείσες κοινότητες [2,3,4].

ΤΙ ΣΗΜΑΙΝΕΙ ΑΥΤΟ ΓΙΑ ΤΟ SHIELD

Ο στόχος δεν είναι η ανάπτυξη «περισσότερης τεχνητής νοημοσύνης» για την ίδια την τεχνητή νοημοσύνη. Ο στόχος είναι η δημιουργία ασφαλέστερης και πιο αξιόπιστης ψηφιακής υποστήριξης για τους νέους που αντιμετωπίζουν κινδύνους στο διαδίκτυο. Τα στοιχεία σχετικά με τους διαλογικούς πράκτορες υποδηλώνουν ότι τα εργαλεία αυτά είναι πιο κατάλληλα όταν σχεδιάζονται ως συμπληρωματικά μέσα της ανθρώπινης φροντίδας και όχι ως υποκατάστατα της επαγγελματικής κρίσης ή των συστημάτων προστασίας [5,6].

Τους επόμενους μήνες, το SHIELD θα συνεχίσει να βελτιώνει τις προδιαγραφές των chatbot, να ενισχύει τις διαδικασίες προστασίας και να συνεργάζεται με τους εταίρους του, προκειμένου να διασφαλίσει ότι ο ηθικός σχεδιασμός θα παραμείνει στο επίκεντρο της ανάπτυξης. Η εμπιστοσύνη δεν θα προέλθει μόνο από την τεχνολογία. Θα προέλθει από την προσεκτική ακρόαση, τις διαφανείς επιλογές και τη συνεχή συνεργασία με τις κοινότητες που το SHIELD στοχεύει να υποστηρίξει.

Αν ένα chatbot τεχνητής νοημοσύνης έχει ως στόχο να υποστηρίξει νέα άτομα που αντιμετωπίζουν διαδικτυακή βία λόγω φύλου, τι δεν πρέπει ποτέ να κάνει λάθος;

Το ζήτημα αυτό αποκτά κεντρική σημασία για το έργο SHIELD. Καθώς οι εταίροι συνεχίζουν να αναπτύσσουν ένα chatbot που βασίζεται στην τεχνητή νοημοσύνη και μια διαδρομή υποστήριξης, η προσοχή δεν εστιάζεται μόνο στο τι μπορεί να κάνει η τεχνητή νοημοσύνη, αλλά και στο τι θα εμπιστεύονταν πραγματικά τα νέα άτομα, οι γονείς, οι εκπαιδευτικοί, οι εργαζόμενοι στον τομέα της νεολαίας και οι ειδικοί σε θέματα έμφυλης βίας. Αυτό έχει σημασία, διότι η έμφυλη βία δεν αποτελεί απλώς ένα τεχνικό πρόβλημα, έχει τις ρίζες της στις άνισες σχέσεις εξουσίας και μπορεί να εκδηλώνεται τόσο στον πραγματικό όσο και στον διαδικτυακό χώρο [1,2].

