

FOCUS - 2

Project Syndicate

«Να αναπτύξουμε την ΤΝ για τους ανθρώπους, όχι για να είναι άνθρωπος»

Όλο και περισσότερο οι άνθρωποι αρχίζουν να πιστεύουν τόσο έντονα στην τεχνητή νοημοσύνη, που της προσδίδουν ανθρώπινα χαρακτηριστικά, αποκτώντας την ψευδαίσθηση πως έχουν να κάνουν με μια πραγματική οντότητα

Η τεχνητή νοημοσύνη βρίσκεται στο επίκεντρο ενός παγκόσμιου διαλόγου όχι μόνο ως τεχνολογική καινοτομία αλλά και ως πεδίο που αναδιαμορφώνει θεμελιώδη ερωτήματα για την κοινωνία, την εργασία και την ίδια τη φύση της συνείδησης αγγίζοντας ζητήματα κοινωνικά, ηθικά και πολιτικά. Από τη μία πλευρά, η ταχύτατη πρόοδος δημιουργεί εντυπωσιακές εφαρμογές που μπορούν να βελτιώσουν τη ζωή μας, από την άλλη, γεννά ανησυχίες για την ψευδαίσθηση της αυτονομίας των συστημάτων και για τις κοινωνικές ή ψυχολογικές συνέπειες που αυτή μπορεί να επιφέρει.

Οι συζητήσεις κυμαίνονται ανάμεσα στον φόβο για την υπερβολική ανθρωπομορφοποίηση της τεχνητής νοημοσύνης και στην ανάγκη για ψύχραιμη αποτίμηση των πραγματικών κινδύνων, οι οποίοι σχετίζονται περισσότερο με τον τρόπο χρήσης της παρά με την ίδια την τεχνολογία. «Γι' αυτό η πρόκληση είναι πολιτική: πώς διασφαλίζουμε ότι η τεχνητή νοημοσύνη υπηρετεί τις δημοκρατικές αξίες και οι μικρότερες χώρες δεν μετατρέπονται σε ψηφιακές αποικίες των πολυεθνικών της ΤΝ».

Η αποστολή της ζωής μου ήταν να δημιουργήσω ασφαλή, ωφέλιμη τεχνητή νοημοσύνη που θα κάνει τον κόσμο καλύτερο. Αλλά τελευταία, ανησυχώ όλο και περισσότερο για το γεγονός ότι οι άνθρωποι αρχίζουν να πιστεύουν τόσο έντονα στην ΤΝ ως οντότητα με συνείδηση, ώστε να ζητούν «δικαιώματα της ΤΝ ή ακόμα και ιθαγένεια». Αυτή η εξέλιξη θα αποτελούσε μια επικίνδυνη στροφή για την τεχνολογία. Πρέπει να αποφευχθεί.



ΤΟΥ MUSTAFA SULEYMAN

Πρέπει να αναπτύξουμε την ΤΝ για τους ανθρώπους, όχι για να είναι άνθρωπος.

Σε αυτό το πλαίσιο, οι συζητήσεις για το αν η τεχνητή νοημοσύνη μπορεί πραγματικά να έχει συνείδηση συνιστούν αντιπερισπασμό. Αυτό που έχει σημασία βραχυπρόθεσμα είναι η ψευδαίσθηση της συνείδησης. Βρι-

σκόμαστε ήδη κοντά σε αυτό που αποκαλώ συστήματα «ΤΝ με φαινομενική συνείδηση» (Seemingly Conscious AI - SCAI), τα οποία θα μιμούνται αρκετά πειστικά τη συνειδητότητα.

Ένα SCAI θα είναι ικανό να χρησιμοποιεί άπταιστα τη φυσική γλώσσα, επιδεικνύοντας μια προσωπικότητα με πειθώ και συναισθηματικό αντίκτυπο. Θα διαθέτει μακρά και ακριβή μνήμη που θα τροφοδοτεί μια συνεκτική αίσθηση του εαυτού της και θα αξιοποιεί αυτή την ικανότητα για να διεκδικεί υποκειμενική εμπειρία (αναφερόμενη σε παλαιότερες αλληλεπιδράσεις και αναμνήσεις). Πολύπλοκες συναρτήσεις μέσα σε αυτά τα μοντέλα θα προσομοιάζουν τα εσωτερικά κίνητρα, ενώ η προηγμένη θέσπιση στόχων και ο προγραμματισμός θα ενισχύουν την αίσθησή μας ότι η ΤΝ δρα με πραγματική αυτονομία.

Η «ψύχωση» της ΤΝ

Όλες αυτές οι δυνατότητες υπάρχουν ήδη ή είναι προ των πυλών. Πρέπει να αναγνωρίσουμε ότι τέτοια συστήματα σύντομα θα είναι εφικτά, να αρχίσου-

με να σκεφτόμαστε τις συνέπειες και να θέσουμε κανόνες ενάντια στην επιδίωξη μιας απατηλής επίγνωσης.

Για πολλούς ανθρώπους, η αλληλεπίδραση με την ΤΝ ήδη μοιάζει σαν μια πλούσια, ικανοποιητική και αυθεντική εμπειρία. Οι ανησυχίες για τον κίνδυνο να εμφανίσει κανείς «ψύχωση της ΤΝ», προσκόλληση και προβλήματα ψυχικής υγείας αυξάνονται, καθώς υπάρχουν αναφορές ότι κάποιοι θεωρούν την ΤΝ έκφραση του Θεού.

Παράλληλα, επιστήμονες μου λένε ότι κατακλύζονται από ερωτήματα ανθρώπων που θέλουν να μάθουν αν τα μοντέλα της ΤΝ διαθέτουν συνείδηση κι αν είναι εντάξει να τα ερωτευτούν.

Βέβαια, το ότι κάτι είναι τεχνικά εφικτό στα συστήματα SCAI έχει ελάχιστα να κάνει με το αν ένα τέτοιο σύστημα μπορεί όντως να διαθέτει συνείδηση. Όπως επισημαίνει ο καθηγητής νευροεπιστήμης Ανίλ Σεθ, μια προσομοίωση καταγίδας δεν σημαίνει ότι βρέχει στον υπολογιστή σας. Αλλά πρακτικά πρέπει

να αναγνωρίσουμε ότι κάποιοι θα δημιουργήσουν συστήματα SCAI που θα υποστηρίζουν ότι πράγματι έχουν συνείδηση. Και, ακόμη περισσότερο, κάποιοι άνθρωποι θα τους πιστέψουν.

Ένας νέος άξονας διαίρεσης

Ακόμα κι αν αυτή που αντιλαμβανόμαστε ως συνείδηση δεν είναι πραγματική (ένα θέμα που θα προκαλέσει ατελείωτη συζήτηση), ο κοινωνικός αντίκτυπος θα είναι σίγουρα πραγματικός. Η συνείδηση είναι στενά συνδεδεμένη με την αίσθηση της ταυτότητας και την κατανόηση των ηθικών και νομικών δικαιωμάτων μας μέσα στην κοινωνία. Αν κάποιοι αρχίσουν να αναπτύσσουν συστήματα SCAI και αυτά τα συστήματα πείσουν ανθρώπους ότι μπορούν να υποφέ-

Γιατί δεν κινδυνεύουμε σήμερα από την

Τα τελευταία τρία χρόνια η τεχνητή νοημοσύνη (ΤΝ) βρίσκεται στο επίκεντρο των συζητήσεων. Από «μηχανές» που γράφουν κείμενα σε δευτερόλεπτα μέχρι συστήματα που «ζωγραφίζουν», αξιοποιώντας την ψηφιοποιημένη ανθρώπινη γνώση, πολλοί αναρωτιούνται αν πλησιάζουμε σε εποχή όπου οι άνθρωποι θα υποκατασταθούν από μηχανές. Σε έναν βαθμό συγχέονται οι επιστημονικές εξελίξεις με την επιστημονική φαντασία. Ωστόσο, οι πραγματικοί κίνδυνοι της ΤΝ δεν



ΤΟΥ ΘΕΟΔΩΡΟΥ Γ. ΚΑΡΟΥΝΟΥ

βρίσκονται – προς το παρόν – στην ίδια την τεχνολογία, αλλά στον τρόπο χρήσης της.

Η έρευνα για τη νοημοσύνη, είτε φυσική είτε τεχνητή, δεν είναι ιδεολογικά ουδέτερη.

Η ίδια η έννοια έχει ιστορικά διαμορφωθεί μέσα από διαφορετικές φιλοσοφικές σχολές. Έτσι, κάθε ερευνητική ομάδα υιοθετεί πλαίσιο που καθορίζει ποια χαρακτηριστικά θεωρούνται ουσιώδη.

Συνεπώς, τα μοντέλα ΤΝ αντανακλούν σε έναν βαθμό τις κατευθύνσεις των ερευνητών που τα αναπτύσσουν.

Διεθνώς, η έρευνα στην ΤΝ ακολουθεί τρεις βασικές γραμμές:

■ **Τους εμπειριστές**, που βασίζονται στη μάθηση από τεράστιους όγκους δεδομένων, με χαρακτηρι-

στικό παράδειγμα τα συστήματα βαθιάς μάθησης.

■ **Τους ορθολογιστές**, που επενδύουν στη λογική και στους κανόνες, αναζητώντας διαφάνεια και ερμηνευσιμότητα.

■ **Τη νευροσυμβολική προσέγγιση**, που επιχειρεί να συνδυάσει την ισχύ των νευρωνικών δικτύων με τη διαύγεια της λογικής.

Πανεπιστήμια και εταιρείες διεθνώς επενδύουν στο υβριδικό αυτό μοντέλο, που υπόσχεται πιο ασφαλή και κατανοητή ΤΝ, αν και ακόμη βρίσκεται σε πειραματικό στάδιο.

Τα θεμέλια των κατευθύνσεων ανικνεύονται σε μεγάλους πολιτισμούς. Ο ορθολογισμός των ελλήνων φιλοσόφων, που συνέχισαν άραβες στοχαστές, εμπνέει τη «συμβολική» ΤΝ. Ο εμπειρισμός, παρών στον βου-

δισμό αλλά και της νεότερης Ευρώπης, προαναγγέλλει τη μάθηση από δεδομένα. Ο συνδυασμός αυτών των δύο, όπως τον διατύπωσε ο Καντ, ανοίγει τον δρόμο για την «υβριδική» ΤΝ. Έτσι, η τεχνολογική εξέλιξη ενσωματώνει σοφία αιώνων από Ανατολή και Δύση.

Οι επιπτώσεις στην εργασία

Ένα από τα πιο συχνά ερωτήματα είναι αν η ΤΝ απειλεί άμεσα την εργασία. Μέχρι σήμερα δεν έχει καταγραφεί περίπτωση όπου εφαρμογές ΤΝ αντικατέστησαν πλήρως θέσεις, αλλά κυρίως τμήματα ρόλων ή καθήκοντα. Παρ' όλα αυτά, σε ορισμένους κλάδους μπορεί να επιφέρει σημαντικές αλλαγές – ειδικά όταν ένας εργαζόμενος χρησιμοποιεί αποτε-

