Νικ Μπόστρομ

Η Τεχνητή Νοημοσύνη είναι πιο επικίνδυνη από την κλιματική αλλαγή

Βιντεοκλήση/Συνέντευξη στην Ηλιάνα Μάγρα

Είναι η εξέλιξη της τεχνητής νοημοσύνης (AI) μεγαλύτερη απειλή για την ανθρώπινη ύπαρξη από την κλιματική αλλαγή; Αυτό πιστεύει ο Νικ Μπόστρομ, καθηγητής Φιλοσοφίας στο Πανεπιστήμιο της Οξφόρδης, όπου το 2005 ίδρυσε και έκτοτε διευθύνει το Future of Humanity Institute.

Με σπουδές Φιλοσοφίας, Μαθηματικών, Φυσικής και Υπολογιστικής Νευροεπιστήμης, ο κ. Μπόστρομ, ο οποίος έχει περισσότερες από 200 δημοσιεύσεις που έχουν μεταφραστεί σε περισσότερες από 30 γλώσσες, δύο φορές βρέθηκε στη λίστα του περιοδικού Foreign Policy με τους 100 κορυφαίους Global Thinkers, μεταξύ άλλων αντίστοιχων κατατάξεων.

Το βιβλίο του «Superintelligence: Paths, Dangers, Strategies» (μτφρ. Υπερνοημοσύνη: Μονοπάτια, Κίνδυνοι, Στρατηγικές), το οποίο εκδόθηκε το 2014, σκαρφάλωσε γρήγορα στη best-sellers λίστα των New York Times και έκανε τον καθηγητή Μπόστρομ γνωστό στο ευρύ κοινό. Το περιοδικό New Yorker έγραψε ένα εκτενές άρθρο για τον Σουηδό φιλόσοφο που προμήνυε τις υπαρξιακές προκλήσεις τις οποίες θα δημιουργήσει στην ανθρωπότητα η τεχνητή νοημοσύνη. «Το προτείνω ανεπιφύλακτα», είπε ο Μπιλ Γκέιτς για το βιβλίο. «Αξίζει να διαβαστεί», είχε δηλώσει τότε ο Ελον Μασκ, «πρέπει να είμαστε πολύ προσεκτικοί με την AI». Ο Σαμ Αλτμαν, διευθύνων σύμβουλος του OpenAI και δημιουργός του ChatGPT, έχει δηλώσει πως ο κ. Μπόστρομ έχει επηρεάσει βαθιά το έργο του, λέγοντας ότι το βιβλίο του είναι ό,τι καλύτερο έχει δει για το θέμα των κινδύνων της AI. Γι' αυτούς τους κινδύνους, μεταξύ άλλων, μιλάει ο Νικ Μπόστρομ στην «Καθημερινή».

«Νομίζω πως κινούμαστε γρήγορα προς υψηλότερα επίπεδα γενικής τεχνητής νοημοσύνης», δηλώνει μέσω βιντεοκλήσης. «Αν συνεχιστεί αυτός ο ρυθμός προόδου, θα περάσουμε σύντομα σε μεταμορφωτικά επίπεδα της AI – θα πρέπει να αρχίσουμε να σκεφτόμαστε σοβαρά τι μπορεί να συμβεί αν φτάσουμε στο επίπεδο της τεχνητής νοημοσύνης που μπορεί να κάνει ό,τι και οι άνθρωποι», σημειώνει. Αν φτάσουμε εκεί, δεν θα αργήσει πολύ η ανάπτυξη της υπερνοημοσύνης – «θα έχουμε πράγματα που μπορούν να κάνουν ό,τι και οι άνθρωποι, αλλά πολύ καλύτερα, και τότε θα αντιμετωπίσουμε διαφορετικές και πολύ σημαντικές προκλήσεις», αναφέρει.

Τρεις κατηγορίες

Οι βασικές εξ αυτών χωρίζονται σε τρεις κατηγορίες, συμπληρώνει ο κ. Μπόστρομ. Πρώτον, στο πώς θα δημιουργήσουμε κλιμακούμενες μεθόδους για τον έλεγχο της AI – «ούτως ώστε να μπορούμε να καθοδηγούμε αυτά τα ισχυρά μυαλά στο να κάνουν αυτό που θέλουμε», εξηγεί.

Η δεύτερη κατηγορία αφορά τη διακυβέρνηση και συγκεκριμένα το πώς σε έναν κόσμο όπου οι άνθρωποι ενδυναμώνονται από αυτές τις ισχυρές νοητικές προσθήκες θα μπορούμε να είμαστε σίγουροι ότι θα χρησιμοποιούνται για θετικούς σκοπούς.

Η τρίτη πρόκληση έχει μέχρι στιγμής λάβει τη λιγότερη προσοχή, λέει ο ίδιος. Αφορά το ηθικό στάτους των ψηφιακών μυαλών. «Μέχρι στιγμής, οι εφαρμογές AI έχουν υπάρξει εργαλεία, όπως ένα σφυρί ή ένα κατσαβίδι», τονίζει. «Αλλά», συμπληρώνει, «κάποια στιγμή αυτά τα αυξανόμενα, περίπλοκα ψηφιακά μυαλά που χτίζουμε θα αρχίσουν πιθανόν να αποκτούν διαφορετικά επίπεδα ηθικού στάτους, που σημαίνει ότι δεν θα έχει μόνο σημασία τι κάνουν σε εμάς ή τι κάνουμε μεταξύ μας χρησιμοποιώντας τα, αλλά τι κάνουμε εμείς σε εκείνα, αυτό θα αποτελέσει σημαντική ηθική πρόκληση». Υπάρχουν ηθικές κατευθυντήριες γραμμές για τα ιατρικά πειράματα σε ποντίκια, λέει. «Αν έχεις συστήματα τεχνητής νοημοσύνης που διαθέτουν την ίδια γνωστική αντίληψη με ένα ποντίκι, είναι πιθανό ότι θα έχουν παρόμοιο ηθικό στάτους, και αν η γνωστική τους αντίληψη είναι μεγαλύτερη, τότε μπορεί να αλλάζει και το στάτους τους», εξηγεί. «Αν έχουν επαρκώς περίπλοκη αντίληψη εαυτού, που επιμένει στον χρόνο· αν έχουν προτιμήσεις και την ικανότητα να ανταποκριθούν στη λογική, να έχουν αμοιβαίες σχέσεις με ανθρώπους, τότε θα έχουν και ηθικό στάτους», τονίζει ο κ. Μπόστρομ – μια κατάσταση που θα μπορούσε να προκύψει ακόμη και με ένα εξελιγμένο ChatGPT.

Φυσικά, άλλοι ηθικοί κανόνες μπορεί να ισχύουν για την AI από ό,τι για εμάς, καθώς οι ανάγκες μας είναι διαφορετικές. Ενας άνθρωπος πρέπει να φάει, παραδείγματος χάριν. «Η AI μπορεί να χρειάζεται συνεχή παροχή ηλεκτρικής ενέργειας», αναφέρει. «Θα πρέπει να ξανασκεφτούμε τον κόσμο εκ βάθρων όταν θα αρχίσουμε να συνυπάρχουμε με αυτά τα εξελιγμένα ψηφιακά μυαλά».

Μπορούμε να πούμε με μεγαλύτερη βεβαιότητα πως η κλιματική αλλαγή θα συμβεί γιατί ήδη συμβαίνει, αλλά στα περισσότερα σενάρια δεν θα προκαλέσει την εξάλειψη του ανθρώπινου είδους, δηλώνει. «Προφανώς θα έχει δυσμενείς επιπτώσεις σε πολλά οικοσυστήματα και σε πολλούς ανθρώπινους πληθυσμούς», λέει, «αλλά αν μιλάμε για σενάρια πραγματικής εξαφάνισης του ανθρώπινου είδους, πιστεύω ότι είναι πιο πιθανό να προκύψει από κακώς χρησιμοποιημένη υπερνοημοσύνη».

Τι μπορεί να πάει λάθος

Ο Νικ Μπόστρομ δεν εννοεί ανθρωπόμορφα ρομπότ που θα καταλάβουν τον κόσμο. Μπορεί να δημιουργήσουμε καινούργια βιολογικά όπλα ή άλλα μέσα τόσο ισχυρά, με τα οποία θα καταστρέψουμε το είδος μας. Μπορεί, με τη βοήθεια της AI, να αναπτύξουμε τρομερά εξελιγμένες μεθόδους παρακολούθησης, λογοκρισίας και χειραγώγησης που θα επιτρέψουν σε απολυταρχικά καθεστώτα να αναπτύξουν αέναη ανοσία σε προσπάθειες ανατροπής τους. «Χώρες που δεν είναι τώρα απολυταρχικές μπορεί να γίνουν με μια αλλαγή στις υποδομές ελέγχου», συμπληρώνει.

Οι πιο καταστροφικές επιπτώσεις της κλιματικής αλλαγής θα πραγματωθούν σε μερικά χρόνια, ίσως και σε έναν αιώνα, λέει. «Με την AI μπορεί να γίνει πολύ πιο σύντομα», σημειώνει, τονίζοντας πως αυτό δεν σημαίνει ότι δεν πρέπει να ασχοληθούμε και με τις δύο προκλήσεις ταυτόχρονα. Υπάρχει φυσικά και η ειδοποιός διαφορά μεταξύ τους. Η κλιματική αλλαγή δεν έχει πλεονεκτήματα – όσο λιγότερο προχωρήσει τόσο το καλύτερο. «Με την AI δεν είναι έτσι, δεν είναι κάτι αρνητικό που πρέπει να αποφύγουμε – είναι κάτι πολύ σημαντικό που μπορεί να λειτουργήσει θετικά ή αρνητικά, αναλόγως του πώς θα χρησιμοποιηθεί».

Στο Future of Humanity Institute, στο Πανεπιστήμιο της Οξφόρδης, προσπαθούν να απαντήσουν σε ερωτήματα που μέχρι πρόσφατα είχαν ως επί το πλείστον αγνοηθεί ακαδημαϊκά. «Προσπαθούμε να σκεφτούμε προσεκτικά τα ερωτήματα της “μεγάλης εικόνας” για το μέλλον του ανθρώπινου πολιτισμού, σε σχέση με πράγματα που μπορούν να αλλάξουν εις βάθος την ανθρώπινη κατάσταση», εξηγεί ο κ. Μπόστρομ στην «Κ». Ασχολούνται πολύ με την AI και τη σχέση της με τη διακυβέρνηση και την ασφάλεια, αλλά όχι μόνο. Δουλεύουν πάνω στη βιοτεχνολογία, στα ηθικά ερωτήματα που προκύπτουν από τις πιθανές αλλαγές στην ανθρώπινη φύση μέσω βιολογικής αναβάθμισης, κάνουν αναλύσεις ρίσκου που αφορούν διαφορετικούς τομείς, ερευνώντας πώς θα μπορούσαν θεσμοί και οργανισμοί να ισχυροποιηθούν απέναντι σε διαφορετικών ειδών αλλαγές. «Εχουμε σκεφτεί λιγάκι και το πολύ μακρινό μέλλον της τεχνολογικής ωριμότητας του πολιτισμού – πόσο μεγάλος θα ήταν, τι τεχνολογίες θα είχε», αναφέρει.

Η προσομοίωση

Ο 50χρονος Νικ Μπόστρομ έγινε αρχικά γνωστός διεθνώς το 2003, όταν δημοσίευσε το επιχείρημα για την προσομοίωση, σύμφωνα με το οποίο μία εκ τριών πιθανοτήτων ισχύει: Είτε, πρώτον, ότι ο πολιτισμός θα αφανιστεί πριν φτάσει στην τεχνολογική ωριμότητα· είτε, δεύτερον, πως αν φτάσουμε σε τεχνολογική ωριμότητα, οι μελλοντικοί πολιτισμοί θα έχουν χάσει κάθε ενδιαφέρον δημιουργίας προσομοιώσεων των προγόνων τους· είτε, τρίτον, ότι αν τίποτα από τα παραπάνω δεν ισχύει, ζούμε αυτή τη στιγμή σε προσομοίωση.

«Ακόμη και αν είμαστε σε προσομοίωση, οι εμπειρίες μας δεν παύουν να είναι αληθινές, η πραγματικότητά μας συνεχίζει να είναι κάτι που πρέπει να αντιμετωπίσουμε, συνεχίζουμε να πρέπει να φάμε για να μην αρρωστήσουμε, παραδείγματος χάριν», σημειώνει. «Το μόνο που αλλάζει όλο αυτό είναι ότι ίσως μας κάνει πιο ταπεινούς όσον αφορά το τι πραγματικά καταλαβαίνουμε για το πολύ μεγάλο πλαίσιο στο οποίο βρισκόμαστε», τονίζει.

Δεν ξέρουμε, δηλώνει, τις λεπτομέρειες της Φυσικής. Ξέρουμε πώς έγινε η Μεγάλη Εκρηξη, ότι η εξέλιξη ξεκίνησε πριν από μισό δισεκατομμύριο χρόνια. «Νομίζω ότι αν το επιχείρημα της προσομοίωσης ισχύει, είναι πιθανό ότι αυτό που βλέπουμε είναι ένα μικρό κομμάτι μιας πολύ μεγαλύτερης πραγματικότητας και έχουμε πολύ μικρότερη ιδέα του πώς μοιάζει αυτή η πραγματικότητα».

Οσον αφορά το γιατί θα μπορούσε να έχει δημιουργηθεί η προσομοίωση, το επιχείρημα το οποίο είχε υποστηρίξει χρησιμοποιώντας κλάσματα και πιθανότητες δεν περιέλαβε τέτοιες εικασίες. «Θα μπορούσε να είναι ανάλογο με κάποιες θρησκευτικές αντιλήψεις του κόσμου, κατά τις οποίες η πραγματικότητά μας θα είχε δημιουργηθεί από μια πολύ ανώτερη νοημοσύνη», σημειώνει ο κ. Μπόστρομ. Θεωρητικά θα μπορούσε να υπάρχει ένας ή περισσότεροι δημιουργοί, κάποιοι να έχουν φτιάξει προσομοιώσεις, άλλοι να είναι μέσα σε αυτές, συμπληρώνει.

Επομένως, ο καθηγητής Μπόστρομ δεν πιστεύει στον Θεό; «Δεν θα το έλεγα αυτό», απαντάει, και για πρώτη φορά κατά τη διάρκεια της συνέντευξης κομπιάζει. «Θα μπορούσε να υπάρχει κάποια ανώτερη δύναμη στον κόσμο και όσο περισσότερο σκέφτεται κανείς αυτές τις μεγάλες ερωτήσεις τόσο μικρότερος νιώθει και τόσο μεγαλύτερη γίνεται ίσως η ανάγκη για μια χείρα βοηθείας σε όσα έχουμε να αντιμετωπίσουμε».

Η προσαρμογή

Οσον αφορά τις προκλήσεις της τεχνητής νοημοσύνης, είμαστε ήδη πίσω. «Και θα συνεχίσουμε να είμαστε πίσω γιατί όταν σε δύο χρόνια φτάσουμε να αντιμετωπίσουμε τις

προκλήσεις της σημερινής AI, εκείνη θα έχει εξελιχθεί περισσότερο», τονίζει. Η συμβουλή του στις νεότερες γενιές είναι να εξοικειωθούν με την τεχνητή νοημοσύνη. Ο κόσμος θα είναι πολύ διαφορετικός όταν κάποιος που είναι τώρα μαθητής γίνει μεσήλικος, λέει. Αλλά αν οι νέοι εξοικειωθούν από νωρίς με την AI, μπορεί να γίνουν πιο ευπροσάρμοστοι στις καινούργιες ικανότητές της, όταν αυτές θα έρθουν.

Το πιο σημαντικό είναι όλοι μας να συνειδητοποιήσουμε ποια στιγμή στον χρόνο ζούμε, τονίζει. «Αν αυτή η εικόνα του κόσμου είναι σωστή, είναι εντυπωσιακό ότι τυχαίνει να υπάρχουμε πολύ κοντά σε ένα σημείο καμπής για όλη την ανθρώπινη Ιστορία», αναφέρει. «Οι περισσότεροι άνθρωποι έχουν ζήσει είτε πολύ πιο νωρίς ή θα ζήσουν πολύ αργότερα – αυτή τη στιγμή μπορεί να είμαστε πολύ κοντά στο σημείο που θα διαμορφώσει το τι δυνητικά θα γίνει σε εκατομμύρια χρόνια από τώρα», λέει. Θα ήταν κρίμα να υπνοβατούμε.

Είναι αισιόδοξος για το μέλλον του ανθρώπινου είδους; «Είμαι αγχωμένα αισιόδοξος», λέει γελώντας. «Είναι συναρπαστικό και έχω ελπίδες, αλλά όχι ανέμελα, γιατί δεν μπορώ να πω με αυτοπεποίθηση πως θα πάει καλά», σημειώνει. Είναι σίγουρος πως πολλά θα αλλάξουν. Δεν είναι σίγουρος αν θα είναι για καλό. «Πρέπει να ελπίζουμε πως θα είναι για καλό. Κι αν ισχύει», καταλήγει στην «Κ», «τότε θα πάει πραγματικά πολύ καλά».

Πηγή: Εφημερίδα Καθημερινή, 07.11.2023