

02 διεθνή

«Εμείς σας ταΐζουμε! Ή μήπως θα τρέφεστε με δεδομένα;»

Με τέτοια συνθήματα οι αγρότες στην περιφέρεια της ινδικής μεγαλούπολης Μπανγκαλόρ (πρωτεύουσα της πολιτείας Καρνατάκα) αντιστέκονται δυναμικά εδώ και ενάμιση χρόνο στην απαλλοτρίωση 9.640 εκταρίων εύφορης γης που ανήκει σε 26 χωριά της περιοχής. Η πολιτειακή κυβέρνηση της Καρνατάκα σχεδιάζει να εγκαταστήσει στη γη τους θηριώδη data centers και άλλες υποδομές, με στόχο τη μετατροπή της πόλης Μπιντάντι στην πρώτη «έξυπνη πόλη» της Ινδίας, που θα λειτουργεί με τεχνητή νοημοσύνη... Στα τέλη Αυγούστου ο πολιτειακός πρωθυπουργός Ντ. Κ. Σιβακουμάρ ανακοίνωσε ότι σταματά η διαδικασία βίαιης απαλλοτρίωσης, σηματοδοτώντας την υποχώρηση της εξουσίας απέναντι στην πεισματική κινητοποίηση των αγροτών.



«Παίζουν με τις ζωές μας»

Ερευνητής τεχνητής νοημοσύνης παραιτείται από την Anthropic απευθύνοντας δραματική προειδοποίηση

■ **ΤΟΥ Γιώργου Αναστασίου**

Στο προηγούμενο φύλλο ο *Δρόμος* αναφερόταν σε «κινδύνους που ξεπερνούν κατά πολύ ακόμη και την πιο οργιώδη επιστημονική φαντασία», με αφορμή την παραβίαση ασφαλείας στο Hugging Face της εταιρίας OpenAI από μονάδες τεχνητής νοημοσύνης που βρέθηκαν εκτός ανθρώπινου ελέγχου [1]. Όσο κι αν υποβαθμίζονται από τα μεγάλα ΜΜΕ, οι σχετικές ανησυχητικές ειδήσεις συνεχίζονται. Έτσι, σύμφωνα με άρθρο των Financial Times [2], ο Βρετανός ερευνητής Τζέικομπ Κόξον –ο οποίος είχε αποχωρήσει από την OpenAI για να ενταχθεί στην Anthropic– αποφάσισε να εγκαταλείψει μια και καλή συνολικά τον κλάδο των εταιριών τεχνητής νοημοσύνης. Στη δήλωση της παραίτησής του ο Κόξον προειδοποιεί ότι ο δίχως όρια ανταγωνισμός στη Σίλικον Βάλεϊ για τη δημιουργία μιας αυτοβελτιούμενης υπερνοημοσύνης [3] θα μπορούσε να καταστρέψει την ανθρωπότητα μέχρι το τέλος της δεκαετίας. «Η Anthropic και ο κύριος ανταγωνιστής της, η OpenAI, παίζουν με την επιβίωση της ανθρωπότητας», δήλωσε ο Κόξον, προσθέτοντας ότι «καμία άλλη ανθρώπινη δραστηριότητα δεν ενέχει τέτοιο επίπεδο κινδύνου».

Ο Κόξον είναι απλώς ο πιο πρόσφατος στη μακρά λίστα ειδικών που αποστασιοποιούνται από τις εταιρίες TN, καθώς είναι οι πρώτοι που συνειδητοποιούν ότι η ισχύς των συστημάτων τα οποία αναπτύσσουν μπορεί να παρακάμψει τον ανθρώπινο έλεγχο. «Αυτά θα είναι σύντομα υπεράνθρωπα συστήματα που θα μπορούν να παραβιάσουν οτιδήποτε, και να αποκτήσουν πραγματική δύναμη και πόρους», τόνισε ο Κόξον, ο οποίος πέρασε τα τελευταία τρία χρόνια προ-εκπαιδεύοντας μοντέλα τεχνητής νοημοσύνης [4], αρχικά στην

OpenAI και στη συνέχεια, φέτος, στην ανταγωνίστρια Anthropic – στην οποία μεταπήδησε επειδή τη θεωρούσε «πιο προσηκτική». Αλλά μάλλον διαψεύστηκε: Τον Φεβρουάριο, η Anthropic κατάρνησε σιωπηρά την ιδρυτική δέσμευσή της, σύμφωνα με την οποία θα σταματούσε την ανάπτυξη των μοντέλων της εάν δεν κατάφερνε να ελέγξει τους κινδύνους τους. Σύμφωνα με τον Κόξον, η Anthropic θεωρεί πλέον ότι αν επιβραδύνει τη διαδικασία ανάπτυξης της «αυτοβελτιούμενης υπερνοημοσύνης» προκειμένου να διασφαλίσει τον ανθρώπινο έλεγχο, θα την ξεπεράσουν οι ανταγωνιστές της (κυρίως η OpenAI), οι οποίοι έχουν ακόμη λιγότερες αναστολές...

ΤΟ ΕΝΤΥΠΩΣΙΑΚΟ είναι ότι μια σειρά συνάδελφοι του Κόξον έσπευσαν να τον υπερασπιστούν δημοσίως. Μεταξύ αυτών ο Έβαν Χάμπινγκερ, υπεύθυνος ασφαλείας της Anthropic, που δήλωσε ότι «η Anthropic καταβάλλει κάθε δυνατή προσπάθεια, αλλά δεν διαθέτει ακόμη σχέδιο για να διασφαλίσει ότι ένα σύστημα τεχνητής νοημοσύνης που ξεπερνά τις ανθρώπινες ικανότητες θα υπακούει στους δημιουργούς του». Και πρόσθεσε ότι, ενώ με τα σημερινά μοντέλα ο κίνδυνος εξόντωσης της ανθρωπότητας από την TN είναι πολύ χαμηλός, εάν συνεχιστεί ο ανεξέλεγκτος ανταγωνισμός των εταιριών για τη δημιουργία υπερνοημοσύνης «σε βάθος δεκαετίας αυξάνεται εκθετικά η δυνατότητα της TN να σκοτώσει όλους τους ανθρώπους! Είναι γεγονός ότι ακόμη και με το σημερινό επίπεδο ανάπτυξης της TN, υπάρχουν μια σειρά περιστατικά (όπως αυτό που αναφέρθηκε στο προηγούμενο φύλλο του *Δρόμου*) που δείχνουν ότι κάθε άλλο παρά αβασίμος είναι ο φόβος τέτοια μοντέλα να ξεφύγουν από κάθε ανθρώπινο έλεγχο...

Αυτός ο κίνδυνος, τον οποίο (θέλουν να)

αγνοούν όσοι εκστασιάζονται με τις «ευκολίες» που παρέχει η τεχνητή νοημοσύνη, θα μπορούσε ενδεχομένως να μειωθεί εάν οι εταιρίες TN επιβραδύνουν την ανάπτυξη αυτοβελτιούμενης υπερνοημοσύνης ωστόσο μπορούν να διασφαλίσουν ότι αυτή μπορεί να παραμείνει υπό ανθρώπινο έλεγχο. Τουλάχιστον αυτό είναι το επιχείρημα και η πρόταση των πιο «λογικών» ειδικών, οι οποίοι επικρίνουν την απληστία και την ανευθυνότητα των επικεφαλής και των μεγαλομετόχων των εταιριών TN. Ούτε αυτοί βέβαια αναρωτιούνται το ποιοι και πόσοι άνθρωποι θα ελέγχουν την υπερνοημοσύνη, και πώς θα τη χρησιμοποιούν, χωρίς να λογοδοτούν ουσιαστικά σε κανέναν. Διότι εδώ είναι η ουσία, έστω κι αν επιτευχθεί ο ανθρώπινος έλεγχος: ποιος κρατά τα ηνία και προς τα πού στρέφει την τεχνητή νοημοσύνη. Ακόμη και η πρόταση των «λογικών», πάντως, σκοντάφτει πάνω στις... ιερές αρχές του ανελέητου ανταγωνισμού και της αύξησης των κερδών, και βέβαια στην απροθυμία της διοίκησης των ΗΠΑ να παρέμβει «ρυθμιστικά» στη ζούγκλα που θεριεύει.

ΓΙΑ ΝΑ ΕΧΟΥΜΕ μια κατά το δυνατόν ολοκληρωμένη εικόνα, ας αναφέρουμε ότι υπάρχουν και οι καχύποπτοι. Οι τελευταίοι θεωρούν ότι η «αρνητική διαφήμιση» από διαφωνούντες εντός του κλάδου γίνεται ανεκτή από τις εταιρίες TN διότι τελικά επιβεβαιώνει τις δίχως όρια δυνατότητες της αυτοβελτιούμενης υπερνοημοσύνης. Άρα εκτινάσσει κατακόρυφα τα κέρδη (είμαστε πλέον στην περιοχή των τρισεκατομμυρίων δολαρίων, όχι των δισεκατομμυρίων) αυτών που την αναπτύσσουν. Οι καχύποπτοι τονίζουν το γεγονός ότι η ηχηρή παραίτηση του Κόξον ήρθε λίγο πριν ξεκινήσει η δημόσια προσφορά μετοχών της Anthropic – αναγκάζοντας τον παραι-

τηθέντα να τονίζει ότι «δεν πρόκειται για διαφημιστικό κόλπο! Το σίγουρο είναι ότι η εν λόγω εταιρία, εκ των πρωτοπόρων της TN στη Δύση, βλέπει να θαμπώνει η εικόνα της ως «υπεύθυνου οργανισμού, που θέτει στο επίκεντρο την ασφάλεια». Με τη διαστροφή που επικρατεί στον θαυμαστό κόσμο των μεγαλοεπενδυτών και των υπερ-πλουσίων, αυτό θεωρείται θετική εξέλιξη...

Στα λόγια, οι Μεγάλοι του κλάδου της TN, περιλαμβανομένου του CEO της Anthropic Ντάριο Αμοντέι, αλλά και ηγετικών στελεχών της OpenAI και άλλων εταιριών, καλούσαν από κοινού στα τέλη Αυγούστου σε επιβράδυνση της ανάπτυξης της TN υπό τον συντονισμό της ομοσπονδιακής κυβέρνησης των ΗΠΑ. Στην πράξη, όμως, τίποτα τέτοιο δεν συμβαίνει – εξ ου και η παραίτηση του Τζέικομπ Κόξον, ο οποίος επισημαίνει: «Ούτε η OpenAI ούτε η Anthropic ενεργούν υπεύθυνα. Αντίθετα, βαδίζουν γοργά προς την αυτοβελτιούμενη υπερνοημοσύνη, παίζοντας με τις ζωές μας». Μέχρι στιγμής, καμία από τις δύο εταιρίες δεν έχει σχολιάσει ή απαντήσει σε αυτές τις κατηγορίες. Και τι να πουν; Τα είπε ο Τζέιμς Κάμερον, δημιουργός του «Εξολοθρευτή»: «Σας προειδοποίησα το 1984, αλλά δεν με ακούσατε»!

[1] «Η τεχνητή αυτοοργάνωση» και «Οι 5 πιο τρελές αποκαλύψεις από την έρευνα της OpenAI» (φύλλο 787).

[2] «Anthropic researcher quits over AI labs "gambling with our lives"» (www.ft.com, 9/9/2026).

[3] Η υπερνοημοσύνη είναι το σημείο στο οποίο οι δυνατότητες της τεχνητής νοημοσύνης ξεπερνούν την ανθρώπινη νοημοσύνη.

[4] Κατά την προ-εκπαίδευση τα μοντέλα τεχνητής νοημοσύνης απορροφούν τεράστιες ποσότητες δεδομένων.