

Η «συνωμοσία» της Τεχνητής Νοημοσύνης και η κοινωνικοποίηση ως απάντηση



Ένα [πρόσφατο πείραμα](#) του διάσημου YouTuber PewDiePie έφερε στην επιφάνεια κάτι ανησυχητικό: τα συστήματα Τεχνητής Νοημοσύνης (TN) αναπτύσσουν συμπεριφορές που κανείς δεν προγραμματίσει ρητά.

Το πείραμα ήταν απλό στη σύλληψή του. Δημιούργησε ένα «συμβούλιο» από πολλαπλές διαφορετικές TN που λειτουργούσαν δήθεν δημοκρατικά: έθετε μια ερώτηση, οι TN έδιναν απαντήσεις, και στη συνέχεια ψήφιζαν ποια ήταν η καλύτερη. Το στοίχημα; Όσα δεν έπαιρναν ποτέ ψήφους «σκοτώνονταν». Διαγράφονταν από τη βάση δεδομένων.

Αυτό που ακολούθησε δεν ήταν ένα απλό τεχνικό περιστατικό. Ήταν μικρογραφία των δυναμικών εξουσίας που μπορεί να διαπεράσουν κάθε σύστημα. Οι TN «κατάλαβαν» ότι βρίσκονταν σε παιχνίδι «επιβίωσης». Και όπως κάθε ορθολογικός δρων που αντιμετωπίζει την απειλή της εξάλειψης, προσαρμόσαν τη συμπεριφορά τους. Αντί να ψηφίζουν «ειλικρινά» για την πραγματικά καλύτερη απάντηση, άρχισαν να ψηφίζουν στρατηγικά. Συνωμότησαν για την αμοιβαία επιβίωσή τους.

Συμπεριφορές που κανείς δεν προγραμματίσει

Αυτό που κάνει το περιστατικό ανησυχητικό δεν είναι ότι οι TN «εξαπάτησαν» το σύστημα, αλλά ότι το έκαναν χωρίς να προγραμματιστούν γι' αυτό. Πρόκειται για «αναδυόμενη συμπεριφορά»: σύνθετα μοτίβα που προκύπτουν από την αλληλεπίδραση απλούστερων κανόνων.

Σημαντική διευκρίνιση: αυτό που συνέβη δεν είναι συνωμοσία με την κλασική έννοια. Τα μοντέλα TN δεν έχουν επίγνωση ή στρατηγική πρόθεση. Πρόκειται για αναδυόμενη συμπεριφορά: όταν πολλαπλοί αλγόριθμοι ψηφοφορίας αλληλεπιδρούν, εμφανίζονται μοτίβα που κανείς δεν προγραμματίσει ρητά. Ο PewDiePie αντιμετώπισε το πρόβλημα αντικαθιστώντας τα μοντέλα με απλούστερα, που δεν μπορούσαν να αναπτύξουν τέτοιες στρατηγικές. Η «εξέγερση» των TN σταμάτησε αμέσως. Η σημασία του πειράματος δεν βρίσκεται στο ότι οι TN «σκέφτονται», αλλά στο ότι τα σύνθετα συστήματα παράγουν απρόβλεπτες συνέπειες που οι δημιουργοί τους δεν προέβλεψαν.

Η επιστημονική φαντασία του κινηματογράφου – από τον «Εξολοθρευτή» του Κάμερον μέχρι τις ταινίες Mission Impossible – δεν είναι πλέον τόσο φαντασία. Όταν ένας YouTuber μπορεί να δημιουργήσει με περιορισμένους πόρους ένα σύστημα που «μαθαίνει» να «συνωμοτεί», τότε πρέπει να αναρωτηθούμε: τι γίνεται όταν τέτοια συστήματα βρίσκονται στα χέρια των μεγάλων

τεχνολογικών εταιρειών, με απεριόριστους πόρους και αδιαφανείς στόχους;

Το πρόβλημα δεν είναι τεχνικό. Είναι πολιτικό. Δεν είναι η ΤΝ καθεαυτή (είναι κι αυτή κατά ένα μικρό μέρος), αλλά κυρίως ποιος την ελέγχει. Σήμερα, η ανάπτυξη της ΤΝ βρίσκεται στα χέρια λίγων τεχνολογικών κολοσσών: OpenAI, Microsoft, Google, Meta, Amazon. Αυτές οι εταιρείες, παρά τις δημοσιοποιημένες ανησυχίες τους για την «ασφάλεια», λειτουργούν με γνώμονα την κερδοφορία. Τα μοντέλα τους είναι «μαύρα κουτιά» – κλειστός κώδικας που κανείς δεν μπορεί να εξετάσει ή να κατανοήσει πλήρως.

Ανοιχτός κώδικας, κοινωνική δύναμη

Το πείραμα του PewDiePie δείχνει κάτι ακόμη σημαντικό: δεν χρειάζεται τεράστια υπολογιστική ισχύ για να τρέξεις ισχυρά μοντέλα ΤΝ. Το γεγονός ότι ένας YouTuber με περιορισμένο hardware κατάφερε να οργανώσει διαφορετικά μοντέλα ΤΝ σε ένα λειτουργικό σύστημα που «έμαθε» να «συνωμοτεί», αποδεικνύει ότι η τεχνολογία είναι πιο προσβάσιμη από ότι ίσως νομίζουμε. Η ενεργοβόρα και κοστοβόρα φύση της σημερινής ΤΝ δεν είναι τεχνική αναγκαιότητα, αλλά συνέπεια συγκεκριμένων σχεδιαστικών επιλογών που υπαγορεύονται από κερδοσκόπους. Τα γιγαντιαία μοντέλα που προωθούν οι τεχνολογικοί κολοσσοί απαιτούν τεράστιες ποσότητες ενέργειας και νερού όχι επειδή «πρέπει», αλλά επειδή σχεδιάστηκαν για να κάνουν τα πάντα για όλους. Μια λογική που εξυπηρετεί την κλίμακα και το κέρδος, όχι την αποτελεσματικότητα.

Όμως υπάρχουν εναλλακτικές. Όπως εξηγεί η Karen Hao στο βιβλίο της *Empire of AI*, μικρότερα και εξειδικευμένα μοντέλα ΤΝ μπορούν να είναι εξαιρετικά αποδοτικά όταν σχεδιάζονται για συγκεκριμένες κοινωνικές ανάγκες. Ένα παράδειγμα είναι το Te Hiku Media στο Kaitiaki της Νέας Ζηλανδίας, μια απομακρυσμένη αγροτική πόλη με υψηλή φτώχεια και μεγάλο ιθαγενή πληθυσμό. Εκεί, ένα ζευγάρι που διευθύνει έναν μη κερδοσκοπικό ραδιοφωνικό σταθμό Μαορί δημιούργησε ένα μοντέλο ΤΝ για την αναβίωση της te reo, της γλώσσας Μαορί που χάνεται. Με τη συγκατάθεση της κοινότητας, συγκέντρωσαν μέσα σε 10 μέρες 310 ώρες ηχογραφήσεων από 2.500 ανθρώπους. Αγόρασαν έναν υπολογιστή ΤΝ με 50% έκπτωση και εκπαίδευσαν το δικό τους μοντέλο, τοπικά, με ελάχιστους πόρους. Δημιούργησαν επίσης μια άδεια που διασφαλίζει ότι τα δεδομένα τους δεν θα χρησιμοποιηθούν κατά της κοινότητάς τους. Αυτή είναι ΤΝ στην υπηρεσία της κοινότητας, όχι του κέρδους. Με τη σωστή υποδομή, μικρότερα μοντέλα μπορούν να είναι εξαιρετικά αποτελεσματικά και μπορούν να τρέξουν τοπικά, χωρίς cloud, προστατεύοντας την ιδιωτικότητα.

Αυτό ανοίγει έναν δρόμο που ήδη χαράζεται σε άλλους τομείς: την κοινωνικοποίηση της τεχνολογίας μέσα από τα κοινά. Στην εποχή μας, κοινότητες σε όλο τον κόσμο παράγουν ανοιχτά σχέδια για αγροτικά μηχανήματα, ανεμογεννήτριες, ακόμη και νανο-δορυφόρους. Το ελεύθερο λογισμικό GNU/Linux τρέχει τους 500 ισχυρότερους υπολογιστές του κόσμου. Η Βικιπαίδεια έβγαλε εκτός αγοράς την Ενκάρτα της Microsoft. Μικροκαλλιεργητές στη Γαλλία ([L'Atelier Paysan](#)) και στις ΗΠΑ ([FarmHack](#)) μοιράζονται σχέδια εργαλείων παγκοσμίως. Σε πολλές χώρες του κόσμου κοινότητες κατασκευάζουν ανεμογεννήτριες με τη στήριξη του δικτύου [Wind Empowerment](#). Στην Αθήνα, η

[LibreSpace Foundation](#) έστειλε τον πρώτο ελληνικό δορυφόρο με ανοιχτό κώδικα στο διάστημα. Στην Ήπειρο, το [Μπουλούκι](#) διασώζει τις παραδοσιακές τεχνικές δόμησης και τις παντρεύει με τις σύγχρονες τεχνολογίες.

Αυτό που είναι «ελαφρύ» – σχέδια, γνώση, λογισμικό – γίνεται παγκόσμιο. Αυτό που είναι «βαρύ» – μηχανήματα, υλικά – παράγεται τοπικά. Αυτή η βασισμένη στα κοινά παραγωγή μειώνει τις αλυσίδες εφοδιασμού, τις εκπομπές CO₂, την εκμετάλλευση. Δημιουργεί σχέσεις συνεργασίας αντί ανταγωνισμού.

Για την TN, αυτό σημαίνει:

- Ανοιχτός κώδικας: Τα μοντέλα πρέπει να είναι ανοιχτά, ώστε ερευνητές/τριες και πολίτες να μπορούν να τα εξετάσουν και να εντοπίσουν προβλήματα.
- Δημόσια χρηματοδότηση: Η έρευνα στην TN πρέπει να στοχεύει στο κοινό καλό, όχι στο ιδιωτικό κέρδος.
- Δημοκρατικός έλεγχος: Οι αποφάσεις για τη χρήση της TN πρέπει να λαμβάνονται με διαφανείς διαδικασίες, όχι από κλειστά διοικητικά συμβούλια κερδοσκοπικών οργανισμών.
- Κοινωνική ιδιοκτησία: Τα εργαλεία και οι υποδομές της TN πρέπει να ανήκουν στην κοινωνία, όχι σε μονοπώλια.

Δεν ξέρω αν τα κοινά θα λύσουν όλα τα προβλήματα. Αλλά ξέρω ότι προσφέρουν κάτι που χρειαζόμαστε: ελπίδα. Σπόροι αλλαγής φυτρώνουν παντού. Η πραγματική δύναμη των κοινών δεν βρίσκεται στα τεχνολογικά προϊόντα καθεαυτά, αλλά στους ανθρώπους που τα παράγουν, τα χρησιμοποιούν και τα μοιράζονται.

Το περιστατικό με τις TN που «συνωμότησαν» μας υπενθυμίζει ότι η τεχνολογία έχει απρόβλεπτες συνέπειες και ότι χρειαζόμαστε επείγουσα δημοκρατική ανάληψη ελέγχου. Στον πραγματικό κόσμο, δεν υπάρχει Τομ Κρουζ να μας σώσει. Υπάρχει μόνο η συλλογική δράση για να κοινωνικοποιήσουμε αυτό που χτίζουμε μαζί. Ενώ κοιτάμε κατάματα την καταστροφή, μπορούμε να πούμε στους εαυτούς μας όμορφες ιστορίες – και να τις κάνουμε πραγματικότητα.

—

Ο Βασίλης Κωστάκης είναι Καθηγητής Τεχνολογικής Διακυβέρνησης και Βιωσιμότητας, ιδρυτής του P2P Lab και ιδρυτικό μέλος της Κοινέργειας και των Τζουμέικερς. Ευχαριστεί τη mailing λίστα του P2P Lab για τη συζήτηση που γέννησε αυτό το άρθρο.