

Ρομπότ, φιλοσοφία και ηθική: Γίνεται ο άνθρωπος "hackable" είδος;



«Το σημαντικότερο γεγονός, που οποιοσδήποτε ζει σήμερα χρειάζεται να γνωρίζει για τον 21ο αιώνα, είναι ότι (οι άνθρωποι) γινόμαστε ζώα που μπορούν να πέσουν θύματα hacking (hackable animals). Όταν συνδυάζεις την αυξανόμενη κατανόησή μας για τη βιολογία -ιδιαίτερα για τις επιστήμες του εγκεφάλου- με την τεράστια υπολογιστική ισχύ, που μας προσφέρουν η μηχανική μάθηση και η Τεχνητή Νοημοσύνη, αυτό που λαμβάνεις είναι η δυνατότητα να χακάρεις τους ανθρώπους, κάτι που σημαίνει να προβλέπεις τις επιλογές τους, να κατανοείς τα συναισθήματά τους, να τους χειραγωγείς και επίσης να τους αντικαθιστάς. Αν μπορείς να χακάρεις κάτι, τότε μπορείς και να το αντικαταστήσεις»: η φράση αυτή του ιστορικού και συγγραφέα Γιουβάλ Νώε Χαράρι* δίνει το στίγμα μιας συζήτησης, που έχει ενταθεί τα τελευταία χρόνια γύρω από την αλματώδη ανάπτυξη της τεχνολογίας και της Τεχνητής Νοημοσύνης και ιδίως την επίδρασή τους στην πορεία της ανθρωπότητας.

Η εξατομίκευση περιεχομένου, αλλά και τα «filter bubbles»***, αποτελούν ήδη καθημερινές εφαρμογές της Ειδικής Τεχνητής Νοημοσύνης**, η οποία μπορεί μεν να προγραμματίζεται από άνθρωπο, αλλά μέσω της μηχανικής μάθησης αποκτά τη δυνατότητα να μαθαίνει μόνη της από τις εμπειρίες της. Το δε ερώτημα, αν η Τεχνητή Νοημοσύνη (TN) θα είναι -κάποια στιγμή στο μακρινό μέλλον- σε θέση να σκεφτεί, με την έννοια της συνειδητής διεργασίας που αποδίδουμε στη σκέψη, συζητιέται πλέον ευρέως όχι μόνο μεταξύ σεναριογράφων ταινιών επιστημονικής φαντασίας, αλλά και ανάμεσα σε εκπροσώπους όλων των επιστημών, από την πληροφορική μέχρι την κοινωνιολογία, την ηθική ψυχολογία και τη νομική.

Παρότι δεν υπάρχουν ξεκάθαρες απαντήσεις, ως προς το πότε η ανθρωπότητα θα μπορούσε ενδεχομένως να πετύχει τη Γενική Τεχνητή Νοημοσύνη** (αν την πετύχει κάποτε), ωστόσο ζητήματα γύρω από τον έλεγχο και την ιδιοκτησία των μηχανών και των προσωπικών δεδομένων των ανθρώπων, που χρησιμοποιούνται -μεταξύ άλλων- για την εκπαίδευση της ΤΝ, ανοίγουν ήδη νέα πεδία συζήτησης.

Το ΑΠΕ-ΜΠΕ συνομίλησε για το θέμα με τον καθηγητή Φιλοσοφίας και Ηθικής του Τεχνικού Πανεπιστημίου του Eindhoven (Τμήμα Σπουδών Καινοτομίας), Βίνσεντ Μίλερ (Vincent C. Müller).

Το σκάκι, το ποδόσφαιρο, ο Χατζηδάκις

Θα μπορούν κάποτε να σκεφτούν οι μηχανές; «Θεωρητικά δεν υπάρχει εμπόδιο για να γίνει κάτι τέτοιο» λέει ο κ.Μίλερ, στη συνέντευξη που έδωσε στο ΑΠΕ-ΜΠΕ σε άπταιστα ελληνικά, εξηγώντας πως η λειτουργία της ανθρώπινης σκέψης βασίζεται σε φυσικά όργανα, όχι σε κάτι μεταφυσικό και, άρα, από τη στιγμή που καταλαβαίνουμε πώς λειτουργεί η διαδικασία της σκέψης, ίσως να είναι κάποια μέρα εφικτό να μεταφερθεί αυτή η λειτουργία στην ΤΝ. Πρακτικά ωστόσο, το ερώτημα είναι αν πρόκειται να αναπτυχθεί επαρκώς η τεχνολογία, για να πετύχουμε μηχανές που σκέφτονται. «Προς το παρόν, με την ειδική ΤΝ, μπορούμε να αναπαράγουμε τεχνητά πολλές λειτουργίες, αρκεί να γνωρίζουμε πώς λειτουργούν τα αντίστοιχα φυσικά συστήματα. Και τότε μπορούμε να βρούμε μέσω της ΤΝ ακόμα και εντελώς νέες λύσεις. Για παράδειγμα, ο άνθρωπος Χατζηδάκις έγραψε υπέροχη μουσική, αλλά ένα σύστημα ΤΝ θα μπορούσε να παράγει ένα καινούργιο μουσικό κομμάτι, που πιθανότατα δεν θα είναι ισάξιο, αλλά θα μοιάζει πολύ με Χατζηδάκι μιας συγκεκριμένης περιόδου, αν εισάγουμε σε αυτό δεδομένα για τη συγκεκριμένη μουσική περίοδο του συνθέτη. Θεωρούμε ότι πολλά πράγματα που κάνουμε οι άνθρωποι είναι πολύ πολύπλοκα για τις μηχανές, αλλά στην πραγματικότητα πράγματα πολύπλοκα για εμάς είναι ευκολότερα για τις μηχανές και αντίστροφα. Για παράδειγμα, ένα ρομπότ με τεχνητή νοημοσύνη μπορεί να παίζει πολύ καλύτερο σκάκι από έναν άνθρωπο, αλλά να δυσκολεύεται να κλωτσήσει μια μπάλα και αυτό δεν οφείλεται στο ότι το σκάκι είναι λιγότερο πολύπλοκο» λέει.

Το κυριότερο εμπόδιο. Και η ζωή στην εξίσωση

Ποιοι είναι λοιπόν σήμερα το κυριότερο εμπόδιο για την ανάπτυξη της ΤΝ μέχρι το σημείο της Γενικής Τεχνητής Νοημοσύνης (ΓΝΤ); Κατά τον κ.Μίλερ, το σημαντικότερο εμπόδιο είναι ότι «η ΤΝ δεν έχει θέληση και δεν αντιλαμβάνεται τα πράγματα συνειδητά. Εμείς και όλα τα υπόλοιπα ζώα αντιλαμβανόμαστε τον

κόσμο με αξιολογήσεις, βάσει του τι μας αρέσει και τι όχι, τι θέλουμε και δεν θέλουμε. Επίσης, τα πράγματα που βλέπουμε έχουν για εμάς συγκεκριμένη σημασία. Αντίθετα, τα σύμβολα που εισάγονται σε ένα σύστημα TN δεν έχουν νόημα για αυτό. Αν το input (εισερχομένα δεδομένα) είναι η εικόνα ενός πιονιού, το μηχάνημα δεν το αντιλαμβάνεται ως πιόνι, αλλά ως μια μαθηματική σειρά με τα σύμβολα του δυαδικού συστήματος, 0 και 1. Παρότι δεν υπάρχει κοινά αποδεκτή απάντηση για το αν και πότε θα μπορούσαμε να έχουμε ΓNT, αν ρωτάτε τη γνώμη μου, αυτό δεν θα συμβεί σύντομα. Πιστεύω ότι, για τους λόγους που προανέφερα, τα σημερινά εμπόδια για την ανάπτυξη TN ξεπερνιούνται μόνο αν στην εξίσωση μπει και η ζωή. Η κατεύθυνση που θα μας οδηγούσε εκεί νομίζω, στη ΓTN δηλαδή, θα ήταν ενδεχομένως μηχανές που συνδέονται με βιολογικά μέρη. Τέτοιου είδους μικτά συστήματα, με μηχανικά και βιολογικά χαρακτηριστικά, τα λεγόμενα cyborgs, θα μπορούσαν ίσως, κατά τη γνώμη μου, να αποκτήσουν ΓTN. Βέβαια, αν συμβεί αυτό, η έννοια της ατομικότητας του ανθρώπου διαλύεται, το πού αρχίζει και πού τελειώνει ο άνθρωπος παύει να είναι σαφές».

“Singularity”. Μικρή πιθανότητα, τεράστιος αντίκτυπος

Κατά πολλούς επιστήμονες, το λεγόμενο “singularity”, η χρονική στιγμή δηλαδή κατά την οποία η ευφυία της TN θα ξεπεράσει την ανθρώπινη, εξακολουθεί να ανήκει στη σφαίρα της επιστημονικής φαντασίας. Κατά άλλους όμως, είναι κάτι που τελικά θα συμβεί. Και μάλιστα ενδεχομένως πριν από το τέλος του 21ου αιώνα. Τι θα σημαίνει το singularity -αν ποτέ συμβεί- για τον άνθρωπο, που ενδεχομένως θα νιώσει ότι χάνει τον ρόλο του ως το ευφυέστερο -και άρα κυρίαρχο- είδος στον πλανήτη Γη; «Ο φόβος ότι θα χάσουμε ως είδος την κυριαρχία έχει μεν κάποια βάση, αλλά κατά τη γνώμη μου όχι πολλή, γιατί η πιθανότητα να υπάρξουν πιο ευφυείς οντότητες από εμάς δεν είναι κατά τη γνώμη μου μεγάλη. Ο Homo Sapiens έχει πράγματι τον έλεγχο στον πλανήτη λόγω της υψηλότερης νοημοσύνης του, σε σχέση με τα υπόλοιπα ζώα. Δεν είμαστε πιο δυνατοί ή πιο γρήγοροι, αλλά πιο έξυπνοι. Άρα, αν κάποια στιγμή αυτή η υπεροχή μας στην ευφυία πάψει να υφίσταται, δεν θα έχουμε τον ίδιο έλεγχο. Η πιθανότητα για μια “έκρηξη” νοημοσύνης στους υπολογιστές είναι μεν μικρή, αλλά αν συνέβαινε, η επίδραση θα ήταν τεράστια, οπότε δεν μπορούμε και να την αγνοήσουμε. Ενώ οι άνθρωποι έχουμε την τάση να βλέπουμε με προσοχή τα πράγματα που έχουν μεγάλη πιθανότητα να συμβούν, ακόμα και αν η βλάβη είναι μικρή, συχνά δεν πράττουμε το αντίθετο, δηλαδή να βλέπουμε με προσοχή πράγματα που έχουν μικρή πιθανότητα να συμβούν, αλλά αν συμβούν η επίδραση θα είναι πολύ μεγάλη. Είναι αρκετά σίγουρο ότι το singularity δεν θα συμβεί τώρα. Χρειάζεται όμως να κοιτάζουμε τα συστήματά μας, για να ξέρουμε πού πηγαίνουμε ανά πάσα στιγμή» προσθέτει.

“Να σκεφτόμαστε πού δίνουμε τον έλεγχο”

Κατά τον κ.Μίλερ, παρότι αυτές οι εξελίξεις δεν μας επηρεάζουν άμεσα, το θέμα του ελέγχου στη διαχείριση των προσωπικών μας δεδομένων, που -μεταξύ άλλων- βοηθούν την ΤΝ να εκπαιδευτεί και αναπτυχθεί, μάς αφορά τώρα. Κι αν το θέμα των δεδομένων έχει πολύ μεγαλύτερες προεκτάσεις, από αυτές που σχετίζονται με την ΤΝ, «ήδη έχουμε μεταφέρει οικειοθελώς σημαντικό μέρος τους σε αμερικανικές εταιρείες, όπως οι Facebook και Google, οι οποίες αντλούν κέρδη από αυτά. Χάσαμε εν ολίγοις τον έλεγχο των δεδομένων μας, χωρίς να έχουμε καν αποφασίσει συνειδητά αν αυτή η απώλεια είναι καλή ιδέα. Αν αυτό συμβεί και σε άλλα επίπεδα, δεν θα είναι εύκολο να κάνουμε πίσω. Χρειάζεται να σκεφτόμαστε πού δίνουμε τον έλεγχο. Παρότι η οικονομία μας εξαρτάται σε μεγάλο βαθμό από τέτοια συστήματα, θα πρέπει να βλέπουμε κάθε φορά τα συν και τα πλην, να μπορούμε να πούμε “δεν θα το προχωρήσουμε άλλο, γιατί δεν θα ελέγχεται”. Απαιτείται συναίσθηση και επίγνωση» προειδοποιεί.

Ιδιοκτησία δεδομένων: “The winner takes it all” και το 1 εκατ. μολύβια

Το ποιος έχει την ιδιοκτησία των μηχανών, σε μια περίοδο που αυτές θα γίνονται ολοένα ευφυέστερες, θεωρείται κρίσιμη παράμετρος ως προς το αν η ΤΝ θα εξελιχθεί τελικά με τρόπο που θα αποβεί προς όφελος της ανθρωπότητας -ή σε βάρος της. Πόσο αισιόδοξος είναι ο κ.Μίλερ ότι θα γίνει ορθή κατανομή αυτής της ιδιοκτησίας (και εξουσίας); «Είμαι αρκετά απαισιόδοξος. Μερικοί, πολλοί λίγοι, χρησιμοποιούν ήδη τις τεχνολογίες ΤΝ για να βγάλουν κέρδος από τα δεδομένα, πλουτίζοντας απ’ τον καθένα μας. Αυτό συμβαίνει και θα συμβαίνει και στο μέλλον. Σήμερα στον κόσμο υπάρχουν πολλές τεχνολογικές αγορές του στυλ “The winner takes it all”, που παράγοντας ένα προϊόν μία φορά, όχι απλά αντλούν τα δεδομένα μας, αλλά μπορούν στη συνέχεια να αντλήσουν επαναληπτικά κέρδος από αυτό, με σχεδόν μηδενικό κόστος. Τι εννοώ... Αν θέλεις να πουλήσεις 1 εκατ. μολύβια, θα πρέπει να παράγεις ένα εκατομμύριο μολύβια, με όποιο κόστος συνεπάγεται αυτό. Αν, αντίθετα, θέλεις να πουλήσεις λογισμικό, θα το παράγεις μία φορά και θα το πουλήσεις ένα εκατομμύριο φορές. Η οικονομική ανάπτυξη που θα έχουμε από τα συστήματα ΤΝ θα είναι τεράστια και άρα είναι πολύ σημαντική η ανάγκη νομικής και πολιτικής ρύθμισης για την πιο ισομερή κατανομή του πλούτου που θα προκύψει» λέει.

‘Ανθρωπος, ένα hackable είδος;

«Στο κοντινό μέλλον θα έχουμε έτσι και αλλιώς πολύ αμεσότερη σχέση με τα μηχανήματα. Ήδη κουβαλάμε συνέχεια πάνω μας έναν υπολογιστή με οθόνη αφής,

το smartphone μας, που γίνεται ολοένα ισχυρότερο και αν το χάσουμε νιώθουμε χαμένοι. Αυτή η επέκταση του σώματός μας θα γίνεται ολοένα εντονότερη... Πόσο hackable θα είμαστε; Τα βιολογικά συστήματα είναι ιδιαίτερα πολύπλοκα και φαίνεται δύσκολο να μας χακάρει κάποιος, αλλά το τι μέλλει γενέσθαι μένει να το δούμε στην πράξη... Ως προς το hacking του μυαλού, ήδη, γίνεται σε έναν βαθμό, αφού τα συστήματα TN γνωρίζουν τις επιθυμίες μας και μας προσφέρουν αυτό που προτιμάμε μέσω content personalization και filter bubbles. Το hacking του μυαλού, με σκοπό το κέρδος, πρέπει οπωσδήποτε να το δούμε νομικά. Η χειραγώγησή μας από το κράτος απαγορεύεται, γιατί να επιτρέπεται από μια ιδιωτική εταιρεία, με σκοπό το κέρδος;» καταλήγει ο κ.Μίλερ._

Αλεξάνδρα Γούτα

***Συνέντευξη στο aljazeera, 24/8/2018**

**Η ειδική ή "στενή" TN έχει προγραμματιστεί να εκτελεί μία μόνο εργασία (π.χ, να παίζει σκάκι ή να εξατομικεύει περιεχόμενο). Η γενική ή "ισχυρή" TN είναι τόσο ανεπτυγμένη, όσο και η ανθρώπινη και μπορεί να εκτελεί με επιτυχία όλες τις πνευματικές εργασίες που ένας άνθρωπος μπορεί.

***Φίλτρα που όταν αναζητάς κάτι στο Διαδίκτυο σου επιστρέφουν κυρίως αποτελέσματα προσαρμοσμένα στο ιστορικό των προηγούμενων προτιμήσεών σου, περιορίζοντας έτσι την πρόβαση σε πιο αντικειμενική πληροφόρηση