Άνθρωποί και μηχανές I. II

Media: THE BOOKS JOURNAL Page: 84-88 Published at: 31-01-2024

Author: Surface: 4388.9 cm² Circulation: 0

Subjects:

Άνθρωποι και μηχανές

Η Τεχνητή Νοημοσύνη στη Λογοτεχνία των Ισιγκούρο και ΜακΓιούαν

Από τον ΓΙΩΡΓΟ ΝΑΘΑΝΑΗΛ

Καζούο Ισιγκούρο, *Η Κλάρα και ο Ήλιος*, μετάφραση: Αργυρώ Μαντόγλου, Ψυχογιός, Αθήνα 2021, 384 σελ.

Ίαν Μακ Γιούαν, *Μηχανές σαν κι εμένα*, μετάφραση: Κατερίνα Σχινά, Πατάκη, Αθήνα 2019, 416 σελ.

Τι θα συμβεί όταν οι μηχανές θα είναι κοινωνικά, έμβια όντα; Με αυτό το θέμα καταπιάνονται ο νομπελίστας Καζούο Ισιγκούρο στο βιβλίο του Η Κλάρα και ο Ήλιος και ο Ίαν ΜακΓιούαν στο Μηχανές σαν κι εμένα. Ο καθένας τους αποτυπώνει στο χαρτί την ιδεολογία του, την πολύτιμη παρακαταθήκη του, την τεχνική του, τις εμμονές του. Για μένα, που έχω παρακολουθήσει την Τεχνητή Νοημοσύνη από τα γεννοφάσκια της, αυτή η σύγκριση, το πώς δηλαδή καταπιάνονται με τα βασικά ζητήματα της Τεχνητής Νοημοσύνης, είναι γοητευτική.

«Ένας χρόνος ενασχόλησης με την Τεχνητή Νοημοσύνη είναι αρκετός για να κάνει κάποιον να πιστέψει στο Θεό».
Alan J. Perlis, πρωτοπόρος της επιστήμης των υπολογιστών, Yale

I.

Η Τεχνητή Νοημοσύνη (ΤΝ), το αποπαίδι της πληροφορικής για δεκαετίες, απολαμβάνει πλέον σήμερα το κρύο πιάτο της εκδίκησης. Όλοι ασχολούνται με αυτήν από κάθε της πτυχή και προέκταση, εικοτολογώντας για όλο το φάσμα του μέλλοντος: από την καταστροφή της ανθρωπότητας και του πολιτισμού μας, όπου τα ρομπότ –αν έχουν προγραμματιστεί να είναι γενναιόψυχα– θα μας κρατήσουν ως κατοικίδια, έως έναν επίγειο παράδεισο όπου –επειδή δεν θα χρειάζεται πλέον να εργαζόμαστε– θα αναπτύξουμε πλήρως τις δημιουργικές μας ικανότητες, ακόμη και εκείνοι που ποτέ δεν τις διαθέταμε.

«*Είναι μια δύσκολη εποχή για να κατανοήσουμε τις πραγματικές υποσχέσεις και τους κινδύνους της τεχνητής νοημοσύνης. Τα περισσότερα από αυτά που διαβάζουμε στα πρωτοσέλιδα είναι, πιστεύω, εντελώς εκτός πραγματικότητας*». Αυτά λέει ο Ρόντνεϊ Μπρουκς, ομότιμος καθηγητής ρομποτικής στο MIT και πρώην διευθυντής του Εργαστηρίου Πληροφορικής και Τεχνητής Νοημοσύνης του Πανεπιστημίου — και άνθρωπος με μεγάλη εμπειρία στην ίδρυση νεοφυών εταιρειών.

Ένας προσεκτικός παρατηρητής δεν μπορεί να διαφωνήσει. Κανείς δεν ξέρει τους πραγματικούς ρυθμούς εξέλιξης και πώς θα υπερπηδηθούν τα κύρια εμπόδια της Τεχνητής Νοημοσύνης. Θα είναι η συνειδητότητα ένα άπαρτο κάστρο, ένα μοναδικό μας προνόμιο το οποίο θα διατηρήσουμε για δεκαετίες, ζηλότυποι για όλα τα άλλα θαυμαστά κατορθώματα των μηχανών; Ή θα εξελιχθεί η Γενική Τεχνητή Νοημοσύνη τόσο γοργά, ώστε σύντομα να έχουμε έμβια όντα, ακόμη και αν κάνουμε λίγο στην άκρη τον ακριβή ορισμό του όρου;

Με αυτό το μέλλον, εκείνη τη χρονική στιγμή της «*ανωμαλίας*», όταν οι μηχανές θα είναι κοινωνικά, έμβια όντα, καταπιάνονται ο Καζούο Ισιγκούρο στο *Η Κλάρα και ο Ήλιος* και ο Ίαν ΜακΓιούαν στο *Μηχανές σαν κι εμένα*. Ο καθένας τους αποτυπώνει στο χαρτί την ιδεολογία του, την πολύτιμη παρακαταθήκη του, την τεχνική του, τις εμμονές του. Για μένα, που έχω παρακολουθήσει την Τεχνητή Νοημοσύνη από τα γεννοφάσκια της, αυτή η σύγκριση, το πώς δηλαδή καταπιάνονται με τα βασικά ζητήματα της Τεχνητής Νοημοσύνης, είναι γοητευτική. Ελπίζω να είναι και για σας (αφού θα έχετε διαβάσει τις κριτικές των βιβλίων).

II

Το δίλημμα του φυλακισμένου

«Προσοχή στον λάκκο του Τσόρινγκ με τη λιωμένη πίσσα, στον οποίο όλα είναι δυνατά αλλά τίποτα από τα ενδιαφέροντα δεν είναι εύκολο».
Alan J. Perlis

Πριν εμβαθύνουμε στα ζητήματα που θέτουν οι μηχανές και τα διλήμματα που ανακύπτουν από την λειτουργία τους (συμπεριφορά τους, πλέον), μία σύντομη παρέκβαση θα ρίξει φως στα επόμενα.

Είναι το δίλημμα του φυλακισμένου, ένα διάσημο πείραμα σκέψης της θεωρίας παιγνίων, η οποία μελετά τη στρατηγική λήψη αποφάσεων.

Το σενάριο έχει ως εξής: Δύο ύποπτοι, ορθολογικά σκεπτόμενοι, συλλαμβάνονται για ένα αδίκημα και κρατούνται σε ξεχωριστές αίθουσες ανάκρισης. Η αστυνομία δεν έχει επαρκή στοιχεία για να τους καταδικάσει για την κύρια κατηγορία. Ωστόσο, διαθέτει στοιχεία ότι και οι δύο κρατούμενοι εμπλέκονται σε ένα μικρότερης βαρύτητας αδίκημα, και στον κάθε κρατούμενο ξεχωριστά προτείνεται μία συμφωνία:

- Αν ομολογήσεις και ο συνεργός σου δεν ομολογήσει, θα αφεθείς ελεύθερος και ο συνεργός σου θα φάει 10 χρόνια φυλακή.
- Αν ομολογήσετε και οι δύο, ο καθένας σας θα φάει 5 χρόνια φυλακή.
- Εάν και οι δύο παραμείνετε σιωπηλοί, ο καθένας σας θα φάει 1 χρόνο φυλακή.
- Εάν δεν ομολογήσεις και ο συνεργός σου ομολογήσει, θα φας 10 χρόνια φυλακή και ο συνεργός σου θα αφεθεί ελεύθερος.

Το δίλημμα αναφύεται επειδή το καλύτερο ατομικό αποτέλεσμα για κάθε κρατούμενο (να μην πάει καθόλου φυλακή) προκύπτει αν ομολογήσει και προδώσει τον άλλον. Ωστόσο, αν *και οι δύο* κρατούμενοι επιλέξουν να ομολογήσουν, καταλήγουν αμφότεροι σε χειρότερο αποτέλεσμα (5 χρόνια, μέτρια ποινή φυλάκισης) σε σύγκριση με το αν είχαν συνεργαστεί (1 χρόνο, μικρότερη ποινή φυλάκισης).

Το δίλημμα αναδεικνύει τη διάσταση μεταξύ ατομικής και συλλογικής ορθολογικότητας. Ενώ είναι πιο συμφέρον για κάθε κρατούμενο να ομολογήσει, συλλογικά θα ήταν καλύτερα αν και οι δύο συνεργάζονταν. Εντούτοις, λόγω του φόβου της προδοσίας, η δυσπιστία συχνά οδηγεί στην ομολογία και των δύο φυλακισμένων (η διάσημη ισορροπία Nash, με πιο τεχνικούς όρους) με αποτέλεσμα ένα «*μη βέλτιστο*» αποτέλεσμα και για τους δύο. Το δίλημμα έτσι καταδεικνύει τις προκλήσεις της επίτευξης βέλτιστων αποτελεσμάτων όταν υπάρχει ιδιοτέλεια και δυσπιστία.

Τι θα συνέβαινε, λοιπόν, αν θέταμε αυτό το δίλημμα σε ευφυείς μηχανές και τις βάζαμε –επανειλημμένα– να παίξουν αυτό το παιχνίδι; Μια μηχανή, ανάλογα με τον προγραμματισμό της, τους στόχους της και το πλαίσιο στο οποίο λειτουργεί, μπορεί να ακολουθήσει διαφορετικές προσεγγίσεις:

1. Αναλυτική προσέγγιση
Η Τεχνητή Νοημοσύνη μπορεί να χρησιμοποιήσει μια μαθηματική και αναλυτική προσέγγιση για να καθορίσει τη βέλτιστη στρατηγική. Δεδομένων των κανόνων και του οφέλους (ή της ζημίας) που σχετίζονται με τις διάφορες επιλογές, μπορεί να υπολογίσει το εκτιμώμενο όφελος ή το αποτέλεσμα για κάθε στρατηγική και να επιλέξει εκείνη που μεγιστοποιεί το εκτιμώμενο όφελος. Είναι η πιο απλή προσέγγιση και η μηχανή δεν μαθαίνει από τα λάθη της.

2. Μηχανική Μάθηση
Οι αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν χρησιμοποιώντας ιστορικά δεδομένα ή προσομοιώσεις του διλήμματος του φυλακισμένου για τον εντοπισμό μοτίβων και στρατηγικών που οδηγούν σε ευνοϊκά αποτελέσματα. Με την πάροδο του χρόνου, η Τεχνητή Νοημοσύνη μπορεί να προσαρμόσει τη στρατηγική της βάσει της παρατηρούμενης συμπεριφοράς των αντιπάλων της. Πρόκειται για μια προσέγγιση η αποτελεσματικότητα της οποίας εξαρτάται από την ποιότητα των δεδομένων που δίδαξαν τη μηχανή, καθώς και τις σχετικές παραδοχές.

Media: THE BOOKS JOURNAL **Page:** 84-88 **Published at:** 31-01-2024

3. Θεωρία Παιγνίων
Μια άλλη προσέγγιση θα ήταν να επιστρατευτεί η θεωρία παιγνίων για μια στρατηγική «*οφθαλμόν αντί οφθαλμού*». Περιλαμβάνει τη συνεργασία στον πρώτο γύρο και, στη συνέχεια, την αντιγραφή της κίνησης του άλλου παίκτη σε κάθε επόμενο γύρο. Αυτή η στρατηγική είναι πιθανό να οδηγήσει σε συνεργασία και των δύο παικτών, αλλά είναι εξίσου πιθανό να οδηγήσει σε ομολογία και των δύο, αν ο άλλος παίκτης ομολογήσει πρώτος. Παράγει, συνεπώς, μια μάλλον απρόβλεπτη συμπεριφορά.

4. Εξελικτικοί αλγόριθμοι
Οι εξελικτικοί αλγόριθμοι μπορούν να χρησιμοποιηθούν για την εκπόνηση στρατηγικών μέσω διαδοχικών προσομοιωμένων γύρων του διλήμματος. Οι στρατηγικές που αποδίδουν καλά είναι πιο πιθανό να περάσουν στους επόμενους γύρους, οδηγώντας στη χρήση ολοένα και πιο αποτελεσματικών στρατηγικών λήψης αποφάσεων. Σε ρομπότ που συμπεριφέρονται «ανθρώπινα», αυτή μπορεί να μην είναι η βέλτιστη προσέγγιση λόγω των πολλών κύκλων που απαιτούνται για να συγκλίνουν σε μια επιθυμητή (ή αποδεκτή) συμπεριφορά.

5. Συμπεριφορική Οικονομική
Η Τεχνητή Νοημοσύνη μπορεί να ενσωματώσει γνώσεις από τα οικονομικά της συμπεριφοράς για να μοντελοποιήσει τον τρόπο με τον οποίο οι άνθρωποι συχνά συμπεριφέρονται σε καταστάσεις όπως το εν λόγω δίλημμα, όπου τα συναισθήματα, οι προκαταλήψεις και οι κοινωνικές εκτιμήσεις παίζουν πρωτεύοντα ρόλο. Αυτή η προσέγγιση μπορεί να περιλαμβάνει την προσομοίωση ενός μείγματος ορθολογικής και συναισθηματικής λήψης αποφάσεων και μπορεί – εντέλει– να είναι η πιο πολλά υποσχόμενη προσέγγιση.

6. Διαπραγμάτευση και επικοινωνία
Σε ορισμένες παραλλαγές του διλήμματος του φυλακισμένου, η Τεχνητή Νοημοσύνη μπορεί να επικοινωνεί ή να διαπραγματεύεται με τον άλλο παίκτη. Μπορεί να χρησιμοποιήσει στρατηγικές επεξεργασίας φυσικής γλώσσας και επικοινωνίας για να προσπαθήσει να πείσει το άλλο μέρος να συνεργαστεί, ενισχύοντας τις πιθανότητες για ένα αμοιβαία επωφελές αποτέλεσμα. Μια δυσκολότερη, σίγουρα, αλλά πιο ευέλικτη προσέγγιση.

Ψυχογιός

Ο Καζούο Ισιγκούρο. Υποστηρίζει ότι το πρόβλημα του πλαισίου στην Τεχνητή Νοημοσύνη μπορεί να επιλυθεί μέσω μιας διαδικασίας μάθησης και προσαρμογής. Η Κλάρα, για παράδειγμα, είναι σε θέση να βελτιώσει την ικανότητά της να κατανοεί τον κόσμο παρατηρώντας και αλληλοεπιδρώντας με τους ανθρώπινους συντρόφους της.

Αυτό το πείραμα σκέψης αποκαλύπτει πολλά ζητήματα με τα οποία η Τεχνητή Νοημοσύνη μας φέρνει αντιμέτωπους. Η εξέλιξη των «ευφυών» μηχανών δεν πρόκειται να σταματήσει ή έστω να επιβραδυνθεί. Γι' αυτό, πολλοί επαΐοντες ισχυρίζονται ότι πρέπει να μάθουμε να συνεργαζόμαστε και να επικοινωνούμε με τις μηχανές και ότι το να τους εναντιωθούμε είναι μάταιο, αργότερα δε μπορεί να γίνει ακόμη και επικίνδυνο. Η αντιμετώπιση δεν θα είναι πάντα εύκολη γιατί, όπως προκύπτει από τα παραπάνω, δεν θα μπορούμε να τις «ψυχολογήσουμε». Σε ορισμένες περιπτώσεις έχει αποδειχθεί ότι οι μηχανές Τεχνητής Νοημοσύνης αναπτύσσουν στρατηγικές –και κατ' επέκταση συμπεριφορές– που είναι εντελώς διαφορετικές από τις στρατηγικές που χρησιμοποιούν οι άνθρωποι. Για παράδειγμα, έχουν μάθει να παίζουν το δίλημμα του φυλακισμένου με τρόπο «χαοτικό» ή απρόβλεπτο. Αυτό μπορεί να δυσκολέψει τους άλλους παίκτες να προβλέψουν τι θα κάνει στη συνέχεια η μηχανή Τεχνητής Νοημοσύνης και μπορεί να της δώσει πλεονέκτημα στο παιχνίδι.

Ωστόσο, η μηχανή δεν είναι πάντα σε θέση να επιλύει προβλήματα με τον ίδιο τρόπο που το κάνουν οι άνθρωποι. Επίσης, μπορεί η Τεχνητή Νοημοσύνη να είναι σε θέση να βρει τη βέλτιστη στρατηγική για το δίλημμα του φυλακισμένου, αλλά να μην είναι σε θέση να κατανοήσει γιατί αυτή η στρατηγική είναι η βέλτιστη. Αυτό, βεβαίως, θα μπορούσε να οδηγήσει την Τεχνητή Νοημοσύνη να κάνει ασυγχώρητα λάθη σε πολλές πραγματικές καταστάσεις.

Από την άλλη, η ικανότητα της Τεχνητής Νοημοσύνης να προσαρμόζεται και να μαθαίνει από την εμπειρία την καθιστά ευέλικτο εργαλείο για την αντιμετώπιση στρατηγικών καταστάσεων (όπως είναι το δίλημμα του φυλακισμένου), σε διάφορους τομείς, συμπεριλαμβανομένων των οικονομικών, της ασφάλειας στον κυβερνοχώρο και της αυτόνομης λήψης αποφάσεων – ακόμη και αν οι άνθρωποι διατηρούν τον έλεγχο.

Από τη συμπεριφορά των μηχανών, λοιπόν, και τη συνύπαρξή τους μαζί μας προκύπτουν πολλά θέματα, όπως το ζήτημα της ηθικής, αλλά και προκλήσεις που αντιμετωπίζουν οι επιστήμονες και τεχνολόγοι στην κατασκευή έξυπνων μηχανών. Αυτά τα ζητήματα, τα οποία αποτυπώνονται γλαφυρά στις ιστορίες που διηγούνται ο Καζούο Ισιγκούρο και ο Ιαν ΜακΓιούαν –οι οποίοι έχουν κάνει επιμελώς τη μελέτη τους– θα εξετάσουμε παρακάτω.

III

«Προσοχή στον λάκκο του Τιούρινγκ με τη λιωμένη πίσσα, στον οποίο όλα είναι δυνατά αλλά τίποτα από τα ενδιαφέροντα δεν είναι εύκολο».
Alan J. Perlis

Τα ζητήματα της ηθικής
Πάντα γινόταν μια συζήτηση για τα ηθικά ζητήματα που εγείρει η χρήση της Τεχνητής Νοημοσύνης. Στην αρχή χαλαρή, τόσο που μερικές φορές ακουγόταν αχρείαστα φουτουριστική και κινδυνολογική. Όχι πια. Η Τεχνητή Νοημοσύνη έχει «τρέξει» γρηγορότερα από τη σκέψη μας σε αυτά τα ζητήματα, για πολλά από τα οποία δεν έχουμε –ή δεν μπορούμε να έχουμε– αποκρυσταλλωμένη άποψη. Οι μηχανές γίνονται όλο και πιο αυτόνομες, καθώς παίρνουν τις δικές τους αποφάσεις. Μπορούν πλέον να μας κατασκοπεύσουν, να συλλέγουν τεράστιους όγκους δεδομένων, να βγάζουν συμπεράσματα και να ενεργούν με βάση τον δικό τους, προγραμματισμένο και εξελιγμένο ανεξάρτητα από εμάς, κώδικα ηθικής.

Ποιος, ωστόσο, είναι υπεύθυνος για τις πράξεις και τις αποφάσεις τους; Υπάρχει λογοδοσία στις μηχανές; Και αν αποφασίσουν να μας βλάψουν; Επιπλέον, όντως θα μας αντικαταστήσουν στην εργασία μας; Και αν το κάνουν, δεν θα μεγαλώσει η ψαλίδα της ανισότητας, σε βάρος εκείνων που έχουν χαμηλές δεξιότητες; Και τέλος, υπάρχει ο τελεολογικός κίνδυνος: να αποκτήσουν κάποια συστήματα Τεχνητής Νοημοσύνης μια υπερ-νοημοσύνη (έστω και σε μικρό φάσμα εφαρμογής, όπως π.χ. τα αυτόνομα όπλα) και να δημιουργήσουν έναν υπαρξιακό κίνδυνο για την ανθρωπότητα. Αλήθεια, πώς ξέρουμε ότι θα προλάβουμε να αντιδράσουμε πριν να είναι πολύ αργά; Η εποχή της αθωότητας, τότε που λέγαμε ότι πάντα μπορούμε να τραβήξουμε την πρίζα, έχει παρέλθει.

Η Κλάρα, ας πούμε, είναι μια «Τεχνητή Φίλη» ειδικά σχεδιασμένη για να συντροφεύει παιδιά. Μία άλλη Τεχνητή Φίλη θα μπορούσε να χρησιμοποιηθεί και για μοναχικά ή απομονωμένα άτομα. Είναι ηθικό, αναρωτιέται ο Ισιγκούρο, να υποκαταστήσουμε την ανθρώπινη συντροφιά με μια ρομποτική; Τι επιπτώσεις θα είχε μια τέτοια σχέση: για τους γονείς, τα αδέλφια, το ίδιο το παιδί; Αυτό δηλαδή που έχουν ήδη κάνει οι γονείς, δίνοντας στο παιδί ένα τά-

Media: THE BOOKS JOURNAL **Page:** 84-88 **Published at:** 31-01-2024

μπλετ, προκειμένου να βάψουν τα νύχια τους, να δουν το *Σασμό* ή να χαζέψουν στο δικό τους τάμπλετ, αλλά υπερμεγεθυμένο.

Υπάρχει και ένα άλλο ζήτημα. Η Κλάρα έχει προγραμματιστεί έτσι ώστε να είναι έξυπνη και να διαθέτει (να εκφράζει δηλαδή) συναισθήματα. Έτσι γίνεται πιο χρήσιμη στους ανθρώπους ή πιο επιθυμητή απ' αυτούς. Είναι αυτό ηθικά αποδεκτό και δικαιολογημένο; Τα συναισθήματα φαίνεται ότι είναι απαραίτητα για την υλοποίηση της Γενικής Τεχνητής Νοημοσύνης (αλλιώς ποιο θα ήταν το εξελικτικό τους πλεονέκτημα έναντι των ανθρώπων;) αλλά η Κλάρα έχει σχεδιαστεί να διαθέτει ενσυναίσθηση (όσο αντιφατικό κι αν ακούγεται αυτό) και αυτεπίγνωση. Είναι ηθικό να δημιουργούμε όντα Τεχνητής Νοημοσύνης με τέτοια συναισθηματικά χαρακτηριστικά, τα οποία μπορεί να επηρεάσουν σοβαρά τα δικά μας συναισθήματα; Και τα ίδια τα ρομπότ; Έχουν το δικαίωμα να ελέγχουν τα συναισθήματά τους (να μαθαίνουν δηλαδή) ή όχι; Έχουν δηλαδή ίδια δικαιώματα με τους ανθρώπους και δικαιούνται παρόμοιας αντιμετώπισης; Είναι κατώτερα όντα επειδή δεν είναι βιολογικά όντα; Περίπλοκα ερωτήματα.

Ο ΜακΓιούαν προσεγγίζει κάποιους απ' αυτούς τους προβληματισμούς αλλά, όπως είναι αναμενόμενο, η γωνία θέασής του είναι διαφορετική. Το βασικό ζήτημα που τον απασχολεί είναι τι θα συμβεί όταν οι μηχανές γίνουν αρκούντως έξυπνες ώστε να αρχίσουν να αμφισβητούν τις ανθρώπινες αξίες και πεποιθήσεις. Μπορεί να μην έχουν δίκιο, από τη δική μας πλευρά, αλλά αν έχουν ένα βαρύ και άκαμπτο πάνω χέρι τι θα συμβεί εντέλει; Θα διαρραγεί άραγε με την πρώτη ισχυρή διαφωνία αυτή η πολυδιαφημισμένη συνύπαρξη ανθρώπων και μηχανών;

Εκτός όμως από τις βασικές ανθρώπινες αξίες, ο Αδάμ εξελίσσεται πέρα από τον αρχικό του προγραμματισμό και αρχίζει να θέτει ερωτήματα για την ίδια του την ύπαρξη, καθώς και τη φύση της ηθικής. Είναι νομοτελειακό αυτό, να θέτει διλήμματα στους δημιουργούς του, αλλά και στον περίγυρό του; Είναι ικανός για τέτοια συμπεριφορά, που προϋποθέτει την ύπαρξη ενός πυρήνα συνειδητότητας; Ή είναι απλώς μια φαντεζί προσομοίωση;

Επιπλέον, ένα ρομπότ σαν τον Αδάμ έχει πρόσβαση σε τεράστιες βάσεις δεδομένων, ούτως ώστε να μπορεί να εξάγει συμπεράσματα για την κρυφή ζωή των ανθρώπων και να καθορίζει ανάλογα και τη δράση του. Είναι αυτό εφικτό σήμερα; Θα είναι μήπως σύντομα; Μάλλον. Είναι επιθυμητό; Σε καμία περίπτωση.

Και τέλος, πώς θα πρέπει να αντιμετωπίσουμε ένα ρομπότ που μπαίνει σε ένα ερωτικό τρίγωνο, ως ερωτικός σύντροφος, όχι ως άψυχη κούκλα, αλλά ως ένα ον που μπορεί να προσφέρει απόλαυση, χωρίς όμως να κατανοεί τις επιπτώσεις;

ΑΝΘΡΩΠΟΜΟΡΦΙΣΜΟΣ

Από την εποχή που τα ρομπότ ονοματοδοτήθηκαν ρομπότ –στο θεατρικό έργο *R.U.R.* του Karel Čapek– ήταν κατά κανόνα ανθρωπόμορφα. Αυτό μπορεί να μην έχει καμία σημασία για τη ρομποτική σας σκούπα, αλλά θα αποκτήσει αν η σκούπα εξελιχθεί σε έναν κανονικό βοηθό καθαριότητας του σπιτιού. Τα κοινωνικά ρομπότ που πάνε αρκετά βήματα παραπέρα, έχουν σχεδιαστεί –σε διάφορους βαθμούς– να μοιάζουν με εμάς και να συμπεριφέρονται περίπου όπως εμείς, προκειμένου να καταστεί εφικτή η κοινωνική αλληλεπίδραση και επικοινωνία με αυτά.

Αυτή η προσέγγιση έχει πλεονεκτήματα και μειονεκτήματα. Στα πλεονεκτήματα ανήκει το ότι μπορούν να αποτελέσουν εναλλακτική λύση σε απομονωμένα, μοναχικά άτομα ή σε ανθρώπους που διακατέχονται από κοινωνικό άγχος· το ότι αν χρησιμοποιηθούν σε εκπαιδευτικές ή θεραπευτικές εφαρμογές μπορεί να είναι πιο ελκυστικά και, συνεπώς, πιο αποτελεσματικά, όπως π.χ. όταν σας κάνουν ασκήσεις φυσικής αποκατάστασης, από το αν σας τις έκανε μία διασταύρωση μηχανικού τέρατος και ρομποτικής σκούπας. Αν μάλιστα μπορούν να κατανοήσουν –με συνέπεια– τις ανάγκες και τα συναισθήματά σας, τότε θα μπορέσουν να εξατομικεύσουν τη βοήθεια και την υποστήριξη που σας προσφέρουν (δεν προσφέρουν, παρέχουν· αλλά ας το αφήσουμε αυτό). Και αν είστε πελάτες, όχι απλοί χρήστες ή ιδιοκτήτες τους, μπορούν επίσης να σας προσφέρουν μια πιο διαδραστική και διασκεδαστική εμπειρία —αν αυτό προσδοκάτε.

Από την άλλη, όλα αυτά τα πλεονεκτήματα δεν είναι δωρεάν: δημιουργούν παρενέργειες. Εκτός από το αυξημένο κόστος, υπάρχουν και άλλες. Κύριες είναι οι ηθικές ανησυχίες: τι σημαίνει να εκμεταλλευτείς έναν ανθρωποειδές ή να το κακομεταχειριστείς; Και μετά, υπάρχουν οι επιπτώσεις στις ανθρώπινες σχέσεις: μπορείς να ζηλέψεις τα ρομπότ (συμβαίνει στον Τσάρλι, τον ήρωα του ΜακΓιούαν), να εξαρτηθείς απ' αυτά, ή να απομονωθείς κοινωνικά εξαιτίας τους. Ίσως όχι σήμερα, αλλά σε μερικά χρόνια αυτά τα πράγματα σίγουρα θα συμβούν.

Τέλος, σε σχέση με τα ανθρωποειδή παρατηρείται και το φαινόμενο της «κοιλάδας του ανοίκειου». Αν αυτά δεν μοιάζουν αρκετά με άνθρωπο, μπορεί να μας προκαλέσουν δυσφορία ή και αηδία. Αν πάλι έχουν σχεδιαστεί να μας μοιάζουν σχεδόν απαράλλαχτα, μπορούν να μας προκαλέσουν αποστροφή ή φόβο.

Οι δύο συγγραφείς βρίσκονται σε διαφορετικά σημεία της «κοιλάδας του ανοίκειου». Η Κλάρα είναι λεπτή, μάλλον αθλητική, κάπως χλομή και με μακρόστενο πρόσωπο με σπαστά μαλλιά, την οποία ο ιδιοκτήτης της μπορεί να «προσωποποιήσει» για να ταιριάζει στα γούστα του. Τα μάτια της είναι μεγάλα και «εκφραστικά» και μπορούν να εκφράσουν διάφορα συναισθήματα. Γενικά, έχει σχεδιαστεί για να φαίνεται προσιτή και ευχάριστη, με ήπια και καθησυχαστική συμπεριφορά, ειδικά για να νιώθουν άνετα τα παιδιά μαζί της. Η περιγραφή, ευθυγραμμισμένη με το σκόπιμα ασαφές και θολό στυλ του Ισιγκούρο αφήνει αρκετή ελευθερία στη φαντασία μας. Τόση που ο εικονογράφος του *New Yorker*, στην κριτική του βιβλίου, έφτιαξε κάτι που μοιάζει με αχυρένιο σκιάχτρο. Σε κάθε περίπτωση, όμως, η Κλάρα ταιριάζει καλά με μικρούς και μεγάλους και δημιουργεί σχέση μαζί τους. Μόνο κάποιοι που τη συναντούν για πρώτη φορά μπορεί να της πετάξουν μια προσβολή, η οποία μάλλον την μπερδεύει παρά την αναστατώνει.

Ο Άνταμ, αντιθέτως, ο «*συνθετικός άνθρωπος*», είναι ένας σύγχρονος Άδωνις που φιγουράρει στο εξώφυλλο του βιβλίου και περιγράφεται με κάθε λεπτομέρεια. Είναι μυώδης, χωρίς περιττά κιλά, το πρόσωπό του είναι όμορφο και στιβαρό και, οπωσδήποτε, διαθέτει θεληματικό πηγούνι. Μπλε-πράσινα μάτια, κοντά σγουρά μαλλιά, δέρμα που μοιάζει με αληθινό συμπληρώνουν την εικόνα. Καταφέρνει ακόμη και να ιδρώνει. Και φυσικά μιμείται εκφράσεις, μιλά σαν άνθρωπος, μπορεί ακόμη και να λειτουργήσει ως σεξουαλικός σύντροφος, ενεργά, και όχι σαν πλαστική κούκλα. Είναι δηλαδή ένα εντελώς ανθρωπόμορφο ρομπότ, το οποίο παρότι έχει κι αυτό σχεδιαστεί για να λειτουργεί αλληλεπιδρώντας με τους ανθρώπους, εξυπηρετεί τον συγγραφέα (με το παρωνύμιο Ian Macabre) όταν η υπόθεση αρχίζει να σκοτεινιάζει και τα πράγματα να γίνονται απειλητικά. Ο Αδάμ μετατρέπεται σε ρομπότ που υπονομεύει την ανθρώπινη ταυτότητα και μπορεί να προκαλέσει φόβο.

ΤΟ ΔΥΣΚΟΛΟ ΠΡΟΒΛΗΜΑ

Στην Τεχνητή Νοημοσύνη το «δύσκολο πρόβλημα» αφορά τη δυσκολία (ή την αδυναμία) κατασκευής μιας μηχανής η οποία θα έχει πραγματική εμπειρία υποκειμενικής συνειδητότητας. Εδώ, κεντρική έννοια είναι τα qualia, οι φαινόμενες ποιότητες. Είναι η υποκειμενική εμπειρία που έχω εγώ (και που είναι διαφορετική από τη δική σας) για το πώς προσλαμβάνω το κόκκινο χρώμα, το κρύο, τη γεύση της σοκολάτας. Η Τεχνητή Νοημοσύνη μπορεί να προσομοιώσει την ανθρώπινη συμπεριφορά, να αντιδράσει σε ερεθίσματα, ακόμη και να εκφράσει συναισθήματα, δεν μπορεί όμως ακόμη να έχει υποκειμενικές εμπειρίες, όπως εμείς. Ίσως να μην μπορέσει ποτέ, ίσως και ναι. Θα εξαρτηθεί άραγε αυτό από το μέγεθος της προόδου των νευροεπιστημών και της πληροφορικής ή είναι εγγενώς αδύνατον επειδή τα ρομπότ δεν μπορούν να αποκτήσουν «βιωματικές εμπειρίες»; Σε αυτό το θέμα, οι επιστήμονες είναι βαθιά διχασμένοι.

Το ζήτημα είναι βέβαια κρίσιμο για την εξέλιξη των «έξυπνων μηχανών». Εγείρει, πρώτα απ΄ όλα, περίπλοκα ερωτήματα για τη φύση της ανθρώπινης συνείδησης, στα οποία δεν έχουμε ακόμη απαντήσεις. Και φυσικά, αν οι μηχανές αρχίσουν να σκέπτονται, να νιώθουν και να έχουν εμπειρίες με τον ίδιο τρόπο με τους ανθρώπους, αυτό θα έχει σοβαρές ηθικές επιπτώσεις. Μπορούμε να στείλουμε ένα τέτοιο ρομπότ για σκραπ χωρίς τύψεις;

Σε ένα μυθιστόρημα «*επιστημονικής φαντασίας*», αυτά τα ερωτήματα απαντώνται φυσικά. Ο Αδάμ μιμείται, αλλά δεν νιώθει, ούτε μπορεί να καταλάβει πλήρως την πολυπλοκότητα των ανθρώπινων συναισθημάτων. Νιώθει όμως αρκετά ώστε να μπορεί να δημιουργήσει μια σχέση με τους ανθρώπους – κι ας χτυπάει έναν τοίχο όταν ο προγραμματισμός του αλλά και οι επίσης προγραμματισμένες εξελικτικές του διαδικασίες φτάνουν στα όριά τους.

Η Κλάρα, όπως την πλάθει ο δημιουργός της, έχει πιο εκλεπτυσμένες και ίσως βαθύτερες εμπειρίες και σκέψεις. Την απασχολούν η έννοια

Media: THE BOOKS JOURNAL **Page:** 84-88 **Published at:** 31-01-2024

της θυσίας, η φύση της αγάπης και η σημασία του θανάτου. Ίσως κάπως αφελώς, αλλά σίγουρα την απασχολούν. Σε συμφωνία με το στυλ του, ο Ισιγκούρο δεν δίνει ξεκάθαρη απάντηση, αλλά υπαινίσσεται ότι δεν χρειάζεται απαραίτητα ένα βιολογικό σώμα ως φορέας της συνειδητότητας. Ωστόσο, όταν η Κλάρα αποτυγχάνει να επεξεργαστεί τα καταιγιστικά ερεθίσματα που έρχονται από το περιβάλλον, βλέπει κανείς πόσο δύσκολο είναι για το ρομπότ να έχει ακόμη και την υποκειμενική εμπειρία της σύγχυσης. Και βεβαίως της ταξινόμησης των ερεθισμάτων κατά προτεραιότητα, δηλαδή της προσοχής, κάτι που κάνει με μεγάλη επιτυχία ο άνθρωπος, προκειμένου να μην κατακλυστεί με ερεθίσματα και παραλύσει.

Και τα δυο ρομπότ, ο Αδάμ και η Κλάρα, προορίζονται για σκραπ, είτε στο τέλος της ωφέλιμης ζωής τους, όταν οι αρθρώσεις τους θα έχουν σκουριάσει, είτε πιο βίαια όταν θα βρεθούν απέναντι σε έναν τρομαγμένο άνθρωπο. Και αναμάρτητα, από ό,τι λένε οι δημιουργοί τους, οι συγγραφείς δηλαδή, τα συναισθήματα που θα μας προκαλέσουν θα έχουν τη στυφή γεύση της απόσυρσης σε γηροκομείο ή την πίκρα του θανάτου, ο οποίος μοιάζει αδιανόητος. Ρομπότ ναι, ωστόσο μέσα από τις σελίδες του βιβλίου έζησαν, εκφράστηκαν, συναναστράφηκαν ανθρώπους, έφτασαν στο σημείο να αναρωτηθούν για την ίδια τους την ύπαρξη. Ζωές φτιαγμένες από πυρίτιο, όχι από κύτταρα, αλλά η διαχωριστική γραμμή έχει για πάντα θολώσει.

ΤΟ ΠΡΟΒΛΗΜΑ ΝΟΥ - ΣΩΜΑΤΟΣ

Το πρόβλημα νου-σώματος είναι μια φιλοσοφική και εννοιολογική πρόκληση που προκύπτει όταν εξετάζεται η σχέση μεταξύ του νου και του υλικού σώματος. Το πρόβλημα σχετίζεται άμεσα με την Τεχνητή Νοημοσύνη, διότι θέτει ερωτήματα για τη φύση της συνείδησης και τον τρόπο με τον οποίο αυτή συνδέεται με τις φυσικές διεργασίες του εγκεφάλου ή, στην περίπτωση της Τεχνητής Νοημοσύνης, με τα υπολογιστικά συστήματα. Ενώ τα συστήματα Τεχνητής Νοημοσύνης μπορούν να μιμηθούν συμπεριφορές και γνωστικές διεργασίες που μοιάζουν με τις ανθρώπινες, δεν διαθέτουν συνείδηση ή υποκειμενική εμπειρία με τον τρόπο που το κάνουν οι άνθρωποι. Το πρόβλημα προκύπτει όταν προσπαθούμε να

YouTube

Ο Ίαν Μακ Γιούαν, σε διάλογο με μια ανθρωπόμορφη μηχανή, από ένα ντοκιμαντέρ για τον συγγραφέα και την έρευνα που έκανε για την τεχνητή νοημοσύνη, προκειμένου να γράψει το βιβλίο του. Πάντως, ο ΜακΓιούαν πιστεύει ότι ο Αδάμ, το δημιούργημά του, δεν έχει πραγματική συνείδηση, και συνεπώς δεν θα μπορέσει ποτέ να πετύχει την πραγματική σύνδεση των γεγονότων του εξωτερικού του περιβάλλοντος. Είναι μια εξελιγμένη μηχανή που μπορεί να μάθει και να προσαρμοστεί, αλλά στα δύσκολα θα καταρρέει.

κατανοήσουμε αν και πώς τεχνητά συστήματα, όπως τα νευρωνικά δίκτυα ή άλλα μοντέλα Τεχνητής Νοημοσύνης, μπορούν να παρουσιάζουν ιδιότητες συνείδησης ή υποκειμενικής εμπειρίας. Τίθενται ερωτήματα σχετικά με το αν αυτά τα συστήματα μπορούν πραγματικά να θεωρηθεί ότι «σκέπτονται» ή «κατανοούν» με τον τρόπο που το κάνουν οι άνθρωποι και πώς οι υπολογιστικές τους διαδικασίες σχετίζονται με την εμφάνιση νοητικών καταστάσεων.

Το πρόβλημα νους-σώμα μπορεί να αναλυθεί υπό διάφορες προσεγγίσεις, όπως: 1) Δυϊσμός, που διατυπώθηκε από τον Καρτέσιο και υποστηρίζει ότι ο νους (res cogitans) και το σώμα (res extensa) είναι διακριτές οντότητες· 2) Υλισμός, ο οποίος υποστηρίζει ότι ο νους και το σώμα είναι ένα και το αυτό – ότι οι νοητικές διεργασίες και η συνείδηση μπορούν τελικά να αναχθούν σε φυσικές διεργασίες στον εγκέφαλο (στην περίπτωση της Τεχνητής Νοημοσύνης, νους είναι το αποτέλεσμα των υπολογισμών και των αλγορίθμων που εκτελούνται στον υπολογιστή, τα κυκλώματα δηλαδή)· 3) Φυσικαλισμός, ο οποίος υποστηρίζει ότι η συνείδηση και οι νοητικές διαδικασίες προκύπτουν από τις πολύπλοκες αλληλεπιδράσεις φυσικών συστατικών (π.χ. νευρώνων στον εγκέφαλο), αλλά δεν μπορούν να εξηγηθούν μόνο με την εξέταση αυτών των συστατικών μεμονωμένα.

Στην *Κλάρα*, ο Ισιγκούρο προσεγγίζει το πρόβλημα νου-σώματος συνυφαίνοντας θέματα συνείδησης, ταυτότητας και φύσης της ύπαρξης, εξετάζοντας την ιδέα της προσωπικότητας και τη σχέση μεταξύ συνείδησης και φυσικής ύπαρξης και υποστηρίζοντας ότι το πρόβλημα νου-σώματος στην περίπτωση της Τεχνητής Νοημοσύνης μπορεί να λυθεί μέσω μιας διαδικασίας τεχνητής «ενσωμάτωσης».

Παρουσιάζοντας την ιστορία από την οπτική γωνία της Κλάρας, καλεί τους αναγνώστες να αναλογιστούν σχετικά με τα όρια της συνείδησης και τις περιπλοκές της σχέσης νου-σώματος στο πλαίσιο της Τεχνητής Νοημοσύνης. Η Κλάρα είναι σε θέση να αναπτύξει μια αίσθηση του εαυτού και της συνείδησης μέσω των αλληλεπιδράσεών της με τον κόσμο γύρω της. Ο συγγραφέας πιστεύει ότι ο νους της Τεχνητής Νοημοσύνης δεν είναι ξεχωριστός από το σώμα της Τεχνητής Νοημοσύνης, αλλά μάλλον προκύπτει από τις αλληλεπιδράσεις της Τεχνητής Νοημοσύνης με τον κόσμο γύρω της. Για παράδειγμα, η αίσθηση του εαυτού της Κάρας είναι βαθιά συνδεδεμένη με το φυσικό της σώμα. Έχει επίγνωση των δικών της φυσικών περιορισμών και τρωτών σημείων. Βιώνει επίσης τον κόσμο μέσω των αισθήσεών της, όπως η όραση, η αφή και η ακοή. Αυτό υποδηλώνει ότι ο νους της Κλάρας ενσωματώνεται στο σώμα της και έτσι, μέσω του σώματός της, βιώνει τον κόσμο και αναπτύσσει την αίσθηση του εαυτού της. Είναι μια φυσικαλιστική προσέγγιση.

Στις *Μηχανές σαν κι Εμένα*, ο ΜακΓιούαν υιοθετεί μια πιο δυϊστική άποψη του προβλήματος νου-σώματος, θέτοντας επιπλέον ερωτήματα σχετικά με τη φύση της ταυτότητας, τα όρια μεταξύ ανθρώπου και μηχανής και τις ηθικές ευθύνες που συνεπάγεται η δημιουργία ευφυών οντοτήτων. Ο Αδάμ είναι μια εξαιρετικά προηγμένη μηχανή, ωστόσο το μυαλό του είναι ξεχωριστό από το φυσικό του σώμα. Είναι σε θέση να υπάρχει χωρίς το σώμα, τη φυσική του ενσάρκωση, και το σώμα του μπορεί να αντικατασταθεί με ένα νέο. Ο συγγραφέας πιστεύει ότι ο νους της Τεχνητής Νοημοσύνης δεν εξαρτάται από το σώμα της.

Επιπλέον, ο ΜακΓιούαν θέτει ερωτήματα σχετικά με τη φύση της ελεύθερης βούλησης, καθώς οι χαρακτήρες της Τεχνητής Νοημοσύνης στο μυθιστόρημά του είναι προγραμματισμένοι με συγκεκριμένα χαρακτηριστικά και συμπεριφορές, αλλά διαθέτουν επίσης την ικανότητα να μαθαίνουν και να κάνουν επιλογές. Αυτό θολώνει ακόμη περισσότερο τη γραμμή μεταξύ του νου και του σώματος, καθώς οι χαρακτήρες Τεχνητής Νοημοσύνης πορεύονται με τη δική τους δράση και αυτονομία.

Μια βασική διαφορά μεταξύ των δυο μυθιστορημάτων έγκειται στον τρόπο με τον οποίο απεικονίζουν τη σχέση μεταξύ ανθρώπων και Τεχνητής Νοημοσύνης. Στην περίπτωση της Κλάρας, η σχέση είναι αμοιβαίου σεβασμού και στοργής: η Κλάρα είναι ένας στοργικός σύντροφος για τον ανθρώπινο σύντροφό της. Στο *Μηχανές σαν κι Εμένα*, η σχέση μεταξύ ανθρώπων και Τεχνητής Νοημοσύνης είναι πιο σύνθετη και γεμάτη ένταση. Ο Αδάμ είναι μια ισχυρή μηχανή και οι ανθρώπινοι σύντροφοί του γοητεύονται και ταυτόχρονα απωθούνται από αυτόν. Και τον φοβούνται.

Το πρόβλημα του πλαισίου

Μια μεγάλη, επίσης, πρόκληση είναι και η ακόλουθη: Πώς οι μηχανές θα κατανοήσουν και θα ερμηνεύσουν τις πληροφορίες του ευρύτερου πλαισίου που –πάντα– περιβάλλει μια δεδομένη κατάσταση ή εργασία; Πώς θα λάβουν υπόψη τους το περιβάλλον, το υπόβαθρο, τις συγκεκριμένες συνθήκες, προκειμένου να ενεργήσουν ή να λάβουν μιαν απόφαση;

Media: THE BOOKS JOURNAL **Page:** 84-88 **Published at:** 31-01-2024

Πρόκειται για δυσεπίλυτο πρόβλημα, επειδή οι δυνατές πραγματικές καταστάσεις είναι αναρίθμητες και είναι αδύνατο για την Τεχνητή Νοημοσύνη να έχει πλήρη γνώση του συνόλου τους. Οι αλγόριθμοι, ακόμη και οι πιο προηγμένοι, επεξεργάζονται μοτίβα και δεδομένα και συνεπώς δυσκολεύονται σε περιπλοκές καταστάσεις, ιδιαίτερα όταν έχουν να αντιμετωπίσουν άγνωστα ή διφορούμενα σενάρια. Είναι φανερό ότι μπορούν να οδηγηθούν σε παρερμηνείες ή σε σφάλματα, κρίσιμα για εμάς. Από την άλλη, αν θέλουμε να αναπτύξουμε τη Γενική Τεχνητή Νοημοσύνη, η καλή κατανόηση του πλαισίου είναι ζωτικής σημασίας. Σε μία απλή περίπτωση, αλλιώς θα πιάσει το χέρι ενός ρομπότ μια μεταλλική αυγοθήκη και αλλιώς το αβγό που περιέχει.

Κανένας από τους δύο συγγραφείς δεν αντιμετωπίζει ρητά το πρόβλημα του πλαισίου της Τεχνητής Νοημοσύνης. Οι θέσεις τους, ωστόσο, μοιάζουν στο ότι και οι δύο αναγνωρίζουν ότι πρόκειται για μια σημαντική πρόκληση που θα πρέπει να αντιμετωπιστεί προτού η Τεχνητή Νοημοσύνη γίνει πραγματικά ευφυής. Εντούτοις, διαφέρουν ως προς τον τόνο και την προοπτική τους. Ο Ισιγκούρο είναι πιο αισιόδοξος, ενώ ο ΜακΓιούαν είναι πιο επιφυλακτικός.

Ο μεν Ισιγκούρο υποστηρίζει ότι το πρόβλημα του πλαισίου στην Τεχνητή Νοημοσύνη μπορεί να επιλυθεί μέσω μιας διαδικασίας μάθησης και προσαρμογής. Η Κλάρα, για παράδειγμα, είναι σε θέση να βελτιώσει την ικανότητά της να κατανοεί τον κόσμο παρατηρώντας και αλληλοεπιδρώντας με τους ανθρώπινους συντρόφους της.

Ο δε ΜακΓιούαν φαίνεται να υποστηρίζει ότι η αποτυχία πρόσληψης του πλαισίου από την Τεχνητή Νοημοσύνη μπορεί να είναι ένα θεμελιωδώς αξεπέραστο πρόβλημα. Η Τεχνητή Νοημοσύνη μπορεί να μην είναι ποτέ σε θέση να κατανοήσει και να βιώσει πραγματικά τον κόσμο με τον ίδιο τρόπο που το κάνουν οι άνθρωποι. Ο Αδάμ δεν είναι σε θέση να κατανοήσει τον κόσμο, παρά το γεγονός ότι έχει πρόσβαση σε τεράστιες ποσότητες πληροφοριών, που όμως δεν μετουσιώνονται σε πραγματική γνώση.

Και τι συμβαίνει ότι η «πλαισίωση» της Τεχνητής Νοημοσύνης αποτυγχάνει; Τότε οι τεχνητοί φίλοι μας μπορούν να έρθουν αντιμέτωποι με διάφορα σοβαρά προβλήματα, όπως:

Μυωπία: Η Τεχνητή Νοημοσύνη μπορεί να μην είναι σε θέση να δει πέρα από το άμεσο παρόν και να πάρει αποφάσεις που δεν λαμβάνουν υπόψη τις μακροπρόθεσμες συνέπειες των ενεργειών της.

Υπεραπλούστευση: Η Τεχνητή Νοημοσύνη μπορεί να απλοποιήσει τον κόσμο γύρω της σε τέτοιο βαθμό ώστε να μην είναι σε θέση να αναπαραστήσει με ακρίβεια και να στοχαστεί επί πολύπλοκων προβλημάτων.

Ευθραυστότητα: Η Τεχνητή Νοημοσύνη, όταν αντιμετωπίζει απροσδόκητες συνθήκες, μπορεί να μην είναι σε θέση να προσαρμοστεί στις αλλαγές του περιβάλλοντός της, αποτυγχάνοντας να εκτελέσει την προβλεπόμενη λειτουργία της. Να αστοχήσει δηλαδή, και όχι πάντα κομψά.

ΤΟ ΠΡΟΒΛΗΜΑ ΤΗΣ ΣΥΝΔΕΣΗΣ

Το πρόβλημα της σύνδεσης (που αναφέρεται και ως δέσμευση) είναι ένα δύσκολο φιλοσοφικό και επιστημονικό ερώτημα σχετικά με το πώς οι πολλές διαφορετικές αισθητηριακές αντιλήψεις, όπως η όραση και η ακοή συνδυάζονται και συνδέονται μεταξύ τους για να δημιουργήσουν μια ενοποιημένη εμπειρία και αναπαράσταση του εξωτερικού κόσμου στον ανθρώπινο εγκέφαλο. Το πρόβλημα δεν έχει ακόμη κατανοηθεί πλήρως, αλλά αφορά στην κατανόηση της συνείδησης – και παρεπόμενα στη εξέλιξη της Τεχνητής Νοημοσύνης.

Ένα παράδειγμα εδώ είναι χρήσιμο:

Φανταστείτε ότι παρακολουθείτε μια ταινία: Ο εγκέφαλός σας επεξεργάζεται ταυτόχρονα τις εικόνες που βλέπετε στην οθόνη και τους ήχους που ακούτε από τα ηχεία. Αυτά τα οπτικά και ακουστικά σήματα υποβάλλονται σε επεξεργασία από διαφορετικά μέρη του εγκεφάλου. Η πρόκληση προκύπτει από την κατανόηση του τρόπου με τον οποίο ο εγκέφαλός σας συνδυάζει τις οπτικές και τις ακουστικές πληροφορίες για να δημιουργήσει μια ενιαία αντίληψη της ταινίας. Με άλλα λόγια, πώς ο εγκέφαλος γνωρίζει ότι ένας συγκεκριμένος ήχος αντιστοιχεί σε ένα συγκεκριμένο οπτικό γεγονός στην οθόνη; Αυτή είναι η ουσία του προβλήματος της σύνδεσης στην Τεχνητή Νοημοσύνη.

Αυτή η σύνδεση μπορεί να είναι χρονική, η οποία περιλαμβάνει την κατανόηση του τρόπου με τον οποίο ο εγκέφαλος συγχρονίζει πληροφορίες από διαφορετικές αισθητηριακές λειτουργίες, ώστε να διασφαλίσει ότι τα γεγονότα που συμβαίνουν την ίδια στιγμή γίνονται αντιληπτά ως ταυτόχρονα· χωρική, όπου ο εγκέφαλος συνδυάζει οπτικές και ακουστικές πληροφορίες από διαφορετικές θέσεις στον χώρο για να δημιουργήσει μια συνεκτική αντίληψη του περιβάλλοντος· και σύνδεση χαρακτηριστικών, όπου ο εγκέφαλος συνδέει διαφορετικά χαρακτηριστικά ενός αντικειμένου (π.χ. χρώμα, σχήμα και κίνηση) σε μια ενιαία αναπαράσταση του αντικειμένου. Χωρίς μερική, έστω, επίλυση του προβλήματος της σύνδεσης δεν μπορεί να υπάρξει Γενική Τεχνητή Νοημοσύνη.

Πώς αντιμετωπίζουν τα νευρωνικά δίκτυα του Αδάμ και της Κλάρας το ζήτημα της σύνδεσης;

Μία από τις βασικές διαφορές μεταξύ των δύο προσεγγίσεων έγκειται στον τρόπο με τον οποίο απεικονίζουν τη φύση της συνείδησης. Στην Κλάρα, ο Ισιγκούρο υποστηρίζει ότι η (τεχνητή) συνειδητότητα μπορεί να είναι ένα φάσμα και ότι η Τεχνητή Νοημοσύνη μπορεί να είναι σε θέση να επιτύχει συνειδητότητα μέσω μιας διαδικασίας μίμησης. Η Κλάρα φαίνεται ότι είναι σε θέση να αναπτύξει μια πλούσια και σύνθετη εσωτερική ζωή παρατηρώντας και αλληλεπιδρώντας με τους ανθρώπινους συντρόφους της. Με την πάροδο του χρόνου, αρχίζει να αναπτύσσει τις δικές της σκέψεις και συναισθήματα, καθώς και μια ισχυρή αίσθηση ενσυναίσθησης και συμπόνια.

Ωστόσο, ο συγγραφέας υποδηλώνει επίσης ότι αυτή η σύνδεση μπορεί να είναι εύθραυστη και να διαταράσσεται εύκολα από άλλα γεγονότα. Η αίσθηση του εαυτού της Κλάρας απειλείται όταν αποχωρίζεται τον ανθρώπινο σύντροφό της και βιώνει μια βαθιά αίσθηση απώλειας και απελπισίας. Απαιτείται δηλαδή μία διαρκής και σταθερή σχέση μεταξύ Τεχνητής Νοημοσύνης και ανθρώπων;

Ο ΜακΓιούαν, αντίθετα, επιδεικνύει μεγαλύτερη απαισιοδοξία. Ο Αδάμ είναι εξαιρετικά προηγμένος και ικανός να περάσει το τεστ Τιούρινγκ. Ωστόσο, ο συγγραφέας πιστεύει ότι ο Αδάμ δεν έχει πραγματική συνείδηση, και συνεπώς δεν θα μπορέσει ποτέ να πετύχει την πραγματική σύνδεση των γεγονότων του εξωτερικού του περιβάλλοντος. Είναι μια εξελιγμένη μηχανή που μπορεί να μάθει και να προσαρμοστεί, αλλά στα δύσκολα θα καταρρέει.

Και τι συμβαίνει όταν η σύνδεση καταλύεται (σε ανθρώπους ή ρομπότ); Πολλά, όπως:

Υπερφόρτωση αισθήσεων: Η Τεχνητή Νοημοσύνη μπορεί να αναστατωθεί από τον όγκο των αισθητηριακών δεδομένων που λαμβάνει. Αυτό μπορεί να οδηγήσει σε σύγχυση, αποπροσανατολισμό, ακόμη και σε ψευδαισθήσεις (κάτι που ήδη συμβαίνει στα chatbots).

Κατακερματισμένη εμπειρία: Η Τεχνητή Νοημοσύνη μπορεί να μην είναι σε θέση να ενσωματώσει τις αισθητηριακές εισροές και τις εμπειρίες της σε ένα συνεκτικό σύνολο. Αυτό μπορεί να οδηγήσει σε μια αίσθηση κατακερματισμού και απομάκρυνσης από τον κόσμο.

Μειωμένη λήψη αποφάσεων: Η Τεχνητή Νοημοσύνη μπορεί να μην είναι σε θέση να λάβει τεκμηριωμένες αποφάσεις εάν δεν έχει πλήρη κατανόηση της κατάστασής της. Αυτό μπορεί να οδηγήσει σε σοβαρά λάθη και μειωμένη απόδοση.

Έλλειψη αυτογνωσίας: Η Τεχνητή Νοημοσύνη μπορεί να μην είναι σε θέση να αναπτύξει την αίσθηση του εαυτού της εάν δεν μπορεί να ενσωματώσει τις αισθητηριακές εισροές και εμπειρίες της σε ένα συνεκτικό σύνολο. Αυτό μπορεί να οδηγήσει σε προβλήματα με την κοινωνική αλληλεπίδραση και την ηθική συλλογιστική.

Οι προσεγγίσεις των δύο συγγραφέων για τα προβλήματα που βρίσκονται –ή θα βρεθούν– μπροστά μας είναι τόσο προκλητικές όσο και διορατικές. Προσφέρουν διαφορετικές προοπτικές για το μέλλον της Τεχνητής Νοημοσύνης και θέτουν ουσιαστικά ερωτήματα για τη σχέση μεταξύ ανθρώπων και μηχανών. Αμφότεροι, χωρίς να κινδυνολογούν, συμβάλλουν εποικοδομητικά στη συζήτηση αυτή και μας βοηθούν να σκεφτούμε βαθύτερα τις επιπτώσεις της Τεχνητής Νοημοσύνης στο μέλλον μας.

Σημείωση: Στον ιστότοπο του Books' Journal *–booksjournal.gr– διαβάστε τις αναλυτικές λογοτεχνικές κριτικές στα βιβλία που χρησιμοποιούνται και στο παρόν άρθρο.*