

Τα «εγώ» των μηχανών. Θα αποκτήσουν ποτέ οι μηχανές συνείδηση του εαυτού τους; Αν αυτό συμβεί θα είναι τελικά θετικό ή αρνητικό;

ΤΟ ΒΗΜΑ: science ΔΗΜΟΣΙΕΥΣΗ: 04/11/2012 05:45



Όλα καλά με την πρόοδο του Curiosity στον Αρη. Μόνο που δεν μπορούμε να το ρωτήσουμε «πώς τα βλέπεις τα πράγματα εκεί πάνω;»

Παρακολουθούμε τις προσπάθειες να χαρτογραφηθεί η επιφάνεια του πλανήτη Αρη, να καταγραφούν οι θερμοκρασίες στις διάφορες εποχές, να ακτινογραφηθεί το εσωτερικό, να καταλάβουμε αν έχει συχνά σεισμούς και αν τελικά μπορεί να υπάρξει ζωή εκεί επάνω. Το ταξίδι ήταν μακρινό, δύσκολο να το βγάλει πέρα ένας άνθρωπος, τα μηχανήματα τα πήγαν μια χαρά. Το μόνο κακό είναι ότι δεν μπορούμε να μιλήσουμε μαζί τους όπως θα το κάναμε αν βρισκόταν εκεί επάνω ένας άνθρωπος. Να τον ρωτήσουμε αυτό το στερεότυπο αλλά τόσο ουσιαστικό: Πώς τα βλέπεις εκεί πάνω τα πράγματα;

Ένας από τους διάσημους του χώρου, ο Ντάνιελ Ντένιτ, τον Μάρτιο του 2009 «φώναξε» μέσα από ένα άρθρο του: Είμαστε ρομπότ, που τα έκαναν ρομπότ, που τα έκαναν ρομπότ, που τα έκαναν ρομπότ... Αυτό σημαίνει ότι ο εγκέφαλος είναι μια καλοβαλμένη αποθήκη με σύμβολα και κανόνες που απεικονίζουν τον εξωτερικό μας κόσμο και αυτό γίνεται με τη βοήθεια του πλέγματος των νευρώνων του. Αρα αρκεί να αποθηκεύσουμε αρκετά στοιχεία σε μια αντίστοιχη κατασκευή και με τη βοήθεια μιας εξαιρετικής μηχανής αναζήτησης να έχουμε μια λειτουργία αντίστοιχη με την εγκεφαλική. Για χρόνια αυτή η τεχνητή νοημοσύνη ήταν η αυταπάτη πολλών επιστημόνων.

Ο «κευραυτός» της αλλαγής

Μια δολοφονική ερώτηση για την παραπάνω πεποίθηση διατυπώθηκε από τον Χιούμπερτ Ντρέιφους: Πώς θα κρίνει ένας υπολογιστής που έχει αποκτήσει μια εικόνα για την πραγματικότητα με βάση τα δεδομένα που του έδωσαν, αν κάτι ξαφνικά αλλάξει, τι ακριβώς αλλάζει και τι όχι στο κοσμοείδωλό του (εννοείται αν δεν του υποδείξουμε εμείς); Μια μηχανή μαθαίνει να αναγνωρίζει τους φοίνικες από το σχήμα της σκιάς τους. Όταν όμως η φωτογραφία έχει τραβηχτεί ακριβώς το μεσημέρι που δεν υπάρχουν σκιές, τι γίνεται; Ή, στην περίπτωση της γλώσσας, πώς θα μπορέσει το μηχάνημα να καταλάβει τη φράση «Αυτή η γυναίκα είναι παγόβουνο»;

Υπάρχει λοιπόν μια «φυλή» ερευνητών που έχει αντιληφθεί ότι σημαντικό ρόλο στο πώς λειτουργεί ο ανθρώπινος εγκέφαλος παίζει η αλληλεπίδρασή μας με το περιβάλλον και η γλώσσα. Ξεκινούν από τον Χάιντεγκερ και συντάσσονται μαζί του όταν κατηγορήσε τη δυτική φιλοσοφία ότι συνηγορεί στο να θεωρούμε τον εαυτό μας κάτι ξεχωριστό από τον περίγυρο. Όταν κάνουμε κάτι με ένα εργαλείο να συνειδητοποιήσουμε ότι είμαστε ένα με το εργαλείο. Είναι δηλαδή υπέρ μια ολιστικής αντίληψης της πραγματικότητας και αντίθετοι στην αντίληψη υποκείμενο-αντικείμενο (κάτι που η ανατολική φιλοσοφία και κυρίως οι άνθρωποι του Ζεν είχαν ήδη ως τρόπο ζωής εδώ και αιώνες). Διότι παραμορφώνει την εμπειρία και δεν θα μας επιτρέψει ποτέ να δημιουργήσουμε

πραγματικά ευφυείς μηχανές. Ο ανθρώπινος εγκέφαλος δεν είναι ένας υπολογιστής από βιολογικό υλικό.

«Κοινοότητες» υπολογιστών;

Βέβαια οι αντιλήψεις αυτές έχουν οδηγήσει μερικούς, στην προσπάθειά τους για έξυπνες μηχανές, να προτείνουν τη δημιουργία μιας «κοινωνίας μηχανών» που θα ανταλλάσσουν, λέει, εμπειρίες. Ας κρατήσουμε προς το παρόν από αυτό την αναγνώριση ότι η κοινωνική συναναστροφή και η επικοινωνία μεταξύ των ανθρώπων είναι δυνατόν να οξύνουν την ευφυΐα τους. Ένας από τους γνωστούς ερευνητές των θεμάτων των σχετικών με την ευφυΐα των μηχανών, ο Ιγκορ Αλεξάντρ, από το 2004 ακόμη είχε πει σε μια συνέντευξή του: *«Αν θέλουν να στείλουν έναν "συνειδητοποιημένο" ρομποτικό εξερευνητή στον Αρη, θα πρέπει αυτός: να φοβάται, να ευχαριστιέται με αυτό που κάνει, να έχει συνείδηση του τι είναι αυτό που βλέπει, να έχει συνείδηση του εαυτού του, μόνο που δεν θα πρέπει να έχει την ανησυχία για το αν θα τον φέρουμε πίσω»*. Αποδεικνύεται ότι ακόμη και για το 2012 ζητούσε πολλά. Αργεί ακόμη η κατασκευή ενός τέτοιου μοναχικού μηχανικού εξερευνητή. Προς το παρόν έχουμε ξεκαθαρίσει το ποια στοιχεία τού λείπουν, αλλά δεν έχουμε ξεκαθαρίσει ποια θέση μπορεί να πάρει τελικά στον κόσμο μας...

ΑΠΟ ΤΟΝ ΚΑΡΤΕΣΙΟ ΣΤΟ ΜΙΤ

Συνοπτική ιστορία της συνείδησης

Την εποχή του Καρτέσιου (1596-1650) αρχίζει η αναζήτηση, στους νεότερους χρόνους, της λύσης στο πρόβλημα τι είναι αυτό που λέμε νόηση. Ο Καρτέσιος επέμενε ότι ο καθένας μας αποτελείται από δύο ενότητες. Ένα υλικό σώμα και μια άυλη νόηση. Το πρόβλημα που αναπήδησε αμέσως από αυτό ήταν το πώς κάτι το μη υλικό, όπως η νόηση, κατάφερνε να έχει χειροπιαστά αποτελέσματα, όπως το να με οδηγήσει να κάνω μια κίνηση, έστω και απλή, π.χ. το να πιάσω ένα ποτήρι. Για αιώνες θεωρίες διατυπώνονταν, ακόμη και ότι μεσολαβεί ο Θεός ανάμεσα στα δύο, αλλά τίποτε δεν κρίθηκε ικανοποιητικό. Το 1949 ο Γκίλμπερτ Ράιλ έκανε λόγο για το φάντασμα που έβαζε μέσα στη μηχανή ο Καρτέσιος όταν σκεφτόταν με αυτόν τον τρόπο. Το λάθος του, κατά τον Ράιλ, ήταν ότι θεωρούσε και τη νόηση ως κάποιο «πράγμα», όπως ήταν και ο εγκέφαλος. Επρεπε να την κατατάξει σε άλλη κατηγορία. Στη συνέχεια εμφανίστηκε η θεωρία του Λειτουργισμού, όπου το πρόβλημα θα λυνόταν αν ταυτίζαμε τις νοητικές διαδικασίες με τις εγκεφαλικές. Και έτσι γεννήθηκε η ιδέα ότι μπορούμε να κατασκευάσουμε έναν εγκέφαλο που να έχει συνείδηση της ύπαρξής του, όπως και ο δικός μας, αλλά από άλλα υλικά, μη βιολογικά. Εμφανίστηκαν όμως άλλοι επιστήμονες που ισχυρίστηκαν ότι θα ήταν λάθος να ταυτίζουμε τον ανθρώπινο εγκέφαλο τον ηλεκτρονικό, γιατί κάτι τέτοιο αφήνει έξω το φαινόμενο της συνειδητότητας, της αίσθησης του "εγώ", του "είμαι εδώ και υπάρχω". Κλασικά αντεπιχειρήματα ήταν το κινεζικό δωμάτιο, όπου ο Τζον Σιρλ περιέγραψε ένα μηχανήμα που όταν του στέλνουμε ερωτήσεις στα κινεζικά ψάχνει σε ένα βιβλίο, βρίσκει τις απαντήσεις και τις δίνει αλλά δεν καταλαβαίνει τίποτε από κινεζικά. Ολα αυτά ήταν αντεπιχειρήματα στον λεγόμενο Φυσικαλισμό, στην ιδέα δηλαδή ότι δεν υπάρχει τίποτε που να μην εξηγείται με τους νόμους της Φυσικής και των άλλων θετικών επιστημών, άρα και ο εγκέφαλος αφού είναι από ύλη θα υπακούει στους ίδιους νόμους. Αλλά τότε, όπως αποκρυσταλλώθηκε με την έννοια των Ζόμπι, είμαστε απλά πλάσματα που ενεργούν χωρίς να έχουν συνείδηση του τι κάνουν; Τελευταίως ο Ντέιβιντ Τσάλμερς διετύπωσε τη θεωρία ότι ως προς το πρόβλημα της συνειδητότητας έχουμε τα «εύκολα», δηλαδή καταλαβαίνουμε γιατί καταλαβαίνουμε τι είναι ένα ολοκλήρωμα ή οι χρόνοι ενός ρήματος, αλλά υπάρχουν και τα «δύσκολα» (The Hard Problem), όπου δεν καταλαβαίνουμε ότι έχουμε και υποκειμενικές, δηλαδή συνειδητές από εμάς, εμπειρίες. Σήμερα, Σιρλ, Νταμάζιο, Πίνκερ, Ντένιτ και άλλοι προσπαθούν να δώσουν απαντήσεις στο πρόβλημα, ενώ η απλή τεχνητή νοημοσύνη που πίστευε ότι μπορούμε να φτιάξουμε ένα είδος τεχνητού εγκεφάλου έχει πέσει σε ανυποληψία, παρ' όλο που υποστηρίχθηκε με φανατισμό από τον Μάρβιν Μίνσκι στο MIT.