ΕΝΟΤΗΤΑ ΠΕΜΠΤΗ: ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

1. Μηχανή και νους

Πολλή συζήτηση γίνεται στην εποχή μας πάνω στο θέμα της τεχνητής νοημοσύνης. Υπάρχουν άραγε "σκεπτόμενες ή έξυπνες μηχανές", "ηλεκτρονικοί εγκέφαλοι", "τεχνητή νοημοσύνη" κτλ.; Έχει νόημα αυτό το ερώτημα ή το μόνο που κάνουμε είναι να αποδίδουμε -χωρίς να έχουμε το δικαίωμα- καθαρά ανθρώπινες ιδιότητες σε μηχανές; Πριν δούμε αν το ερώτημα νομιμοποιείται να απαντηθεί, θα πρέπει να καθορίσουμε τι εννοούμε με τους όρους "μηχανή" και "νόηση".

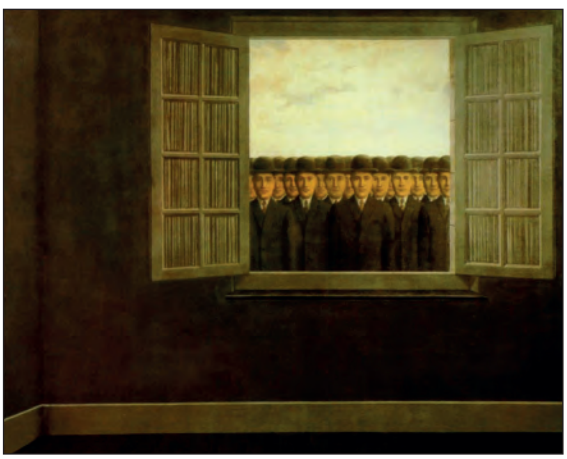
Τζιόρτζιο ντε Κίρικο, *Ο άσωτος υιός*, 1922, Μιλάνο, Δημοτικό Μουσείο Σύγχρονης Τέχνης. Στο πίνακα οι δυο μορφές που ανήκουν στον κόσμο της *Μεταφυσικής Ζωγραφικής* συμβολίζουν: τον πατέρα που φορά πέτρινο μανδύα και το γιο-ανδρείκελο ο οποίος είναι εντελώς τεχνητός, εκτός από το χέρι που ακουμπά στον ώμο του πατέρα του.

Αν, για παράδειγμα, το ουσιαστικό χαρακτηριστικό μιας μηχανής είναι ότι έχει **κατασκευαστεί** από τον άνθρωπο, τότε θα μπορούσαμε να ισχυριστούμε ότι ένας κλωνοποιημένος άνθρωπος δεν είναι άνθρωπος αλλά μηχανή. Λίγοι όμως σήμερα θα υποστήριζαν κάτι τέτοιο. Είναι άραγε τόσο ουσιαστικό να αποτελείται η μηχανή από μέταλλα και πλαστικό και όχι από σάρκα και οστά; Τι είναι, από την άλλη μεριά, ο άνθρωπος; Ο Αριστοτέλης όριζε τον άνθρωπο ως "έλλογο ον", δηλαδή ένα ον με **νοημοσύνη**. Ας υποθέσουμε ότι υπάρχουν εξωγήινα όντα με νοημοσύνη. Είναι αυτά άνθρωποι; Ας υποθέσουμε επίσης ότι υπάρχουν μηχανές με νοημοσύνη. Θα πρέπει να θεωρηθούν άνθρωποι, σύμφωνα με τον ορισμό μας, ή θα πρέπει να αλλάξουμε τον ορισμό; Όλα αυτά είναι σημαντικά ερωτήματα, που ίσως θα χρειαστεί να τα αντιμετωπίσουμε κάποτε.

Επειδή όλη η συζήτηση στις μέρες μας για το θέμα της τεχνητής νοημοσύνης εστιάζεται σε ένα ιδιαίτερο είδος μηχανών, στους ηλεκτρονικούς υπολογιστές ή εγκεφάλους, μπορούμε να μείνουμε μόνο σ' αυτές τις μηχανές. Ας περάσουμε όμως στο θέμα της νοημοσύνης. Με τον όρο "νοημοσύνη" εννοούμε ένα σύνολο ανθρώπινων ικανοτήτων που εστιάζονται κυρίως στον ανθρώπινο εγκέφαλο. Τέτοιες ικανότητες είναι η υπολογιστική και η λογική ικανότητα, η γλωσσική ικανότητα, η φαντασία, η κρίση κτλ. Ας υποθέσουμε ότι εμείς γνωρίζουμε ή πιστεύουμε βαθιά μέσα μας ότι έχουμε αυτές τις ικανότητες στις οποίες βασίζονται η γνώση και οι πράξεις μας. Μπορούμε όμως να γνωρίζουμε ότι τις έχουν και οι άλλοι συνάνθρωποί μας; Η απάντηση είναι ότι φαίνεται πως

μπορούμε, αφού και οι άλλοι ανθρωποι χρησιμοποιούν τις ίδιες λέξεις, εκφράσεις, χειρονομίες μ' εμάς και συμπεριφέρονται εν γένει με παρόμοιο μ' εμάς τρόπο σε αντίστοιχες περιπτώσεις. Έστω τώρα ότι είναι δυνατόν να κατασκευαστεί μια μηχανή που θα μιλά, θα επικοινωνεί, θα αντιδρά και θα συμπεριφέρεται όπως ένας άνθρωπος. Μπορούμε να πούμε ότι μια τέτοια μηχανή έχει νοημοσύνη ή, για να το θέσουμε αλλιώς, είναι δυνατόν μια μηχανή να **μιμείται** άριστα όλα τα εξωτερικά γνωρίσματα της νοημοσύνης μας, αλλά να μην έχει πραγματικά νοημοσύνη; Από την άλλη πλευρά, υπάρχει άλλος τρόπος να αναγνωρίσουμε αν ένα ον έχει νοημοσύνη πέρα από την οποιαδήποτε συμπεριφορά του; Αν δεν υπάρχει άλλος τρόπος, τότε μια μηχανή που έχει τη δυνατότητα να μιμείται ακριβώς τον άνθρωπο ως νοήμον ον θα πρέπει και αυτή να έχει νοημοσύνη.

2. Μπορεί η μηχανή να σκέφτεται; Μπορεί η τεχνητή νοημοσύνη να είναι ίδια με την ανθρώπινη;



Ρενέ Μαγκρίτ, *Ο μήνας του τρύγου*, 1959. Όπως και στην Γκολκόντα εμφανίζονται οι ανώνυμοι και χωρίς διαφορές μεταξύ τους άνθρωποι με καπέλο. Εμποδίζουν τη θέα από το παράθυρο και έτσι αποκτούν μια τρομακτική διάσταση παρά την παθητικότητά τους.

Μήπως, όταν μιλάμε για νοημοσύνη, είμαστε αναγκαστικά προσκολλημένοι σε κάτι αποκλειστικά ανθρώπινο; Αν είναι έτσι, τότε μπορούμε να ισχυριστούμε ότι μόνο ο άνθρωπος έχει νοημοσύνη και οτιδήποτε μοιάζει με την ανθρώπινη νοημοσύνη δεν είναι παρά απομίμησή της - όπως ο παπαγάλος μιμείται την ανθρώπινη ομιλία, αλλά δε μιλά πραγματικά. Αν αποδεχτούμε αυτόν τον ισχυρισμό, τότε δεν τίθεται θέμα συζήτησης· η νοημοσύνη είναι καθαρά ανθρώπινο φαινόμενο. Αν όμως πράγματι θέλουμε να συγκρίνουμε την ανθρώπινη νοημοσύνη με κάποιες λειτουργίες μηχανών, τότε το τι είναι νοημοσύνη θα πρέπει να εκφραστεί με όρους που δεν παραπέμπουν στον άνθρωπο αλλά σε ανεξάρτητες *και* -αν είναι δυνατόν- μετρήσιμες και συγκρίσιμες λειτουργίες. Για παράδειγμα, όσον αφορά την ικανότητα μαθηματικών υπολογισμών, ξέρουμε καλά ότι οι μηχανές όχι μόνο μπορούν και εκτελούν τέτοιους υπολογισμούς, αλλά το κάνουν ταχύτερα από τον άνθρωπο και χωρίς λάθη. Τα τελευταία χρόνια μάλιστα έχουν κατασκευαστεί υπολογιστές και έχουν σχεδιαστεί περίπλοκα προγράμματα που είναι σε θέση να νικήσουν ακόμη και τους καλύτερους σκακιστές.

Ο μεγάλος μαθηματικός και πρωτοπόρος σε θέματα σκεπτόμενων μηχανών Αλαν Τιούρινγκ έγραψε το 1950 ένα άρθρο πάνω σ' αυτό το ζήτημα. Προσπάθησε να βρει τρόπους, ώστε να καταστεί δυνατόν να συγκρίνει την ανθρώπινη νοημοσύνη με τις ικανότητες λειτουργίας μιας εξελιγμένης μηχανής. Κατέληξε ότι η μηχανή είναι σε θέση να πραγματοποιήσει όλες τις ανθρώπινες λειτουργίες που αποκαλούμε νοητικές. Το μόνο σημείο στο οποίο, κατά τη γνώμη του, διαφέρουν ο άνθρωπος και οι μηχανές που αυτός κατασκευάζει είναι η

εξωαισθητηριακή αντίληψη και τα **παραφυσικά φαινόμενα**, όπως η τηλεπάθεια, η πρόγνωση του μέλλοντος, η τηλεκίνηση κτλ. Ο Τιούρινγκ πιστεύει ότι τέτοια φαινόμενα -παρ' όλο που δεν μπορούμε να τα εξηγήσουμε επιστημονικά- παρατηρούνται στον άνθρωπο (ή σε κάποιους ανθρώπους), αλλά όχι στις μηχανές. Από την άλλη πλευρά όμως αυτά τα παραφυσικά φαινόμενα δεν είναι πλήρως επιβεβαιωμένα επιστημονικά και πολλοί επιστήμονες έχουν σοβαρές αμφιβολίες για την ύπαρξή τους. Αν λοιπόν δεν υπάρχουν τέτοια φαινόμενα, τότε, κατά τον Τιούρινγκ, δεν υφίσταται και διαφορά μεταξύ ανθρώπινης νόησης και νόησης της μηχανής.

Υπάρχει βέβαια η αντίρρηση ότι ο άνθρωπος έχει ψυχή, κάτι που φαίνεται πως δεν μπορούμε να δεχτούμε για τη μηχανή. Τι είναι όμως η ψυχή; Μπορούμε να τη δούμε ή να την εξετάσουμε επιστημονικά με κάποιον τρόπο; Αν οι υλιστές φιλόσοφοι, για τους οποίους μιλήσαμε παραπάνω, έχουν δίκιο, τότε οι νοητικές καταστάσεις δεν είναι τίποτα περισσότερο από ηλεκτροχημικές αντιδράσεις ή, όπως υποστηρίζουν οι λειτουργιστές, από ένα σύνολο λειτουργιών που επεξεργάζονται πληροφορίες του περιβάλλοντος και επιτρέπουν στον οργανισμό μας να αντιδρά με τον κατάλληλο τρόπο. Αν είναι έτσι, τότε ο εγκέφαλός μας δεν είναι παρά μια μηχανή και επομένως είναι δυνατόν να φτιάξουμε μηχανές που να μιμούνται τις λειτουργίες του εγκεφάλου. Σήμερα πολλοί επιστήμονες της τεχνητής νοημοσύνης δουλεύουν πάνω στις ομοιότητες του ανθρώπινου νευρικού συστήματος και της μετάδοσης πληροφορίας και ανάδρασης στις μηχανές.

Ο άνθρωπος όμως έχει αυτοσυνειδησία, συνείδηση των νοητικών λειτουργιών του, στην οποία, όπως είδαμε, στηρίζεται σε μεγάλο βαθμό η αίσθηση της προσωπικής του ταυτότητας. Είναι δυνατόν η μηχανή να έχει αυτή τη δυνατότητα αυτοσυνειδησίας; Από την άλλη πλευρά, ο άνθρωπος, αν και έχει άμεση συνείδηση των δικών του νοητικών λειτουργιών, δεν μπορεί να έχει άμεση συνείδηση των νοητικών λειτουργιών των άλλων ανθρώπων. Παρ' όλα αυτά δέχεται ότι και οι άλλοι άνθρωποι έχουν νου. Γιατί λοιπόν να μη δεχτεί το ίδιο και για τις μηχανές;

Η μηχανή είναι προγραμματισμένη από τον άνθρωπο και επιτελεί μόνο εκείνες τις λειτουργίες για τις οποίες είναι προγραμματισμένη. Δεν μπορεί η ίδια να αυτοπρογραμματίζεται. Αυτό είναι αλήθεια σε μεγάλο βαθμό για τις σημερινές μηχανές. Γίνονται όμως συστηματικά πειράματα για αυτοπρογραμματιζόμενες μηχανές και έχουν ήδη ξεκινήσει να κατασκευάζονται μηχανές που "μαθαίνουν" να αυτοπρογραμματίζονται.

Ο άνθρωπος μπορεί να κρίνει και να ελέγχει τη νοημοσύνη του. Μπορεί όμως να κάνει το ίδιο και η μηχανή; Εδώ τα πράγματα είναι πιο δύσκολα για τη μηχανή· και κανείς δεν μπορεί να μας εγγυηθεί ότι αύριο θα έχουμε μηχανές που θα έχουν αυτή την ικανότητα.

Τελικά, μπορεί η μηχανή να έχει νοημοσύνη όμοια με την ανθρώπινη; Από όσα αναφέραμε φαίνεται ότι το ερώτημα δεν μπορεί να απαντηθεί στη γενικότητά του. Εξαρτάται από το πώς θα εννοήσουμε τη μηχανή, τον άνθρωπο και τον νου του. Όσον αφορά συγκεκριμένες λειτουργίες που θεωρούνταν παλιά προνόμιο του ανθρώπου, σήμερα διαπιστώνουμε ότι η μηχανή μπορεί να τις επιτελέσει και μάλιστα αποτελεσματικότερα. Υπάρχουν ωστόσο και ανθρώπινες νοητικές λειτουργίες που είναι αμφίβολο αν ποτέ η μηχανή θα μπορέσει να τις επιτελέσει ή έστω να τις μιμηθεί. Αναφερό-

μαστε βέβαια σε ψυχικές καταστάσεις, όπως τα αισθήματα και τα συναισθήματα, και στην ικανότητα για αισθητική απόλαυση ή για ηθική κρίση, που φαίνεται πως μια μηχανή δεν μπορεί να διαθέτει. Το ερώτημα όμως στο οποίο δεν μπορούμε ακόμη να απαντήσουμε και το οποίο συναρπάζει τους φίλους της επιστημονικής φαντασίας είναι αν ένα ρομπότ -εξοπλισμένο με τον τελειότερο κατά το δυνατόν υπολογιστή- μπορεί να μετάσχει σε μια ανθρώπινη μορφή ζωής, να κοινωνικοποιηθεί και να ενεργήσει σαν ανθρώπινο υποκείμενο.

ΚΕΙΜΕΝΑ

1. "Κάνοντας μια αναδρομή στην τελευταία δεκαετία των ερευνών στην τεχνητή νοημοσύνη, μπορούμε να πούμε πως το βασικό ζήτημα που αναδεικνύεται είναι ότι, εφόσον η νοημοσύνη είναι υποχρεωτικά ενταγμένη σε μια κατάσταση, δεν μπορεί να διαχωριστεί από το υπόλοιπο του ανθρώπινου βίου. Η πεισματική απόρριψη αυτής της φαινομενικά προφανούς άποψης δεν μπορεί ωστόσο να αποδοθεί ολόκληρωτικά στην τεχνητή νοημοσύνη. Έχει ως αφετηρία τον διαχωρισμό που εισήγαγε ο Πλάτων μεταξύ της νόησης ή ελλογης ψυχής και του σώματος με τις δεξιότητες, τα αισθήματα και τις ορμές του. Ο Αριστοτέλης συνέχισε αυτή την άστοχη διχοτόμηση διαχωρίζοντας το θεωρητικό από το πρακτικό και ορίζοντας τον άνθρωπο ως έλλογο ζώο - λες και μπορεί κανείς να διαχωρίσει το λογικό του ανθρώπου από τις ζωώδεις ανάγκες και επιθυμίες του. Αν αναλογιστεί κανείς τη σπουδαιότητα των αισθησιοκινητικών δεξιοτήτων στην ανάπτυξη της ικανότητάς μας να αναγνωρίζουμε και να αντιμετωπίζουμε τα αντικείμενα, τον ρόλο των αναγκών και των επιθυμιών στη δόμηση όλων των κοινωνικών καταστάσεων ή, τέλος, το όλο πολιτισμικό υπόβαθρο της ανθρώπινης αυτοερμηνείας που ενέχεται στην απλή πρακτική γνώση τού πώς να διακρίνουμε και να χρησιμοποιούμε τις καρέκλες, η άποψη ότι μπορούμε απλώς να αγνοήσουμε αυτές τις πρακτικές γνώσεις κατά την τυποποίηση της νοητικής μας ικανότητας σε ένα πολύπλοκο σύστημα γεγονότων και κανόνων εμφανίζεται ανεδαφική".

(Hubert Dreyfus, *Τι δεν μπορούν ακόμη να κάνουν οι υπολογιστές; Κριτική της τεχνητής νοημοσύνης*, μτφρ. Π. Καρλέτσα, Πανεπιστημιακές Εκδόσεις Κρήτης, 3η έκδ., Ηράκλειο 1998, σ. 163-164)

2. "Σήμερα όχι μόνο η ψυχολογία αλλά και πολλοί συγγενείς επιστημονικοί κλάδοι αντιμετωπίζουν την αγωνία της γέννησης μιας μεγάλης διανοητικής επανάστασης. Και η επιτομή του όλου δράματος είναι η Τεχνητή Νοημοσύνη, η νέα συναρπαστική προσπάθεια κατασκευής σκεπτόμενων υπολογιστών. Ο θεμελιώδης στόχος αυτής της έρευνας δεν είναι η απλή μίμηση της νοημοσύνης ή η δημιουργία μιας έξυπνης μηχανής. Τίποτα απ' αυτά. Η τεχνητή νοημοσύνη επιδιώκει το αυθεντικό: μηχανές προικισμένες με νόηση, με την πλήρη και κυριολεκτική σημασία. Δεν πρόκειται για επιστημονική φαντασία, αλλά για πραγματική επιστήμη, βασισμένη σε μια θεωρητική αντίληψη που είναι τόσο βαθιά όσο και τολμηρή: στο ότι δηλαδή είμαστε κατά βάση και οι ίδιοι υπολογιστές. [...] Δεν υπάρχει καμιά αντίρρηση ότι οι μηχανές μπορούν να "αισθανθούν" το περι-

βάλλον τους, αν το μόνο που εννοούμε με αυτό είναι η ικανότητα διάκρισης - δηλαδή η παροχή διαφορετικών συμβολικών απαντήσεων σε διαφορετικές περιπτώσεις. Τα ηλεκτρονικά μάτια, τα ψηφιακά θερμόμετρα, οι αισθητήρες επαφής κ.ά. χρησιμοποιούνται ευρύτατα ως όργανα εισόδου σε κάθε είδους συσκευές, από τα ηλεκτρονικά παιχνίδια ως τα βιομηχανικά ρομπότ. Πολύ δύσκολα ωστόσο θα πιστεύαμε ότι τα συστήματα αυτά αισθάνονται πραγματικά κάτι όταν αντιδρούν στα εισερχόμενα ερεθίσματα. Αν και το πρόβλημα είναι γενικό, η διαίσθηση αυτή είναι σαφέστερη στην περίπτωση του πόνου: πολλά εξελιγμένα συστήματα μπορούν να ανιχνεύσουν εσωτερικές βλάβες, ακόμη και να πάρουν διορθωτικά μέτρα. Αισθάνονται όμως πραγματικά πόνο; Φαίνεται απίθανο, αλλά τι ακριβώς τους λείπει; Όσο περισσότερο σκέφτομαι αυτή την ερώτηση, τόσο περισσότερο πείθομαι πως δεν ξέρω καν τι σημαίνει (αυτό δε σημαίνει ότι θεωρώ πως δεν έχει νόημα). […] Το να μιλάμε για τη μανία και την απόλαυση ενός ρομπότ, για να μην αναφέρουμε τη σεξουαλική επιθυμία και την οργή, είναι κάτι που δεν το αποδεχόμαστε το ίδιο εύκολα με τον ηλεκτρονικό πόνο ή την ηλεκτρονική όραση [...] ίσως τα πάθη είναι κατά βάση σύνθετες οντότητες, με ένα συστατικό που έχει τις ρίζες του στη φυσιολογία των θηλαστικών, ένα άλλο στην ικανότητα για αισθήματα […] και ένα τρίτο γνήσια γνωστικό".

(John Haugeland, *Τεχνητή νοημοσύνη*, μτφρ. Σ. Ζαχαρίου, εκδ. Κάτοπτρο, Αθήνα 1985, σ. 11, 320-321)

ΕΡΩΤΗΣΕΙΣ

1. Ποιες κατά τη γνώμη σας είναι οι σημαντικότερες αντιρρήσεις που μας εμποδίζουν να θεωρήσουμε πως οι ηλεκτρονικοί υπολογιστές και τα ρομπότ μπορούν να εξομοιωθούν με τους ανθρώπους;

2. Να αναζητήσετε ιστορίες επιστημονικής φαντασίας στη λογοτεχνία και τον κινηματογράφο που τονίζουν τις ομοιότητες μεταξύ ανθρώπων και ρομπότ.

3. Είναι δυνατόν να μπορούμε να μιλήσουμε για "βούληση" ή πολύ περισσότερο για "ελευθερία βούλησης" ακόμα και του πιο εξελιγμένου υπολογιστή; Αιτιολογήστε την απάντησή σας.