

Σπύρος Μανουσέλης

Η από μηχανής... νοημοσύνη

Πόσο εφικτή επιστημονικά και επιθυμητή τεχνολογικά είναι η προσπάθεια να αναπαραχθεί μέσω μηχανών μια πλήρης και λειτουργική προσομοίωση του ανθρώπινου νου;

Επειδή είναι αδύνατο να εξαντληθούν όλες οι πτυχές ενός τόσο σοβαρού επιστημονικού και κοινωνικού ζητήματος στο πλαίσιο ενός άρθρου, σκεφτήκαμε ότι θα είχε ίσως ενδιαφέρον να εστιάσουμε σε ορισμένα επίμαχα ζητήματα σχετικά με την Τ.Ν. και τις αδιαφανείς τεχνολογικές εφαρμογές της, ενώ στο επόμενο άρθρο μας θα παρουσιάσουμε τις εξελίξεις στη Ρομποτική.

Η ανάγκη για συστηματική ενημέρωση και ευρύτατη κοινωνική διαβούλευση γύρω από αυτά τα θέματα είναι σήμερα επιτακτική επειδή όλοι αναγνωρίζουν ότι βρισκόμαστε σε μια

ιδιαίτερα κρίσιμη καμπή του τεχνολογικού-επιστημονικού πολιτισμού μας. Και οι τελικές αποφάσεις για τις εφαρμογές της τεχνοεπιστήμης οφείλουν να λαμβάνονται από τους πολίτες και όχι βέβαια από τους ειδικούς.

Πράξη γέννησης του ερευνητικού προγράμματος της Τεχνητής Νοημοσύνης (T.N.) θεωρείται από πολλούς ιστορικούς της επιστήμης η δημοσίευση του άρθρου «Υπολογιστικές μηχανές και νοημοσύνη» που έγραψε το 1950 ο μεγάλος Βρετανός μαθηματικός Άλαν Τιούρινγκ (Alan Turing).

Σε αυτό το κείμενο ο πατέρας της σύγχρονης υπολογιστικής επιστήμης εξετάζει εξαντλητικά το πανάρχαιο ερώτημα «Μπορούν οι μηχανές να σκέφτονται;» και προτείνει μια εμπειρική μέθοδο, μια δοκιμασία που μας επιτρέπει να αποφασίζουμε αν και πότε μια υπολογιστική μηχανή μπορεί να επιδεικνύει νοημοσύνη ανάλογη με αυτή των ανθρώπων (βλ. σχετικό πλαίσιο).

Τη μεγάλη επιρροή που άσκησε το επονομαζόμενο σήμερα «τεστ Τιούρινγκ» μάς την αποκαλύπτει και ο κλασικός ορισμός της τεχνητής νοημοσύνης που διατυπώθηκε έπειτα από πολλά χρόνια:

«T.N. είναι ο κλάδος της επιστήμης των υπολογιστών (computer science) που ασχολείται με τον σχεδιασμό νοημών συστημάτων, συστημάτων δηλαδή τα οποία επιδεικνύουν μια συμπεριφορά που αν τη συναντούσαμε σε έναν άνθρωπο, θα τον χαρακτηρίζαμε νοήμονα». Ο ορισμός ακολουθεί κατά γράμμα τη στρατηγική που πρότεινε ο Τιούρινγκ: δεν ορίζει τι είναι η νόηση, συγκρίνει απλώς τις εξωτερικές εκδηλώσεις της. Είναι δηλαδή συμπεριφοριστικός και ανθρωποκεντρικός.

Σκεπτόμενες μηχανές, ένα ανεκπλήρωτο όνειρο

Τη δεκαετία του 1960, χάρη στην ανάδυση της Τεχνητής Νοημοσύνης και της Γνωσιακής Επιστήμης, άρχισε να διαφαίνεται, πρώτη φορά, η προοπτική της δημιουργίας μιας νέας επιστήμης-τεχνολογίας των νοητικών φαινομένων. Οι δύο νέοι, ιδιαίτερα δυναμικοί διεπιστημονικοί κλάδοι φιλοδοξούσαν να συνθέσουν σε ένα ενιαίο εξηγητικό σχήμα τις επιμέρους γνώσεις και τεχνολογίες που είχαν αρχίσει να συσσωρεύονται με εκρηκτικούς ρυθμούς γύρω από τη δομή και τη λειτουργία του ανθρώπινου νου.

Τα επόμενα χρόνια, η «νέα επιστήμη και τεχνολογία του νου» θα συγκροτηθεί πάνω σε δύο θεμελιώδεις –αν και κάθε άλλο παρά προφανείς– παραδοχές:

- 1) κάθε νοητικό φαινόμενο είναι γνωσιακό φαινόμενο και
- 2) κάθε γνωσιακό φαινόμενο δεν είναι παρά ένα υπολογιστικό φαινόμενο.

Ετσι εξηγείται γιατί, από τα πρώτα βήματά τους μέχρι σήμερα, τόσο η Τ.Ν. όσο και οι άλλες γνωσιακές επιστήμες υιοθέτησαν το «υπολογιστικό μοντέλο» για την κατανόηση του νου και ανήγαγαν τους ηλεκτρονικούς υπολογιστές σε βασικό εργαλείο ελέγχου της αλήθειας των θεωριών τους. Η παραδοχή όμως ότι μια υπολογιστική μηχανή (π.χ. ένας ψηφιακός ή αναλογικός υπολογιστής) μπορεί, κατ' αρχήν, να αναπαράγει όλες τις λειτουργίες ενός βιολογικού εγκεφάλου, οδηγεί στο ταυτολογικό συμπέρασμα ότι ο εγκέφαλός μας δεν είναι

τίποτε άλλο από μια υπολογιστική μηχανή!

Η θεμελιώδης και, μέχρι πρόσφατα, κυρίαρχη υπόθεση εργασίας όσων εργάζονται στο φιλόδοξο πρόγραμμα κατασκευής ευφυών μηχανών ήταν η παραδοχή πως κάθε ανθρώπινη νοητική συμπεριφορά είναι μια υπολογιστική διεργασία: αλγοριθμικός χειρισμός συμβόλων βάσει κανόνων. Πρόκειται για το περίφημο και τεχνολογικά παραγωγικότατο μοντέλο της «υπολογιστικής θεωρίας της νόησης» (computationalism).

Συνέπεια αυτής της παραδοχής, ότι δηλαδή κάθε νοητική διεργασία είναι ουσιαστικά μια υπολογιστική διεργασία, ήταν πως το επιστημονικό πρόγραμμα της Τ.Ν. προϋποθέτει -και συνεπώς επιβάλλει- την «καινοφανή» ιδέα ότι σε κάθε υπολογιστική μηχανή υπάρχει μια εγγενής και απόλυτη διάκριση του λογισμικού (software) από το υλισμικό (hardware) τμήμα της μηχανής, αδιάφορο αν η μηχανή αυτή είναι βιολογική ή ηλεκτρονική.

Υπό αυτήν ακριβώς την έννοια, η αποδοχή του υπολογιστικού μοντέλου συνεπάγεται την εξαΰλωση του νου! Μόνον αν οι νοητικές λειτουργίες αποκοπούν από το βιολογικό τους υπόστρωμα μπορούν να μετακομίσουν ελεύθερα από τους βιολογικούς στους ηλεκτρονικούς εγκεφάλους.

Εχει περάσει πάνω από μισό αιώνας συστηματικών ερευνών και εντυπωσιακών εφαρμογών της Τ.Ν. και όμως το αποφασιστικό ερώτημα αν οι μηχανές μπορούν ή αν θα μπορέσουν κάποτε να σκέφτονται παραμένει αναπάντητο. Για την ακρίβεια, δεν υπάρχει ούτε καν συμφωνία μεταξύ των ειδικών για το πώς θα έπρεπε να εννοηθούν οι βασικοί όροι του ερωτήματος: τι ακριβώς είναι μια «μηχανή» και τι σημαίνει να διαθέτει «νοημοσύνη» ή «σκέψη»;

Όλα εξαρτώνται από το πώς ορίζει κανείς αυτές τις έννοιες και από το πόσο αυστηρά ή ευέλικτα είναι τα επιστημονικά κριτήρια που υιοθετεί. Για παράδειγμα, ανάλογα με το μοντέλο της νοημοσύνης που υιοθετεί κανείς, οι σύγχρονοι πανίσχυροι υπολογιστές μπορεί κάλλιστα να θεωρηθούν εντελώς «ανόητοι» ενώ, αντίθετα, μια απλή αριθμομηχανή μπορεί να χαρακτηριστεί «νοήμων».

Τα όρια της εξαϋλωμένης «σκέψης» των υπολογιστών

Ουδείς αμφιβάλει σήμερα για τις προόδους που έχουν σημειωθεί στην επιστήμη των υπολογιστών, και ειδικότερα για τις αποφασιστικής σημασίας εφαρμογές της Τ.Ν. Όμως, αυτές οι εντυπωσιακές τεχνολογικές εξελίξεις θα οδηγήσουν αργά ή γρήγορα και αναπότρεπτα στην εκπλήρωση του τεχνολογικού ονείρου της δημιουργίας σκεπτόμενων μηχανών; Σε αυτό το σημείο οι απόψεις των ειδικών δίστανται.

Οι οπαδοί της «σκληρής Τ.Ν.» (hard A.I.) υποστηρίζουν ότι η δημιουργία σκεπτόμενων μηχανών είναι αναπόφευκτη. Δεδομένου ότι τα υπολογιστικά προγράμματα διαρκώς θα βελτιώνονται, θα καταφέρουμε σύντομα να προσομοιώσουμε τις βασικές ανθρώπινες νοητικές λειτουργίες. Και μόλις αυτά τα υπολογιστικά προγράμματα αποκτήσουν επαρκή πολυπλοκότητα, ανάλογη με αυτήν των ανθρώπινων νοητικών λειτουργιών, τότε η νοημοσύνη του υπολογιστή θα αναδυθεί αυτομάτως.

Ωστόσο, η εμφανής αδυναμία υλοποίησης των στόχων του σκληρού προγράμματος οδήγησε σταδιακά σε μια πιο «ήπια» και «πραγματοστική» εκδοχή της T.N. (soft A.I.). Αυτή υποστηρίζει ότι η αναζήτηση και η έρευνα της T.N. εκτός από την πρακτική χρησιμότητά της αποτελεί και ένα ιδιαίτερα χρήσιμο επιστημολογικό εργαλείο.

Με άλλα λόγια, τα υπολογιστικά μοντέλα αποτελούν το ιδανικό εργαλείο για τον έλεγχο, δηλαδή για την επαλήθευση ή τη διάψευση των διαφορετικών προσεγγίσεων και εξηγήσεων των νοητικών φαινομένων. Έτσι, για παράδειγμα, μπορούμε να αντιπαραβάλουμε το κλασικό υπολογιστικό μοντέλο του νου με το πιο πρόσφατο «συνδεδιστικό» μοντέλο του νου, το οποίο βασίζεται στην αρχιτεκτονική των βιολογικών εγκεφάλων και όχι στο γραμμικό υπολογιστικό μοντέλο επεξεργασίας των πληροφοριών.

Η εμμονή, τις δύο τελευταίες δεκαετίες, στο πρόγραμμα της «σκληρής T.N.» δικαιολογείται μόνον από τις εντυπωσιακές τεχνολογικές εφαρμογές της. Σήμερα, πάντως, θεωρείται επαρκώς επιβεβαιωμένο το γεγονός ότι κάθε αφηρημένη και αμιγώς υπολογιστική προσέγγιση του ανθρώπινου νου, μολονότι οδηγεί στη δημιουργία αξιοθαύμαστων και ιδιαίτερα ισχυρών μηχανών, δεν καταφέρνει να τις προικίσει με ανθρώπινου τύπου νοημοσύνη ούτε βέβαια με κάποια μορφή αυτόνομης σκέψης.

Αυτή η αποτυχία της «σκληρής T.N.» να δημιουργήσει μηχανές με πραγματική νοημοσύνη δεν σημαίνει βέβαια ότι η ανθρώπινη γλώσσα, η συνείδηση, η φαντασία, η έλλογη σκέψη, τα συναισθήματα, όλες αυτές οι «ανώτερες» και αινιγματικές εκδηλώσεις του ανθρώπινου νου είναι ανεξήγητα ή υπερφυσικά φαινόμενα, δημιουργήματα κάποιων σκοτεινών δυνάμεων (της άυλης ψυχής, του Θεού κ.ο.κ.). Γιατί, όπως θα δούμε στο επόμενο άρθρο μας, η νέα επιστήμη των αυτομάτων μάς επιτρέπει να δημιουργούμε εντυπωσιακά βιοκυβερνητικά ομοιώματά μας.

Το τεστ Τιούρινγκ για την «ευφυΐα» των μηχανών

Στο διάσημο κείμενο με τίτλο «Υπολογιστικές μηχανές και νοημοσύνη» (Computing Machinery and Intelligence), που έγραψε το 1950, ο Άλαν Τιούρινγκ εξετάζει εξαντλητικά το ερώτημα αν «οι μηχανές μπορούν να σκέφτονται». Καταλήγει μάλιστα να προτείνει μια φαινομενικά παιγνιώδη δοκιμασία που, όπως ήλπιζε, θα μπορούσε να προσφέρει ένα ασφαλές κριτήριο για να αποφασίζουμε αν οι υπολογιστικές μηχανές -που είχαν μόλις αρχίσει να εμφανίζονται- διαθέτουν ή όχι ανθρώπινη νοημοσύνη.

Το κείμενο του Τιούρινγκ περιλαμβάνεται και σχολιάζεται διεξοδικά στο εξαιρετικό βιβλίο των Ντάγκλας Χόφσταντερ (Douglas Hofstadter) και Ντάνιελ Ντέννετ (Daniel Dennett) «Το Εγώ της Νόησης», που κυκλοφορεί στα ελληνικά από τις εκδ. Κάτοπτρο.

Η όλη δοκιμασία του Τιούρινγκ βασίζεται σε ένα πανούργο παιχνίδι μίμησης που παίζεται από τρεις παίκτες. Οι δύο από αυτούς είναι άνθρωποι, ενώ ο τρίτος είναι ένας ισχυρός ψηφιακός υπολογιστής. Ο πρώτος αναλαμβάνει τον ρόλο του «ανακριτή»: είναι ένας άνθρωπος που θέτει ερωτήματα στους άλλους δύο προκειμένου να καταλάβει και να αποφασίσει ποιος από τους άλλους δύο παίκτες είναι η υπολογιστική μηχανή.

Ο δεύτερος άνθρωπος οφείλει να απαντά πάντοτε ειλικρινά για να βοηθήσει τον ανακριτή,

ενώ ο τρίτος, ένας κατάλληλα προγραμματισμένος υπολογιστής, μιμείται όσο καλύτερα μπορεί την ανθρώπινη λεκτική συμπεριφορά με σκοπό να ξεγελάσει τον ανακριτή.

Το «μαύρο κουτί» ψεύδεται

Στη δοκιμασία Τιούρινγκ η ικανότητα χρήσης της ανθρώπινης γλώσσας αποτελεί το επαρκές και, σε τελευταία ανάλυση, το αποφασιστικό κριτήριο για να διακρίνουμε αν πράγματι ένα σύστημα (ανθρώπινο ή τεχνητό) διαθέτει νοημοσύνη. Συνεπώς, πρόκειται για ένα σαφώς εξωγενές συμπεριφοριστικό κριτήριο που αδιαφορεί παντελώς για το υλικό υπόστρωμα ή τις δομές που παράγουν τη νοήμονα συμπεριφορά.

Είτε πρόκειται για ανθρώπινους βιολογικούς εγκεφάλους είτε για προγράμματα ηλεκτρονικών υπολογιστών, αυτό που ενδιαφέρει είναι μόνον η γλωσσική ικανότητα που επιδεικνύουν και όχι η εγκεφαλική ή ηλεκτρονική δομή που την παράγει, η οποία θεωρείται ένα αδιαφανές «μαύρο κουτί». Με αυτήν ακριβώς την έννοια, η δοκιμασία του Τιούρινγκ εξετάζει αποκλειστικά και μόνο τις γλωσσικές απαντήσεις των παικτών στα γλωσσικά ερεθίσματα-ερωτήσεις που δέχονται.

Σύμφωνα με αυτή τη θεμελιώδη –αλλά μέχρι σήμερα ανεπιβεβαίωτη– παραδοχή του Αλαν Τιούρινγκ και των σύγχρονων οπαδών της σκληρής εκδοχής της Τεχνητής Νοημοσύνης (hard A.I.), η ανθρώπινη σκέψη δεν είναι τίποτε άλλο και τίποτα περισσότερο από λογικός υπολογισμός, δηλαδή χειρισμός συμβόλων βάσει κανόνων, άρα δεν εξαρτάται ούτε

περιορίζεται από το υλικό υπόστρωμα (hardware) στο οποίο πραγματοποιούνται αυτοί οι υπολογισμοί, και συνεπώς μπορεί κάλλιστα να εκδηλώνεται τόσο σε βιολογικούς όσο και σε ηλεκτρονικούς εγκεφάλους!

Πρόκειται για την αφηρημένη, υπολογιστική και συμπεριφοριστική «σκληρή» εκδοχή της Τ.Ν. που έχει δεχτεί πλήθος επικρίσεων και διαψεύσεων, τόσο αμιγώς επιστημονικών όσο και φιλοσοφικών. Από επιστημονική άποψη, είναι αδύνατον, π.χ., να παραβλέψουμε ή να υποτιμάμε τις βιολογικές, εγκεφαλικές προϋποθέσεις της νοημοσύνης. Ενώ από φιλοσοφική άποψη, δεν μπορούμε να αποφασίζουμε εάν ένα τεχνητό σύστημα διαθέτει ή όχι νοημοσύνη βασιζόμενοι αποκλειστικά και μόνο στην οπτική γωνία ενός τρίτου προσώπου (του ανακριτή) και όχι, π.χ., σε αυτήν του πρώτου προσώπου: εγώ σκέπτομαι, άρα είμαι νοήμων!

Το βέβαιο πάντως είναι ότι, από τη σκοπιά του πρώτου προσώπου, οι εκδηλώσεις της νόησης προϋποθέτουν την παρουσία συνείδησης και κατανόησης, των τυπικών δηλαδή γνωρισμάτων του ανθρώπινου νου.

[Πηγή: Εφημερίδα Των Συντακτώ ν, 07.03.2015](#)