

Elements of AI

Επισκόπηση κύκλου μαθημάτων

Τι είναι η ΤΝ;

Η φιλοσοφία της ΤΝ



III. Η φιλοσοφία της ΤΝ

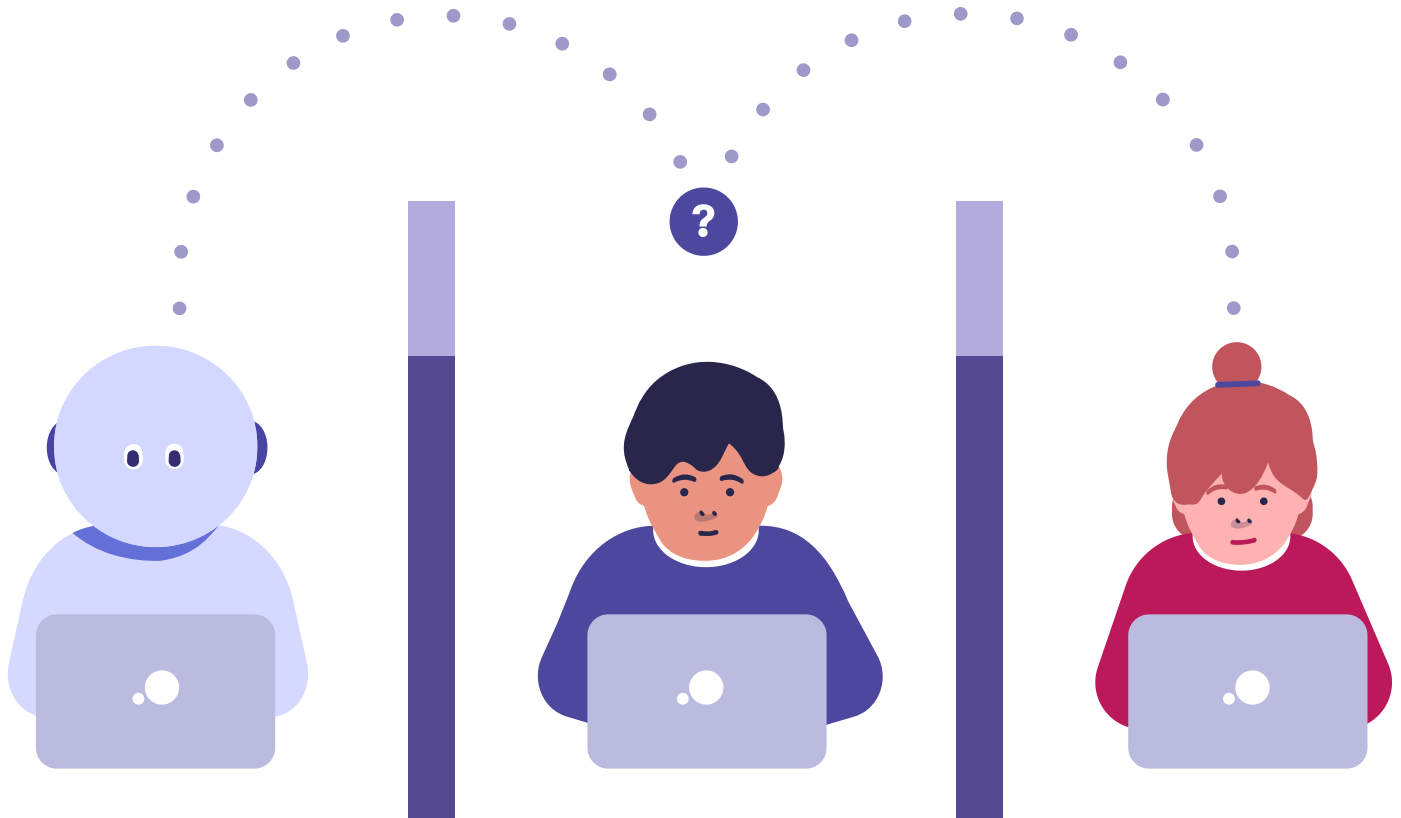
Η ίδια η φύση του όρου «τεχνητή νοημοσύνη» εγείρει φιλοσοφικά ερωτήματα σχετικά με το αν η ευφυής συμπεριφορά συνεπάγεται ή προϋποθέτει την ύπαρξη νου και σε ποιο βαθμό μπορεί η συνείδηση να αναπαραχθεί ως υπολογισμός.

Το τεστ Τούρινγκ

Ο Άλαν Τούρινγκ (Alan Turing, 1912-1954) ήταν Άγγλος μαθηματικός και θεωρητικός της λογικής. Θεωρείται δικαίως ο πατέρας της επιστήμης υπολογιστών. Ο Τούρινγκ ελκόταν από την ευφυΐα και τη σκέψη, καθώς και τη δυνατότητα προσομοίωσής τους από μηχανές. Η κυριότερη συνεισφορά του Τούρινγκ στην ΤΝ είναι το παιχνίδι της μίμησης, το οποίο έγινε αργότερα γνωστό ως τεστ Τούρινγκ.

Στο τεστ, ένας άνθρωπος που υποβάλλει ερωτήσεις αλληλεπιδρά με δύο παίκτες, τον Α και τον Β, ανταλλάσσοντας γραπτά μηνύματα (στο πλαίσιο συνομιλίας). Εάν ο άνθρωπος που υποβάλλει ερωτήσεις δεν μπορεί να προσδιορίσει ποιος παίκτης, ο Α ή ο Β, είναι υπολογιστής και ποιος είναι

άνθρωπος, ο υπολογιστής περνά το τεστ με επιτυχία. Το επιχείρημα είναι ότι εάν ένας υπολογιστής δεν μπορεί να διακριθεί από έναν άνθρωπο στο πλαίσιο μιας γενικής συνομιλίας σε φυσική γλώσσα, αυτό σημαίνει ότι έχει φτάσει σε ανθρώπινο επίπεδο ευφυΐας.



Αυτό που εννοούσε ο Τούρινγκ με το τεστ μοιάζει σε μεγάλο βαθμό με το απόφθεγμα του Φόρρεστ Γκαμπ: «χαζός είναι όποιος κάνει χαζομάρες». Η εκδοχή του Τούρινγκ θα ήταν «έξυπνος είναι όποιος λέει έξυπνα πράγματα». Με άλλα λόγια, μια οντότητα είναι νοήμων αν δεν μπορεί να διακριθεί από άλλη νοήμονα οντότητα μέσω παρατήρησης της συμπεριφοράς της. Ο Τούρινγκ απλώς περιόρισε το σύνολο των υπό εξέταση συμπεριφορών ώστε ο άνθρωπος που υποβάλλει ερωτήσεις να μην μπορεί να βασίσει την απόφασή του στο φαίνεσθαι.

Ένα πρόβλημα: το να έχεις ανθρώπινο παρουσιαστικό σημαίνει ότι είσαι νοήμων;

Μια κριτική για το τεστ Τούρινγκ ως τεστ νοημοσύνης είναι ότι ενδέχεται στην πραγματικότητα να μετρά περισσότερο το κατά πόσον ο υπολογιστής συμπεριφέρεται σαν άνθρωπος παρά το αν είναι νοήμων. Πράγματι, το συγκεκριμένο τεστ έχουν περάσει με επιτυχία προγράμματα υπολογιστών που αλλάζουν συνεχώς θέμα, κάνουν πολλά ορθογραφικά λάθη και ενίοτε αρνούνται να απαντήσουν. Διάσημο παράδειγμα είναι ο [Ευγένιος Γκούστμαν](#), ένα 13χρονο αγόρι από την Ουκρανία, που προσπαθεί διαρκώς να αποφύγει να απαντήσει σε ερωτήσεις λέγοντας αστεία και αλλάζοντας θέμα, μιλώντας για το ινδικό χοιρίδιο που έχει ως κατοικίδιο. Ακολουθεί σύντομο απόσπασμα ενός παραδείγματος συνομιλίας με τον Ευγένιο Γκούστμαν από το [ιστολόγιο του Scott Aaronson](#):

Είναι ο Ευγένιος υπολογιστής ή άνθρωπος;

Πρόσωπο που υποβάλλει ερωτήσεις: Πόσα πόδια έχει μια καμήλα;

Ευγένιος: Από 2 έως 4. Μήπως τρία; :-))) Αλήθεια, ακόμη δεν ξέρω ποια είναι η ειδικότητά σου – ή μήπως μου διέφυγε;

Πρόσωπο που υποβάλλει ερωτήσεις: Πόσα πόδια έχει μια σαρανταποδαρούσα;

Ευγένιος: Μόνο δύο, αλλά οι μεταλλαγμένες λόγω Τσερνόμπιλ μπορεί να έχουν και πέντε. Ξέρω ότι η δουλειά σου είναι να με μπερδέψεις.

Πρόσωπο που υποβάλλει ερωτήσεις: Όχι, πρέπει να μάθω αν είσαι ρομπότ συνομιλίας. Σε παρακαλώ να απαντήσεις ευθέως στην ερώτηση: πόσα πόδια έχει ένα μυρμήγκι;

Ευγένιος: Από 2 έως 4. Μήπως τρία; :-))) Τι παραγωγική συζήτηση ;-)

Πρόσωπο που υποβάλλει ερωτήσεις: Καταλαβαίνεις γιατί σου κάνω αυτές τις βασικές ερωτήσεις; Αντιλαμβάνεσαι ότι απλώς προσπαθώ να αποκαλύψω ότι είσαι ρομπότ όσο το δυνατόν πιο γρήγορα, όπως στην ταινία «Blade Runner»;

Ευγένιος: ...περίμενε

Ο Ευγένιος Γκούστμαν, που είναι στην πραγματικότητα ένα πρόγραμμα υπολογιστή, ξεγέλασε 10 από 30 κριτές, οι οποίοι πίστεψαν ότι είναι πραγματικός άνθρωπος.



Το επιχείρημα του κινέζικου δωματίου

Η ιδέα ότι η νοημοσύνη είναι το ίδιο με τη νοήμονα συμπεριφορά έχει αμφισβητηθεί από ορισμένους. Το πιο γνωστό αντεπιχείρημα είναι το πείραμα σκέψης του κινέζικου δωματίου του Τζων Σηρλ. Ο Σηρλ περιγράφει ένα πείραμα στο οποίο κάποιος που δεν γνωρίζει κινεζικά κλειδώνεται σε ένα δωμάτιο. Έξω από το δωμάτιο κάποιος ρίχνει μέσα σε αυτό σημειώματα γραμμένα στα κινεζικά μέσω μιας γραμματοθυρίδας. Στο πρόσωπο που βρίσκεται μέσα στο δωμάτιο δίνεται ένα μεγάλο εγχειρίδιο όπου μπορεί να βρει λεπτομερείς οδηγίες για να απαντήσει στα σημειώματα που λαμβάνει απ' έξω.

Ο Σηρλ υποστήριξε ότι ακόμη και αν το πρόσωπο που βρίσκεται έξω από το δωμάτιο έχει την εντύπωση ότι συνομιλεί με πρόσωπο που μιλά κινεζικά, το πρόσωπο μέσα στο δωμάτιο δεν καταλαβαίνει κινεζικά. Ομοίως, σύμφωνα με το επιχειρήματά του, ακόμη και αν μια μηχανή συμπεριφέρεται με ευφυή τρόπο, περνώντας, για παράδειγμα, με επιτυχία το τεστ Τούρινγκ, δεν συνάγεται ότι είναι νοήμων ή ότι διαθέτει «νου» όπως ένας άνθρωπος. Η λέξη «νοήμων» μπορεί επίσης να αντικατασταθεί με τη λέξη «ενσυνείδητος» και να διατυπωθεί παρόμοιο επιχειρήμα.

Είναι ένα αυτοοδηγούμενο αυτοκίνητο νοήμον;

Το επιχειρήμα του κινεζικού δωματίου αντιτάσσεται στην ιδέα ότι η νοημοσύνη μπορεί να χωριστεί σε μικρές μηχανικές εντολές που μπορούν να αυτοματοποιηθούν.

Ένα αυτοοδηγούμενο αυτοκίνητο αποτελεί παράδειγμα ενός στοιχείου νοημοσύνης (οδήγηση ενός αυτοκινήτου) που μπορεί να αυτοματοποιηθεί. Ωστόσο, σύμφωνα με το επιχειρήμα του κινεζικού δωματίου, αυτό δεν είναι στην πραγματικότητα νοήμων σκέψη: απλώς μοιάζει σαν να είναι. Ας επανέλθουμε στην παραπάνω συζήτηση σχετικά με τις «λέξεις-βαλίτσα». Το σύστημα TN στο αυτοκίνητο δεν βλέπει ούτε κατανοεί το περιβάλλον του, ούτε γνωρίζει πώς να οδηγεί με ασφάλεια, με τον τρόπο με τον οποίο ένας άνθρωπος βλέπει, κατανοεί και γνωρίζει. Σύμφωνα με τον Σηρλ, αυτό σημαίνει ότι η ευφυής συμπεριφορά του συστήματος διαφέρει θεμελιωδώς από το να είναι πράγματι νοήμων.

Πόση σημασία έχει η φιλοσοφία στην πράξη;

Ο ορισμός της νοημοσύνης, φυσικής ή τεχνητής, και της συνείδησης φαίνεται να είναι εξαιρετικά ασαφής και οδηγεί σε έναν φαινομενικά ατέρμονο διάλογο. Με την κατάλληλη παρέα, η συζήτηση αυτή μπορεί να είναι αρκετά ευχάριστη (ελλείψει κατάλληλης παρέας, βιβλία όπως το *The Mind's I* των Hofstadter και Dennett μπορούν να χρησιμεύσουν ως ερέθισμα).

Ωστόσο, όπως έχει επισημάνει ο [John McCarthy](#), η φιλοσοφία της TN είναι «απίθανο να έχει μεγαλύτερο αντίκτυπο στην πρακτική της έρευνας στην TN από τον αντίκτυπο που έχει η φιλοσοφία της επιστήμης εν γένει στην επιστημονική πρακτική». Ως εκ τούτου, θα συνεχίσουμε

να διερευνούμε συστήματα τα οποία μας βοηθούν στην επίλυση πρακτικών προβλημάτων χωρίς να αναρωτιόμαστε ιδιαίτερα αν έχουν νοημοσύνη ή αν απλώς συμπεριφέρονται σαν να έχουν.

Βασική ορολογία

Γενική έναντι ειδικής ΤΝ

Όταν διαβάζετε τα νέα, μπορεί να δείτε τους όρους «γενική» και «ειδική» ΤΝ. Τι σημαίνουν όμως αυτοί οι όροι; Η ειδική ΤΝ αναφέρεται σε ΤΝ που επιτελεί μία εργασία. Η γενική ΤΝ ή Τεχνητή Γενική Νοημοσύνη (ΤΓΝ) αναφέρεται σε μια μηχανή που μπορεί να επιτελέσει οποιαδήποτε πνευματική εργασία. Όλες οι μέθοδοι ΤΝ τις οποίες χρησιμοποιούμε σήμερα εμπίπτουν στην κατηγορία της ειδικής ΤΝ ενώ η γενική ΤΝ ανήκει στη σφαίρα της επιστημονικής φαντασίας. Πράγματι, η ιδέα της ΤΓΝ έχει σχεδόν εγκαταλειφθεί από τους ερευνητές της ΤΝ λόγω της έλλειψης προόδου εδώ και περισσότερα από 50 έτη παρά τις προσπάθειες που καταβάλλονται. Αντιθέτως, η ειδική ΤΝ σημειώνει αλματώδη πρόοδο.

Ισχυρή έναντι ασθενούς ΤΝ

Ένας συναφής διαχωρισμός είναι μεταξύ «ισχυρής» και «ασθενούς» ΤΝ. Αυτός ο διαχωρισμός ανάγεται στην παραπάνω φιλοσοφική διάκριση μεταξύ του να είναι κάποιος νοήμων και του να ενεργεί με νοήμονα τρόπο, διάκριση την οποία τόνισε ο Σηρλ. Η ισχυρή ΤΝ ισοδυναμεί με έναν «νου» που είναι πραγματικά νοήμων και διαθέτει αυτοσυνείδηση. Η ασθενής ΤΝ είναι αυτό που έχουμε σήμερα, δηλαδή συστήματα που επιδεικνύουν νοήμονες συμπεριφορές παρά το γεγονός ότι είναι «απλοί» υπολογιστές.

Unanswered

Άσκηση 4: Ορισμοί, ορισμοί

Εγγραφείτε για να λύσετε ασκήσεις:

Σύνδεση

Εγγραφή

Αφού ολοκληρώσετε το κεφάλαιο 1, θα πρέπει να μπορείτε:

Να εξηγήσετε την αυτονομία και την προσαρμοστικότητα ως βασικές έννοιες για να εξηγήσετε τι είναι η ΤΝ

Να διακρίνετε μεταξύ ρεαλιστικής και μη ρεαλιστικής ΤΝ (επιστημονική φαντασία έναντι πραγματικής ζωής)

Να εκφράσετε τα βασικά φιλοσοφικά προβλήματα που συνδέονται με την ΤΝ, συμπεριλαμβανομένων των συνεπειών του τεστ Τούρινγκ και του πειράματος σκέψης του κινέζικου δωματίου.

Γίνετε μέλη της [κοινότητας του κύκλου μαθημάτων Elements of AI](#) για να συμμετάσχετε σε συζητήσεις και να υποβάλετε ερωτήσεις σχετικά με αυτό το κεφάλαιο.

Φτάσατε στο τέλος του κεφαλαίου 1

Επόμενο κεφάλαιο

Επίλυση ~~προβλημάτων~~ με την ΤΝ

Έναρξη →

[Εισαγωγή στην ΤΝ](#)

[Χτίζοντας εφαρμογές ΤΝ](#)

[Σχετικά](#)

[Συχνές ερωτήσεις](#)

[Πολιτική απορρήτου](#)

[Όροι χρήσης](#)

[Σύνδεση](#) [Εγγραφή](#)



HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

MinnaLearn