

Heidegger και Τεχνητή Νοημοσύνη¹

Άλκης Γούναρης

Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών

Abstract: Μια εισαγωγή στην Χαϊντεγκεριανή Τεχνητή Νοημοσύνη με αφορμή την μετάφραση στα ελληνικά της εισήγησης του Hubert Dreyfus (2007): «Γιατί Απέτυχε η Χαϊντεγκεριανή Τεχνητή Νοημοσύνη και πώς για την επιδιόρθωσή της θα πρέπει να γίνει πιο Χαϊντεγκεριανή».

Keywords: Artificial Intelligence, Human Cognition, Embodiment

Στο δεύτερο μισό του 20^{ου} αιώνα και κυρίως από τα τέλη της δεκαετίας του '60 και μετά, δυο επιστημονικοί κλάδοι αναπτύχθηκαν παράλληλα συνάπτοντας μεταξύ του αμοιβαία δάνεια με αποτέλεσμα να είναι δύσκολο να αποτιμήσουμε πλέον ποιος χρωστάει σε ποιόν. Πρόκειται για τον κλάδο της Τεχνητής Νοημοσύνης και τον κλάδο των Νευροεπιστημών.

Η ιδέα των ερευνητών της Τεχνητής Νοημοσύνης ήταν να δημιουργήσουν ένα τεχνητό υπολογιστικό σύστημα που θα προσομοιώνει την λειτουργία του ανθρώπινου εγκεφάλου ενώ παράλληλα η ιδέα των νευροεπιστημόνων ήταν να περιγράψουν την λειτουργία του εγκεφάλου ως ένα πολύπλοκο υπολογιστικό σύστημα, μια μηχανή που «παράγει» τον νου, θα λέγαμε απλοϊκά.

Η διατύπωση ότι ο νους λειτουργεί ως υπολογιστικό σύστημα, έχει γίνει ήδη από τον Hobbes το 1651, αλλά θα μπορούσε να ανατρέξει κανείς και σε αντίστοιχες σκέψεις του Σωκράτη ή του Αριστοτέλη. Ο Hobbes υποστήριζε ότι πραγματοποιούμε νοητικούς υπολογισμούς, βασιζόμενοι σε μηχανιστικές αρχές που μπορούν να συγκριθούν με τους αριθμητικούς κανόνες. Ο συλλογισμός, σύμφωνα με τον Hobbes δεν είναι τίποτε άλλο παρά υπολογισμός που προσθέτει και αφαιρεί, ως επακόλουθο

¹ Γούναρης, Α. (2010). Heidegger και Τεχνητή Νοημοσύνη. Πρόλογος στην ελληνική μετάφραση της δημοσίευσης: Dreyfus, H. (2007). Why Heideggerian AI Failed and How Fixing It Would Require Making it More Heideggerian. *Philosophical Psychology*. Routledge. Vol. 20, No. 2, 247-268. Εργαστήριο Dasein Lab, Αθήνα. Ανακτήθηκε 24/01/2019 από <https://alkisgounaris.gr/gr/research/heidegger-and-artificial-intelligence/>

των λέξεων που έχουν προσυμφωνηθεί, ώστε να σηματοδοτούνται και να αποκτούν σημασία οι σκέψεις μας.

Η αρχή για την σύγχρονη διατύπωση μιας θεωρίας που υποστηρίζει ότι η νόηση είναι προϊόν ενός λειτουργικού υπολογιστικού μηχανισμού μπορεί να αποδοθεί στον πάπα της Τεχνητής Νοημοσύνης, τον άγγλο μαθηματικό και θεωρητικό της λογικής, Alan Turing. Το 1950, λίγα χρόνια πριν πεθάνει, ο Turing δημοσίευσε ένα νοητικό πείραμα που το ονόμασε παιχνίδι της μίμησης με το οποίο ήθελε να δείξει, ότι όπως σε ένα παιχνίδι μεταξύ ανθρώπων οι παίχτες σκέφτονται και πράττουν, έτσι και ένας υπολογιστής που μπορεί να παίξει αποτελεσματικά σε αυτό το παιχνίδι, σκέφτεται και πράττει σαν ένα ον με ανθρώπινη νοημοσύνη.

Στις αρχές του '60 ο Hilary Putnam στην προσπάθειά του να αντικρούσει τον αναλυτικό συμπεριφορισμό (ένα ρεύμα που κυριαρχούσε τότε στους φιλοσοφικούς και ψυχολογικούς κύκλους) υποστήριξε ότι οι νοητικές λειτουργίες είναι αντίστοιχες με τις υπολογιστικές διαδικασίες των ηλεκτρονικών υπολογιστών. Η θεωρία εξελίχθηκε στη δεκαετία του '70 από τον Jerry Fodor ο οποίος επισήμανε τον ρόλο της γλώσσας στις νοητικές λειτουργίες, προτείνοντας μάλιστα την υπόθεση ότι οι νοητικές λειτουργίες συνιστούν μια γλώσσα, η οποία έχει δικό της συντακτικό και νοηματικό περιεχόμενο και επηρεάζει αιτιωδώς τις νοητικές αναπαραστάσεις.

Ο παράγων «γλώσσα» έπαιξε στο εξής θεμελιώδη ρόλο στην ανάπτυξη της Τεχνητής Νοημοσύνης όπως η αποκρυπτογράφηση του λειτουργικού μηχανισμού που παράγει τον νου, αποτέλεσε το άγιο δισκοπότηρο των νευροπιστημόνων ως τις μέρες μας.

Μέχρι στιγμής και οι δύο κλάδοι φαίνεται να είναι μακριά από τον στόχο τους. Οι μεν έρευνες για την Τεχνητή Νοημοσύνη απέχουν ακόμα πολύ από την κατασκευή ενός νοήμονος ρομπότ που να «σκέφτεται» και να συμπεριφέρεται σαν ανθρώπινο ον, οι δε νευροεπιστήμονες όσο και να σκανάρουν ή να ανατέμνουν τους εγκεφάλους δεν μπορούν να εξηγήσουν πως παράγεται και τι είναι η νόηση, τι είναι η συνείδηση, πως συγκροτείται το νόημα και πως δικαιολογείται το υποκειμενικό βίωμα της αισθητηριακής εμπειρίας.

Τα τελευταία δέκα – δεκαπέντε χρόνια έχει αρχίζει να «ακούγεται» η πιθανότητα ότι οι ερευνητικές προσπάθειες αλλά και τα εξηγητικά μοντέλα μπορεί να καταλήγουν σε αδιέξοδα, επειδή τα βασικά αξιώματα, οι βασικές αρχές που έχουν υιοθετηθεί και βάσει των οποίων πορεύονται οι έρευνες και οι εφαρμογές, είναι λάθος. Για παράδειγμα, μπορεί να είναι λάθος η παραδοχή ότι υπάρχουν νοητικές αναπαραστάσεις ή η παραδοχή ότι αποθηκεύονται στην μνήμη δεδομένα και πληροφορίες, κ.α. Ακόμα μπορεί να μας απομακρύνει από την κατανόηση του φαινομένου της νόησης η σύλληψη των ίδιων των στοιχειωδών μονάδων πληροφορίας, και των αισθητηριακών δεδομένων. Είναι σχεδόν βέβαιο δε, ότι η ίδια η γλώσσα μάς έχει παγιδέψει, στο βάθος του χρόνου, σε ένα πλαίσιο όπου ορισμένοι όροι αναπαράγονται και θεωρούνται δεδομένοι ως ισχύουσες καταστάσεις, ξεχνώντας συχνά ότι απλώς αποτελούν κατασκευασμένες σε κάποια εποχή και κάποιο θεωρητικό πλαίσιο έννοιες. Η διατήρηση των ψυχολογικών όρων για παράδειγμα, η οντολογική αναφορά στον νου, η υπερτίμηση του ρόλου και η θεώρηση της ύπαρξης της συνείδησης, η αναφορά στον εαυτό και στην αυτοσυνείδησία, είναι μερικές μόνον περιπτώσεις.

Και τι δουλειά έχει ο Heidegger με όλα αυτά;

Ο Heidegger (και ορισμένοι επηρεασμένοι από αυτόν φαινομενολόγοι όπως ο Merleau-Ponty) διατύπωσαν μια διαφορετική θεώρηση για την ανθρώπινη νόηση και την σχέση της με την γλώσσα και το περιβάλλον, η οποία για πολλά χρόνια παρέμενε στη σκιά. Συγκεκριμένα ο Heidegger αντέκρουσε σε πολλά σημεία την καρτεσιανή υπόθεση (για την νόηση, την προθετικότητα ή κατευθυντικότητα, τις νοητικές αναπαραστάσεις κλπ) που φαίνεται ότι στοιχειώνει τα περισσότερα εξηγητικά μοντέλα ακόμα και σήμερα, την οποία μάλιστα κατηγόρησε ως υπεύθυνη για πολλές στοιχειώδεις παρανοήσεις.

Ο Hubert Dreyfus, πριν 35 περίπου χρόνια εισήγαγε τον Heidegger στα μαθήματα επιστήμης ηλεκτρονικών υπολογιστών του MIT (του Τεχνολογικού Πανεπιστημίου Μασαχουσέτης) με αποτέλεσμα να αρχίζει να αλλάζει ο τρόπος που «νοείται η νόηση» και ως εκ τούτου να σημειωθούν σημαντικά επιτεύγματα στην έρευνα της Τεχνητής Νοημοσύνης τα τελευταία 15 χρόνια. Ένα καίριο ζήτημα που τροποποιεί το πώς αντιλαμβανόμαστε την νοητική διαδικασία, είναι το ζήτημα των αναπαραστάσεων. Ο Dreyfus υποστηρίζει ότι: συνήθως δεν έχουμε ανάγκη την αναπαράσταση του κόσμου στον νου μας, αφού ο καλύτερος τρόπος για να αντιληφθούμε την εκάστοτε κατάσταση των πραγμάτων, είναι να στραφούμε στον κόσμο όπως τον βιώνουμε. Για παράδειγμα, βιώνω το σφυρί που χρησιμοποιώ όχι σαν ένα εργαλείο με κάποιες ιδιότητες που μπορεί να αναπαρασταθούν τυπικά, αλλά σαν αυτό που χρησιμοποιώ για να καρφώσω π.χ. ένα καρφί. Ο Dreyfus σχολιάζει, ότι η ιδέα της νοητικής αναπαράστασης είναι κατάλοιπο της Καρτεσιανής σκέψης, όπως η ιδέα της λειτουργίας του μυαλού ως υπολογιστικής μηχανής είναι κατάλοιπο της σκέψης του Hobbes και η ιδέα των καθολικών χαρακτηριστικών απομεινάρι από τον Leibniz. Όλα αυτά συνδέονται σε μια αναπαραστατική και υπολογιστική θεωρία, σύμφωνα με την οποία, η ανθρώπινη νοημοσύνη και η τεχνητή νοημοσύνη εξετάζονται με τους ίδιους λίγο πολύ όρους.

Στη παρούσα δημοσίευση, αυτό που κάνω είναι να παρουσιάσω μεταφρασμένη μια εξαιρετική κατά τη γνώμη μου εισήγηση του Hubert Dreyfus η οποία έγινε το 2007 στο MIT και δημοσιεύθηκε στη συνέχεια σε πολλά περιοδικά φιλοσοφίας του νου, τεχνητής νοημοσύνης και νευροεπιστημών, με τον τίτλο: Γιατί Απέτυχε η Χαϊντεγκεριανή Τεχνητή Νοημοσύνη και πώς για την επιδιόρθωσή της θα πρέπει να γίνει πιο Χαϊντεγκεριανή.

Επέλεξα να μεταφράσω και να παρουσιάσω αυτή την εισήγηση διότι έχει άμεση σχέση με τη διατριβή μου και γιατί προσφέρει με πολύ απλό και προσπελάσιμο τρόπο, μεγάλο όγκο πληροφοριών, ανάλυση και διευκρίνιση εννοιών και επιχειρημάτων και πλούσιο ακαδημαϊκό παρασκήνιο που βοηθούν να γίνει κατανοητή η διαφορετική προσέγγιση της νόησης από τον Heidegger αλλά ταυτόχρονα, το πώς η προσέγγιση αυτή μπορεί να έχει πρακτική εφαρμογή στην έρευνα για την Τεχνητή Νοημοσύνη και στην συγκρότηση ενός εξηγητικού μοντέλου της νόησης στις Νευροεπιστήμες.

Λόγω της μεγάλης έκτασης της εισήγησης του Dreyfus δημοσιεύεται σε τρία μέρη:

Στο πρώτο μέρος, [Link](#), παρουσιάζονται: η προϊστορία και η σχέση της έρευνας για την τεχνητή νοημοσύνη με την φιλοσοφία, καθώς και οι αποτυχίες της συμβολικής – αναπαραστασιακής τεχνητής νοημοσύνης και της πρώιμης χαϊντεγκεριανού τύπου TN που βασίστηκε στη συμπεριφορά.

Στο δεύτερο μέρος, [Link](#), παρουσιάζεται η χαϊντεγκεριανή TN ως μια προσπάθεια προγραμματισμού της «προχειρότητας», αλλά και ως παρευρισκόμενη γνώση ή ως ενταγμένος - ενσωματωμένος και επεκτάσιμος νους.

Και στο τρίτο μέρος [Link](#) παρουσιάζεται μια σύγχρονη εκδοχή της χαϊντεγκεριανής TN μέσα από το νευροδυναμικό (Χαϊντεγκεριανής – Μερλω-Ποντιανής εκδοχής) μοντέλο του Walter Freeman, καθώς και τα ζητήματα που εγείρονται εκ νέου.