

Εισαγωγή

Η Τεχνητή Νοημοσύνη (στο εξής: ΤΝ)¹ είναι πλέον πανταχού παρούσα στη ζωή μας. Παράλληλα, οι μελέτες που ασχολούνται με αυτή την τεχνολογία πληθαίνουν με αυξανόμενους ρυθμούς, κάτι που φαίνεται να συμβαίνει και στην ελληνική βιβλιογραφία.² Εντούτοις, λόγω της σημασίας που έχει αποκτήσει η ΤΝ στις σύγχρονες κοινωνίες, καθώς και του μύθου που έχει χτιστεί γύρω της, φαίνεται πια κάπως παρωχημένο να γράφει κανείς για αυτή την τεχνολογική εξέλιξη. Αυτό δεν συμβαίνει επειδή έχουμε κατανοήσει πλήρως την ΤΝ και την έχουμε αναλύσει σε βάθος, αλλά επειδή οι δύο λέξεις «Τεχνητή» και «Νοημοσύνη» έχουν ακουστεί τόσο πολύ τα τελευταία χρόνια, που κάθε νέο κείμενο πάνω σε αυτές μοιάζει να επαναλαμβάνει τα ίδια: κάποια μελλοντική δυστοπία όπου ρομπότ ΤΝ έχουν κατακτήσει τον κόσμο· τα οφέλη που προσφέρει και το πώς θα βελτιώσει την καθημερινότητα· η αντικατάσταση θέσεων εργασίας από λογισμικά ΤΝ και οι νέες εργασιακές συνθήκες· πώς οι αλγόριθμοι παραβιάζουν τα προσωπικά μας δεδομένα και τι μπορούμε να κάνουμε γι' αυτό· αν η ΤΝ συμβάλλει στην κλιματική κρίση ή πώς μπορεί να βοηθήσει στην καταπολέμησή της· πόσα είναι τα κέρδη των εταιρειών που αναπτύσσουν ΤΝ· αν το Gemini της Google είναι πιο «έξυπνο» από το ChatGPT της OpenAI· ή για κάποιο ατύχημα με αυτοκίνητο της Tesla, και ούτω καθεξής. Δημοσιεύματα, άρθρα, βιβλία και

1. Όπως φανερώνεται ήδη από τα εισαγωγικά στον τίτλο, δεν θεωρώ πως η λέξη «νοημοσύνη» ταιριάζει σε αυτό το πεδίο έρευνας. Πώς είναι δυνατό να υποθέσουμε ότι μπορούμε να φτιάξουμε «τεχνητή» «νοημοσύνη» τη στιγμή που δεν έχουμε λύσει το μυστήριο της νοημοσύνης στα ίδια τα ανθρώπινα όντα; Το ερώτημα αυτό διαπερνά το παρόν βιβλίο σε όλο του το εύρος.

2. Βλ. ενδεικτικά: Αλέξανδρος Σχισμένος, *Τεχνολογία και Βαρβαρότητα. Κριτική της Τεχνητής Νοημοσύνης*, Αθήνα: Athenschool, 2024· Αντώνης Μαυρόπουλος, *Τεχνητή Νοημοσύνη. Άνθρωπος, φύση, μηχανές*, Αθήνα: Τόπος, 2024· Γιώργος Χατζηβασιλείου, *Φιλοσοφία της Τεχνητής Νοημοσύνης. Ένα ταξίδι στο μέλλον*, Αθήνα: Διόπτρα, 2024· Κώστας Θεολόγου, *Φιλοσοφία και Τεχνολογία*, Αθήνα: Ελληνοεκδοτική, 2023· Γιώργος Γιαννακόπουλος, *Τεχνητή Νοημοσύνη. Μία διακριτική απομυθοποίηση*, Θεσσαλονίκη: Ροπή, 2020.

κείμενα που επεξεργάζονται ζητήματα όπως τα παραπάνω εμφανίζονται συνεχώς στα μέσα κοινωνικής δικτύωσης και σε διάφορους ιστοτόπους.

Τι άλλο μένει να ειπωθεί για την ΤΝ χωρίς να επαναλαμβάνεται η ίδια αφήγηση; Ή, για να το θέσουμε αλλιώς, υπάρχει κάτι που δεν έχει ήδη ειπωθεί ή που αποφεύγεται να συζητηθεί σχετικά με την ΤΝ;

Στην παραπάνω ερώτηση θα απαντούσα καταφατικά. Υπάρχει κάτι που δεν είναι «χιλιοειπωμένο», που δεν αναπαράγει δυστοπικά αφηγήματα και δεν επικεντρώνεται στις προφανείς αλλαγές που φέρνει ή θα φέρει η ΤΝ. Και αυτό το σχεδόν αθέατο ή παραμελημένο στοιχείο στην ιστορία του πεδίου της ΤΝ αποτελεί το υπόβαθρο όλων των γνωστών αφηγήσεων, είτε αυτές είναι δυστοπικές, είτε σχετίζονται με το περιβάλλον, είτε αφορούν στην αντικατάσταση των ανθρώπων από ρομπότ στην εργασία. Συγκεκριμένα, αυτό που συχνά θεωρείται δεδομένο, αν και δεν έχει εξεταστεί αρκετά, και που διαπερνά αθόρυβα τα εκατοντάδες δημοσιεύματα και τα δεκάδες βιβλία είναι το ζήτημα των θεμελίων της ΤΝ: τα φιλοσοφικά θεμέλια που στηρίζουν τη λογική πίσω από την ΤΝ και τον τρόπο με τον οποίο αυτό το πεδίο έρευνας ερμηνεύει τη φύση και τον άνθρωπο. Με άλλα λόγια, «ο δρόμος που οδήγησε στην ΤΝ».

Τα προαναφερθέντα βιβλία στην ελληνική γλώσσα που επιχειρούν να ασχοληθούν με την ΤΝ, παρά τη σημασία τους ως μελέτες και τις ενδιαφέρουσες αναλύσεις τους, δεν αναπτύσσουν μια κριτική των φιλοσοφικών θεμελίων³ που καθορίζουν την εξέλιξη και τις εφαρμογές

3. Από τη μία, «μερικώς» εξαίρεση αποτελεί το βιβλίο του Μαυρόπουλου, στο περιεχόμενο του οποίου μπορούν να εντοπιστούν περιπτώσεις κριτικής σε θεμελιακές αρχές της ΤΝ. Αναφέρω τη λέξη «μερικώς» όχι για να υπονομεύσω το βιβλίο, μιας και εκτιμώ την πληθώρα πληροφοριών, τη δομή αλλά και την κριτική κοινωνικοπολιτική του προσέγγιση της τεχνολογίας. Αντιθέτως, η λέξη «μερικώς» αναφέρεται μονάχα στο γεγονός ότι στόχος του εν λόγω βιβλίου δεν είναι να εξετάσει τα θεμέλια της ΤΝ συστηματικά, αλλά να εξερευνήσει κριτικά πολλαπλά ζητήματα που σχετίζονται με την τεχνολογία αυτή. Αναμφισβήτητο, το εγχείρημα του Μαυρόπουλου είναι έξοχο και ο συγγραφέας πετυχαίνει να γράψει ένα σημαντικό για την ελληνική βιβλιογραφία βιβλίο με κοινωνική κατεύθυνση, κόντρα στις δυστοπικές ή θετικιστικές προσεγγίσεις στην ΤΝ.

Από την άλλη, «μερικώς» εξαίρεση αποτελεί το βιβλίο του Σχισμένου, το οποίο μπορεί επίσης να μην στρέφεται συστηματικά στα θεμέλια της ΤΝ, ωστόσο η πλούσια ανάλυσή του και η κοινωνική και πολιτική κριτική του στο τεχνούργημα της ΤΝ είναι πολύτιμη. Σε αυτό, ο Σχισμένος εξετάζει τη νοημοσύνη από αφετηρίες μη συννηθισμένες (π.χ. από τη νοημοσύνη των ζώων ή από τη χρονικότητα της ψηφιακότητας ή από εννοιολογήσεις της βαρβαρότητας), πετυχαίνοντας να διανοίξει νέους ορίζοντες για μια κριτική της ΤΝ, ασκώντας κριτική στις θετικίζουσες προοπτικές, καθώς και σε εκείνες που επιχειρούν να εξετάσουν την ΤΝ μέσα από μία

του πεδίου της επιστήμης των υπολογιστών που επιδιώκει να κατασκευάσει μηχανές προσομοιώνοντας τον άνθρωπο και τις λειτουργίες του (δηλαδή του πεδίου της ΤΝ).⁴ Η σε βάθος εξέταση του τρόπου λειτουργίας της ΤΝ, των επιρροών της σε κοινωνικούς τομείς, ή ακόμη και της δυϊστικής της βάσης, όταν δεν καταλήγει σε μια κριτική των θεμελίων της, δεν αρκεί για να αντιμετωπίσει τον μύθο της δυστοπίας που την περιβάλλει.⁵ Παραμένει μια απλή ανάλυση του φαινομένου,

στείρα νευροεπιστημονική και βιολογική οπτική. Στην ανάλυσή του διαφαίνονται στοιχεία που εμφανίζονται και στο ανά χείρας βιβλίο και θα ήταν ενδιαφέρον να μελετήσει καμιά τα δύο έργα παράλληλα.

Ένα επίσης εξαιρετικό και πρωτότυπο βιβλίο είναι: Μανώλης Ανδριωτάκης, *Homo Automaton. Η τεχνητή νοημοσύνη κι εμείς*, Αθήνα: Garage Books, 2020. Το έργο αυτό υποστηρίζει πως η Τεχνητή Νοημοσύνη και οι αρχές της (όχι απλώς οι εφαρμογές της) διαμορφώνουν έναν νέο ανθρωπότυπο (τον Homo Automaton) με χαρακτηριστικά που έχουν άμεσες συνέπειες σε κοινωνικό, πολιτικό και εντέλει οντολογικό επίπεδο. Ο συγγραφέας αναφέρεται σε οντολογικά στοιχεία που καθιστούν δυνατή την ΤΝ, σε επιστημολογικά, ακόμη και στον συμπεριφορισμό, τα media και τις εταιρείες κολοσσούς της τεχνολογίας. Θα μπορούσε, δηλαδή, να πει κανείς, ότι το ανά χείρας βιβλίο συνδέεται άρρηκτα με εκείνο του Ανδριωτάκη, καθώς και τα δύο εξερευνούν τον ανθρωπότυπο που κατέστησε δυνατή την ανάδυση της ΤΝ (και ότι, άρα, αυτή δεν αναδύθηκε εν κενώ, ως ουδέτερη τεχνολογία, ούτε και εφαρμόζεται πάνω σε ένα κοινό που δεν είχε κανέναν θύλακα ήδη διαμορφωμένο για τη συγκεκριμένη της μορφή, όπως αυτή υπάρχει σήμερα).

4. Όπως θα δούμε και στη συνέχεια, και όπως υποστηρίζει και ο Λουτσιάνο Φλορίντι στο βιβλίο του (Luciano Floridi, *Εισαγωγή στη Φιλοσοφία της Πληροφορικής*, μτφρ. Χρ. Γεμελιάρης, Γ. Καραμπέλας, Αθήνα: Νήσος, 2008, σ. 292–300), από κάποια ιστορική στιγμή και έπειτα το πεδίο της ΤΝ ξεκίνησε να αναπτύσσει λογισμικά χωρίς να προσπαθεί να μιμηθεί τις ανθρώπινες νοητικές λειτουργίες (αυτό που ο ίδιος ονομάζει MTN). Προσπάθησε να πετύχει τους στόχους των ανθρώπινων λειτουργιών αλλά διαμέσου διαφορετικών οδών. Παρόλο που αυτό πράγματι συνέβη, τα θεμέλια που υποβαστάζουν ακόμη και αυτήν τη φάση ανάπτυξης της ΤΝ κρίνουμε ότι παρέμειναν ίδια. Άλλωστε, αυτό φανερώνει και το γεγονός ότι ο τίτλος του πεδίου έρευνας («Τεχνητή Νοημοσύνη») διατήρησε μία έννοια που χαρακτηρίζει εκείνο το on το οποίο αυτό το πεδίο επιχειρεί να προσομοιάσει.

5. Η άποψή μου περί *κριτικής* είναι επηρεασμένη από τον Μισέλ Φουκώ, του οποίου το κείμενο-ομιλία του 1978 «Τι είναι κριτική;» (Qu'est-ce que la critique?) –διάθεσιμο στην ελληνική: Michel Foucault, *Τι είναι κριτική;*, μτφρ. Θ. Λάγιος, Αθήνα: Πλέθρον, 2016– διέπεται από τη σκέψη πως η κριτική δεν συνίσταται στο να λέει κανείς ότι τα πράγματα δεν είναι καλά όπως είναι, αλλά στο να δείχνει πάνω σε τι είδους παραδοχές, τι είδους οικείες και αδιαμφισβήτητες μορφές σκέψης βασίζονται όλα όσα αποδεχόμαστε ως αληθή· δηλαδή τα θεμέλια της ίδιας μας της ύπαρξης και γνώσης. Χωρίς ανάδειξη και αμφισβήτηση των θεμελίων, των παραδοχών και των αξιωμάτων, δεν μπορεί να υφίσταται κριτική. Κατ' αντιστοιχία, χωρίς βαθιά γνώση των επιστημολογικών και οντολογικών θεμελίων που κατέστησαν δυνατή την ΤΝ, δεν μπορεί να ασκηθεί ούτε ηθική, ούτε πολιτική, ούτε περιβαλλοντική κριτική σε αυτήν.

χωρίς αναφορά στο ουσιώδες υπόβαθρο, δηλαδή τις οντολογικές, επιστημολογικές και ψυχολογικές παραδοχές που χαρακτηρίζουν το εν λόγω πεδίο. Αυτή η προσέγγιση μπορεί να οδηγήσει σε σφάλματα, παρωχημένες αφηγήσεις, δυστοπικές προβολές για το μέλλον και έναν γενικευμένο πανικό, ο οποίος μπορεί να αποδειχθεί θεωρητικά αλλά και πρακτικά επιβλαβής.

Αντίστοιχα αποτελέσματα μπορεί να προκύψουν και από ομιλίες, διαλέξεις, εκλαϊκευμένα επιστημονικά ή φιλοσοφικά κείμενα, όταν προέρχονται από ανθρώπους και κλάδους που δεν έχουν βαθιά γνώση του ευρύτερου πλαισίου της ΤΝ. Ειδικά μετά την εμφάνιση του ChatGPT και τη συζήτηση που αυτό προκάλεσε παγκοσμίως, υπήρξε ένα τεράστιο κύμα ανάλυσης και επεξήγησης της ΤΝ από ποικίλους παράγοντες, συχνά σε ασταθή θεμέλια. Αυτή η τάση ενισχύει τη δυστοπική φαντασιοπληξία, καθώς η ΤΝ δεν αξιολογείται με βάση τα πραγματικά θεμέλιά της, αλλά μέσα από προσεγγίσεις ή απόψεις επιφανειακού χαρακτήρα.⁶

Τα παραπάνω φαινόμενα προκύπτουν επειδή τα πολιτικά, κοινωνικά, οικονομικά, αλλά και αισθητικά ζητήματα που εγείρει η ΤΝ, συχνά συνοδευόμενα από καταστροφολογικές προβλέψεις και πανικό, πηγάζουν τόσο από τον τρόπο ανάπτυξης του πεδίου όσο και από τον τρόπο με τον οποίο ερμηνεύουμε –και προσπαθούμε να αναπαράγουμε με τεχνητά μέσα– τη νοημοσύνη. Δηλαδή, από τον τρόπο που το πεδίο της ΤΝ αντιλαμβάνεται τον κόσμο, τον άνθρωπο, την πραγματικότητα και τη φύση. Σημειώνουμε εδώ ότι δεν μιλάμε για μια απλή ανάλυση της «νοημοσύνης» της ΤΝ (αν, για παράδειγμα, είναι δυϊστική, βασίζεται σε αλγορίθμους ή παράγει νέους), αλλά για μια ανάλυση των θεμελιωδών παραδοχών που καθόρισαν το πεδίο ώστε αυτό να λάβει τη σημερινή του μορφή. Με άλλα λόγια, δεν εξετάζουμε την ΤΝ μόνο στο πλαίσιο μιας φιλοσοφίας του νου απομονωμένης από άλλες συνδηλώσεις και πλαίσια, αλλά θέτοντας οντολογικά και επιστημολογικά ερωτήματα, που επιτρέπουν την ανάπτυξη κοινωνικής και πολιτικής κριτικής. Κατά τη γνώμη μας, χωρίς μια ουσιαστική κατανόηση των οντολογικών και επιστημολογικών βάσεων του πεδίου της ΤΝ, είναι σχεδόν αδύνατο να εξαχθούν στιβαρά κοινωνικά, οικονομικά και πολιτικά συμπεράσματα που πηγάζουν από αυτήν. Μια φιλοσοφία της ΤΝ, ειδικά αν πρόκειται

6. Λέγοντας αυτό δεν υπονοώ ότι μόνο κάποιου είδους ειδήμονες *πρέπει* να συζητούν για την ΤΝ, αλλά ότι υπάρχει ανάγκη ενός βιβλίου (όπως προσπάθησα να είναι το ανά χείρας) που να αναδεικνύει τα πραγματικά θεμέλια του πεδίου, για την περαιτέρω εκτίμηση, κρίση και ανάπτυξή του.

να ασκήσει οποιαδήποτε μορφή κριτικής, οφείλει να ξεκινά από αυτήν τη βάση, πριν επιχειρήσει συνδέσεις με παραδόσεις, μελλοντικές υποθέσεις ή απλές περιγραφές σχετικά με τρέχουσες εξελίξεις, εταιρείες, προϊόντα κ.ο.κ.

Επιπλέον, όσοι και όσες είναι, λιγότερο ή περισσότερο, εξοικειωμένοι με τα ζητήματα που θα αναλυθούν στα κεφάλαια αυτού του βιβλίου, ίσως διαβάζουν με θαυμασμό (ή απέχθεια) τις δημοσιεύσεις και τα βιβλία που παρουσιάζουν την ΤΝ με δυστοπικό τρόπο. Αυτό δεν σημαίνει, αντίστοιχα στην κάθε περίπτωση, ότι η ΤΝ δεν μπορεί να επιφέρει αρνητικές συνέπειες ή ότι την εξετάζουμε με τεχνολατρεία. Αντιθέτως, υποστηρίζουμε ότι η άποψη πως αυτές οι τεχνολογίες μπορούν να συγκριθούν με τον άνθρωπο ή ακόμη και να τον υπερβούν είτε δεν ευσταθεί, καθώς τα θεμέλιά της είναι άστατα, είτε ευσταθεί μόνο υπό την προϋπόθεση να ορίσουμε τον «άνθρωπο» με έναν πολύ συγκεκριμένο τρόπο (τον οποίο θα εξετάσουμε στη συνέχεια του βιβλίου). Γίνεται, συνεπώς, κατανοητό ότι τα εν λόγω ζητήματα συνδέονται, αφενός, με το ερώτημα «τι είναι ο άνθρωπος» (τον οποίο η ΤΝ υποτίθεται ότι μπορεί να υπερβεί), και αφετέρου, με το γεγονός ότι τα προβλήματα αφανισμού, κατάληψης θέσεων εργασίας και κοινωνικής παρακμής προκύπτουν από τον τρόπο κοινωνικοπολιτικής και οικονομικής διαχείρισης και σχεδιασμού της τεχνολογίας αυτής, και όχι από την ίδια τη φύση της.⁷

Το βιβλίο που κρατάτε στα χέρια σας αποτελεί πνευματικό αποκύημα της μελέτης, πρωτίστως, του μοναδικού σε βάθος και ποιότητα ανάλυσης έργου του Χιούμπερτ Ντρείφους *Τι δεν μπορούν ακόμη να κάνουν οι υπολογιστές. Μία κριτική της Τεχνητής Νοημοσύνης*.⁸ Σε αυτό, η ΤΝ παρουσιάζεται –κατά την προσωπική μου γνώμη– όπως θα έπρεπε, δηλαδή στη βάση των φιλοσοφικών παραδοχών-θεμελίων της, της εξέλιξής της και των αποφάσεων που καθόρισαν τον σχεδιασμό της. Θα ήθελα να τονίσω ότι το έργο του Ντρείφους αποτελεί το ιδανικό

7. Το βασικό μου επιχείρημα σε αυτό το βιβλίο είναι ότι κατά την ανάπτυξη της ΤΝ, κατά την ανάπτυξη των διανθρωπιστικών οραμάτων και κατά την άρθρωση καταστροφολογικών προβλέψεων, λαμβάνεται ως δεδομένο ένα από τα νήματα περί του τι είναι ο άνθρωπος (και μάλιστα ένα κάπως απαρχαιωμένο θα λέγαμε). Το αν αυτό είναι το ευρύτερα αποδεκτό νόημα περί ανθρώπου, ή αν πρέπει να το θεωρούμε ως δεδομένο, δεν τίθεται ως ερώτημα, διότι τότε θα υπέσκαπτε εμμέσως τα θεμέλια του ίδιου του πεδίου της ΤΝ.

8. Hubert L. Dreyfus, *Τι δεν μπορούν ακόμη να κάνουν οι υπολογιστές. Κριτική της Τεχνητής Νοημοσύνης*, μτφρ. Π. Καρλέτσα, Αθήνα-Ηράκλειο: Πανεπιστημιακές εκδόσεις Κρήτης, 2001.

ανάγνωσμα για οποιαδήποτε επιθυμεί να κατανοήσει το πεδίο που συχνά αποκαλείται, ίσως καταχρηστικά, «Τεχνητή Νοημοσύνη».

Τα κεφάλαια που ακολουθούν αποτελούν μια επέκταση ορισμένων από τις θεμελιώδεις θέσεις του Ντρέιφους, με στόχο να εξεταστεί εάν τα κριτικά επιχειρήματά του διατηρούν την ισχύ τους ακόμη και σήμερα. Επιπρόσθετα, όπως θα φανεί στη συνέχεια, σκοπός μας εδώ είναι να εμπλουτίσουμε το κύριο επιχειρήμα του Ντρέιφους –το οποίο είναι έντονα επηρεασμένο από τη φαινομενολογική παράδοση– προσθέτοντας στοιχεία από άλλες φιλοσοφικές σχολές σκέψης, ώστε να αναδυθούν περισσότερες ατραποί κριτικής της δυστοπικής τάσης της ΤΝ.

Ο κεντρικός στόχος είναι η επανεξέταση των θεμελίων της ΤΝ, με σκοπό να διερευνηθεί εκ νέου αν αυτό που ονομάζουμε «Νοημοσύνη» στην ΤΝ είναι όντως νοημοσύνη, και αν ναι, τι είδους. Από αυτήν τη βάση, μπορούμε στη συνέχεια να προσεγγίσουμε ζητήματα όπως το αν μία τεχνητή οντότητα, όπως αυτές που κατασκευάζουμε και ερμηνεύουμε ως ΤΝ, μπορεί να αυτονομηθεί ή να γίνει «πιο έξυπνη» – ανησυχίες που συχνά πυροδοτούν τον πανικό σε μεγάλο μέρος των ανθρώπων.

Χωρίς την κατανόηση των θεμελίων αυτού που ονομάζουμε «ΤΝ», είναι αδύνατο να γίνουν οι κατάλληλες συνδέσεις με κοινωνικά, πολιτικά, οικονομικά και περιβαλλοντικά ζητήματα. Αυτός είναι και ο βασικός λόγος για τον οποίο οι σύγχρονες κοινωνικές και πολιτικές αναλύσεις καταλήγουν συχνά σε δυστοπικές προβλέψεις για το μέλλον ή περιορίζονται σε ηθικολογικές και νομικές προσεγγίσεις, αναζητώντας κανόνες και νομοθετικά πλαίσια για τον περιορισμό της λειτουργίας της υπάρχουσας ΤΝ.⁹

9. Πράγματι, πληθώρα στοχαστών διεθνώς, αλλά και στην Ελλάδα, αναφέρονται στον ηθικό σχεδιασμό της ΤΝ μέσα από την εισαγωγή ηθικών αξιών ήδη από τα στάδια του προγραμματισμού της, προκειμένου στο μέλλον να αποφευχθεί η δυστοπία. Αγνοείται, ωστόσο, το γεγονός ότι αν η ανάπτυξη μίας τεχνολογίας, ακόμη και της ΤΝ, προχωρά στη βάση συγκεκριμένων θεμελίων, νοημάτων και ερμηνειών (κάτι που την καθιστά πολιτική· βλ. τον επίλογο του παρόντος βιβλίου), οι νομοθεσίες, ή ακόμη και η αυθαίρετη εισαγωγή αξιών διαφορετικών από εκείνες των θεμελίων τους, απορρίπτονται ως προσπάθεια παρεμπόδισης της τεχνολογικής εξέλιξης. Μπορούμε να εντοπίσουμε σημάδια αυτού του φαινομένου στο γεγονός ότι οι εταιρείες ανάπτυξης ΤΝ σήμερα αποφεύγουν την Ευρώπη ως τόπο εφαρμογής των προγραμμάτων τους λόγω του ευρωπαϊκού AI Act. Πώς μπορούμε να ζητάμε από εταιρείες όπως αυτές να «εισάγουν πιο ηθικές αξίες στην ΤΝ» όταν τα συστήματα αυτά θεωρείται ότι κυριαρχούν επί της φύσης και του ανθρώπου και ότι ο AI Act απλώς παρεμποδίζει τη φυσική τους εξέλιξη; Καταλήγουμε, λοιπόν, στο ότι το ζήτημα της ενσωμάτωσης διαφορετικών και πιο «ηθικών» αξιών στην ΤΝ και σε κάθε τεχνούργημα είναι εγγενώς πολιτικό – δηλαδή, σχετίζεται με, εξαρτάται από

Από την άλλη πλευρά, αυτό που προτείνεται στο παρόν βιβλίο είναι πως μια βαθύτερη κατανόηση των θεμελίων που οδήγησαν στην ΤΝ, στη σημερινή της μορφή, καθώς και των εγγενών ορίων που επιβάλλει η ίδια η φύση αυτών των προγραμμάτων, θα συνέβαλε στην καλύτερη αξιολόγηση του κατά πόσον η ΤΝ αποτελεί πραγματικά απειλή για την ανθρωπότητα. Εντούτοις, όπως θα φανεί στα επόμενα κεφάλαια, αυτή η επανεξέταση μάς οδηγεί επίσης στην ανάγκη να επανεκτιμήσουμε τα θεμέλια της ίδιας της ανθρώπινης νοημοσύνης, όπως και τις αντιλήψεις για την έννοια του ανθρώπου, τις γνωστικές και σωματικές του συνδελιώσεις, τόσο στο παρόν όσο και στο παρελθόν.¹⁰

Το θετικό στοιχείο της ΤΝ είναι ότι επανέφερε στο προσκήνιο τη συζήτηση για τα θεμέλια της ανθρώπινης ύπαρξης, της νοημοσύνης και της αλληλεπίδρασης. Το αρνητικό, όμως, είναι ότι αυτή η συζήτηση, λόγω της πολυπλοκότητας και του εύρους της –ακόμη και για τους επιστήμονες ή φιλοσόφους– συχνά παραγκωνίζεται, δίνοντας τη θέση της σε επιφανειακές προσεγγίσεις που αποδέχονται άκριτα τις κυρίαρχες παραδοχές σχετικά με τον κόσμο και τον άνθρωπο.

Παρότι οι αναλύσεις που έπονται θα εξετάσουν παραδοσιακά φιλοσοφικά ζητήματα που άπτονται της φιλοσοφίας του νου, το βιβλίο δεν ισχυρίζεται ότι αντιμετωπίζει πλήρως τις προβληματικές αυτής της φιλοσοφικής περιοχής. Στα επόμενα κεφάλαια θα γίνουν πολλαπλές αναφορές σε φιλοσόφους όπως ο Ρενέ Ντεκάρτ, ο Τζον Λοκ, και ο Ντέιβιντ Χιουμ, στους οποίους συχνά επιστρέφουν οι φιλόσοφοι του νου για να εξετάσουν έννοιες όπως η νοημοσύνη, η αντίληψη και η μνήμη. Εντούτοις, ο σκοπός δεν είναι να διερευνήσουμε πώς αντιλαμβάνεται ο άνθρωπος ή σε ποιον βαθμό η μνήμη είναι συστατικό της νοημοσύνης, ούτε και να παρουσιάσουμε σε ολοκληρωμένη μορφή τη θεωρία του καθενός εκ των προαναφερθέντων, ή άλλων, φιλοσόφων. Αντίθετα, το βιβλίο επιδιώκει να αναδείξει τις οντολογικές και ψυχολογικές παραδοχές πάνω στις οποίες, σιωπηρά ή φανερά, βασίστηκαν οι μελέτες της επιστήμης των υπολογιστών, των μαθηματικών και της γνωσιακής επιστήμης που διαμόρφωσαν το πεδίο της ΤΝ κατά τον 20ό αιώνα. Όπως θα εξηγή-

και καθορίζει τα νοήματα που διαμορφώνουν το υπόβαθρο των αξιών που εισάγονται στο τεχνούργημα. Βλ. Giannis Perperidis, “Designing Ethical A.I. Under the Current Socio-Economic Milieu: Philosophical, Political and Economic Challenges of Ethics by Design for A.I.”, *Philosophy & Technology*, 37(84), 2024.

10. Ένα εξαιρετικό βιβλίο αφιερωμένο σε μία τέτοιου είδους εξέταση (και επανεξέταση) είναι το Karolina Hübner (ed.), *Human. A History*, Οξφόρδη: Oxford University Press, 2022.

σουμε στη συνέχεια, αυτά τα θεμέλια άρχισαν να τοποθετούνται ήδη από την αυγή της νεωτερικότητας, με την Επιστημονική Επανάσταση και την ανάδυση του νεωτερικού «γνωστικού υποκειμένου», καθώς και με την προσπάθεια κυριαρχίας επί της φύσης μέσω της μαθηματικοποίησης και γεωμετρικοποίησης της πραγματικότητας. Έτσι, στόχος του βιβλίου τίθεται η «απομάγευση» της ΤΝ στη βάση της επανεξέτασης της σχέσης του ανθρώπου, της φύσης και της τεχνολογίας (δηλαδή των οντολογικών θεμελίων της τελευταίας), όπως υποδεικνύει ο τίτλος. Στο επόμενο μέρος της εισαγωγής θα εξηγήσουμε γιατί χρησιμοποιούμε τον όρο «απομάγευση».

Προς μια «απομάγευση» της Τεχνητής Νοημοσύνης

Ο όρος «απομάγευση», όπως τον χρησιμοποιούμε εδώ, δεν υποδηλώνει την πλήρη αφαίρεση του νοήματος από την ΤΝ. Δεν αποσκοπούμε στην υπονόμηση ή την κατάργηση της ΤΝ ως πεδίου έρευνας, δεν υποστηρίζουμε ότι η ΤΝ δεν έχει καμία αξία για την ανθρώπινη ύπαρξη, ούτε υποστηρίζουμε την πλήρη απορρόφηση της «ψυχής» (εν είδει νοήματος) της τεχνολογίας, κάτι που θα στόχευε στην εγκαθίδρυση ενός παν-ανθρωπισμού, με την έννοια ότι τα πάντα, απομαγεμένα και αποκαθαρμένα από κάθε άλλου είδους νόημα, στέκουν αποκλειστικά υπό το νόημα που τους αποδίδει ο άνθρωπος.¹¹ Επίσης, δεν χρησιμοποιούμε τον εν λόγω όρο με την έννοια της πραγμάτωσης μιας πλήρους υπολογιστικής κυριαρχησης στο πεδίο, όπως θα το εννοούσε ενδεχομένως ο Μαξ Βέμπερ.¹² Αντιθέτως, χρησιμοποιούμε τον όρο «απομάγευση» σε συνδυασμό με την εικόνα τού να «φέρουμε την ΤΝ ενώπιον του εαυτού της», σαν σε έναν καθρέφτη, για να τονίσουμε την ανάγκη να δούμε το

11. Βλ. τον τρόπο με τον οποίο αναφέρεται στην απομάγευση ως έλευση του παν-ανθρωπισμού ο φιλόσοφος της τεχνολογίας Αγκοστίνιο Τσέρα: Agostino Cera, «Homo Deus ή Homo Datum; Προλεγόμενα για μια φαινομενολογία της δεδομενοκρατίας», στο Γκ. Μαγγίνη και Γ. Περπερίδης (επιμ.), *Μια φιλοσοφική ανατομία του ψηφιακού. Επιστημολογία, ανθρωπολογία, πολιτική*, Αθήνα: Καρδαμίτσα-Ινστιτούτο του βιβλίου, 2025.

12. Με τον όρο «απομάγευση» ο Βέμπερ εννοεί «τη γνώση ή την πεποίθηση ότι, εφόσον το θελήσουμε, μπορούμε ανά πάσα στιγμή να κατανοήσουμε την πραγματικότητα και –κατ' αρχήν– να κυριαρχήσουμε επί των πάντων μέσω υπολογισμού» (Max Weber, «Science as a Profession and Vocation», στο *Max Weber: Collected Methodological Writings*, Λονδίνο & Νέα Υόρκη: Routledge, 2012, σ. 335–353, σ. 342).

«πραγματικό της πρόσωπο», το οποίο συνήθως αποκρύπτεται από ένα μαγικό πέπλο –σαν μανδύας–, και να συγκρίνουμε τη διαστρεβλωμένη εικόνα που έχουμε γι’ αυτή με την πραγματικότητα. Με άλλα λόγια, να την απογυμνώσουμε από τις υπερβολές και τις παρανοήσεις, οι οποίες πολλές φορές παρουσιάζονται ως «κάτι μαγικό» (μιας και καμιά δεν καταλαβαίνει πλήρως πώς λειτουργεί και πώς έχει φτάσει να είναι αυτό που είναι), ώστε να χτίσουμε τις τρέχουσες και μελλοντικές μας αναλύσεις και αφηγήσεις πάνω στην αλήθεια της πραγματικής της υπόστασης και λειτουργίας.

Αυτήν τη στιγμή, η ΤΝ παρουσιάζεται στον κόσμο σαν νθοποιός ή τηλεοπτικός παρουσιαστής, φορώντας έντονο μακιγιάζ. Το μακιγιάζ αυτό καλύπτει τις αδυναμίες, τους περιορισμούς, τα σφάλματα και τη λειτουργία της στον πραγματικό κόσμο. Εφαρμόζεται από τις εταιρείες που την αναπτύσσουν με σκοπό κυρίως το κέρδος, αλλά και από επιστήμονες που είτε εξιδανικεύουν τα αποτελέσματά τους, παρουσιάζοντας τις επιδόσεις των αλγορίθμων τους ως εφάμιλλες με εκείνες του ανθρώπου, είτε, αφελώς, πιστεύουν ακόμα σε μία ουδετερότητα της επιστήμης και της τεχνολογίας. Η διογκωμένη εικόνα, αυτή η ψευδαίσθηση, που καλύπτει τις αδυναμίες και τα όρια της ΤΝ, δημιουργείται κυρίως από τους οικονομικούς παράγοντες που επιδιώκουν να επωφεληθούν από την ΤΝ. Η απομάγευση, όπως την εννοούμε εδώ, είναι η διαδικασία κατά την οποία «θέτουμε την ΤΝ ενώπιον του πραγματικού της εαυτού». Είναι, με άλλα λόγια, η απόπειρα να «αφαιρεθεί το μακιγιάζ» της, ώστε να φανεί τι πραγματικά είναι, πέραν της διογκωμένης εικόνας των οικονομικών παραγόντων και πίσω από την εξιδανίκευση της επιστημονικής κοινότητας. Στη βάση του πραγματικού της προσώπου, της πραγματικής λειτουργίας και των πραγματικών συνθηκών υπό τις οποίες η ΤΝ αναδύεται, θα πρέπει να διαμορφώνονται οι κρίσεις μας, είτε αυτές είναι πολιτικές είτε οικονομικές είτε κοινωνικές. Επιπλέον, πρέπει πάνω σε αυτό το πρόσωπο να στηριχθούν τα οποιαδήποτε προτεινόμενα ρυθμιστικά νομοθετικά πλαίσια.

Αναφέρω τη λέξη «απομάγευση», και όχι «απομυθοποίηση» ή «κριτική», συνειδητά και στοχευμένα. Το «μαγικό» συνήθως αναφέρεται σε κάτι που μας φαίνεται απλό, αλλά δεν μπορούμε να κατανοήσουμε πώς πραγματικά συμβαίνει ή λειτουργεί. Επίσης, η μαγεία (με την έννοια μιας παράστασης μαγείας, δηλαδή ενός magic show) συχνά μας εξαπατά, στρέφοντας την προσοχή μας όπου αυτή επιθυμεί, ώστε να μην κατανοήσουμε τι πραγματικά συμβαίνει. Το «μαγικό» είναι, επομένως, κάτι που δεν κατανοούμε πλήρως, αλλά μας εντυπωσιάζει και μας αρέσει

να το βλέπουμε ή να το χρησιμοποιούμε. Στο παρελθόν, οι ταχυδακτυλουργοί χρησιμοποιούσαν τεχνολογικά τρικ για να εντυπωσιάσουν το κοινό, εκμεταλλευόμενοι την ασάφεια που περιβάλλει τη λειτουργία της τεχνολογίας. Με ανάλογο τρόπο θεωρώ ότι παρουσιάζεται σήμερα η ΤΝ: Λίγοι κατανοούν πώς λειτουργεί, πολλοί όμως τη χρησιμοποιούν στην καθημερινότητά τους, χωρίς να κατανοούν πλήρως τι είναι, ενώ την ίδια στιγμή εντυπωσιάζονται από το αποτέλεσμα. Για παράδειγμα, η πρώτη επαφή κάποιου με το ChatGPT μπορεί να φανεί «μαγική»: *Πώς μπορεί ένα λογισμικό να δημιουργήσει ένα τόσο καλό κείμενο σε τόσο λίγο χρόνο; Μαγικό...*¹³

Η απομάγευση της ΤΝ σημαίνει, συνεπώς, να αποκαλύψουμε τις βάσεις στις οποίες στηρίζεται, ώστε να κατανοήσουμε πώς είναι δυνατόν να παραχθεί ένα κείμενο από έναν υπολογιστή τόσο γρήγορα και γιατί αυτό συγκρίνεται συχνά με ένα κείμενο που έχει παραχθεί από άνθρωπο, θεωρούμενο, μερικές φορές, ακόμη και ως ανώτερο. Η απομάγευση δεν περιορίζεται στην *απομυθοποίηση* κάποιου μύθου που έχει δημιουργηθεί γύρω από την ΤΝ, δεν στοχεύει απλώς να δείξει ότι ο μύθος γύρω από κάτι είναι υπερβολικός ή λανθασμένος· αντιθέτως, η απομάγευση είναι κάτι βαθύτερο που στοχεύει στο ίδιο το αντικείμενο, όχι σε ό,τι έχει διαμορφωθεί γύρω του.

Ο μαγικός χαρακτήρας της ΤΝ ενισχύεται από την ιδέα του τεχνολογικού ντετερμινισμού, που έχει βαθιά ριζώσει στο κοινωνικό φαντασιακό της όψιμης νεωτερικότητας, δηλαδή της εποχής μας. Θεωρούμε πως η τεχνολογία εξελίσσεται αυτόνομα, βελτιώνοντας συνεχώς τις δικές της δομές και λειτουργίες, με την εξέλιξή της να ακολουθεί έναν γραμμικό δρόμο προς όλο και ανώτερα (πιο εξελιγμένα) επίπεδα. Σε αυτήν τη διαδικασία, μόνο οι ειδικοί φαίνεται να έχουν λόγο, ενώ οι «πολλοί»,

13. Σε ένα πρόσφατο δημοσίευσμά του, ο ιδρυτής της OpenAI, Σαμ Άλτμαν, αναφέρει ακριβώς: «Στις επόμενες δεκαετίες, θα είμαστε σε θέση να κάνουμε πράγματα που θα φάνταζαν μαγικά στους παππούδες και τις γιαγιάδες μας». Βλ. <https://ia.samaltman.com>. Βέβαια, στη θέση των «παππούδων και γιαγιάδων» μας μπορούν να τοποθετηθούν και οι γονείς μας, αλλά και πολλοί και πολλές της δικής μας, πιο εξοικειωμένης τεχνολογικά, γενιάς. Αν, ωστόσο, η τεχνολογία επεξηγούνταν πλήρως από αυτούς που την αναπτύσσουν (όπως π.χ. ο Άλτμαν), ίσως να μην φαίνονταν και τόσο «μαγική» εντέλει, αλλά να ήταν κατανοητή όχι απλώς ως προς τη χρήση της και ως προς τον ευρύτερο αντίκτυπό της, αλλά και ως προς πιθανές ανατροπές και αναδιαμορφώσεις της. Η «μαγεία» (με την έννοια της μυστηριώδους επένδυσης που την καθιστά ακατανόητη – και έτσι αντικείμενο επεξεργασίας και σχεδιασμού μονάχα από μερικούς ειδήμονες, δηλαδή κλειστής) γύρω από την τεχνολογία, ειδικά γύρω από την ΤΝ, εξυπηρετεί ανθρώπους σαν τον Άλτμαν και τις *Big Tech* (εταιρείες τεχνολογίας-κολοσσοί) που ο ίδιος αντιπροσωπεύει.

οι ανειδίκευτοι, παραμένουν απλοί παρατηρητές, χωρίς δυνατότητα παρέμβασης στην πορεία της τεχνολογικής εξέλιξης.

Σήμερα η τεχνολογία έχει φτάσει στο σημείο να δημιουργήσει την ΤΝ με τρόπο που οι περισσότεροι από εμάς δεν μπορούμε να κατανοήσουμε. Αποδεχόμαστε αυτή την εξέλιξη και στηριζόμαστε στις διαβεβαιώσεις ή τις προειδοποιήσεις των ειδικών, ακριβώς όπως οι θεατές ενός ταχυδακτυλουργικού τρικ ακολουθούν τις οδηγίες του μάγου εστιάζοντας εκεί που αυτός θέλει, ώστε να μην κατανοήσουν τι πραγματικά συμβαίνει *κάπου αλλού*. Η τεχνολογία, λοιπόν, διαμορφώνει και μεταλλάσσει όλες τις πτυχές της ζωής μας, αλλά παραμένει ένα ασαφές και δυσνόητο «κάτι» που μας προσφέρεται έτοιμο, χωρίς να φαίνεται ότι μπορούμε να ασκήσουμε κάποιον έλεγχο πάνω του. Αυτός είναι και ο λόγος που αυτή θεωρείται «μαγική». Η φιλοσοφία της τεχνολογίας, σε μερικές εκφάνσεις της, επιχείρησε να διαλευκάνει κάπως τον εν λόγω προβληματισμό ως προς την τεχνολογία γενικότερα. Με το παρόν βιβλίο επιχειρούμε να κάνουμε το ίδιο για την ΤΝ συγκεκριμένα.

Τι μπορεί, λοιπόν, να προσφέρει ο φιλοσοφικός αναστοχασμός στο πλαίσιο της ΤΝ; Επιστρέφοντας στους στοχαστές των απαρχών της νεωτερικότητας, μπορεί να αναδείξει τα οντολογικά και επιστημολογικά θεμέλια που διαμόρφωσαν την εποχή και που αργότερα ελήφθησαν ως δεδομένα από την επιστήμη των υπολογιστών του 20ού αιώνα. Έτσι, στα επόμενα κεφάλαια θα εξετάσουμε τη μετατόπιση της σκέψης προς τη συνείδηση, αλλά και την ανάδυση του γνωστικού υποκειμένου που παρατηρεί τη φύση και την ύλη, μαθηματικοποιώντας την. Θα εξετάσουμε τη σκέψη του Ρενέ Ντεκάρτ και τις «σαφείς και διακριτές ιδέες» του, καθώς και τη θεώρησή του για τον εκατό χαρακτήρα της ύλης. Θα συναντήσουμε το αιτιοκρατικό σύμπαν του Ντέιβιντ Χιουμ και τις αιτιακές σχέσεις μεταξύ διακριτών εντυπώσεων και ιδεών. Επιπλέον, θα επικεντρωθούμε στην ανάδυση, κατά τη νεωτερικότητα, της πολύ σημαντικής (για την ΤΝ μετέπειτα) ιδέας ότι αυστηροί κανόνες (*μέθοδοι*) διέπουν την ύλη και τον στοχασμό, και μέσω αυτών μπορεί να ξεκλειδωθεί κάθε πτυχή του κόσμου. Θα δούμε, στη συνέχεια, πώς αυτή η ιδέα οδήγησε αργότερα στην αναζήτηση μεθόδων στους ψηφιακούς υπολογιστές (υπό τη μορφή ευρετικών προγραμμάτων). Θα μελετήσουμε τη μοναδολογία του Βίλχελμ Λάιμπνιτς, τον αιτιοκρατικό τρόπο αναφοράς στους πιθανούς κόσμους, και την αρχή του αποχρώντος λόγου που ανέπτυξε. Θα παρουσιάσουμε τον Τζον Λοκ και τη θεώρηση των πραγμάτων ως οντοτήτων που αποτελούνται από μικρά σωματίδια με πρωτογενείς και δευτερογενείς ποιότητες, και τον αιτιακό τρόπο αλλη-

λεπίδρασής τους με τα αισθητηριακά όργανα του ανθρώπου. Επίσης, θα συζητήσουμε, εν συντομία, την πολιτική και επιστημολογική προσέγγιση του Τόμας Χομπς, ο οποίος αναφέρεται σε εξατομικευμένα άτομα, οι αιτιακές σχέσεις των οποίων, σε συνδυασμό με τον ανθρώπινο λόγο ως υπολογισμό, οδηγούν στη συγκρότηση πολιτικών κοινωνιών.

Οι αναγνώστριες και οι αναγνώστες, παρά τη συνεχή παράθεση φιλοσοφικών όρων (κάτι που μπορεί να ξενίσει όσες και όσους δεν είναι εξοικειωμένες με τη φιλοσοφία, όμως πρέπει να γίνει κατανοητό ότι πρέπει αναπόφευκτα να αναμετρηθούμε μαζί τους στο πλαίσιο που εξετάζουμε), θα κατανοήσουν ότι μέσω των θεωρήσεων των εν λόγω στοχαστών υφαίνεται ένας συγκεκριμένος τρόπος αντίληψης της φύσης, του κόσμου και του ανθρώπου. Αυτός ο τρόπος αντίληψης βοήθησε στην ανάδυση της επιστήμης της ΤΝ μερικούς αιώνες αργότερα. Επιστρέφοντας στους φιλοσόφους αυτούς, θα κατανοήσουμε ποιος είναι αυτός ο τρόπος ερμηνείας της φύσης και του ανθρώπου, ποια είναι τα όριά του και κατά πόσον αποτελεί τη μοναδική ορθή ερμηνεία ή όχι.

Η απομάγευση που προτείνεται μέσα από το παρόν βιβλίο θα μας βοηθήσει να εκτιμήσουμε καλύτερα το τεχνολογικό μας μέλλον. Σήμερα, η έρευνα γύρω από την ΤΝ παρουσιάζεται συχνά διαστρεβλωμένη, είτε αφελώς, λόγω της μη αναφοράς ή κατανόησης των θεμελίων της και προκρίνοντας μια τεχνολογική ουδετερότητα που θα έπρεπε να έχουμε απορρίψει από καιρό, είτε σκοπίμως, για να εξυπηρετήσει οικονομικές και πολιτικές επιδιώξεις. Είμαι βέβαιος ότι (ότ)αν οι κβαντικοί υπολογιστές αρχίσουν να χρησιμοποιούνται σε ιδιωτικές ή δημόσιες υπηρεσίες για την επίλυση καθημερινών προβλημάτων, από την οποία κάποιες εταιρείες θα κερδίζουν δισεκατομμύρια, θα χρειαστεί ένα παρόμοιο βιβλίο για να προσγειώσει τη συζήτηση που, όπως και με την ΤΝ, θα έχει διογκωθεί γύρω από την εν λόγω τεχνολογική εξέλιξη. Μόνο όταν κατανοήσουμε τι ακριβώς είναι αυτό που ονομάζουμε ΤΝ, θα μπορέσουμε να αντιληφθούμε ότι δεν «έρχεται το τέλος του κόσμου» διαμέσου αυτής. Και μόνο τότε θα αρχίσουμε να συνειδητοποιούμε ότι ίσως ο όρος «νοημοσύνη» είναι κάπως καταχρηστικός ή και αποπροσανατολιστικός. Επίσης, ίσως τότε ξεκινήσουμε να κατανοούμε ότι παίζεται ένα πολύ μεγαλύτερο πολιτικό και οικονομικό παιχνίδι, όχι απλώς γύρω από την τεχνολογία της ΤΝ, αλλά γύρω από το να πιστέψει η πλειοψηφία του κόσμου ότι η ΤΝ είναι όντως «Ν» (δηλαδή νοημοσύνη). Η ερμηνεία αυτής της τεχνολογίας ως πραγματικής νοημοσύνης είναι το μεγαλύτερο διακύβευμα για τις εταιρείες, και ο αντίκτυπος της εν λόγω τεχνολογίας διογκώνεται όσο περισσότερο πεύθονται οι άνθρω-

ποι σχετικά με αυτό. Η ΤΝ δεν θα είχε τον ίδιο κοινωνικό, πολιτικό και οικονομικό αντίκτυπο αν δεν θεωρούνταν νοημοσύνη.¹⁴ Σε αυτήν τη βάση, λοιπόν, η απομάγευση της ΤΝ, σήμερα, μέσα από τη φιλοσοφική ανάλυση των θεμελίων της, είναι πιο αναγκαία από ποτέ.¹⁵

Δομή του βιβλίου

Ο Ντρέιφους αφιερώνει σημαντικό μέρος του βιβλίου του στην ανάλυση των «παραδοχών» του πεδίου της ΤΝ που επηρέασαν τον τρόπο με τον οποίο ερμηνεύθηκαν τα ευρήματά της. Οι τέσσερις βασικές παραδοχές που εξετάζει είναι: α) η βιολογική, β) η ψυχολογική, γ) η γνωσιολογική και δ) η οντολογική. Από αυτές, η ψυχολογική και η οντολογική παραδοχή θεωρούνται οι πιο σημαντικές, με την τελευταία να λαμβάνει ακόμη μεγαλύτερη σημασία. Η οντολογική παραδοχή καθορίζει τον τρόπο με τον οποίο οι επιστήμονες της ΤΝ αντιλαμβάνονται τον κόσμο – και κατ' επέκταση, πώς οι ψηφιακοί υπολογιστές και τα συστήματα ΤΝ «αντιλαμβάνονται» την πραγματικότητα. Παρά την αναφορά ορισμένων πληροφοριών για τις ρίζες της εν λόγω οντολογίας, ο Ντρέιφους δεν εμβαθύνει στην ανάλυση, αφήνοντας ένα κενό σε αυτό το κρίσιμο ζήτημα – δηλαδή τον τρόπο με τον οποίο διαμορφώθηκε το οντολογικό υπόβαθρο που η ΤΝ έλαβε ως δεδομένο.

14. Στο πλαίσιο της Φιλοσοφίας της Τεχνολογίας γνωρίζουμε το στοιχείο αυτό μέσα από το (εδώ και σαράντα χρόνια) έργο στοχαστών όπως ο Άντριου Φίνμπεργκ, οι ανθρωπολόγοι της τεχνολογίας (Πφάφενμπεργκερ, Λεμονιέ, Λερούα-Γκουράν κ.ά.), καθώς και οι θεωρητικοί του πεδίου της κατασκευασιοκρατίας των Σπουδών Επιστήμης και Τεχνολογίας (Λω, Μπάικερ, Πιντς κ.ά.). Η αγνόηση του στοιχείου αυτού, ότι δηλαδή οφείλουμε να διαυγάσουμε την ερμηνεία (ή τη συμβολικότητα, σύμφωνα με κάποιους) που αποδίδεται στα τεχνουργήματα *πριν ή από κοινού με* τον αντίκτυπο που αυτά επιφέρουν, συνιστά σοβαρό μεθοδολογικό σφάλμα και βιβλιογραφική ασυνέπεια για μία μελέτη στο πεδίο της Φιλοσοφίας της Τεχνολογίας.

15. Μία παρόμοια απομάγευση επιχειρεί ο Αλέξανδρος Σχισμένος στο κείμενό του «Η μεταφυσική της χρονομηχανής» (*Kaboom. Ημερολόγια πριν από τη μεγάλη έκρηξη*, τ. 5, 2018, σ. 231-254) σχετικά με τον μύθο γύρω από το (φανταστικό) τεχνούργημα της χρονομηχανής. Ασχέτως του κατά πόσον μπορεί να κατασκευαστεί ένα τέτοιο τεχνούργημα, για την ώρα, το ίδιο, ακόμη και θεωρητικά, θεμελιώνεται σε μεταφυσικές-οντολογικές παραδοχές που δεν είναι στέρεες, αλλά έχουν εκληφθεί ως τέτοιες: τον χρόνο ως διακριτές και εξατομικευμένες «στιγμές». Αν ο χρόνος δεν συνιστά, εντέλει, ότι η Φυσική του 20ού αιώνα θεώρησε ότι είναι, αν οι πιο επίκαιρες αντιλήψεις περί χρόνου αποδειχθούν ορθές, τότε ένα τεχνούργημα όπως η χρονομηχανή απορρίπτεται εξ ολοκλήρου ως αθεμελιώτο ή τουλάχιστον απομαγεύεται και γειώνεται στη βάση των περιορισμών του. Κάτι παρόμοιο προσπαθούμε να κάνουμε κι εμείς στο παρόν βιβλίο, με αναφορά στην ΤΝ.

Τα οντολογικά θεμέλια της ΤΝ έχουν επηρεάσει τη διαμόρφωση των λογισμικών της, καθώς αυτά –όπως και οι επιστήμονες που εργάζονται σε αυτό το πεδίο– αντιλαμβάνονται τον κόσμο και τη γνώση με έναν συγκεκριμένο τρόπο. Πάνω σε αυτή την αντίληψη έχει οικοδομηθεί το επιστημονικό εποικοδόμημα της ΤΝ έως και σήμερα. Το παρόν βιβλίο στοχεύει να διαφωτίσει τα οντολογικά και ψυχολογικά θεμέλια που η επιστήμη της ΤΝ λαμβάνει ως δεδομένα, συχνά παραβλέποντας ότι ο άνθρωπος ίσως να μην αντικατοπτρίζεται εξ ολοκλήρου σε αυτά: μία παράβλεψη αυτού του είδους έχει τονιστεί κριτικά από φιλοσοφικές παραδόσεις του 20ού και του 21ου αιώνα.

Το πρώτο κεφάλαιο του παρόντος βιβλίου εξετάζει τα οντολογικά θεμέλια της ΤΝ, τα οποία δεν τέθηκαν τον 20ό ή τον 19ο αιώνα, αλλά πολύ νωρίτερα, κατά τον 15ο και 16ο αιώνα. Αυτό που μας ενδιαφέρει είναι να εντοπίσουμε πότε και πώς η πραγματικότητα, η φύση και η γνώση άρχισαν να γίνονται αντιληπτές ως σαφείς, διακριτές και ποσοτικοποιήσιμες έννοιες – ό,τι σήμερα θα ονομάζαμε «πληροφορία». Αυτή η κοσμοθεώρηση, που αναπτύχθηκε αιώνες πριν, είναι εκείνη που τελικά οδήγησε στην αντίληψη ότι μπορούμε να ψηφιοποιήσουμε τον κόσμο, να τον μετατρέψουμε σε διακριτές πληροφορίες και να τις εισαγάγουμε σε έναν ψηφιακό υπολογιστή.

Η ανωτέρω διαδικασία άνοιξε τον δρόμο για την ανάπτυξη της ΤΝ. Οι έρευνες στον τομέα αυτόν βασίζονται σε έναν πολύ συγκεκριμένο τρόπο αντίληψης του κόσμου, ο οποίος, αν και θεμελιωμένος σε θεωρίες στοχαστών της νεωτερικότητας, μπορεί να είναι ελλιπής και μη ολοκληρωμένος, όπως υπέδειξαν μεταγενέστερες φιλοσοφικές παραδόσεις. Σε αυτό το κεφάλαιο θα εστιάσουμε στις απαρχές αυτής της θεώρησης του κόσμου, γεγονός που θα μας βοηθήσει να κατανοήσουμε καλύτερα πώς λειτουργεί η ΤΝ και ποια είναι τα πραγματικά της όρια.

Στο πλαίσιο που σκιαγραφείται παραπάνω, θα επανεξετάσουμε τις θεωρίες ρασιοναλιστών στοχαστών όπως ο Ρενέ Ντεκάρτ και ο Γκότφριντ Βίλχελμ Λάιμπνιτς, καθώς και εμπειριστών όπως ο Τζον Λοκ και ο Ντέιβιντ Χιουμ, με κάποιες αναφορές στον Τζορτζ Μπέρκλεϊ. Επιπλέον, θα αναφερθούμε στην επιστημολογική και πολιτική θεώρηση του Τόμας Χομπς, ο οποίος πρώτος θέτει το ζήτημα του ανθρώπινου λόγου ως «υπολογιστή» (calculator), καθώς και σε ορισμένες από τις επιστημονικές κρίσεις του Γαλιλαίου, στη σκέψη του οποίου απαντάται για πρώτη φορά η μαθηματικοποίηση του κόσμου. Με αυτόν τον τρόπο, θα προσπαθήσουμε να συνθέσουμε τον ιστό της οντολογίας που διαμορφώθηκε κατά τους αιώνες της «πρώιμης» νεωτερικότη-

τας, η οποία φτάνει μέχρι τους επιστήμονες της ΤΝ, όπως δείχνει ο Ντρέιφους στο βιβλίο του.¹⁶

Τέλος, πρέπει να υπογραμμίσουμε ότι την ίδια σκιαγράφιση του οντολογικού υποβάθρου, με τις όποιες διαβαθμίσεις και μικρές μεταβολές του, θα μπορούσαμε να αναζητήσουμε και στους μεταγενέστερους φιλοσόφους του Διαφωτισμού, όπως τον Ιμμάνουελ Καντ και τους επιγόνους του, τον Γκέοργκ Χέγκελ και τους στοχαστές της Θεολογικής Σχολής του Τίμπινγκεν,¹⁷ καθώς και στον λογικό θετικισμό, την αναλυτική φιλοσοφία και τη φιλοσοφία της γλώσσας, έως και τον Μπέρτραντ Ράσελ ή τον Λούντβιχ Βίτγκενσταϊν. Ωστόσο, μία τέτοια θεματική θα ξεπερνούσε τη στόχευση του παρόντος βιβλίου, δηλαδή τη σκιαγράφιση της γένεσης του εν λόγω οντολογικού υποβάθρου στην περίοδο που εννοούμε ως απαρχές της νεωτερικότητας. Το οντολογικό υπόβαθρο και η αντίστοιχη παραδοχή διατηρούνται, με κάποιες αλλαγές, στους φιλοσόφους που αναφέρθηκαν. Εντούτοις, οι κατοπινές εποχές μέχρι τα μέσα του 20ού αιώνα, οπότε και ξεκινά, εν ολίγοις, η μελέτη της ΤΝ, μπορούν να αποτελέσουν το αντικείμενο μεταγενέστερης μελέτης.

Το δεύτερο κεφάλαιο καταπιάνεται με μια παρόμοια με το πρώτο κεφάλαιο ανάλυση. Σε αυτό θα παρουσιαστεί η λεγόμενη, σύμφωνα με τον Ντρέιφους, «ψυχολογική παραδοχή» του πεδίου της ΤΝ, δηλαδή η

16. Αν φαντάζει περίεργο ή πολύ «θεωρητικό» το να επιστρέψουμε στους φιλοσόφους της πρωτο-νεωτερικότητας για να εξετάσουμε τις αρχές και το υπόβαθρο της ΤΝ, θα πρότεινα στις αναγνώστριες και τους αναγνώστες να αναζητήσουν το βιβλίο *Artificial Intelligence. The very Idea* (1989) του Τζον Χόγκελαντ και να κοιτάζουν τις προσωπογραφίες που περιλαμβάνονται στις αρχικές του σελίδες. Αυτό διότι, ο Χόγκελαντ σε ένα βιβλίο για την «ιδέα της Τεχνητής Νοημοσύνης» συμπεριλαμβάνει τις εικόνες των Νικόλαου Κοπέρνικου, Τόμας Χομπς, Ρενέ Ντεκάρτ, Γαλιλαίου και Ντρείβιντ Χιουμ, δίπλα από (σε συνάρτηση με) εκείνες των Τσαρλς Μπάμπατζ, Τζον φον Νόιμαν, Άλεν Νιούελ και Άλαν Τούρινγκ. Η παράθεση των εν λόγω φωτογραφιών δεν είναι καθόλου τυχαία, και το βιβλίο που κρατάτε στα χέρια σας επιχειρεί να δείξει γιατί συμβαίνει αυτό. (Δεν παραπέμπω στην ελληνική μετάφραση του εν λόγω βιβλίου εδώ –John Haugeland, *Τεχνητή Νοημοσύνη. Σχεδιάζοντας τη νόηση: από την υπολογιστική θεωρία στις σύγχρονες ευφυείς μηχανές*, μτφρ. Στ. Ζαχαρίου, Αθήνα: Κάτοπτρο, 1992–, καθώς σε αυτήν, οι εν λόγω εικόνες που θα έπρεπε να κοσμούν τις σελίδες πριν από την εισαγωγή του συγγραφέα, δεν έχουν συμπεριληφθεί, κάτι που κατά τη γνώμη μου αφαιρεί από τη σημασία του βιβλίου.)

17. Όπως, άλλωστε, αναφέρει η Γκόλφω Μαγγίν: «Η μαθηματικοποίηση της φύσης είχε ήδη ξεκινήσει με τον Ντεκάρτ, αλλά είναι η καντιανή κριτική του Λόγου που οδηγεί στην αδιαφιλονίκητη φιλοσοφική δικαίωσή της» (Γκόλφω Μαγγίν, *Για μια ερμηνευτική του τεχνικού κόσμου. Από τον Χάιντεγκερ στη σύγχρονη τεχνοεπιστήμη*, Αθήνα: Πατάκη, σ. 264).

αντίληψη περί του ανθρώπου ως υπολογιστή, εν προκειμένω ως ψηφιακού υπολογιστή που επεξεργάζεται πληροφορίες. Σύμφωνα με αυτή την ανθρωπολογική προσέγγιση, η γνώση παράγεται όπως ακριβώς και σε έναν ψηφιακό υπολογιστή, ο οποίος συνιστά έναν επεξεργαστή πληροφοριών.

Όπως και στην οντολογική παραδοχή, έτσι και εδώ οι μελετητές της επιστήμης των υπολογιστών, γνωστής ως «ΤΝ», έλαβαν ως δεδομένη μια ερμηνεία για το τι είναι ο άνθρωπος και πώς παράγει γνώση, η οποία ήταν ταιριαστή για τη δημιουργία υπολογιστών που παράγουν τέτοιου είδους γνώση. Εντούτοις, όπως δείχνει και ο Ντρέιφους, η παραδοχή τους αυτή δεν υποδεικνύει τίποτε περισσότερο από την αποδοχή μιας υπόθεσης ως στέρεου και αυταπόδεικτου δεδομένου: της υπόθεσης ότι ο άνθρωπος λειτουργεί ως επεξεργαστής διακριτών και ποσοτικοποιημένων δεδομένων, λαμβάνοντας πληροφορίες ως εισροές/ερεθίσματα, επεξεργάζοντάς τις μέσω εσωτερικών νοητικών λειτουργιών και παράγοντας εκροές/αντιδράσεις.

Στο κεφάλαιο αυτό θα προσπαθήσουμε να ξετυλίξουμε το κουβάρι που θα μας οδηγήσει στις απαρχές αυτής της θεώρησης. Στην αρχή, εντοπίζουμε τον Βίλχελμ Βουντ και την πειραματική ψυχολογία που ο ίδιος προσπάθησε να διαμορφώσει. Στη συνέχεια, και ίσως σε πολύ πιο έντονο βαθμό, θα συναντήσουμε τον συμπεριφορισμό, ο οποίος, στις αρχές του 20ού αιώνα, κατέστη κυρίαρχη ερμηνεία ως προς τη λειτουργία του ανθρώπινου ψυχισμού και νου. Αυτές οι προσεγγίσεις της ψυχολογίας βοήθησαν να ερμηνευθεί ο άνθρωπος ως ένα είδος «μηχανής», η οποία επεξεργάζεται ένα συγκεκριμένο είδος πληροφορίας (όπως τα ερεθίσματα στον συμπεριφορισμό) και, ανάλογα με τις καταστάσεις, αντιδρά με έναν επίσης συγκεκριμένο τρόπο.

Όπως αναφέρει ο Ντρέιφους, η οντολογική παραδοχή διαπερνά και την ψυχολογική και, στην ουσία, την καθιστά δυνατή. Πράγματι, ο άνθρωπος φαίνεται να θεωρείται ως επεξεργαστής πληροφοριών, δεδομένου ότι ο κόσμος έχει γίνει αντιληπτός ως σαφείς και διακριτές μονάδες πληροφορίας. Αν δεν υπήρχαν τέτοιου είδους πληροφορίες (αν, δηλαδή, δεν νοούνταν ο κόσμος ως διακριτές ψηφίδες έτοιμες προς επεξεργασία), δεν θα είχε νόημα να θεωρηθεί ο άνθρωπος ως επεξεργαστής τους – θα ήταν σαν να προσπαθούμε να φτιάξουμε μια μηχανή επεξεργασίας γάλακτος, χωρίς να υπάρχει στον κόσμο το γάλα. Το κεφάλαιο αυτό θα προσπαθήσει, συνεπώς, να δείξει ποιες μελέτες και απόψεις επηρέασαν την ερμηνεία του ανθρώπου ως «ψηφιακού υπολογιστή» πριν καν δημιουργηθούν ψηφιακοί υπολογιστές τεχνολογικά.

Το τρίτο κεφάλαιο του παρόντος βιβλίου αναδεικνύει μερικές από τις συνέπειες που απορρέουν από την ερμηνεία της ΤΝ υπό την οντολογική και ψυχολογική παραδοχή. Στο πλαίσιο αυτό, θα εξετάσουμε τις ήδη εκδηλωμένες δυσκολίες και προκλήσεις που επιφέρει η ΤΝ, αφενός με τη σύμπραξή της με τον λεγόμενο «καπιταλισμό της επιτήρησης» και την ολοένα και μεγαλύτερη ανάγκη δεδομενοποίησης και ποσοτικοποίησης του κόσμου, και αφετέρου με την ανάγκη για δημιουργία ολοένα και περισσότερων κέντρων δεδομένων (data centers), και το περιβαλλοντικό αποτύπωμα αυτών. Επιπλέον, θα ασκήσουμε κριτική σε πασίγνωστες δυστοπικές αφηγήσεις που έχουν αναδυθεί τις τελευταίες δεκαετίες και επηρεάζουν τη σκέψη των περισσότερων ανθρώπων γύρω από την ΤΝ. Πιο συγκεκριμένα, θα παρουσιάσουμε το έργο του Ρέι Κέρζουελ, ο οποίος θεωρείται σήμερα γκουρού για τις προβλέψεις του μέλλοντος σχετικά με την ΤΝ και την τεχνολογία γενικότερα, και θα προσπαθήσουμε να ασκήσουμε κριτική στη βάση του ότι και αυτός, προκειμένου να κάνει τις προβλέψεις του, αποδέχεται άκριτα μία μορφή του κόσμου και του ανθρώπου, η οποία δεν πρέπει να θεωρείται δεδομένη, αλλά αντιθέτως ως μία από τις πολλές υποθέσεις (δηλαδή, όπως οι ερευνητές της ΤΝ του 20ού αιώνα, αποδέχεται άκριτα την οντολογική παραδοχή).

Το τέταρτο κεφάλαιο παρουσιάζει τον πραγματικό κίνδυνο που εγείρεται γύρω από και μέσω της ΤΝ, όπως αυτή αναπτύσσεται σήμερα – δηλαδή, με βάση τις παραδοχές που σχολιάζουμε. Σε αυτό το κεφάλαιο βασιζόμαστε στην έννοια της «περιτύλιξης» που εισάγει ο φιλόσοφος της τεχνολογίας και της πληροφορίας Λουτσιάνο Φλορίντι και τη συνδυάζουμε με την κριτική που έχει ασκήσει κατά της δυνατότητας ανάπτυξης νοημοσύνης από τους ηλεκτρονικούς υπολογιστές ο Τζον Σιρλ ήδη από τη δεκαετία του 1990. Στη βάση αυτήν, επιδιώκουμε να δείξουμε ότι υπάρχουν συγκεκριμένοι τομείς της ζωής (εκπαίδευση, φιλοσοφία, τέχνη) που δεν μπορούν να προσομοιωθούν από την ΤΝ και ότι η προσπάθεια υλοποίησης αυτής της προσομοίωσης ίσως εγκυμνεί κινδύνους για το παρόν και το μέλλον (σήμερα είναι ορατή στο περιβάλλον, όπως θα δούμε).

Το πέμπτο κεφάλαιο εξετάζει φιλοσοφικές προοπτικές και παραδόσεις που ασκούν κριτική στην οντολογική παραδοχή και προσπαθούν να την υπερβούν. Αυτές οι παραδόσεις είναι γνωστές για άλλους λόγους στον τομέα της φιλοσοφίας, αλλά επειδή ξεκινούν με μια κριτική της επιστημολογίας και επεκτείνονται σε κοινωνικές και πολιτικές κριτικές, είναι σημαντικό να τις επανεξετάσουμε σήμερα στο πλαίσιο της ΤΝ.

Ακολουθώντας το παράδειγμα του Ντρέιφους, ο οποίος προτείνει τη φαινομενολογική προοπτική ως απάντηση στην ΤΝ, θα προσπαθήσουμε να εξετάσουμε εάν και άλλες φιλοσοφικές θεωρήσεις μπορούν να υποδείξουν διαφορετικούς τρόπους του Είναι και του νοείν, τους οποίους η ΤΝ μέχρι σήμερα δεν μπορεί να μιμηθεί, να προσομοιώσει ή να αναπαράγει. Στο κεφάλαιο αυτό θα παραμείνουν ανοιχτά ερωτήματα σχετικά με το κατά πόσον ένα λογισμικό ΤΝ θα μπορούσε να λειτουργήσει με βάση τη λογική που προτείνεται από τις θεωρίες που θα εξεταστούν.

Στο ανωτέρω πλαίσιο, θα επανεξετάσουμε όψεις της φαινομενολογίας, δηλαδή τον τρόπο με τον οποίο αυτή η παράδοση αναφέρεται στην ενσώματη εμπειρία, σε αντίθεση με την αντίληψη του νου που προτείνει η οντολογική παραδοχή της ΤΝ. Επιπλέον, θα αναφερθούμε στην κριτική του εργαλειακού λόγου, όπως αυτή αναπτύσσεται από τους Μαξ Χορκχάιμερ και Τέοντορ Αντόρνο στη *Διαλεκτική του Διαφωτισμού*, καθώς παραδεχόμαστε ότι ο λόγος έτσι όπως τον θεωρεί η Κριτική Θεωρία τροφοδοτεί μεθοδολογικά την έρευνα της ΤΝ. Τέλος, περνώντας στον 21ο αιώνα, θα εξετάσουμε το θεωρητικό έργο της Κορίν Πελλυσόν, το οποίο συνδυάζει επιστημολογία, ηθική και πολιτική σκιαγραφώντας μια ανθρώπινη νοημοσύνη άλλου είδους, και γι' αυτό μπορεί να προσφέρει γόνιμο έδαφος για μια νέα κριτική στην ΤΝ.

Οδεύοντας προς το τέλος της εισαγωγής, δεν θα μπορούσε να διατυπωθεί ο στόχος του παρόντος βιβλίου με καλύτερο τρόπο απ' ό,τι με τα ίδια τα λόγια του Ντρέιφους:

Η ώθηση [εννοεί την οντολογική παραδοχή] που έχει δημιουργηθεί από την αμοιβαία ενίσχυση δύο χιλιετιών παράδοσης και του προϊόντος της [ως «προϊόν» εννοεί την ψυχολογική παραδοχή], της ισχυρότερης συσκευής που επινόησε ποτέ ο άνθρωπος, είναι απλώς πολύ μεγάλη για να αναχαιτιστεί, να εκτραπεί ή έστω να κατανοηθεί πλήρως. Αυτό που μπορεί κανείς να ελπίζει στην καλύτερη περίπτωση είναι να αντιληφθούμε πως η κατεύθυνση που έχουν πάρει τα πράγματα, αν και αναπόφευκτη, δεν είναι η μόνη δυνατή· πως οι παραδοχές στις οποίες βασίζεται η βεβαιότητα για το εφικτό της τεχνητής λογικής είναι παραδοχές και όχι αξιώματα – εν ολίγοις ότι μπορεί να υπάρχει ένας εναλλακτικός τρόπος αντίληψης της ανθρώπινης λογικής ο οποίος εξηγεί τόσο γιατί το παράδειγμα του υπολογιστή είναι ακαταμάχητο όσο και γιατί είναι καταδικασμένο να αποτύχει.¹⁸

18. Dreyfus, *Τι δεν μπορούν ακόμη να κάνουν οι υπολογιστές*, όπ.π., σ. 405-406.

Εν συνόλω, το βιβλίο θα επιχειρήσει να αναδείξει ένα σημαντικό μέρος αυτής της «ώθησης», καθώς και ένα μέρος του «προϊόντος», ενώ στη συνέχεια θα επιχειρήσει να δείξει εναλλακτικούς τρόπους, με βάση τους οποίους η ΤΝ σήμερα ούτε «έξυπνη» θα θεωρούνταν ούτε *νοημοσύνη*. Με αυτόν τον τρόπο, το βιβλίο συμβάλλει, κατά τη γνώμη μας, στην καταπολέμηση βεβιασμένων και θεμελιωμένων σε ασυνεπείς βάσεις δυστοπικών φαντασιοκοπημάτων και υποθέσεων για το μέλλον.

Δύο τελευταίες επισημάνσεις προτού εισέλθουμε στο κυρίως σώμα του βιβλίου:

Πρώτον, παρόλο που φαίνεται ότι θα περιηγηθούμε σε θεωρίες φιλοσόφων σαν σε εισαγωγή στη φιλοσοφία ή σαν να αναλύουμε μονάχα τη σκέψη και το έργο τους, ο σκοπός μας εδώ δεν είναι αυτός. Δεν θα παρουσιάσουμε ούτε θα αναλύσουμε εξαντλητικά το έργο και τις απόψεις των φιλοσόφων στους οποίους θα αναφερθούμε. Στόχος είναι να σκιαγραφηθεί το οντολογικό υπόβαθρο που ξεκίνησε να διαμορφώνεται κατά την πρώιμη νεωτερικότητα και το οποίο, μερικούς αιώνες μετά, αποτέλεσε τη βάση των ερευνών της ΤΝ. Τα τμήματα των θεωριών των φιλοσόφων στους οποίους θα αναφερθούμε θα βοηθήσουν να ξετυλιχθεί το κουβάρι αυτού του υποβάθρου. Αυτό σημαίνει ότι δεν θα επιχειρήσουμε μια διεξοδική παρουσίαση των απόψεων του Ντεκάρτ ή του Λοκ, αλλά τη σκιαγράφηση του υποβάθρου που προϋποτίθεται για την ανάπτυξη των ιδεών τους, καθώς και του υποβάθρου που οι ίδιοι διαμόρφωσαν και καθόρισαν με τις σκέψεις τους.

Δεύτερον, έχουμε ήδη αναφερθεί πολλές φορές στο βιβλίο και την ανάλυση του Ντρεϊφους σαν να την προϋποθέτουμε σε μεγάλο βαθμό, χωρίς, ωστόσο, να παρουσιάσουμε αναλυτικά το επιχείρημά του. Θα κλείσουμε την παρούσα εισαγωγή με ένα σύντομο τμήμα σχετικά με την κριτική του Ντρεϊφους στην ΤΝ. Το τμήμα αυτό είναι σύντομο, διότι, από τη μία, θα επανερχόμαστε διαρκώς στην κριτική του σε διάφορα μέρη των επόμενων κεφαλαίων, κάτι που θα βοηθήσει στη σκιαγράφηση των απόψεών του σε μεγαλύτερο βάθος· από την άλλη, διότι προσωπικά ελπίζω το παρόν βιβλίο να αποτελέσει εφελτήριο για τη μελέτη του έργου του Ντρεϊφους (ή την ανάγνωσή του ξανά υπό νέα πρίσματα, από αυτές και αυτούς που το έχουν μελετήσει ήδη).

Σύντομη παρουσίαση της κριτικής του Χιούμπερτ Ντρεΐφους στην Τεχνητή Νοημοσύνη

Η κριτική του Ντρεΐφους στην ΤΝ είναι εξαιρετικά σημαντική, καθώς εγείρει σοβαρά ερωτήματα για τα όρια και τις δυνατότητες των υπολογιστών στη μίμηση της ανθρώπινης σκέψης. Ο Ντρεΐφους, ως φαινομενολόγος, ήταν από τους πρώτους που αμφισβήτησαν την ιδέα ότι οι υπολογιστές μπορούν να αναπαραστήσουν ρητά και να προσομοιώσουν συνολικά την ανθρώπινη νοημοσύνη με ουσιαστικό τρόπο.

Στα έργα του *Τι δεν μπορούν να κάνουν οι υπολογιστές* (1972) και *Τι δεν μπορούν ακόμη να κάνουν οι υπολογιστές* (1992), ο Ντρεΐφους ασκεί δριμεία κριτική στις θεμελιώδεις παραδοχές των ερευνών στον τομέα της ΤΝ. Όπως έχουμε ήδη πει παραπάνω, ένα από τα κεντρικά σημεία της κριτικής του είναι η έννοια της «οντολογικής παραδοχής». Με αυτή την έννοια, υποστηρίζει ότι οι θεωρίες της ΤΝ βασίζονται στην παραδοχή ότι όλη η γνώση και η κατανόηση του κόσμου μπορούν να αναπαρασταθούν με λογικά σύμβολα, μαθηματικούς κανόνες και αλγόριθμους, με απόλυτη ακρίβεια. Η θεώρηση αυτή προϋποθέτει ότι το ζητούμενο για την ανάπτυξη νοημοσύνης είναι να βρεθεί το κατάλληλο «ευρετικό πρόγραμμα» και να εισαχθούν σε αυτό τα κατάλληλα δεδομένα. Αυτή η άποψη έχει τις ρίζες της στη φιλοσοφία του 17ου αιώνα, με φιλοσόφους όπως ο Ντεκάρτ, ο Λοκ, ο Χιουμ και ο Λάιμπνιτς να διαμορφώνουν μια κοσμοαντίληψη που θεωρεί ότι ο κόσμος και η ανθρώπινη σκέψη αναλύονται σε μεμονωμένα και σαφώς ορισμένα στοιχεία.¹⁹

Η εναντίωση του Ντρεΐφους έχει τις ρίζες της στη φαινομενολογία. Ο ίδιος υποστηρίζει ότι η ανθρώπινη γνώση και εμπειρία δεν μπορούν να περιοριστούν σε σύνολα ρητών και αντικειμενικών δεδομένων. Επιπλέον, ότι η ανθρώπινη νοημοσύνη δεν λειτουργεί μέσω ακριβούς και συνειδητής εφαρμογής κανόνων, όπως πιστεύει η έρευνα της ΤΝ, αλλά είναι πάντοτε εγκιβωτισμένη σε ένα ευρύτερο πλαίσιο ενσώματης εμπειρίας, ενστίκτων, διαισθητικής κατανόησης και κοινωνικών αλληλεπιδράσεων. Ο άνθρωπος δεν κατανοεί τον κόσμο ως μια συλλογή δεδομένων στη βάση ρητών κανόνων, αλλά μέσω της ενσώματης και βιωματικής του σχέσης με το περιβάλλον.

Για την ανάπτυξη της κριτικής του στην ΤΝ, ο Ντρεΐφους αντλεί από τη φαινομενολογία των Έντμουντ Χούσερλ, Μάρτιν Χάιντεγκερ και Μωρίς Μερλώ-Ποντύ, υποστηρίζοντας ότι η ανθρώπινη ύπαρξη και

19. Στο ίδιο, σσ. 371-398.

κατανόηση είναι προσανατολισμένες προς τον κόσμο με τρόπους που δεν μπορούν να αναπαραχθούν από μια μηχανή. Για παράδειγμα, όταν κάποιος χειρίζεται ένα εργαλείο ή αλληλεπιδρά με άλλους, δεν ακολουθεί συνειδητά μια λίστα κανόνων, αλλά βασίζεται σε μία διαισθητική και ενσώματη γνώση, η οποία είναι βαθιά ριζωμένη στην καθημερινή του εμπειρία. Αυτός ο τύπος γνώσης, ο εξαρτημένος από το περιβάλλον και, τις περισσότερες φορές, μέσα από αλληλεπίδραση με άρρητες εκφάνσεις του, δεν μπορεί να κωδικοποιηθεί και να αναπαρασταθεί με πλήρη και ρητό τρόπο, όπως το απαιτεί η ΤΝ.

Ολοκληρώνοντας, είναι σημαντικό να υπογραμμίσουμε πως ο Ντρέιφους επεσήμανε τα όρια των προγραμμάτων της ΤΝ ως προς τη διαχείριση της σημασιολογίας και της κατανόησης του πλαισίου. Σύμφωνα με τον ίδιο, οι υπολογιστές μπορούν να επεξεργάζονται σύμβολα, αλλά δεν μπορούν να κατανοήσουν το νόημά τους με τον τρόπο που το κάνει ο άνθρωπος, επειδή τους λείπει η ενσωμάτωση στον πραγματικό κόσμο (τον βίοκοσμο – *Lebenswelt*) και η βιωμένη εμπειρία σε αυτόν. Η ΤΝ αποτυγχάνει να προσεγγίσει τη ρευστή και εύπλαστη φύση της ανθρώπινης σκέψης, η οποία είναι προσαρμοσμένη σε καταστάσεις και κοινωνικούς δεσμούς αλληλεπίδρασης και όχι σε μια στεία επιλογή μεταβλητών μέσα από έναν πίνακα υποθέσεων και πιθανών αντιδράσεων.

Και εδώ, όπως θα συμβαίνει και στα επόμενα κεφάλαια σε σχέση με θεωρίες όπως εκείνη του Ντεκάρτ ή του Χιουμ, η παρουσίασή μου δεν επιχειρεί να καλύψει όλες τις διαστάσεις της σκέψης του Ντρέιφους, αλλά να δώσει μια γεύση της κριτικής του. Είμαι της άποψης ότι ο αναγνώστης και η αναγνώστρια των οποίων το ενδιαφέρον θα καταφέρω να κεντρίσω με το παρόν βιβλίο, θα ανατρέξουν σε κάποιο από τα βιβλία στα οποία αναφέρομαι για περαιτέρω μελέτη και επέκταση της προβληματικής. Αυτός είναι, άλλωστε, ο κρυφός πόθος που με ώθησε στη συγγραφή του παρόντος βιβλίου: η επέκταση της κριτικής σκέψης ολοένα και μεγαλύτερου τμήματος του κοινού, σχετικά με τη νέα τεχνολογία, με την οποία καλούμαστε σήμερα, και θα κληθούμε ίσως πολύ περισσότερο αύριο, να αναμετρηθούμε.