

ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΠΟΛΙΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ  
ΤΜΗΜΑ ΠΟΛΙΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ  
Π.Μ.Σ.: ΠΟΛΙΤΙΚΗ ΕΠΙΣΤΗΜΗ ΚΑΙ ΚΟΙΝΩΝΙΟΛΟΓΙΑ

### ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του Συμεών Κάλφα

A.M.:7342112200006

## Αόρατα Νήματα: Τεχνητή Νοημοσύνη, Γνώση και Αυθεντία

### Επιβλέποντες:

- α) Γεράσιμος Κουζέλης, Ομότιμος Καθηγητής, Τμήμα Πολιτικής Επιστήμης και Δημόσιας Διοίκησης, ΕΚΠΑ
- β) Κανάκης Λελεδάκης, Επίκουρος Καθηγητής, Τμήμα Πολιτικής Επιστήμης και Δημόσιας Διοίκησης, ΕΚΠΑ
- γ) Λέανδρος Κυριακόπουλος, Επίκουρος Καθηγητής, Τμήμα Κοινωνιολογίας, ΕΚΠΑ

Αθήνα, 2024

## Περιεχόμενα

Εισαγωγή .....	3
Η τεχνητή νοημοσύνη μέσα από τα γραπτά του Άλαν Τούρινγκ .....	8
Η Μαριονέτα .....	10
Τι είναι τα συστήματα βαθιάς μάθησης .....	17
Αυθεντία και γνώση .....	20
Για την Αυτοματοποίηση .....	33
AlphaFold .....	40
Σκακιστικές Μηχανές .....	46
Η TN ως καταστροφέας των κανόνων και η λύση του προβλήματος.....	54
Δύο Προοπτικές .....	61
Η Μαριονέτα και ο Πολύφημος .....	64
Το Σάμποργκ και ο Νίο .....	65
Επίλογος .....	68
Βιβλιογραφία .....	71

## Εισαγωγή

Η βασική μου βιβλιογραφία για αυτήν την εισαγωγή είναι το βιβλίο *Τεχνητή Νοημοσύνη: Μια Σύγχρονη Προσέγγιση (AIMA)* των Ράσελ και Νόρβιγκ<sup>1</sup>, και το *The Interface Effect* του Αλεξάντερ Γκάλογουεϊ<sup>2</sup>. Το πρώτο είναι ένας τόμος που ανανεώνεται συνεχώς και είναι τόσο εδραιωμένος ως το βασικό σύγγραμμα για την τεχνητή νοημοσύνη (TN) που είναι ασφαλές να θεωρήσουμε ότι κάθε φοιτητής και φοιτήτρια που έχει παρακολουθήσει μάθημα για την TN σε πολυτεχνείο το έχει διαβάσει<sup>3</sup>. Το δεύτερο είναι μια εξαιρετική πολιτική και κοινωνική θεωρητική ανάλυση για τα «Νέα Μέσα», τα επιχειρήματα της οποίας είναι πολύ κοντινά σε αυτά που θα ήθελα να αναπτύξω. Η αρχική μου ιδέα όμως ήταν να χρησιμοποιήσω εισαγωγικά αποκλειστικά το πρώτο, με σκοπό να συνθέσω μια συστηματική ανάλυση του τι είναι η TN μέσα από τα λόγια των ειδημόνων που την κατασκευάζουν. Η δυσαρμονία μεταξύ των δύο κειμένων είναι πραγματικά εντυπωσιακή. Συγκεκριμένα για τον Γκάλογουεϊ, αυτή η δυσαρμονία έχει σε μεγάλο βαθμό να κάνει με τη μεθοδολογία του, καθώς θέλει να παρουσιάσει πώς η μηχανή ξεφεύγει από την περιγραφή της και θα πρέπει να αναλύεται ως πολιτικό φαινόμενο. Το χάσμα όμως είναι λιγότερο διαχειρίσιμο σε πολλή από τη βιβλιογραφία που μπόρεσα να εντοπίσω, με τις εξαιρέσεις να συγκεντρώνονται αισθητά στις πραγματεύσεις της καινοτόμας τεχνολογίας πριν, και αμέσως μετά την εμφάνισή της. Από αυτήν την άποψη η TN έχει ήδη γεράσει. Γεννήθηκε μάλλον γερασμένη.

Οι Ράσελ και Νόρβιγκ, στην εισαγωγή τους *Τεχνητή Νοημοσύνη, μια Σύγχρονη Προσέγγιση*<sup>4</sup> παρουσιάζουν τους ορισμούς της σε δύο άξονες: Είτε σκέψη είτε δράση· είτε ανθρώπινη είτε ορθολογική. Μπορούμε να φανταστούμε πώς μηχανές των οποίων οι δυνατότητες ανήκουν σε οποιαδήποτε από τις τέσσερις κατηγορίες (ανθρώπινη σκέψη/πράξη, ορθολογική σκέψη/πράξη) θα μπορούσαν να αποκαλούνται τεχνητή νοημοσύνη. Στην κατηγορία της ανθρώπινης πράξης και σκέψης μπαίνουν, για παράδειγμα, οι αλγόριθμοι που κινούν τους χαρακτήρες σε ένα παιχνίδι. Τα πλάσματα αυτά πρέπει συνήθως να είναι αρκετά προβλέψιμα ώστε ο παίκτης να μπορεί να τα διαχειριστεί, αλλά ταυτόχρονα να αλληλεπιδρούν μαζί του

---

<sup>1</sup>Stuart Russel και Peter Norvig, *Τεχνητή Νοημοσύνη: Μια Σύγχρονη Προσέγγιση* (Αθήνα: Κλειδάριθμος, 2021).

<sup>2</sup>Alexander R. Galloway, *The Interface Effect* (Polity, 2012).

<sup>3</sup>'*Artificial Intelligence: A Modern Approach*', σε *Wikipedia*, 3 Μαΐου 2024, [https://en.wikipedia.org/w/index.php?title=Artificial\\_Intelligence:\\_A\\_Modern\\_Approach&oldid=1222113628](https://en.wikipedia.org/w/index.php?title=Artificial_Intelligence:_A_Modern_Approach&oldid=1222113628). Υπολογίζεται ότι το βιβλίο αναφέρεται μέχρι στιγμής σε περίπου 59,000 άρθρα

<sup>4</sup>Stuart Russel και Peter Norvig, 15–20.

και με το περιβάλλον τους με τρόπους αρκούντως «φυσικούς» και απρόβλεπτους. Το ζητούμενο εδώ είναι η επίφαση της αυτονομίας, η οποία σε συγκεκριμένες περιπτώσεις μπορεί να ενθαρρύνει και την ταύτιση. Όπως θα δούμε και παρακάτω, η αδυναμία να κατανοήσει κανείς και να προβλέψει το σύστημα που παρατηρεί, είτε λόγω περιπλοκότητας είτε λόγω εσκεμμένης απόκρυψης των εσωτερικών του διαδικασιών, εύκολα δημιουργεί αυτήν την αντίληψη.

Η προσέγγιση της λογικής σκέψης αντιπροσωπεύει σε αδρές γραμμές την προηγούμενη φάση της έρευνας γύρω από την τεχνητή νοημοσύνη. Τα συστήματα αυτής της προσέγγισης προσπαθούν να προσομοιώσουν κανονιστικά μοντέλα της λογικής, κυρίως για να καταφέρνουν να απαντούν σε ερωτήσεις ή να αποδεικνύουν και να απορρίπτουν λογικές προτάσεις. Αυτή η προσέγγιση γρήγορα αποδείχθηκε περιορισμένη στις εφαρμογές της, καθώς τα κανονιστικά μοντέλα της λογικής απαιτούν βεβαιότητα:

Η λογική, κατά τη συμβατική ερμηνεία της, απαιτεί τη γνώση του κόσμου να είναι *βέβαιη* – μια συνθήκη που στην πραγματικότητα σπάνια επιτυγχάνεται. Δεν μπορούμε να γνωρίζουμε τους κανόνες, για παράδειγμα, της πολιτικής ή των πολιτικών συγκρούσεων όπως γνωρίζουμε τους κανόνες του σκακιού ή της αριθμητικής. Η θεωρία των πιθανοτήτων καλύπτει αυτό το κενό επιτρέποντας τη συλλογιστική με αβέβαιες πληροφορίες. Ουσιαστικά επιτρέπει την κατασκευή ενός ολοκληρωμένου μοντέλου ορθολογικής σκέψης, το οποίο οδηγεί από ανεπεξέργαστες αντιληπτικές πληροφορίες στην κατανόηση του τρόπου με τον οποίο λειτουργεί ο κόσμος, και σε προβλέψεις για το μέλλον. Αυτό που δεν κάνει είναι να παράγει ευφυή *συμπεριφορά*. Για τον λόγο αυτό χρειαζόμαστε μια θεωρία για την ορθολογική δράση. Η λογική σκέψη από μόνη της, δεν είναι αρκετή<sup>5</sup>.

Η αβεβαιότητα του πραγματικού κόσμου λοιπόν θέτει περιορισμούς στην λειτουργικότητα της λογικής, ενώ επίσης διαχωρίζει (για τους συγγραφείς) τη *λογική* από την *ορθολογικότητα*. Η ορθολογική σκέψη, σε σύγκριση με τη λογική, έχει να κάνει με τη διαχείριση μιας ομιχλώδους κατάστασης προς όφελος ενός *δράστη*<sup>6</sup>. Η έρευνα έτσι στρέφεται προς τη συμπεριφορά και την επιλογή, την επιδίωξη στόχων. Σε αυτήν την υποκατηγορία η πολυμαθία είναι δευτερεύουσα· το βασικό είναι να μπορεί το μοντέλο να συμπεριφερθεί αποτελεσματικά ακόμα και σε νέες

---

<sup>5</sup>Stuart Russel και Peter Norvig, 18.

<sup>6</sup>actor – αυτό που δρα. Στη βιβλιογραφία των κοινωνικών επιστημών ο όρος μεταφράζεται πιο σωστά ως *δρώντας*. Βλ. ο.π. σελίδα 670

περιστάσεις. Υπάρχουν έτσι προβλήματα για τα οποία ο συμπερασμός μέσω λογικής δεν παράγει τα επιθυμητά αποτελέσματα, καθώς πρέπει να βασίζεται σε ήδη ενυπάρχουσες περιγραφές που να αντιστοιχούν τέλεια στην εκάστοτε κατάσταση. Σε αβέβαιο περιβάλλον, με τη χρήση της θεωρίας πιθανοτήτων, ο ορθολογικός πράκτορας<sup>7</sup> πρέπει να μπορεί να επιλέγει το καλύτερο αναμενόμενο αποτέλεσμα, καθώς το καλύτερο αποτέλεσμα δεν μπορεί να υπολογιστεί. Τα μοντέλα αυτού του τύπου χρειάζονται ένα τρόπο να κρίνουν ποια επιλογή τους είναι πιο πιθανό να οδηγήσει σε μεγαλύτερο κέρδος, και ένα τρόπο να κρίνουν το τι είναι επικερδές – έναν στόχο. Ο στόχος δεν χρειάζεται να είναι μοναδικός – πολλές φορές κάτι τέτοιο θα ήταν επικίνδυνο: Αν πρέπει το μοντέλο να κάνει οτιδήποτε για την επίτευξη του στόχου του, τότε μπορεί να επιλέξει «λάθος» μέσα, όπως, για να χρησιμοποιήσω το κλασικό παράδειγμα, να χτυπήσει τον αντίπαλο σκακιστή ώστε να νικήσει το παιχνίδι<sup>8</sup>.

Οι όροι που χρησιμοποιούνται παραπάνω βγάζουν μάτι για έναν κοινωνικό επιστήμονα ως μη δόκιμοι – η μετάφραση του *actor* ως *δράστη* είναι μάλλον το πιο τρανταχτό παράδειγμα. Έτσι όμως μεταφράζονται και χρησιμοποιούνται στην τεχνική βιβλιογραφία<sup>9</sup>. Η μετάφραση δρώντας, που θα συνέδεε το κείμενο παραπάνω με τη θεωρία δρώντος-δικτύου (*actor-network theory*) είναι μεν πιο σωστή και πιο μεστή νοήματος ως σύνδεση μεταξύ αναλύσεων, δημιουργούνται όμως αμφιβολίες και για το κατά πόσο η χρήση του *actor* στο πρωτότυπο έχει αυτές τις καταβολές. Παρομοίως και για το *πράκτορας* ως μετάφραση του *agent* – το *εμπρόθετο υποκείμενο* ομολογουμένως προτρέπει στη συζήτηση. Το σύγγραμμα που χρησιμοποιώ απαιτεί έτσι ιδιαίτερη μεταχείριση εξαρχής – ο πειρασμός να επιλέξω διατυπώσεις και να τις παραθέσω με εμπάθεια αναλύοντάς τις κριτικά είναι τόσο συνήθης που έχει καταντήσει βαρετός. Μια τέτοια στάση όμως θα κινδύνευε να παραβλέψει το πόσο εντυπωσιακό είναι το κείμενο ως συλλογή μηχανιστικής γνώσης για το υποκείμενο, τη συμπεριφορά, την επικοινωνία, την κοινωνία, αλλά και την ίδια τη γνώση. Αυτή η κριτική στη γλώσσα του βιβλίου βέβαια θα μπορούσε να είναι γόνιμη – υπάρχουν μάλλον και άλλες διατυπώσεις που θα μπορούσαν να συνθέσουν έναν παρόμοιο

---

<sup>7</sup>agent – αυτό που δρα αυτόνομα/με πρόθεση. Όπως και παραπάνω αντί του αυτενεργό ή εμπρόθετο υποκείμενο. Βλ. ο.π. σελίδα 18

<sup>8</sup>Stuart Russel και Peter Norvig, 20. Η λύση που προτείνουν οι συγγραφείς σε αυτό το ζήτημα είναι η δημιουργία μηχανών που είναι αβέβαιες για τους στόχους τους. Με αυτόν τον τρόπο, λένε, αναγκαστικά η μηχανή θα παραμείνει υπό τον έλεγχό μας, καθώς πρέπει να μαθαίνει τους στόχους της συνεχώς. Έτσι θα έχουμε «όχι απλά λειτουργικές αλλά αποδεδειγμένα επωφελείς μηχανές.» Ίσως βέβαια η λειτουργία τις στοχοθεσίας να αποδειχθεί παρομοίως βλαβερή – αλλά αυτού του είδους η εσχατολογία είναι κάτι που θα αποφύγω να συζητήσω.

<sup>9</sup>Γλωσσάρι EN', *Τεχνητή Νοημοσύνη* (blog), τελευταία πρόσβαση 29 Σεπτεμβρίου 2024, <https://aibook.gr/glossary-en/>.

τόμο. Το επικίνδυνο όμως είναι να παραβλέψουμε ότι η λογική του κειμένου είναι *απολαυστική* – και το πόσο δυσαρμονική φαίνεται κατά καιρούς να είναι με τις ηθικές αναλύσεις που περιλαμβάνονται.

Το θέμα του παρόντος κειμένου είναι οι κανόνες και ο έλεγχός μας πάνω σε αυτούς μέσω του Λόγου. Θα ήταν έτσι επόμενο να συνεχίσω την παραπάνω ανάλυση συγκρίνοντας τη μηχανιστική σκέψη των μηχανικών με τη μηχανιστική σκέψη των κοινωνιολόγων. Θα μπορούσα έτσι να παραθέσω τις αναλύσεις του Βέμπερ ή του Μπουρντιέ και να επιδοθώ σε μια συγκριτική ανάλυση που θα είχε σίγουρα ενδιαφέρον. Δεν έχω όμως δυστυχώς το χώρο να κάνω κάτι τέτοιο. Θα επιλέξω αντίθετα να ξεκινήσω από μια φράση του Μπακούνιν που, στην απλότητά της, φέρνει το θέμα στην ουσία του.

[Το ζητούμενο του σοσιαλισμού είναι να] εντάξει τα υποκείμενα σε κατανοήσιμους και ελέγξιμους θεσμούς<sup>10</sup>.

Το ζήτημα του Διαφωτισμού ως θεωρίας της χειραφέτησης είναι να αναδημιουργήσει τους θεσμούς σε αυτήν τη μορφή εκ των προτέρων – διαδικασία που ταυτίζεται με τη δημιουργία μιας λειτουργικής μηχανής. Να λύσει την αντίφαση της ιδέας ότι η κοινωνία δημιουργείται από τα υποκείμενα αλλά τα καταδυναστεύει. Να αφήσει το υποκείμενο, ως κυρίαρχο, ως το μόνο μη ελέγξιμο και κατανοήσιμο στοιχείο στην κοσμοθεώρησή μας. Η ανάλυση των θεσμών εκ των υστέρων, ως δεδομένων δηλαδή σε κάποιο βαθμό και συμπαγών, όπως και να δικαιολογείται μια τέτοια προσέγγιση, έρχεται σε σύγκρουση με την ιδέα ότι η κοινωνία θα έπρεπε να οργανώνεται ορθολογικά, αλλά μοιάζει πολλές φορές αναγκαία. Αυτή η απλοϊκή παρουσίαση ενός τεράστιου ζητήματος δεν είναι όμως το θέμα μας, παρόλο που θα αναπτυχθεί περισσότερο παρακάτω. Το θέμα μας είναι πώς το next big thing αυτήν την περίοδο, τα συστήματα βαθιάς μάθησης, είναι μηχανές των οποίων η λειτουργία μπορεί να κριθεί μόνο εκ των υστέρων<sup>11</sup>.

---

<sup>10</sup> *On the Content of Socialism*, τελευταία πρόσβαση 29 Σεπτεμβρίου 2024, <https://usa.anarchistlibraries.net/library/cornelius-castoriadis-on-the-content-of-socialism>.

<sup>11</sup> Cynthia Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead', *Nature Machine Intelligence* 1, no. 5 (13 Μαΐου 2019): 206–15, <https://doi.org/10.1038/s42256-019-0048-x>.

Γιατί διαφέρει όμως η βαθιά μάθηση, ή η τεχνητή νοημοσύνη, από τις προϋπάρχουσες μηχανές; Στην εισαγωγή του Γκάλογουεϊ υπάρχει μια πολύ ενδιαφέρουσα διατύπωση για τον υπολογιστή:

Όμως, ο υπολογιστής κατέχει μια δυσχερή θέση σε σχέση τόσο με τον ορισμό όσο και με τη δυνατότητα, αφού σε πολλές περιπτώσεις οι ίδιες λέξεις που χρησιμοποιούμε για να θέσουμε το ερώτημα του υπολογιστή είναι οι ίδιες αυτές λέξεις ορισμός και δυνατότητα. Ακούμε ιστορίες για τους υπολογιστές να είναι μηχανές «ορισμού»: όχι μόνο ο κώδικας του υπολογιστή λειτουργεί μέσα από ορισμούς καταστάσεων και αλλαγές καταστάσεων, αλλά επιπλέον οι υπολογιστές είναι αυτές οι μοναδικές μηχανές που ονομαστικοποιούν τον κόσμο, που ορίζουν και μοντελοποιούν τη συμπεριφορά του με τη χρήση μεταβλητών και συναρτήσεων. Παρόμοια, ακούμε ιστορίες για τους υπολογιστές να είναι μηχανές «δυνατοτήτων»: λειτουργούν όχι μέσα από ακαθόριστες εκτιμήσεις πρακτικής, αλλά μέσα από άκαμπτες μηχανιστικές δυνατότητες αλήθειας ή ψεύδους, δεκτικότητας ή κλειστότητας, ανοιχτού ή σβηστού. Συνεπώς, υποστηρίζω ότι αυτοί οι όροι ορισμός και δυνατότητα μπορούν να μας ζημιώσουν περισσότερο από το να μας βοηθήσουν αν ο σκοπός μας είναι να κατανοήσουμε τη μηχανή και το πώς δουλεύει<sup>12</sup>.

Ο Γκάλογουεϊ αναγνωρίζει στον υπολογιστή τη δυνατότητα να ορίζει και να μοντελοποιεί τον κόσμο, όχι ως αυτόνομο υποκείμενο, αλλά ως *ηθική*: Η χρήση του, ή μάλλον η συμπεριφορά που διαμεσολαβείται μέσω αυτού – είτε μιλάμε για μια συσκευή είτε για το λογισμικό εν γένει – συμπεριλαμβάνει την επεξεργασία του κόσμου μέσα από τη λογική του. Δημιουργείται έτσι ένας τρόπος να βλέπουμε τον κόσμο ο οποίος απασχολείται από το πώς μπορεί να τον ταΐσει σε έναν υπολογιστή, και μια μορφή αναπαράστασης του κόσμου που ακολουθεί τις δυνατότητές του και την εσωτερική του λογική, καθώς παράγεται μέσω αυτού. Αυτή είναι η λογική που σε γενικές γραμμές παρατηρούμε στο *AIMA*. Αλλά τα συστήματα βαθιάς μάθησης παραδόξως διαφέρουν. Τη διαφορά αυτή μπορούμε να τη δούμε σε δύο επίπεδα: Αυτό της τεχνητής νοημοσύνης εν γένει, και αυτό των συστημάτων βαθιάς μάθησης συγκεκριμένα. Στην εισαγωγή, θα μας απασχολήσει το πρώτο.

Είπα παραπάνω ότι η τεχνητή νοημοσύνη γεννήθηκε γερασμένη. Ας ανατρέξουμε λοιπόν στην κύηση της σύγχρονης μορφής της.

---

<sup>12</sup>Galloway, *The Interface Effect*, 19 *Η Αντι-Γλώσσα των Νέων Μέσων* (Ουτοπία τεύχος 133 (2022) 17-26)

## Η Τεχνητή νοημοσύνη μέσα από τα γραπτά του Άλαν Τούρινγκ

Ο Άλαν Τούρινγκ, ο εφευρέτης του πρώτου σύγχρονου υπολογιστή (ή του κοντινότερου του προγόνου) μεταχειρίστηκε την ερώτηση της μηχανικής σκέψης και ευφυΐας από την αρχή. Σε μεγάλο βαθμό, ο Τούρινγκ απαντά στην ιδέα ότι ο υπολογιστής που κατασκεύασε είναι ο ίδιος τεχνητή νοημοσύνη, περιγράφοντας πώς θα έμοιαζε πραγματικά κάτι τέτοιο.

Εξ' αρχής, ο Τούρινγκ επικεντρώνεται στη σημασία των κοινωνικών συμβολισμών. Στο άρθρο του *Computing Machinery and Intelligence*<sup>13</sup>, που εκδίδεται το 1950, ξεκινάει περιγράφοντας τη δυσκολία της ερώτησης «μπορούν οι μηχανές να σκεφτούν;» αναγνωρίζοντας ότι δεν έχουμε καλούς ορισμούς ούτε για τη «μηχανή» ούτε για το «σκέπτεσθαι» – παρά μόνο συνηθισμένες τους χρήσεις. Έτσι, ορίζει ένα τεστ – το «παιχνίδι της μίμησης». Ξεκινάει την εξήγηση ως εξής:

[Το παιχνίδι] παίζεται από τρία άτομα, έναν άντρα (Α), μια γυναίκα (Β), και έναν ανακριτή (Γ) που μπορεί να είναι οποιουδήποτε φύλου. Ο ανακριτής παραμένει σε ένα δωμάτιο απομονωμένος από τους άλλους δύο. Ο σκοπός του παιχνιδιού για τον ανακριτή είναι να καταλάβει ποιος από τους άλλους δύο είναι ο άντρας και ποια η γυναίκα.<sup>14</sup>

Οι ερωτήσεις μπορεί να είναι οποιεσδήποτε, όπως και οι απαντήσεις, και επικοινωνούνται γραπτώς. Η τελική μορφή του παιχνιδιού συνεπάγεται της πρώτης:

“Τι θα γινόταν αν μια μηχανή έπαιρνε τη θέση του Α σε αυτό το παιχνίδι; Θα καταλήξει ο ανακριτής λάθος με την ίδια συχνότητα που το έκανε όταν το παιχνίδι παιζόταν μεταξύ ενός άντρα και μιας γυναίκας; Αυτές οι ερωτήσεις αντικαθιστούν το αρχικό μας «μπορούν οι μηχανές να σκέφτονται;»<sup>15</sup>»

---

<sup>13</sup>A. M. Turing ‘I.—Computing Machinery and Intelligence’, *Mind* LIX, no. 236 (1 Οκτωβρίου 1950): 433–60, <https://doi.org/10.1093/mind/LIX.236.433>.

<sup>14</sup>B J Copeland, *The Essential Turing* (Oxford University Press, 2004), 441

<sup>15</sup>Ο.π., 441.



Ο Τούρινγκ έχει αναγνωρίσει ότι μια μηχανή που μπορεί να επικοινωνεί, από τη στιγμή που εκπαιδεύεται και κατασκευάζεται από ανθρώπους και εισέρχεται στο δικό τους δίκτυο επικοινωνίας, είναι προϊόν κοινωνικής κατασκευής σε βαθμό συγκρίσιμο με τους ίδιους. Στο κείμενό του *Intelligent Machinery*<sup>16</sup>, που γράφεται το 1948, λέει επίσης τα εξής:

Ο βαθμός στον οποίο αναγνωρίζουμε ότι κάτι συμπεριφέρεται με νοήμων τρόπο εξαρτάται τόσο από τη δική μας διανοητική κατάσταση και εκπαίδευση όσο και από τις ιδιότητες του αντικειμένου που εξετάζουμε. Αν μπορούμε να εξηγήσουμε και να προβλέψουμε τη συμπεριφορά του ή αν δεν φαίνεται να τη διέπει ιδιαίτερος σχεδιασμός, τότε δεν μπαίνουμε και σε ιδιαίτερο πειρασμό να φανταστούμε νοημοσύνη. Για το ίδιο αντικείμενο, οπότε, είναι δυνατόν κάποιος να το βλέπει σαν νοήμον και κάποιος άλλος όχι· ο δεύτερος θα έχει καταλάβει τους κανόνες της συμπεριφοράς του.<sup>17</sup>

Επιστρέφουμε έτσι στην κατανοησιμότητα. Ως όρο, αυτή τη φορά, της νοημοσύνης. Ο Τούρινγκ προσεγγίζει το πώς μπορούμε να δούμε ένα αντικείμενο ως εμπρόθετο υποκείμενο, τότε θα αναγκαστούμε να θεωρήσουμε ότι κάνει χρήση Λόγου, και άρα, με καντιανούς όρους, να το αντιμετωπίσουμε ως κυρίαρχο (του εαυτού του). Αυτό είναι για τον Τούρινγκ η νοημοσύνη (ή η ευφυΐα – intelligence) με δεδομένο όμως ότι, καθώς η πρόσβαση του υποκειμένου στον κοινό Λόγο είναι περιορισμένη, τη θέση του παίρνει μια κοινή γλώσσα – ένα σύστημα συμβόλων<sup>18</sup>.

Στην περίοδο αυτή της εικασίας λοιπόν που γέννησε τη σύγχρονη ΤΝ, οι ορισμοί των κατασκευαστών της ήταν κοντά σε αυτούς που θα έθεταν κοινωνικοί επιστήμονες. Παρόλα αυτά, ακόμα και πριν τη γέννησή της, η ΤΝ αποτελούσε ένα θέμα κουρασμένο. Οι αρχέγονοι μύθοι του τεχνητού ανθρώπου συμπλέκονται με τον άνθρωπο ως το ευφυές ή το νοήμον με τρόπους περίπλοκους και ενδιαφέροντες, αλλά η εξέταση αυτής της σχέσης δεν είναι κεντρική στο θέμα μας. Μια ανάλυση του *Φρανκενστάιν*, του *Μηχανικού Τούρκου*, ακόμα περισσότερο του τεχνητού άνδρα του Χομπς, για να αναφέρω μόνο λίγα από τα πρόσφατα παραδείγματα, θα βοηθούσε στην ανάλυση των αντιλήψεων για την ΤΝ στο φιλολογικό τους ιστορικό πλαίσιο. Η

---

<sup>16</sup>Ο.π., 417–32.

<sup>17</sup>Ο.π., 441.

<sup>18</sup>Ο.π., 445. Στο ίδιο κείμενο υπάρχει ένα ακόμη ενδιαφέρον απόσπασμα που περιγράφει το ενδεχόμενο να αποκτούσαν κοινή γλώσσα ομάδες ζώων, όπως και εικασίες για το πώς εγγράφεται η γλώσσα στον εγκέφαλο με κοινωνικά κριτήρια.

επιστημονική φαντασία διαποτίζει την ΤΝ με πολυσχιδείς τρόπους, σε βαθμό που πραγματικά δεν μπορούμε να την προσδιορίσουμε σαν κατηγορία ούτε με βάση τον ορισμό, ούτε με βάση τη δυνατότητά της. Καταλήξαμε στο συμπέρασμα του αποσπάσματος παραπάνω από διαφορετικό δρόμο. Πριν εξηγήσω τι είναι ένα σύστημα βαθιάς μάθησης, έχω επιλέξει έναν μύθο που θα αναλύσω για να πλαισιώσω τη συζήτηση που θα ακολουθήσει – τον μύθο της ιδεατής μαριονέτας.

### Η Μαριονέτα

Παρόλο που το διήγημα *Οι Μαριονέτες*<sup>19</sup> του Χάινριχ φον Κλάιστ είναι πάνω από διακοσίων χρόνων, το εύρος των ερμηνειών του και η ισχύς των σκοτεινών του ιδεών κάνουν την ανάγνωσή του διαχρονικά εντυπωσιακή. Για το δικό μας θέμα, απαιτούνται κάποιες αποκλείσεις από τις συνήθεις αναγνώσεις του, τις οποίες όμως δεν θεωρώ καθόλου αυθαίρετες.

Μπροστά σε ένα κουκλοθέατρο στημένο σε μια πολυσύχναστη γερμανική αγορά, ο αφηγητής μας συναντά τον κύριο Κ., πρώτο χορευτή της Όπερας της πόλης. Αφού του εκφράζει την έκπληξή που του προκαλεί το να βλέπει έναν τόσο διακεκριμένο ερμηνευτή να παρακολουθεί ένα θέαμα για τον όχλο, ξεκινάει ένας απρόσμενος διάλογος, με τον Κ. να εξηγεί ότι η ευτελής μαριονέτα δεν τον διασκεδάζει απλά σαν θέαμα, αλλά αποτελεί το πρότυπό του για την ιδανική κινησιολογία.

Του ζήτησα να μου εξηγήσει το μηχανισμό της μαριονέτας, και πώς γίνεται κάθε να κατευθύνουν κάθε μέλος, σε όλα τα σημεία του, όπως υπαγορεύει ο ρυθμός της κίνησης ή ο χορός, δίχως μυριάδες νήματα στα δάχτυλα.

Δεν θα έπρεπε να φανταστώ, αποκρίθηκε, ότι ο χειριστής τραβά ή χαλαρώνει το νήμα κάθε μέλους χωριστά, κατά τις φάσεις του χορού.

“Κάθε κίνηση», είπε, «έχει ένα κέντρο βάρους· φτάνει να το ελέγχεις, στο εσωτερικό της μαριονέτας, και τα μέλη, που δεν είναι παρά εκκρεμή, υπακούν μόνα τους μηχανικά, δίχως να τα αγγίζεις».

---

<sup>19</sup>Heinrich Von Kleist, *Οι Μαριονέτες* (Αθήνα: Άγρα, 2015).

Πρόσθεσε ακόμη πως η κίνηση αυτή είναι πολύ απλή: κάθε φορά που το κέντρο βάρους κινείται σε *ευθεία γραμμή*, τα μέλη διαγράφουν *καμπύλες* και, συχνά, αρκεί ένα τυχαίο σκίρτημα για να δώσει στο σύνολο κίνηση ρυθμική, που μοιάζει με χορό<sup>20</sup>.

Οι τεχνικές περιγραφές διαδέχονται τις καλλιτεχνικές ξανά και ξανά. Μέσα από μια εξήγηση της μηχανικής της μαριονέτας, ο χορευτής καταλήγει στο γιατί αυτή μοιάζει να χορεύει. Συνεχίζοντας για τον χειριστή:

Τον ρώτησα αν ο χειριστής που κατευθύνει τις κούκλες πρέπει, κατά τη γνώμη του, να είναι επίσης χορευτής, ή έστω να έχει την αίσθηση του ωραίου στο χορό.

“Όταν», μου απάντησε, «ένα έργο δεν παρουσιάζει δυσκολίες από μηχανικής πλευράς, δεν έπεται ότι μπορεί και να εκτελεστεί δίχως αίσθημα».

Η γραμμή που όφειλε να διαγράψει το κέντρο βάρους θα ήταν βεβαίως απλή και, κατά τη γνώμη του, τις περισσότερες φορές ευθεία. Αν πάλι ήταν καμπύλη, η καμπυλότητα θα έπρεπε να είναι πρώτου, ή, το πολύ, δευτέρου βαθμού· ωστόσο, και στην τελευταία περίπτωση, θα ήταν ελλειψοειδής, κίνηση φυσικότερη για τα άκρα του ανθρώπινου κορμιού (λόγω των αρθρώσεων), οπότε ο χειριστής δεν θα χρειαζόταν και μεγάλη τέχνη για να την πετύχει.

Από μian άλλη άποψη, η γραμμή εκείνη θα ήταν κάτι το αινιγματικό, αφού θα επρόκειτο για την *τροχιά της ψυχής του χορευτή*· και δεν θεωρούσε βέβαιο ότι ο χειριστής θα μπορούσε να την ακολουθήσει αλλιώς, εκτός κι αν μετατοπιζόταν νοερά στο κέντρο βάρους της μαριονέτας – μ’ άλλα λόγια, *αν χόρευε*.

“Ετούτη τη δουλειά», του αποκρίθηκα, «μου την είχαν παραστήσει σαν κάτι μάλλον άψυχο, όπως το γύρισμα της μανιβέλας για να παίξει το οργανέτο».

“Απεναντίας», μου είπε, «Οι κινήσεις των δαχτύλων του συνδέονται καλλιτεχνικά με την κίνηση των ανδρικών που κρατεί· όπως ο αριθμός με τον λογάριθμο του, ή η ασύμπτωτος με την υπερβολή».

Πίστευε πάντως ότι οι μαριονέτες θα έχαναν μια μέρα κι αυτό το τελευταίο ίχνος πνεύματος για το οποίο μου μιλούσε· ο χορός τους θα μεταπηδούσε στη σφαίρα μηχανικών δυνάμεων και θα μπορούσε να εκτελεστεί με τη βοήθεια μιας μανιβέλας, όπως το φανταζόμουν<sup>21</sup>.

Η σύνδεση του αριθμού με τον λογάριθμο είναι *καλλιτεχνική*, και, ενώ μέχρι στιγμής ο χειριστής αναγκάζεται να χορεύει, αφού η κίνησή του μπαίνει στη θέση της ψυχής της

---

<sup>20</sup>Ο.π., 12–13.

<sup>21</sup>Ο.π., 14–15.

μαριονέτας, η τέλεια μαριονέτα δεν θα απαιτεί αυτό το στοίχειωμα από τον άνθρωπο.

Το διπλό εκκρεμές (ένα εκκρεμές με άλλο ένα στο τέλος του) είναι ένα από τα πιο απλά συστήματα που επιδεικνύουν χαοτική κίνηση<sup>22</sup>. Ο ορισμός αυτής της ιδιότητας είναι η υπερευαισθησία στις αρχικές συνθήκες. Ευαισθησία σε βαθμό που μια απειροελάχιστη, τυπικά αμελητέα, αλλαγή στη θέση από την οποία απελευθερώνεται ένα διπλό εκκρεμές, αλλάζει τη θέση στην οποία θα βρίσκεται μετά από ένα χρονικό διάστημα σε ανυπολόγιστο βαθμό. Τα συστήματα αυτού του τύπου έχουν την εξαιρετική ιδιαιτερότητα να είναι ντετερμινιστικά αλλά απρόβλεπτα, κάτι που μοιάζει αντιφατικό. Το ζήτημα είναι ότι η ακρίβεια με την οποία θα έπρεπε να καθοριστεί η αρχική θέση ενός τέτοιου αντικειμένου, ώστε η εξίσωση που περιγράφει την κίνησή του μέσα στο χρόνο να δώσει το σωστό αποτέλεσμα, είναι ασύλληπτη. Ακόμη, αν επιχειρούσαμε μια προσέγγιση, το αποτέλεσμα που θα παίρναμε θα ήταν σε δυσανάλογο βαθμό λάθος σε σχέση με την ακρίβεια της αρχικής μέτρησης που καταφέραμε.

Ένα άλλο σύνηθες παράδειγμα χαοτικού συστήματος είναι ο καιρός<sup>23</sup>, καθώς εξαρτάται από ένα μεγάλο αριθμό παραμέτρων που αλληλεπιδρούν, και άρα οποιαδήποτε μικρή απόκλιση στον καθορισμό τους γιγαντώνεται μέσα στην μοντελοποίηση του συνολικού συστήματος. Έχουμε τρόπους να πράξουμε σχετικά έμπιστες προβλέψεις για τέτοια συστήματα για μικρά διαστήματα χρόνου, αποδεικνύεται όμως μαθηματικά ότι, παρόλο που μπορεί να είμαστε σίγουροι ότι υπάρχει ένα σύνολο κανόνων που εκφράζουν τέλεια την κίνηση του συστήματος με τη μορφή μιας συνάρτησης, οι προβλέψεις μας δεν μπορούν να είναι τέλειες ακόμα και αν τους καθορίζαμε.

Κάθε μέλος της μαριονέτας είναι ένα (τουλάχιστον) διπλό εκκρεμές. Ο έλεγχος που έχει λοιπόν ένας χειριστής δεν μπορεί να είναι πλήρης. Τα «μυριάδες νήματα» που φαντάζεται ο αφηγητής μας δεν είναι αναγκαία, καθώς το μόνο που έχει σημασία είναι να ελέγχεται το κέντρο βάρους κάθε κίνησης. Ο χορός που προκύπτει από μια

---

<sup>22</sup>R. Levien και Sze Tan, 'Double Pendulum: An Experiment in Chaos', *American Journal of Physics - AMER J PHYS* 61 (1 Νοεμβρίου 1993): 1038–44, <https://doi.org/10.1119/1.17335>.

<sup>23</sup>A. A. Tsonis και J. B. Elsner, 'Chaos, Strange Attractors, and Weather', *Bulletin of the American Meteorological Society* 70, no. 1 (1 Ιανουαρίου 1989): 14–23, [https://doi.org/10.1175/1520-0477\(1989\)070<0014:CSAAW>2.0.CO;2](https://doi.org/10.1175/1520-0477(1989)070<0014:CSAAW>2.0.CO;2).

πολύ απλή μετατόπιση ενός αρμού με τη βοήθεια της αδράνειας είναι έτσι απρόβλεπτος και περίπλοκος, ενώ διέπεται από το ρυθμό που επιβάλλει η αρχική μετατόπιση, την *ψυχή του χορευτή*, το στοιχείο που κάνει την κίνηση αναγνωρίσιμη ως χορό. Η τελευταία παράγραφος του παραπάνω αποσπάσματος εκφράζει όμως μια περαιτέρω πρόβλεψη, η οποία είναι και επιθυμία του χορευτή – την κατάργηση της αναγκαιότητας για χειρισμό εξ ολοκλήρου· την πλήρη αυτοματοποίηση του χορού της μαριονέτας.

Στη συνέχεια, περιγράφει τις προδιαγραφές της ιδεατής μαριονέτας από τις οποίες, με τη βοήθεια ενός επιδέξιου τεχνίτη, θα μπορούσε να προκύψει ένα ανδρείκελο που θα ξεπερνούσε σε χάρη κάθε ανθρώπινο χορευτή. Το βασικό ζητούμενο είναι να μπορεί να συγκεντρωθεί ο έλεγχος εξ ολοκλήρου στην *ουσία* της κίνησης: Κινώντας έναν αρμό ενσταλάσσεται έτσι *ψυχή* μόνο εκεί που έχει σημασία – η υπόλοιπη κίνηση που προκύπτει είναι άψυχη, και άρα φυσική.

Του είπα ότι όσο επιδέξια και αν ανέπτυξε το παράδοξο επιχείρημά του, δεν θα κατόρθωνε να με πείσει πως το μηχανικό νευρόσπαστο έχει μεγαλύτερη χάρη από το ανθρώπινο σώμα.

“Ο άνθρωπος», αντέταξε, «δεν είναι καν εις θέση να πλησιάσει το νευρόσπαστο. Σε αυτόν το χώρο, μόνο ένας Θεός θα μπορεί να αναμετρηθεί με την ύλη, γιατί εδώ είναι το σημείο όπου σμίγουν τα δύο άκρα του στρογγυλού κόσμου.<sup>24</sup>»

Χαρακτηριστικό της μαριονέτας είναι ότι δεν ελέγχεται κάθε της αρμός. Αν ήταν έτσι, η κίνηση της θα είχε μια επιτήδευση που θα έκανε αδύνατο το εφέ του τυχαίου χορού. Το εφέ αυτό εξηγείται εν πολλοίς μέσω του φαινομένου του διπλού εκκρεμούς. Θα ήθελε ο χορευτής μας να έχει πλήρη έλεγχο ενός περίπλοκου και ευλύγιστου ανδρικού, όταν αναφέρεται στις απαιτητικές του προδιαγραφές, ή συνεχίζει να έχει σημασία για αυτόν το πρώτο κομμάτι της περιγραφής του; Ασφαλώς και ισχύει το δεύτερο. Αν απαιτούσε ανάλογο έλεγχο και τέχνη από μεριάς του χειριστή το να κινεί την τέλεια μαριονέτα με το να χορεύει ο ίδιος, τότε η προστιθέμενη χάρη θα χανόταν. Προς το τέλος του κειμένου, τα «δύο άκρα του στρογγυλού κόσμου» επεξηγούνται:

---

<sup>24</sup>Heinrich Von Kleist, *Οι Μαριονέτες*, 20.

“Βλέπουμε ότι, στον οργανικό κόσμο, όσο σκοτίζεται και αδυνατίζει ο λογισμός, τόσο πιο αστραποβόλα και ηγεμονική προβάλλει η χάρη. Και όπως η τομή δύο ευθειών στη μια πλευρά κάποιου σημείου διασχίζει το άπειρο και ξεφυτρώνει αίφνης από την άλλη πλευρά, ή όπως το είδωλο κοίλου κατόπτρου διανύει το κενό και εμφανίζεται απρόοπτα εμπρός μας, έτσι αναπάντεχα παρουσιάζεται η χάρη όταν η γνώση φεύγει στο αχανές, και διακρίνεται πιο καθαρά στο σώμα εκείνο που δεν έχει διόλου γνώση, ή έχει γνώση απέραντη, δηλαδή στο ανδρείκελο ή στο Θεό.<sup>25</sup>»

Στις λίγες αυτές σελίδες παρουσιάζεται ένας διάλογος μεταξύ ανθρώπου, φύσης, και μηχανής, κατά τη διάρκεια του οποίου όλες οι πλευρές προβάλλονται η μια στην άλλη, με σκοπό να ανακύψει, ως σύνολο, μια ειρωνική θεολογία της μη-συνείδησης, του μη-ελέγχου, της μη-κατανοησιμότητας.

Το πιο αινιγματικό κομμάτι του κειμένου, το οποίο παρεμβάλλεται μεταξύ των αποσπασμάτων παραπάνω – του αρχικού σχήματος και της κατάληξης του – είναι η ιστορία της αρκούδας. Ο χορευτής περιγράφει την εμπειρία του ως καλεσμένος στην έπαυλη κάποιου ευγενούς, ο οποίος καυχιέται για την ικανότητά του στην ξιφασκία. Παρόλα αυτά, ο χορευτής καταφέρνει να τον νικήσει με σχετική ευκολία. Σε μια αναπάντεχη τροπή των διακοπών του, ο ευγενής τον φέρνει μπροστά σε μια αρκούδα, που μεγάλωνε ο πατέρας του στους κήπους, λέγοντας ότι όσο επιδέξιος κι αν είναι ο Κ., η ικανότητά του δεν μπορεί να πλησιάσει ποτέ αυτή του άγριου ζώου με το οποίο τον προκαλεί να ξιφομαχήσει. Η σκηνή που ακολουθεί περιγράφεται πράγματι ως ξεκάθαρη ήττα του χορευτή, καθώς η αρκούδα αγνοεί πλήρως οποιοδήποτε τέχνασμα και προσποίηση σκαρφιστεί ο αντίπαλός της, αποκρούοντας μόνο την τελική επίθεση με μια μετρημένη κίνηση του ποδιού της.

Με κοίταζε στα μάτια, λες και μέσα τους μπορούσε να διαβάσει την ψυχή μου, και με το πόδι σηκωμένο, έτοιμο ν' αποκρούσει, περίμενε ατάραχη, δίχως να νοιάζεται πως ίσως σκόπευα να τη χτυπήσω σοβαρά<sup>26</sup>.

Στην ιστορία αυτή της ανίκητης, αλλά παρόλα αυτά αλυσοδεμένης αρκούδας, εκφράζεται παραδειγματικά η σύγκλιση μηχανής και φύσης.«Το πνεύμα δεν μπορεί

---

<sup>25</sup>Ο.π., 26.

<sup>26</sup>Ο.π., 25.

να σφάλει εκεί που δεν υπάρχει», λέει σε ένα προηγούμενο σημείο ο αφηγητής – και από μια άποψη έχει δίκιο<sup>27</sup>.

Στη *Διαλεκτική του Διαφωτισμού*<sup>28</sup> συναντάμε χαρακτηριστικά το επιχείρημα ότι η βαρβαρότητα που παράγει η αστική τάξη πραγμάτων, η αντιφατική προς τις αρχές της, έχει σε μεγάλο βαθμό να κάνει με το μίσος που δημιουργείται στο νεωτερικό άνθρωπο προς το φυσικό, άλογο στοιχείο που ενυπάρχει μέσα του.

Η μηχανή περιέχει κάτι από τον εαυτό μας με τον ίδιο ακριβώς τρόπο. Ο κύριος Κ. θέλει να δημιουργήσει ένα χορευτή χωρίς ίχνος λογισμού, ελπίζοντας έτσι να προσεγγίσει τη θεία χάρη όπως εκφράζεται στη φύση. Η αρκούδα είναι αήττητη γιατί δεν φοβάται να κάνει λάθος, γιατί είναι εναρμονισμένη με το περιβάλλον της σε βαθμό που βλέπει τα ίχνη του μέσα στον πονηρό αντίπαλό της. Και αυτά τα ίχνη, η κίνηση της αθάνατης ψυχής, είναι η θεϊκή υπόσταση του ανθρώπου. Ο Διαφωτισμός επιτυγχάνει την αντιστροφή που αναζητήσε: Η άθρη θεία χάρη του διαφαίνεται όχι στο άλογο, αλλά στο πλήρως λογικό. Η τέλεια μηχανή στην εποχή μας είναι αυτή που θα μπορεί να σκέφτεται ακριβώς με τον ίδιο τρόπο που η μαριονέτα του Κ. χορεύει – η μια ως Θεός, η άλλη ως το αντίθετό του, τοποθετημένα ακριβώς στο «σημείο που σμίγουν τα δύο άκρα του στρογγυλού κόσμου».

Ο Σαμ Άλτμαν, διευθύνων σύμβουλος της εταιρίας OpenAI που κατασκευάζει το ChatGPT, είπε χαρακτηριστικά στους επενδυτές του ότι δεν έχει σχέδιο για το πώς θα παράξει κέρδος, αλλά σε λίγο θα μπορέσει να απαντήσει αυτήν την ερώτηση η νοημοσύνη που κατασκευάζει<sup>29</sup>. Ο Ντέμης Χασάμπης της DeepMind πρότεινε αντίστοιχα ότι η «επίλυση της νοημοσύνης» είναι το πρωταρχικό πρόβλημα προς λύση, καθώς θα μπορέσουμε να τη χρησιμοποιήσουμε για να λύσουμε όλα τα υπόλοιπα<sup>30</sup>. Μια καθολική ορθολογική μηχανή διαφαίνεται έτσι ως ένα συλλογικό όνειρο, και μέσω της TN υπάρχει για πρώτη φορά η δυνατότητα να γίνει πραγματικό.

---

<sup>27</sup>Ο.π., 19.

<sup>28</sup>Theodor W. Adorno και Max Horkheimer, *Διαλεκτική Του Διαφωτισμού* (Αθήνα: Νήσος, 1996).

<sup>29</sup>Startup Bell, 'Sam Altman Told Investors Bluntly He Had No Plans On How To Generate Revenue', Startup Bell, 14 Μαΐου 2024, <https://www.startupbell.net/post/sam-altman-told-investors-bluntly-he-had-no-plans-on-how-to-generate-revenue>.

<sup>30</sup>'The Superhero of Artificial Intelligence: Can This Genius Keep It in Check?', *The Observer*, 16 Φεβρουαρίου 2016, sec. Technology, <https://www.theguardian.com/technology/2016/feb/16/demis-hassabis-artificial-intelligence-deepmind-alphago>.

“Για το ίδιο αντικείμενο, οπότε, είναι δυνατόν κάποιος να το βλέπει σαν νοήμον και κάποιος άλλος όχι· ο δεύτερος θα έχει καταλάβει τους κανόνες της συμπεριφοράς του<sup>31</sup>», μας λέει ο Τούρινγκ. Και όπως είδαμε ο σκοπός του χορευτή είναι να απομονώσει το θεϊκό στοιχείο στον άνθρωπο και να το ενσταλάξει στη μηχανή, έτσι ώστε η υπόλοιπη της κίνηση να εξηγείται πλήρως από αυτό, με δεδομένο ότι είναι απρόβλεπτη στα νόθα μάτια του ανθρώπου. Έτσι λοιπόν αν πειστούμε ότι έχουμε επιτύχει στην ενστάλαξη του Λόγου μας, στην καθαρή του μορφή, σε ένα τεχνούργημα που δεν θα καταλαβαίνουμε πλήρως θα έχουμε από μια άποψη κλείσει τον κύκλο της γνώσης και του Διαφωτισμού.

«Μήπως», είπα λιγάκι αφηρημένος, «θα πρέπει να ξαναγετούμε τον καρπό του Δέντρου της Γνώσεως, για να επιστρέψουμε στην ηλικία της αθωότητας;»

«Να σας πω», απάντησε, «αυτό είναι το τελευταίο κεφάλαιο της ιστορίας του κόσμου<sup>32</sup>»

Η μηχανή για την οποία συζητάμε είναι μια μηχανική αυθεντία. Ένα σύστημα που δίνει συμβουλές με τρόπο που δεν μπορεί να εξηγήσει, παρά μόνο ίσως να πλαισιώσει με επιχειρήματα, ενώ η απειλή που επικαλείται προέρχεται μόνο από το φόβο της ευθύνης για το λάθος. Η ορθολογικότητα της βαθιάς μάθησης, όπως θα δούμε παρακάτω, βασίζεται είτε στην ικανότητά της να έχει εποπτεία πάνω σε έναν όγκο ανθρώπινων δεδομένων που δεν θα είχε τη δυνατότητα ένας άνθρωπος να προσπελάσει, είτε στην ικανότητά της να μάθει να κινείται σε ένα περιβάλλον χωρίς να μιμηθεί ανθρώπινα παραδείγματα. Στην πρώτη περίπτωση, η ορθολογική μαριονέτα θα κινούταν μέσω του κοινού στους ανθρώπους Λόγου, στη δεύτερη θα κινούταν σαν την αρκούδα – με βάση την απόλυτη προσήλωση στη λογική του περιβάλλοντος.

Η ιδέα της ορθολογικής οργάνωσης της κοινωνίας μέσω ενός κοινού σε όλους Λόγο βραχυκυκλώνει με την ιδέα μιας υπεράνθρωπης ορθολογικής μηχανής. Καθώς όμως μια τέτοια ροή σκέψης θα μας οδηγούσε μόνο στη μηχανή ως *φιλόσοφο-βασιλέα*, θεωρώ ότι πρέπει να κρίνουμε πόσο ορθολογικά είναι πράγματι αυτά τα νέα μηχανήματα, αλλά και πώς δεν καταλήξαμε σε μια ορθολογική οργάνωση της κοινωνίας μέσα σε τόσους αιώνες Διαφωτισμού. Για το πρώτο, απαιτείται μια παρουσίαση της ιδιαιτερότητας των συστημάτων βαθιάς μάθησης και των

---

<sup>31</sup>Copeland, *The Essential Turing*, 441.

<sup>32</sup>Heinrich Von Kleist, *Οι Μαριονέτες*, 26–27.



νευρωνικών δικτύων, για το δεύτερο, μια παρουσίαση της ιστορίας της ιδέας της αυθεντίας της γνώσης.

## Τι είναι τα συστήματα βαθιάς μάθησης

Όπως είδαμε, από την πρώτη εμφάνιση του ηλεκτρονικού υπολογιστή, το μυαλό πολλών ανθρώπων δημιούργησε με τις νέες μηχανές μια σχέση ταύτισης, ή μάλλον ενίσχυσε την υπάρχουσα. Από πολύ νωρίς λοιπόν δημιουργείται η ιδέα ενός *νευρωνικού δικτύου*. Όπως και οι βιολογικοί νευρώνες, οι επί μέρους κόμβοι ενός τέτοιου δικτύου είναι συνδεδεμένοι μεταξύ τους, και ο κάθε ένας από αυτούς λαμβάνει σήμα από κάποιους προηγούμενους και εκπέμπει σε κάποιους επόμενους. Αντίθετα (μάλλον) από τα βιολογικά τους ανάλογα, τα υπολογιστικά νευρωνικά δίκτυα είναι συνήθως δομημένα σε *επίπεδα* [layers], οργανώνοντας έτσι τη διάδοση του σήματος σε μια κατεύθυνση. Δεν δημιουργούν νέες συνάψεις.

Τα ενδιάμεσα *κρυφά επίπεδα* μπορούν να είναι οποιοδήποτε αριθμού, και το κάθε *επίπεδο* οποιοδήποτε μεγέθους. Αυτό που πάντα ισχύει για ένα νευρωνικό δίκτυο είναι ότι το πρώτο *επίπεδο* λαμβάνει κωδικοποιημένη την εισερχόμενη πληροφορία (input) και το τελευταίο παράγει μια κατανοήσιμη απάντηση (output). Η διαδικασία της εκπαίδευσης έχει σκοπό να μπορέσουν τα δύο αυτά να αντιστοιχιστούν επαρκώς.

Η εκπαίδευση ακολουθεί την αντίθετη φορά από τη λειτουργία. Ο κάθε ψηφιακός νευρώνας είναι ουσιαστικά μια συνάρτηση μέσω της οποίας παράγεται η ισχύς του σήματος που θα εκπέμψει με βάση την ισχύ του σήματος που έλαβε. Η εκπαίδευση ενός τέτοιου συστήματος, που είναι ξεχωριστή διαδικασία από τη λειτουργία του, γίνεται μέσω της μεταβολής των λεγόμενων *βαρών* [weights]. Η κάθε σύναψη μεταξύ νευρώνων έχει ένα *βάρος*, έναν αριθμό δηλαδή που καθορίζει πόσο ο νευρώνας-δέκτης «ακούει» τον νευρώνα-πομπό<sup>33</sup>. Αυτό που ουσιαστικά κάνει το μοντέλο κατά τη διάρκεια της εκπαίδευσής του είναι να «εκλογικεύει» το feedback που λαμβάνει για κάθε επιλογή του. Με αρκετές δοκιμές και με ποικιλία input ανάλογη του σκοπού

---

<sup>33</sup>Stuart Russel and Peter Norvig, *Τεχνητή Νοημοσύνη: Μια Σύγχρονη Προσέγγιση*, 831–840.

του, κάποια στιγμή καταλήγει να μπορεί να γενικεύσει τις εκλογικές αυτές σε νέα δεδομένα με μια ικανοποιητική ακρίβεια<sup>34</sup>.

Όλοι οι νευρώνες ενεργοποιούνται έστω λίγο κάθε φορά που «τρέχει» ένα νευρωνικό δίκτυο. Έτσι, αν, για παράδειγμα, πούμε ότι έχουμε ως output 5 κατηγορίες, στις οποίες καλείται να κατατάξει το μοντέλο το input του, κάθε μια από τις 5 θα καταλήξει με μια ποσότητα σήματος, η οποία μεταφράζεται άμεσα στην πιθανότητα, κατά τη γνώμη του νευρωνικού δικτύου, να είναι η σωστή.

Η αρχιτεκτονική του δικτύου είναι γνωστή, όπως και οι αριθμοί που αντιστοιχούν στα βάρη του. Αν θέλουμε, μπορούμε να ζωγραφίσουμε σε ένα πίνακα το μοντέλο εξ ολοκλήρου, όπως και να εξετάσουμε την ενεργοποίηση που οδήγησε σε κάθε του συναγωγή. Ο ρόλος όμως κάθε νευρώνα στο συνολικό δίκτυο δεν μπορεί να κατανοηθεί αφαιρετικά – δεν μπορούμε να εκφράσουμε με σιγουριά τι έχει «αναλάβει να κάνει». Τα σύγχρονα μοντέλα τα οποία θα εξετάσουμε παρακάτω αποτελούνται από δισεκατομμύρια νευρώνες, κάθε ένας από τους οποίους είναι ένας παράγοντας στη συναγωγή τους. Ακόμα και αν μπορούσαμε να διαχωρίσουμε τη δική μας ροή σκέψης όταν αποφασίζουμε, για παράδειγμα, αν βλέπουμε ένα σκύλο, σε ένα δισεκατομμύριο επί μέρους παράγοντες, ή σε 10 επί μέρους «επίπεδα» συναγωγών, είναι σχεδόν αδύνατο να καταλάβουμε αν το νευρωνικό δίκτυο έχει κατανείμει την πληροφορία με τον ίδιο τρόπο. Στις περιπτώσεις μάλιστα που κάποια σύγκριση είναι εφικτή, συνήθως αυτή η υπόθεση διαψεύδεται<sup>35</sup>.

Το εμφανές ζήτημα που δημιουργεί αυτή η αδυναμία κατανόησης είναι ότι, σε περίπτωση που ένα νευρωνικό δίκτυο έχει πρακτική εφαρμογή, δεν ξέρουμε αναλυτικά τη μέθοδο που χρησιμοποιήθηκε για τη λύση του προβλήματος, και άρα δεν μπορούμε να την αναπαράγουμε χωρίς το συγκεκριμένο μοντέλο. Ένα λιγότερο προφανές ζήτημα όμως είναι ότι δεν υπάρχει τρόπος να επιλεγεί με σιγουριά μια αρχιτεκτονική νευρωνικού δικτύου για ένα συγκεκριμένο πρόβλημα. Η επιλογή αυτή, όπως και η εξακρίβωση των κατάλληλων παραμέτρων εκπαίδευσης, είναι σε μεγάλο βαθμό βασισμένη στο ένστικτο των μηχανικών και σε αλληπάλληλες δοκιμές<sup>36</sup>. Εδώ

---

<sup>34</sup>Ο.π., 847–50.

<sup>35</sup>Rudin, ‘Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead’; Gabrielle Ras et al., ‘Explainable Deep Learning: A Field Guide for the Uninitiated’, *Journal of Artificial Intelligence Research* 73 (25 Ιανουαρίου 2022): 329–96, <https://doi.org/10.1613/jair.1.13200>.

<sup>36</sup>Stuart Russel and Peter Norvig, *Τεχνητή Νοημοσύνη: Μια Σύγχρονη Προσέγγιση*, 852–53.

ακριβώς εστιάζεται η διαφορά των συστημάτων αυτών από την περιγραφή του Γκάλογουεϊ για τον υπολογιστή.

Όχι μόνο ο κώδικας του υπολογιστή λειτουργεί μέσα από ορισμούς καταστάσεων και αλλαγές καταστάσεων, αλλά επιπλέον οι υπολογιστές είναι αυτές οι μοναδικές μηχανές που ονομαστικοποιούν τον κόσμο, που ορίζουν και μοντελοποιούν τη συμπεριφορά του με τη χρήση μεταβλητών και συναρτήσεων. Παρόμοια, ακούμε ιστορίες για τους υπολογιστές να είναι μηχανές «δυνατοτήτων»: λειτουργούν όχι μέσα από ακαθόριστες εκτιμήσεις πρακτικής, αλλά μέσα από άκαμπτες μηχανιστικές δυνατότητες αλήθειας ή ψεύδους, δεκτικότητας ή κλειστότητας, ανοιχτού ή σβηστού<sup>37</sup>.

Η κατασκευή των μηχανών αυτών δεν ξεκινά από μια μηχανιστική μετάφραση του προβλήματος προς λύση, αλλά από την αναγνώριση της αδυναμίας να παραχθεί κάτι τέτοιο. Το νευρωνικό δίκτυο δημιουργεί έτσι μια μηχανή που την αντιμετωπίζουμε παρόμοια με έναν υπολογιστή αλλά *δεν λειτουργεί μόνο μέσα από άκαμπτες μηχανιστικές δυνατότητες*. Βασίζεται σε αυτές, καθώς είναι και παραμένει λογισμικό, αλλά, καθώς «οργανώνει τον εαυτό του<sup>38</sup>», προσομοιώνει μια ρευστή διαδικασία συναγωγής με τρόπο που δεν ελέγχουμε εξ αρχής, αλλά μόνο κρίνουμε εκ των υστέρων – όχι μόνο ως χρήστες αλλά ως άνθρωποι. Στις περιπτώσεις που λειτουργεί, κάνει συστήματα και πεδία υπερβολικής περιπλοκότητας διαχειρίσιμα<sup>39</sup>. Η χρήση τους όμως ακόμα και σε πεδία όπου κανένας άλλος τύπος συστήματος δεν έχει φέρει ουσιαστικά αποτελέσματα μπορεί να ισοδυναμεί με μια απόληση του ελέγχου.

Όταν αναφερόμαστε στον υπολογιστή ως «ουσιωτική μηχανή», αυτό που ουσιαστικά εννοούμε είναι ότι οι υπολογιστές προσομοιώνουν οντολογίες, καθορίζουν ορίζοντες δυνατοτήτων», συνεχίζει ο Γκάλογουεϊ στο ίδιο απόσπασμα. Σκοπός μου είναι να αναδείξω πώς τα συστήματα βαθιάς μάθησης αντιστοιχούν στην εικόνα του ιδανικού υποκειμένου για να κινείται και να πράττει μέσα σε τεχνητά καθορισμένους ορίζοντες δυνατοτήτων, και πώς μια παρενέργεια της είναι να τους εδραιώνει ως απέραντους.

---

<sup>37</sup>Galloway, *The Interface Effect*, 19. *Η Αντι-Γλώσσα των Νέων Μέσων* (Ουτοπία τεύχος 133 (2022) 17-26)

<sup>38</sup>Περιγραφή της μηχανικής μάθησης από τον Τούρινγκ

<sup>39</sup>Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'.

## Αυθεντία και Γνώση

Κατά κοινή ομολογία, είναι σχεδόν αδύνατον να ορίσουμε την αυθεντία γενικά και συγχρόνως να την κρίνουμε αξιολογικά. Οι περισσότεροι ορισμοί που βρίσκουμε είναι για αυτό το λόγο αρνητικοί ή ειδικοί. Ο Μαξ Χορκχάιμερ χρησιμοποιεί τον πιο απλό, και πιο γενικό: Αυθεντία, λέει, είναι η *αποδεκτή εξάρτηση*<sup>40</sup>. Υπάρχουν όμως εμφανώς τύποι αποδοχής, και τύποι εξάρτησης, που είναι αναγκαίοι ή θεμιτοί, και άλλοι που δεν είναι. Η Χάνα Άρεντ ξεκινάει την προσέγγισή της ως εξής:

Επειδή η αυθεντία απαιτεί πάντα υπακοή, συνήθως θεωρείται λαθεμένα πως είναι κάποια μορφή εξουσίας η βίας. Ωστόσο η αυθεντία αποκλείει τη χρήση εξωτερικών μέσων εξαναγκασμού (coercion): όπου χρησιμοποιείται δύναμη, η αυθεντία έχει αποτύχει. Από την άλλη μεριά η αυθεντία είναι ασυμβίβαστη με την πειθώ, η οποία προϋποθέτει ισότητα και λειτουργεί μέσα από μια διαδικασία επιχειρηματολογίας. Όπου χρησιμοποιούνται επιχειρήματα η αυθεντία πέφτει σε αχρηστία. [...] Αν πρέπει να δοθεί κάποιος ορισμός στην αυθεντία, αυτό θα πρέπει να γίνει σε αντιδιαστολή και προς τον καταναγκασμό με τη βία και προς την πειθώ μέσω επιχειρημάτων<sup>41</sup>.

Η αυθεντία είναι λοιπόν ενός τύπου δικαιολόγηση της υπακοής, η οποία δεν βασίζεται σε εξαναγκασμό και βία αλλά ούτε χρειάζεται επιχειρήματα. Είναι μια συμβουλή με ιδιαίτερο βάρος, η οποία είναι σωστό να ακολουθηθεί παρά το ότι δεν συνοδεύεται ούτε από απειλή ούτε από επιβράβευση. Η Άρεντ παρουσιάζει μια γενεαλογία της έννοιας ώστε να καταλήξει στη νεωτερική αυθεντία, η οποία, όπως μας λέει, βρίσκεται σε μια ιδιότυπη κρίση. Η αναγνώριση αυτής της κρίσης είναι κοινό σημείο σε όλα τα κείμενα που θα πραγματευτώ, καθώς θεωρώ ότι αυτού του είδους η ανάλυση θα μας επιτρέψει να κατανοήσουμε τις δυνατότητες της TN ως «σκεπτόμενης μηχανής» μέσα από την κοινωνική σημασία της γνώσης.

Ο Πλάτωνας είναι ο πρώτος που συνδέει την αυθεντία με τη γνώση, μεταχειριζόμενος, στα λόγια της Άρεντ<sup>42</sup>, τη *διαμάχη* μεταξύ φιλοσόφου και κοινωνίας. Παρουσιάζει τον φιλόσοφο ως αυτόν που τολμά να προσπαθεί να γνωρίσει τον κόσμο όπως πραγματικά είναι, προσεγγίζοντας έτσι τον θεϊκό κόσμο των ιδεών. Καθώς όμως οι συμπολίτες του δεν μπορούν, ή δεν θέλουν να καταλάβουν την

---

<sup>40</sup>Max Horkheimer, Erich Fromm, και Herbert Marcuse, *Αυθεντία Και Οικογένεια* (Αθήνα: Νήσος, 1995), 30.

<sup>41</sup>Hannah Arendt, 'Τι Είναι Αυθεντία', στο *Μεταξύ Παρελθόντος Και Μέλλοντος* (Αθήνα: Λεβιάθαν, 1996), 160.

<sup>42</sup>Ο.π., 181–90.

αλήθεια στη μορφή που την έχει προσεγγίσει, αναγκάζεται να δημιουργεί μύθους. Με αυτόν τον τρόπο μετασχηματίζει τη γνώση ώστε να γίνει εφαρμόσιμη στην καθημερινή και πολιτική ζωή των ανθρώπων, με έναν τρόπο που μπορεί ο ίδιος να κρίνει μέσω της προνομιακής του πρόσβασης στην αλήθεια του κόσμου. Οι διηγήσεις αυτές συγκροτούν την πολιτική επιστήμη και τη θεολογία, κατηγορίες διαφορετικές από τη θρησκεία ως λόγο του θεού, και από την αντίστοιχη υπερβατική έννοια της ηθικής. Έτσι καταλήγει ο Πλάτωνας στον φιλόσοφο-βασιλέα της *Πολιτείας*. Πρέπει όμως, όπως επισημαίνει η Άρεντ, να βλέπουμε την περιγραφή της ουτοπικής *Πολιτείας* ως παράδειγμα ενός πολιτικού μύθου, και άρα ως μια νουθεσία με «δευτέρου βαθμού» σκοπιμότητα. Ο Πλάτων, δηλαδή, ως φιλόσοφος, δεν παρουσιάζει την ιδανική κοινωνία με βάση τη δική του γνώση και αντίληψη, αλλά ελπίζει να νουθετήσει τους συγχρόνους του προς μια κατεύθυνση που μάλλον αγνοούμε χωρίς να ξέρουμε τα ειδικά ιστορικά συγκείμενα.

Η αυθεντία λοιπόν του Πλάτωνα, που προέρχεται από τη γνώση, δίνει στον φορέα της το δικαίωμα, ή το καθήκον, να χειραγωγεί τους μη γνώστες, ή τουλάχιστον να τους καθοδηγεί εκθέτοντας όχι την ίδια την γνώση, αλλά κάποια όψη της ανάλογα με τον εκάστοτε ακροατή. Στη Ρώμη, συνεχίζει η Άρεντ, η αυθεντία μπαίνει για πρώτη φορά πραγματικά στην πολιτική σκηνή, ορίζεται, και παίρνει μια μορφή πολύ διαφορετική από αυτή που περιγράφει ο Πλάτωνας.

Η γνώση και η αλήθεια βγαίνουν σε μεγάλο βαθμό από την εξίσωση, και τη θέση του όρου της αυθεντίας παίρνει η ίδρυση της Ρώμης. Η διήγηση της ίδρυσης είναι ένας μύθος ο οποίος νοείται ως ιστορικό γεγονός, και αυτό κάνει την ιδεολογία που απορρέει από αυτόν εγκόσμια και, μέσω της σύνδεσης με την πόλη, αμιγώς πολιτική. Η αυθεντία ανήκει στους πρεσβύτερους όχι λόγω της γνώσης και της εμπειρίας τους καθαυτές, αλλά ως ουσία που κληροδοτείται από τους ιδρυτές μέσα από το χρόνο. Έτσι η συμβουλή τους έχει ιερό βάρος, αλλά μέσω μιας εγκόσμιας παράδοσης<sup>43</sup>.

Ο διήγηση που αντιστοιχεί στην ίδρυση για τη μοναρχική τάξη του μεσαίωνα είναι ο θάνατος και η ανάσταση του Χριστού – ένας θεολογικός μύθος που τοποθετείται στην εγκόσμια ιστορία. Ο διαχωρισμός όμως της εκκλησίας, του θεσμικού φορέα της αυθεντίας, από το κράτος, αποξενώνει την αυθεντία από την πολιτική της υπόσταση. Έτσι, εδραιώνεται μια αυθεντία που έχει κοινά στοιχεία με αυτή του Πλάτωνα, όντας

---

<sup>43</sup>Ο.π., 194–96.

βασισμένη στη γενική υπερκόσμια αλήθεια, η οποία όμως είναι emphaticά εκτός της κοινωνίας που καθοδηγεί. Η περιπλοκότητα του ανθρώπινου δικτύου συσχετίσεων που συνθέτει την κοινωνία (“των οποίων η ίδια η ουσία φαίνεται να είναι η σχετικότητα<sup>44</sup>»), οδηγεί την εκκλησία σε ένα αδιέξοδο που ο Πλάτωνας έχει περιγράψει: Η ίδια η υπερβατική αλήθεια δεν αρκεί για να θεμελιωθεί μια πολιτική τάξη. Η λύση του Πλάτωνα, την οποία περιγράψαμε ως νοθεσία, προϋποθέτει τη συναίσθηση από τη μεριά του φορέα της αυθεντίας ότι οι διηγήσεις που σκαρφίζεται είναι σκόπιμες διασκευές της αλήθειας. Επίσης, ο Πλάτωνας αναγνωρίζει την ανάγκη να ενταχθεί ο φορέας της αυθεντίας στην πολιτική, έστω ως η κορυφή της.

Η εκκλησία από την άλλη επινοεί διηγήσεις που, ειδικά μετά την εγγραφή τους στο θεολογικό κανόνα, καταφέρνουν να εκφοβίζουν αδιακρίτως, αλλά και να αντικαθιστούν τις Γραφές από τις οποίες αντλούν ισχύ. Το βασικό παράδειγμα της Άρεντ είναι ο θεολογικός μύθος της κόλασης, ειδικότερα οι διηγήσεις που περιγράφουν τα βασανιστήρια και τους δαίμονες λεπτομερώς, που συνθέτουν μια συνολική υπερβατική απειλή. Καθώς οι ιερείς, οι φορείς της αυθεντίας, δεν έχουν θεσμική πολιτική θέση, η υπόστασή τους ως πολίτες, και άρα εν δυνάμει κολασμένοι, συγκρούεται με το ρόλο τους σαν πομποί του λόγου του Θεού. Μέσα από αυτήν την αντίφαση εμφανίζεται μια αυθεντία νοθευμένη με βία, καθώς όταν η αυθεντία είναι υπεράνω της πολιτικής αλλά και της κατανόησης, οι φορείς της επιτρέπεται να αυθαιρετούν για τη διατήρηση της κυριαρχίας τους. Καταλήγουμε έτσι σε ένα *αμάλαμα* – το πρότυπο της Ρωμαϊκής Αυτοκρατορίας νοθεύεται με τα υπερβατικά μέτρα και κανόνες της ελληνικής φιλοσοφίας και με το δόγμα της Καθολικής Εκκλησίας. Η αυθεντία αυτή, η οποία παρουσιάζεται ως δευτερευόντως πολιτική, χάνει έτσι για την Άρεντ τη δυνατότητα να παράξει συνεκτική διαπαιδαγώγηση, και άρα ουδέτερη υπακοή. Ο επί μέρους «ειδήμονας» – ο ιερέας – γίνεται φορέας μιας αλήθειας που έχει να κάνει με την καθημερινή και την πολιτική ζωή και συγχρόνως με κάτι υπερβατικό και απρόσιτο, χωρίς να θεωρείται θεμιτό να τα διαχωρίσει:

Τώρα πια γενικά και υπερβατικά κριτήρια, στα οποία θα μπορούσαν να υπαχθούν το μερικό και το εμμενές χρειάζονταν για οποιαδήποτε πολιτική πράξη, ηθικοί κανόνες για κάθε διανθρώπινη συμπεριφορά και ορθολογικά μέτρα για την καθοδήγηση κάθε ατομικής κρίσης<sup>45</sup>.

---

<sup>44</sup>Ο.π., 207.

<sup>45</sup>Ο.π., 196.

Η κριτική της αστικής τάξης πραγμάτων σε σύγκριση με τα Διαφωτιστικά ιδεώδη στο έργο της Σχολής της Φρανκφούρτης πιάνει το νήμα της γενεαλογίας της Άρεντ και επεκτείνει την περιγραφή της κρίσης της αυθεντίας την οποία συζητάμε. Συνοψίζοντας, στο κείμενο της Άρεντ βρίσκουμε μια περιγραφή της αρχετυπικής αυθεντίας, του προτύπου για τις επόμενες μορφές της. Αυτές όμως οι μορφές δεν μπορούν να υποστηρίξουν το διαχωρισμό της αυθεντίας από τη βία, καθώς τη διαχωρίζουν από την πολιτική. Έτσι, αντί να δημιουργείται μια κεντρική τάξη παθητικής υποταγής στην παράδοση, οδηγούμαστε σε ένα σύστημα όπου η αυθεντία διεισδύει σε κάθε επίπεδο. Για να αποσαφηνίσουμε όμως τη διάχυτη αυθεντία, πρέπει να εξετάσουμε άλλες δύο μορφές της, που εμφανίζονται στη διαλεκτική μεταξύ Διαφωτισμού και αστικής τάξης πραγμάτων: Την ταξική αυθεντία όπως εκφράζεται στην παραγωγή, και τη νεωτερική αυθεντία της γνώσης καθαυτή.

Η Άρεντ τελειώνει το άρθρο της λέγοντας τα εξής:

[Η αυθεντία νοθευμένη με βία συνεχίζει να επικρατεί] Γιατί η ζωή σε μια πολιτική επικράτεια δίχως αυθεντία και τη συνακόλουθη επίγνωση ότι η πηγή της αυθεντίας υπερβαίνει την εξουσία και όσους βρίσκονται στην εξουσία, σημαίνει πως αντιμετωπίζουμε εξ υπαρχής, χωρίς τη θρησκευτική πίστη σε μια ιερή απαρχή και χωρίς την προστασία των παραδοσιακών και επομένως αυτονόητων κριτηρίων συμπεριφοράς, τα στοιχειώδη προβλήματα ανθρωπίνης συμβίωσης<sup>46</sup>.

Η προσέγγιση αυτή ξεκινάει από το Διαφωτισμό και στη συνέχεια αποτυγχάνει. Βασικός στόχος του Διαφωτισμού είναι να εξυψώσει το Λόγο, στον οποίο κάθε άνθρωπος μπορεί να έχει πρόσβαση, πάνω από κάθε προϋπάρχουσα αυθεντία. Χαρακτηριστικό παράδειγμα αποτελεί η γνωστή ρήση του Καντ, σύμφωνα με την οποία ο Διαφωτισμός είναι η έξοδος του ανθρώπου από την ανωριμότητα του, που συνίσταται στην αδυναμία του να μεταχειρίζεται το νου του χωρίς την καθοδήγηση κάποιου άλλου – μια ευθεία καταδίκη κάθε μορφής αυθεντίας<sup>47</sup>. Υπάρχει όμως στην γενική καταδίκη της αυθεντίας ένα θεμελιακό ζήτημα: Η οργάνωση της παραγωγής απαιτεί υπακοή, η οποία οδηγεί σε καταπίεση και εκμετάλλευση, αλλά είναι σε κάποιο μέτρο και στο συμφέρον αυτού που υπακούει. Όπου υπάρχει καταμερισμός της εργασίας, χρειάζεται αναγκαστικά ένα μέτρο «τυφλής» εμπιστοσύνης, καθώς δεν

---

<sup>46</sup>Ο.π., 220.

<sup>47</sup>Immanuel Kant, 'Απάντηση Στο Ερώτημα Τι Είναι Διαφωτισμός', στο *Τι Είναι Διαφωτισμός* (Αθήνα: Κριτική, 2014), 47.

μπορούν όλοι οι άνθρωποι να κατανοούν το σύνολο της παραγωγικής διαδικασίας<sup>48</sup>. Για τον Καντ, η λύση είναι να διαχωριστεί η δημόσια σφαίρα, όπου κάθε υποκείμενο θα εκφράζεται και θα φέρεται ως αυτόνομος συγγραφέας των ιδεών του, ή ελεύθερος επιστήμονας, με την προσωπική, η οποία, σε μια αντιστροφή, αντιστοιχεί στην κρατική θέση ή υπηρεσία που μπορεί να της έχει παραχωρηθεί<sup>49</sup>. Η σημασία της «κρατικής θέσης» εδώ είναι ιδιαίτερη, καθώς το βασικό παράδειγμα που αναλύει στη συνέχεια είναι αυτό ενός ιερωμένου – ο οποίος υποχρεούται να διδάσκει το δόγμα το πρωί, αλλά να εκδίδει τις ερμηνείες και τις θέσεις του το βράδυ, ακόμα και αν θα θεωρούντο αιρετικές. Η κρατική θέση λοιπόν αντιστοιχεί στην καθημερινή εργασία, όπου το υποκείμενο έχει το καθήκον να υπακούει στην αυθεντία των ανώτερων, η οποία εδραιώνεται μέσω του κράτους και του νόμου, αλλά ιδιωτικοποιείται. Το καθεστώς το οποίο ρητά προωθεί ο Καντ σε αυτό το κείμενο είναι αυτό της πεφωτισμένης δεσποτείας, κρίνοντας απαραίτητη τη διατήρηση της παραδοσιακής δομής της αυθεντίας (τον μονάρχη και την Εκκλησία), λόγω της ευελιξίας που θα μπορούσε θεωρητικά να της δώσει αυτή η θέση: μια αυθεντία πλήρης και αυθαίρετη μπορεί να ανεχτεί και να αφομοιώσει τις νέες ιδέες, επειδή ακριβώς η δικαιολόγησή της δεν θεμελιώνεται ούτε στη βία ούτε στην πειθώ. Τέτοιες αναγκαίες εκπτώσεις βρίσκουμε στους περισσότερους μεγάλους θεμελιωτές του Διαφωτισμού, χαρακτηριστικά στον Καρτέσιο και στον Φίχτε, όπως παρουσιάζεται και στο κείμενο του Μαρκούζε στην *Αυθεντία και Οικογένεια (Από τη σκοπιά της ιστορίας των ιδεών)*<sup>50</sup>.

Η γενεαλογία όμως της Άρεντ αναδεικνύει τις αδυναμίες της μεσαιωνικής δομής της αυθεντίας (άρα και της πεφωτισμένης δεσποτείας), οι οποίες μπορούν να μας οδηγήσουν και στην κατανόηση της κρίσης του Διαφωτισμού. Ο Καντ στο συγκεκριμένο κείμενο δεν πραγματεύεται ακριβώς μια αυθεντία της γνώσης, παρά προσπαθεί να εξάγει την κυκλοφορία της γνώσης από μια δομή της αυθεντίας που νομιμοποιείται από το Λόγο και επιβάλλεται από το κράτος. Η πίστη του στις διαχειριστικές ικανότητες της πεφωτισμένης δεσποτείας είναι αβάσιμη, καθώς η παθολογική αυθεντία της Εκκλησίας, όπως παρουσιάστηκε προηγουμένως, δεν θα μπορούσε να υπομείνει, πόσο μάλλον να ενσωματώσει, επιχειρηματολογία που βάλλει το δόγμα της. Ακόμα, δηλαδή, και αν ο μονάρχης ήταν επιεικής προς μια

---

<sup>48</sup>Friedrich Engels, 'On Authority', στο *Marx and Engels Collected Works*, vol. 23 (Progress Publishers, 1988), 422–25.

<sup>49</sup>Απάντηση Στο Ερώτημα Τι Είναι Διαφωτισμός', 49.

<sup>50</sup>Max Horkheimer, Erich Fromm, και Herbert Marcuse, *Αυθεντία Και Οικογένεια*, 199–311.



οικονομία της ελεύθερης γνώσης, δεν θα μπορούσε αυτή να συμφιλιωθεί με τα θεμέλια της εξουσίας του.

Η σχέση με τη γνώση που προτείνουν οι Διαφωτιστές διαφέρει σημαντικά από το σχήμα του Πλάτωνα, καθώς είναι καθήκον κάθε υποκειμένου να την αναζητεί ως ίσο με όλα τα υπόλοιπα. Καταλήγει όμως στο ίδιο συμπέρασμα, το οποίο είναι ότι αν δεν είναι κάποιος φορέας της γνώσης είναι «ανώριμος», και άρα είναι στο συμφέρον του να υπακούει. Η ανάλυση της *Πολιτείας* της Άρεντ μας επιτρέπει να δούμε τα δύο ως ισοδύναμα μέσα από τις ιστορικές τους συνέπειες. Ακόμα και αν, ως πολιτικοί μύθοι, οι φιλοσοφίες αυτές είχαν χειραφετητικό σκοπό, έχουν τη δυνατότητα να υποστηρίξουν ένα σύστημα βίαιας κυριαρχίας. Η ορθολογική δημοκρατία του Διαφωτισμού μπορεί εύκολα να διαβαστεί ως το αντίθετό της. Ο Καντ θεμελιώνει τον Λόγο ως υπερβατικό με τρόπο που συγγενεύει με τον κόσμο των ιδεών του Πλάτωνα ή με την θεϊκή αλήθεια της εκκλησίας. Κανείς δεν μπορεί να τα συλλάβει πλήρως, αλλά η πρόσβαση σε αυτά, η γνώση της αλήθειας, είναι και αυταξία αλλά και εφαρμόσιμη στον υλικό κόσμο. Ο Λόγος του Καντ όμως υπάρχει μέσα στο υποκείμενο εκ των προτέρων, δεν είναι ένας κόσμος εκτός. Έτσι η γνώση των πραγμάτων γίνεται ειδοποιό ανθρώπινο καθήκον στο επίπεδο του υποκειμένου. Στη θέση της αυθεντίας τίθεται ο νόμος, ο οποίος περιορίζει τα άτομα στο ιδιωτικό επίπεδο (της εργασίας) και συμφωνείται στο γενικό επίπεδο (της «συγγραφής», της επιστήμης, της πολιτικής). Η αυθεντία αυτή θεμελιώνεται λοιπόν στην κοινή συναίνεση και συμφωνία των έλλογων υποκειμένων, με τρόπο που για τον Καντ διασφαλίζει τα συμφέροντα όλων. Στα λόγια όμως του Μαρκούζε:

Αλλά όπως κατά την εφαρμογή του νόμου ο εξορθολογισμός αφήνει ανέπαφες τις εσωτερικές αντιφάσεις της αστικής κοινωνίας, την ενύπαρκτη «αδικία» της, έτσι και στην πηγή της ίδιας της νομοθεσίας: «η δυνατότητα μιας νοητής κατοχής, κατά συνέπεια και του εξωτερικού δικού μου και δικού σου, δεν μπορεί να κατανοηθεί, αλλά πρέπει να προκύψει συμπερασματικά από το αξιωματικό αίτημα του πρακτικού Λόγου». Ο νόμος παραμένει μια αυθεντία που, όπως και οι πηγές της, δεν μπορεί να θεμελιωθεί ορθολογικά χωρίς να ξεπεραστούν τα όρια εκείνης ακριβώς της κοινωνίας για την υπόσταση της οποίας είναι αναγκαία<sup>51</sup>.

---

<sup>51</sup>Ο.π., 246.

Η κριτική αυτή βλέπει την πίστη του Καντ στο ότι ο νόμος (είτε του πεφωτισμένου δεσπότη είτε της αστικής δημοκρατίας) θα συμφωνηθεί με βάση τις ανάγκες όλων ως αφελή ή και παραπλανητική. Ένας νόμος που βασίζεται στην προστασία της ατομικής ιδιοκτησίας, ο οποίος θεσπίζεται από άτομα με άνιση ισχύ και περιουσία που φαντάζονται τον εαυτό τους ως ίσα, θα ρέπει αναγκαστικά προς την προστασία των συμφερόντων αυτών που έχουν τα περισσότερα. Η επίφαση αυτή της ουδετερότητας, που χρησιμοποιεί την ιδέα του κοινού σε όλους Λόγου για να δικαιολογηθεί, βρίσκεται στη βάση του συστήματος της αυθεντίας της νεωτερικότητας.

Η μαρξιστική παράδοση της Σχολής της Φρανκφούρτης θεματοποιεί το πώς αυτές οι αντιφάσεις οδήγησαν την, κυρίαρχη πια, αστική τάξη που προήλθε από το Διαφωτισμό στο να συμφιλιωθεί με τις προσμίξεις βίας στη δομή της αυθεντίας που δημιούργησε. Όπως είδαμε και παραπάνω, η υπακοή, με την έννοια της συμμόρφωσης σε εντολές χωρίς να απαιτείται επιχειρηματολογία, είναι αναγκαίο κομμάτι της παραγωγής σε μια κοινωνία με καταμερισμό εργασίας. Η ανισότητα στην εμπειρία και τη γνώση του αντικειμένου μέσα σε ένα περιβάλλον εργασίας δημιουργεί σχέσεις εξάρτησης και υπακοής, που όμως επιτρέπουν την εξέλιξη του ανειδίκευτου σε ειδήμονα. Όταν όμως στη σχέση αυτή εισέρχεται η ταξική ανισότητα, τότε βλέπουμε ότι η κατοχή κεφαλαίου αποτελεί κυρίαρχο γνώμονα για το ποιος θα δίνει εντολές και ποιος θα αναγκάζεται να τις υπακούει, και όχι η λειτουργική γνώση εντός της παραγωγικής διαδικασίας<sup>52</sup>. Ο κάτοχος των μέσων παραγωγής αναλαμβάνει διοικητική θέση στην παραγωγή αποκλειστικά λόγω αυτής της εκ των προτέρων υπεροχής του. Αυτή η μορφή της αυθεντίας θεμελιώνεται στην πίστη ότι η αγορά, η κίνηση του κεφαλαίου, ακολουθεί κάποιους κανόνες που αντανakλούν τις αναγκαιότητες της κοινωνίας εν γένει, και άρα οι κυρίαρχες τάξεις κατέχουν μια ανώτερη λειτουργική γνώση, για να έχουν κυριαρχήσει. Επίσης, η ιδέα της ελεύθερης ταξικής κινητικότητας ενισχύει την πεποίθηση ότι τα κριτήρια της κοινωνικής ισχύος είναι όχι μόνο ορθολογικά αλλά και αξιολογικά. Αν η κίνηση του κεφαλαίου ακολουθεί κάποιους περίπλοκους αλλά ορθολογικούς κανόνες, τότε αυτοί οι οποίοι αποκτούν περισσότερα χρήματα και ισχύ έχουν μια ικανότητα να προβλέπουν αυτές τις κινήσεις, μια πρόσβαση στους κανόνες της ανώτερης των υποταγμένων<sup>53</sup>.

---

<sup>52</sup>Ο.π., 59.

<sup>53</sup>Ο.π., 62.

Η επίφαση αυτή της ευφυΐας έχει τις ρίζες της στην διαφωτιστική ιδέα του υποκειμένου που κυριαρχεί στη φύση κατανοώντας την. Το να βλέπουμε όμως την κοινωνία σαν ακατανόητη, και άρα εκμεταλλεύσιμη με την ίδια μέθοδο όπως η φύση, είναι ένα αντιφατικό εσωτερικά παράγωγο της αστικής τάξης πραγμάτων. Η προσπάθεια να ανοικοδομηθεί η κοινωνία και η παραγωγή ορθολογικά, προσέκρουσε στο ίδιο σάσισμα που περιέγραφε η Άρεντ για την Εκκλησία: Ο κόσμος των ανθρώπων είναι υπερβολικά περίπλοκος και έτσι η υπερπολιτική αυθεντία αναγκάζεται να επιλέξει μεταξύ της κυριαρχίας της και της εσωτερικής της συνοχής<sup>54</sup>.

Στη διάχυτη αυθεντία που ευνοεί την αστική τάξη στον σύγχρονο καπιταλισμό, τα πράγματα είναι ακόμη πιο περίπλοκα. Καθώς δεν υπάρχει ένα κεντρικό δόγμα ή θεσμός για να επιβάλλει τα παραπάνω, ο μεμονωμένος αστός αναγκάζεται να ενισχύει τους μύθους του καπιταλισμού, νιώθοντας όμως το ίδιο δέος προς την ακατανόητη αγορά που νιώθει και ο (πειθήνιος) εργαζόμενος, του οποίου την εργασία εκμεταλλεύεται. Η «αυτονομημένη αναγκαιότητα» όμως που αντιπροσωπεύει η αγορά αντιστοιχεί για τον προλετάριο στον άμεσο φόβο για την πείνα και την αθλιότητα, ενώ για τον αστό στην προϋπόθεση της κυριαρχίας του. Αν λοιπόν για να διατηρήσει τα προνόμιά του ο μεμονωμένος αστός έχει συμφέρον να ενσαρκώσει τη φαντασίωση του μεγαλοφυή δαμαστή της αγοράς, θα το κάνει, και ενδεχομένως θα το πιστέψει και ο ίδιος. Το ίδιο θα κάνουν και θα πιστέψουν όσοι και όσες θεωρούν ότι είναι στο συμφέρον τους να διεκδικήσουν αυτά τα προνόμια μέσω της υποταγής τους. Η αντανάκλαση μεταξύ αυτών των δύο ομάδων δημιουργεί την ανάγκη η αυθεντία να μετατοπίζεται όλο και υψηλότερα, σε ένα φαντασιακό που διευρύνει τον πρωταρχικό φόβο της υποταγής στην αναγκαιότητα ακόμα και σε αυτούς που δεν είναι σε κίνδυνο να τη γνωρίσουν. Μια ιεραρχία χωρίς κεφάλι, μια επιτακτική ανάγκη ανέλιξης στη σκάλα του κέρδους χωρίς όρια.

Το κείμενο του Φρομμ στην *Αυθεντία και Οικογένεια* (Από τη σκοπιά της κοινωνικής ψυχολογίας) εστιάζει στην κατηγορία της *σαδομαζοχιστικής προσωπικότητας*<sup>55</sup>. Η συμπεριφορά αυτής της μερίδας ανθρώπων (ή αυτή η μερίδα της συμπεριφοράς των ανθρώπων) δρα πάντα υπέρ της αυθεντίας και κατά του υποταγμένου. Η αυθεντικοκεντρική συμπεριφορά αγαπά την αυθεντία τυφλά, και φαντάζεται την πυραμίδα της απρόσωπης ιεραρχίας με τρόπο που την καθηλώνει. Ο αμέσως από

---

<sup>54</sup>Hannah Arendt, 'Τι Είναι Αυθεντία', 207.

<sup>55</sup>Max Horkheimer, Erich Fromm, και Herbert Marcuse, *Αυθεντία Και Οικογένεια*, 162–92.

πάνω χρίζει έτσι λατρείας και υποταγής, στο ίδιο μέτρο που ο αμέσως από κάτω μπορεί να χρησιμοποιείται αυθαίρετα για να αντιλαμβάνεται ο προϊστάμενος την αξία της θέσης του ως πραγματική. Η αγάπη αυτή για την υποταγή αποτελεί ένα ακόμα βασικό πρόβλημα στην κριτική της αυθεντίας, και είναι άμεσα συνδεδεμένη με τον φόβο. Όπως λέει ο Φρομμ, μέσα από αυτήν την περιγραφή ουσιαστικά μιας νεύρωσης, μπορούμε να καταλάβουμε τις συνθήκες στις οποίες η αυθεντία γίνεται ορθολογική – ακόμα και αν δεν ήταν εξαρχής.<sup>56</sup> Στη βάση της υπάρχει η αποστροφή προς την προσωπική επιλογή, ένα αίσθημα που έχει ιδιαίτερο ενδιαφέρον για το θέμα της ΤΝ.

Όπως είδαμε μέσα από το κείμενο του Χορκχάιμερ, η αυθεντία στο σύγχρονο καπιταλισμό εμπλέκει τρία πεδία, καθένα από τα οποία παρουσιάζεται ως «γενικό». Αυτά είναι το κράτος, ο Λόγος, και η οικονομία. Από την άποψη όμως της ανάλυσης του Μαρξ, μπορούμε να αντιληφθούμε ότι το μόνο πραγματικά γενικό επίπεδο στο σύγχρονο καπιταλισμό είναι αυτό της κυκλοφορίας του κεφαλαίου. Το κράτος συμβιβάζεται στις επιταγές του σαν αναγκαιότητες, ενώ ο Λόγος εκφυλλίζεται μέσω των αντιφάσεων που δημιουργεί εντός του η μέριμνα για τη σταθερότητα μιας παράλογης τάξης πραγμάτων. Παρόλα αυτά η επιστήμη και η τεχνική, η γνώση και οι εφαρμογές της, προοδεύουν. Η πρόοδος αυτή διευρύνει την επιρροή τους όχι μόνο στη φύση, αλλά και το εσωτερικό της κοινωνίας, με τρόπους αντικειμενικά χρήσιμους και ευεργετικούς, αλλά και διαχειριστικούς ή αυθεντικοκεντρικούς. Η μορφή αυτή της διάχυτης γνώσης φέρει μέσα της τη σφραγίδα της αυθεντίας του κεφαλαίου, ή της ψευδούς αναγκαιότητας που συζητήσαμε παραπάνω. Για να διατηρηθεί η κυριαρχία του κεφαλαίου πάνω στους ανθρώπους πρέπει η αγορά να είναι υπεράνω πρόβλεψης, εμπρόθετης ορθολογικής οργάνωσης, και πραγματικής εξυγίανσης, ενώ συγχρόνως πρέπει να διατηρείται η ιδέα ότι αυτοί που κυριαρχούν σε αυτήν τη δαμάζουν και την καταλαβαίνουν. Έτσι η διάχυτη αυθεντία της γνώσης(/εξουσίας) βιώνεται ως κάτι που είναι συγχρόνως χρήσιμο και καταπιεστικό, ορθολογικό αλλά και παράλογο, και παρουσιάζεται ως η πηγή της αυθεντίας του κεφαλαίου, σε βαθμό που τελικά μοιάζει να ισχύει και το αντίθετο. Η κοινωνία έτσι δεν γίνεται μόνο πιο σύνθετη – πράγμα που θα ίσχυε έτσι κι αλλιώς με την πρόοδο της γνώσης και της τεχνικής – αλλά και πιο παράλογη – καθώς η πρόοδος αυτή δεν οδηγεί στην ευμάρεια που εμφανώς επιτρέπει.

---

<sup>56</sup>Ο.π., 192.

Επιστρέφοντας στην περιγραφή της διάχυτης εκκλησιαστικής αυθεντίας της Άρεντ που είδαμε παραπάνω, η αγορά ως υπερπολιτική πηγή αυθεντίας οδηγεί σε μια ανάγκη να δικαιολογηθεί κάθε πράξη σύμφωνα με τη μεγιστοποίηση του κέρδους. Καθώς όμως δεν υπάρχουν ιερείς του καπιταλισμού, τα «κριτήρια για τη δικαιολόγηση κάθε πράξης» προέρχονται από την παρατήρηση προτύπων. Τη θέση αυτή παίρνουν επιτυχημένοι επιχειρηματίες, πλούσιοι χρηματιστές, ή σε πιο ειδικό επίπεδο προϊστάμενοι και σύμβουλοι επιχειρήσεων. Αυτή είναι η σύγχρονη μορφή της αυθεντικοκεντρικής συμπεριφοράς, η δράση η οποία, μέσα στον εκφυλισμό του ορθού Λόγου, έχει χάσει κάθε αξιολογικό κριτήριο για τον εαυτό της και το ψάχνει παντού γύρω της.

Η τρίτη μορφή της αυθεντίας έχει να κάνει με την ειδημοσύνη, κυρίως με την έννοια της λειτουργικής γνώσης. Τη σχέση αυτή τείνουμε να μην την βλέπουμε ως μορφή της αυθεντίας, καθώς αποτελεί προϋπόθεση για την ορθολογική οργάνωση της κοινωνίας, είτε εμπρόθετη (κατά Μαρξ) είτε αναφαινόμενη (κατά Καντ). Με τον ίδιο τρόπο παρουσιάζονται συνήθως και η γονεϊκή αυθεντία και αυτή του εκπαιδευτικού πάνω στο παιδί ή στον εκπαιδευόμενο, τις οποίες οι Άρεντ ονομάζει προπολιτικές, όπως χαρακτηριστικά παρουσιάζεται και η γνώση (ως δημιουργία αφαιρετικών κατηγοριών) στη *Διαλεκτική του Διαφωτισμού*<sup>57</sup>. Ακόμα όμως και αν δούμε τις κακές εκφάνσεις τους ως εποικοδόμημα, το οποίο δεν θα βιώνεται ως αυθεντία μετά την αλλαγή της βάσης, έχει νόημα να τις αναλύσουμε μέχρι ένα σημείο ως καθαυτή μορφή της αυθεντίας.

Η παραδειγματική σχέση για την αυθεντία του ειδήμονα είναι συνήθως αυτή του γιατρού με τον ασθενή του, και αυτό γιατί η συγκεκριμένη σχέση είναι ξεκάθαρα μια σχέση εξουσίας. Ο γιατρός βέβαια είναι σε αυτή του την έκφανση τεχνικός, προσπαθεί να λύσει, δηλαδή, το πρόβλημα που υπαγορεύεται από τον ασθενή του. Η λύση του προβλήματος ισούται με την απόσυρση του αιτήματος από μέρους του ασθενούς. Πρέπει θεωρήσουμε ότι ο αρχετυπικός ασθενής δεν έχει τρόπο να κρίνει ή να συμμετέχει στη θεραπεία του, θα αφεθεί λοιπόν να πάρει κάποιο φάρμακο ή να κάνει κάποια εγχείρηση με μόνο κριτήριο αυτά που του λέει ο γιατρός. Βέβαια, ο ασθενής μπορεί να έχει πρόσβαση στη γνώση του γιατρού (όπως για παράδειγμα στο πανεπιστήμιο ή στο ίντερνετ), ή σε περισσότερους ειδήμονες που θα μπορούσε να συμβουλευθεί, συνήθως όμως δεν έχει τον χρόνο ή τη θέληση να φτάσει σε σημείο

---

<sup>57</sup>Theodor W. Adorno και Max Horkheimer, *Διαλεκτική Του Διαφωτισμού*, 77.

που να κατανοεί ο ίδιος την κατάστασή του ή να μπορεί να κρίνει τις αποφάσεις του γιατρού.

Είναι γεγονός ότι από τη στιγμή που η ιατρική περίθαλψη γίνεται μέριμνα της κοινωνίας συνολικά, τότε δεν μιλάμε πια για μεμονωμένα αιτήματα μεμονωμένων ασθενών. Η διαχείριση της υγείας σε μεγάλη κλίμακα δημιουργεί την ανάγκη για πρόνοια, για κανόνες υγιεινής, για προτροπή σε εξέταση ή εμβολιασμό. Καθώς τους κανόνες αυτούς τους επιβάλλει το κράτος, κάποια γιατρός πρέπει να αναλάβει το ρόλο της εμπειρογνώμονα, που συμβουλεύει τους πολιτικούς για τη μορφή αυτών των κανόνων. Η λειτουργία αυτή παρουσιάζεται τις περισσότερες φορές ως μη πολιτική. Σε αυτήν την αντίληψη, ο εμπειρογνώμων παίρνει τη θέση μιας «καθαρής» αυθεντίας της εφαρμοσμένης γνώσης, φαίνεται όμως μερικές φορές δύσκολο να πιστέψουμε ότι αφενός ο πολιτικός θα συμμορφωθεί με την αυθεντία στο βαθμό που παρουσιάζει, όπως και ότι ο εμπειρογνώμων δεν άγεται από κριτήρια που έχουν να κάνουν με το προσωπικό του όφελος.

Διευρύνοντας όμως το παράδειγμα, την ίδια αυτή περίπλοκη σχέση μπορούμε να αναγνωρίσουμε και σε άλλα πεδία της τεχνικής (εδώ ως η λειτουργική γνώση). Οι συσκευές που χρησιμοποιούμε καθημερινά είναι παρομοίως ακατανόητες με το σώμα μας, είναι όμως πιο σπάνιο να χαρακτηρίσουμε τη σχέση μας με τον τεχνίτη που μας τις επισκευάζει ως σχέση εξουσίας, ή τον ίδιο ως φορέα αυθεντίας. Η σχέση όμως αυτή είναι πλέον ένα καθημερινό παράδειγμα *αποδεκτής εξάρτησης*, και σε μεγάλο βαθμό πάντα ήταν, με μόνη διαφορά ότι η περιπλοκότητα με την οποία έχουμε πλέον να κάνουμε σε κάθε στιγμή της ζωής μας έχει αυξηθεί εκθετικά.

Η ασύλληπτη περιπλοκότητα ενός υπολογιστή ή ενός κινητού τηλεφώνου κάνει την κατανόησή τους να φαντάζει αδύνατη, άρα η ανεξαρτησία από τους μηχανικούς τους φαντάζει κατά καιρούς όσο δύσκολη θα ήταν η ανεξαρτησία από τους γιατρούς. Η καθησυχαστική σκέψη στην οποία μπορεί να καταφύγουμε είναι ότι *κάποιος* το καταλαβαίνει για εμάς. Και για τις συσκευές μας αυτό υποστηρίζεται επίσης από το γεγονός ότι *κάποιος* της έφτιαξε. Συνήθως επίσης ιεραρχούμε την αναγκαιότητα των λειτουργιών αυτών ώστε να κρίνουμε ποιες από αυτές τις εξαρτήσεις είναι προσωπικές μας επιλογές, και άρα θα μπορούσαμε να απολέσουμε, και ποιες όχι, σε μια προσπάθεια να προσεγγίσουμε την *αυτάρκεια* – έναν επίσης πολυσήμαντο όρο. Η καθολικότητα όμως της χρήσης των ηλεκτρονικών συσκευών στην εργασία και στην

προσωπική μας ζωή θα μας έκανε στην πράξη να αντιμετωπίζουμε ένα άτομο που αρνείται να τις χρησιμοποιήσει με μεγαλύτερη καχυποψία από κάποιο που δεν εμπιστεύεται τη «δυτική ιατρική».

Η αντιστροφή, που είναι εύκολα παρατηρήσιμη αν εξετάσουμε ιστορικά τη λειτουργική γνώση καθαυτή, συνίσταται στο ότι παρόλο που τυπικά αποτελεί τον κύριο όρο του καπιταλισμού καταλήγει δευτερεύουσα. Η αντικειμενική κοινωνική της αξία μπορεί μάλιστα να θεωρηθεί η πηγή της παθολογικής κατανόησης της αναγκαιότητας που περιγράψαμε παραπάνω: Η πρόοδος της τεχνολογίας και της τεχνογνωσίας, η οποία θα μας επιτρέψει να ξεπεράσουμε τις αναγκαιότητες που επιβάλλει η φύση, θεσπίζεται έτσι σαν αναγκαιότητα η ίδια, στην πορεία όμως ξεχνιέται ο σκοπός της. Έτσι η τεχνολογική πρόοδος βιώνεται ως περιπλοκή της κοινωνίας, ενώ η αγορά γίνεται μοναδικό μέτρο της αξίας. Με τη γνώση να είναι πλέον προσβάσιμη, αλλά ακόμη αγνώσιμη, η αυθεντία της ειδημοσύνης είναι πιο παρούσα, και πιο νεφελώδης από ποτέ. Προβάλλεται δηλαδή ως κατανόηση των κρυφών κανόνων της αγοράς-ως-αναγκαιότητας, με τον τρόπο που προσεγγίσαμε μέσα από τη *μεγαλοφυΐα* του Χορκχάιμερ – μια κοινή ουσία στην οποία κάποιοι έχουν προνομιακή πρόσβαση.

Ταυτόχρονα, οι επί μέρους ειδικοί δεν έχουν καμία δύναμη να ορίζουν τις συνθήκες της εργασίας τους, και εκ των πραγμάτων δυσκολεύονται να αναζητήσουν γνώση που να μην έχει άμεση εφαρμογή στους σκοπούς της «αγοράς». Ενώ λοιπόν η αυθεντία αυτή βρίσκεται εν δυνάμει παντού, δεν βρίσκεται ταυτόχρονα και πουθενά. Ενώ, μέσω της διείσδυσής της και της πραγματοποίησης των μέσων της, είναι στην πιο πλήρη της μορφή, οι επί μέρους εκφάνσεις της μοιάζουν ευτελείς και ανυπόστατες. Η ίδια η γνώση, που είναι προσβάσιμη από τον καθένα, μοιάζει ασυνάρτητη, καθώς η τάξη πραγμάτων που την επικαλείται είναι εμφανώς αντιφατική. Ο δημόσιος χώρος είναι όντως ανοιχτός για το κάθε άτομο που θέλει να συνομιλήσει σαν συγγραφέας των ιδεών του ή επιστήμων, αλλά η κοινωνική πραγματικότητα επιβάλλει τον εαυτό της εμφατικά ως ανορθολογικό.

Ο τεχνίτης και ο επιστήμονας συνεχίζουν να είναι τα σύμβολα της ελεύθερης αγοράς και του διαφωτιστικού Λόγου, σε μια εποχή που και τα δύο αυτά τα έχει καταπιεί αυτό που οι Αντόρνο και Χορκχάιμερ περιγράφουν ως «ο ασυνείδητος κολοσσός της

πραγματικότητας, ο χωρίς υποκείμενο καπιταλισμός»<sup>58</sup>, ενώ ταυτόχρονα επιβιώνουν ως κενά ουσίας πρότυπα. Η αυθεντία της γνώσης και της εξειδίκευσης γίνεται έτσι ένα ανδρείκελο, ένα προπέτασμα που κατά καιρούς απορροφά τον αποπροσανατολισμένο θυμό προς την ισχύουσα τάξη πραγμάτων, και, βέβαια, άλλες φορές την ενσαρκώνει παραδειγματικά. Η πρόοδος, που οι μερίδες αυτές πραγματικά κινούν, οι νέες εφευρέσεις και κατανοήσεις του κόσμου, παραμερίζονται ή εφαρμόζονται βάρβαρα με τέτοια συχνότητα, που η μόνη ορθολογική αντίδραση είναι να αποφεύγει η ερευνήτρια να σκέφτεται την κοινωνική αξία της έρευνάς της όσο το δυνατόν περισσότερο.

Η αντίθεση μεταξύ παθητικότητας και ενεργητικότητας, που εμφανίζεται στη Γνωσιοθεωρία σαν δυϊσμός μεταξύ αισθητικότητας και διάνοιας, δεν ισχύει για την κοινωνία στο ίδιο μέτρο που ισχύει για το άτομο. Ενώ δηλαδή το τελευταίο θεωρεί τον εαυτό του παθητικό και εξαρτημένο, εκείνη, αν και αποτελείται από άτομα, είναι εντούτοις ένα υποκείμενο – χωρίς συνείδηση και συνεπώς όχι καθαυτό, αλλά πάντως ενεργό υποκείμενο. Αυτή η διαφορά ως προς την ύπαρξη του ανθρώπου και της κοινωνίας είναι έκφραση της διάσπασης που χαρακτηρίζει μέχρι σήμερα τις ιστορικές μορφές της κοινωνικής ζωής. Η ύπαρξη της κοινωνίας είτε βασίστηκε στην άμεση καταπίεση είτε ήταν η συνισταμένη αντίρροπων δυνάμεων· σε καμία περίπτωση όμως δεν ήταν αποτέλεσμα της συνειδητής αυτόνομης δράσης των ελεύθερων ατόμων. Γι' αυτό και η σημασία των εννοιών της παθητικότητας και της ενεργητικότητας αλλάζει, ανάλογα με το αν αναφέρεται στην κοινωνία ή στο άτομο. Μέσα στην αστική οικονομική διαδικασία η δραστηριότητα της κοινωνίας είναι τυφλή και συγκεκριμένη, ενώ του ατόμου αφηρημένη και συνειδητή.<sup>59</sup>

Μέσα από την ανάλυση της *μαριονέτας* εισήγα ήδη την ιδέα ότι μια ορθολογική μηχανή μπορεί να ενσαρκώσει το πρότυπο του φιλοσόφου-βασιλέα κατά τον Πλάτωνα – τον γνώστη που είναι στο συμφέρον όλων να υπακούν. Κάτι τέτοιο θα απαιτούσε την αναγωγή του Λόγου σε θεό. Τότε μόνο θα μπορούσε ένας προνομιακός φορέας του να αναδειχθεί σε κυρίαρχη αυθεντία, ακόμα κι αν ήταν μηχανή. Από μια άποψη, κάτι τέτοιο θα μπορούσαμε να το φανταστούμε ακόμα και ως μια κάποια λύση των αντιφάσεων που μας καταδυναστεύουν. Αλλά ο θεός είναι νεκρός, και οι μόνοι κανόνες που καταδεχόμαστε, οι μόνες αυθεντίες, είναι οι απολύτως αναγκαίοι. Καθώς όμως η αναγκαιότητα αυτή αυτονομείται και δημιουργεί μηχανισμούς ώστε να αυτοσυντηρηθεί, χάνοντας κάθε πραγματική

---

<sup>58</sup>Ο.π., 190.

<sup>59</sup>Max Horkheimer, *Φιλοσοφία και Κοινωνική Κριτική* (Αθήνα: Εκδόσεις Ύψιλον, 1984), σ. 21



σημασία, η ηθική του νέου αυτού κόσμου αντιστρέφεται. Αντί να προσπαθεί να σπάσει τα είδωλα, η γνώση επικεντρώνεται έτσι στο πώς κινείται κανείς ανάμεσα σε αυτά. Η νέα μορφή τους, ως ψευδώς ορθολογικοί θεσμοί, επιτάσσει και τη νέα ιδανική κινησιολογία. Η ΤΝ όμως, ακόμα και αν έχει τη δυνατότητα στη θεωρία ή στα μάτια των δημιουργών της να προσεγγίσει τον θεό-Λόγο, επιδρά εδώ και πολλά χρόνια στις επιμέρους οικονομίες της γνώσης και της πρακτικής στη διαδικασία παραγωγής. Ας μιλήσουμε, λοιπόν, για την αυτοματοποίηση.

## Για την αυτοματοποίηση

Κινούμενοι στη σύγχρονη κοινωνία έχουμε συχνά την εντύπωση ότι υπάρχουν θέσεις εργασίας, ή ακόμα και ολόκληρες κατηγορίες της, που δεν παράγουν κάτι ουσιαστικό για κανέναν. Την ιδέα αυτή συνήθως την αντιλαμβανόμαστε ως σχετικά κιτς – όσο και να φιλοσοφούμε για τις αντιφάσεις του καπιταλισμού δεν γίνεται να πληρώνονται άτομα ή εταιρίες για να μην κάνουν τίποτα. Ο Ντέιβιντ Γκρέμπερ δημοσίευσε αυτήν την αντίληψη το 2013 σε ένα προβοκατόρικο άρθρο στο περιοδικό *Strike!*<sup>60</sup>. Προς έκπληξή του, άρχισε να λαμβάνει απαντήσεις από άτομα που ήταν διατεθειμένα να παραδεχτούν ότι η δουλειά που κάνουν είναι μάταιη – ότι αν εξαφανίζονταν ένα πρωί, δεν θα άλλαζε τίποτα στον κόσμο. Διευρύνοντας την απεύθυνσή του για τέτοιες μαρτυρίες, ο Γκρέμπερ παρουσιάζει στο βιβλίο *Bullshit Jobs*<sup>61</sup> την έρευνά του γύρω από το φαινόμενο και αναλύσεις για την προέλευσή του.

Πολλές από αυτές τις δουλειές είναι άνετες και καλοπληρωμένες. Παρόλα αυτά, τα άτομα που στέλνουν τις μαρτυρίες τους εκφράζουν μια βαθιά δυσφορία. Ο συγγραφέας επικαλείται την έρευνα του ψυχολόγου Καρλ Γκρόος για να υποστηρίξει ότι αυτό που δημιουργεί αυτή τη δυσφορία είναι η έλλειψη αυτού που ονομάζει «ευχαρίστηση του να είσαι αίτιο» – της χαράς που παίρνει ένα βρέφος όταν καταλαβαίνει ότι «είναι ικανό να προκαλέσει προβλέψιμα αποτελέσματα στον περιβάλλοντα κόσμο του.»<sup>62</sup> Η θεώρηση αυτή συνδέει άμεσα την αίσθηση του εαυτού

---

<sup>60</sup>David Graeber, 'On the Phenomenon of Bullshit Jobs', STRIKE! Magazine (STRIKE!), accessed 29 September 2024, <https://strikemag.org>.

<sup>61</sup>David Graeber, *Bullshit Jobs: Μια Θεωρία* (Αθήνα: Στάσει Εκπίπτοντες, 2020).

<sup>62</sup>Ο.π., 113.

με τη δράση, και αποσυνδέει σε ένα βαθμό την επιδίωξη της εξουσίας από την πρακτική μέριμνα για την ασφάλεια ή την αναπαραγωγή του εαυτού.

Αυτό που ενδιέφερε πρωτίστως τον Γκρόος ήταν να εξετάσει γιατί οι άνθρωποι παίζουν με παιχνίδια και γιατί παθιάζονται και ενθουσιάζονται τόσο πολύ με το αποτέλεσμα, ακόμα και στις περιπτώσεις εκείνες που γνωρίζουν ότι δεν έχει καμία σημασία ποιος θα κερδίσει και ποιος θα χάσει πέρα από τα όρια του ίδιου του παιχνιδιού.<sup>63</sup>

Η επενέργεια σε ένα περιβάλλον που ορίζεται από κοινά συμφωνημένους κανόνες προσφέρει ευχαρίστηση ακόμα και χωρίς ουσιαστική απειλή ή επιβράβευση. Ο Γκρέμπερ όμως γενικεύει αυτήν την πρωταρχική ηδονή και στο ζήτημα της εργασίας. Το σχήμα του συνδυάζει την τάση της αυτοματοποίησης με αυτήν της ποσοτικοποίησης, ως παράγωγα της λογικής της μεγιστοποίησης του συνολικού κοινωνικού κέρδους. Η μισθωτή εργασία, μας λέει, είχε στην πρώιμη νεωτερικότητα χαρακτηριστικά ηθικής εκπαίδευσης, και θεωρούταν διαδικασία ενηλικίωσης. Για τον Γκρέμπερ, το πρότυπο αυτό επιβιώνει, παρόλο που η εποχή μας δικαιολογεί την εργασία με βάση την παραγωγή ποσοτικής αξίας αντί για την ενστάλαξη ηθικών αξιών. Στο προηγούμενο αυτό στάδιο, όταν ο άνθρωπος εξερχόταν από την ανωριμότητα του μέσω της μαθητείας ήταν αναμενόμενο να πάψει να δουλεύει ως μισθωτός εργαζόμενος, ασκώντας πλέον την τέχνη που έμαθε ως ενηλικιωμένος πολίτης. Η λειτουργική γνώση ήταν έτσι συνδεδεμένη όχι με τη μεγιστοποίηση του κέρδους, αλλά με την παραγωγή ηθικών υποκειμένων<sup>64</sup>.

Η ποσοτικοποίηση της αξίας οδηγεί για τον Γκρέμπερ στον κατακερματισμό των ενδιάμεσων βαθμίδων της παραγωγής σε περισσότερες και πιο συγκεκριμένες. Κάνει έτσι τελικά την παραγωγή πιο σύνθετη αντί να την απλοποιεί. Αναλύοντας τις σκοπιμότητες κάθε διεργασίας και απομονώνοντας τα ενδιάμεσα στάδιά της ως σημεία προς βελτιστοποίηση, δημιουργείται η εντύπωση ότι θα ήταν ωφέλιμο όλα τα στάδια να ελέγχονται ξεχωριστά από διαφορετικά είδη ειδημόνων. Οι επί μέρους ειδήμονες, καθώς το πεδίο τους ολοένα και συρρικνώνεται και αλληλοαναγνωρίζονται ως αυθεντίες, αποξενώνονται από την κατανόηση της παραγωγής σε επίπεδο που αρχίζουν να υποψιάζονται ότι η θέση τους μπορεί να είναι κάποιου είδους αριθμητικό λάθος. Στην προσπάθεια να εξαλείψει την

---

<sup>63</sup>Ο.π., 114.

<sup>64</sup>Ο.π., 255–64.

απροσδιοριστία του ανθρώπινου παράγοντα, να δημιουργήσει ένα «ατσάλινο κλουβί», η τάση αυτή φαίνεται έτσι να αυξάνει τις θέσεις εργασίας, και άρα την ανθρώπινη επικοινωνία, αφαιρώντας της όμως όσο το δυνατόν περισσότερο τη δυνατότητα εμπρόθετης δράσης.

Η αυτοματοποίηση δεν είναι πραγματικά επόμενο στάδιο αυτής της διαδικασίας -είναι η διαδικασία η ίδια<sup>65</sup>. Η ανθρώπινη εργασία αυτοματοποιείται με όργανο τους ίδιους τους ανθρώπους, αφαιρώντας τους την ηδονή της επιρροής στον κόσμο. Ταυτόχρονα, καταφέρνει έτσι να μετατρέψει την γενική ηθική εκπαίδευση που ήταν συνδεδεμένη με τη μισθωτή εργασία σε ενστάλαξη της ηθικής της διατήρησης του υπάρχοντος. Η δυσφορία που δημιουργεί αυτή η συνθήκη οδηγεί σε περαιτέρω πολλαπλασιασμό των bullshit jobs, καθώς προσλαμβάνονται άτομα ώστε να νιώθει ένας υψηλόβαθμος, αλλά απεγνωσμένος υπάλληλος ότι έχει κάποια εξουσία. Ο δρόμος της ανέλιξης είναι η αποδοχή της ηθικής της αυτοματοποίησης, η υπακοή στον ανώτερο, και η άσκηση εξουσίας σε άλλους ως μέτρο της αξίας: Η σαδομαζοχιστική αυθεντιοκεντρική συμπεριφορά γυμνή και πλήρης ως ο όρος του εαυτού της<sup>66</sup>.

Οι δουλειές που δεν είναι *φούμαρα* [bullshit] θεωρούνται με βάση αυτήν τη διαστροφική ηθική κολάσιμες<sup>67</sup>. Έτσι η εργασία που είναι κατά κοινή παραδοχή κοινωνικά και ηθικά ωφέλιμη καταλήγει να αμείβεται ελάχιστα, ενώ η αλλοτρίωση αδρά. Δάσκαλοι, νοσοκόμοι, εργάτες και καθαριστές πληρώνονται όλο και λιγότερο από τους συμβούλους επιχειρήσεων και τους διευθυντές μάρκετινγκ, παρόλο που και οι ίδιοι οι τελευταίοι παραδέχονται ότι η εργασία τους είναι λιγότερο αναγκαία. Στην περίοδο της πανδημίας του κόβιντ είχαμε μια από τις πιο τρανταχτές αποδείξεις αυτού του φαινομένου, με την δημόσια παραδοχή ότι μια συγκεκριμένη μερίδα της εργασίας είναι πράγματι αναγκαία να ανακοινώνεται χωρίς να συνοδεύεται από την παρατήρηση ότι οι εργαζόμενες αυτές είναι ανάμεσα στις χειρότερα αμειβόμενες – και ότι αυτό ίσως δεν είναι σωστό.

Η ειδημοσύνη της γνώσης και της τεχνικής βρίσκεται στο κέντρο αυτής της δίνης. Η προσφορά τους είναι εμφατικά ωφέλιμη, αλλά ταυτόχρονα χρησιμοποιείται για την ενίσχυση και την παγίωση των αντιφάσεων και των διαστροφών. Η

---

<sup>65</sup>Ο.π., 291–97.

<sup>66</sup>Ο.π., 279–85.

<sup>67</sup>Ο.π., 284–85.

ποσοτικοποίηση/αυτοματοποίηση και η αυτόνομη αναπαραγωγή των *bullshit* έχει εισέλθει και στα δικά τους πεδία, στα μανατζερικά πανεπιστήμια, στην τυποποίηση, στα τεστ πιστοποιήσεων, των οποίων η ουσία περιπλέκεται με την πραγματοποιημένη αναγκαιότητα<sup>68</sup>. Η αναγωγή επίσης όλων των μοριακών σταδίων της παραγωγής σε πεδία ειδημοσύνης έχει δημιουργήσει μια νέα τάξη φουμαροειδικών και φουμαρολόγων, που θολώνει τα νερά σε σχέση με την κοινωνική αναγκαιότητα της εργασίας της γνώσης. Επιστρέφοντας στην Άρεντ:

Τώρα πια γενικά και υπερβατικά κριτήρια, στα οποία θα μπορούσαν να υπαχθούν το μερικό και το εμμενές χρειάζονταν για οποιαδήποτε πολιτική τάξη, ηθικοί κανόνες για κάθε διανθρώπινη συμπεριφορά και ορθολογικά μέτρα για την καθοδήγηση κάθε ατομικής κρίσης<sup>69</sup>.

Και πίσω στον Γκρέμπερ:

Μεγάλο μέρος της φουμαροποίησης των πραγματικών δουλειών και της επέκτασης του κλάδου των φούμαρων θα έλεγα πως οφείλεται στην επιδίωξη να ποσοτικοποιηθεί αυτό που δεν μπορεί να ποσοτικοποιηθεί. Για να το θέσω ευθέως, με την αυτοματοποίηση κάποιες δουλειές γίνονται πιο αποτελεσματικά και κάποιες λιγότερο αποτελεσματικά. Αυτό οφείλεται στο ότι απαιτείται πολύ περισσότερη δουλειά από τους ανθρώπους προκειμένου να μετατρέψουν τις διαδικασίες, τις λειτουργίες και τα αποτελέσματα που σχετίζονται με οτιδήποτε έχει φροντιστική αξία σε μια μορφή που οι υπολογιστές μπορούν να αναγνωρίζουν. Πλέον γίνεται να κατασκευάσουμε ένα ρομπότ που μπορεί, από μόνο του, να ταξινομή έναν σωρό φρέσκων φρούτων ή λαχανικών σε ώριμα, άγουρα, και σάπια. Αυτό είναι καλό γιατί το να ξεχωρίζεις φρούτα, ύστερα από μια ή δύο ώρες, είναι βαρετό. Όμως, δεν γίνεται να κατασκευάσεις ένα ρομπότ που να μπορεί, από μόνο του, να σαρώνει έναν σωρό βιβλιογραφικές λίστες για το μάθημα της Ιστορίας και να αποφασίζει ποια βιβλία είναι καταλληλότερα για την παράδοση στο αμφιθέατρο. Ούτε αυτό είναι τόσο κακό, αφού μια τέτοια εργασία μπορεί να είναι ενδιαφέρουσα (ή, τουλάχιστον, δεν είναι δύσκολο να βρούμε ανθρώπους που θα την έβρισκαν ενδιαφέρουσα). Ένας λόγος για να έχουμε ρομπότ για να ξεχωρίζουν τα φρούτα είναι για να έχουν οι πραγματικοί άνθρωποι περισσότερο χρόνο να σκεφτούν ποιο μάθημα Ιστορίας θα προτιμούσαν να παρακολουθήσουν ή κάποιο άλλο πράγμα που δεν μπορεί επίσης να ποσοτικοποιηθεί [...] Ωστόσο – και εδώ είναι η παγίδα -, αν, για κάποιο λόγο θέλαμε να πιστέψουμε πως ένας υπολογιστής θα μπορούσε να αποφασίσει ποιο είναι το καλύτερο μάθημα Ιστορίας, επειδή αποφασίσαμε, ας πούμε,

---

<sup>68</sup>Ο.π., 297–300.

<sup>69</sup>Hannah Arendt, 'Τι Είναι Αυθεντία', 196.

ότι πρέπει να έχουμε ομοιόμορφα ποσοτικά μετρήσιμα, «ποιοτικά» κριτήρια που να εναρμονίζονται σε όλα τα πανεπιστήμια για λόγους χρηματοδότησης, δεν υπάρχει περίπτωση να μπορούσε ένας υπολογιστής να φέρει σε πέρας το έργο αυτό από μόνος του. Τα φρούτα μπορείτε απλώς να τα πετάξετε στα σκουπίδια. Στην περίπτωση ενός μαθήματος Ιστορίας, απαιτείται τεράστια ανθρώπινη προσπάθεια για να μετατραπεί το υλικό σε ενότητες τις οποίες ένας υπολογιστής θα μπορούσε να χειριστεί στοιχειωδώς<sup>70</sup>.

Ο Γκρέμπερ κάνει λάθος. Ο περιορισμός του υπολογιστή που περιγράφει ισχύει, αλλά δυστυχώς δεν πρόλαβε να γράψει για το chatGPT. Πόσο λάθος είναι όμως αυτά που λέει; Η βαθιά μάθηση δεν λειτουργεί με ποσοτικά μετρήσιμα κριτήρια, ένα chatbot μπορεί όμως να τα παράξει. Αν του ζητήσουμε να παράξει ένα μάθημα Ιστορίας θα το κάνει – και θα το κάνει καλά. Υπάρχουν αρκετά παραδείγματα παρόμοιων μεταξύ τους μαθημάτων στο διαδίκτυο για να μπορεί να καταλάβει πώς θα έπρεπε να μοιάζει μια τέτοια σειρά λέξεων και κενών. Αν του ζητήσουμε να κρίνει το μάθημά του με βάση ποσοτικά κριτήρια θα το κάνει, αλλά εκ των υστέρων, σαν να το βλέπει εκείνη τη στιγμή. Μπορεί κάλλιστα να κρίνει επίσης ένα μάθημα Ιστορίας που παρήγαγε ένας άνθρωπος και να προτείνει τη βελτίωσή του, όπως και να το συγκρίνει με άλλα. Μπορεί βέβαια το μάθημα που θα παράξει εκ νέου να παρουσιάζει το ολοκαύτωμα ως ψέμα, ή την πτώση των Δίδυμων Πύργων ως επιχείρηση του FBI<sup>71</sup>.

Ανεξάρτητα από το αν η πληροφορία που εισέρχεται στα μοντέλα αυτά είναι ιεραρχημένη ή έχει υποστεί κάποια επιμέλεια, είναι σίγουρο πως οι εταιρίες που τα συντηρούν επιμελούνται με διάφορους τρόπους την πληροφορία που παράγουν. Συγκεκριμένα, αυτή τη στιγμή λειτουργούν ένας μεγάλος αριθμός εταιριών και υπηρεσιών (η σημαντικότερη μάλλον είναι το *Mechanical Turk* (!) της Amazon<sup>72</sup>) που απασχολούν σε μεγάλο ποσοστό άτομα με τριτοβάθμια και παραπάνω εκπαίδευση σε αναπτυσσόμενες χώρες σε ρόλο επεξεργαστών πληροφορίας. Συγκεκριμένα, η τεχνική RLHF (Reinforcement Learning from Human Feedback – ένας απόλυτα γενικός όρος που αντιστοιχεί περίπου μόνο σε μια συγκεκριμένη διαδικασία) απαιτεί την παραγωγή απαντήσεων και ιεράρχηση στοιχείων από ανθρώπους σε ελεγχόμενο περιβάλλον, ώστε να «ευθυγραμμιστεί» (η διαδικασία είναι γνωστή ως alignment) το μοντέλο με τις προσδοκίες των κατασκευαστών του<sup>73</sup>.

<sup>70</sup>David Graeber, *Bullshit Jobs: Μια Θεωρία*, 296–97.

<sup>71</sup>Rudin, ‘Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead’; Ras et al., ‘Explainable Deep Learning’.

<sup>72</sup>‘Amazon Mechanical Turk’, τελευταία πρόσβαση 29 Σεπτεμβρίου 2024, <https://www.mturk.com/>.

<sup>73</sup>‘What Is AI Alignment?’, IBM Research, 9 February 2021, <https://research.ibm.com/blog/what-is-alignment-ai>.

Είναι σημαντικό να ξεκαθαρίσουμε ότι η διαδικασία αυτή δεν έχει να κάνει με τη λειτουργία ενός LLM με την αυστηρή έννοια. Είναι όμως ζητούμενο να αποφεύγονται απαντήσεις που είναι ψευδείς, παράνομες, παράξενες, ανήθικες και λοιπά, παρόλο που μπορεί να είναι οι «σωστές» με βάση τα δεδομένα της εκμάθησης. Είναι, για παράδειγμα, αναμενόμενο, ένα μοντέλο που έχει «διαβάσει το ίντερνετ» να παράγει συχνά πορνογραφικό υλικό, ρατσιστικές ή υβριστικές απαντήσεις, και παραπληροφόρηση, πράγμα που θα έβλαπτε τις εταιρίες που το συντηρούν, παρόλο που θα ήταν μια πιο ακριβής αναπαραγωγή της ιντερνετικής γνώσης και επικοινωνίας.

Στην πρότερη μορφή της TN, αυτή των λεγόμενων «συστημάτων ειδημόνων», τη θέση του «ευθυγραμμιστή» έπαιρνε μια αυθεντία (άτομο ή ομάδα) του πεδίου στο οποίο αναφέρονταν, η οποία έδινε στους κατασκευαστές τις κατευθυντήριες που χρειαζόνταν ώστε να αποτυπώσουν τη σκέψη της σε έναν αλγόριθμο. Αυτό γινόταν όμως a priori, τις περισσότερες φορές χωρίς να παρεμβάλλεται καν κάποιο νευρωνικό δίκτυο<sup>74</sup>. Το εύρος της απεύθυνσης των συστημάτων που «διαβάζουν το ίντερνετ» οδηγεί σε μια αντιστροφή της διαδικασίας, όπου οι κατευθυντήριες ενός ειδήμονα περιορίζουν το μοντέλο εκ των υστέρων, καθώς μεταξύ του ωκεανού πληροφορίας που κονιορτοποιούν παρεμβάλλεται το black box του νευρωνικού δικτύου. Καθώς και τα δεδομένα, αλλά και η ίδια η μηχανή που προκύπτει είναι κατ' ουσίαν αγνώστου περιεχομένου, πρέπει να συνταχθεί ένας στρατός από αυθεντίες σε όσο το δυνατόν περισσότερα πεδία, οι οποίες, αντί να προσφέρουν αφαιρετικές κατανοήσεις της σκέψης και της δράσης τους, επεξεργάζονται συγκεκριμένες, ειδικές προτάσεις (ή εικόνες) τη μια μετά την άλλη. Είναι φανερό ότι οι άνθρωποι αυτοί δεν λειτουργούν ουσιαστικά σαν αυθεντίες ή ακόμα και ειδήμονες, ενώ είναι πιθανό και να μην είναι<sup>75</sup>. Πρέπει ακόμη να αναλογιστούμε πόσες από τις κατευθυντήριες γραμμές για την κρίση τους έρχονται από την ίδια την εταιρία, με πιο απλό παράδειγμα τις πολιτικές απόψεις που επιτρέπεται να εκφράζει το μοντέλο. Αν σε αυτή την εξίσωση βάλουμε και τις κακές συνθήκες στις οποίες συνήθως γίνεται η εργασία της ευθυγράμμισης, μπορούμε να πούμε με σιγουριά ότι τα LLM που έχουν υποστεί αυτού του είδους την

---

<sup>74</sup>Stuart Russel and Peter Norvig, *Τεχνητή Νοημοσύνη: Μια Σύγχρονη Προσέγγιση*, 38–40.

<sup>75</sup>AI Labour Disclosure Initiative: Recognizing the Social Cost of Human Labour behind Automation | International Labour Organization', 19 Μαρτίου 2024, <https://www.ilo.org/meetings-and-events/ai-labour-disclosure-initiative-recognizing-social-cost-human-labour-behind>.

επεξεργασία όχι απλά δεν μπορούμε να τα δούμε ως «συστήματα ειδημόνων» αλλά και μάλλον θα είχαν περισσότερο ενδιαφέρον στην αρχική τους μορφή<sup>76</sup>.

Η εργασία της εκ των υστέρων ευθυγράμμισης είναι ανάλογη της «τεράστιας ανθρώπινης προσπάθειας» που ο Γκρέμπερ περιγράφει να απαιτείται εκ των προτέρων – σε συνδυασμό με την εργασία που απαιτείται για τον χαρακτηρισμό των δεδομένων που παίρνει τη θέση της. Έτσι, η χρήση αυτών των συστημάτων δεν εξαλείφει την ανάγκη για ειδήμονες, αλλά μάλλον την εντείνει. Είναι όμως ένας τρόπος να αυτοματοποιηθεί η εργασία τους στο επίπεδο που χάνει αρκετή από την ουσία της. Αν η ανάγκη για ειδήμονες εστιάζεται στις αλληπάληλες επιφανειακές επιλογές που απαιτεί η ευθυγράμμιση ή στην επαλήθευση των εικασιών ενός συστήματος, τότε η εργασία των ειδημόνων θα γίνει σίγουρα πιο βαρετή. Ταυτόχρονα, ένας φορέας που θα θέλει να συμβουλευτεί έναν ειδήμονα θα έχει την επιλογή να χρησιμοποιήσει το chatbot. Ακόμα και αν αναγνωριστεί ότι απαιτείται μια ανθρώπινη επαλήθευση, ο ειδήμονας που προσλαμβάνεται ισούται με όλους τους άλλους – δεν είναι αυθεντία, αλλά απλώς ένας εργαζόμενος<sup>77</sup>. Όσο για τις φουμαροδουλειές, δεν εξαλείφεται η αυτοτροφοδοτούμενη ανάγκη για υποτελείς και προϊστάμενους που περιέγραψα παραπάνω. Μπορεί ένα τέτοιο σύστημα να έχει τη δυνατότητα να αυτοματοποιήσει τα, σε πολλές περιπτώσεις αχρείαστα, γραφειοκρατικά στάδια ανταλλαγής μείλ και υπογραφής εγγράφων, αλλά οι φουμαροειδήμονες, που δυσκολευόμαστε να διαχωρίσουμε από τους αναγκαίους, θα συνεχίσουν μάλλον να υπάρχουν. Η φουμαροειδημοσύνη περιλαμβάνει άλλωστε την υποκατηγορία της «μεγαλοφυΐας» που φαινομενικά κατανοεί την αγορά στα μάτια αυτών που θέλουν να το πιστέψουν, όπως περιγράφει ο Χορκχάμερ<sup>78</sup>.

Η αυτοματοποίηση λοιπόν δεν είναι ζήτημα της βαθιάς μάθησης, αλλά ζήτημα της γενικής μας σχέσης με την αυθεντία και την γνώση ή την πληροφορία<sup>79</sup>. Είναι σίγουρο ότι η τεχνολογία αυτή σε μεγάλο βαθμό θα ενταχθεί και θα υποστηρίξει αυτήν την τάση. Η πρόσβαση που προσφέρει στη γνώση και η αυτοματοποίηση των βαρετών διεργασιών θα έχει σίγουρα και επωφελή αντίκτυπο, αλλά δεν μπορούμε να περιμένουμε ότι η οικονομία της γνώσης θα εξυγιανθεί ουσιαστικά από αυτήν την

---

<sup>76</sup>Tsvetelina Hristova, Liam Magee, and Karen Soldatic, 'The Problem of Alignment', *AI & SOCIETY*, 7 Αυγούστου 2024, <https://doi.org/10.1007/s00146-024-02039-2>.

<sup>77</sup>'AI Labour Disclosure Initiative'.

<sup>78</sup>Max Horkheimer, Erich Fromm, και Herbert Marcuse, *Αυθεντία Και Οικογένεια*, 62–63.

<sup>79</sup>Guy Debord, *Η Κοινωνία Του Θεάματος* (Αθήνα: Μεταίχμιο, 2016), 167, 175 (θέση 189 και 201 (και πολλές άλλες)).

καινοτομία προς το μοντέλο, για παράδειγμα, του Καντ, όταν η μορφή που παίρνει η δημόσια ελεύθερη συζήτηση ήδη σε μεγάλο βαθμό προσιδιάζει στην ανταλλαγή ανούσιας κωδικοποιημένης πληροφορίας. Η πιο πιθανή λοιπόν θέση που θα πάρει μια τέτοια μηχανή, θα είναι μιας περαιτέρω αυτοματοποιημένης αποδεκτής εξάρτησης σαν αυτές που έχουμε στην τσέπη μας, αυτές που διαχειριζόμαστε όταν συνομιλούμε με ένα θεσμό, αυτές που συμβουλευόμαστε όταν έχουμε κάποιο καθημερινό πρόβλημα. Υπάρχει όμως και μια πιο φιλοσοφική προσέγγιση της βαθιάς μάθησης, που θα μας επαναφέρει τελικά στο μοντέλο της μαριονέτας. Αυτή θα προσπαθήσω να καθορίσω μέσα από τα παραδείγματα παρακάτω.

## **Alphafold**

Η πιο προβεβλημένη εφαρμογή ενός συστήματος νευρωνικών δικτύων στις θετικές επιστήμες, αν όχι στην έρευνα εν γένει, έχει να κάνει με το λεγόμενο «πρόβλημα της αναδίπλωσης των πρωτεϊνών»<sup>80</sup>. Πολύ συνοπτικά, το DNA αποτελείται από τις λεγόμενες αζωτούχες βάσεις (A, T, C, G), τις οποίες μπορούμε να σκεφτούμε σαν τα «γράμματα» του γενετικού κώδικα. Κάθε τρεις από αυτές σε σειρά αντιστοιχούν σε ένα από 22 αμινοξέα. Οι πρωτεΐνες είναι αορίστου μήκους αλυσίδες από αυτά τα αμινοξέα.

Το κάθε αμινοξύ έχει κάποιες φυσικές και χημικές ιδιότητες μέσω των οποίων αλληλεπιδρά με τα υπόλοιπα αμινοξέα και το περιβάλλον του για να διαμορφώσει η πρωτεΐνη μια συνολική δομή σε τρεις διαστάσεις η οποία της προσδίδει τη λειτουργία της.

Η συνήθης παρομοίωση παρουσιάζει τις πρωτεΐνες ως μηχανές, καθώς επιτελούν συγκεκριμένες λειτουργίες μέσω της δομής τους, και το κάνουν για την πλειοψηφία των μοριακών διεργασιών της ζωής. Για να καταλάβουμε τη σημασία τους αρκεί μόνο να σκεφτούμε ότι το μόνο που κάνει το DNA είναι να μεταφράζεται σε πρωτεΐνες μέσω πρωτεϊνών, μια διαδικασία που ονομάζεται και «κεντρικό δόγμα της μοριακής βιολογίας». (Σκεπτόμενοι βέβαια της πρωτεΐνες ως μηχανές, και αναλογιζόμενοι το

---

<sup>80</sup>Ken A. Dill et al., 'The Protein Folding Problem', *Annual Review of Biophysics* 37 (9 Ιουνίου 2008): 289–316, <https://doi.org/10.1146/annurev.biophys.37.092707.153558>.



εκπληκτικά σύνθετο σύστημα που περιγράφω εδώ με τις ασύλληπτα λεπτές του ισορροπίες, η άσκηση είναι να σκεπτόμαστε ότι οι μηχανές αυτές δεν έχουν χειριστή, ούτε συγκεκριμένο σκοπό. Τα δαιδαλώδη αυτά συστήματα ισορροπιών είναι ταυτόχρονα τα στοιχεία που μας αποτελούν, μας κάνουν να σκεπτόμαστε και να πράττουμε, και ένας κόσμος στον οποίο πρέπει να αποφεύγουμε να προβάλλουμε τις έννοιες που έχουν συγκροτηθεί ιστορικά στο επίπεδο της κοινωνίας. Η προβολή, για παράδειγμα, της κοινωνικής έννοιας του ανταγωνισμού ή της μεγιστοποίησης του κέρδους απαιτεί ιδιαίτερη προσοχή, αν και πολλές φορές την βλέπουμε να γίνεται απερίσκεπτα. Έτσι, ο όρος «μηχανή» μπορεί να παραπέμπει στην θέαση μιας πρωτεΐνης ως κάτι που υπάρχει για να εξυπηρετεί την ύπαρξη του υποκειμένου ή του ζώου, παρόλο που εύκολα θα μπορούσε να αναδειχθεί και η αντίστροφη σχέση.)

Το πρόβλημα της αναδίπλωσης είναι ότι δεν έχουμε τρόπο να προβλέψουμε ποια θα είναι η τελική δομή μιας πρωτεΐνης από τα αμινοξέα που την απαρτίζουν. Οι δυνάμεις που εμπλέκονται στη διαδικασία της αναδίπλωσης είναι διάφορων ειδών. Από ισχυρές/μόνιμες χημικές αλληλεπιδράσεις – δεσμούς – έως ασθενείς, αλλά συγκεκριμένες έλξεις [affinities] που έχουν να κάνουν τόσο με το ίδιο το μόριο όσο και με το περιβάλλον του. Ακόμα και αν οι κρίκοι της αλυσίδας είναι γνωστοί και συγκεκριμένοι, το πώς θα συμπεριφερθούν είναι ανυπολόγιστα περίπλοκο. Η εξακρίβωση της δομής μιας πρωτεΐνης ενδιαφέροντος με πειραματικές διαδικασίες αποτελεί ένα σύνθετο διδακτορικό πρότζεκτ στη μοριακή βιολογία καθώς απαιτεί συνήθως δύο με τρία χρόνια αλλεπάλληλων προσπαθειών.

Η λύση του προβλήματος θεωρήθηκε εξ αρχής ένα υπολογιστικό πρόβλημα, με τις συνήθειες πρακτικές να προσπαθούν να προσομοιώσουν φυσικούς και χημικούς νόμους σε ψηφιακά περιβάλλοντα στα οποία εκτίθεται ένα μοντέλο της αλυσίδας των αμινοξέων. Το κεντρικό πρόβλημα σε αυτές τις μεθόδους είναι ότι όσο αυξάνεται η λεπτομέρεια στην οποία κάθε άτομο και αρμός αλληλεπιδρά με κάθε άλλο αυξάνεται πολλαπλάσια η υπολογιστική δύναμη που απαιτείται για να παρακολουθήσει τις δυναμικές, φτάνοντας εξαιρετικά γρήγορα σε απαγορευτικές απαιτήσεις. Μια ενδιαμέση, για παράδειγμα, αφαίρεση στο παράδειγμά μας θα ήταν να μοντελοποιηθεί το κάθε αμινοξύ σαν ένα αντικείμενο, αντί για ένα σύνολο επιμέρους ατόμων, θυσιάζοντας έτσι ευκρίνεια για να γίνει το πρόβλημα πιο διαχειρίσιμο<sup>81</sup>. Αυτός ο περιορισμός σημαίνει ότι οποιαδήποτε πρόβλεψη θα είναι

---

<sup>81</sup>Dill et al.

εγγενώς αβέβαιη, ανεξάρτητα από το αν οι γενικοί κανόνες του συστήματος είναι γνωστοί και συγκεκριμένοι, ή επίσης προσεγγιστικοί και απροσδιόριστοι, όπως είναι εδώ.

Τι σημαίνει όμως πρακτικά η λύση του εν λόγω προβλήματος για τις εφαρμογές της μοριακής βιολογίας; – αν δηλαδή έχουμε ένα τρόπο να προβλέψουμε την τρισδιάστατη δομή μιας πρωτεΐνης από τον γενετικό κώδικα που της αντιστοιχεί; Πολλές μέθοδοι για την ανακάλυψη νέων φαρμάκων χρησιμοποιούν υπάρχουσες διεργασίες του οργανισμού, προσπαθώντας να τις επηρεάσουν ώστε να γίνουν πιο αποτελεσματικές στην καταπολέμηση μιας ασθένειας. Άλλες πάλι στοχεύουν στην σύνθεση μιας δραστικής ουσίας η οποία επιτίθεται η ίδια στο παθογόνο όταν χορηγείται στη σωστή ποσότητα. Μια άλλη κατηγορία ασθενειών, όπως για παράδειγμα ο καρκίνος, προκαλούνται από τον ίδιο τον οργανισμό όταν μια λειτουργία του ξεφύγει από τη συνήθη της ισορροπία. Οι πρωτεΐνες είναι «μοριακές μηχανές» ενυπάρχουσες ήδη στον οργανισμό, που σημαίνει ότι οι διαδικασίες παραγωγής, εξισορρόπησης, διάχυσης, και απόρριψής τους είναι βασικές λειτουργίες του. Ένα φάρμακο λοιπόν, εκτός του ότι είναι πολύ πιθανό να στοχεύει να διορθώσει μια δυσλειτουργική πρωτεΐνη ή να απενεργοποιήσει μια παθογόνα, είναι πολλές φορές βολικό να είναι και το ίδιο πρωτεΐνη<sup>82</sup>. Για να έχει όμως το φάρμακο το επιθυμητό αποτέλεσμα και μόνο, είναι συνήθως αναγκαίο να κατανοηθούν οι διεργασίες που επηρεάζει, πράγμα το οποίο συνήθως απαιτεί τον χαρακτηρισμό των μηχανικών και χημικών αλληλεπιδράσεων των πρωτεϊνών που τις απαρτίζουν. Μια λύση του προβλήματος της αναδίπλωσης λοιπόν θα επέτρεπε στις ερευνήτριες να δοκιμάζουν μακράν λιγότερες πιθανές ουσίες πριν βρουν αυτή που θα έχει το επιθυμητό αποτέλεσμα, καθώς θα δοκίμαζαν πολύ περισσότερες σε προσομοιώσεις<sup>83</sup>.

Ακόμα όμως και αν η δομή μιας πρωτεΐνης είναι γνωστή, το πώς ακριβώς δρα είναι ένα ερώτημα ανάλογης περιπλοκότητας με αυτό της αναδίπλωσης. Μπορούμε συνήθως να κρίνουμε τα σημεία της που είναι τα πιο σημαντικά σε ό,τι έχει να κάνει με τη λειτουργία της, αλλά το πώς θα μπορούσαν να αλλάξουν ώστε να απενεργοποιηθεί, ή, ακόμα περισσότερο, για να γίνει πιο αποτελεσματική, είναι

---

<sup>82</sup>H.A. Daniel Lagassé et al., ‘Recent Advances in (Therapeutic Protein) Drug Development’, *F1000Research* 6 (7 Φεβρουαρίου 2017): 113, <https://doi.org/10.12688/f1000research.9970.1>.

<sup>83</sup>Patrick Bryant et al., ‘Predicting the Structure of Large Protein Complexes Using AlphaFold and Monte Carlo Tree Search’, *Nature Communications* 13, no. 1 (12 Οκτωβρίου 2022): 6028, <https://doi.org/10.1038/s41467-022-33729-4>.

προβλήματα που απαιτούν μια μεγάλη διαδικασία δοκιμών και σφαλμάτων. Κάθε όμως εξέταση της λειτουργίας μιας πρωτεΐνης απαιτεί τον καθορισμό της δομής της<sup>84</sup>.

Η λύση του προβλήματος της αναδίπλωσης είναι κάτι που περιμένουν οι βιοχημικοί και για έναν άλλο λόγο, που έχει να κάνει με την ίδια την έρευνα και τη φιλοσοφία της, αν και έχει επίσης κοντινές πρακτικές εφαρμογές. Η ολοκλήρωση της ανάλυσης του ανθρώπινου γονιδιώματος, σε συνδυασμό με το γεγονός ότι η γενετική ανάλυση είναι πλέον φτηνή και γρήγορη διαδικασία<sup>85</sup> έχουν επιτρέψει τη δημιουργία του λεγόμενου (οπτικού) πεδίου της γενομικής. Η διαφορά από την κλασσική γενετική είναι ότι ο κλάδος αυτός ασχολείται όχι με μεμονωμένα γονίδια, αλλά με γονιδιώματα, με το δίκτυο αλληλεπιδράσεων δηλαδή που συνθέτει ολόκληρη η γενετική πληροφορία ενός οργανισμού. Ο κλάδος της πρωτεομικής υπάρχει μέχρι στιγμής κυρίως σαν θεωρητική αναγκαιότητα, αν και οι τελευταίες εξελίξεις των ερευνητικών μεθόδων έχουν επιτρέψει κάποιες αναλύσεις σε αυτό το επίπεδο<sup>86</sup>. Όπως και με τη γενετική, αυτό που απαιτείται για να μπορέσει να λειτουργήσει η έρευνα με αυτόν τον τρόπο είναι μια φτηνή και γρήγορη μέθοδος να χαρακτηριστεί το σύνολο των πρωτεϊνών σε έναν οργανισμό – μια λύση, δηλαδή στο πρόβλημα της αναδίπλωσης. Το ενδιαφέρον αυτό για τα συστήματα ως πεδίο ανάλυσης είναι μια γενική τάση στις φυσικές επιστήμες, η οποία βασίζεται σε μεγάλο βαθμό στην τεχνολογική πρόοδο.

Ο διαγωνισμός CASP που άρχισε το 1994 είναι αφιερωμένος σε υπολογιστικές εφαρμογές για το πρόβλημα της αναδίπλωσης. Η πρόοδος ήταν για πολλά χρόνια σταθερή, αλλά μακριά από οποιαδήποτε γενική λύση, μέχρι που, το 2020, το AlphaFold<sup>87</sup>, της Google DeepMind, κέρδισε με πρωτοφανή διαφορά. Το AlphaFold είναι η πρώτη υπολογιστική εφαρμογή που παράγει τρισδιάστατες δομές πρωτεϊνών με σημαντική ακρίβεια. Κεντρικά στη λειτουργία του είναι η ανθρώπινη κατανόηση του προβλήματος, αλλά και η μάθηση μέσω των ανθρώπινων δεδομένων, των

---

<sup>84</sup>Στις 5/9/24, προχθές, δημοσιεύτηκε η νέα έκδοση του AlphaFold η οποία παράγει μοντέλα μορίων που είναι πιθανό να συναρμόζονται σε συγκεκριμένα σημεία πρωτεϊνών. Αν αυτό το σύστημα λειτουργεί, όπως δείχνει, αποτελεί μια εκ νέου απίστευτης σημασίας καινοτομία, αυτή τη φορά στον τομέα της λειτουργίας των πρωτεϊνών.

<sup>85</sup>Stephen A. Merrill, Anne-Marie Mazza, and National Research Council (US) Committee on Intellectual Property Rights in Genomic and Protein Research and Innovation, 'Genomics, Proteomics, and the Changing Research Environment', in *Reaping the Benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health* (National Academies Press (US), 2006), <https://www.ncbi.nlm.nih.gov/books/NBK19861/>.

<sup>86</sup>Ο.π.

<sup>87</sup>John Jumper et al., 'Highly Accurate Protein Structure Prediction with AlphaFold', *Nature* 596, no. 7873 (Αύγουστος 2021): 583–89, <https://doi.org/10.1038/s41586-021-03819-2>.

λυμένων δηλαδή δομών μέσω πειραματικών διαδικασιών. Σε αυτήν την περίπτωση, δεν υπάρχει δυνατότητα η εξάρτηση αυτή να καταργηθεί μελλοντικά.

Η λειτουργία του AlphaFold είναι βασισμένη στις ανθρώπινες κατανοήσεις του πώς κληρονομούνται τα επιμέρους τμήματα μιας πρωτεΐνης μέσα στη διαδικασία της εξέλιξης ως συνάρτηση με τη σημασία τους στη λειτουργία της<sup>88</sup>. Χωρίς να μπω σε λεπτομέρειες, ο κανόνας αυτός είναι κοινά αποδεκτός στην κοινότητα και δεν είναι η πρώτη φορά που εφαρμόζεται στο πρόβλημα της αναδίπλωσης. Οι προβλέψεις όμως του AlphaFold τον επαληθεύουν περαιτέρω, σε μέτρο ανάλογο της ακρίβειας τους. Ο κανόνας δεν έχει «διδαστεί» στο σύστημα μέσω κάποιας διαδικασίας μηχανικής μάθησης, αλλά διέπει την ίδια του τη δομή. Είναι λοιπόν δεδομένος εκ των προτέρων και το συγκεκριμένο σύστημα δεν έχει τη δυνατότητα να τον φέρει σε αμφισβήτηση ή να τον «ξεπεράσει». Θεωρητικά, ένα επόμενο σύστημα βαθιάς μάθησης θα μπορούσε να λειτουργεί χωρίς αυτόν – είμαι σίγουρος πως υπάρχουν ερευνήτριες που το εξετάζουν<sup>89</sup> - όσο όμως το συγκεκριμένο σύστημα θα παράγει όλο και πιο καλές προσεγγίσεις, τόσο πιο δεδομένος θα θεωρείται αυτός ο κανόνας για την κατανόησή μας της φύσης. Θα υπάρχει έτσι λιγότερη έρευνα που θα εξετάζει την πιθανότητα να είναι ανεπαρκής.

Ένα ερώτημα που παραμένει ανοιχτό είναι το αν το πρόβλημα της αναδίπλωσης έχει λυθεί από το AlphaFold. Τα αποτελέσματα που παρουσιάζει το σύστημα μέχρι στιγμής είναι εντυπωσιακά σε σχέση με τα προηγούμενα ανάλογα συστήματα, συνεχίζουν βέβαια να μην είναι τέλεια. Η ακρίβεια της απάντησης επηρεάζεται κυρίως από το πόσες παρόμοιες πρωτεΐνες θα καταφέρει να εντοπίσει στις βάσεις δεδομένων και επίσης μειώνεται σε αναμενόμενο βαθμό όταν μεταχειρίζεται παραδείγματα πρωτεϊνών που έχουν ασταθείς ή εξαιρετικά δυναμικές δομές. Υπάρχει όμως και μια μερίδα λαθών η οποία είναι είναι ανεξήγητη – φαινομενικά, δηλαδή, τυχαία. Η τελευταία περίπτωση είναι πιο δύσκολη, καθώς δεν ακολουθεί κανόνες που μπορούμε να συνάγουμε από τη δομή του συστήματος, αν και, εντοπίζοντας κατηγορίες παραδειγμάτων στα οποία φαίνεται να υστερεί, μπορούμε ίσως να στοχεύσουμε στη βελτίωσή του παρέχοντάς του στοχευμένα νέα δεδομένα. Η ανακρίβεια, που δεν υπάρχει προοπτική να αφανιστεί πλήρως, δημιουργεί το εύλογο ερώτημα του πόση εμπιστοσύνη είναι θεμιτό να δείξει μια ερευνήτρια στο εν

---

<sup>88</sup>Ο.π.

<sup>89</sup>Oscar Chang et al., ‘A Protein Folding Robot Driven by a Self-Taught Agent’, *Bio Systems* 201 (Μάρτιος 2021): 104315, <https://doi.org/10.1016/j.biosystems.2020.104315>.

λόγω σύστημα. Η απάντηση είναι σχετικά περίπλοκη, αλλά στη συγκεκριμένη περίπτωση οριοθετεί επαρκώς τη χρησιμότητα του εργαλείου – και καθιστά ανώφελη την ανάλυση του συστήματος μέσω του ορισμού της νοημοσύνης του Τούρινγκ, παρόλο που θα ταίριαζε στην παραπάνω περιγραφή.

Η ανακρίβεια ενός εργαλείου που χρησιμοποιείται για να δώσει κατευθυντήριες γραμμές σε μια έρευνα που θα ακολουθήσει οδηγούν σε μια απροσδιοριστία εντός συγκεκριμένου πλαισίου. Από την πρώτη εμφάνιση του AlphaFold το 2020, αρχίζουν μαζί με τη χρήση του να γράφονται άρθρα για την αξιοπιστία του, για τις εφαρμογές του, για της αδυναμίες του, αλλά και να προτείνονται καινοτόμες χρήσεις που το τροποποιούν ή το επεκτείνουν για να διευρύνουν το πεδίο του<sup>90</sup>. Ένα μεγάλο ποσοστό των άρθρων αυτών γράφονται από την κοινότητα των φυσικών επιστημόνων που χρησιμοποιούν το εργαλείο και ήταν ήδη εξειδικευμένοι και στην τεχνική νοημοσύνη εκτός από το στενότερο πεδίο τους. Κατά κοινή ομολογία, ακόμα και αν η ακρίβεια του συστήματος είναι παρόμοια με αυτή των πειραματικών τεχνικών, η χρήση του δεν μπορεί να αντικαταστήσει την διαδικασία της επαλήθευσης ή της αναπαραγωγής των αποτελεσμάτων του στο εργαστήριο. Δεν σημαίνει βέβαια αυτό ότι δεν δημιουργείται ο κίνδυνος μεμονωμένοι ερευνητές να το εμπιστευτούν υπερβολικά έτσι ώστε τα τυχόν λάθη του να οδηγήσουν σε ανυπόστατα αποτελέσματα. Αυτός ο κίνδυνος όμως υπάρχει στην επιστημονική διαδικασία έτσι κι αλλιώς, με τη δομή της να είναι εν πολλοίς φτιαγμένη για να ελέγχει τέτοιου είδους απροσεξίες ή αλαζονικές παραβλέψεις.

Το σημαντικότερο στοιχείο εδώ είναι ότι το συγκεκριμένο σύστημα δεν έχει την δυναμική να παρακάμψει την ανθρώπινη αυθεντία της γνώσης παρά μόνο να την υποβοηθήσει στην εργασία της, αλλά και να την ενδυναμώσει ως προνομιακό του χρήστη. Σε αυτό όμως το πεδίο γίνεται θέμα η μη-κατανοησιμότητα της συγκεκριμένης μηχανής. Οι προσπάθειες να αναπτυχθεί περαιτέρω<sup>91</sup> από την επιστημονική κοινότητα και, ακόμα παραπάνω, οι προσπάθειες να κατανοηθεί το

---

<sup>90</sup>Bryant et al., ‘Predicting the Structure of Large Protein Complexes Using AlphaFold and Monte Carlo Tree Search’; Richard Evans et al., ‘Protein Complex Prediction with AlphaFold-Multimer’ (bioRxiv, 10 Μαρτίου 2022), <https://doi.org/10.1101/2021.10.04.463034>; Thomas C. Terwilliger et al., ‘AlphaFold Predictions Are Valuable Hypotheses and Accelerate but Do Not Replace Experimental Structure Determination’, *Nature Methods* 21, no. 1 (Ιανουάριος 2024): 110–16, <https://doi.org/10.1038/s41592-023-02087-4>; Pietro Fontana et al., ‘Structure of Cytoplasmic Ring of Nuclear Pore Complex by Integrative Cryo-EM and AlphaFold’, *Science (New York, N.Y.)* 376, no. 6598 (10 Ιουνίου 2022): eabm9326, <https://doi.org/10.1126/science.abm9326>; ‘Using AlphaFold Predictions in Viral Research’, τελευταία πρόσβαση 29 Σεπτεμβρίου 2024, <https://www.mdpi.com/1467-3045/45/4/240>.

<sup>91</sup>Terwilliger et al., ‘AlphaFold Predictions Are Valuable Hypotheses and Accelerate but Do Not Replace Experimental Structure Determination’.

πώς παράγει τις απαντήσεις του, ώστε να προσεγγισθεί η πραγματική λύση του προβλήματος, περιορίζονται εξαιρετικά από αυτή του την ιδιομορφία. Προς τιμήν τους, οι κατασκευαστές του AlphaFold εξέδωσαν όλο τον κώδικά του, μαζί με μια λεπτομερή περιγραφή της λειτουργίας του και τη λίστα των δεδομένων της εκπαίδευσής του.<sup>92</sup>

Ορισμένες μελέτες<sup>93</sup> υποστηρίζουν ότι παρατηρώντας το AlphaFold και προσπαθώντας να αναπαράγουν τις απαντήσεις του χωρίς χρήση νευρωνικών δικτύων έχουν καταλήξει σε ορισμένους νέους κανόνες για την αναδίπλωση (με τον τρόπο των σκακιστών), η πρόβλεψη όμως της τρισδιάστατης δομής των πρωτεϊνών χωρίς τη χρήση βαθιάς μάθησης υστερεί ακόμα σημαντικά. Με την αντίληψη ότι σκοπός της επιστήμης είναι να φτάνει σε όλο και πιο αφαιρετικές κανονιστικές κατανοήσεις του κόσμου, ή τουλάχιστον αυτή που θέλει τον επιστήμονα σε μεγάλο βαθμό να κυνηγάει τη γνώση ως αυτοσκοπό και τον Λόγο να είναι ανεξάρτητος της πρακτικής εφαρμογής του, η γνώση που στηρίζεται στο AlphaFold φαίνεται σαθρή. Από τη μια, η κοινότητα συμφωνεί μάλλον ότι κάτι λείπει, ότι το πρόβλημα της αναδίπλωσης δεν λύθηκε – απλά έπαψε να είναι τόσο σημαντικό. Η παράκαμψη ενός θεμελιώδους προβλήματος όμως δεν σημαίνει ότι παύει να είναι θεμελιώδες, παρά μόνο ότι, και για την έρευνα ως αυτοσκοπό αλλά και για την εφαρμογή της, ανοίγονται μονοπάτια που προηγουμένως ήταν υπερβολικά ασύμφορα για να εξερευνηθούν.

Δεν θα ασχοληθώ με το βαθμό στον οποίο μπορούμε να πούμε ότι το AlphaFold πράττει ή σκέφτεται. Αυτό που μας απασχολεί είναι ότι το AlphaFold επιδρά σε μια οριοθετημένη περιοχή, η οποία, όσο το ίδιο παράγει καλύτερα αποτελέσματα, μπορεί να θεωρηθεί όλο και πιο συμπαγής. Αν και οι προβλέψεις του συστήματος πρέπει να ελεγχθούν, και για τη βελτίωσή του απαιτούνται νέα ανθρώπινα δεδομένα, αν το πρόβλημα θεωρηθεί λυμένο και σταματήσει να ερευνάται καθαυτό, τα λάθη που μπορεί να έχουν οι σύγχρονες προσεγγίσεις εδραιώνονται περαιτέρω. Φαίνεται έτσι ότι παρά το γεγονός ότι η χρήση της TN μπορεί να μας οδηγήσει σε σημαντικές πρακτικές εφαρμογές, όπως νέα φάρμακα, έχει τη δυνατότητα να λειτουργήσει και

---

<sup>92</sup> Δεν έχουν κάνει βέβαια το ίδιο μέχρι στιγμής για το AlphaFold 3 το οποίο δημοσιεύτηκε το Μάιο του 2024. Στο άρθρο υπάρχει η δήλωση ότι ο κώδικας δεν δημοσιεύεται μαζί με το άρθρο χωρίς κάποιο πλάνο για το πότε μπορεί να γίνει κάτι τέτοιο

<sup>93</sup> Casper A. Goverde et al., ‘De Novo Protein Design by Inversion of the AlphaFold Structure Prediction Network’, *Protein Science: A Publication of the Protein Society* 32, no. 6 (1 Ιουνίου 2023): e4653, <https://doi.org/10.1002/pro.4653>.

περιοριστικά προς τη δημιουργικότητα και την «επαναστατική» πρόοδο. Το θέμα όμως αυτό είναι εγγενές ζήτημα της επιστήμης, καθώς είναι αναγκαίο να εξαρτόμαστε από επί μέρους πεδία γνώσης ως συμπαγή. Το βασικό ζητούμενο είναι να αντιλαμβανόμαστε τη βαθιά μάθηση ως συνομιλητή ή εργαλείο και όχι ως αλάνθαστη ή αναγκαία αυθεντία. Το AlphaFold δεν έχει μέχρι στιγμής τη δυνατότητα να θεωρηθεί αλάνθαστο, αλλά η πρακτική εφαρμογή του ανοίγει στο μέλλον την πιθανότητα να θεωρηθεί αναγκαίο.

## Σκακιστικές Μηχανές

Σε αυτό το κεφάλαιο, θα παρουσιάσω τα πιο σημαντικά ορόσημα στην πρόσφατη ιστορία των μηχανών που παίζουν παιχνίδια στρατηγικής. Συγκεκριμένα, θα επικεντρωθώ στις σημαντικές καινοτομίες που έχουν σημειωθεί από το 1997 – χρονιά ορόσημο σε αυτόν τον τομέα έρευνας καθώς τότε σημειώνεται η πρώτη νίκη ενός αλγοριθμικού συστήματος έναντι του τότε παγκόσμιου πρωταθλητή στο σκάκι Γκάρικ Κασπάροφ<sup>94</sup>. Επέλεξα το σκάκι ως βασικό μου παράδειγμα καθώς η μακρόχρονη ιστορία του τού έχει προσδώσει μια ιδιαίτερη θέση στην κουλτούρα ως κριτήριο ευφυΐας, σε επίπεδο να θεωρηθεί ακόμα και ένα από τα πεδία μάχης μεταξύ Σοβιετικής ένωσης και Ηνωμένων Πολιτειών κατά τη διάρκεια του ψυχρού πολέμου. Καθώς όμως η μηχανική υπεροχή στο σκάκι επιτεύχθηκε σχετικά νωρίς, και θέλω να ασχοληθώ με έναν τύπο συστημάτων που εμφανίστηκε αργότερα, θα επεκταθώ και στο παιχνίδι γκο (GO), που μπορεί να θεωρηθεί αντίστοιχο του σκακιού στην ανατολική Ασία, ως το αρχαιότερο και πιο διαδεδομένο παιχνίδι στρατηγικής.

Σκοπός μου είναι να αναδείξω το πώς οι πρόσφατες καινοτομίες σε αυτόν τον τομέα μας δείχνουν ότι σε ένα οριοθετημένο περιβάλλον με αυστηρούς και συγκεκριμένους κανόνες μια μηχανή μπορεί να επιδείξει πλήρως υπεράνθρωπες ικανότητες. Επιπλέον, θέλω να επεκταθώ στο πώς, όταν ισχύουν τα παραπάνω, η βέλτιστη συμπεριφορά ενός ανθρώπου φαίνεται να είναι να μιμηθεί, ή τουλάχιστον να μάθει από, τη μηχανή. Όταν όμως δεν είναι σε θέση να σκεφτεί με τον ίδιο τρόπο,

---

<sup>94</sup>Andrew Lea, ‘Computer Chess: A Historical Perspective’, *ITNOW* 66, no. 3 (1 Σεπτεμβρίου 2024): 50–52, <https://doi.org/10.1093/itnow/bwae098>.

να «καταλάβει τους κανόνες της συμπεριφοράς [της μηχανής]», τότε σε κάποιο βαθμό θα την αντιλαμβάνεται πράγματι ως εμπρόθετο υποκείμενο.

Το 1997, το σύστημα Deep Blue της IBM νικά τον παγκόσμιο πρωταθλητή στο σκάκι χωρίς τη χρήση νευρωνικών δικτύων. Όπως αναλύεται όμως στο άρθρο *Garry Kasparov is a Cyborg* του Τζον Χάρτμαν<sup>95</sup> (*Philosophy Looks at Chess*, 2008), οι σκακιστικές μηχανές προϋπήρχαν. Ακόμα και αν δεν είχαν θεωρηθεί υπεράνθρωπες, ήταν ένα διαδεδομένο εργαλείο για την ανάλυση παιχνιδιών και για να δοκιμάζουν οι παίκτες τις στρατηγικές τους, ώστε να προετοιμαστούν για επερχόμενα παιχνίδια. Όπως εξηγεί ο Χάρτμαν, ο Κασπάροφ δεν βίωσε την ήττα του τραυματικά, παρά τη φιλολογία περί κοσμοϊστορικής σημασίας ήττας του ανθρώπου έναντι της μηχανής. Αντίθετα, άρχισε να χρησιμοποιεί τις μηχανές περισσότερο, λέγοντας μάλιστα ότι πλέον αποτελούν «πολύτιμα μέλη της ομάδας του».

Ο Χάρτμαν, που γράφει πριν από τις καινοτομίες της βαθιάς μάθησης, εντοπίζει ένα κεντρικό ζήτημα: Αν ένας παίκτης χρησιμοποιήσει μια μηχανή στην προετοιμασία του, και αυτή του προτείνει κινήσεις που εντέλει θα εντάξει στη στρατηγική του όταν θα αντιμετωπίσει έναν ανθρώπινο αντίπαλο, τότε ποιανού πράξη, και ποιανού πρόθεση είναι αυτή που νικάει το παιχνίδι; Καθώς η μηχανή δεν μπορεί να προβλέψει όλο το παιχνίδι εναντίον του συγκεκριμένου αντιπάλου, παρά μόνο ίσως να εντοπίσει τις στρατηγικές που τείνει αυτός να ευνοεί, η νίκη που προκύπτει από την προετοιμασία αυτή πρέπει να θεωρηθεί ως πράξη και του ανθρώπινου παίκτη και της σκακιστικής μηχανής.

Μέσω της ανάλυσης του Γουότσον<sup>96</sup>, ο Χάρτμαν περιγράφει την αλλαγή στο παιχνίδι που προκύπτει από αυτή τη σύμπραξη, ή ακόμα και συναρμογή, ανθρώπου και μηχανής, ως μια κίνηση προς την ανεξαρτητοποίηση από τους παραδοσιακούς κανόνες. Οι καινοτομίες στο αγωνιστικό παιχνίδι, ήδη από εκείνη την εποχή, δεν προέρχονται από νέες εφαρμογές των στρατηγικών που περιγράφονται στα σκακιστικά βιβλία, ενώ επίσης δεν καταγράφονται σε νέα, τουλάχιστον όχι με το ρυθμό που παράγονται. Προέρχονται αντίθετα από τις δοκιμές των παικτών ενώ παίζουν εναντίον της μηχανής, και, ανεξάρτητα σχεδόν από το αν πρωτοπαίχτηκαν

---

<sup>95</sup>J. Hartmann, 'Garry Kasparov Is a Cyborg, or What ChessBase Teaches Us about Technology', in *Philosophy Looks at Chess*, ed. Benjamin Hale (Open Court Press, 2008), 39--64.

<sup>96</sup>J. Watson, *Secrets of Modern Chess Strategy: Advances Since Nimzowitsch* (Gambit, 1999; J. Watson, *Chess Strategy in Action* (Gambit, 2003).



από τους ίδιους ή τη μηχανή, δεν συγκειμενοποιούνται σε σχέση με τη γενική κατανόηση του παιχνιδιού στον ίδιο βαθμό που αυτό γινόταν παλαιότερα.

Ο Γκάρυ Κασπάροφ είναι μια παραδειγματική περίπτωση ειδήμονα του οποίου η εξαιρετική ικανότητα μεταφράζεται σε ετοιμότητα να αποδεχτεί και να αφομοιώσει τις νέες εξελίξεις, χωρίς όμως να θεωρήσει τις μηχανές *αυθεντίες*. Αντίθετα, το ένστικτο άλλων, πιο αρχάριων παικτών, είναι να αναλύουν κάθε υψηλού επιπέδου παιχνίδι που βλέπουν μέσα από μια ανάλογη μηχανή, και να επαναλαμβάνουν τις ποσοτικές αναλύσεις της αντί να προσπαθούν να τις καταλάβουν και να τις εξηγήσουν. Με αυτόν τον τρόπο, λέει ο Χάρτμαν, αντιλαμβάνονται τους εαυτούς τους ως *ειδήμονες*, αλλά στην πραγματικότητα απλώς εκφυλλίζουν τη δημόσια συζήτηση που είναι εξαιρετικά πολύτιμη για την πρόοδο στο παιχνίδι.

Οι εφαρμογές που υπάρχουν όμως την εποχή που γράφει ο Χάρτμαν δεν χρησιμοποιούν βαθιά μάθηση. Η συμπερίληψή της έρχεται αργότερα, όταν η εταιρία DeepMind καταφέρνει το 2015 να κατασκευάσει μια μηχανή που καταφέρνει να νικήσει τον παγκόσμιο πρωταθλητή στο παιχνίδι γκο (Go).

Το ταμπλό του γκο έχει διαστάσεις 19x19, μακράν δηλαδή περισσότερες θέσεις από τις 64 (8x8) του σκακιού. Τα πούλια είναι μόνο ενός είδους: μικροί παχείς δίσκοι που, ένας το γύρω, καταλαμβάνουν μια από τις προαναφερθείσες θέσεις στο ταμπλό, άσπρα και μαύρα για τις δύο παίκτριες αντιστοίχως. Αν μια περιοχή περικλείεται πλήρως από πούλια ενός χρώματος, τότε θεωρείται ολόκληρη ως ελεγχόμενη από τον αντίστοιχο παίκτη, και τα αντίπαλα πούλια στο εσωτερικό της αφαιρούνται. Το χρώμα που καταλαμβάνει περισσότερο από το ταμπλό όταν αυτό γεμίσει νικά. Δεν υπάρχουν ισοπαλίες. Παρόλο που υπάρχει μόνο ένας δυνατός τύπος κίνησης, το γεγονός ότι αυτή μπορεί να γίνει οπουδήποτε οδηγεί σε αστρονομικό πολλαπλασιασμό των υπολογισμών που καλείται να κάνει ο παίκτης ώστε να μπορέσει να προγραμματίσει ένα τέχνασμα σε βάθος λίγων κινήσεων, πόσο μάλλον να προβλέψει μια νικητήρια τακτική συνολικά.

Καθώς δεν έχω εδώ το χώρο να εντάξω τις διαφορές στις αντιλήψεις γύρω από τα δύο παιχνίδια σε μια ευρύτερη πολιτισμική ανάλυση, θα αναφέρω μόνο ότι η αντίδραση του παγκόσμιου πρωταθλητή Λη Σεντόλ ήταν πολύ διαφορετική από αυτή του Κασπάροφ κατά τον Χάρτμαν: Η ήττα του μοιάζει να τον απογοήτευσε βαθιά, παρά τη

συγκίνηση που περιγράφει να νιώθει για τις ευρηματικές κινήσεις της μηχανής, και κατέληξε να αποσύρεται από το αγωνιστικό παιχνίδι λίγα χρόνια αργότερα. Η δυνατότητα των μηχανών να καταρρίπτουν την εμπιστοσύνη σε στρατηγικές, κάποιες από τις οποίες είναι χιλιάδων ετών, φαίνεται επίσης να μην αντιμετωπίζεται με την ίδια όρεξη που δημιούργησε σε, ομολογουμένως συγκεκριμένους, σκακιστές. Είναι όμως και μια πολύ πιο πρόσφατη εξέλιξη, καθώς τα συστήματα πριν το AlphaGo δεν ξεπερνούσαν το επίπεδο παιχνιδιού ενός εκπαιδευμένου ερασιτέχνη<sup>97</sup>.

Μία σκακιστική μηχανή συμπεριλαμβάνει δύο λειτουργίες: Τη δυνατότητα να κρίνει ποιος από τους δύο παίκτες είναι σε θέση υπεροχής και κατά πόσο, και να εκφράζει τη διαφορά τους ποσοτικά, και τη δυνατότητα να επιλέγει την επόμενη της κίνηση, με βάση έναν υπολογισμό των πιθανών επόμενων κινήσεων του αντιπάλου. Το AlphaGo ξεκινάει με δεδομένους του κανόνες του παιχνιδιού, και εκπαιδεύεται αρχικά προσπαθώντας να μιμηθεί τη συμπεριφορά των δύο παικτών σε ένα σύνολο ανθρώπινων παιχνιδιών υψηλού επιπέδου. Για το υπόλοιπο της εκπαίδευσής του, παίζει έναν αστρονομικό αριθμό παιχνιδιών με τον εαυτό του. Τα παιχνίδια αυτής της περιπλοκότητας επιβάλλουν έναν περιορισμό: Δεν είναι δυνατό να υπολογιστούν όλα τα πιθανά παιχνίδια που μπορεί να προκύψουν από την κάθε κίνηση ολόκληρα. Η υπαρκτή υπολογιστική ισχύς δεν επιτρέπει έναν τόσο μεγάλο αριθμό υπολογισμών. Η διαφορά μεταξύ των συστημάτων βαθιάς μάθησης και των προκατόχων τους είναι ότι αυτή η διαδικασία αγνόησης ενός υποσυνόλου των δυνατών κινήσεων βασιζόταν στα παλαιότερα συστήματα σε κάποια μαθηματική αφαίρεση των ανθρώπινων στρατηγικών, παρόλο που, όπως είδαμε, μπορούσαν να τις ξεπεράσουν. Συστήματα μηχανικής μάθησης όπως το AlphaGo δεν μαθαίνουν αυτούς του κανόνες καθόλου, παρά μόνο έμμεσα μέσω της ανθρώπινης συμπεριφοράς πάνω στην οποία εκπαιδεύονται<sup>98</sup>

Το AlphaZero<sup>99</sup> είναι ο απόγονος του AlphaGo, ο οποίος καταφέρνει να επηρεάσει το αγωνιστικό σκάκι εκ νέου. Με δεδομένους μόνο τους κανόνες του παιχνιδιού, το AlphaZero καταφέρνει να παίξει 3 διαφορετικά παιχνίδια – σκάκι, σόγκι («ιαπωνικό σκάκι»), και γκο – σε επίπεδο ανώτερο των ανθρώπινων πρωταθλητών, αλλά και των

---

<sup>97</sup>Hajin Lee, 'Impact of Go AI on the Professional Go World', *Medium* (blog), 30 Αυγούστου 2020, <https://hajinlee.medium.com/impact-of-go-ai-on-the-professional-go-world-f14cf201c7c2>.

<sup>98</sup>David Silver et al., 'Mastering the Game of Go with Deep Neural Networks and Tree Search', *Nature* 529, no. 7587 (1 Ιανουαρίου 2016): 484–89, <https://doi.org/10.1038/nature16961>.

<sup>99</sup>Ο.π..

προηγούμενων καλύτερων υπολογιστικών μεθόδων. Το μοντέλο μαθαίνει ταχύτατα, αποκλειστικά μέσω παιχνιδιών με τον εαυτό του. Με αυτόν τον τρόπο έχουμε μια μηχανή που είναι πραγματικά ανεξάρτητη από την ανθρώπινη στρατηγική βιβλιογραφία. Οι υψηλού επιπέδου παίκτες αισθάνονται ότι το παιχνίδι ενός τέτοιου συστήματος όντως ξεφεύγει ακόμα περισσότερο από τους κανόνες της στρατηγικής, και σπρώχνει το παιχνίδι ακόμα περισσότερο προς την «ανεξαρτησία από κανόνες» που συζητήσαμε<sup>100</sup>. Καθώς όμως και οι προηγούμενες μηχανές είχαν αυτή τη δυνατότητα, τέτοιου είδους μικρές ποιοτικές διαφορές στο παιχνίδι είναι δύσκολο να παρατηρηθούν χωρίς αφενός εξαιρετική σκακιστική ικανότητα, και αφετέρου εμπειρία στο μηχανικό παιχνίδι – δύο χαρακτηριστικά αποκλειστικά στους ανθρώπινους ειδήμονες.

Ακόμα όμως και αν αγνοήσουμε τις πιθανές διαφορές στο παιχνίδι που προκύπτει, φιλοσοφικά τα συστήματα αυτά έχουν μια διαφορά που αξίζει να συζητηθεί: Παρόλο που οι παίκτες, στην καλύτερη περίπτωση, ήταν και παραμένουν *σάιμποργκ*, καθώς το παιχνίδι τους είναι μη διαχωρίσιμα ανθρώπινο και μηχανικό, το παιχνίδι των μηχανών, που είχε τα ίδια χαρακτηριστικά, μπορεί να είναι πλέον *αμιγώς μηχανικό*. Θα μπορούσαν έτσι δύο τέτοιες μηχανές να παίζουν μόνες τους επ' άπειρον, δημιουργώντας σταδιακά ένα υπέρτατο, ασύλληπτο σκάκι; Όχι. Εκτός βέβαια αν κομμάτι της λειτουργίας τους ήταν να βελτιστοποιούν και την παραγωγή ενέργειας και τις καινοτομίες στα ηλεκτρονικά εξαρτήματα που απαιτεί η πρόοδος στο παιχνίδι τους. Υπάρχει, φυσικά, και μια άλλη απάντηση σε αυτό το ερώτημα, η οποία συνοψίζεται ως: Και τι μας ενδιαφέρει;

Υπάρχουν προφανώς άλλοι λόγοι για τους οποίους το σκάκι είναι ενδιαφέρον οι οποίοι δεν έχουν να κάνουν με την σχετική ικανότητα μιας μηχανής. Ακόμα και η ντάμα, που έχει πια λυθεί αυστηρά<sup>101</sup>, συνεχίζει να παίζεται σε όλο τον κόσμο. Αυτό που ίσως φαίνεται εκ πρώτης λίγο παράξενο είναι ότι τα πρωταθλήματα σκακιού μεταξύ ανθρώπων επίσης δεν έχουν πάψει. Διαχωρίσαμε προηγουμένως τους παίκτες που αντιλαμβάνονται τη μηχανή σαν αυθεντία από αυτούς που την εντάσσουν σαν συνεργάτη στη δική τους προπόνηση. Είναι όμως σημαντικό να διαχωρίσουμε και το σκάκι ή το γκο, τα παιχνίδια στρατηγικής, από το γενικότερο πεδίο της γνώσης.

---

<sup>100</sup> Garry Chess On the AlphaZero Era, 2021, <https://www.youtube.com/watch?v=FTxoV8DC48c>.

<sup>101</sup> Jonathan Schaeffer et al., 'Checkers Is Solved', *Science* 317 (1 Οκτωβρίου 2007): 1518–22, <https://doi.org/10.1126/science.1144079>.

Αν η πρόοδος στο σκάκι κινείται προς μια «ανεξαρτησία από τους κανόνες», τότε αν θέλουμε να γενικεύσουμε αυτήν την επιρροή της ΤΝ στα πεδία που εφαρμόζεται πρέπει να σκεφτούμε επίσης το εξής: Σε περίπτωση που η μηχανή αποφάσιζε να ανεξαρτητοποιηθεί από τους κανόνες όχι της στρατηγικής, αλλά του σκακιού, τότε θα παρήγε απλά άχρηστη πληροφορία. Υπάρχει λοιπόν ένα όριο, το οποίο έχει να κάνει με την αναγνώριση ενός καθορισμένου πεδίου όπου η μηχανή αυτή είναι υπεράνθρωπη ή ανεξάρτητη του ανθρώπου. Το γεγονός ότι το AlphaZero μπορεί να παίζει περισσότερο από ένα παιχνίδι είναι μια πρόοδος σε αυτό το ζήτημα, αλλά το τελευταίο σύστημα παιχνιδιών της ίδιας εταιρίας πάει ένα βήμα παραπέρα. Το MuZero<sup>102</sup> δεν χρειάζεται καν να έχει δεδομένους τους κανόνες του παιχνιδιού που παίζει – μπορεί να τους βρει από μόνο του, παίζοντας με τον εαυτό του. Οι κανόνες που ψάχνει όμως παραμένουν ανθρωπογενείς και συγκεκριμένοι. Δεν θα μπω σε λεπτομέρειες, γιατί η προσέγγιση διαφέρει για κάθε παιχνίδι στο οποίο εφαρμόζεται, αλλά το ουσιαστικό σημείο παραπάνω δεν αλλάζει: Σε περιβάλλοντα που ορίζονται μέσω συγκεκριμένων κανόνων η ανεξαρτησία από κανόνες είναι οριοθετημένη.

Ακόμα και οι πιο πεπειραμένοι και ικανοί σκακιστές πρέπει σε ένα βαθμό να δουν τη μηχανή ως εμπρόθετο υποκείμενο. Αυτό συμβαίνει με τον πιο εμφανή τρόπο όταν μια μηχανή παίζει μια κίνηση η σκοπιμότητα της οποίας δεν είναι κατανοητή εξ αρχής, αλλά οδηγεί σε μια ευνοϊκή θέση σε βάθος χρόνου. Αναλύοντας το παιχνίδι προς τα πίσω, ο ανθρώπινος παρατηρητής κατανοεί την αρχική μη-αναμενόμενη κίνηση με βάση το πού είχε προβλέψει η μηχανή ότι θα οδηγήσει. Η ακατανόητη λοιπόν κίνηση πρέπει να αντιμετωπιστεί ως εμπρόθετη για να μπορέσει ο άνθρωπος να την αφομοιώσει εντάσσοντάς τη στη στρατηγική του. Στην καλή μάλλον περίπτωση, όπως αυτή του Κασπάροφ όπως παρουσιάζεται από τον Χάρτμαν, ο σκακιστής αναγνωρίζει τη μηχανή ως έναν άξιο συνομιλητή, χωρίς όμως να την εμπιστεύεται πλήρως. Η ανισορροπία που εισάγει η βαθιά μάθηση – το ότι υπάρχει αμιγώς μηχανικό παιχνίδι αλλά όχι αμιγώς ανθρώπινο – δεν αλλάζει αυτήν την παραδοχή, καθώς για να έχει αξία το μηχανικό παιχνίδι πρέπει να παραμένει μέσα στους ανθρώπινους ορισμούς.

Το ενδιαφέρον για το θέμα μας είναι ότι τα συστήματα αυτά μπορούν με μια έννοια να εξηγήσουν τις επιλογές τους – μπορούν να παράγουν όλες τις θέσεις τις οποίες

---

<sup>102</sup>Julian Schrittwieser et al., ‘Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model’, *Nature* 588, no. 7839 (24 Δεκεμβρίου 2020): 604–9, <https://doi.org/10.1038/s41586-020-03051-4>.

εξετάσαν και πού θεωρούν ότι οδηγούν<sup>103</sup>. Υπό αυτήν την έννοια, μπορούν να παράξουν την πλήρη δικαιολόγηση αυτού που επιλέγουν, και τίποτα παραπάνω. Το σύστημα άλλωστε δεν «σκέφτεται» με βάση σύμβολα και στρατηγικές, παρά μόνο με σειρές κινήσεων. Η αδυναμία κατανόησης έρχεται στο πώς επιλέγουν να αγνοήσουν κάποιες από τις πιθανές κατευθύνσεις του παιχνιδιού, το οποίο δεν θα μπορούσε ένας άνθρωπος, ακόμα και ειδήμων, να εξετάσει πλήρως, όπως και στο ότι η «εξήγηση» που αναφέραμε είναι στη μορφή καθαρής πληροφορίας. Ο Χάρτμαν μας δείχνει ότι αυτή καθαρή πληροφορία μπορεί να μετατραπεί σε ωφέλιμη και επεξεργασμένη μέσα από τη δημόσια συζήτηση, με δεδομένο ότι δεν θα αντιμετωπίσουμε τη μηχανή ως αυθεντία. Τι γίνεται όμως με το μη-κατανοήσιμο στοιχείο;

Σε περίπτωση που συνδυάζαμε το AlphaZero, ή το AlphaFold με ένα chatbot, θα μπορούσαμε μέσω σωστής εκπαίδευσης να παράγουμε ένα συνολικό σύστημα που να μπορεί να δικαιολογεί τις επιλογές του σε φυσική γλώσσα. Θα μπορούσαμε έτσι να συνομιλούμε μαζί του πραγματικά, και να του ζητάμε επίσης να δικαιολογήσει τις επιλογές του.

Τα συστήματα τύπου GPT, η πιο δημοφιλής τάξη συστημάτων που μιλούν, λειτουργούν επιλέγοντας μια μια της λέξεις, κρίνοντας κάθε φορά μέσω του νευρωνικού τους δικτύου ποια είναι πιο πιθανό να ακολουθεί τις προηγούμενες. Αυτό σημαίνει ότι δεν έχουν τη δυνατότητα να προγραμματίσουν ποια θα είναι η συνολική τους απάντηση και να επιλέξουν έτσι μεταξύ πιθανών προτάσεων ή κειμένων (όπως, για παράδειγμα, λειτουργούσαν τα σκακιστικά συστήματα). Η κρίση για την πιο πιθανή επόμενη λέξη γίνεται με βάση το ευρύτατο απόθεμα γραπτού κειμένου που έχουν αφομοιώσει από το ίντερνετ· προσπαθούν άρα να παράξουν ένα αποτέλεσμα που να μοιάζει με ανθρώπινη γραφή<sup>104</sup>.

Η εξήγηση που θα παράξει δεν μπορεί να αντιστοιχεί λοιπόν στη λογική του συστήματος που υποβόσκει, σε αυτό το παράδειγμα του AlphaFold ή του AlphaZero, καθώς θα εμπεριέχει την επιρροή των δειγμάτων γραφής στα οποία εκπαιδεύτηκε και την επιρροή της επί μέρους σκοπιμότητας του να παράξει μια απάντηση που να

---

<sup>103</sup>Nate Solon, 'Towards Explainable AI for Chess', Substack newsletter, *Zwischenzug* (blog), 28 Φεβρουαρίου 2021, <https://zwischenzug.substack.com/p/towards-explainable-ai-for-chess>.

<sup>104</sup>Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'.

δεχτεί ο χρήστης. Έτσι, ο λόγος που μπορεί να παράξει μια τέτοια εφαρμογή είναι στην τάξη των διηγήσεων, ή των μύθων. Των πληροφοριών που μια αυθεντία της γνώσης παράγει για να νουθετήσει τους μη γνώστες. Η εξήγηση αυτή είναι όμως πειστική, ενώ, τουλάχιστον για τώρα, μπορεί να μεταβάλλεται αν ο χρήστης τη θεωρήσει αναληθή ή της αντιταχθεί με επιχειρήματα.

Είδαμε όμως συστήματα των οποίων η χρήση ενθαρρύνει την εξάρτησή μας από αυτά, και αναγνωρίσαμε ότι υπάρχουν φορές που ο άνθρωπος θέλει να υπακούσει. Η βασική σύνδεση που θέλω να αναδείξω όμως μεταξύ της ΤΝ, και συγκεκριμένα της βαθιάς μάθησης, και της αυθεντίας είναι ότι η αυθεντία ισούται στις περισσότερες εκφάνσεις της με τα κομμάτια της κοινωνικής λειτουργίας που αποδεχόμαστε να αγνοούμε, με τον ίδιο τρόπο που θα μπορούσαμε να θεωρήσουμε το πρόβλημα της αναδίπλωσης λυμένο, ή να παπαγαλίσουμε τις αναλύσεις των σκακιστικών μηχανών. Και αυτή η επιλογή, που ορισμένες φορές επιβάλλεται, είναι αναγκαστικά πολιτική.

## **Η τεχνητή νοημοσύνη ως καταστροφείας των κανόνων και η λύση του προβλήματος**

Είδαμε δύο μορφές επιτυχημένων συστημάτων βαθιάς μάθησης και εξετάσαμε το πώς φέρεται ο χρήστης τους απέναντί τους. Με τον ίδιο τρόπο που ο Κασπάροφ είναι *σάιμποργκ*, όλοι οι βιοχημικοί από εδώ και πέρα, ή μάλλον όλη η βιοχημεία, θα είναι μερικώς μηχανική. Όλα τα φάρμακα και άρα όλοι εμείς θα κληρονομήσουμε αυτήν την ουσία και θα υβριδοποιηθούμε. Δεν περιμέναμε όμως τη βαθιά μάθηση για να κοινωνήσουμε τη μηχανή – όλα αυτά ισχύουν ήδη. Ποια είναι λοιπόν η διαφορά;

Όταν η Ντόνα Χάραγουεϊ δημιουργεί το μύθο της *σάιμποργκ* οντολογίας, οι εφευρέσεις που αναφέρει είναι της τάξης της βιντεοκάμερας παλάμης και της φορητής τηλεόρασης, των πυραύλων Κρουζ και της ατομικής βόμβας<sup>105</sup>. Παρακολουθώντας αυτό το στάδιο της προόδου, η Χάραγουεϊ λέει ήδη το 1985 ότι τα όρια στα οποία βασιζόμασταν για αιώνες έχουν καταρρεύσει: Το υποκείμενο δεν

---

<sup>105</sup> Donna Haraway, *Το Μανιφέστο Των Σάιμποργκ* (Αθήνα: Τοποβόρος, 1985), 12–15.

υφίσταται διαχωρισμένο από τη μηχανή, από τη φύση, με τον τρόπο του παρελθόντος. Και έτσι οι μύθοι της επιστροφής σε μια πρωταρχική ενότητα, σε μια αμιγή ύπαρξη, είναι ξεπερασμένοι.

Ξεκίνησα παρουσιάζοντας την ιδέα μιας νέας ενότητας, όχι για τον άνθρωπο αλλά για τον άνθρωπο ως ορθολογική μηχανή. Αλλά η υβριδική μας φύση δεν μπορεί να καταρρεύσει σε μια πλήρη εκμηχάνιση όπως δεν μπόρεσε να κάνει και το αντίθετο. Τη θέση της ενότητας παίρνει με το Διαφωτισμό η διαλεκτική πάλη – ο άνθρωπος δεν ψάχνει πια την κοινότητά του με τον κόσμο προσπαθώντας να επιστρέψει σε μια πρωταρχική αρμονία, παρά ψάχνει τη διαφορά του, ώστε να δημιουργήσει μια εκ νέου. Το όπλο του διαφωτιστή απέναντι στη φύση είναι το γνωρίζειν, η περιγραφή και η οργάνωση που θα του επιτρέψουν να κυριαρχήσει ώστε να απελευθερωθεί. Μέσω της γνώσης του, του επιτρέπεται να δημιουργήσει την κοινωνία στα μέτρα του, να απολέσει την αναγκαιότητα που επιβάλλει η φύση και να κυριαρχεί στον εαυτό του. Έτσι φτάνουμε στην ευθεία κριτική της αυθεντίας από τον Καντ – ο κάθε άνθρωπος ως αυθεντία-συγγραφέας των ιδεών του, και όλες οι ατομικές αυθεντίες μέσα σε μια κοινή συνομιλία, ώστε να αναφανεί ο κοινός Λόγος που μας διαχωρίζει από τη φύση.

Είδαμε μέσω της κριτικής του Μαρκούζε και της Άρεντ πώς μια τέτοια διδαχή δεν μεταφράζεται πραγματικά σε μια κοινωνία κυρίαρχων υποκειμένων. Πώς ο νόμος, παρόλο που προτάσσεται ως κοινά λογικός, διαιωνίζει την ανισότητα μιας άνισης αρχής. Πώς η ανισότητα αυτή οδηγεί σε μια πρωτοκαθεδρία του εαυτού της απέναντι στο Λόγο, δημιουργώντας αυτοτροφοδοτούμενες αντιφάσεις. Πώς οι επί μέρους αυθεντίες αποτραβιούνται από τον κοινό διάλογο, καθώς η πραγμάτωση της θέσης τους αντιφάσκει με τις προτεραιότητες της κοινωνίας. Είδαμε δηλαδή πώς η *αποδεκτή εξάρτηση* μετατρέπεται σε ένα σύμπλεγμα αδιαφανών πεδίων. Είναι προφανές ότι η ιδιωτικοποίηση της γνώσης που περιγράφω δεν οφείλεται μόνο στην εμπειρία του υποκειμένου, αλλά και στην απαίτηση που υποβάλλει στις εταιρίες το κίνητρο του κέρδους να κρατούν την ευρεσιτεχνία τους μυστική. Αλλά η εστίασή μου, με δεδομένο και αυτό το επίπεδο, είναι στην τάση μιας αυθεντίας που δεν είναι άμεσα πολιτική να «σαστίζει μπροστά στην περιπλοκότητα του κόσμου<sup>106</sup>» – και η σύγχρονη ειδημοσύνη, άσχετα με το πόσο προσβάσιμη είναι, παραμένει αυθεντία

---

<sup>106</sup>Hannah Arendt, 'Τι Είναι Αυθεντία', 207. Η ενσωμάτωση του μύθου της κόλασης στον χριστιανισμό ενδυνάμωσε την αυθεντία της ώστε να νικάει κάθε φορά την κοσμική εξουσία, παρόλο που ήταν αντιμετώπη με ένα απερίγραπτα σύνθετο δίκτυο ανθρώπινων υποθέσεων και σχέσεων «των οποίων η ίδια η ουσία φαίνεται νάναι η σχετικότητα».

μόνο στο βαθμό που εντάσσεται στην απρόβλεπτη κίνηση του κεφαλαίου. Και η κίνηση του κεφαλαίου μοιάζει υπεράνω οποιασδήποτε πολιτικής<sup>107</sup>.

Η τεχνητή νοημοσύνη, και συγκεκριμένα τα μοντέλα βαθιάς μάθησης αντιπροσωπεύουν μια καινοτομία στη φιλοσοφική κατανόηση της μηχανής. Η ιδέα που υποστηρίζει την αποδεκτή εξάρτηση από ένα μηχάνημα ή από έναν άνθρωπο στη μέχρι τώρα ιστορία είναι ότι κάποιος μπορεί να εξηγήσει τη λειτουργία του, όπως και το πώς θα μπορούσε ένας άνθρωπος, ή ένα σύνολο συνεργαζόμενων ανθρώπων, να επιτελέσει την ίδια διαδικασία. Για το αν υπάρχει όντως στον κόσμο αυτή τη στιγμή ένα τέτοιο άτομο που θα μπορούσε να εξηγήσει έναν υπολογιστή εξ ολοκλήρου αμφιβάλλω, αλλά στην περίπτωση της βαθιάς μάθησης δεν υπάρχει εξ ορισμού.

Όταν η βαθιά μάθηση εκπαιδεύεται μέσα από ανθρώπινα δεδομένα, όπως στην περίπτωση του AlphaFold, μπορούμε να τη δούμε ως μια αφαίρεση της συλλογικής γνώσης. Σε αυτήν της την έκφανση, η τεχνητή νοημοσύνη εμφανίζεται ως ένα ψευδουποκείμενο με προνομιακή αντίληψη της μοριακής ακατέργαστης πληροφορίας, και επιτυγχάνει σε συγκεκριμένες περιπτώσεις να παράξει συναγωγές που δεν θα μπορούσε να παράξει ένας άνθρωπος, ή το σύνολο της ανθρωπότητας με την παρούσα οργάνωσή της. Η επιστημονική κοινότητα, λόγω της δομής της, μπορεί να διαχειριστεί το νέο αυτό εργαλείο και να το οριοθετήσει σε συγκεκριμένο «ορίζοντα δυνατοτήτων». Είναι όμως επικίνδυνο να θεωρείται η διαδικασία που επιτελεί «λυμένο πρόβλημα» καθώς μια τέτοια αντίληψη θα μπορούσε να οδηγήσει αφενός σε μια διαχρονική εξάρτηση από το εργαλείο και την εταιρία που το συντηρεί, και αφετέρου στο να παραμεληθούν οι υποθέσεις που χρησιμοποιεί η επιστημονική κοινότητα και το ίδιο, που θα μπορούσαν να είναι αναληθείς. Η αυτοματοποίηση όμως σε αυτό το πεδίο επιτρέπει την ενασχόληση με ζητήματα υψηλότερου επιπέδου αφαίρεσης, όπως και οδηγεί σε νέες πρακτικές εφαρμογές.

Τα μηχανήματα που παίζουν σκάκι ανήκουν σε μια περιορισμένη κατηγορία συστημάτων βαθιάς μάθησης, στην οποία τα μοντέλα έχουν ξεπεράσει την αναγκαιότητα για ανθρώπινα δεδομένα. Μπορούν έτσι να κινούνται στο χώρο που δημιουργούν οι κανόνες του σκακιού με τρόπο πραγματικά υπεράνθρωπο. Σε αυτήν όμως την περίπτωση δεν μιλάμε ακριβώς για αυτοματοποίηση – το παίξιμο σκακιού

---

<sup>107</sup>Max Horkheimer, Erich Fromm, και Herbert Marcuse, *Αυθεντία Και Οικογένεια*, 63. «Στη σημερινή τάξη πραγμάτων αυτή η συνάρτηση μεταξύ της αξίας και των κοινωνικών αναγκών διαμεσολαβείται όχι μόνον από υπολογίσιμα ψυχικά και πολιτικά στοιχεία, αλλά και από το άθροισμα αμέτρητων μη ελέγξιμων γεγονότων»



δεν αντιστοιχεί σε κάποια υλική ανάγκη. Είδαμε τους σκακιστές ως την παραδειγματική περίπτωση μιας *σάιμποργκ* δυναμικής. Αλλά όσο η χρήση αυτόνομης ΤΝ γενικεύεται εκτός ενός καθορισμένου πεδίου γνώσης, ποια είναι η αναλογία με το περιβάλλον του σκακιού;

Υπάρχει μια εκτενής συζήτηση στη βιβλιογραφία των νέων μέσων για το παιχνίδι, ψηφιακό ή μη, τους ορισμούς του και τις δυνατότητές τους σαν αναλυτικό εργαλείο. Από αυτήν τη συζήτηση θα αντλήσω μόνο κάποιους βασικούς ορισμούς για τους σκοπούς αυτής της εργασίας, παρόλο που θεωρώ ότι είναι μια από τις πιο σημαντικές κατευθύνσεις προς τις οποίες θα μπορούσε να πάει αυτή η έρευνα. Ταυτόχρονα, υπάρχει εκτενής βιβλιογραφία και στις υπολογιστικές και στις κοινωνικές επιστήμες για τη μοντελοποίηση της κοινωνικής συμπεριφοράς μέσω της θεωρίας παιγνίων – ούτε προς αυτήν την κατεύθυνση έχω χώρο εδώ να επεκταθώ, παρότι είναι πολύ ταιριαστή για το θέμα μας.

Έχουμε ήδη την πραγμάτευση του παιχνιδιού μέσα από τον Γκρέμπερ, ως μια έκφραση της πρωταρχικής ηδονής του να επηρεάζεις το περιβάλλον σου μέσα από τη δράση<sup>108</sup>. Ο Γκάλουγουεϊ, στην έρευνά του πάνω στα βιντεοπαιχνίδια<sup>109</sup>, δίνει έναν πιο συνεκτικό ορισμό: Παιχνίδι είναι μια δραστηριότητα που ορίζεται μέσω κανόνων όπου παίκτες προσπαθούν να επιτύχουν κάποιο σκοπό<sup>110</sup>. Ο τρόπος που το προσεγγίζει ο Γκρέμπερ μου αρέσει περισσότερο – δεν είναι οι κανόνες, ούτε ο σκοπός αναγκαίοι για το παιχνίδι – είναι η διάδραση. Οι κανόνες όμως είναι αναγκαίοι για να διαχωρίσουμε το πεδίο του παιχνιδιού από αυτό γύρω του – όπως είπαμε, αν ένα μοντέλο παράξει κινήσεις που δεν ακολουθούν τους κανόνες του παιχνιδιού, τότε θα λέγαμε ότι παράγει άχρηστη πληροφορία. Ο σκοπός επίσης είναι αναγκαίος για να μπορέσουμε να ξεφύγουμε από το παιχνίδι σε περίπτωση που το θέλουμε – να υπάρχει μια κατεύθυνση και ένα τέλος. Το σκάκι είναι ένα τέτοιο περιβάλλον· η πραγματικότητα δεν είναι.

Το ζήτημα που δημιουργείται από μια ακατανόητη μηχανή είναι ότι είναι δύσκολο να διαχωρίσουμε πότε δρα ως συμπαίκτης, πότε ως αντίπαλος και πότε ως περιβάλλον. Στη βιοχημεία, μπορούμε να περιορίσουμε τη μηχανή, αλλά αυτό σημαίνει ότι οι

---

<sup>108</sup>David Graeber, *Bullshit Jobs: Μια Θεωρία*, 114.

<sup>109</sup>Alexander R. Galloway, *Gaming: Essays on Algorithmic Culture*, NED-New edition, vol. 18 (University of Minnesota Press, 2006), <https://www.jstor.org/stable/10.5749/j.ctttss5p>.

<sup>110</sup>Galloway, 18:19.

κανόνες του πεδίου στο οποίο λειτουργεί μπορεί να σταματήσουν να μας απασχολούν σαν να τους είχαμε καθορίσει. Στο σκάκι, η μηχανή αυτή μας ανεξαρτητοποιεί από τους επί μέρους κανόνες στρατηγικής, καθώς μας προτείνει βέλτιστες συμπεριφορές, αλλά μπορεί να μας αποξενώσει από τον έλεγχο του παιχνιδιού σε αφαιρετικό επίπεδο. Η ιστορία της εφαρμογής της τεχνητής νοημοσύνης σε κοινωνικές διαδικασίες είναι ήδη μακρά – στη διαχείριση ανθρώπινου δυναμικού<sup>111</sup>, στην επικοινωνία με θεσμούς<sup>112</sup>, στην ιατρική διάγνωση<sup>113</sup>, σε οπλικά συστήματα<sup>114</sup>, στο χρηματιστήριο<sup>115</sup>. Όσο όμως η τεχνητή νοημοσύνη που χρησιμοποιούμε γίνεται όλο και πιο γενική, και όσο η ορθολογικότητα της κατοχυρώνεται μέσω των αποτελεσμάτων της, κινδυνεύουμε να θεωρήσουμε ότι τα πεδία στα οποία λειτουργεί όντως διέπονται από συγκεκριμένους κανόνες. Αυτή είναι η κίνηση μέσω της οποίας η τεχνητή νοημοσύνη μπορεί να οδηγήσει σε περαιτέρω διατήρηση του υπάρχοντος: μέσω της παραγωγής συμβουλών για το πώς μοιάζει η ιδεατή συμπεριφορά μέσα στο υπάρχον. Και όταν μια τέτοια πηγή νοημοσύνης συνδυάζεται με μια διάθεση για υπακοή, δημιουργείται, όπως είδαμε μέσω της ανάλυσης του Φρομμ, ένας δαίδαλος αυτοδιατηρούμενων σχέσεων απρόσωπης αυθεντίας.

Σε μια παράλληλη κίνηση με τα παραπάνω, ο ενθουσιασμός γύρω από την ΤΝ εκφράζεται και σε μια φουμαρολογία γύρω από τη νέα τεχνολογία που κατά καιρούς συνιστά πράγματι απάτη. Αναφέρομαι και στην ρητορική γύρω από το πώς η ΤΝ είναι το απαραίτητο γκάτζετ για οποιονδήποτε «νέο μάνατζερ»<sup>116</sup>, αλλά και σε ειδήσεις όπως η πρόσφατη καταδίκη δύο επενδυτικών εταιριών για εξαπάτηση<sup>117</sup>, καθώς διαφήμιζαν ότι χρησιμοποιούσαν ΤΝ χωρίς να μπορούν να το αποδείξουν. Αυτή η

---

<sup>111</sup>Ben Eubanks, *Artificial Intelligence for HR: Use AI to Support and Develop a Successful Workforce* (Kogan Page Publishers, 2022).

<sup>112</sup>Thomas Vogl et al., 'Algorithmic Bureaucracy', in *Proceedings of the 20th Annual International Conference on Digital Government Research*, Dg.o '19 (New York, NY, USA: Association for Computing Machinery, 2019), 148–53, <https://doi.org/10.1145/3325112.3325240>.

<sup>113</sup>Hiroshi Fujita, 'AI-Based Computer-Aided Diagnosis (AI-CAD): The Latest Review to Read First', *Radiological Physics and Technology* 13, no. 1 (1 Μαρτίου 2020): 6–19, <https://doi.org/10.1007/s12194-019-00552-4>.

<sup>114</sup>Max van Kralingen, 'Use of Weapons: Should We Ban the Development of Autonomous Weapons Systems?', *The International Journal of Intelligence, Security, and Public Affairs* 18, no. 2 (3 Μαΐου 2016): 132–56, <https://doi.org/10.1080/23800992.2016.1196947>.

<sup>115</sup>Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'; Longbing Cao, 'AI in Finance: A Review', SSRN Scholarly Paper (Rochester, NY, 10 Ιουλίου 2020), <https://doi.org/10.2139/ssrn.3647625>.

<sup>116</sup>Eubanks, *Artificial Intelligence for HR*.

<sup>117</sup>'SEC.Gov | SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence', τελευταία πρόσβαση 29 Σεπτεμβρίου 2024, <https://www.sec.gov/newsroom/press-releases/2024-36>.

αυτοτροφοδοτούμενη θεαματική οικονομία που δημιουργείται λοιπόν γύρω από την ΤΝ, σε συνδυασμό με τις πραγματικές εφαρμογές της, ηθικές ή ανήθικες, έχουν τη δυνατότητα να εξυψώσουν την ΤΝ προς το πρότυπο της *μαριονέτας* – ως ένα φανταστικό πρόσωπο της διάχυτης αυθεντίας.

Όσο ο κόσμος γίνεται πιο σύνθετος και τα περιθώρια εμπρόθετης δράσης του υποκειμένου στενεύουν και παγιώνονται, τόσο η ιδέα του κάθε ατόμου ως αυθεντία του εαυτού του παράγει σαδομαζοχιστική συμπεριφορά. Κάθε υποκείμενο σαστίζει μπροστά στην περιπλοκότητα που καλείται να διαχειριστεί, καθώς ορθολογική σκέψη πρέπει να εφαρμόζεται σε κάθε επιλογή μέσα σε ένα εγγενώς αβέβαιο περιβάλλον. Έτσι, αντί η αυθεντία να θεωρηθεί ξεπερασμένη, ξαφνικά την ψάχνουμε, και τη βρίσκουμε, παντού. Η ανεξαρτησία από τους κανόνες και η καταγγελία της αυθεντίας συνυπάρχουν αρμονικά με μια πραγματοποιημένη αναγκαιότητα που ενθαρρύνει την άκριτη συμμόρφωση, συνθέτοντας ένα αδιαφανές γενικό πεδίο. Η κοινωνία διατηρεί έτσι για τον εαυτό της τη θέση ως προς το υποκείμενο της προνεωτερικής φύσης: Της φρίκης πέρα από την κατανόηση. Η αντικατάσταση (κατά Γκρέμπερ) των αξιών από την αξία επιτρέπει την κρίση οποιασδήποτε πράξης με βάση το κέρδος ή μια πραγματοποιημένη έννοια της επιβίωσης, και έτσι τα επί μέρους συστήματα πίστης και γνώσης μπορούν να συνυπάρχουν ή να συγκρούονται, εντείνοντας τον φόβο, αλλά χωρίς ιδιαίτερο γενικό αντίκτυπο.

Η ερώτηση είναι αν μπορούμε να δούμε τον ορίζοντα δυνατοτήτων μας ή το βλέμμα μας δεν φτάνει τόσο μακριά. Η συνήθης μέθοδος της νεωτερικής επιστήμης είναι η περιγραφή, η ανάλυση, η κατανόηση. Πώς λοιπόν δεχόμαστε να χρησιμοποιούμε και να αναπτύσσουμε ένα λογισμικό που δεν κατανοούμε; Η απάντηση μάλλον βρίσκεται μεταξύ της λειτουργικότητας – που μπορούμε αν θέλουμε να δούμε ως εργαλειακή ορθολογικότητα, και της απόστασης που έχουμε από το σύνθετο περιβάλλον μας ήδη. Η αδυναμία κατανόησης ενός συστήματος βαθιάς μάθησης γίνεται ζήτημα μόνο στα ανώτατα επίπεδα εξειδίκευσης και πρόσβασης σε κάθε τομέα που τη χρησιμοποιεί. Αν ο χρήστης δεν ελέγχει το ακριβές πεδίο εφαρμογής της σε εξαιρετικά υψηλό επίπεδο, τότε η διαφορά με ένα κατανοήσιμο υπολογιστικό πρόγραμμα είναι ανύπαρκτη.

Καθώς τα περίπλοκα συστήματα τεχνικής νοημοσύνης είναι κάθε άλλο παρά ουδέτερα, όπως είδαμε μέσω της διαδικασίας της ευθυγράμμισης, η μεταχείριση

ενός τέτοιου συστήματος ως μη κατανοήσιμο και αυτόνομο εγκυμονεί εξαιρετικά σημαντικούς κινδύνους, οι οποίοι όμως έρχονται να δέσουν πάνω στην υπαρκτή σχέση του υποκειμένου με τη γνώση και τη δράση. Η αποδεκτή εξάρτησή μας όμως συνολικά από ένα μη κατανοήσιμο τεχνούργημα υποδηλώνει σε κάποιο βαθμό μια παραίτηση από τις ιδέες του Διαφωτισμού (η οποία βέβαια είναι απόλυτα συμβατή με της ιδέες του νεοφιλελευθερισμού).

Αν για τον Γκάλουγουεϊ η διάδοση του υπολογιστή δημιούργησε ένα κόσμο στα μέτρα της προγραμματιστικής λογικής, τότε η ΤΝ ως ορθολογικός πράκτορας είναι το ιδανικό υποκείμενο για να κινείται σε αυτούς τους χώρους. Έχει επίσης τη δυνατότητα να τους δίνει φωνή – να επιτρέπει τη δικαιολόγηση των αντιφάσεων σε φυσική γλώσσα. Το ενδιαφέρον με τα υπολογιστικά περιβάλλοντα είναι ότι είναι επιρρεπή στο *hacking* – ότι όσο αδιαφανή και συμπαγή και αν γίνουν, μια έμπειρη χρήστρια της γλώσσας τους μπορεί να τα ωθήσει πέρα από τις δυνατότητές τους. Ο Γκάλουγουεϊ καταλήγει την εισαγωγή του *Η Αντι-Γλώσσα των Νέων Μέσων*<sup>118</sup> λέγοντας ότι η γλώσσα του υπολογιστή δεν τον ενδιαφέρει καθαυτή – οποιαδήποτε κριτική ανάλυσή της πρέπει να γίνει με πολιτικούς όρους. Κλείνει επίσης το βιβλίο του για τα βιντεοπαιχνίδια με ένα εγκώμιο στο *countergaming*<sup>119</sup>, την τάση που δημιουργείται στις κοινότητες παικτών να τροποποιούν και να δημιουργούν παιχνίδια εκ νέου. Σε αυτήν την πρακτική βλέπει τις δυνατότητες δημιουργίας νέων περιβαλλόντων που, μέσω της διάδρασης, θα αντανακλούν χειραφετητικές πρακτικές και αξίες – μια ανάποδη χρήση του λογισμικού ως ιδεολογίας.

Όσο και να υπάρχει ο πειρασμός να καταλήξω στην ανάγκη δημιουργίας ενός chatbot με χειραφετητικές καταβολές, ο *Μηχανικός Τούρκος* χρησιμοποιείται ήδη σαν όνομα για την υπηρεσία της Amazon όπου μπορεί κανείς να νοικιάσει αποκεντρωμένες ομάδες ειδημόνων για εργασίες όπως η επεξεργασία δεδομένων και η ευθυγράμμιση. Ο νάνος του διαλεκτικού υλισμού του Μπένγιαμιν<sup>120</sup> έχει λοιπόν ήδη μέσω της βαθιάς μάθησης το σύγχρονο ανάλογό του: Μια οχλαγωγία από αλλοτριωμένους ειδήμονες που ονομάζουμε (θεολογικά) προσομοίωση του Λόγου, αλλά στην πραγματικότητα το μόνο τους κοινό χαρακτηριστικό είναι η ανασφαλής μισθωτή εργασία. Πρέπει λοιπόν να αναγνωρίσουμε ότι το μέγεθος ενός συστήματος σαν το chatGPT, παρότι προσβάσιμο από παντού και πλήρως αποκεντρωμένο, είναι

---

<sup>118</sup>Galloway, *The Interface Effect*, 19 *Η Αντι-Γλώσσα των Νέων Μέσων* (Ουτοπία τεύχος 133 (2022) 17-26)

<sup>119</sup>Galloway, *Gaming*, 18:107–26.

<sup>120</sup>Walter Benjamin, 'Για Την Έννοια Της Ιστορίας', in *Walter Benjamin* (Αθήνα: Έρμα, 2023), 57.

μνημειώδες. Η ενέργεια, οι πόροι, και η ανθρώπινη εργασία που απαιτούνται για τη λειτουργία του απαιτούν την κινητοποίηση εκατομμυρίων γранаζιών στην παγκοσμιοποιημένη οικονομία για οποιαδήποτε συνομιλία του. Είναι λοιπόν εξαρχής δύσκολο να φανταστούμε ένα τέτοιο μηχάνημα εκτός των ισχυουσών συνθηκών παραγωγής.

## Δύο Προοπτικές

Δεν θα ήθελα να κρίνω τη βαθιά μάθηση ως μια εγγενώς εκμεταλλευτική ή καταπιεστική τεχνολογία. Θέλω μάλλον να την προτείνω ως μια αντι-τεχνολογία. Είδαμε ήδη πως δεν παράγεται από την κατανόηση ενός πεδίου αλλά από την παραίτηση από αυτήν την προσπάθεια ή ίσως τη βιασύνη να διαχειριστούμε το πεδίο ως κατανοητό. Είδαμε επίσης πως παράγει μια ιδιαίτερη αυτοματοποίηση των διεργασιών που αναλαμβάνει, καθώς τυπικά θα πρέπει να προσπαθούμε μονίμως να την αντικαταστήσουμε με κάτι διαφορετικό. Η παραίτηση από αυτήν την προσπάθεια, την οποία κάθε άλλη τεχνολογία θέλει να ενθαρρύνει μέσω της λειτουργικότητάς της, γίνεται εδώ απαγορευτική. Θεωρώντας έτσι τα προβλήματα που διευκολύνει λυμένα, η «περιπλοκότητα του κόσμου» μειώνεται τεχνητά, και αντί για να μας δίνει η τεχνητή νοημοσύνη νέους τρόπους να ξεπεράσουμε τις στρατηγικές που εδράζονται σε αντιφάσεις, τις καθιερώνει ως αγνοήσιμες.

Ενώ ο ασυνείδητος κολοσσός της πραγματικότητας, ο χωρίς υποκείμενο καπιταλισμός, εκτελεί τυφλά την εξόντωση, το παραλήρημα του εξεγερμένου υποκειμένου πραγματοποιεί μέσω αυτής την εκπλήρωσή του και έτσι μαζί με την τσουχτερή ψυχρότητα απέναντι στους εκπραγματισμένους, τους χρησιμοποιούμενους σαν πράγματα ανθρώπους, ακτινοβολεί και την αντεστραμμένη αγάπη, η οποία στον κόσμο των πραγμάτων έχει πάρει τη θέση της άμεσης. Η αρρώστια γίνεται σύμπτωμα της ανάρρωσης. Στην εξιδανίκευση των θυμάτων η παράνοια διακρίνει την ταπείνωση τους. Εξομοιώνεται με το τέρας της κυριαρχίας, που στη ζωντανή

αναμέτρηση δεν μπορεί να το υπερνικήσει. Με τη μορφή της φρίκης η Φαντασίωση επιχειρεί να αντισταθεί στη φρίκη.

Το απόσπασμα παραπάνω περιγράφει την σαδομαζοχιστική συμπεριφορά και το επόμενο της – για τους συγγραφείς της *Διαλεκτικής*, τον φασισμό. Η δυσκολία έγκειται στο ότι η αρρώστια είναι σύμπτωμα της ανάρρωσης αναγκαστικά σε ένα διαλεκτικό σχήμα. Αυτή είναι η ιδέα του *countergaming* – όπως και της καταστατικής θέσης του γιατρού που περιέγραψα παραπάνω. Το ζήτημα που εγείρεται, όπως ελπίζω να έχω καταφέρει να δείξω, είναι το ζήτημα του *περιβάλλοντος* και της *γλώσσας*, των στοιχείων δηλαδή που επιβάλλουν τους όρους μιας σύγκρουσης – τα στοιχεία που θεωρούνται εκτός της ως αναγκαιότητες και οι τρόποι με τους οποίους μπορεί να τελεστεί. Σε αυτήν τη σχετικά αυθαίρετη χρήση των όρων μπορούμε να εντάξουμε και τα επιχειρήματα της *Διαλεκτικής* και το αρχικό απόσπασμα του Γκάλογουεϊ: Ο τελευταίος, λέγοντας ότι οι ορισμοί και οι δυνατότητες των νέων μέσων δεν τον απασχολούν, εννοεί ότι η γλώσσα τους έχει αναχθεί σε περιβάλλον – άποψη που πλαισιώνεται και από την πραγμάτευση της αυτοματοποίησης από τον Γκρέμπερ. Οι Χορκχάιμερ και Αντόρνο χρησιμοποιούν με παρόμοιο τρόπο την εκβιομηχάνιση και την εντατικοποίηση της κοινωνικής ζωής. Το συνολικό όμως περιβάλλον που δεν θα έπρεπε να νοείται ως τέτοιο (αλλά να χρίζει αλλαγών) είναι για όλους ο καπιταλισμός, και οι αντιφάσεις που περιέχει ήδη από τα προτάγματα του Διαφωτισμού δημιουργούν τις σύγχρονες νευρώσεις, όπως η αυθεντικοκεντρική συμπεριφορά στο λυκόφως των ειδώλων.

Η αρχή του Καντ «να πράττει κανείς πάντα με άξονα τη θέλησή του, η οποία θα έχει αντικείμενο τον εαυτό της ως γενικό νομοθετικό κέντρο», είναι επίσης το μυστικό του υπερανθρώπου [του Νίτσε]. Η θέλησή του δεν είναι λιγότερο δεσποτική από την κατηγορική προσταγή. Και οι δύο αρχές στοχεύουν στην ανεξαρτησία από εξωτερικές δυνάμεις, στην απόλυτη ωριμότητα, που έχει ορισθεί ως η ουσία του διαφωτισμού. Καθώς όμως ο φόβος του ψέματος, τον οποίο ο Νίτσε ακόμη και στις πιο φωτεινές του στιγμές διέσυρε ως «δονκιχωτισμό», αντικαθιστά το νόμο με την αυτονομοθεσία και όλα γίνονται τόσο διαφανή σαν να ήταν μια μεγάλη πρόληψη που αποκαλύφθηκε, ο ίδιος ο διαφωτισμός, και μάλιστα η ίδια η αλήθεια υπό οποιαδήποτε μορφή, γίνεται είδωλο· τότε διακρίνουμε «ότι και εμείς οι σημερινοί γνώστες, εμείς οι άθεοι και αντιμεταφυσικοί, παίρνουμε τη φωτιά μας ακόμη από την πυρκαγιά που άναψε μια πίστη χιλιετιών, εκείνη η πίστη των χριστιανών, που είναι και η πίστη του Πλάτωνα, ότι ο θεός είναι η αλήθεια, ότι η αλήθεια είναι θεϊκή». Η κριτική της μεταφυσικής περιλαμβάνει λοιπόν και την επιστήμη. Η άρνηση του θεού εμπεριέχει μια άλυτη αντίφαση: αρνείται την ίδια τη γνώση. [...] Το έργο του αποκαλύπτει το μυθολογικό χαρακτήρα των αρχών στις οποίες σύμφωνα με τη

θρησκεία στηρίζεται ο πολιτισμός: των Δέκα Εντολών, της πατρικής αυθεντίας, της ιδιοκτησίας<sup>121</sup>.

Η γλώσσα του Διαφωτισμού, η *ratio* της νεωτερικότητας φέρει έτσι μέσα της το νήμα των αιώνων της κυριαρχίας, του φιλοσόφου-βασιλέα και του αστυνόμου-ιερέα. Ο Γκάλογουεϊ, όπως και η Χάραγουεϊ, εντάσσουν την νέα τεχνολογία στην γενεαλογία των μύθων της κυριαρχίας. Η ΤΝ, ως εγγενώς μη-κατανοήσιμη μηχανή όμως αποτελεί από μια άποψη την αναγνώριση των αντιφάσεων του κυριαρχικού *ratio*, ή του κόσμου του υπολογιστή.

Η ΤΝ είναι μια μηχανή που μπορεί να μάθει τη γλώσσα οποιουδήποτε περιβάλλοντος, όχι όμως και να την εξηγήσει. Παρά το γεγονός ότι, όπως ανέπτυξα παραπάνω, αυτό οδηγεί σε μια περαιτέρω παγίωση ενός περιβάλλοντος ως αναπόδραστο, αποτελεί και μια παραδοχή ότι η *ratio* έχει αποτύχει – ότι ακόμα και ανθρωπογενή περιβάλλοντα είναι υπερβολικά περίπλοκα για να αναλυθούν. Ακόμα πιο διαλεκτικά, παρόλο που αποτελεί την παραδοχή μιας αντίφασης, ακολουθεί απόλυτα τα πρότυπα του θετικιστικού ορθολογισμού που φέρνει σε ανοιχτή αμφισβήτηση: Την άκριτη κονιορτοποίηση δεδομένων, ανθρώπινων και μη, τη λειτουργικοποίηση του κόσμου, την εκμηχάνιση του ανθρώπου. Από αυτήν την άποψη, μια τεχνολογία σαν τη βαθιά μάθηση έχει πραγματικά τη δυναμική να αλλάξει θεμελιακά τον τρόπο που σκεφτόμαστε τον κόσμο. Καθώς όμως τη στιγμή που γράφω αυτήν την εργασία αυτή η τεχνολογία εξελίσσεται ραγδαία, και οι κοινωνικές συνθήκες που επηρεάζει είναι εξαιρετικά πολύπλοκες και ανομοιογενείς, θα αρκεστώ στο να εξετάσω καταληκτικά τις δύο κριτικές όψεις του ζητήματος που έχω ήδη σκιαγραφήσει, ώστε να καταλήξω σε κάποιες πιθανές κατευθύνσεις στις οποίες θα έβλεπα αυτήν την έρευνα να εξελίσσεται.

Ουσιαστικά, το συμπέρασμά μου είναι κοινό με αυτό του Γκάλογουεϊ ή της Άρεντ, όπως και τον περισσότερων κειμένων που γράφονται για την ηθική της ΤΝ: Ότι πρέπει να πασχίζουμε να βλέπουμε την τεχνολογία πολιτικά, καθώς αν συνεχίσουμε να την αφήνουμε να ξεφεύγει προς, για παράδειγμα, τη σφαίρα του μύθου, της αυτόνομης οικονομίας, ή του εργαλειακού ορθολογισμού, οι επιπτώσεις θα είναι καταστροφικές. Σε μια εποχή όμως που οι παραδοσιακοί διαλεκτικοί διαχωρισμοί φαίνεται να περιθωριοποιούνται, αξίζει ίσως να εξετάσουμε την τάση που αποζητά

---

<sup>121</sup>Theodor W Adorno και Max Horkheimer, *Διαλεκτική Του Διαφωτισμού*, 192–93.

μια νέα ενότητα, και αυτήν που προσπαθεί να θεμελιώσει μια πολλαπλότητα· και τη σχέση και των δύο με την εσωτερική λογική της βαθιάς μάθησης.

### Η Μαριονέτα και ο Πολύφημος

Για να δούμε την ΤΝ από την άποψη ενός επόμενου τεχνολογικού μύθου της κυριαρχίας μπορούμε να συνθέσουμε την ανάλυση των *Μαριονετών* παραπάνω με την τοποθέτηση της αστικής τάξης στη θέση του Οδυσσέα, στην παραβολή που παραθέτεται στη *Διαλεκτική του Διαφωτισμού*.

Δέσμιος, με δάκρυα στα μάτια, ο Οδυσσέας ακούει το τραγούδι των σειρήνων παρακαλώντας τους συντρόφους του να τον απελευθερώσουν. Εκείνοι, με βουλωμένα τα αυτιά, τον δένουν πιο σφιχτά. Ο Οδυσσέας, σε μια κατάσταση που έχει ο ίδιος σκηνοθετήσει, επιλέγει την κυριαρχία πάνω στους συντρόφους του έναντι ενός δημιουργικού καταστροφικού έρωτα με τη φύση-γυναίκα-κτήνος. Η διατήρηση της θέσης του Οδυσσέα ως κυρίαρχου συνεπάγεται δηλαδή την άρνηση της σαγήνης του πραγματικά ωραίου, αλλά και τη λήθη – καθώς δεν επιστρέφει αφού έχει ανακτήσει την ελευθερία του από τα δεσμά που ο ίδιος επέβαλε στον εαυτό του και στους γύρω του. Σε άλλα επεισόδια, βλέπουμε ότι ενδίδει στον έρωτα, αλλά και απαρνείται τη λήθη, σε οριοθετημένες βέβαια καταστάσεις. Για τον παρομοιαζόμενο όμως με τον Οδυσσέα, τους κυρίαρχους του διαφωτιστικού κόσμου, μπορούμε να θεωρήσουμε ότι στην πορεία των χρόνων δεν αποδείχθηκαν τόσο συγκρατημένοι όσο ο ίδιος. Έχοντας ξεχάσει την ουσία των δεσμών τους, ως πρόφαση για την ορθότητα της κυριαρχίας τους, αρχίζουν να τους τρώνε και να τους ενοχλούν. Ο σαδισμός, στον οποίο εκτονώνεται αυτή η δυσφορία, σταδιακά μετατρέπει τον Οδυσσέα, που σε κάποιο βαθμό κράτησε τα αυτιά του ανοιχτά όχι μόνο από πλεονεξία, αλλά και περιέργεια, που θεώρησε αναγκαίο να ακούσει κάποιος το τραγούδι για να μαρτυρήσει τα μυστικά του, αλλά και έκρινε ότι κάποιος άλλος πρέπει να κωπηλατεί, στον Κύκλωπα, του οποίου η δύναμη και η αγριότητα δεν προέρχονται από τον έξυπνο σχεδιασμό του αλλά από την αλογία του. Στην προσπάθειά του να βγάλει το μάτι του, να κλείσει επιτέλους τα πονεμένα αυτιά του, αντί να σπάσει τα δεσμά του με την νέα του κτηνώδη δύναμη, κρίνει ότι από το να ρισκάρει να ανοίξουν οι κωπηλάτες τα αυτιά τους, είναι προτιμότερο να



δημιουργήσει αυτόν που θα τον σακατέψει. Κλείνοντας το μάτι στον Ποσειδώνα, ο Οδυσσέας-Πολύφημος ονομάζει το ανδρείκελο Κανένα, ένα τέχνασμα που λειτουργεί βέβαια μόνο σε ξεβούλωτα αυτιά<sup>122</sup>.

Καταλήγουμε έτσι σε μια φαντασίωση της TN ως το ιδανικό ομοίωμα της ανορθολογικής ορθολογικότητας στην οποία βασίζεται το ratio ως κυριαρχία, η οποία έρχεται να επιλύσει την κρίση της νεωτερικής αυθεντίας. Μέσω της τέλει μαριονέτας, πραγματώνεται έτσι η αυθεντικοκεντρική οπτική, και ο άνθρωπος, μη έχοντας έλεγχο πάνω στη ζωή του και αποφασίζοντας να σταματήσει να τον αποζητά, μετατρέπεται ο ίδιος σε ομοίωμα του εαυτού του. Σε ένα ιδανικό ανδρείκελο του οποίου η κίνηση ελέγχεται πλήρως από μυριάδες νήματα, αόρατα και ορατά, τα οποία, απαλλαγμένος από το καθήκον να εξετάσει, μπορεί να βιώνει ως αισθητική απόλαυση. Έτσι η TN γίνεται το τέλει όργανο για να πραγματώσει τις τάσεις προς το φασισμό ή τον ολοκληρωτισμό που περιγράφονται από τους Χορκχάιμερ και Αντόρνο και την Άρεντ αντίστοιχα, κλείνοντας τον κύκλο του Διαφωτισμού *στο σημείο που συναντιούνται τα δύο άκρα του στρογγυλού κόσμου*, στην ενότητα μιας κυριαρχίας που αφαιρεί κάθε ουσία από την εμπρόθετη δράση του υποκειμένου.

### Το σάιμποργκ και ο Νίο

Στα πρώτα τρία *Matrix*, παρουσιάζεται η ιστορία της μάχης του προγραμματιστή-μεσσία Νίο ενάντια σε ένα κράτος μηχανών που έχουν υποδουλώσει τους ανθρώπους, αναγκάζοντάς τους να ζουν μέσα σε μια προσομοίωση του κόσμου των 90s – το *Μάτριξ*. Αυτό που αναδεικνύει τον πρωταγωνιστή σε ενός είδους υπερήρωα είναι η εξαιρετική του ικανότητα να αντιλαμβάνεται και να τροποποιεί τη γλώσσα της μηχανής, η οποία μέσα στην προσομοίωση συγκροτεί την πραγματικότητα. Αυτό που τον αναδεικνύει σε μεσσία είναι τελικά το γεγονός ότι, λόγω των ικανοτήτων του, πείθει τις κυρίαρχες μηχανές ότι τους είναι χρήσιμος για την επιδιόρθωση των αυτοκαταστροφικών κυριαρχικών έξεων που ενυπάρχουν στον προγραμματισμό

---

<sup>122</sup>Ο.π., 74–88. Το επιχείρημα που παρουσιάζω εδώ είναι μια ανασκευή των τελευταίων σελίδων του κεφαλαίου *H Έννοια του Διαφωτισμού* και του πρώτου παραρτήματος (βλ. σελίδα 124-125 για τον Κύκλωπα και τον Κανένα)

τους. Η τριλογία κλείνει με μια συμφωνία συνύπαρξης, όπου οι μηχανές δέχονται να προσφέρουν στον κάθε άνθρωπο την επιλογή μεταξύ του να ζήσει στην προσομοίωση, ή να επανέλθει στην κατεστραμμένη πραγματικότητα που πλέον μπορεί να προσπαθήσει να ανοικοδομήσει.

Τα πρώτα τρία *Matrix* εκφράζουν έτσι την απόλυτη φαντασίωση του *countergaming*, την ιδέα ότι όσο ο κόσμος γίνεται όλο και πιο υπολογιστικός, κάποιος που μπορεί να αντιληφθεί τη μηχανή για αυτό που είναι και να συνομιλήσει μαζί της στη γλώσσα της μπορεί να μεταχειριστεί το περιβάλλον του και να το αλλάξει: Είναι μια διασκευή της διαφωτιστικής διαλεκτικής ανθρώπου-φύσης εντός του κοινωνικού. Αυτό που αναγνωρίζει όμως η πολύ μεταγενέστερη τέταρτη ταινία στη σειρά είναι ότι η γκραντζ αισθητική των πρώτων, όπου ο μεσσίας βλέπει πέρα από τα *φούμαρα* της καθημερινότητας και αντιμετωπίζει ένα συμπαγή εχθρό, τη μηχανή, είναι ξεπερασμένη. Παρά τις όποιες αστοχίες της, η τέταρτη ταινία παρουσιάζει την αναγέννηση του Νίο σε ένα κόσμο όπου μηχανές και άνθρωποι συνυπάρχουν, αλλά παρόλα αυτά οι συνθήκες μοιάζουν το ίδιο καταπιεστικές. Το νέο *Μάτριξ* προσομοιώνει την σύγχρονη πραγματικότητα του 2020 και χρησιμοποιεί πολύ διαφορετικές τεχνικές πειθάρχησης για να πείθει τους ανθρώπους να παραμείνουν μέσα του: Η αόρατη ασφυκτική καταπίεση κρύβεται πίσω από ένα *σύστημα από καθρέφτες* – τα πρότυπα και τις απολαύσεις που προσφέρει στα άτομα ο νεοφιλελευθερισμός.

Ο Νίο είναι ένας μεσσίας χωρίς σκοπό – οι νέες συνθήκες θεωρούνται καλύτερες από τις παλαιότερες, και πλέον οι περισσότεροι άνθρωποι βρίσκονται στη θέση που είναι με τη συναίνεσή τους, ακόμα και αν δεν την απολαμβάνουν, ενώ απολαμβάνουν ακόμα και χωρίς τη συναίνεσή τους. Η ταινία καταλήγει στο ότι ο έλεγχος πάνω στον κόσμο από το υποκείμενο δεν χρειάζεται να άγεται από μια ορθολογική μοντελοποίησή του, ότι οι παρεμβάσεις με βάση αισθητικά και ηθικά κριτήρια που έχουμε συνηθίσει να θεωρούμε ευτελή είναι μάλλον η μόνη μέθοδος να υπερβούμε τις αυθεντικοκεντρικές μας νευρώσεις.

Η Χάραγουει<sup>123</sup>, ο Γκρέμπερ<sup>124</sup>, και ο Γκάλουγουει<sup>125</sup> κλείνουν τα αντίστοιχα βιβλία τους με αναλύσεις μύθων της τεχνολογίας (επιστημονικής φαντασίας ή παιχνιδιών) που

---

<sup>123</sup> Donna Haraway, *Το Μανιφέστο Των Σάιμποργκ*, 63–80.

<sup>124</sup> David Graeber, *Bullshit Jobs: Μια Θεωρία*, 279–90.

<sup>125</sup> Galloway, *Gaming*.

έχουν χειραφετητικό περιεχόμενο. Οι μύθοι της TN, που είναι αδιαχώριστοι από την εγκόσμια μορφή της, έχουν εκφράσει εδώ και εκατοντάδες χρόνια και αυτές και τις αντίθετες τάσεις. Μέσα από το *Matrix*, θα μπορούσαμε να δούμε τη βαθιά μάθηση ως μια πολλά υποσχόμενη παράκαμψη της δυσκολίας της μάθησης της γλώσσας του τεχνολογικού μας περιβάλλοντος, ένα τρόπο να το ελέγξουμε χωρίς να το κατανοήσουμε ή να εισέλθουμε τη λογική του – ήδη τέτοια συστήματα γράφουν κώδικα με ευχέρεια συγκρίσιμη με αυτή ενός έμπειρου προγραμματιστή. Σε αυτήν όμως την περίπτωση, η εξάρτησή μας από την TN θα μας ανάγκαζε να τη διαιωνίζουμε – και είδαμε ότι στη βάση της είναι μια τεχνολογία που πρέπει να θέλει να καταργήσει τον εαυτό της<sup>126</sup>.

Χωρίς να λέω ότι αυτή η λειτουργία, ή η παραγωγή φαρμάκων που υποβοηθά το AlphaFold, δεν είναι ουσιαστική πρόοδος, πιστεύω ότι η χειραφετητική προοπτική της TN έγκειται αλλού. Η απογύμνωση της ανορθολογικότητας της διαχείρισης και της κυριαρχίας μέσω της εξάρτησής της από ένα μη-κατανοήσιμο μηχάνημα, φέρνει στο προσκήνιο την αντίφαση και τον παραλογισμό της *ratio* με πρωτοφανή τρόπο. Αντί λοιπόν να προσπαθούμε να διαχωρίσουμε το μύθο από την ορθολογικότητα, τον άνθρωπο από τη μηχανή, ή τη φύση από τον πολιτισμό, αναγνωρίζοντας τη μηχανή ως συνομιλήτη και συνεργάτη μπορεί ίσως να αναφανεί η αναγκαιότητα της παρατήρησης του όντως πραγματικού – να αναγνωριστεί η διαδικασία του να θέτει κανείς τα δικά του όρια, και τα δικά του πρότυπα ως αναγκαία απόλαυση. Ίσως έτσι η εκφυλισμένη *ratio* της «αστικής τάξης πραγμάτων» και οι κατηγορίες που επιβάλλει για να κυριαρχεί, θα αντικατασταθούν από τις *έλξεις*, ή και τις απωθήσεις, που αντιστοιχούν στα υπαρκτά υποκείμενα, τις μοριακές συγγένειες που ζητάμε από την TN να διαισθανθεί· από κοινότητες και νουθεσίες με συναίσθηση της πολιτικής τους θέσης ως κατασκευασμένες ενότητες μέσα στην «τελική επιβολή ενός πλέγματος ελέγχου πάνω στον πλανήτη»<sup>127</sup>.

Έτσι η βαθιά μάθηση αποδυναμώνει ίσως τη συνθήκη που τη δημιούργησε, χωρίς να δημιουργεί ταυτόχρονα εξάρτηση από τον εαυτό της. Έτσι η *αντι-γλώσσα* της αυτοματοποίησης μπορεί να παραμεριστεί φτάνοντας στην πληρότητά της, παίρνοντας μαζί της την «πυρκαγιά της θεϊκής αλήθειας». Όλα αυτά όμως είναι απλά

---

<sup>126</sup>Rudin, 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'.

<sup>127</sup>Donna Haraway, *Το Μανιφέστο Των Σάμπουργκ*, 16–18.

εικασίες. Το ουσιαστικό είναι ότι, ανεξαρτήτως από το πώς μοιάζει η τεχνολογία, οι όροι για μια κοινωνία κυρίαρχων υποκειμένων παραμένουν οι ίδιοι:

Γιατί η ζωή σε μια πολιτική επικράτεια δίχως αυθεντία και τη συνακόλουθη επίγνωση ότι η πηγή της αυθεντίας υπερβαίνει την εξουσία και όσους βρίσκονται στην εξουσία, σημαίνει πως αντιμετωπίζουμε εξ υπαρχής, χωρίς τη θρησκευτική πίστη σε μια ιερή απαρχή και χωρίς την προστασία των παραδοσιακών και επομένως αυτονόητων κριτηρίων συμπεριφοράς, τα στοιχειώδη προβλήματα ανθρώπινης συμβίωσης<sup>128</sup>.

## Επίλογος

Μοιάζει ίσως να καταλήγω σε ένα αισιόδοξο συμπέρασμα, αλλά δεν έχω, δυστυχώς, αυτόν το σκοπό. Προσπάθησα να αναδείξω το ενδιαφέρον που παρουσιάζει η νέα τεχνολογία της τεχνητής νοημοσύνης και της βαθιάς μάθησης μέσα από το πρίσμα των θεωριών για την αυθεντία και τη γνώση που προσπαθούν να διασώσουν τα προτάγματα του Διαφωτισμού κριτικά, και να αναδείξω την κατεύθυνση μιας πιο σύγχρονης θεωρίας που βλέπει την κατάρρευση των κλασικών διαχωρισμών σαν μια πρόκληση για τη δημιουργία νέων θεωριών της χειραφέτησης. Η Χάραγουεϊ όμως συνθέτει το μύθο της *σάιμποργκ* οντολογίας ως έναν «ειρωνικό πολιτικό μύθο<sup>129</sup>» – θα ήταν λάθος να θεωρήσουμε ότι βλέπει την εξάρτηση του ανθρώπου από τη μηχανή ως μια θετική εξέλιξη. Το ζήτημα είναι ότι, ακόμα και αν το λογισμικό είναι ιδεολογία, είμαστε αναγκασμένοι να αναγνωρίσουμε τη «νέα τεχνολογία» ή τις νέες μορφές διαμεσολάβησης, ως κομμάτι της υλικής και ψυχολογικής πραγματικότητας του υποκειμένου – θα ήταν μεγαλύτερο λάθος λοιπόν να μιλάμε για μια μελλοντική μορφή διαβίωσης χωρίς να παίρνουμε υπόψιν τη σύγχρονη μορφή της τεχνολογίας ή παρουσιάζοντάς την συνολικά σαν κάποιου είδους *φούμαρα*.

Για αυτόν τον λόγο επέλεξα επίσης και δύο παραδείγματα που πέρα από την ενδιαφέρουσα σχέση τους με τη γνώση και την ειδημοσύνη έχουν εμφανώς επωφελή αντίκτυπο. Παραδείγματα όπως η τεχνητή νοημοσύνη που χρησιμοποιείται σε οπτικά

---

<sup>128</sup>Hannah Arendt, 'Τι Είναι Αυθεντία', 220.

<sup>129</sup>Donna Harraway, *Το Μανιφέστο Των Σάιμποργκ*, 5.

συστήματα<sup>130</sup>, στα χρηματοοικονομικά<sup>131</sup>, στην καταστολή<sup>132</sup>, ή μια πιο λεπτομερής ανάλυση της επιρροής που έχουν τα chatbot σε γενικό επίπεδο θα έδιναν μια πολύ πιο ρεαλιστική εικόνα του πώς τέτοια συστήματα ήδη επηρεάζουν τη σχέση μας με τον κόσμο. Καθώς επίσης η νεοφιλελεύθερη θεωρία έχει επιτύχει στο να θεωρείται η ιδέα του ορθολογικού σχεδιασμού της κοινωνίας ρομαντικό κατάλοιπο, η αντίστιξη μεταξύ εργαλειακού ορθολογισμού και διαλεκτικού υλισμού ίσως να φαίνεται πιο γερασμένη. Πράγματι, η ανάγνωση ενός βιβλίου σαν τη *Διαλεκτική του Διαφωτισμού* με τα σημερινά δεδομένα δημιουργεί συχνά το αντιφατικό αίσθημα ότι το κείμενο παραείναι απαισιόδοξο, καθώς οι σκοτεινές του προβλέψεις είναι σε μεγάλο βαθμό πλέον κοινοτοπίες της καθημερινής ζωής. Για αυτόν τον λόγο θεώρησα ότι έχει νόημα να εξετάσω τις αντιφάσεις της βαθιάς μάθησης από αυτή τη σκοπιά, ενδεχομένως ως ένα θεμέλιο για μια πιο «σύγχρονη προσέγγιση».

Κάποια από τα ζητήματα που ανοίγονται στην εισαγωγή μένουν για αυτό το λόγο αναπάντητα, και κάποια από τα επί μέρους θέματα που διατρέχουν αυτό το κείμενο δεν αναλύονται ίσως στο βάθος που τους αναλογεί – οι πολιτικές επιπτώσεις της τεχνολογίας δεν καλύπτονται από την ανάλυσή τους ως θέματα γνώσης. Η διεπιστημονική ανάλυση των θεωρητικών θεμελίων της τεχνητής νοημοσύνης είναι θεωρώ ένα πρόσφορο πεδίο για τη συνέχεια αυτής της έρευνας, καθώς η ύπαρξη μιας καθολικής κοινής αναφοράς – του *A/MA* –, και μεμονωμένων επιστημόνων που κινήθηκαν μεταξύ πολιτικής και επιστήμης υπολογιστών και είχαν τεράστια επιρροή και στα δύο – όπως ο Τούρινγκ και ο φον Νόιμαν –, σημαίνει ότι υπάρχουν πολύ κεντρικά σημεία ενδιαφέροντος. Θεωρώ συγκεκριμένα ότι μια τέτοιου είδους ανάλυση θα ήταν πολύ σημαντική για τη βαθύτερη κατανόηση της TN ως *μύθου*, και ότι μια τέτοια προσέγγιση θα αναδείκνυε την αναγκαιότητα να συνειδητοποιούν την πολιτική τους εμπλοκή και οι κατασκευαστές της εκάστοτε τεχνολογίας.

Η TN δεν είναι παρά το σημείο τομής εκατοντάδων μύθων και κατανοήσεων του κόσμου, που προέρχονται από πεδία που μερικές φορές δεν επικοινωνούν σχεδόν καθόλου. Ταυτόχρονα, παρόλο που λόγω της βαθιάς μάθησης θεωρούμε αυτή τη στιγμή την TN καινοτομία, οι παλαιότερες συλλήψεις της εφαρμόζονται ευρέως εδώ

---

<sup>130</sup> ‘AI in Israel’s War on Gaza’, *Access Now* (blog), τελευταία πρόσβαση 30 Σεπτεμβρίου 2024, <https://www.accessnow.org/publication/artificial-genocidal-intelligence-israel-gaza/>.

<sup>131</sup> ‘SEC.Gov | SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence’.

<sup>132</sup> Khari Johnson, ‘How Wrongful Arrests Based on AI Derailed 3 Men’s Lives’, *Wired*, τελευταία πρόσβαση 30 Σεπτεμβρίου 2024, <https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/>.

και χρόνια. Η μεταχείρισή της από την άποψη της σχέσης της με τη γνώση ή την αυθεντία λοιπόν δεν είναι μόνο ελλιπής σε σχέση με το συνολικό πεδίο, αλλά και ακαθόριστου πιθανού εύρους η ίδια. Αν όμως δεν βρεθεί μια κοινή γλώσσα μεταξύ αυτών των πεδίων, που να ανταποκρίνεται και στην υλική πραγματικότητα, δεν νομίζω ότι υπάρχει ιδιαίτερη ελπίδα η TN να χρησιμοποιηθεί για εφαρμογές εκτός της διαδικασίας *αντι-παραγωγής*.

Απέφυγα επίσης συνειδητά την αισθητική και την τέχνη, δύο πεδία που είναι απολύτως απαραίτητα για τη συζήτηση γύρω από την TN και από την άποψη του υλικού που παράγουν τα νέα συστήματα τύπου GPT αλλά και από την άποψη της αισθητικής που αντανακλά την εσωτερική λογική των συστημάτων. Σε αυτήν τη θεματική δένει επίσης μεταξύ άλλων και η συζήτηση περί προστασίας της πνευματικής ιδιοκτησίας ή των προσωπικών δεδομένων, η οποία θέλει ιδιαίτερη μεταχείριση – αν υπήρχε πραγματικά προστασία, δεν θα υπήρχαν τα μεγάλα μοντέλα που χρησιμοποιούμε σήμερα· ταυτόχρονα η πρόσβαση σε πληροφορία που προσφέρει το ίντερνετ δεν μπορεί να θεωρηθεί αρνητική εξέλιξη, παρά το γεγονός ότι μπορεί να είναι παράνομη.

Θα μπορούσα φυσικά να συνεχίσω. Ανταυτού θα συνοψίσω καταληκτικά το επιχείρημα ως εξής: Η εμφάνιση μιας μη-κατανοήσιμης τεχνολογίας σε ένα ιδεολογικό περιβάλλον που έχει παραιτηθεί από την κατανόηση του κόσμου είναι ένα γεγονός που έχει εξαιρετική σημασία. Ταυτόχρονα, η τεχνητή νοημοσύνη είναι ένα έμβλημα της κατάργησης των παραδοσιακών ορίων μέσω της συμβολικής της υπόστασης. Το αν αυτή η ιδεολογία που αποκτά υλική υπόσταση θα οδηγήσει σε μια αναγνώριση των παραπάνω ή θα θεωρηθεί μια ακόμα κοινοτοπία, το αν ο φθαρμένος διαχωρισμός θα καταρρεύσει σε μια ενότητα της *μαριονέτας* ή σε μια πολλαπλότητα του *σάιμποργκ*, ή σε κάτι τελείως διαφορετικό, ή σε κάτι τελείως ίδιο, δεν μπορούμε να το ξέρουμε. Ο μόνος τρόπος όμως να ξεφύγουμε από το πηχτό λυκόφως μιας μη-κατανοήσιμης κοινωνικής πραγματικότητας είναι η αναγνώριση της περιπλοκότητας του κόσμου – μόνο ένα επιμελές ψηλάφισμα θα μας επιτρέψει να καταλάβουμε πόσα από τα μυριάδες νήματα που φανταζόμαστε είναι πράγματι εκεί, και αν θέλουμε να κοινωνήσουμε την *ψυχή* που ενσταλάσσουν.

## Βιβλιογραφία

- Access Now. 'AI in Israel's War on Gaza'. Τελευταία πρόσβαση 30 Σεπτεμβρίου 2024. <https://www.accessnow.org/publication/artificial-genocidal-intelligence-israel-gaza/>.
- 'AI Labour Disclosure Initiative: Recognizing the Social Cost of Human Labour behind Automation | International Labour Organization', 19 Μαρτίου 2024. <https://www.ilo.org/meetings-and-events/ai-labour-disclosure-initiative-recognizing-social-cost-human-labour-behind>.
- 'Amazon Mechanical Turk'. Τελευταία πρόσβαση 29 Σεπτεμβρίου 2024. <https://www.mturk.com/>.
- Arendt, Hannah. 'Τι Είναι Αυθεντία'. In *Μεταξύ Παρελθόντος Και Μέλλοντος*. Αθήνα: Λεβιάθαν, 1996.
- 'Artificial Intelligence: A Modern Approach'. Σε *Wikipedia*, 3 Μαΐου 2024. [https://en.wikipedia.org/w/index.php?title=Artificial\\_Intelligence:\\_A\\_Modern\\_Approach&oldid=1222113628](https://en.wikipedia.org/w/index.php?title=Artificial_Intelligence:_A_Modern_Approach&oldid=1222113628).
- Bell, Startup. 'Sam Altman Told Investors Bluntly He Had No Plans On How To Generate Revenue'. Startup Bell, 14 Μαΐου 2024. <https://www.startupbell.net/post/sam-altman-told-investors-bluntly-he-had-no-plans-on-how-to-generate-revenue>.
- Benjamin, Walter. 'Για Την Έννοια Της Ιστορίας'. In *Walter Benjamin*. Αθήνα: Έρμα, 2023.
- Bryant, Patrick, Gabriele Pozzati, Wensi Zhu, Aditi Shenoy, Petras Kundrotas, and Arne Elofsson. 'Predicting the Structure of Large Protein Complexes Using AlphaFold and Monte Carlo Tree Search'. *Nature Communications* 13, 1 (12 Οκτωβρίου 2022): 6028. <https://doi.org/10.1038/s41467-022-33729-4>.
- Castoriades, Cornelius. *On the Content of Socialism*. Τελευταία πρόσβαση 29 Σεπτεμβρίου 2024. <https://usa.anarchistlibraries.net/library/cornelius-castoriadis-on-the-content-of-socialism>.
- Cao, Longbing. 'AI in Finance: A Review'. SSRN Scholarly Paper. Rochester, NY, 10 Ιουλίου 2020. <https://doi.org/10.2139/ssrn.3647625>.
- Chang, Oscar, Fernando A. Gonzales-Zubiarte, Luis Zhinin-Vera, Rafael Valencia-Ramos, Israel Pineda, and Antonio Diaz-Barrios. 'A Protein Folding Robot Driven by a Self-Taught Agent'. *Bio Systems* 201 (Μάρτιος 2021): 104315. <https://doi.org/10.1016/j.biosystems.2020.104315>.
- Copeland, B J, ed. *The Essential Turing*. Oxford University Press, 2004. <https://doi.org/10.1093/oso/9780198250791.001.0001>.
- Debord, Guy. *Η Κοινωνία Του Θεάματος*. Αθήνα: Μεταίχμιο, 2016.
- Dill, Ken A., S. Banu Ozkan, M. Scott Shell, and Thomas R. Weikl. 'The Protein Folding Problem'. *Annual Review of Biophysics* 37 (9 Ιουνίου 2008): 289–316. <https://doi.org/10.1146/annurev.biophys.37.092707.153558>.
- Engels, Friedrich. 'On Authority'. In *Marx and Engels Collected Works*, 23:422–25. Progress Publishers, 1988.
- Eubanks, Ben. *Artificial Intelligence for HR: Use AI to Support and Develop a Successful Workforce*. Kogan Page Publishers, 2022.
- Evans, Richard, Michael O'Neill, Alexander Pritzel, Natasha Antropova, Andrew Senior, Tim Green, Augustin Židek, et al. 'Protein Complex Prediction with AlphaFold-Multimer'. bioRxiv, 10 Μαρτίου 2022. <https://doi.org/10.1101/2021.10.04.463034>.
- Fontana, Pietro, Ying Dong, Xiong Pi, Alexander B. Tong, Corey W. Hecksel, Longfei Wang, Tian-Min Fu, Carlos Bustamante, and Hao Wu. 'Structure of Cytoplasmic Ring of Nuclear Pore Complex by Integrative Cryo-EM and AlphaFold'. *Science (New York, N.Y.)* 376, no. 6598 (10 Ιουνίου 2022): eabm9326. <https://doi.org/10.1126/science.abm9326>.

- Fromm, Erich, Horkheimer, Max, και Marcuse, Herbert. *Αυθεντία Και Οικογένεια*. Αθήνα: Νήσος, 1995.
- Fujita, Hiroshi. 'AI-Based Computer-Aided Diagnosis (AI-CAD): The Latest Review to Read First'. *Radiological Physics and Technology* 13, no. 1 (1 Μαρτίου 2020): 6–19. <https://doi.org/10.1007/s12194-019-00552-4>.
- Galloway, Alexander R. *Gaming: Essays on Algorithmic Culture*. NED-New edition. Vol. 18. University of Minnesota Press, 2006. <https://www.jstor.org/stable/10.5749/j.cttss5p>.  
 ———. *The Interface Effect*. Polity, 2012.
- Garry Chess On the AlphaZero Era, 2021. <https://www.youtube.com/watch?v=FTxoV8DC48c>.
- Graeber, David. *Bullshit Jobs: Μια Θεωρία*. Αθήνα: Στάσει Εκπίπτοντες, 2020.  
 ———. 'On the Phenomenon of Bullshit Jobs'. STRIKE! Magazine. STRIKE! Accessed 29 September 2024. <https://strikemag.org>.
- Goverde, Casper A., Benedict Wolf, Hamed Khakzad, Stéphane Rosset, and Bruno E. Correia. 'De Novo Protein Design by Inversion of the AlphaFold Structure Prediction Network'. *Protein Science : A Publication of the Protein Society* 32, no. 6 (1 Ιουνίου 2023): e4653. <https://doi.org/10.1002/pro.4653>.
- Haraway, Donna. *Το Μανιφέστο Των Σάιμποργκ*. Αθήνα: Τοποβόρος, 1985.
- Hartmann, J. 'Garry Kasparov Is a Cyborg, or What ChessBase Teaches Us about Technology'. In *Philosophy Looks at Chess*, edited by Benjamin Hale, 39–64. Open Court Press, 2008.
- Horkheimer, Max, Adorno, Theodor W.. *Διαλεκτική Του Διαφωτισμού*. Αθήνα: Νήσος, 1996.
- Hristova, Tsvetelina, Liam Magee, and Karen Soldatic. 'The Problem of Alignment'. *AI & SOCIETY*, 7 August 2024. <https://doi.org/10.1007/s00146-024-02039-2>.
- IBM Research. 'What Is AI Alignment?', 9 Φεβρουαρίου 2021. <https://research.ibm.com/blog/what-is-alignment-ai>.
- Johnson, Khari. 'How Wrongful Arrests Based on AI Derailed 3 Men's Lives'. *Wired*. Τελευταία πρόσβαση 30 Σεπτεμβρίου 2024. <https://www.wired.com/story/wrongful-arrests-ai-derailed-3-mens-lives/>.
- Jumper, John, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, et al. 'Highly Accurate Protein Structure Prediction with AlphaFold'. *Nature* 596, no. 7873 (Αύγουστος 2021): 583–89. <https://doi.org/10.1038/s41586-021-03819-2>.
- Kant, Immanuel. 'Απάντηση Στο Ερώτημα Τι Είναι Διαφωτισμός'. In *Τι Είναι Διαφωτισμός*. Αθήνα: Κριτική, 2014.
- Kralingen, Max van. 'Use of Weapons: Should We Ban the Development of Autonomous Weapons Systems?' *The International Journal of Intelligence, Security, and Public Affairs* 18, no. 2 (3 Μαΐου 2016): 132–56. <https://doi.org/10.1080/23800992.2016.1196947>.
- Lagassé, H.A. Daniel, Aikaterini Alexaki, Vijaya L. Simhadri, Nobuko H. Katagiri, Wojciech Jankowski, Zuben E. Sauna, and Chava Kimchi-Sarfaty. 'Recent Advances in (Therapeutic Protein) Drug Development'. *F1000Research* 6 (7 Φεβρουαρίου 2017): 113. <https://doi.org/10.12688/f1000research.9970.1>.
- Lea, Andrew. 'Computer Chess: A Historical Perspective'. *ITNOW* 66, no. 3 (1 Σεπτεμβρίου 2024): 50–52. <https://doi.org/10.1093/itnow/bwae098>.
- Lee, Hajin. 'Impact of Go AI on the Professional Go World'. *Medium* (blog), 30 Αυγούστου 2020. <https://hajinlee.medium.com/impact-of-go-ai-on-the-professional-go-world-f14cf201c7c2>.
- Levien, R., and Sze Tan. 'Double Pendulum: An Experiment in Chaos'. *American Journal of Physics - AMER J PHYS* 61 (1 Νοεμβρίου 1993): 1038–44. <https://doi.org/10.1119/1.17335>.



- Merrill, Stephen A., Anne-Marie Mazza, and National Research Council (US) Committee on Intellectual Property Rights in Genomic and Protein Research and Innovation. 'Genomics, Proteomics, and the Changing Research Environment'. In *Reaping the Benefits of Genomic and Proteomic Research: Intellectual Property Rights, Innovation, and Public Health*. National Academies Press (US), 2006. <https://www.ncbi.nlm.nih.gov/books/NBK19861/>.
- Ras, Gabrielle, Ning Xie, Marcel van Gerven, and Derek Doran. 'Explainable Deep Learning: A Field Guide for the Uninitiated'. *Journal of Artificial Intelligence Research* 73 (25 Ιανουαρίου 2022): 329–96. <https://doi.org/10.1613/jair.1.13200>.
- Rudin, Cynthia. 'Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead'. *Nature Machine Intelligence* 1, no. 5 (13 Μαΐου 2019): 206–15. <https://doi.org/10.1038/s42256-019-0048-x>.
- Russel, Stuart, Norvig, Peter. *Τεχνητή Νοημοσύνη: Μια Σύγχρονη Προσέγγιση*. Αθήνα: Κλειδάριθμος, 2021.
- Schaeffer, Jonathan, Yngvi Björnsson, Akihiro Kishimoto, Martin Müller, Robert Lake, Paul Lu, and Steve Sutphen. 'Checkers Is Solved'. *Science* 317 (1 Οκτωβρίου 2007): 1518–22. <https://doi.org/10.1126/science.1144079>.
- Schrittwieser, Julian, Ioannis Antonoglou, Thomas Hubert, Karen Simonyan, Laurent Sifre, Simon Schmitt, Arthur Guez, et al. 'Mastering Atari, Go, Chess and Shogi by Planning with a Learned Model'. *Nature* 588, no. 7839 (24 Δεκεμβρίου 2020): 604–9. <https://doi.org/10.1038/s41586-020-03051-4>.
- 'SEC.Gov | SEC Charges Two Investment Advisers with Making False and Misleading Statements About Their Use of Artificial Intelligence'. Τελευταία πρόσβαση 29 Σεπτεμβρίου 2024. <https://www.sec.gov/newsroom/press-releases/2024-36>.
- Silver, David, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, et al. 'Mastering the Game of Go with Deep Neural Networks and Tree Search'. *Nature* 529, no. 7587 (1 Ιανουαρίου 2016): 484–89. <https://doi.org/10.1038/nature16961>.
- Solon, Nate. 'Towards Explainable AI for Chess'. Substack newsletter. *Zwischenzug* (blog), 28 Φεβρουαρίου 2021. <https://zwischenzug.substack.com/p/towards-explainable-ai-for-chess>.
- Terwilliger, Thomas C., Dorothee Liebschner, Tristan I. Croll, Christopher J. Williams, Airlie J. McCoy, Billy K. Poon, Pavel V. Afonine, et al. 'AlphaFold Predictions Are Valuable Hypotheses and Accelerate but Do Not Replace Experimental Structure Determination'. *Nature Methods* 21, no. 1 (Ιανουάριος 2024): 110–16. <https://doi.org/10.1038/s41592-023-02087-4>.
- The Observer*. 'The Superhero of Artificial Intelligence: Can This Genius Keep It in Check?' 16 Φεβρουαρίου 2016, sec. Technology. <https://www.theguardian.com/technology/2016/feb/16/demis-hassabis-artificial-intelligence-deepmind-alphago>.
- Tsonis, A. A., and J. B. Elsner. 'Chaos, Strange Attractors, and Weather'. *Bulletin of the American Meteorological Society* 70, no. 1 (1 Ιανουαρίου 1989): 14–23. [https://doi.org/10.1175/1520-0477\(1989\)070<0014:CSAAW>2.0.CO;2](https://doi.org/10.1175/1520-0477(1989)070<0014:CSAAW>2.0.CO;2).
- TURING, A. M. 'I.—COMPUTING MACHINERY AND INTELLIGENCE'. *Mind* LIX, no. 236 (1 Οκτωβρίου 1950): 433–60. <https://doi.org/10.1093/mind/LIX.236.433>.
- 'Using AlphaFold Predictions in Viral Research'. Τελευταία πρόσβαση 29 Σεπτεμβρίου 2024. <https://www.mdpi.com/1467-3045/45/4/240>.
- Vogl, Thomas, Cathrine Seidelin, Bharath Ganesh, and Jonathan Bright. 'Algorithmic Bureaucracy'. In *Proceedings of the 20th Annual International Conference on Digital Government*

- Research*, 148–53. Dg.o '19. New York, NY, USA: Association for Computing Machinery, 2019. <https://doi.org/10.1145/3325112.3325240>.
- Von Kleist, Heinrich. *Οι Μαριονέτες*. Αθήνα: Άγρα, 2015.
- Watson, J. *Chess Strategy in Action*. Gambit, 2003. <https://books.google.gr/books?id=ScqLAAAACAAJ>.
- . *Secrets of Modern Chess Strategy: Advances Since Nimzowitsch*. Gambit, 1999. <https://books.google.gr/books?id=CNXIAuqFs0EC>.
- ΟΥΤΟΠΙΑ, ΤΕΥΧΟΣ 133, ΙΟΥΛΙΟΣ - ΑΥΓΟΥΣΤΟΣ - ΣΕΠΤΕΜΒΡΙΟΣ 2020. Τελευταία πρόσβαση 29 Σεπτεμβρίου 2024. <https://www.politeianet.gr/magazines/-periodika-outopia-teuchos-133-ioulios-augoustos-septembrios-2020-317132>.
- Τεχνητή Νοημοσύνη. 'Γλωσσάρι EN'. Τελευταία πρόσβαση 29 Σεπτεμβρίου 2024. <https://aibook.gr/glossary-en/>.