

# Τεχνητή Νοημοσύνη: Τὸ τέλος τῆς Ἐποχῆς τοῦ Ἀνθρώπου;

Σεργίου Θεοδωρίδη\*

Θὰ ἦταν κοινοτοπία νὰ ἐπαναλάβουμε ὅτι ἡ ἐποχὴ μας χαρακτηρίζεται ἱστορικὰ ὡς ἡ ἀρχὴ τῆς τέταρτης βιομηχανικῆς ἐπανάστασης. Ἡ πρώτη βιομηχανικὴ ἐπανάσταση ταυτίζεται μὲ τὴ χρήση τοῦ ἀτμοῦ γιὰ τὴ λειτουργία τῶν μηχανῶν καὶ τὸ ξεκίνημά της τοποθετεῖται κάπου στὸ τέλος τοῦ 18ου αἰῶνα. Ἡ δεύτερη βιομηχανικὴ ἐπανάσταση δρομολογήθηκε μὲ τὴ χρήση τοῦ ἠλεκτρισμοῦ ἀπὸ τὴ μία καὶ τὴν ἐπιστημονικὰ ὀργανωμένη μαζικὴ παραγωγὴ ἀπὸ τὴν ἄλλη. Ἡ τρίτη βιομηχανικὴ ἐπανάσταση οἰκοδομεῖται πάνω στὴν αὐτοματοποίηση τῶν μηχανῶν, τὴ χρήση τῆς ἠλεκτρονικῆς καὶ τοῦ τρανζίστορ καὶ τὴν ἐπεξεργασία τῆς πληροφορίας ποὺ ἐνέχουν τὰ δεδομένα. Ἡ τέταρτη βιομηχανικὴ ἐπανάσταση εἶναι μετεξέλιξη τῆς τρίτης καὶ ἀποκύημα τῆς σύγκλισης τῶν τεχνολογιῶν τῶν ὑπολογιστῶν καὶ τῶν ἐπικοινωνιῶν (internet). Ἡ εἰδοποιὸς διαφορὰ ἀπὸ τὶς προηγούμενες τρεῖς εἶναι ὅτι στὸ πλαίσιο τῆς τέταρτης ἐπανάστασης, πέρα ἀπὸ τὴ χειρωνακτικὴ ἐργασία, ἔχουμε καὶ τὴν ἀντικατάσταση τῆς πνευματικῆς ἐργασίας, ποὺ μέχρι τώρα ἐπιτελοῦσε ὁ ἄνθρωπος, ἀπὸ «μηχανές». Αὐτὴ εἶναι μία ποιοτικὰ σημαντικὴ διαφορὰ ποὺ δὲν περιορίζει τὴ χρήση «μηχανῶν» στὸν χῶρο τῆς βιομηχανικῆς παραγωγῆς (ἐργοστάσια) μόνον, ἀλλὰ διαχέει τὴ χρήση τῶν «μηχανῶν» σὲ ὅλο τὸ φάσμα τῆς οἰκονομικῆς μας ζωῆς (ἐμπόριο, ὑπηρεσίες) καὶ ταυτόχρονα ἐνσωματώνει τίς «μηχανές» στὸν τρόπο μὲ τὸν ὁποῖο διασκεδάζουμε (μουσικὴ, ταινίες),

---

\* Ὁ Σέργιος Θεοδωρίδης εἶναι Ὁμότιμος Καθηγητὴς Μηχανικῆς Μάθησης καὶ Ἐπεξεργασίας Δεδομένων τοῦ Τμήματος Πληροφορικῆς καὶ Τηλεπικοινωνιῶν τοῦ Ἐθνικοῦ καὶ Καποδιστριακοῦ Πανεπιστημίου Ἀθηνῶν, Distinguished Professor στὸ Aalborg University (Denmark) καὶ Μέλος τῆς Ἐθνικῆς Ἐπιτροπῆς Βιοηθικῆς καὶ Τεχνοηθικῆς.

στὸν τρόπο μὲ τὸν ὁποῖο πραγματώνουμε τὴν κοινωνική μας ὑπόσταση ὡς κοινωνικά ὄντα (κοινωνικά δίκτυα), στὸν τρόπο μὲ τὸν ὁποῖο λειτουργοῦμε καὶ ἐνημερωνόμαστε ὡς πολῖτες (ΜΜΕ), στὸν τρόπο ποὺ ἐνσωματώνουμε τὶς νέες γενιές στὴν κοινωνία (Παιδεία) ἀλλὰ καὶ δημιουργοῦμε τὸ μέλλον μας (ἔρευνα, ἰδεολογίες, πολιτισμός). Μὲ ἄλλα λόγια, ἡ δυναμικὴ ποὺ ἀναβλύζει ἀπὸ τὴν τέταρτη βιομηχανικὴ ἐπανάσταση δημιουργεῖ καὶ πλάθει, ἄμεσα καὶ ὄχι ἔμμεσα, ἓνα νέο περιβάλλον μέσα στὸ ὁποῖο ὁ Ἄνθρωπος, ὡς κοινωνικὸ ὄν, καλεῖται νὰ ζήσει καὶ νὰ δημιουργήσει. Καὶ τὸ νέο αὐτὸ περιβάλλον χαρακτηρίζεται ἀπὸ τὴ σύγκλιση καὶ ἐνοποίηση τῆς σφαίρας τοῦ πραγματικοῦ, τοῦ ψηφιακοῦ καὶ τοῦ βιολογικοῦ «κόσμου».

Ἔχουν συγγραφεῖ δεκάδες βιβλίων, ἑκατοντάδες ἄρθρων στὸν ἔντυπο καὶ ἠλεκτρονικὸ τύπο, καὶ ἔχουν γίνῃ πολλές συζητήσεις στὴν τηλεόραση, σὲ σχετικὲς συνεδρίες καὶ ὑπάρχει πληθώρα videos στὸ διαδίκτυο. Ἡ συντριπτικὴ ὅμως πλειονότητα τῶν σχετικῶν τοποθετήσεων ἀφορᾷ τὶς ἐπιπτώσεις τῶν νέων τεχνολογιῶν στὴν ἀγορὰ ἐργασίας καὶ τὶς ἐργασιακὲς σχέσεις, τοὺς φόβους ποὺ σχετίζονται μὲ τὴν ἀνεργία ἀπὸ τὴ μία, ἀλλὰ καὶ τὶς δυνατότητες γιὰ νέες θέσεις ἐργασίας ἀπὸ τὴν ἄλλη. Γίνονται συζητήσεις σχετικὰ μὲ τοὺς κινδύνους ποὺ ἐνέχει τὸ νέο αὐτὸ περιβάλλον στὴ λειτουργία τῆς δημοκρατίας καὶ τῶν θεσμῶν, τὴν ιδιωτικότητα, τὴν ἀσφάλεια, τὶς δυνατότητες χαλιναγώγησης τῆς κοινῆς γνώμης ἀπὸ τὴ μία, ἀλλὰ καὶ τὶς νέες δυνατότητες γιὰ περισσότερη ἄμεση Δημοκρατία καὶ πιὸ ἐνημερωμένους πολῖτες μέσα ἀπὸ μία πιὸ ἐλεύθερη διακίνηση ἰδεῶν ἀλλὰ καὶ μὲ τὸν ταυτόχρονο κίνδυνο πλαστογράφησης τῶν γεγονότων μέσα ἀπὸ τὰ κοινωνικά δίκτυα. Συζητήσεις διεξάγονται σχετικὰ μὲ τὰ (τεχνο)φοβικὰ αἰσθήματα ποὺ δημιουργοῦνται στὴν αὐγὴ τῆς νέας αὐτῆς ἐποχῆς καὶ γίνονται ἀφορμὴ ἐκμετάλλευσης ἀπὸ «ἰδεολογικούς» γυρολόγους, ποὺ εὐαγγελίζονται τὴ θαλπωρὴ μέσα ἀπὸ ἀκραῖα ἰδεολογικο/θρησκευτικὰ συστήματα ἀπὸ τὴ μία, ἀλλὰ καὶ ταυτόχρονα τὴν ἀπελευθερωτικὴ ὁρμὴ καὶ τὶς δυνατότητες ποὺ ἀναφαίνονται ἀπὸ τὶς νέες τεχνολογίες, ὅπως γιὰ παράδειγμα τὴ μείωση τῶν ὠρῶν ἐργασίας, τὸ ἐγγυημένο ἐλάχιστο εἰσόδημα καὶ τὶς δυνατότητες ἐπιμήκυνσης τοῦ προσδόκιμου τῆς ζωῆς, μέσα ἀπὸ τὶς ἐξελίξεις τῆς ἰατρικῆς καὶ τὶς δυνατότητες αὐτοῦ ποὺ ὀνομάζεται «προσωποποιημένη» ἰατρική.

Πολύ λιγότερα όμως έχουν γραφεί και έχουν συζητηθεί σχετικά με την ίδια την Τεχνολογία που ώθει τις παραπάνω εξελίξεις, την τεχνολογία δηλαδή της τεχνητής νοημοσύνης και των νευρωνικών δικτύων αρχιτεκτονικών βάθους (*deep neural networks*). Στο παρόν άρθρο θα εστιάσουμε σε αυτήν την πτυχή και σε όρισμένες «φιλοσοφικές» προεκτάσεις που απορρέουν από αυτή, με απλά, πολύ απλά λόγια. Στόχος μας είναι η «άπομυθοποίηση» των τεχνολογιών αυτών. Μία τέτοια άπομυθοποίηση είναι προϋπόθεση για να αντιμετωπίσουμε τις προκλήσεις που τίθενται στη νέα αναδυόμενη εποχή. Άς μη ξεχνούμε ότι οι πολιτικές θεωρίες, οι ιδεολογίες και κατά συνέπεια ο ίδιος ο ανθρώπινος πολιτισμός και τα ρεύματα που τον διαμορφώνουν στον ροή του χρόνου αποτελούν «ἀφηγήματα». Είναι το σύνολο των αξιών, των αντιλήψεων και των συμπεριφορών, μέσω των οποίων αντιλαμβανόμαστε τον υπάρχοντα κόσμο και συνδεόμαστε μαζί του. Και οι «μύθοι» παίζουν κυρίαρχο ρόλο στη διαμόρφωσή τους.

Οι «μύθοι» έχουν πολλαπλές αναγνώσεις. Μπορεί να ενεργοποιούν και να δημιουργούν όράματα από τη μία, αλλά και να φανατίζουν και να αδρανοποιούν από την άλλη. Πολλές φορές ταυτόχρονα και τα δύο. Ανάλογα με την εποχή, κάποιος κυρίαρχος «μύθος» μπορεί να καλλιεργεί αισθήματα μειονεξίας, αισθήματα ασημαντότητας και μία τάση αποδοχής των γεγονότων ως αναπόφευκτων. Μπορεί να κλονίζει την αυτοεκτίμηση του πολίτη και κατά συνέπεια την αισιοδοξία του –που αποτελεί προϋπόθεση για το μέλλον του– και έτσι να του στερεί την ικανότητα για κριτική στάση και παρέμβαση. Ακόμη χειρότερα, ίσως το καλλιεργεί και φοβικά συναισθήματα. Αφ' έτερου, ο ίδιος «μύθος» μπορεί να λειτουργεί και σαν όπλο στα χέρια των ισχυρών της ύπαρχουσας ή της αναδυόμενης νέας κυρίαρχης τάξεως, σε εποχές μεγάλων ανακατάξεων, όπως είναι η δική μας. Ο νέος «μύθος», που πάντοτε δημιουργείται στις νέες «εποχές», συνειδητά ή ασυνείδητα, με πρόθεση ή χωρίς πρόθεση, μπορεί να δρᾶ ανασταλτικά και σαν τροχοπέδη στις κοινωνικές παρεμβάσεις που απαιτούνται για να επιτευχθούν οι νέες ισορροπίες και θεσμικές θωρακίσεις. Και βέβαια, τους μύθους αυτούς χρησιμοποιούν τα εκάστοτε αναδυόμενα «ίερατεία» και οι εκάστοτε ταγοί του «συστήματος», που σήμερα εκπροσωπούνται –σημασιακά– από τη Silicon Valley στις Η.Π.Α. και

τὸ Shenzhen στὴν Κίνα. Κάθε νέα ἐποχὴ ἐνέχει καὶ τὰ σπέρματα ἐνός «φρονταμενταλισμοῦ», ποὺ συχνά-πυκνά καθορίζει καὶ τὶς ἐξελίξεις.

Ὁ παρὼν «μῦθος», ἢ ἂν θέλετε μία πτυχὴ τοῦ νέου «μῦθου», καλλιεργεῖ τὴν ἄποψη ὅτι οἱ μηχανές ποὺ προαναφέρθηκαν ἔχουν «νοημοσύνη». Ὁ παρὼν μῦθος «διαδίδει» ὅτι οἱ μηχανές αὐτὲς θὰ ἐκτοπίσουν τὸν Ἄνθρωπο, γιατί θὰ γίνουν ἀνώτερες. Ὁ παρὼν μῦθος προβλέπει ὅτι οἱ μηχανές αὐτὲς θὰ ἐπιβληθοῦν στὸν Ἄνθρωπο λόγω τῆς ἀνωτερότητάς τους καὶ τῆς νοημοσύνης ποὺ παράγουν. Ὁ παρὼν μῦθος εὐαγγελίζεται ὅτι σύντομα τὰ ἐπιστημονικὰ ἐπιτεύγματα τῆς τεχνητῆς νοημοσύνης (TN) θὰ παράξουν «μηχανές» μὲ ὑπεράνθρωπη νοημοσύνη καὶ ὅτι ἡ ἐποχὴ τοῦ Ἀνθρώπου τελειώνει. Naί, ὅλες αὐτὲς οἱ ἀπόψεις ἔχουν διατυπωθεῖ ἀπὸ τοὺς λεγόμενους «γκουρού» τῆς νέας ἐποχῆς. Ἄρα, ὑπόρρητα, ὁ Ἄνθρωπος παρομοιάζεται σὰν μία ἄλλη «μηχανή» καὶ μάλιστα ὑποδεέστερη ἀπὸ αὐτὲς ποὺ ἔχει ὁ ἴδιος δημιουργήσει. Δὲν ξέρουμε τί θὰ μπορούσε νὰ γίνῃ στὸ μέλλον, καὶ τὸ παρὸν ἄρθρο δὲν εἶναι ἄρθρο ἐπιστημονικῆς φαντασίας. Εἶναι ἄρθρο πανεπιστημιακοῦ δασκάλου, ποὺ ἀσχολεῖται μὲ τὸ θέμα αὐτὸ γιὰ κάποιες δεκαετίες καὶ τὰ σχόλιά μας ἀφοροῦν αὐτὰ ποὺ μπορούν νὰ προβλεφθοῦν μὲ βάση τὸ ἐπιστημονικὸ παρὸν καὶ τεχνολογικὸ γίνεσθαι τοῦ σήμερα. Καὶ παραφράζοντας αὐτὸ ποὺ εἶπε ὁ Wittgenstein, γράφω γι' αὐτὰ ποὺ μπορῶ νὰ γράψω. Γιὰ τὰ ὑπόλοιπα θὰ σιωπήσω.

Ἰσχυρίζομεθα ὅτι ὁ ὅρος τεχνητὴ νοημοσύνη δὲν εἶναι δόκιμος καὶ μᾶλλον εἶναι παραπλανητικός. Ἡ νοημοσύνη ποὺ ἐνσωματώνουν οἱ ὑπολογιστὲς τοῦ σήμερα μόνον μὲ τὴ νοημοσύνη τοῦ ἀνθρωπίνου εἴδους πρὶν ἀπὸ μερικὰ ἑκατομμύρια χρόνια μπορεῖ νὰ συγκριθεῖ, τότε ποὺ ἐξελικτικὰ τὸ εἶδος μας ἦταν ἀκόμη πιὸ κοντὰ στὰ ἄλλα ζῶα. Οἱ «μηχανές» τεχνητῆς νοημοσύνης σήμερα δὲν μπορούν νὰ κάνουν τίποτε παραπάνω ἀπὸ τὸ νὰ «προβλέπουν» μὲ βάση τὴν πρότερη ἐμπειρία ποὺ ἔχουν, ὅπως αὐτὴ εἶναι ἐνσωματωμένη στὰ δεδομένα ποὺ ἔχουν συλλεγεῖ καὶ χρησιμοποιοῦνται γιὰ τὴ λεγόμενη ἐκπαίδευσή τους: ὅπως δηλαδὴ «προβλέπει» μία γάτα τί θὰ γίνῃ ὅταν δεῖ ἓνα σκύλο καὶ μὲ βάση τὴν πρότερη ἐμπειρία τῆς φεύγει, ὅπως ὁ πρωτόγονος ἄνθρωπος μπορούσε νὰ προβλέπει ὅτι ὁ ἥλιος θὰ ἀνατείλει, ἐπειδὴ ἔτσι ἔκανε χθὲς καὶ προχθὲς κ.λπ., ἀλλὰ ὅμως θὰ πανικοβληθεῖ τὴν ἡμέρα τῆς ἐκλειψῆς τοῦ ἡλίου καὶ θὰ ἐπικαλεσθεῖ τὰ πνεύματα. Ἡ νοημοσύνη

των μηχανών σήμερα περιορίζεται σε αυτό που είναι γνωστό ως επαγωγικός συλλογισμός. Είναι ο τύπος νοημοσύνης που κυριαρχεί στο ζωικό βασίλειο.

Αυτό όμως που διαχωρίζει την ανθρώπινη νοημοσύνη από εκείνη των ζώων είναι η ικανότητα έλλογης σκέψης, η ικανότητα λογικών συσχετισμών και πάνω από όλα η ικανότητα του Άνθρώπου να έρμηνεύει και να εξηγεί: να δημιουργεί υποθέσεις και να τις επαληθεύει ή να τις διαψεύδει. Η ανθρώπινη νοημοσύνη χαρακτηρίζεται κυρίως από την ικανότητα του νοῦ να δημιουργεί έννοιες και ιδέες που κωδικοποιούν γενικεύσεις και να τις χειρίζεται. Ο Κόσμος γύρω μας είναι ιδιαίτερος πολύπλοκος για να μπορούμε να τον αντιλαμβανόμαστε με όλες τις λεπτομέρειές του. Μέσα από τα εκατομμύρια χρόνια εξέλιξης, η ανθρώπινη νοημοσύνη απέκτησε την ικανότητα να χειρίζεται αφηρημένες έννοιες/ιδέες, αναπαραστάσεις στον νοῦ που δεν απορρέουν άμεσα από τις εμπειρίες των αισθήσεών μας. Καμμία μαθηματική «εὐθεία» δεν υπάρχει στον κόσμο. Ο κόσμος αυτός των Ἰδεῶν προκάλεσε την περιέργεια του Πλάτωνα και ἔφθασε στην υπόθεση που ἔφθασε. Οἱ αφηρημένες αυτές έννοιες παρέχουν στον Ἄνθρωπο την ικανότητα να ἀπλοποιεῖ και ταυτόχρονα να «κωδικοποιεῖ» την πεμπτούσια του ἀντικειμένου τῆς σκέψης. Ὅπως λέμε στη μηχανική μάθηση, ὁ ἀνθρώπινος νοῦς ἔχει τὴ δυνατότητα να γενικεύει και να ἐκπροσωπεῖ «κλάσεις» ἀντικειμένων και ὄντοτήτων. Μὲ ἄλλα λόγια, ἡ ιδέα/κλάση τῆς «γάτας» ἐκπροσωπεῖ ὅλες τὶς γάτες, ἀνεξάρτητα ἀπὸ τὸ μέγεθος, τὸ χρῶμα, τὴν ἡλικία, και κάθε ἰδιαιτερότητα που σχετίζεται μὲ κάθε συγκεκριμένη γάτα. Καὶ πάνω ἀπ’ ὅλα, εἶναι ἡ ικανότητα τοῦ νοῦ, μὲ τὴ βοήθεια τῆς γλώσσας, να ἀντιλαμβάνεται τὸν Κόσμο γύρω του και να τὸν «ἐπενδύει» μὲ σημασίες/νοήματα και να δημιουργεῖ.

Ἡ δημιουργία εἶναι ἡ πεμπτούσια τῆς Ἀνθρώπινης νοημοσύνης. Ὅχι μόνο μὲ τὴν ἔννοια τῆς δημιουργίας ἔργων Τέχνης ἢ ἐπιστημονικῶν ἀνακαλύψεων ἀλλὰ και μέσα ἀπὸ τὴν καθημερινότητά μας. Ὁ κάθε ἄνθρωπος αὐτοδημιουργεῖται καθημερινά. Οἱ ἀποφάσεις μας για ὀτιδήποτε δὲν εἶναι προκαθορισμένες/προβλέψιμες. Κανείς δὲν μπορεῖ να προβλέψει, ξεκινώντας μία σχέση, πῶς θὰ εἶναι αὐτὴ μετὰ ἀπὸ ἕναν μῆνα. Κανείς δὲν μπορεῖ να προβλέψει τὸ ἐργασιακό του μέλλον σὲ ἕναν χρόνο ἀπὸ τώρα και τί ἀποφάσεις θὰ πρέπει να πάρει

τότε. Ὁ ἀνθρώπινος νοῦς μαθαίνει καὶ ἀποφασίζει σὲ ἓνα συνεχῶς μεταβαλλόμενο περιβάλλον. Καὶ πάνω ἀπὸ ὅλα, οἱ ἀποφάσεις του εἶναι ἀποφάσεις βασισμένες τόσο σὲ ἔλλιπῆ ὅσο καὶ ἀσαφῆ γνώση. Γιατὶ ὁ ἀνθρώπινος νοῦς, μέσα ἀπὸ τὰ ἑκατομμύρια χρόνια ἐξέλιξης, ἀνέπτυξε μηχανισμοὺς εὐφυΐας/νοημοσύνης ποὺ χρησιμοποιεῖ ὄχι μόνο τὴ λογικὴ του ἀλλὰ ταυτόχρονα καὶ μέσα ἀπὸ μία διαδικασίᾳ, ποὺ ἀκόμη δὲν ἔχουμε ἐπιστημονικὰ κατανοήσει πλήρως, τὸ συναίσθημα καὶ τὰ αἰσθημάτά του (πέρα βεβαίως ἀπὸ τὰ ἔνστικτά του). Γιατὶ ὁ Ἄνθρωπος χρησιμοποιεῖ τὸ σῶμα (τὶς αἰσθήσεις του δηλαδή) γιὰ νὰ μαθαίνει. Γιατὶ ὁ Ἄνθρωπος δὲν εἶναι ἓνας «ἐγκέφαλος» σὲ μία μπουκάλα, ἀλλὰ ζεῖ μέσα στὸν Κόσμο.

Ὁ Καρτεσιανὸς δυϊσμός «νοῦς-σῶμα» ἔχει δώσει τὴ θέση του στὶς νέες θεωρίες, μὲ σκαπανεῖς τοὺς φιλόσοφους Merleau-Ponty, Polanyi, Dreyfus καὶ Todes, ποὺ ἔχουν ἐπιβεβαιωθεῖ καὶ ἀπὸ τὰ εὐρήματα στὸ πλαίσιο τῆς νευροεπιστήμης. Ἡ κίνησή μας στὸν τριδιάστατο (αὐτὸν δηλαδή ποὺ ἄμεσα ἀντιλαμβανόμαστε) κόσμο ποὺ μᾶς περιβάλλει, τὰ ἐργαλεῖα ποὺ χρησιμοποιοῦμε καὶ οἱ ἀντίστοιχες δεξιότητες ποὺ «μαθαίνουμε» καὶ ἀποτυπώνονται στὸν νοῦ, εἶναι ἀπόρροια συνεργασίας νοῦ καὶ σώματος/αἰσθήσεων. Πάνω ἀπ' ὅλα, δὲν μαθαίνουμε δεξιότητες μιμούμενοι ἀπλῶς, ὅπως κάνουν οἱ πίθηκοι. Δὲν μαθαίνουμε συγκεκριμένες κινήσεις καὶ δὲν ἀπομονώνουμε ἐρεθίσματα. Ἀντίθετα, μέσα ἀπὸ ἓνα ἐνιαῖο πλαίσιο, μέσα ἀπὸ τὸ Ὅλο, καὶ μὲ συνέργεια τῶν διαφορετικῶν μας αἰσθήσεων, μαθαίνουμε ἓναν γενικὸ τύπο ἐπίλυσης ποὺ ἀντιστοιχεῖ σὲ ἓναν γενικὸ τύπο προβλημάτων. Ἔτσι, μέσα ἀπὸ τὴ γενικότητα αὐτῆ τῆς ἐκμάθησης, μπορεῖ ὁ Ἄνθρωπος νὰ ἀντιμετωπίζει καὶ νὰ ἐπιλύει προβλήματα διαφορετικὰ ἀπὸ αὐτὰ ποὺ ἀποτελοῦν μέρος τῆς προτέρας του ἐμπειρίας. Μπορεῖ νὰ δημιουργεῖ καὶ νὰ μεταφέρει γνώση σὲ νέες καταστάσεις. Οἱ καταστάσεις αὐτὲς μπορεῖ νὰ διαφέρουν πολὺ ἢ μία ἀπὸ τὴν ἄλλη, καὶ νὰ διαφέρουν ἀκόμη καὶ στὶς ἐπιμέρους λεπτομέρειές τους, ἀρκεῖ νὰ ἔχουν ἓνα κοινὸ σημεῖο σημαντικότητας, ποὺ νὰ τὶς καθιστᾷ μέρος μιᾶς κοινῆς ἀντιμετώπισης. Μὲ ἄλλα λόγια, τὸ κύριο χαρακτηριστικὸ τῆς ἀνθρώπινης νοημοσύνης δὲν εἶναι ἡ (ἐπαγωγικὴ) πρόβλεψη μὲ βάση τὴν πρότερη ἐμπειρία ἀλλὰ ἡ δημιουργία, ποὺ χρησιμοποιεῖ τὴν πρότερη γνώση, ἀλλὰ ὄχι σὰν μίμηση παρὰ ὡς ἐφαλτήριο γιὰ νὰ γενικεύσει, νὰ νοηματοδοτήσῃ καὶ νὰ ἐξηγήσῃ.

Σε αντίθεση με την ανθρώπινη νοημοσύνη, θα λέγαμε ότι οι «μηχανές» που υλοποιούν σήμερα τεχνικές μηχανικής μάθησης είναι υπαρκτικά «ήλιθες». Στην ουσία, οι τεχνικές μηχανικής μάθησης είναι στατιστική. Μέσα από τα δεδομένα παρατηρήσεων μαθαίνουν τις στατιστικές ιδιότητες και μετά προβλέπουν. Για παράδειγμα, εάν έχουμε δει πολλές φορές γκρίζα σύννεφα και έχουμε ακούσει βροντές, προβλέπουμε ότι θα βρέξει. Όμοιως τα πουλιά. Είναι ακριβώς αυτό που κάνουν τα στατιστικά μοντέλα και τα όποια διδάσκουμε τους σπουδαστές μας επί δεκαετίες. Τί άλλαξε; Πρώτον, ο αριθμός των δεδομένων που χρησιμοποιούμε για να «έκπαιδευσουμε» τις «μηχανές», ο αριθμός δηλαδή των δεδομένων που μαθαίνουμε τις αντίστοιχες στατιστικές ιδιότητες, είναι άσυγκρίτως μεγαλύτερος από αυτόν που είχαμε, ως ποῦμε, 10-15 χρόνια πριν. Δεύτερον, η υπολογιστική ισχύς είναι πολύ μεγαλύτερη από αυτή που είχαμε 10-15 χρόνια πριν. Έτσι τώρα μπορούμε να χρησιμοποιούμε κάποια πιο πολύπλοκα μοντέλα για να μοντελοποιούμε τις στατιστικές ιδιότητες των δεδομένων. Η πεμπτουσία της μεθοδολογίας όμως παραμένει ή ίδια για δεκαετίες. Τα αποκαλούμενα νευρωνικά δίκτυα, που τα έχουμε εμπνευστεί από τη δομή με την οποία λειτουργεί ο εγκέφαλος (που παρεμπιπτόντως ή λειτουργία τους πολύ μικρή σχέση έχει με το πώς λειτουργεί ο εγκέφαλος), είναι μαθηματικά μοντέλα που δίνουν τη δυνατότητα να κωδικοποιούν με αποδοτικό τρόπο την πληροφορία που υπάρχει στα δεδομένα. Με άλλα λόγια, τα μοντέλα αυτά, μέσα από μία ιεραρχία αναπαραστάσεων και μαθηματικών μετασχηματισμών της πληροφορίας, διευκολύνουν την εξαγωγή των στατιστικών ιδιοτήτων, σε σχέση με τα παλαιότερα μοντέλα που χρησιμοποιούσαμε.

Έχοντας εκπαιδευτεί το στατιστικό μοντέλο, έχοντας δηλαδή μάθει τις στατιστικές ιδιότητες των δεδομένων, τότε όταν έχουμε κάποια νέα δεδομένα μπορούμε να κάνουμε προβλέψεις με κάποιες πιθανότητες. Έπειδή τα μοντέλα αυτά είναι ισχυρά και αρκετά γενικά, τα λάθη στις προβλέψεις είναι πολύ λίγα. Οι προβλέψεις αυτές έχουν έντυπωσιακά ποσοστά επιτυχίας. Αυτό βέβαια υπό την προϋπόθεση ότι η στατιστική δομή και πληροφορία που υπάρχει στα νέα δεδομένα, για τα όποια κάνουμε τις προβλέψεις, είναι πολύ ὅμοια με αυτή που υπάρχει στα δεδομένα πάνω στα όποια εκπαιδεύθηκε το δίκτυο. Όταν ένα δίκτυο προβλέπει ότι το αντικείμενο σε μία εικόνα είναι, για παράδειγμα,

μία γάτα, τὸ δίκτυο δὲν ἔχει δημιουργήσει κατὰ τὴ διάρκεια τῆς ἐκπαίδευσής του τὴν «ἔννοια» τῆς γάτας. Ἀπλῶς βασιζέται στὰ στατιστικὰ χαρακτηριστικὰ ποὺ ἔχει «μάθει» σχετικὰ μὲ τὴ γάτα. Χάριν ἀπλουστεύσεως, ἄς ποῦμε ὅτι τέτοια χαρακτηριστικὰ εἶναι τὸ μῆκος τῆς, τὸ χρῶμα τῆς, ἡ ὑφή τοῦ τριχώματός τῆς, τὰ τέσσερα πόδια τῆς, ἡ οὐρά τῆς, τὸ σχῆμα τῆς μύτης τῆς, κ.λπ. Γιὰ νὰ ἐκπαιδευθεῖ ἓνα δίκτυο ἀπαιτεῖ χιλιάδες δεδομένα ἐκπαίδευσης, δηλαδὴ εἰκόνες μὲ γάτες, μὲ διαφορετικὰ χρώματα, σὲ διαφορετικὲς στάσεις, μὲ διαφορετικὰ μεγέθη, κ.λπ. Σὲ ἀντίθεση, ἓνα παιδί μαθαίνει νὰ ἀναγνωρίζει μία γάτα μόλις δεῖ μία, ἄντε δύο. Ἀμέσως δημιουργεῖ τὴν «ἔννοια» τῆς γάτας καὶ πάει καὶ τελειώσε. Κάτι ἀνάλογο γίνεται καὶ μὲ τὴ μετάφραση ἀπὸ μία γλῶσσα σὲ ἄλλη. Μέσα ἀπὸ ἑκατομμύρια κείμενα μεταφρασμένα ἀπὸ τὴ μία γλῶσσα στὴν ἄλλη, ἡ «μηχανή» μαθαίνει τὴν πιθανότητα ἢ ἐπόμενη λέξη νὰ εἶναι, ἄς ποῦμε, «νερό», ὅταν οἱ προηγούμενες λέξεις ἦταν «τὸ ποτήρι εἶναι γεμάτο μὲ». Σὲ ἀντίθεση, ὁ Ἴανθρωπος ἀφ' ἑνὸς καταλαβαίνει τὸ «νόημα» τῆς κάθε λέξης καὶ ἀφ' ἑτέρου ἔχει τὴν ἰκανότητα νὰ ἐστιάζει στὰ πιὸ σημαντικὰ μέρη μιᾶς πρότασης, ἔχοντας πρῶτα ἀντιληφθεῖ τὸ «ὄλο» νόημά τῆς, περὶ τίνος δηλαδὴ πρόκειται στὴ φράση.

Μὲ ἄλλα λόγια, ἡ τεχνολογία τῆς μηχανικῆς μάθησης σήμερα μᾶς θυμίζει τὸν μαθητὴ ἐκεῖνο ποὺ δὲν ἔχει καταλάβει τίποτε, ἀλλὰ λόγῳ τῆς πολὺ καλῆς του μνήμης ἔχει ἀποστηθίσει τὰ πάντα. Ἔτσι, ἂν τὸν ρωτήσουν κάτι ἀπὸ αὐτὰ ποὺ ἔχει ἀποστηθίσει, ἢ κάτι πολὺ κοντὰ σὲ αὐτὰ, τότε ἀπαντᾷ σωστά (πρόβλεψη μὲ βάση τὴ στατιστικὴ). Ἐὰν ἡ ἐρώτηση εἶναι λίγο διαφορετικὴ, ἢ οἱ λέξεις μὲ τις ὁποῖες αὐτὴ διατυπώνεται εἶναι διαφορετικὲς ἀπὸ τις λέξεις ποὺ ἔχει αὐτὸς ἀποστηθίσει, τότε δίνει λανθασμένη ἀπάντηση. Τὸ ἴδιο ἰσχύει καὶ μὲ τις τεχνικὲς μηχανικῆς μάθησης. Μπορεῖ νὰ ἀναγνωρίζουν ἀντικείμενα σὲ εἰκόνες καὶ πρόσωπα μὲ ἐκπληκτικὴ ἐπιτυχία, ἀλλὰ ταυτόχρονα ἂν ἀλλάξουμε λίγο τὰ δεδομένα, μὲ τρόπο ποὺ κανένας Ἴανθρωπος δὲν θὰ μπορούσε νὰ ἀντιληφθεῖ μία τέτοια ἀλλαγὴ, τότε ἀντὶ γιὰ «γάτα» ἢ «μηχανή» μπορεῖ νὰ κάνει τὴν πρόβλεψη, ἄς ποῦμε, «πολυκατοικία» (αὐτὰ εἶναι τὰ λεγόμενα στὴ μηχανικὴ μάθηση «ἀντίπαλα» παραδείγματα-“*adversarial examples*”). Ὁ Ἴανθρωπος χειρίζεται ἔννοιες καὶ νοήματα καὶ οἱ μηχανεὲς στατιστικὴ χωρὶς κατανόηση.

Ἱστορικά, ἡ περιοχὴ τῆς τεχνητῆς νοημοσύνης, κάπου ἐκεῖ στὴ δεκαετία τοῦ 1950, ξεκίνησε φιλόδοξα. Ἦταν τότε ποὺ ὁ Turing ἔθεσε τὸ



ἐρώτημα: «Μποροῦν οἱ μηχανές νὰ σκεφτοῦν;». Τότε ἦταν ποῦ ἄρχισαν νὰ ὀραματίζονται καὶ νὰ κατασκευάζουν μηχανές ποῦ νὰ μποροῦν νὰ σκέπτονται, μηχανές βασισμένες σὲ συντακτικούς κανόνες καὶ λογικὴ ἐπεξεργασία συμβόλων. Ἄν καὶ τὸ ἔναυσμα δόθηκε ἀπὸ τοὺς πρώτους τότε ὑπολογιστὲς ποῦ κατασκευάσθηκαν καὶ τὶς θεωρίες τοῦ Turing, τὸ ὄραμα καὶ ὁ στόχος αὐτὸς ἀποτελοῦσαν συνέχεια μιᾶς φιλοσοφικῆς παράδοσης αἰώνων στὸ πλαίσιο τῆς δυτικῆς φιλοσοφίας ἀπὸ τὸν Πλάτωνα καὶ ἐντεῦθεν. Εἶναι ἡ τάση νὰ ἐρμηνευθεῖ ὁ «νοῦς» σὰν μία ἀλγοριθμικὴ διαδικασία ποῦ ἐπεξεργάζεται σύμβολα, ἡ θεωρία δηλαδὴ ὅτι ὅλη ἡ σκέψη δὲν εἶναι τίποτε παραπάνω ἀπὸ πράξεις πάνω σὲ καλῶς ὀρισμένα σύμβολα/ἔννοιες στὸ πλαίσιο ἐνὸς κατάλληλα ὀρισμένου μαθηματικοῦ φορμαλισμοῦ. Κάτι ἀνάλογο μὲ τοὺς νόμους τῆς φυσικῆς, ποῦ μποροῦν νὰ ἐκφραστοῦν μέσα ἀπὸ μαθηματικὰ μοντέλα καὶ σύμβολα, ἔτσι καὶ ἡ ἀνθρώπινη σκέψη θὰ μποροῦσε νὰ διατυπωθεῖ μέσα ἀπὸ συντακτικούς λογικούς κανόνες. Χαρακτηριστικά, ὁ Hobbes εἶχε πεῖ ὅτι «ἡ σκέψη (*reason*) δὲν εἶναι τίποτε ἄλλο ἀπὸ ὑπολογισμός (*reckoning*)». Ἀργότερα, ὁ Leibnitz προσπάθησε νὰ ἀναπτύξει ἓνα σύστημα συμβολικῆς γλώσσας ὅπου κάθε ἔννοια (*concept*) θὰ εἶχε τὸν πλήρη ὀρισμὸ τῆς καὶ τὸ σύμβολό τῆς καὶ θὰ μποροῦσε νὰ ἐκφρασθεῖ σὰν συνάρτηση κάποιων βασικῶν ἐννοιῶν, καὶ ἔτσι ὅλη ἡ γνώση θὰ μποροῦσε νὰ ἀναπαραχθεῖ «ἀντικειμενικά». Ἡ φιλοσοφικὴ αὐτὴ γραμμὴ φθάνει στὸ ἀπόγειό τῆς μὲ τὸν λογικὸ ἀτομισμὸ τοῦ Russell καὶ τὸν πρώιμο Wittgenstein, ὁ ὁποῖος στὸ κλασικὸ πόνημά του *Tractatus* θεωρεῖ ὅτι ἡ γνώση συνίσταται ἀπὸ «ἀτομικά» γεγονότα ποῦ μποροῦν νὰ ἐκφραστοῦν μὲ τὴ βοήθεια ἀνεξάρτητων λογικῶν προτάσεων. Σύμφωνα μὲ αὐτὴ τὴ φιλοσοφικὴ τάση, ὁ ἐγκέφαλος δὲν ἦταν παρὰ ἓνας ἀλγόριθμος. Ὁ Minsky, ἀπὸ τοὺς πρωτοπόρους τῆς TN, παρομοίωσε τὸν ἐγκέφαλο μὲ «κρεατομηχανή» (*meat machine*).

Παρὰ τὶς ἀρχικὲς ἐντυπωσιακὲς ἐπιτυχίες, ποῦ σημειώθηκαν κατὰ τὶς δεκαετίες τοῦ '50 καὶ '60, ὁ κλάδος αὐτὸς τῆς TN δὲν προχώρησε. Σήκωνε τὰ χέρια ψηλὰ κάθε φορά ποῦ τὰ προβλήματα γίνονταν πιὸ πολύπλοκα. Δύο εἶναι οἱ κύριοι λόγοι:

Πρῶτον, ἀκόμη καὶ ἐὰν ἡ διαδικασία τῆς σκέψης θὰ μποροῦσε νὰ τυποποιηθεῖ καὶ νὰ ἀναχθεῖ σὲ διαχείριση λογικῶν συντακτικῶν κανόνων στὸ πλαίσιο μιᾶς τυπικῆς γλώσσας, ὁ ἀριθμὸς τῶν ἀπαιτούμενων συνδυασμῶν καὶ λογικῶν σχέσεων ποῦ θὰ ἀπαιτοῦνταν

για να φθάσουμε σε μία απόφαση ξεφεύγει από τις δυνατότητες που έχουν οι ψηφιακοί υπολογιστές. Μικρής μόνο κλίμακας προβλήματα μπορούν να επιλυθούν. Αυτό οφείλεται στην πολυπλοκότητα της πραγματικότητας και του μεγάλου αριθμού πιθανών ενδεχομένων στη ζωή. Είναι σαν να ζητούμε από έναν σκακιστή να υπολογίσει όλους τους δυνατούς συνδυασμούς όλων των μελλοντικών κινήσεων. Αυτό όμως είναι αδύνατον. Ένας σκακιστής μόνον έναν σχετικά μικρό συνδυασμό μελλοντικών κινήσεων μπορεί να υπολογίσει: οι καλοί σκακιστές περισσότερες και οι λιγότερο έμπειροι/καλοί λιγότερες. Άλλωστε και οι υπολογιστές που παίζουν σκάκι, κάθε άλλο παρά τη συνδυαστική αυτή μέθοδο ακολουθούν. Πολύ σύντομα, μετά από ένα σχετικά μικρό αριθμό μελλοντικών κινήσεων, ο απαιτούμενος αριθμός υπολογισμών θα χρειάζονταν τόσο χρόνο, όσο είναι και η ζωή που έχει το σύμπαν μας!

Δεύτερον, οι έννοιες δεν μπορούν να περιγραφούν με διακριτά σύμβολα. Κάθε έννοια αποκτά τη σημασία της μέσα από τα συμφραζόμενά της. Δεν έχει κάποια αντικειμενική υπόσταση. Δεν υπάρχουν ακριβείς όρισμοί. Φιλοσοφικά αυτό προβληματίσε τον Πλάτωνα και τον Άριστοτέλη και είναι γνωστό ήδη από τα χρόνια εκείνα.

Ο άνθρωπος νους αντιμετωπίζει την πρώτη δυσκολία, που σχετίζεται με τη διαχείριση μεγάλου αριθμού δυνατών ενδεχομένων, με το να αποφασίζει για το τί είναι «σημαντικό» ή όχι σε κάθε συγκεκριμένη κατάσταση. Ο μηχανισμός της προσοχής/έστιασης επιστρατεύεται ώστε να φωτίσει την πιο σημαντική πληροφορία. Από την πληθώρα των πιθανών ενδεχομένων αποκλείεται, συνειδητά αλλά και ασυνείδητα, ένας μεγάλος αριθμός από αυτά. Ο αποκλεισμός ενδεχομένων είναι αποτέλεσμα νευρο-βιολογικών διεργασιών που εξελίσσονται στον εγκέφαλο, διεγείροντας ή απενεργοποιώντας νευρώνες σε διάφορες περιοχές του, και όπου αισθήσεις, αισθήματα, συναισθήματα, συστήματα αξιών, εμπειρίες αποτυπωμένες στη ενεργό μνήμη αλλά και στο υποσυνείδητο, και ό,τι άλλο συνιστά αυτό που ονομάζουμε παρούσα κατάσταση του νου συνεργάζονται, διεγείροντας και τις αντίστοιχες φυσικές αναπαραστάσεις στους νευρώνες του εγκεφάλου, από τις οποίες ή παρούσα κατάσταση του νου καθορίζεται. Με άλλα λόγια, το ανώτερο εξελικτικά τμήμα του εγκεφάλου (νεοφλοιός), που

διεγείρεται με τις διαδικασίες σκέψης και αντίληψης (*rational*) είναι σε πλήρη συνεργασία με το πιο πρωτόγονο, μη λογικό (*irrational*) τμήμα του (ύποθάλαμος, έγκεφαλικό στέλεχος και μεταιχμιακό σύστημα), για να αποκλειστούν ένδεχόμενα και έτσι να μπορούμε να κάνουμε συλλογισμούς πάνω σε ένα υποσύνολο πιθανών ένδεχομένων, ώστε τελικά να μπορούμε να φθάσουμε σε μία απόφαση σε μικρό χρόνο. Συχνά μιλούμε για «διαίσθηση» για να εκφράσουμε το σύνολο όλων αυτών των ασύνειδων διεργασιών που βοηθεί τη σκέψη στο να φθάνει σε αποφάσεις, που είναι και το κύριο ζητούμενο της σκέψης.

Όπως αναφέρει ο Damasio στο σχετικό βιβλίο του, μετά από την καταστροφή τμημάτων του έγκεφάλου που σχετίζονται με τα συναισθήματα, οί ασθενείς χάνουν την ικανότητά τους να αποφασίζουν. Μείωση των συναισθηματικών ικανοτήτων σχετίζεται και με έκδηλώσεις μη-λογικά έρμηνεύσιμων συμπεριφορών. Άλλωστε, το γνωρίζουμε όλοι μας και από την προσωπική μας έμπειρία. Όταν είμαστε συναισθηματικά φορτισμένοι, οί λογικοί συνειρμοί και οί αποφάσεις που έπονται δεν ακολουθούν την ίδια «διαδρομή» με αυτήν που θα ακολουθούσαν, εάν ήμασταν σε διαφορετική ψυχική κατάσταση. Επίσης, οί έμπειρίες και οί εκάστοτε προσωπικές αξίες και τα κοινωνικά στερεότυπα καθορίζουν σε μεγάλο βαθμό το πώς θα αντιδράσουμε και τί αποφάσεις θα πάρουμε κατά περίπτωση. Αυτό που συμβαίνει στον έγκέφαλο είναι ότι όλα αυτά ρυθμίζουν αυτόματα (ασύνειδα) τα δίκτυα αλληλεπίδρασης των νευρώνων έτσι ώστε να περιορίζουν τον χώρο/άριθμο των πιθανών ένδεχομένων που δυνητικά θα έπρεπε να εξετάσουμε, έτσι ώστε να μπορούμε να φθάσουμε σε μία απόφαση.

Τίποτε από τα παραπάνω δεν θυμίζει «άλγόριθμο». Ένας αλγόριθμος συνίσταται από προκαθορισμένα λογικά βήματα που δρούν πάνω σε καλώς όρισμένα σύμβολα. Όσοι λέγουν ότι ο έγκέφαλος είναι ένας αλγόριθμος, ή δεν γνωρίζουν τί είναι αλγόριθμος ή/και πώς λειτουργεί ο έγκέφαλος. Οί φυσικο/χημικές διεργασίες, που λαμβάνουν χώρα στον έγκέφαλο, δεν είναι αλγοριθμικές. Και βέβαια, όταν μιλούμε για φυσικο/χημικές διεργασίες, δεν μιλούμε για τον νοῦ αλλά για το φυσικό υπόστρωμα που συνίσταται από τους νευρώνες. Ο νοῦς είναι ή ροή υποκειμενικών έμπειριών, είναι σύλληψη δημιουργίας και φαντασίας. Ο έγκέφαλος είναι δισεκατομμύρια νευρώνων, συνάψων

καὶ ἠλεκτροχημικῶν διεργασιῶν. Ὁ νοῦς εἶναι συνείδηση. Τὰ τελευταῖα εἴκοσι χρόνια ἔχουν γίνει σημαντικὲς ἔρευνες σχετικὰ μὲ τὴ λειτουργία τοῦ ἐγκεφάλου καὶ σήμερα γνωρίζουμε πολὺ περισσότερα ἀπὸ ὅ,τι μόλις πρὶν ἀπὸ λίγα χρόνια. Μέχρι στιγμῆς ὅμως δὲν ἔχουμε τὴν παραμικρὴ ιδέα πῶς ἐγκέφαλος καὶ νοῦς ἀλληλοεπιδρῶν. Σύμφωνα μὲ τὸν φιλόσοφο Chalmers, ἡ μελέτη τοῦ ἐγκεφάλου εἶναι τὸ «εὐκόλο» πρόβλημα. Τὸ «δύσκολο» πρόβλημα εἶναι νὰ καταλάβουμε πῶς αὐτὸ τὸ σύμπλεγμα ἀλληλοεπιδρῶντων νευρῶνων καὶ ἀνταλλαγῆς ἰόντων σχετίζεται μὲ αὐτὸ ποὺ ὀνομάζουμε νοῦ καὶ συνείδηση. Ἡ ἐπιστήμη δὲν ἔχει δώσει ἀκόμη ἀπαντήσεις καὶ οἱ φιλόσοφοι ἐρίζουν ὡς πρὸς τὴν ἐρμηνεία.

Τὸ δεύτερο σκέλος, τὸ ὁποῖο διαφοροποιεῖ τὴν ἀνθρώπινη σκέψη ἀπὸ τὴ φιλοσοφικὴ ἄποψη ποῦ «ἔβλεπε» τὸν νοῦ σὰν ἀλγόριθμο, εἶναι ὅτι ὁ ἀνθρώπινος νοῦς δὲν χειρίζεται διακριτὰ καὶ καλῶς ὀρισμένα σύμβολα ἀλλὰ ἀσαφεῖς ἔννοιες. Ποιὸς ἄραγε μπορεῖ νὰ ὀρίσει τίς ἔννοιες τῆς «εὐτυχίας», τῆς «ἐλευθερίας» ἢ τῆς «εὐσέβειας»; Ἡ τελευταία ἔννοια καὶ ὁ ὀρισμὸς της ἀπασχολεῖ τὸν Σωκράτη στὸν *Εὐθύφρονα* τοῦ Πλάτωνος. Κατὰ τὴ διάρκεια τῶν ἑκατομμυρίων χρόνων ἐξέλιξης, ὁ ἐγκέφαλος/νοῦς ἀνέπτυξε στρατηγικὲς γιὰ νὰ μπορεῖ νὰ διαχειρίζεται τίς ἀσάφειες. Τὸ «ὄλον» νοηματοδοτεῖ τὸ ἐπιμέρους καὶ δὲν δομεῖται ἀπὸ ἐπιμέρους ἀνεξάρτητα τμήματα. Σύμφωνα μὲ τὸν φιλόσοφο Husserl καὶ τὴ φαινομενολογικὴ του προσέγγιση, ὁ τρόπος ἀντίληψης σύνθετων προτύπων εἶναι μέσα ἀπὸ τὴ δημιουργία ἐνὸς ἀσαφοῦς καὶ μὴ καλὰ ὀρισμένου «ὄλου» ἀρχικά, ποὺ στὴ συνέχεια συμπληρώνεται σταδιακὰ μὲ τίς λεπτομέρειές του, οἱ ὁποῖες ἀντλοῦνται ἀπὸ τὴν πρότερη ἐμπειρία μας καὶ τίς μνήμες μας. Μία λέξη (μέρος) ἀποκτᾶ τὸ νόημά της στὸ πλαίσιο μιᾶς φράσης (ὄλον) καὶ τῆς γενικώτερης πρότερης ἐμπειρίας μας. Αὐτὸ τὸ γνωρίζουν πολὺ καλὰ οἱ γλωσσολόγοι. Στὴ μουσικὴ κάθε νότα ἀποκτᾶ τὴ μουσικὴ ἀξία της μέσα ἀπὸ τὸ σύνολο τῆς μελωδίας. Καὶ ὅπως μᾶς λέει ὁ Wittgenstein, «ἓνα στόμα χαμογελάει μόνο στὸ πλαίσιο ἐνὸς προσώπου». Μὲ ἄλλα λόγια, ἡ ἐρμηνεία τοῦ «μέρους» ἐξαρτᾶται ἀπὸ τὸ «ὄλον» ἐντὸς τοῦ ὁποῖου ὑπάρχει. Ἐν ἀντιθέσει, σὲ ἓναν ψηφιακὸ ὑπολογιστὴ ποὺ ἡ λειτουργία του βασιζέται στὴ χρῆση τυπικῶν γλωσσῶν, τὸ «ὄλον» κατασκευάζεται ἀπὸ τὰ ἐπιμέρους μὲ τὴ βοήθεια συνδυασμοῦ λογικῶν πράξεων καὶ συντακτικῶν κανόνων.

Μὲ τὴν προηγούμενη συζήτηση προσπαθήσαμε νὰ καταστήσουμε σαφὲς ὅτι αὐτὸ ποὺ ὀνομάζουμε ἀνθρώπινη νοημοσύνη, πρὸς τὸ παρόν,

οὐδεμία σχέση ἔχει μὲ αὐτὸ ποὺ ὀνομάζουμε τεχνητὴ νοημοσύνη καὶ πολὺ περισσότερο μὲ τὸν κλάδο τῆς ΤΝ, ποὺ εἶναι γνωστὸς ὡς μηχανικὴ μάθηση καὶ ποὺ σήμερα καθοδηγεῖ τὶς ἐξελίξεις, ἔστω καὶ ἂν φαινομενικὰ μπορεῖ νὰ μοιάζουν καὶ νὰ ἐπιτελοῦν παρόμοιες λειτουργίες. Ἡ μηχανικὴ μάθηση δανεῖζεται τὴ δύναμή της ἀπὸ τὴν τεχνολογία τῶν ὑπολογιστῶν, ὥστε νὰ ἀποθηκεύονται τεράστιες ποσότητες δεδομένων καὶ νὰ ἐκτελοῦνται ἀριθμητικοὶ ὑπολογισμοὶ πολὺ γρήγορα. Ἐδῶ ὁ Ἄνθρωπος ὑστερεῖ, ὅπως ὑστερεῖ καὶ στὴ δύναμη τοῦ χεριοῦ του σὲ σχέση μὲ μία μπουλντόζα. Ἀπὸ τὴν ἄλλη, ἢ ἀνθρώπινη νοημοσύνη χτίζεται πάνω στὴν ἀφαιρετικὴ δύναμη τῶν ἐννοιῶν καὶ νοημάτων καὶ τὴ συνεκτικότητα ποὺ τοὺς παρέχει ἢ ἔμφυτη ἐξελικτικὰ δημιουργικὴ ὁρμή. Οἱ ὑπολογιστὲς εἶναι μὴ νοήμονα ἀντικείμενα. Ἡ ἔμφαση στοὺς ὑπολογιστὲς εἶναι στὸ ἐπιμέρους, στὴ λεπτομέρεια, στὸ ἐξειδικευμένο, στὰ συγκεκριμένα δεδομένα πάνω στὰ ὁποῖα βελτιστοποιοῦν τὴν ἀπόδοσή τους. Οἱ ὑπολογιστὲς βασίζονται σὲ ἀλγόριθμους ποὺ βελτιστοποιοῦν μὲ ἔμφαση τὰ ὑπάρχοντα δεδομένα, δηλαδὴ τὸ «παρελθόν». Βασίζονται σὲ «γεγονότα». Στὴν ἀνθρώπινη νοημοσύνη, ἢ ἔμφαση εἶναι σὲ ἓνα συνεχῶς μεταβαλλόμενο «ὄλον». Ὅπως μᾶς λέγει χαρακτηριστικὰ ὁ Heidegger, ὁ «ἀπώτερος σκοπὸς μιᾶς ἀπόφασης στὸν Ἄνθρωπο δὲν εἶναι νὰ πετύχει ἓναν στόχο ποὺ σηματοδοτεῖ τὸ τέλος μιᾶς πορείας. Κάθε στιγμή ἀναπροσαρμόζουμε τὸ σύνολο τῆς ἐμπειρίας μας καὶ σχεδιάζουμε τὶς ἐνέργειές μας ἐστιάζοντας καὶ ἐπιλέγοντας σύμφωνα μὲ τὴν παροῦσα κατάσταση καὶ τὶς ἐκάστοτε ἀνάγκες μας». Ὁ Ἄνθρωπος δὲν εἶναι γεγονὸς ἢ σύνολο γεγονότων. Ὁ Ἄνθρωπος δημιουργεῖ τὰ γεγονότα καὶ τὸν ἑαυτὸ του ζώντας μέσα στὸν Κόσμο. Ὅπως χαρακτηριστικὰ μᾶς λέει ὁ Piaget, ἡ νοημοσύνη μας δημιουργεῖται μέσα ἀπὸ «ἐννοιολογικὲς (conceptual) ἐπαναστάσεις».

Ἡ διαφορὰ μεταξὺ ἀνθρώπινης δημιουργίας καὶ ὑπολογιστικοῦ μοντέλου εἶναι ὅτι αὐτὸ ποὺ εἴμαστε «σήμερα» δὲν θὰ μπορούσε νὰ προβλεφθεῖ ἀπὸ αὐτὸ ποὺ ἤμασταν «χθές». Σὲ ἀντίθεση μὲ τὰ μοντέλα τῆς μηχανικῆς μάθησης, τὸ «σήμερα» μπορεῖ νὰ προβλεφθεῖ ἀπὸ τὸ «χθές». Ἐδῶ ὁ ὑπολογιστὴς ὑστερεῖ. Θὰ μπορούσαμε νὰ ποῦμε ὅτι ἡ ἔννοια τοῦ «παρόντος» ἔχει μία πολὺ διαφορετικὴ σημασία στὸν Ἄνθρωπο καὶ τὴ «μηχανή». Στίς «μηχανές» τὸ παρὸν ἀποκτᾶ τεράστια σημασία γιατί βελτιστοποιοῦμε τὰ μοντέλα μας, τώρα/παρόν, γιὰ ἓναν συγκεκριμένο στόχο. Ὁ Ἄνθρωπος, ὅπως μᾶς εἶπε καὶ ὁ Pascal, ξεχνᾷ

τό «παρόν» και, ὅποτε τὸ θυμᾶται, εἶναι γιὰ νὰ δεῖ τὴν ἀχτίδα τοῦ φωτὸς ποὺ αὐτὸ ἀφῆνει γιὰ νὰ φωτίσει τὰ σχέδιά μας γιὰ τὸ Μέλλον.

Ἐνα ἐρώτημα ποὺ τίθεται εἶναι τὸ κατὰ πόσον θὰ μπορούσαν οἱ «μηχανές» νὰ ἀποκτήσουν νοημοσύνη ὅπως αὐτὴν τοῦ Ἄνθρώπου. Ἡ ἀπάντησή μας εἶναι: ἴσως. Καὶ αὐτὴ τὴ στιγμή ὑπάρχει ἔρευνα καὶ ἐπενδύσεις πρὸς αὐτὴν τὴ κατεύθυνση, ποὺ μὲ κάπως πιὸ τεχνικούς ὄρους ὀνομάζεται Ἰσχυρὴ Τεχνητὴ Νοημοσύνη (*Strong AI*). Προϋπόθεση βέβαια γιὰ κάτι τέτοιο εἶναι νὰ καταλάβουμε πῶς πράγματι λειτουργεῖ ἡ ἀνθρώπινη νοημοσύνη. Γνωρίζουμε κάποια χαρακτηριστικά της καὶ προϋποθέσεις. Δηλαδή μνήμη, συνείδηση, νόημα. Ἀλλὰ δὲν γνωρίζουμε πῶς ἀκριβῶς αὐτὰ δημιουργοῦνται καὶ πῶς λειτουργοῦν. Πῶς δηλαδή ἀπὸ τοὺς φυσικοὺς νευρῶνες καὶ τὶς βιοχημικὲς διεργασίες ποὺ λαμβάνουν χώρα στὸν ἐγκέφαλο χτίζεται ἡ συνείδηση στὸν νοῦ, πῶς δημιουργοῦνται οἱ ἔννοιες, οἱ ἰδέες καὶ τὰ νοήματα, πῶς δημιουργεῖται αὐτὸ ποὺ ἀποκαλοῦμε κατανόηση. Ἄν μᾶς ἐπιτρέπεται ἡ ἀναλογία, γνωρίζουμε πολὺ καλὰ τὴ χημεία τοῦ DNA καὶ τῶν γονιδίων ποὺ συνιστοῦν στοιχεῖα τῆς Ζωῆς. Ἀλλὰ δὲν ἔχουμε δημιουργήσει Ζωή. Πρὸς τὸ παρόν, ὅ,τι νοημοσύνη ὑπάρχει στὶς μηχανές ἔχει ἀναπτυχθεῖ ἀπὸ τὸν Ἄνθρωπο. Εἶναι μέρος τῆς νοημοσύνης καὶ τῆς δημιουργικότητας τοῦ Ἄνθρώπου ποὺ ἔχει «ἐμφυτευθεῖ» στὴ «μηχανή», ἡ ὁποία ἀπλῶς ἐκτελεῖ τὴν ἀριθμητικὴ – τὶς πράξεις ἂν θέλετε. Κατὰ τὴν ἄποψή μας, θὰ μπορούσαμε νὰ μιλήσουμε γιὰ «μηχανές» μὲ νοημοσύνη, μόνον ἂν αὐτὲς θὰ μπορούσαν νὰ παραξοῦν νέα γνώση κατὰ τὴ διάρκεια τῆς ἐκτέλεσης τοῦ προγράμματος καὶ ὄχι μόνον κατὰ τὴ διαδικασία ἀνάπτυξης τοῦ προγράμματος ἀπὸ τὸν προγραμματιστή. Ταυτόχρονα, ἡ «μηχανή» θὰ πρέπει νὰ εἶναι σὲ θέση νὰ «ἐξηγεῖ» τὸ γιατί καὶ τὸ πῶς. Ἡ ἀπλὴ πρόβλεψη δὲν συνιστᾷ νοημοσύνη, ἂν δὲν συνοδεύεται ἀπὸ κάποια ἐξήγηση. Βελτιώνοντας τίς «μηχανές» ὥστε νὰ προσποιοῦνται ὅτι σκέφτονται, δὲν εἶναι τὸ ἴδιο μὲ τὸ νὰ σκέφτονται.

Ἀπὸ ὅλα τὰ παραπάνω, ἴσως κάποιοι ὑποθέσουν ὅτι προσπαθοῦμε νὰ ὑποβαθμίσουμε τὰ ἐπιτεύγματα τῆς Τεχνολογίας. Κάθε ἄλλο. Τὰ ἐπιτεύγματα εἶναι τεράστια καὶ ἀξιοθαύμαστα, ὄχι μόνον ἐξ ἀπόψεως ἐφαρμογῶν ἀλλὰ καὶ ἀπὸ τὴν ἄποψη τῆς ἐπιστημονικῆς δημιουργίας. Αὐτὸ ποὺ θέλουμε νὰ τονίσουμε εἶναι ὅτι ἡ τεχνολογία αὐτὴ εἶναι –ἢ τουλάχιστον πρέπει νὰ εἶναι– ὅ,τι καὶ οἱ προηγούμενες, ἓνα δηλαδή

Άλλο «έργαλειο» στην ύπηρεσία του Άνθρώπου και των αναγκών του. Είναι ένα επιπλέον έργαλειο στην ιστορική δημιουργική διαδρομή του που λέγεται Πολιτισμός. Και όπως κάθε νέα τεχνολογία, δημιουργεί και ένα νέο περιβάλλον στο οποίο ο Άνθρωπος θα πρέπει να λειτουργήσει και να ζήσει. Στο νέο αυτό περιβάλλον θα δημιουργήσουμε τις νέες μας αξίες και τις «σημασίες» μας, το νέο μας αφήγημα, την υπαρξιακή μας σφαίρα, όπως θα μάς έλεγε ο Kierkegaard.

Αυτό που επιχειρήσαμε να τονίσουμε με όσα προαναφέρθηκαν είναι ότι ο κίνδυνος δεν είναι από τη νοημοσύνη των «μηχανών» που θα ξεπεράσει τη δική μας. Ο κίνδυνος είναι να υποβιβάσουμε τη δική μας νοημοσύνη στο αλγοριθμικό/στατιστικό επίπεδο της τεχνητής νοημοσύνης και μηχανικής μάθησης. Κάποιοι πρόγονοί μας πίστευαν ότι τα τότε είχαν ζωσα ενέργεια και υπεράνθρωπες δυνατότητες και προσεύχονταν σε αυτά. Εάν αποδώσουμε χαρακτηριστικά υπεράνθρωπης νοημοσύνης στις «μηχανές» που εμείς κατασκευάζουμε, αυτό θα οφείλεται στο ότι εμείς υποβαθμίσαμε την αξία της δικής μας «ύπαρξης». Με άλλα λόγια, ο κίνδυνος είναι από «υπάρξεις» δημιουργίας να υποβιβασθούμε σε «αλγορίθμους» ή υπορουτίνες κάποιων αλγορίθμων. Ο κίνδυνος είναι να πιστέψουμε ότι είμαστε «κρεατομηχανές» και να ταυτίσουμε τον «νοῦ» με δισεκατομμύρια νευρώνες και τρισεκατομμύρια συνάψεις που διεγείρονται από ηλεκτροχημικές διεργασίες, αυτό δηλαδή που είναι ο «έγκεφαλος», και να αφήσουμε να παραλύσει αυτό που είναι πράγματι ο νοῦς, δηλαδή η ροή των υποκειμενικών ἐμπειριῶν, ὁ θυμός, ἡ λύπη, ἡ ἀγάπη, ὁ φόβος, ἡ χαρά, ἡ ἀπόλαυση, ὁ φθόνος –ναί, ὅλα αὐτὰ μαζί– και ταυτόχρονα συνείδηση, συναίσθηση και σύλληψη δημιουργίας και φαντασίας.

Ἐκφράζουμε τὸν φόβο πὼς ἡ τάση τὰ τελευταῖα δέκα χρόνια εἶναι πρὸς τὴν ἀλγοριθμοποίησή μας· τὴ μετάλλαξή μας ἀπὸ προγραμματιστὲς σὲ προγραμματιζόμενους «έγκεφάλους». Ἀπὸ χρῆστες πὸ ἐλέγχουμε τὴν τεχνολογία, μεταμορφωνόμαστε, ὅλο και περισσότερο, σὲ ἀντικείμενα πὸν χρησιμοποιεῖ ἡ Τεχνολογία στὸ νέο περιβάλλον, μέσα στὸ ὁποῖο ὑπάρχουμε, εἴμαστε συνδεδεμένοι (online) 365 μέρες τὸν χρόνο και παρέχουμε συνεχῶς δεδομένα/πληροφορία γιὰ τὶς συνήθειές μας και τὴν προσωπική μας ζωή. Τὰ δεδομένα αὐτὰ χρησιμοποιοῦνται γιὰ νὰ μάς διαμορφώσουν σύμφωνα μὲ μία εἰκόνα πὸν τὰ δεδομένα και οἱ

αλγόριθμοι προβλέπουν για μᾶς. Οἱ πληροφορίες πού λαμβάνουμε, τὰ νέα πού βλέπουμε, ἡ μουσική πού μᾶς προτείνεται διαφέρουν για τὸν καθένα μας. Ἐξατομικευόμαστε, ἀποκοινωνικοποιούμαστε καὶ ἔτσι εὐκολώτερα χαλιναγωγούμαστε. Ἐκπαιδευόμαστε για νὰ εἴμαστε παθητικοὶ καταναλωτὲς ἀντὶ για ἐνεργοὶ πολῖτες.

Μὲ ἀφετηρία τὰ ὅσα εἰπώθηκαν παραπάνω, ἡ δύναμη καὶ τὸ πλεονέκτημα τοῦ Ἀνθρώπου σὲ σχέση μὲ τὶς μηχανὲς εἶναι ἡ δημιουργικότητά του καὶ ἡ ικανότητά του νὰ ἐξηγεῖ καὶ νὰ νοηματοδοτεῖ τὸν Κόσμο μὲ σημασίες. Ὅχι νὰ ἀντιδρᾷ προβλέψιμα καὶ ἀ-νόητα. Εἶναι αὐτὰ τὰ στοιχεῖα πού θὰ πρέπει νὰ δημιουργήσουν τὸν νέο «μῦθο» τῆς νέας ἐποχῆς. Καὶ ἡ δημιουργικότητα ἔχει δύο πλευρές: τὴν ικανότητα κριτικῆς καὶ ἐλεύθερης σκέψεως ἀπὸ τὴ μία καὶ τὴν κοινωνικότητα ἀπὸ τὴν ἄλλη. Ὁ Καρτέσιος εἶχε πεῖ «σκέφτομαι ἄρα ὑπάρχω» (*Cogito ergo sum*). Ἡ ρήση αὐτὴ κατέρρευσε στιγμιαία ἀπὸ μία ἄλλη ρήση, αὐτὴ τοῦ Wittgenstein: «Ἡ ἔλλογη σκέψη προϋποθέτει τὴ γλῶσσα καὶ ἡ γλῶσσα προϋποθέτει κοινωνία». Ἐπομένως, ἂν μᾶς ἐπιτρέπεται, θὰ λέγαμε: «Σκέφτομαι ἄρα ὑπάρχω στὴν κοινωνία». Δὲν λέμε ὅχι στίς «μηχανές», ἀλλὰ ἀναδεικνύουμε τὴν ἀνωτερότητα τοῦ Ἀνθρώπου ἀπέναντι στίς μηχανὲς καὶ θέλουμε μία κοινωνία, ὅπου θὰ τὶς χειριζόμαστε καὶ δὲν θὰ μᾶς χειρίζονται καὶ πού θὰ εἴμαστε ἐλεύθεροὶ νὰ ἐκφράζουμε τὴ δημιουργικότητά μας καὶ τὴν κοινωνικότητά μας. Ἡ νέα γνώση δὲν συνεπάγεται αὐτὴ καθ' αὐτὴν ἓνα σύστημα ἀξιών. Ὑπάρχει αὐτὸ πού γνωρίζουμε, ἀλλὰ ὑπάρχει καὶ ἡ κατεύθυνση πρὸς τὴν ὁποία θέλουμε νὰ πορευθοῦμε· αὐτὸ δηλαδὴ πού ἐπιδιώκουμε. Οἱ ἀξίες δὲν ὑπάρχουν στὴ φύση, δημιουργοῦνται στὸ φαντασιακό μας.

Τὰ τελευταῖα χρόνια ὑπάρχει ἡ τάση νὰ διδάσκεται ἡ τεχνολογία καὶ οἱ ὑπολογιστὲς καὶ ὁ κώδικας καὶ ἡ ρομποτικὴ ὅσο ἐνωρίτερα γίνεται στὰ σχολεῖα. Ναι, προφανῶς καὶ δὲν λέμε ὅχι. Ἀλλὰ ἡ τάση αὐτὴ γίνεται στὸ πλαίσιο προετοιμασίας για τὴ μελλοντικὴ ἀγορὰ ἐργασίας· για τὴν ἀνταγωνιστικότητά μας. Δὲν τοποθετούμαστε ἀρνητικὰ σὲ αὐτό. Εἶναι μία σημαντικὴ παράμετρος τῆς ἐκπαίδευσης. Ὑπάρχει ὅμως καὶ ἡ ἄλλη, ἡ πιὸ σημαντικὴ παράμετρος τῆς Παιδείας, πού γαλουχεῖ πολῖτες. Μὲ αὐτὴν τὴ δεύτερη ματιά, ἡ διδασκαλία τῶν σχετικῶν αὐτῶν μαθημάτων θὰ πρέπει νὰ γίνεται καὶ ἀπὸ τὴν ἄλλη τους πλευρά, στὸ πλαίσιο τῶν Κοινωνικῶν ἐπιστημῶν, πού παράλληλα



μὲ μηχανικούς λογισμικοῦ θὰ προετοιμάζει πολῖτες ποὺ θὰ ζήσουν στὸ νέο περιβάλλον· πολῖτες μὲ κριτική ἰκανότητα ποὺ θὰ ἔχουν αὐξημένη εὐαισθησία μπροστὰ στὸ ψεύτικο καὶ τὸ ἀληθινό· πολῖτες μὲ δημιουργική φαντασία, γιατί ἐδῶ βρίσκεται τὸ πλεονέκτημά τους ἀπέναντι στίς «μηχανές»· πολῖτες ποὺ ἐνῶ θὰ βιώνουν καὶ θὰ ζοῦν τὴν ψηφιακὴ πραγματικότητα (τὰ κοινωνικὰ δίκτυα καὶ ἡ πραγματικότητά τους εἶναι ἐδῶ), ταυτόχρονα δὲν θὰ ἀποδυναμώνουν τὴν ἐνεργὸ παρουσία καὶ ἐπαφή τους μὲ τὴν «πραγματικὴ» πραγματικότητα· πολῖτες, ποὺ ἡ ψηφιακὴ πραγματικότητα θὰ δρᾷ συμπληρωματικὰ τῆς πραγματικῆς πραγματικότητος τοῦ «ἔξω κόσμου», ποὺ δὲν θὰ τοῦ παρέχει εὐτελῆ ὑποκατάστατα διεξόδου σὲ ἐπιθυμίες καὶ προσδοκίες. Καὶ ὁ «ἔξω κόσμος» εἶναι ἡ δική του κοινωνία, ὁ δικός του περίγυρος, μὲ τὶς δεσμεύσεις, τὶς ἐξαρτήσεις, τὶς ὑποχρεώσεις, δηλαδὴ ἡ κοινωνικὴ του ὑπόσταση. Καὶ κοινωνικὴ ὑπόσταση σημαίνει συναίσθηση, ἀνοχή καὶ σεβασμὸς τοῦ διαφορετικοῦ. Σὲ ἀντίθεση μὲ τὴν ψηφιακὴ πραγματικότητα ποῦ, ὅπως ἔχει σήμερα δομηθεῖ τὸ μοντέλο βάσει τοῦ ὁποῖου λειτουργοῦν οἱ ἀντίστοιχες πλατφόρμες, εἶναι ἰδιαίτερα εὐάλωτη στὸ νὰ προάγει τὸν φανατισμό, νὰ καλλιιεργεῖ τὰ πάθη, νὰ ὑψώνει φράγματα, νὰ διχάζει καὶ νὰ χαλιναγωγεῖ. Γιατὶ οἱ ἀλγόριθμοί της δὲν ἀναπτύχθηκαν μὲ γνώμονα νὰ ὑπηρετοῦν τὴ Δημοκρατία καὶ τὴν ἀνοικτὴ κοινωνία, ἀλλὰ νὰ ἐξυπηρετοῦν συμφέροντα καὶ νὰ δημιουργοῦν πελάτες καὶ νὰ ἐξισώνουν, ὅσο μποροῦν, τὸν Ἄνθρωπο μὲ τὴ «μηχανή».

Εἶναι αὐτὸ τὸ ἄρθρο ἓνα ἄρθρο ἀπισιοδοξίας; Ἄρθρο ποὺ ἀποπνέει νεο-Λουδισμό, ποῦ παραπέμπει στοὺς Λουδίτες τοῦ 19ου αἰῶνα ποὺ ἔσπαγαν τὶς μηχανές; Τὸ ἀντίθετο, καὶ αὐτὸς εἶναι ὁ λόγος ποὺ συνεισφέρουμε τὸ ἄρθρο αὐτό. Νὰ μοιραστοῦμε μὲ τοὺς ἀναγνώστες τὶς ἀπόψεις μας ποὺ δὲν εἶναι φόβος γιὰ τὴν Τεχνολογία, ἀλλὰ ἀντίθετα ἐσωτερικὴ δύναμη αἰσιοδοξίας γιὰ τὸν Ἄνθρωπο, ὁ ὁποῖος θὰ ἐξουσιάζει καὶ δὲν θὰ ἐξουσιάζεται ἀπὸ τὴν Τεχνολογία. Αἰσιοδοξία δὲν εἶναι νὰ προφητεύουμε ἓνα μέλλον ρόδινο ποὺ νὰ ὑπόσχεται εὐτυχία (τὸ ἔχουμε κι αὐτὸ διαβάσει ἀπὸ τοὺς «γκουρού» τοῦ εἴδους, ποὺ ὀραματίζονται νὰ ἐμφυτεύουν στὸν ἐγκέφαλο τσιπάκια εὐτυχίας – ἓνα εἶδος ἠλεκτρονικοῦ ὀπίου). Αἰσιοδοξία σημαίνει ὅτι ἀγκαλιάζω τὴν Τεχνολογία ἀλλὰ τὴ χειρίζομαι καὶ δὲν μὲ χειρίζεται· αἰσιοδοξία ση-

μαίνει ότι έχω κριτική στάση απέναντι σε πιθανόν λάθος επιλογές και πάνω από όλα ή αισιοδοξία μου απορρέει από την πίστη μου, τη βαθειά μου πίστη, ότι τό «κακό» οφείλεται σε στρεβλή ή έλλιπη ένημέρωση. Και είμαστε αισιόδοξοι γιατί δέν πιστεύουμε σε λύσεις από μάγους και μαγικά ραβδιά και σε «έτοιμες» λύσεις που ξεθάπτονται από τό παρελθόν, αλλά σε κοινωνίες ανοικτές και ελεύθερες που μαθαίνουν από τὰ λάθη τους, συζητούν, επαναπροσδιορίζουν τούς στόχους τους και δημιουργούν τό μέλλον τους, όπως ακριβώς και ή λειτουργία τής Ανθρώπινης νόησης.

Ἡ Ἱστορία άλλωστε εἶναι Δημιουργική Ἐξέλιξη, όπως μάς εἶπε ὁ Bergson.

## SUMMARY

Artificial Intelligence: Is this the End of the Age of Humans?

By Sergios Theodoridis

*Professor Emeritus of Signal Processing and Machine Learning  
Dept. of Informatics and Telecommunications  
National and Kapodistrian University of Athens*

The goal of this article is to discuss the essence of the difference between what we call human intelligence and what is known as artificial intelligence and machine learning. The essential difference could be compressed in a single word: creativity.

Machines that build upon the machine learning philosophy can only perform predictions. Based on, usually, massive collection of data, they “learn” to predict. These predictions are based on certain characteristics, which the machine “learns” from the collected data, e.g., color, length, height of, say, an object. In contrast, human intelligence builds upon concepts. In a similar vein, machines that are built so that to imitate the reasoning mechanism of the human mind fail when they are confronted with complex situations. This is because human reasoning cannot be reduced to a set of formal logical rules that work on well

defined concepts. Humans can reason on concepts that are not well defined and the logical part of reasoning is just a part of the procedure. During reasoning, feelings, sensations, human values and beliefs are also mobilized to help humans to reach decisions.

The above major differences make human intelligence capable of creativity. In other words, the human mind is able to generalize to previously unseen situations. Humans predict but also do create. The results of creation are not predictable from what existed before. Humans exist in a world which, at the same time, is created by the humans. This is the cultural world. And it is culture that distinguishes humans from animals. Animals and machines can predict. They can also reason to some extent, yet without language and conceptual understanding. In contrast, humans not only predict and reason with concepts and language but more important create the sphere of their existence. Moreover, language presupposes our existence as social beings and therefore as beings with moral values that keep societies together. Above all, the human mind is governed by what we call consciousness. When we think, we are conscious that we are thinking. We are conscious of our consciousness!

Thus, the only danger that we face from the machines is to view them as something superior, something magical, as something truly intelligent and from simple useful tools, whose purpose is make our life easier, to upgrade them to entities that govern our lives. The danger comes from degrading our mind, consciousness and existence from creative entities to non-creative beings that live a predictable life, which is decided and predicted by machines. And of course, what a machine “decides” has been pre-decided by those who program the machines and embed in them their own intelligence and, also, according to their own values for their own benefits. To live together with the machines, which we have to do, this is the reality, and use the machines on our benefit, we have to work for and demand as citizens more democracy, more freedom and for machines that we understand how they work and how they “decide”. The strength of human intelligence is creativity, it is not the number crunching and memorization, the strength of human intelligence is conceptual understanding, the strength of human mind and consciousness lies in our social existence. These should be the

prevailing values and the goals, which any educational system should focus on in the new era.

### Πηγές

- Bergson H., *Creative Evolution*, McMillan, London 1922.
- Chalmers D., *The Conscious Mind*, Oxford University Press, Oxford 1996.
- Damasio A., *Descartes' Error: Emotion, Reason, and the Human Brain*, Penguin, 2005.
- Deutsch D., *The Fabric of Reality*, Penguin, 1998.
- Deutsch D., *The Beginning of Infinity: Explanations That Transform the World*, Allen Lane, 2011.
- Dreyfus H. (ed.), *Husserl, Intentionality, and Cognitive Science*, MIT Press, Cambridge 1982.
- Dreyfus H., *What Computers Still Can't Do*, M.I.T. Press, 1992.
- Heidegger M., *Being and Time*, Translated into English, Blackwell, 1962.
- Merleau-Ponty M., *Phenomenology of Perception*, Routledge & Kegan Paul, 2014.
- Minsky L., *Computation: Finite and Infinite Machines*, Prentice Hall, 1967.
- Russell S., Norvig P., *Artificial Intelligence*, 3rd ed., Prentice Hall, 2010.
- Pearl J., Mackenzie D., *The Book of Why: The New Science of Cause and Effect*, Basic Books, 2018.
- Πλάτωνος, *Εὐθύφρων, Εἰσαγωγή-Μετάφραση-Σχόλια*: Ν. Σκουτερόπουλος, ἐκδ. Καρδαμίτσα, Ἀθήνα 1982.
- Polanyi M., *Personal Knowledge: Towards a postcritical philosophy*, Routledge & Kegan Paul, 1962.
- Piaget J., *The Psychology of Intelligence*, Routledge Classics, 2001.
- Popper K., *The Logic of Scientific Discovery*, Routledge Classics, 2002.
- Schwarz N., Skurnik I., "Feeling and thinking: implications for problem solving", στό: Davidson J. E., Sternberg R. (eds.), *The Psychology of Problem Solving*, Cambridge University Press, 2003.
- Theodoridis S., *Machine Learning: A Bayesian and Optimization Perspective*, 2nd Edition, Academic Press, 2020.
- Todes S., *Body and World*, MIT Press, 2001.
- Turing A., "Computing machinery intelligence", *MIND*, τ. 49 (1950), σσ. 433-460.
- Wittgenstein L., *Φιλοσοφικές Έρευνες, μετάφρ. Π. Χριστοδουλίδης*, ἐκδ. Πατάκης, Ἀθήνα 1977.