Κεφάλαιο 1

Εισαγωγή

«Η επιτυχημένη δημιουργία [γενικευμένης] Τεχνητής Νοημοσύνης θα είναι το μεγαλύτερο γεγονός στην ανθρώπινη ιστορία. Δυστυχώς, ίσως είναι και το τελευταίο εάν δε μάθουμε πώς να αποφεύγουμε τα ρίσκα.» — Stephen Hawking

1.1 Επισκόπηση του Κόσμου της Τεχνητής Νοημοσύνης

Είναι πλέον γεγονός, η Τεχνητή Νοημοσύνη (Artificial Intelligence - AI) εντοπίζεται σε πολλές εκφάνσεις της καθημερινότητας των περισσότερων ανθρώπων [1]. Δεν αποτελεί απλά έναν ακόμα μοδάτο όρο που καταχράται στον χώρο της αγοραστικής (marketing). Εν αντιθέσει, διαδραματίζει καθοριστικό ρόλο στην εργασία μας, στη μετακίνησή μας και στην ψυχαγωγία μας. Για παράδειγμα, εντοπίζεται στις μηχανές αναζήτησης όπως η Google, στους ηλεκτρονικούς χάρτες πλοήγησης αλλά και στα συστήματα συστάσεων (Recomender systems) όπως αυτό του YouTube, του Twitter και του Netflix [2] που εξατομικεύουν το προβαλλόμενο περιεχόμενο στα ενδιαφέροντα του χρήστη.

Η επιρροή που έχει η Τεχνητή Νοημοσύνη είναι ακόμα πιο ευδιάκριτη αν τη μελετήσει κανείς υπό μια συλλογική σκοπιά. Για παράδειγμα, στον χώρο του επιχειρείν, η αξιοποίηση τεχνολογιών Τεχνητής Νοημοσύνης έχει αποδειχθεί ότι αυξάνει την επιχειρηματική αξία (business value) μέσα από τη βελτίωση της επίδοσης τόσο στο οικονομικό (financial), αγοραστικό (marketing) και διοικητικό (administrative) επίπεδο όσο και στο επίπεδο επιχειρηματικών διαδικασιών (business process) [3]. Χαρακτηριστικό παράδειγμα αποτελεί η χρήση των συστημάτων συστάσεων αφού με το να φιλτράρουν το περιεχόμενο και να παρουσιάζουν στον χρήστη μόνο αυτό που του είναι οικείο και θεμιτό, αυξάνουν τον βαθμό ενασχόλησής του με κίνδυνο την παγίδευσή του σε μια «φούσκα προκατειλημμένου φιλτραρίσματος» («biased filter bubble») [4]. Άλλωστε, όπως δηλώνει η ομάδα ανάπτυξης του εν λόγω συστήματος για λογαριασμό της συνδρομητικής υπηρεσίας streaming, Netflix, «Πιστεύουμε ότι αθροιστικά, η επίδραση της εξατομίκευσης και των συστάσεων μας εξοικονομεί ένα δισεκατομμύριο δολάρια τον χρόνο» [5].

Στον εργασιακό χώρο, πολλές δουλειές που περιλαμβάνουν επαναλαμβανόμενες, προβλέψιμες εργασίες κυρίως στον τομέα της βιομηχανίας και της γεωργίας αντικαθίστανται από αυτοματι-

σμούς Τεχνητής Νοημοσύνης (AI automations) εκτοπίζοντας έτσι τον άνθρωπο. Η μείωση των διαθέσιμων θέσεων εργασίας στους τομείς αυτούς δοκιμάζει τα όρια του κοινωνικού οικοδομήματος: τα «εκτοπισμένα» άτομα καλούνται να αποκτήσουν νέες δεξιότητες προκειμένου να βρουν απασχόληση στις πιο δημιουργικές (και συνάμα λιγότερο τυποποιήσιμες) θέσεις του εξελισσόμενου τομέα των υπηρεσιών [6]. Βέβαια, η μείωση των θέσεων εργασίας επαναλαμβανόμενης φύσης είναι μόνο η μια πλευρά του νομίσματος. Σύμφωνα με αναλύσεις, μέχρι την επόμενη δεκαετία οι εφαρμογές της Τεχνητής Νοημοσύνης εν δυνάμει θα αυξήσουν το παγκόσμιο Ακαθάριστο Εθνικό Προϊόν (Gross Domestic Product - GDP) κατά 26% (δεκαπέντε τρισεκατομμύρια δολάρια) [7]. Αυτό, με τη σειρά του, θα οδηγήσει στη δημιουργία πολλών νέων θέσεων εργασίας έτσι ώστε να μην παρατηρηθεί αύξηση στους δείκτες ανεργίας [7, 8].

Τέλος, δε θα μπορούσαμε να παραλείψουμε την επιρροή που έχει η Τεχνητή Νοημοσύνη στον χώρο της υγείας. Οι εφαρμογές είναι ατελείωτες: από συστήματα για πρόωρη διάγνωση ασθενειών μέχρι ρομποτικά συστήματα υποβοήθησης χειρουργείου [9]. Αξιοσημείωτη είναι επίσης η επιτυχημένη εφαρμογή της για την πρόβλεψη της τρισδιάστατης δομής των πρωτεϊνών [10] - ένα θέμα με σημαντικές προεκτάσεις που απασχολούσε την επιστημονική κοινότητα για 50 χρόνια. Παρόλα αυτά, μαζί με την προσπάθεια για αξιοποίηση των νέων τεχνολογιών στην κλινική πράξη προκύπτουν νέες προκλήσεις. Μια πρώτη δυσκολία είναι η ανάπτυξη συστημάτων Τεχνητής Νοημοσύνης που απαιτούν μεγάλο όγκο δεδομένων σε χώρους προβλημάτων όπου αυτά σπανίζουν (όπως για παράδειγμα, στην περίπτωση μιας ασυνήθιστης ασθένειας όπου ο αριθμός των ιατρικών υποθέσεων είναι ελάχιστος). Αυτή η δυσκολία εντείνεται αφενός λόγω της έλλειψης ασφαλών υποδομών για τη συλλογή ιατρικών δεδομένων [11] και αφετέρου λόγω της απόρρητης φύσης αυτών, κάτι που δυσχεραίνει τον ελεύθερο διαμοιρασμό τους. Μια τελευταία δυσκολία αποτελεί το γεγονός ότι πολλά συστήματα Τεχνητής Νοημοσύνης που έχουν αναπτυχθεί σε περιβάλλον εργαστηρίου (Lab setting) δεν παρέχουν αρκετά κίνητρα για μεταστροφή της καθιερωμένης κλινικής πράξης [11]. Για να επιτευχθεί κάτι τέτοιο, μεταξύ άλλων θα πρέπει τα συστήματα να αποδίδουν αποδεδειγμένα τόσο καλά όσο και το καταρτισμένο προσωπικό στο συγκεκριμένο πεδίο εφαρμογής τους και να παρέχουν πληροφορίες που θα τα καθιστούν περισσότερο έμπιστα π.χ. αιτιολογώντας την απόφασή τους (explainability), παρέχοντας μια μετρική αβεβαιότητας (uncertainty) ή δίνοντας τη δυνατότητα αλληλεπίδρασης [12].

Αντιλαμβανόμενοι το εύρος των εφαρμογών της Τεχνητής Νοημοσύνης, η εκτίμηση από την International Data Corporation - IDC πως οι Ευρωπαϊκές δαπάνες σε τέτοιες εφαρμογές θα έχουν σχεδόν τριπλασιαστεί μέσα στα επόμενα τρία χρόνια δε θα πρέπει να μας εκπλήσσει. Ωστόσο, με τη μεγάλη ισχύ έρχεται και μεγάλη ευθύνη. Είναι αλήθεια, η Τεχνητή Νοημοσύνη είναι ήδη εδώ και θα συνεχίζει να επιδρά όλο και εντονότερα στην καθημερινότητα μας και στην κοινωνία. Βέβαια, η σημερινή Τεχνητή Νοημοσύνη είναι εντελώς διαφορετική από αυτό που φαντάζονταν η κοινή γνώμη τις προηγούμενες δεκαετίες (σαφώς επηρεασμένη από ταινίες επιστημονικής φαντασίας όπως το Terminator). Θα λέγαμε, αντίθετα, πως εμφανίζεται περισσότερο με μια περιορισμένη μορφή στην εκάστοτε συγκεκριμένη εφαρμογή (narrow AI). Έτσι, ένα «εφυές» σύστημα για μια εργασία δεν μπορεί να «γενικεύσει» και να εφαρμοστεί σε άλλο χώρο προβλημάτων. Ούτε λόγος δε για αισθήματα και υπαρξιακή συνείδηση· αυτά (ακόμα) ανήκουν στην επιστημονική φαντασία. Αυτό όμως δε σημαίνει ότι η επιπόλαιη χρήση της Τεχνητής Νοημοσύνης δεν ελοχεύει κινδύνους. Σύμφωνα με το περιοδικό Spectrum της IEEE [13] προτού

επιτευχθεί Τεχνητή Νοημοσύνη επιπέδου ανθρώπου (human-like Artificial Inteligence) - αυτή στην οποία αναφέρεται ο Stephen Hawking - υπάρχουν ήδη πολλά σενάρια όπου εφαρμογές της μπορούν να αποβούν μοιραίες. Ενδεικτικά, ένα από αυτά είναι τα deepfakes - ψεύτικα πολυμέσα βίντεο και εικόνας κατασκευασμένα από εφαρμογές Τεχνητής Νοημοσύνης - έχουν υπονομεύσει την εμπιστοσύνη στα συστήματα πληροφόρησης. Επιπρόσθετα, ένα ακόμα καταστροφικό σενάριο σχετίζεται με την ιδιωτικότητα (privacy) και την ελεύθερη βούληση (free will). Με την παραχώρηση ευαίσθητων δεδομένων σε επιχειρήσεις και κυβερνήσεις τους παρέχουμε τη δυνατότητα να μας εποπτεύουν ακόμα και να μας χειραγωγούν. Ένα τελευταίο σενάριο για το οποίο διαδραματίζουν άμεσο ρόλο τα κοινωνικά δίκτυα είναι αυτό του μειωμένου διαστήματος προσοχής (short attention span) ως απόρροια της εκμετάλλευσης του μηχανισμού επιβράβευσης του εγκεφάλου ώστε οι χρήστες να εθίζονται σε αυτά. Το περιοδικό καλεί τον αναγνώστη να αναλογιστεί τις συνέπειες της συνεχόμενης βελτίωσης των μηχανισμών που μας καθηλώνουν από τη νέα τεχνολογία. Συμπερασματικά, η Τεχνητή Νοημοσύνη αν και δεν προσομοιάζει την ανθρώπινη νοημοσύνη δεν παύει να αποτελεί μια πολύ ισχυρή τεχνολογία που μπορεί να αποβεί είτε σωτήρια είτε μοιραία ανάλογα με τον τρόπο αξιοποίησής της.

Είναι λοιπόν απαραίτητη η εξασφάλιση της συνετής χρήσης αυτών των τεχνολογιών μέσω μιας σειράς κανονισμών. Στην Ευρωπαϊκή Ένωση, μια σειρά από διατάξεις επιχειρούν να θέσουν ένα νομοθετικό πλαίσιο ώστε να ωθήσουν στην αξιοποίηση της Τεχνητής Νοημοσύνης διασφαλίζοντας παράλληλα την ασφάλεια των θεμελιωδών δικαιωμάτων [14]. Άλλωστε, σύμφωνα με την von der Lein [15], «Η Τεχνητή Νοημοσύνη πρέπει να εξυπηρετεί τους ανθρώπους και συνεπώς, πρέπει πάντα να συμμορφώνεται με τα δικαιώματά τους. Αυτός είναι ο λόγος που ένα άτομο πρέπει πάντα να έχει τον έλεγχο στην περίπτωση κρίσιμων αποφάσεων[...] Εφαρμογές της Τεχνητής Νοημοσύνης που μπορεί να παρέμβουν στα ανθρώπινα δικαιώματα θα πρέπει να ελέγχονται και να πιστοποιούνται πριν φτάσουν στην Ευρωπαϊκή αγορά. » Αν και οι αυστηροί διακανονισμοί καθυστερούν τη μετάβαση εφαρμογών Τεχνητής Νοημοσύνης από το εργαστήριο στην αγορά, εξασφαλίζουν την ασφάλειά τους συμβάλλοντας στην αξιοπιστία τους.

Με την παραπάνω σύντομη εισαγωγή καλύψαμε εμπεριστατωμένα πολλά από τα μη τεχνικά θέματα που σχετίζονται με την Τεχνητή Νοημοσύνη (Artificial Inteligence). Αναλυτικότερα, αρχίσαμε από παραδείγματα εντοπισμού της στην καθημερινή ζωή σε ατομικό και σε συλλογικό επίπεδο με τα οποία αντιληφθήκαμε τη σημασία της. Συνεχίσαμε με τους κινδύνους που ελοχεύει η απερίσκεπτη εφαρμογή της σε συγκεκριμένες εργασίες επισημαίνοντας ταυτόχρονα ότι η τεχνητή νοημοσύνη απέχει από την ανθρώπινη. Κλείσαμε, με μερικές από τις προσπάθειες που γίνονται σε Ευρωπαϊκό επίπεδο για την αποφυγή αυτών των κινδύνων. Πολλά από τα προαναφερθέντα στοιχεία πιθανότατα να είναι ήδη γνωστά σε έναν έμπειρο αναγνώστη. Εντούτοις, εξυπηρετούν σε μια ομαλή εισαγωγή για τον αρχάριο και σε μια υπενθύμιση για τον έμπειρο αναγνώστη του κόσμου της Τεχνητής Νοημοσύνης. Στην επόμενη υπο-ενότητα αυτού του κεφαλαίου θα παρουσιάσουμε μια ιστορική αναδρομή της τεχνητής νοημοσύνης. Έτσι, ο αναγνώστης θα κατανοήσει σε βάθος την έννοια γύρω από την οποία εκτυλίσσεται η παρούσα διπλωματική προτού εισαχθεί στο συγκεκριμένο τεχνικό της θέμα. Έπειτα, περιγράφεται το κίνητρο που με ωθούσε καθ΄όλη τη διάρκεια συγγραφής του έργου. Τέλος, αναφερόμαστε στην τεχνική συνεισφορά της παρούσας εργασίας και στην οργάνωση του τόμου.

1.2 Ιστορική Αναδρομή Τεχνητής Νοημοσύνης

Οι εφευρέτες οραματίζονται εδώ και χιλιετίες τη δημιουργία μηχανών που σκέφτονται. Ήδη, γύρω στο 700 π.Χ. αναφέρεται από τον Ησίοδο ο Τάλος: ο μυθικός χάλκινος γίγαντας φτιαγμένος από τον Ήφαιστο με αποστολή να προστατεύσει το νησί της Κρήτης από τους επιδρομείς [16]. Παρόμοια παραδείγματα αποτελούν αυτά της Πανδώρας και της Γαλάτειας. Η μακρόβια επιθυμία για απομίμηση της νοημοσύνης μαρτυρά την αξία που της δίνει ο άνθρωπος. Γεγονός απόλυτα δικαιολογημένο αφού η νοημοσύνη - η νοητική ικανότητα που μας επιτρέπει να σκεπτόμαστε λογικά, να επιλύουμε προβλήματα και να μαθαίνουμε - έχει συμβάλει σημαντικά στην επιβίωση του είδους από τον διαειδικό ανταγωνισμό (interspecific competition)[1]. Σε τελική ανάλυση, ο όρος homo sapiens - άνθρωπος ο σοφός - οφείλεται στη σημασία που έχει η νοημοσύνη στη ζωή μας.

Το πρώτο ευρέως αναγνωρισμένο έργο προς την επίτευξη Τεχνητής Νοημοσύνης είναι αυτό της μαθηματικής μοντελοποίησης της λειτουργίας ενός νευρώνα από τους Warren McCulloch και Walter Pitts (1943) [20]. Αναλυτικότερα, βασιζόμενοι στην υπόθεση ότι η κατάσταση λειτουργίας ενός νευρώνα είναι δυαδική («all-or-none») αναπαρέστησαν κάθε νευρώνα ενός δικτύου ως μια πρόταση (proposition) της προτασιακής λογικής (propositional logic). Όπως περιγράφουν, η διέγερση (excitation) του μοντέλου ενός νευρώνα είναι ταυτόσημη με το να είναι η πρόταση του νευρώνα θετική, κάτι που εξαρτάται από την κατάσταση των γειτονικών νευρώνων. Όσο περισσότεροι, διεγερτικά διασυνδεδεμένοι (excitatory connected) προσυναπτικοί νευρώνες (presynaptic neurons) είναι ενεργοποιημένοι τόσο πιο πιθανή είναι η ενεργοποίηση του μετασυναπτικού νευρώνα (postsynaptic neuron). Μεταξύ άλλων, απέδειξαν ότι κάθε συνάρτηση που μπορεί να υπολογιστεί από μια μηχανή Turing μπορεί να υπολογιστεί και από ένα δίκτυο από διασυνδεδεμένους τεχνητούς νευρώνες ύστερα από την κατάλληλη παραμετροποίησή του. Το έργο των Warren McCulloch και Walter Pitts ήταν πρωτοπόρο για την εποχή του αφού έθεσε τις βάσεις όχι μόνο για την σημερινή Τεχνητή Νοημοσύνη αλλά και για την Υπολογιστική Νευρωεπιστήμη (Computational Neuroscience). Ωστόσο, οι συγγραφείς δεν παρουσίασαν κανέναν αλγόριθμο για την αλλαγή της τοπολογίας και των παραμέτρων του τεχνητού νευρωνικού δικτύου αν και φαίνεται πως αναγνώριζαν τη σημασία του για τη μάθηση.

Στο βιβλίο του [21] ο D. Hebb το 1949 επιδίωξε να ενώσει τις αποκλίνουσες θεωρίες της ψυχολογίας και της νευροεπιστήμης θέτοντας κοινές βάσεις για την ερμηνεία της ανθρώπινης συμπεριφοράς. Προηγούμενες θεωρίες απέφευγαν να δώσουν εξήγηση στις διεργασίες του εγκεφάλου ως μεσάζοντα μεταξύ του αισθητηριακού ερεθίσματος (sensory stimuli) και της πιθανής, καθυστερημένης απόκρισης. Αντίθετα, κατέφευγαν στη φιλοσοφία για την ανάλυση των χαρακτηριστικών της ανθρώπινης συμπεριφοράς όπως η προσοχή (attention), το ενδιαφέρον (interest) και η «προσδοκία» (expectancy). Ο Hebb όμως, εργάστηκε στο να αποδείξει ότι η ανθρώπι-

[1]Για την ακρίβεια, ενώ σύμφωνα με τη Δαρβινική θεώρηση η αφηρημένη νοημοσύνη μπορεί να προκύψει άμεσα από τη θεωρία της εξέλιξης των ειδών, νεότερες έρευνες το διαψεύδουν αφού το χαρακτηριστικό της αφηρημένης σκέψης ήταν αχρείαστο στην πραγματιστική παλαιολιθική εποχή. Στην προσπάθειά τους να ερμηνεύσουν την εμφάνιση νοημοσύνης στους ανθρώπους ορίζουν τον όρο «διανοητική βιοθέση» (cognitive niche) για να περιγράψουν όλα τα ζωολογικά ασυνήθιστα χαρακτηριστικά (zoologically unusual traits) που εμφανίζει ο άνθρωπος με τα κυριότερα να είναι η κοινωνικότητα και η λογική αιτίου - αποτελέσματος (cause-and-effect reasoning) [17,18]. Υποστηρίζουν λοιπόν, ότι η εμφάνιση της διανοητικής βιοθέσης αποτέλεσε τον καταλύτη για την εξέλιξη της ανθρώπινης νοημοσύνης [19]

νη συμπεριφορά μπορεί να γίνει κατανοητή υπό το πρίσμα της φυσιολογίας ("[expectancy] can be a physiologically intelligible process"). Φαινομενικά, ενώ το έργο του απασχολεί μόνο την επιστήμη της Νευροφυσιολογίας (Neurophysiology) συνεισέφερε σημαντικά στο κίνημα του διασυνδετισμού (Connectionism) - το κίνημα μελέτης των διανοητικών διεργασιών με τη χρήση τεχνητών νευρωνικών δικτύων. Για παράδειγμα, η θεώρηση της διαδικασίας μάθησης ως της επαναλαμβανόμενης, ταυτόχρονης πυροδότησης γειτονικών νευρώνων με αποτέλεσμα [την ενδυνάμωση των δεσμών και] τη διαμόρφωση νευρικών συστάδων (cell assemblies) είναι η λογική πίσω από πολλούς αλγορίθμους μάθησης τεχνητών νευρωνικών δικτύων. Επίσης, ένα στοιχείο που θα μας φανεί χρήσιμο στη συνέχεια είναι η παρατήρησή του ότι η κίνηση των ματιών δεν είναι τυχαία αλλά σχετίζεται με τη διαδικασία αντίληψής των θωρούμενων αντικειμένων. Τα παραπάνω δύο έργα έθεσαν το απαραίτητο θεωρητικό υπόβαθρο εμπνέοντας τους ερευνητές στην πειραματική υλοποίηση της Τεχνητής Νοημοσύνης.

Αρκετές ήταν οι απόπειρες δημιουργίας Τεχνητής Νοημοσύνης. Ο πρώτος υπολογιστής τεχνητών νευρωνικών δικτύων ονομάζονταν SNARC (Stochastic Neural Analog Reinforcement Calculator) και κατασκευάστηκε από τους Marvin Minsky και Dean Edmonds το 1950 [22]. Χρησιμοποιώντας 3000 λυχνίες κενού και έναν μηχανισμό αυτόματου πιλότου εξομοίωσαν τη λειτουργία 40 διασυνδεδεμένων νευρώνων. Χρησιμοποιώντας έναν απλό μηχανισμό επιβράβευσης τα «βάρη» του δικτύου - υπο τη μορφή ποτενσιόμετρων - προσαρμόζονταν στο πρόβλημα του λαβυρίνθου στο οποίο δοκιμάστηκε. Ακόμα ένα παράδειγμα τεχνητής νοημοσύνης μπορεί να θεωρηθεί το πρόγραμμα του Christopher Strachey στον υπολογιστή Manchester Mark 1 [23] που αργότερα θεωρήθηκε το πρώτο βιντεοπαιχνίδι. Ήταν ένα παιχνίδι ντάμας που πιθανώς χρησιμοποιούσε κάποιον μη πλήρη (incomplete) αλγόριθμο αναζήτησης στον χώρο των επιτρεπτών ακολουθιών κινήσεων (action sequences). Παρόλα αυτά, η δυνατότητά του να ανταγωνίζεται αποτελεσματικά τον άνθρωπο οδήγησε στη θεώρησή του ως Τεχνητή Νοημοσύνη. Άλλωστε, η σύγχυση για το θέμα ήταν ακόμα μεγαλύτερη την εποχή εκείνη.

Ο αρχικός ενθουσιασμός για το επιστημονικό πεδίο τράβηξε τα βλέμματα πολλών επιφανών ερευνητών της εποχής. Ο πατέρας της Επιστήμης των Υπολογιστών (Computer Science) και της Τεχνητής Νοημοσύνης, Alan Turing, στην προσπάθειά του να διασαφηνίσει το ερώτημα του «αν οι μηχανές σκέφτονται» επινόησε το επονομαζόμενο Turing Test. Σύμφωνα με τη δημοσίευσή του [24] το 1950, πρόκειται για μια δοκιμασία που εμπλέκει έναν «ανακριτή» ο οποίος διατυπώνει γραπτές ερωτήσεις σε δύο «μάρτυρες»: έναν άνθρωπο και μια μηχανή. Η δοκιμασία θεωρείται επιτυχής όταν ο ανακριτής - χωρίς να έχει οπτική επαφή με τους «μάρτυρες» - δεν μπορεί να ξεχωρίσει τον άνθρωπο από τη μηχανή. Στην ίδια δημοσίευση τόνισε τη σημασία της μάθησης για την ανάπτυξη της Τεχνητής Νοημοσύνης. Υποστήριζε ότι αντί να επιχειρείται η εξονυχιστική συγγραφή ενός προγράμματος που θα μοιάζει με τη σκέψη ενός ώριμου ενήλικα (με αμέτρητες προγραμματισμένες εντολές) είναι προτιμότερη και ταχύτερη η προσομοίωση της νοημοσύνης ενός παιδιού που μέσα από μηχανισμούς εκπαίδευσης, έμμεσα, αποκτά ώριμη σκέψη. Επίσης, εναπόθεσε τους σπόρους για τους γενετικούς αλγορίθμους, ενώ σε επόμενη δημοσίευσή του [25] μελέτησε τους τρόπους με τους οποίους μια μηχανή με νοημοσύνη θα μπορούσε να λειτουργεί. Ένα ακόμα δημοφιλές όνομα, ο John von Neumann συνεισέφερε στον χώρο αναπτύσσοντας τα «τεχνητά αυτόματα» (artificial automata) [26] ενώ η συμβολή του πιθανώς θα ήταν ακόμα μεγαλύτερη αν προλάβαινε να ολοκληρώσει το βιβλίο του «Ο Υπολογιστής και το

Μυαλό» (The Computer and the Brain). Μια τελευταία απόδειξη της προσοχής που έλαβε η Τεχνητή Νοημοσύνη διαφαίνεται στα δέκα μέλη σχετικού σεμιναρίου (workshop) που έλαβε χώρα το καλοκαίρι του 1955 στο Dartmouth College [27]. Ίσως το ποιο σημαντικό πόρισμα αυτής της συνάθροισης ήταν η ανάπτυξη του Logic Theorist από τους Allen Newell και Herbert Simon, ενός συστήματος για την απόδειξη θεωρημάτων στα μαθηματικά.

Η δεκαετία που ακολούθησε χαρακτηρίζεται από έντονη αισιοδοξία για τις δυνατότητες της Τεχνητής Νοημοσύνης: γενναιόδωρες επενδύσεις σε ερευνητικά προγράμματα ενθάρρυναν τη δημιουργία ποικίλων προγραμμάτων που υποδείκνυαν κάποια μορφή νοημοσύνης. Πιο συγκεκριμένα, αν εξαιρέσουμε την εξέχουσα δουλειά του Arthur Samuel όπου ανέπτυξε ένα παιχνίδι ντάμας χρησιμοποιώντας ενισχυτική μάθηση (Reinforcement Learning), οι περισσότερες προσπάθειες εστίασαν στον χώρο της μίμησης της ανθρώπινης συλλογιστικής (reasoning). Η ιδέα είναι ότι με μια τυπική γλώσσα (formal language) για την αναπαράσταση της γνώσης (knowledge representation) στον υπολογιστή μαζί με την εφαρμογή απλών κανόνων λογικής συμπερασματολογίας (logical inference) σε αυτή καθιστούν δυνατή την εξαγωγή πορισμάτων. Καθοριστικό ρόλο σε αυτή τη «σχολή» είχε η - αναχρονιστική - υπόθεση ότι η νοημοσύνη είναι άρρηκτα συνδεδεμένη με τη δυνατότητα χειρισμού συμβόλων οργανωμένων σε δομές δεδομένων (physical symbol system hypothesis).

Σε αυτήν την κατεύθυνση, δηλαδή της συμβολικής Τεχνητής Νοημοσύνης (Symbolic Artificial Inteligence), εργάστηκαν αρκετοί επιστήμονες της εποχής. Για παράδειγμα, οι δύο ερευνητές πίσω από το Logic Theorist επινόησαν το General Problem Solver. Πρόκειται ουσιαστικά για έναν αλγόριθμο ο οποίος δέχεται σαν είσοδο μια τυποποιημένη περιγραφή του προβλήματος και το επιλύει ακολουθώντας μια στρατηγική ευρετικής αναζήτησης (heuristic search) της λύσης [27]. Στο ίδιο μήκος κύματος εργάστηκε και ο John McCarthy. Εκτός από το ότι ανέπτυξε τη γλώσσα προγραμματισμού Lisp, ειδικά φτιαγμένη για εφαρμογές Τεχνητής Νοημοσύνης, εξέλιξε το πεδίο με το δημοσίευμά του «Programs with Common Sense» (1958) στο οποίο περιέγραφε το Advice Taker. Αυτό ήταν ένα πρόγραμμα για την επίλυση προβλημάτων μέσω της εφαρμογής «κοινής λογικής» σε προτάσεις διατυπωμένες σε τυπική γλώσσα. Για παράδειγμα, δοθέντος μιας σειράς υποθέσεων σχετικά με το περιβάλλον του προβλήματος διατυπωμένων σε τυπική γλώσσα (π.χ. «Εγώ είμαι στο γραφείο.», «Θέλω να πάω αεροδρόμιο.» κτλ.) ο αλγόριθμος εξήγαγε ένα πλάνο με τα βήματα που πρέπει να ακολουθηθούν για τη μετάβαση στο αεροδρόμιο [28]. Το έργο του συνέχισε ο J. A. Robinson όπου και επινόησε μια πλήρη μέθοδο επίλυσης (complete resolution method) για προβλήματα εκφρασμένα σε λογική πρώτης τάξης. Οι εφαρμογές του ήταν πολλές: από συστήματα μαθηματικού λογισμού (James Slagle's SAINT program [29] και Daniel Bobrow's STUDENT program) μέχρι εφαρμογές ερωταπαντήσεων (Cordell Green's question-answering and planning systems) και ρομποτικής (Shakey Robotics Project). Τέλος, πολλές εφαρμογές της Συμβολικής Τεχνητής Νοημοσύνης αναπτύχθηκαν για το «παιχνίδι» blocks world: ένα περιβάλλον αποτελούμενο από τουβλάκια που αποσκοπούσε στον πειραματισμό συστημάτων αναπαράστασης γνώσης και συλλογιστικής [30].

Την ίδια εποχή, ειδικά για τον χώρο των νευρωνικών δικτύων υπήρξε σημαντική πρόοδος με τα έργα Perceptron και ADALINE. Το πρώτο συγγράφηκε από τον F. Rosenblatt το 1958 και αποτέλεσε το πρώτο μοντέλο νευρωνικού δικτύου με δυνατότητα επιβλεπόμενης μάθησης supervised learning. Πιο συγκεκριμένα, το Perceptron είναι ένας ταξινομητής γραμμικά διαχωρίσιμων

προτύπων με ένα μεμονωμένο τεχνητό νευρώνα του οποίου οι ελεύθεροι παράμετροι - τα προσυναπτικά βάρη (presynaptic weights) και η πόλωση (bias) - προσαρμόζονται στα δεδομένα εισόδου σύμφωνα με έναν αλγόριθμο μάθησης (perceptron rule) [31]. Σε μια εκτενή παρουσίαση του έργου [32], ο Rosenblatt βασίστηκε στη θεωρία του D. Hebb και την επέκτεινε προτείνοντας ένα μοντέλο (το Perceptron) με το οποίο η συμπεριφορά (καμπύλη εκμάθησης) μπορεί να προβλεφθεί από τη νευροφυσιολογία του συστήματος (τα συναπτικά βάρη). Παρόμοιο ήταν και το έργο ADALINE (Adaptive Linear Neuron) του B. Widrow στο οποίο περιγράφεται και πάλι ένας αλγόριθμος μάθησης για την προσαρμογή των βαρών. Αυτή τη φορά όμως, είναι ο (γνωστός) αλγόριθμος στοχαστικής κατάβασης βαθμίδας stochastic gradient descent που χρησιμοποιείται ακόμα και σήμερα στον αλγόριθμο γραμμικής παλινδρόμησης (linear regression). Μια ακόμα αξιοσημείωτη διαφορά έγκειται στη συνάρτηση ενεργοποίησης όπου ενώ στο πρώτο έργο είναι η βηματική συνάρτηση (step function), στο δεύτερο έργο είναι η γραμμική συνάρτηση (linear activation function - identity function) που καθιστά τον αλγόριθμο κατάλληλο για την πρόβλεψη πραγματικών τιμών [33, 34]. Συνεπώς, ενώ το πρώτο έργο αποτελεί όπως προαναφέρθηκε έναν αλγόριθμο ταξινόμησης, το δεύτερο έργο ανήκει στην κατηγορία αλγορίθμων γραμμικής παλινδρόμησης. Βέβαια, μολονότι και τα δύο έργα είχαν καθοριστική σημασία στην εξέλιξη της Τεχνητής Νοημοσύνης με τη μορφή που τη συναντάμε σήμερα, όπως θα δούμε στη συνέχεια, η έντονη κριτική που ακολούθησε τα επισκίασε για μια ολόκληρη δεκαετία.

Γύρω στο 1970, το επιστημονικό πεδίο της Τεχνητής Νοημοσύνης διήλθε μια εποχή «χειμώνα» (AI winter). Η χρηματοδότηση ερευνητικών προγραμμάτων πάγωσε και έτσι το ενδιαφέρον στράφηκε αλλού. Κοιτώντας πίσω, είναι εύκολο να εντοπίσει κανείς τα αίτια αυτού του «χειμώνα». Καταρχάς, ένας λόγος ήταν (και είναι) το ελλιπές επιστημονικό υπόβαθρο σε ό,τι αφορά την ανθρώπινη νοημοσύνη [35]. Αναμενόμενο, αφού για μια επιτυχημένη μίμηση της ανθρώπινης νοημοσύνης απαιτείται πρώτα η κατανόησή της. Βέβαια, ίσως ο σημαντικότερος λόγος ήταν η απογοήτευση που προκλήθηκε όταν φιλόδοξες υποσχέσεις για τις δυνατότητες της Τεχνητής Νοημοσύνης στο εγγύς μέλλον δεν μπόρεσαν να ικανοποιηθούν. Όπως αποδείχθηκε, η μετάβαση της Τεχνητής Νοημοσύνης από εφαρμογές παιδικών κόσμων όπως το blocks world σε πραγματικά προβλήματα δεν ήταν απλώς ζήτημα γραμμικής αύξησης της υπολογιστικής δύναμης. Για παράδειγμα, στην περίπτωση της συμβολικής Τεχνητής Νοημοσύνης, με την ωρίμανσή της θεωρίας πολυπλοκότητας (computational complexity αναδείχθηκε το θέμα της συνδυαστικής έκρηξης (combinatorial explosion) αποκαλύπτοντας έτσι τη δυσεπίλυτη (intractable) φύση πολλών προβλημάτων του αληθινού κόσμου. Αντίστοιχα εμπόδια έκαναν την εμφάνισή τους στον χώρο των Νευρωνικών Δικτύων. Το σημαντικότερο ήταν αυτό της αδυναμίας ενός μεμονωμένου Perceptron με δύο εισόδους να αναπαραστήσει πολλές συναρτήσεις όπως τη συνάρτηση XOR [27] - περιγράφηκε από το βιβλίο Perceptrons των M. Minsky και S. Papert το 1969. Συνολικά, αν και ορισμένα από τα ανωτέρω αίτια είναι ακόμα σε ισχύ, το ενδιαφέρον σύντομα αναζωπυρώθηκε.

Παρά τις ανωτέρω αδυναμίες της Τεχνητής Νοημοσύνης την εποχή εκείνη, αυτό δεν εμπόδισε την αξιοποίησή της σε εξειδικευμένες εφαρμογές. Πιο συγκεκριμένα, τις δεκαετίες του 1970 και 1980 αναπτύχθηκαν τα «έμπειρα συστήματα» (expert systems). Πρόκειται για προγράμματα που εφαρμόζουν κανόνες συλλογιστικής σε εξειδικευμένη βάση γνώσης (domain-specific knowledge base) μιμούμενοι τη διαδικασία λήψης αποφάσεων ενός εμπειρογνώμονα. Η εξειδικευμένη βάση γνώσης περιόριζε σημαντικά τον χώρο αναζήτησης λύσεων (search space) έτσι ώστε η συνδυαστι-

κή έκρηξη να μην αποτελεί πρόβλημα. Αυτό, επέτρεψε να αναπτυχθούν πολλά έμπειρα συστήματα για εμπορική χρήση - κυρίως στον χώρο της υγείας - αποδεικνύοντας για πρώτη φορά έμπρακτα τα οφέλη της Τεχνητής Νοημοσύνης. Ενδεικτικά, δύο δημοφιλή παραδείγματα είναι το MYCIN και το INTERNIST. Το MYCIN αποσκοπούσε στη διάγνωση βακτηριακών μολύνσεων μέσω ενός αλγορίθμου συλλογιστικής που μοντελοποιούσε την αβεβαιότητα των λογικών υποθέσεων και συμπερασμάτων. Το INTERNIST από την άλλη βοηθούσε στη διάγνωση ασθενειών μετά από την περιγραφή των εκδηλούμενων συμπτωμάτων [36]. Αν και τα έμπειρα συστήματα ανανέωσαν το ενδιαφέρον για την Τεχνητή Νοημοσύνη, αυτό δε διήρκησε πολύ λόγω προβλημάτων που εμφάνιζαν με το κυριότερο να είναι η έλλειψη «κοινής λογικής» [37].

Η έρευνα στον χώρο της Τεχνητής Νοημοσύνης αποκαταστάθηκε στα συνήθη υψηλά επίπεδα σε σύντομο χρονικό διάστημα. Αυτό μπορεί σε μεγάλο βαθμό να αποδοθεί σε ένα μεμονομένο έργο· το Parallel Distributed Processing που συγγράφηκε από τους David E. Rumelhart et al. και δημοσιεύτηκε το 1986 [38]. Οι συγγραφείς, μεταξύ των οποίων και ο Geoff Hinton εμπνεόμενοι από τα παλαιότερα έργα πάνω στη γνωστική νευροεπιστήμη (cognitive neuroscience) έστρεψαν την έρευνα του χώρου από πειραματικές «αλχημείες» (π.χ. αυτή της συμβολικής λογικής) σε μια πιο επίσημη, φορμαλιστική διαδικασία βασιζόμενη λιγότερο στη φιλοσοφία και περισσότερο στις θετικές επιστήμες[2]. Με αυτόν τον τρόπο, η σχολαστική συγγραφή αναρίθμητων κανόνων προτασιακής λογικής για τη δημιουργία βάσεων γνώσης εγκαταλείφθηκε, μαζί της και η θεωρία της συμβολικής Τεχνητής Νοημοσύνης. Άλλωστε, η τεχνολογία των έμπειρων συστημάτων είχε παρακμάσει αφού όπως φάνηκε από την απουσία «κοινής λογικής» σε αυτά, ήταν εξαιρετικά περιοριστική η χρήση προτασιακής λογικής για την περιγραφή του πραγματικού, αβέβαιου κόσμου [27, 39].

Τη δεκαετία του 1980 τη θέση της συμβολικής Τεχνητής Νοημοσύνης πήρε το κίνημα του κονεκτιβισμού (connectionist movement). Όπως θα δούμε, αυτό συνέβαλλε καθοριστικά στη διαμόρφωση του σημερινού κλάδου των νευρωνικών δικτύων. Τυπικά, ο κονεκτιβισμός είναι το κίνημα της γνωστικής επιστήμης (cognitive science) που επιχειρεί να εξηγήσει τις διανοητικές διεργασίες με τη χρήση ενός δικτύου με βάρη (weighted network) που διασυνδέει απλές μονάδες επεξεργασίες [40]. Σε αυτήν τη θεωρία καταπιάστηκαν και οι συγγραφείς του έργου Parallel Distributed Processing [38] οι οποίοι εδραίωσαν τις ιδέες που σήμερα θεμελιώνουν τη θεωρία των νευρωνικών δικτύων. Η πρώτη σημαντική ιδέα που περιγράφεται λεπτομερώς στο έργο είναι η κατανεμημένη αναπαράσταση (distributed representation) σύμφωνα με την οποία κάθε είσοδος στο σύστημα αναπαρίσταται από πολλά χαρακτηριστικά κατανεμημένα στο δίκτυο και ανάποδα, δηλαδή κάθε μεμονωμένο χαρακτηριστικό μπορεί να αποτελεί μέρος της περιγραφής πολλών, ετερογενών εισόδων. Μια ακόμα σημαντική ιδέα είναι αυτή της μηχανικής μάθησης (machine learning) με την οποία η επίδοση ενός συστήματος βελτιώνεται (μαθαίνει) από την εμπειρία. Για τον σκοπό αυτό, παρουσιάζουν έναν επαναστατικό αλγόριθμο ο οποίος αυτοματοποιεί τη διαδικασία μηχανικής μάθησης στα νευρωνικά δίκτυα. Πρόκειται για τον αλγόριθμο ανάστροφης διάδοσης σφάλματος (back propagation) ο οποίος, παραδόξως, ενώ είχε αναπτυχθεί περίπου το 1960 γνώρισε ευρεία χρήση από το 1980 και μετά. Συνεπώς, η σημασία του κονεκτιβισμού είναι καθοριστική αφού αναβίωσε τις ιδέες της γνωστικής επιστήμης επανατοποθετώντας κατά αυτόν τον τρόπο τον κλάδο της Τεχνητής Νοημοσύνης σε μια πιο επιστημονική τροχιά.

[2] Στη βιβλιογραφία αυτό το γεγονός αναφέρεται σαν «η νίκη των καθαρών» (victory of the neats) [27].

Η στροφή σε μια πιο επιστημονική προσέγγιση του κλάδου της Τεχνητής Νοημοσύνης το 1980 δε συνέβαλλε μόνο στην ανάπτυξη του κλάδου των νευρωνικών δικτύων. Για παράδειγμα, ο κλάδος της επεξεργασίας φυσικής γλώσσας επωφελήθηκε σημαντικά από την επιτυχημένη μοντελοποίηση ακολουθιών με τη χρήση κρυφών Μαρκοβιανών μοντέλων (hidden Markov models) και αργότερα, το 1997 με μονάδες μακράς-βραχέας μνήμης (Long-Short Term Memory block - LSTM). Επίσης, ο χώρος της όρασης υπολογιστών επωφελήθηκε από τη σύγκλιση της Τεχνητής Νοημοσύνης με τις θετικές επιστήμες. Την πρόοδο μαρτυρά η εμφάνιση των πρώτων εφαρμογών οπτικής αναγνώρισης χαρακτήρων (optical character recognition) τη δεκαετία του 1980 και ύστερα, των τυποποιημένων βάσεων δεδομένων για ανάπτυξη και μέτρηση απόδοσης οπτικών συστημάτων αναγνώρισης μοτίβων (π.χ. MNIST). Επίσης, σημαντικά αναπτύχθηκε ο χώρος της ταξινόμησης προτύπων με τη δημιουργία ή βελτίωση αρκετών μοντέλων όπως οι μηχανές διανυσματικής υποστήριξης με μέθοδο πυρήνα (Support Vector Machines with kernel trick) και τα δίκτυα ακτινικής βάσης (Radial Basis Networks). Ακόμα και ο χώρος της συλλογιστικής για τον οποίο κάναμε λόγο σε προηγούμενες παραγράφους εμπλουτίστηκε με μια σχολαστική και αποδοτική - αυτή τη φορά - μοντελοποίηση αβεβαιότητας της γνώσης μέσω της ανάπτυξης Μπεϋζιανών δικτύων (Bayesian networks). Τέλος, ο χώρος της στατιστικής γνώρισε πρόοδο αφού στις παραδοσιακές τεχνικές συμπερασματολογίας προστέθηκαν αυτές βασιζόμενες σε μηχανική μάθηση.

Προς τα τέλη της δεκαετίας του 1990 και τις αρχές του 2000 η έρευνα είχε εστιάσει σε κλάδους της Τεχνητής Νοημοσύνης που δε σχετίζονταν με τα νευρωνικά δίκτυα. Η κατάσταση αυτή όμως σύντομα αναστράφηκε. Αρχικά, η μεγάλη υπολογιστική ισχύ που απαιτούσαν αλγόριθμοι εκπαίδευσης βαθιών νευρωνικών δικτύων δε διευκόλυναν τον πειραματισμό [33]. Παρόλα αυτά, χάρη σε τρεις ερευνητές (Geoffrey Hinton, Yann LeCun και Yoshua Bengio) που χρηματοδοτούνταν από το Καναδικό Ινστιτούτο για προηγμένη έρευνα (CIFAR) η ενασχόληση με τα νευρωνικά δίκτυα κρατήθηκε ζωντανή οδηγώντας τελικά σε αξιοσημείωτη πρόοδο. Το πρώτο ορόσημο των βαθιών νευρωνικών δικτύων ήταν το 2006 όπου οι Hinton et al. [41] απέδειξαν ότι ένα είδος βαθέος νευρωνικού δικτύου - τα βαθιά δίκτυα πίστης (deep belief networks) - μπορούν να εκπαιδευτούν αποδοτικά και γρήγορα μέσω ενός άπληστου (greedy) αλγορίθμου. Το δεύτερο ορόσημο που απέδειξε τις προοπτικές των βαθιών νευρωνικών δικτύων ήταν η επιτυχία αυτής της τεχνολογίας σε διαγωνισμούς ταξινόμησης εικόνων της βάσης ImageNet το 2010 και το 2012. Στη δημοσίευσή τους ImageNet Classification with Deep Convolutional Neural Networks, οι A. Krizhevsky et al. [42] περιγράφουν μια νέα αρχιτεκτονική νευρωνικών δικτύων (βαθιά συνελικτικά δίκτυα) αλλά και πρωτοπόρες μεθόδους για αποδοτική εκπαίδευση (dropout, ReLU activation function κλπ.). Έκτοτε, το ενδιαφέρον απογειώθηκε και σε συνδυασμό με την αυξημένη διαθεσιμότητα δεδομένων, εξειδικευμένων συσκευών για παράλληλη επεξεργασία και αποδοτικών αλγορίθμων εκπαίδευσης δημιούργησαν το πιο πρόσφορο έδαφος για την άνθιση της Τεχνητής Νοημοσύνης.

Σήμερα, η εξέλιξη της Τεχνητής Νοημοσύνης και δη των νευρωνικών δικτύων είναι ραγδαία. Για παράδειγμα, με τη χρήση μοντέλων νευρωνικών δικτύων βασισμένων στον μηχανισμό προσοχής (attention-based neural networks) όπως τα λεγόμενα transformers, σημειώθηκε αξιοσημείωτη πρόοδος τόσο στον κλάδο της επεξεργασίας φυσικής γλώσσας (natural language processing) (βλέπε GPT -3) όσο και στον χώρο της όρασης υπολογιστών (βλέπε Vision Transformer και CoAtNet). Αναλυτικότερα για το Generative Pre-trained Transformer 3 - GPT

3, πρόκειται για το γλωσσικό μοντέλο που μπορεί να δημιουργήσει κείμενο σε φυσική γλώσσα ακόμα και να διατηρήσει για εύλογο χρόνο συζήτηση με έναν άνθρωπο. Ακόμα, σημαντική πρόοδος παρατηρείται στη μέθοδο μάθησης με αυτο-επίβλεψη (self-supervision) επιτρέποντας έτσι την εκπαίδευση δικτύων χωρίς να απαιτείται η σχολαστική και χρονοβόρα ανάθεση ετικετών στα δεδομένα εκπαίδευσης. Τέλος, μεταξύ άλλων, σπουδαία εξέλιξη υπήρξε πρόσφατα στις εφαρμογές όπου δοθέντων μερικών εικόνων από ένα αντικείμενο συνθέτουν εικόνες αυτού από νέες γωνίες θέασης (Novel View Synthesis). Ενδεικτικά, πρωτοπόρα έργα στον χώρο αυτό είναι το NeRF και το GIRAFFE [43].

Κλείνοντας το ιστορικό αυτό σημείωμα, δεν μπορώ παρά να αντικρίσω με δέος το μέλλον που επιφέρει η τεχνολογία της Τεχνητής Νοημοσύνης. Ανύπαρκτη πριν έναν αιώνα, σήμερα είναι μέρος της καθημερινότητά μας με τεράστια ορμή που δε φαίνεται να κατευνάζει. Αν μέσα σε μερικές δεκαετίες έχει τόσες δυνατότητες, στο εγγύς μέλλον η δύναμή της θα είναι τεράστια. Δύναμη που θα δημιουργήσει μια επίγεια ουτοπία ή θα αποτελέσει ένα ακόμα τουβλάκι στην οικοδόμηση της κοινωνίας του ρίσκου [3](;)... ο χρόνος θα δείξει.

[3]Ο όρος «κοινωνία του ρίσκου» (risk society) είναι δανεισμένος από το ομώνυμο βιβλίο του Ulrich Beck [44] όπου περιγράφεται το χαρακτηριστικό των μοντέρνων κοινωνιών να οργανώνονται γύρω από νέες μορφές, απαρατήρητου ρίσκου (όπως αυτό της κλιματικής αλλαγής).