ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΠΜΣ "Δημόσια Διοίκηση - Δημόσιο Μάνατζμεντ"

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

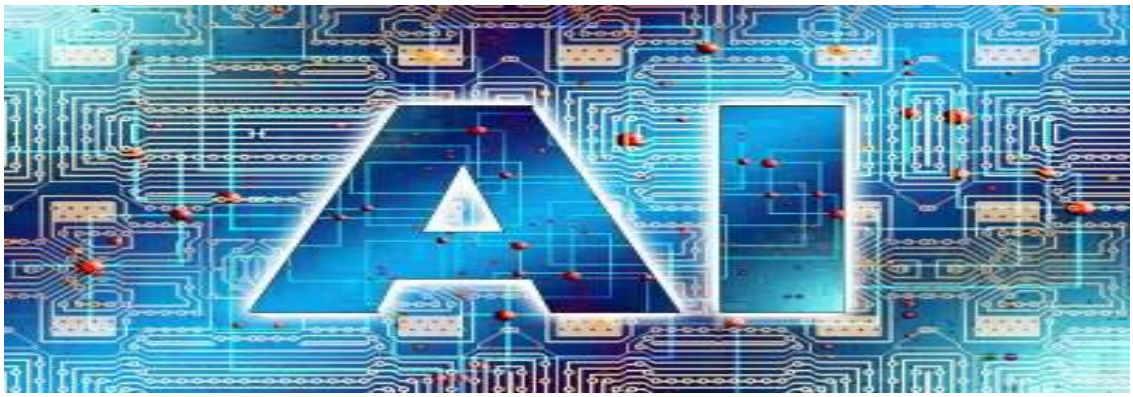
της

Βαρβάρας Ζαχαράκη

Α.Μ.: ΔΜ2088

Θέμα: «Τεχνητή Νοημοσύνη, λήψη αποφάσεων με χρήση αλγορίθμων, ηθική και δεοντολογία στο Δημόσιο Τομέα»

"Artificial Intelligence, Decision-Making using Algorithms, Ethics in the Public Sector"



Επιβλέπων καθηγητής: Νικόλαος Τσότσολας

Ιούνιος 2022

Μέλη Τριμελούς Επιτροπής

Nikolaos Tsotsolas

Digitally signed by Nikolaos Tsotsolas
Date: 2022.11.05 03:01:05 +02'00'

APOSTOLOS MANTHOS

Digitally signed by APOSTOLOS MANTHOS
Date: 2022.11.05 18:56:47 +02'00'

Dimitrios Stranis

Digitally signed by Dimitrios Stranis
Date: 2022.11.05 19:26:36 +02'00'

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Η κάτωθι υπογεγραμμένη ΒΑΡΒΑΡΑ ΖΑΧΑΡΑΚΗ του ΝΙΚΟΛΑΟΥ, με αριθμό μητρώου ΔΜ2088 φοιτήτρια του Προγράμματος Μεταπτυχιακών Σπουδών ΔΗΜΟΣΙΑ ΔΙΟΙΚΗΣΗ-ΔΗΜΟΣΙΟ ΜΑΝΑΤΖΜΕΝΤ του Τμήματος Διοίκησης Επιχειρήσεων της Σχολής Διοικητικών, Οικονομικών και Κοινωνικών Επιστημών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

VARVARA ZACHARAKI
Oct 31 2022 10:59 AM
VARVARA ZACHARAKI

Ευχαριστίες

Η παρούσα διπλωματική εργασία είναι το αποτέλεσμα μιας δημιουργικής διαδρομής ανακάλυψης νέας γνώσης που αντανακλά μια εκσυγχρονισμένη αντίληψη για την οργάνωση, διοίκηση και λειτουργία της Δημόσιας Διοίκησης που ξεκίνησε από την παρακολούθηση του *Μεταπτυχιακού Προγράμματος Σπουδών στη Δημόσια Διοίκηση – Δημόσιο Μάνατζμεντ* και ολοκληρώθηκε μέσω της εξαιρετικής προσπάθειας των καθηγητών μου και την υποστήριξη των συμφοιτητών μου.

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή, κ. Νικόλαο Τσότσολα, ο οποίος με βοήθησε να αντιμετωπίσω το θέμα της διπλωματικής μου εργασίας με έναν πολυεπίπεδο και αναλυτικό τρόπο και τους αγαπημένους μου συμφοιτητές που λειτούργησαν τόσο ως διαδικτυακοί όσο κι ως πραγματικοί φίλοι .

Η διπλωματική μου εργασία είναι αφιερωμένη στην οικογένειά μου για την υπομονή, την κατανόηση και την ηθική συμπαράσταση που μου πρόσφερε στο δύσκολο αυτό εγχείρημα.

Περίληψη

Η *Τεχνητή Νοημοσύνη (Artificial Intelligence)* αποτελεί την πιο καθοριστική τεχνολογία της τελευταίας δεκαετίας και πιθανώς και της επόμενης. Ωστόσο, το μέλλον μας με την Τεχνητή Νοημοσύνη θα εξαρτηθεί από την ικανότητά μας να κατανοήσουμε πλήρως, την πολυπλοκότητα των συστημάτων και των εφαρμογών της, ώστε να είμαστε σε θέση να αντιμετωπίσουμε με τον βέλτιστο δυνατό τρόπο τον αρνητικό αντίκτυπο σε άτομα και κοινωνίες. Θα πρέπει να παραμείνουμε προσεκτικοί αν θέλουμε να διατηρήσουμε τον έλεγχο και να βεβαιωθούμε ότι η Τεχνητή Νοημοσύνη θα είναι συμπληρωματική με την ανθρώπινη νοημοσύνη και θα εκπληρώσει τις θετικές της δυνατότητες, θα βελτιώσει τη ζωή μας και τελικά θα ωφελήσει την ανθρωπότητα.

Η επικινδυνότητα της Τεχνητής Νοημοσύνης δεν θα πρέπει να θεωρείται αποτρεπτικός παράγοντας για τη χρήση της στη διαχείριση και στην παροχή δημόσιων υπηρεσιών, αλλά η εφαρμογή της στο δημόσιο τομέα θα πρέπει να αποτελεί μια στοχαστική και στρατηγική πορεία δράσης για την αξιοποίηση των μεγάλων ευκαιριών που υπόσχεται και τελικά τη δημιουργία αξίας από αυτές.

Οι *ηθικές ανησυχίες (ethical concerns)* που προκύπτουν από την *αλγοριθμική λήψη αποφάσεων (algocracy)* και η λειτουργία των αλγοριθμικών συστημάτων με τεχνικές και μεθόδους *μαύρου κουτιού (black box)*, θα μπορούσαν να μεταφραστούν σε κινδύνους συστημικού χαρακτήρα, οι οποίοι, εάν αποτύχουμε να διασφαλίσουμε την εφαρμογή επαρκών μέτρων, μπορεί να διαβρώσουν τα ηθικά, πολιτικά και πολιτιστικά θεμέλια της κοινωνίας και να υπονομεύσουν το δημοκρατικό πολιτικό σύστημα και μαζί του την ατομική μας ελευθερία, αυτονομία και ικανότητα αυτοδιάθεσης.

Σκοπός της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι να μελετηθεί η έννοια της Τεχνητής Νοημοσύνης ως ενός συνόλου τεχνολογιών που συνδυάζουν δεδομένα, αλγορίθμους και υπολογιστική ισχύ και να απεικονιστεί η εξελικτική πορεία της τεχνολογίας, εστιάζοντας στην υιοθέτηση και χρήση εφαρμογών Τεχνητής Νοημοσύνης στο δημόσιο τομέα, στις ηθικές ανησυχίες που προκύπτουν από το μείζον θέμα της λήψης αποφάσεων με χρήση αλγορίθμων, καθώς και στους προταθέντες θεσμικούς τρόπους αντιμετώπισης των προβλημάτων που εγείρονται.

Abstract

Title: “Artificial Intelligence, Decision-Making using Algorithms, Ethics in Public Sector.”

Artificial Intelligence is the most crucial technology of the last decade and probably the next. However, our future with Artificial Intelligence will depend on our ability to fully understand the complexity of its systems and applications, so that we can best deal with the negative impact on individuals and societies. We must remain vigilant if we are to maintain control and make sure that Artificial Intelligence is complementary to human intelligence and fulfills its positive potential, improves our lives and ultimately benefits humanity.

The riskiness of Artificial Intelligence should not be considered as a deterrent factor to its use in the management and provision of public services, but its application in the public sector should be a thoughtful and strategic course of action to take advantage of the great opportunities it promises and ultimately creating value from them.

Ethical concerns arising from *algorithmic decision-making (algocracy)* and the operation of algorithmic systems using *black box* techniques and methods could translate into systemic risks, which, if we fail to safeguard the implementation of adequate measures, can erode the moral, political and cultural foundations of society and undermine the democratic political system and with it our individual freedom, autonomy and capacity for self-determination.

The purpose of this master's thesis is to study the concept of Artificial Intelligence as a set of technologies that combine data, algorithms and computing power and to illustrate the evolutionary path of the technology, focusing on the adoption and use of Artificial Intelligence applications in the public sector, ethical concerns arising from the major issue of algorithmic decision-making, as well as the proposed institutional ways of dealing with the problems that arise.

Περιεχόμενα

1. Εισαγωγή 9
2. Τεχνητή Νοημοσύνη 12
   2.1 Ορισμός 12
      2.1.1 Ταξινομία Παρατηρητηρίου για την Τεχνητή Νοημοσύνη - (AI Watch) 14
   2.2 Η εξελικτική πορεία της Τεχνητής Νοημοσύνης 17
      2.2.1 Συμβολική Τεχνητή Νοημοσύνη (Symbolic AI) 17
      2.2.2 Μηχανική Μάθηση (Machine Learning ML) και Δεδομένα 19
      2.2.3 Μελλοντικές θεωρητικές προσεγγίσεις εξέλιξης Τεχνητής Νοημοσύνης 29
3. Τεχνητή Νοημοσύνη και Δημόσιος Τομέας 35
   3.1 Περιπτώσεις χρήσης Τεχνητής Νοημοσύνης στο Δημόσιο Τομέα 39
      3.1.1 Εθνική Στρατηγική Τεχνητής Νοημοσύνης – 40
      Το παράδειγμα της Φινλανδίας 40
      3.1.2 Τεχνητή Νοημοσύνη και Μεταφορές - Το παράδειγμα του Καναδά “bomb-in-the-box” 42
      3.1.3 Τεχνητή Νοημοσύνη και Αντιμετώπιση Διαφθοράς – 44
      Το παράδειγμα της Βραζιλίας 44
      3.1.4 Τεχνητή Νοημοσύνη και Αντιμετώπιση Διαφθοράς - Το παράδειγμα της Κίνας 46
      3.1.5 Τεχνητή Νοημοσύνη και Εμπλοκή Πολιτών - Το παράδειγμα του Βελγίου 47
      3.1.6 Τεχνητή Νοημοσύνη και Εμπλοκή Πολιτών - Το παράδειγμα της Νιγηρίας 49
      3.1.7 Τεχνητή Νοημοσύνη και Τελωνεία - Το παράδειγμα των Ηνωμένων Πολιτειών 51
      3.1.8 Τεχνητή Νοημοσύνη και Υγεία - Η περίπτωση της Σιγκαπούρης 52
      3.1.9 Τεχνητή Νοημοσύνη και Δικαστικός Τομέας - Το παράδειγμα της Κίνας 53
      3.1.10 Τεχνητή Νοημοσύνης και Δικαστικός Τομέας – 54
      Το παράδειγμα του Ηνωμένου Βασιλείου 54
      3.1.11 Τεχνητή Νοημοσύνη και Δημόσιες Προμήθειες – 55
      Το παράδειγμα των Ηνωμένων Πολιτειών 55
      3.1.12 Τεχνητή Νοημοσύνη και Δημόσιες Προμήθειες – 56
      Το παράδειγμα της Νότιας Κορέας 56
      3.1.13 Τεχνητή Νοημοσύνη και Φορολογική Συμμόρφωση – 57
      Το παράδειγμα της Αρμενίας 57
      3.1.14 Τεχνητή Νοημοσύνης και Φορολογική Συμμόρφωση – 59
      Το παράδειγμα των Ηνωμένων Πολιτειών 59
4. Λήψη αποφάσεων με χρήση αλγορίθμων 61

4.1 Αλγόριθμοι 61
4.2 Αλγοκρατία 62
4.2.1 Η απειλή της Αλγοκρατίας 63
4.3 Γιατί ανησυχούμε για την αλγοριθμική λήψη αποφάσεων; 69
3.3.1 Ανησυχίες που σχετίζονται με τις διαδικασίες 70
4.3.2 Ανησυχίες με βάση το αποτέλεσμα 79
4.3.3 Πρόβλεψη βάσει δεδομένων και εξατομικευμένες υπηρεσίες πληροφόρησης 82
5. Ηθική και Δεοντολογία στο Δημόσιο Τομέα 87
5.1 «Κατευθυντήριες Γραμμές Δεοντολογίας για Αξιόπιστη Τεχνητή Νοημοσύνη» - Ευρωπαϊκή Επιτροπή 87
5.2 «Σύσταση για την Ηθική της Τεχνητής Νοημοσύνης» - UNESCO 104
Συμπεράσματα 120
Βιβλιογραφία 124

1. Εισαγωγή

Η Τεχνητή Νοημοσύνη αλλάζει σταδιακά τις ζωές μας και μεταμορφώνει τις κοινωνίες μας. Αυτή η αλλαγή θα είναι πιθανότατα η βαθύτερη και η ταχύτερη που έχει βιώσει ποτέ η ανθρωπότητα. Αποτελεί κοινό τόπο ότι η Τεχνητή Νοημοσύνη θα δημιουργήσει πολλές νέες ευκαιρίες και ότι οι περισσότερες αλλαγές που θα επιφέρει θα πρέπει να είναι θετικές προς όφελος της ανθρωπότητας. Υπάρχει, όμως, ευρεία σύγκλιση απόψεων ότι η ανάπτυξη της Τεχνητής Νοημοσύνης συνεπάγεται πολλούς κινδύνους που πρέπει να αντιμετωπιστούν πολύ προσεκτικά. Ωστόσο, αυτό που συχνά υποτιμάται είναι ότι τόσο οι θετικές όσο και οι αρνητικές επιπτώσεις των εφαρμογών της Τεχνητής Νοημοσύνης θα είναι απολύτως ανατρεπτικές.

Η Τεχνητή Νοημοσύνη είναι, ήδη, σε θέση να εκτελεί πολλές εργασίες και να λαμβάνει αποφάσεις, συχνά, καλύτερα από την ανθρώπινη νοημοσύνη. Τα επόμενα χρόνια, ο αριθμός αυτών των εργασιών και κρίσιμων αποφάσεων θα αυξηθεί. Κάποια μέρα, οι μηχανές μπορεί να φτάσουν ακόμη και στο επίπεδο αυτού που πολλοί ειδικοί αποκαλούν «Γενική Τεχνητή Νοημοσύνη», δηλαδή, μιας ευφυίας γενικού σκοπού που προσεγγίζει αυτή των ανθρώπων. Αλλά ακόμα κι αν δεν φτάσουμε ποτέ στη Γενική Τεχνητή Νοημοσύνη, η απίστευτα γρήγορη ανάπτυξη εφαρμογών Τεχνητής Νοημοσύνης αρχίζει να εγείρει μείζονα θεμελιώδη ερωτήματα και ηθικές ανησυχίες για την ανθρωπότητα, συμπεριλαμβανομένου του τι σημαίνει να είσαι άνθρωπος στην εποχή της Τεχνητής Νοημοσύνης.

Πρέπει να παραμείνουμε πολύ προσεκτικοί αν θέλουμε να διατηρήσουμε τον έλεγχο και να βεβαιωθούμε ότι η Τεχνητή Νοημοσύνη θα είναι συμπληρωματική με την ανθρώπινη νοημοσύνη και ότι θα εκπληρώσει τις θετικές της δυνατότητες, θα βελτιώσει τη ζωή μας και τελικά θα ωφελήσει την ανθρωπότητα. Ωστόσο, το μέλλον μας με την Τεχνητή Νοημοσύνη θα εξαρτηθεί πολύ από την ικανότητά μας να κατανοήσουμε πλήρως όλες τις επιπτώσεις των τεχνολογιών με δυνατότητα Τεχνητής Νοημοσύνης που αναπτύσσονται και των εφαρμογών τους, η οποία είναι εξαιρετικά περίπλοκη. Είναι σημαντικό να γνωρίζουμε ότι οι τεχνολογίες Τεχνητής Νοημοσύνης μπορούν να έχουν όχι μόνο σκόπιμα, αλλά και ακούσια αποτελέσματα. Τα επιδιωκόμενα αποτελέσματα μπορούν να βασίζονται σε καλές προθέσεις, όπως η βελτίωση της υγειονομικής περίθαλψης ή η βελτίωση της ασφάλειας ή της ποιότητας των υπηρεσιών. Μπορεί, επίσης, να έχουν κακόβουλους σκοπούς, όπως παραπληροφόρηση ή διαδικτυακή χειραγώγηση ή να έχουν ανεπιθύμητες συνέπειες με αρνητικές επιπτώσεις,

όπως ενίσχυση των προκαταλήψεων και των διακρίσεων, απώλεια ιδιωτικότητας, περιορισμοί του δικαιώματος της ελευθερίας της έκφρασης και της ανθρώπινης αξιοπρέπειας.

Η χαρτογράφηση του τοπίου της Τεχνητής Νοημοσύνης αποτελεί απαραίτητη προϋπόθεση για την παρακολούθηση της ανάπτυξης, της υιοθέτησης, της χρήσης και του αντικτύπου των τεχνολογιών, των συστημάτων και των εφαρμογών της. Ο αρνητικός αντίκτυπος σε άτομα και κοινωνίες συνεπάγεται την αξιολόγηση και επαναξιολόγηση των επιπτώσεων καθώς και τη συνεννόηση με όλα τα ενδιαφερόμενα μέρη, ώστε να διασφαλίζεται η ανάπτυξη και υιοθέτηση της ηθικής και αξιόπιστης Τεχνητής Νοημοσύνης, η οποία θα λειτουργεί στην υπηρεσία των ανθρώπων και θα αποτελεί θετική δύναμη για την κοινωνία.

Οι δημόσιες διοικήσεις έχουν ήδη αρχίσει να υιοθετούν την Τεχνητή Νοημοσύνη σε διάφορους και διαφορετικούς τομείς του δημόσιου τομέα και δεν διερευνούν απλώς τις δυνατότητές της με πιλοτικές λύσεις ή περιβάλλοντα δοκιμών, αλλά αρκετές λύσεις Τεχνητής Νοημοσύνης έχουν ήδη αναπτυχθεί και χρησιμοποιούνται σε καθημερινές λειτουργίες (Misuraca & van Noordt, 2020. Molinari et al., 2021). Η εφαρμογή, ωστόσο, της Τεχνητής Νοημοσύνης στο δημόσιο τομέα απαιτεί μια στοχαστική και στρατηγική πορεία δράσης για την αξιοποίηση των μεγάλων ευκαιριών που υπόσχεται και τελικά τη δημιουργία αξίας από αυτές, γιατί από τη μια πλευρά οι κυβερνήσεις μπορούν να χρησιμοποιήσουν τις δυνατότητες των εφαρμογών Τεχνητής Νοημοσύνης για να βελτιώσουν τις δημόσιες υποθέσεις και να αυξήσουν την αποτελεσματικότητα των εσωτερικών διαδικασιών, ενώ από την άλλη, οι απειλές υποδεικνύουν ότι η Τεχνητή Νοημοσύνη χρειάζεται πολιτικές και ρυθμίσεις βασισμένες σε αρχές και βασικές κοινωνικές αξίες προκειμένου να αποφέρει οφέλη σε όλους (Wirtz et al., 2018. Mehr, 2017. Boyd et Wilson, 2017).

Για πολλά χρόνια η έρευνα σχετικά με τις ηθικές επιπτώσεις της Τεχνητής Νοημοσύνης υστερούσε σε σχέση με την ανάπτυξη της τεχνολογικής γνώσης. Αυτή η κατάσταση άρχισε να αλλάζει μόλις το 2017, με το 90% της σχετικής γνώσης να έχει δημοσιευτεί μετά από αυτή την ημερομηνία. Απόρροια της έρευνας είναι η σύνταξη πάνω από ογδόντα (80) επίσημων κείμενων ηθικών αρχών και κατευθυντήριων γραμμών για την Τεχνητή Νοημοσύνη που έχουν αναπτυχθεί από την κοινωνία των πολιτών, τον ιδιωτικό τομέα, τις κυβερνήσεις, τους διακυβερνητικούς και διεθνείς οργανισμούς και αφορούν μια σειρά κοινών θεμάτων προς αντιμετώπιση, όπως το απόρρητο, η λογοδοσία, η ασφάλεια, η διαφάνεια και η επεξήγηση, η δικαιοσύνη και η μη διάκριση, ο ανθρώπινος έλεγχος της

τεχνολογίας, η επαγγελματική ευθύνη και η προώθηση των ανθρώπινων αξιών. Κύριο στόχο τους δεν αποτελεί μόνο η προαγωγή και η διασφάλιση ηθικής και αξιόπιστης Τεχνητής Νοημοσύνης βάσει των δεοντολογικών αρχών, αλλά και η παροχή κατευθύνσεων σχετικά με τους τρόπους με τους οποίους αυτές οι αρχές είναι δυνατόν να υλοποιηθούν επιχειρησιακά στο πλαίσιο των κοινωνικοτεχνικών συστημάτων (European Commission, 2021).

Σκοπός της παρούσας μεταπτυχιακής διπλωματικής εργασίας είναι να μελετηθεί η έννοια της Τεχνητής Νοημοσύνης ως ενός συνόλου τεχνολογιών που συνδυάζουν δεδομένα, αλγορίθμους και υπολογιστική ισχύ και να απεικονιστεί η εξελικτική πορεία της τεχνολογίας, εστιάζοντας στην υιοθέτηση και χρήση εφαρμογών Τεχνητής Νοημοσύνης στο δημόσιο τομέα, στις ηθικές ανησυχίες που προκύπτουν από το μείζον θέμα της λήψης αποφάσεων με χρήση αλγορίθμων, καθώς και στους προταθέντες θεσμικούς τρόπους αντιμετώπισης των προβλημάτων που εγείρονται.

- ✓ Στο πρώτο κεφάλαιο εισάγεται ο σκοπός της διπλωματικής εργασίας και το πλαίσιο βάσει του οποίου δομείται.
- ✓ Στο δεύτερο κεφάλαιο γίνεται αναφορά στην έννοια της Τεχνητής Νοημοσύνης και στην εξελικτική της πορεία.
- ✓ Το τρίτο κεφάλαιο αναφέρεται στον τρόπο με τον οποίο οι εθνικές στρατηγικές στοχεύουν να ενισχύσουν τη χρήση της Τεχνητής Νοημοσύνης στο δημόσιο τομέα και παρουσιάζει παραδείγματα χρήσης της τεχνολογίας που έχουν ήδη υιοθετηθεί σε διάφορες χώρες του κόσμου για την αντιμετώπιση συγκεκριμένων προκλήσεων και ζητημάτων του δημόσιου τομέα.
- ✓ Το τέταρτο κεφάλαιο εστιάζει στις ηθικές ανησυχίες που προκύπτουν από την εφαρμογή της Τεχνητής Νοημοσύνης και τη λήψη αποφάσεων με χρήση αλγορίθμων.
- ✓ Το πέμπτο κεφάλαιο παρουσιάζει τις δύο πιο σημαντικές διεθνώς θεσμικές προσπάθειες αντιμετώπισης των σχετικών με την Τεχνητή Νοημοσύνη ηθικών ανησυχιών και ζητημάτων με την παροχή κατευθύνσεων και συστάσεων για την επίτευξη ηθικής και αξιόπιστης Τεχνητής Νοημοσύνης.

2. Τεχνητή Νοημοσύνη

2.1 Ορισμός

Η Τεχνητή Νοημοσύνη αποτελεί την πιο καθοριστική τεχνολογία της τελευταίας δεκαετίας και πιθανώς και της επόμενης. Είναι ένας τομέας στρατηγικής σημασίας και έχει αναγνωριστεί ως δυνητικός βασικός μοχλός οικονομικής ανάπτυξης. Ωστόσο, περιγράφεται υπεραπλουστευμένα σχετιζόμενη, συνήθως, με την ανθρώπινη νοημοσύνη και τη νοημοσύνη γενικά, με ορισμούς, οι οποίοι αναφέρονται σε μηχανές που συμπεριφέρονται σαν άνθρωποι ή είναι ικανές για ενέργειες που απαιτούν ευφυΐα. Στην πραγματικότητα, η έννοια της νοημοσύνης, αν και έχει μελετηθεί εκτενώς από ψυχολόγους, βιολόγους και νευροεπιστήμονες, συνεχίζει να αποτελεί μια ασαφή έννοια, δύσκολα προσδιορίσιμη και μετρήσιμη. Η ασάφεια της έννοιας «νοημοσύνη» και ο προσδιορισμός κάτι τόσο υποκειμενικού και αφηρημένου, οδηγεί σε ορισμούς που δίνουν την εντύπωση μιας ακρίβειας που δεν μπορεί να επιτευχθεί, προτείνοντας έναν ιδανικό στόχο και όχι μια μετρήσιμη ερευνητική ιδέα.

Οι πολλαπλές πτυχές της Τεχνητής Νοημοσύνης και η υιοθέτηση διαφορετικών προσεγγίσεων από εθνικά κράτη, Ευρωπαϊκή Ένωση, δημόσιο και ιδιωτικό τομέα, παγκόσμιους οργανισμούς και ακαδημαϊκή κοινότητα, έχει ως αποτέλεσμα την ύπαρξη πολλαπλών ορισμών της, οι οποίοι αντικατοπτρίζουν την διαφορετική στρατηγική στοχοθέτηση των αντίστοιχων πολιτικών, οικονομικών, πολιτιστικών και κοινωνικών συστημάτων. Η ανυπαρξία, ωστόσο, ενός καθολικά αποδεκτού τυπικού ορισμού για το τι περιλαμβάνει στην πραγματικότητα η Τεχνητή Νοημοσύνη, συχνά δυσκολεύει μια κοινή κατανόηση του τομέα, των δυνατοτήτων, του πεδίου και των επιπτώσεων εφαρμογής του.

Η Ευρωπαϊκή Επιτροπή στην προσπάθειά της για κοινή κατανόηση του εν λόγω επιστημονικού κλάδου αποδίδει έναν πρώτο ορισμό κατανοητό και συγχρόνως τυπικό για την Τεχνητή Νοημοσύνη:

«Η Τεχνητή Νοημοσύνη (ΤΝ) αναφέρεται σε συστήματα που χαρακτηρίζονται από ευφυή συμπεριφορά, αναλύοντας το περιβάλλον τους και ενεργώντας – με κάποιο βαθμό αυτονομίας – για την επίτευξη συγκεκριμένων στόχων.

Τα συστήματα που λειτουργούν βάσει Τεχνητής Νοημοσύνης μπορούν να βασίζονται αποκλειστικά σε λογισμικό, ενεργώντας στον εικονικό κόσμο (π.χ. βοηθοί φωνής, λογισμικό ανάλυσης εικόνας, μηχανές αναζήτησης, συστήματα αναγνώρισης ομιλίας και προσώπου) ή η

τεχνητή νοημοσύνη μπορεί να ενσωματωθεί σε συσκευές υλισμικού (π.χ. προηγμένα ρομπότ, αυτόνομα αυτοκίνητα, δρόνοι ή εφαρμογές του Διαδικτύου των Πραγμάτων).

Χρησιμοποιούμε την τεχνητή νοημοσύνη σε καθημερινή βάση, π.χ. για να μεταφράσουμε γλώσσες, να υποτιτλίσουμε βίντεο ή να μπλοκάρουμε ανεπιθύμητη ηλεκτρονική αλληλογραφία.

Πολλές τεχνολογίες Τεχνητής Νοημοσύνης απαιτούν δεδομένα προκειμένου να βελτιώσουν τις επιδόσεις τους. Από τη στιγμή που θα λειτουργήσουν σωστά, μπορούν να υποστηρίξουν τη βελτίωση και την αυτοματοποίηση της λήψης αποφάσεων στον ίδιο τον τομέα. Για παράδειγμα, ένα σύστημα Τεχνητής Νοημοσύνης θα εκπαιδευτεί και ακολούθως, θα χρησιμοποιηθεί για να εντοπίζει κυβερνοεπιθέσεις βάσει των δεδομένων του οικείου δικτύου ή συστήματος».[1]

Ο ορισμός αυτός βελτιώθηκε περαιτέρω από την Ομάδα Εμπειρογνωμόνων Υψηλού Επιπέδου (HLEG) για την Τεχνητή Νοημοσύνη, η οποία απέδωσε έναν περισσότερο τεχνικό ορισμό:

«Τα συστήματα Τεχνητής Νοημοσύνης (AI) είναι συστήματα λογισμικού (και πιθανώς και υλισμικού) σχεδιασμένα από ανθρώπους που, δεδομένου ενός πολύπλοκου στόχου, δρουν στη φυσική ή ψηφιακή διάσταση αντιλαμβανόμενοι το περιβάλλον τους μέσω απόκτησης δεδομένων, ερμηνείας των συλλεγόμενων δομημένων ή μη δεδομένων, συλλογισμών σχετικά με τη γνώση ή την επεξεργασία των πληροφοριών που προέρχονται από αυτά τα δεδομένα και την απόφαση για τις καλύτερες ενέργειες που πρέπει να γίνουν για την επίτευξη του δεδομένου στόχου. Τα συστήματα AI μπορούν είτε να χρησιμοποιήσουν συμβολικούς κανόνες είτε να μάθουν ένα αριθμητικό μοντέλο και μπορούν επίσης να προσαρμόσουν τη συμπεριφορά τους αναλύοντας πώς επηρεάζεται το περιβάλλον από τις προηγούμενες ενέργειές τους. Ως επιστημονικός κλάδος, η Τεχνητή Νοημοσύνη περιλαμβάνει διάφορες προσεγγίσεις και τεχνικές, όπως η μηχανική μάθηση (εκ των οποίων συγκεκριμένα παραδείγματα είναι η βαθιά μάθηση και η ενισχυτική μάθηση), η μηχανική συλλογιστική (που περιλαμβάνει σχεδιασμό, προγραμματισμό, αναπαράσταση και συλλογιστική γνώσης, αναζήτηση και βελτιστοποίηση) και η ρομποτική (η οποία περιλαμβάνει έλεγχο, αντίληψη, αισθητήρες και ενεργοποιητές, καθώς και την ενσωμάτωση όλων των άλλων τεχνικών σε κυβερνοφυσικά συστήματα)».[2]

[1] COM(2018) 237 final
[2] COM(2020) 65 final

Με βάση τον εν λόγω ορισμό, η Ομάδα Εμπειρογνωμόνων Υψηλού Επιπέδου ομαδοποίησε τις τεχνικές και τους επιμέρους κλάδους της Τεχνητής Νοημοσύνης σε δύο σκέλη σε σχέση με τις δυνατότητες των συστημάτων:

1. **Συλλογιστική και λήψη αποφάσεων**

   Η πρώτη ομάδα δυνατοτήτων περιλαμβάνει τη μετατροπή δεδομένων σε γνώση, με τη μετατροπή των πληροφοριών του πραγματικού κόσμου σε κάτι κατανοητό και χρησιμοποιήσιμο από μηχανές και τη λήψη αποφάσεων, ακολουθώντας μια οργανωμένη διαδρομή σχεδιασμού, αναζήτησης λύσεων και βελτιστοποίησης. Αυτό το σκέλος καλύπτει υποτομείς Τεχνητής Νοημοσύνης όπως της Αναπαράστασης και του Συλλογισμού της Γνώσης (συνήθως χρησιμοποιώντας συμβολικούς κανόνες για την αναπαράσταση και την εξαγωγή γνώσεων) και του Προγραμματισμού (συμπεριλαμβανομένου του Σχεδιασμού και του Προγραμματισμού, της Αναζήτησης και της Βελτιστοποίησης).

2. **Μάθηση και αντίληψη**

   Η δεύτερη ομάδα δυνατοτήτων αναπτύσσεται ελλείψει συμβολικών κανόνων και περιλαμβάνει μάθηση, δηλαδή, εξαγωγή πληροφοριών και επίλυση προβλημάτων με βάση δομημένα ή μη δομημένα αντιληπτά δεδομένα (γραπτή ή προφορική γλώσσα, εικόνα, ήχος κλπ), προσαρμογή και αντίδραση σε αλλαγές, πρόβλεψη συμπεριφοράς κλπ. Αυτό το δεύτερο σκέλος καλύπτει υποτομείς Τεχνητής Νοημοσύνης που σχετίζονται με τη μάθηση, την επικοινωνία και την αντίληψη, όπως η Μηχανική Μάθηση, η Επεξεργασίας Φυσικής Γλώσσας, Υπολογιστική Όραση.

2.1.1 Ταξινομία Παρατηρητηρίου για την Τεχνητή Νοημοσύνη - (AI Watch)[3]

Το Παρατηρητήριο για την Τεχνητή Νοημοσύνη, λαμβάνοντας ως σημείο εκκίνησης τον προαναφερθέντα ορισμό της Ομάδα Εμπειρογνωμόνων Υψηλού Επιπέδου (HLEG) για την Τεχνητή Νοημοσύνη και τη σχετική κατηγοριοποίηση των δυνατοτήτων των συστημάτων Τεχνητής Νοημοσύνης, προχώρησε στη δημιουργία ενός λειτουργικού ορισμού της

[3] Το AI-Watch , το οποίο αναπτύχθηκε από το Κοινό Κέντρο Ερευνών της Ευρωπαϊκής Επιτροπής, έχει ευρωπαϊκή εστίαση στο παγκόσμιο τοπίο της Τεχνητής Νοημοσύνης. Σκοπός του είναι η παρακολούθηση της εφαρμογής του «Συντονισμένου σχεδίου για την Τεχνητή Νοημοσύνη» που αφορά στην ανάπτυξη της ΤΝ στην Ευρωπαϊκή Ένωση, με στόχευση στην παρακολούθηση της βιομηχανικής, τεχνολογικής και ερευνητικής ικανότητας καθώς και των πολιτικών πρωτοβουλιών των κρατών-μελών, της υιοθέτησης και των τεχνικών εξελίξεων της ΤΝ και του αντικτύπου της εφαρμογής της στην Ευρώπη.

Τεχνητής Νοημοσύνης, μέσω της δημιουργίας μιας ταξινομίας[4] και μιας λίστας λέξεων-κλειδιών που είναι απαραίτητα για τη χαρτογράφηση του οικοσυστήματος της Τεχνικής Νοημοσύνης και την ανίχνευση εφαρμογών Τεχνητής Νοημοσύνης σε άλλους τεχνολογικούς τομείς, όπως η ρομποτική, τα Μεγάλα Δεδομένα, οι τεχνολογίες Ιστού, οι υπολογιστές υψηλής απόδοσης, τα ενσωματωμένα συστήματα, το Διαδίκτυο των Πραγμάτων. Επομένως, η ταξινόμηση του Παρατηρητηρίου βασίζεται στους κύριους τομείς Τεχνητής Νοημοσύνης που προσδιορίζονται από την Ομάδα Εμπειρογνωμόνων Υψηλού Επιπέδου και επεκτείνεται για να καλύψει πρόσθετες διαστάσεις όπως:

- την έννοια των λογικών πρακτόρων, ως οντοτήτων που λαμβάνουν αποφάσεις και ενεργούν σε σχέση με το περιβάλλον τους, συμπεριλαμβανομένης της αλληλεπίδρασης με άλλους πράκτορες,
- την έρευνα και τις βιομηχανικές εξελίξεις και άλλες εφαρμογές Τεχνητής Νοημοσύνης, όπως τα μοντέλα υπηρεσιών cloud που προσφέρονται από εταιρίες παροχής υπηρεσιών για την επιτάχυνση της πρόσληψης της Τεχνητής Νοημοσύνης ,
- άλλες σημαντικές πτυχές που σχετίζονται με την Τεχνητή Νοημοσύνη και προκύπτουν ως σημαντικά θέματα στα έγγραφα πολιτικής και στον κοινωνικό διάλογο και αφορούν σε ηθικά ζητήματα, όπως η διαφάνεια, η λογοδοσία, η επεξηγησιμότητα, η δικαιοσύνη και η ασφάλεια, όπως επίσης και σε φιλοσοφικά ζητήματα που αναφέρονται στη φύση της Τεχνητής Νοημοσύνης και στην εξέλιξή της.

Το Παρατηρητήριο για την Τεχνητή Νοημοσύνη, βάσει των ανωτέρω και ενσωματώνοντας στον σχεδιασμό του τα τέσσερα (4) βασικά χαρακτηριστικά, τα οποία εντοπίστηκαν ύστερα από επεξεργασία και ανάλυση πενήντα πέντε (55) ορισμών της Τεχνητής Νοημοσύνης που αναπτύχθηκαν μεταξύ 1955 και 2019 και αφορούν στην:

1. αντίληψη του περιβάλλοντος και στην πολυπλοκότητα του πραγματικού κόσμου,
2. επεξεργασία πληροφοριών που αφορά τόσο στη συλλογή όσο και στην επεξεργασία εισροών σε μορφή δεδομένων,

[4] Μια ταξινομία είναι ένα ιεραρχικό σύστημα ταξινόμησης που λειτουργεί ως τρόπος οργάνωσης γνώσεων: ένα δέντρο που ξεκινάει από μια γενική ιδέα ρίζας και σταδιακά χωρίζεται σε πιο συγκεκριμένες έννοιες. Οι κόμβοι μιας ταξινομίας αντιστοιχούν σε έννοιες που συνδέονται με κλάδους ή άκρα που κατευθύνονται από τον κόμβο ρίζας προς τους κόμβους των φύλλων. Οι ταξινομίες αντιπροσωπεύουν μια συλλογή θεμάτων με σχέσεις "is-a".

3. λήψη αποφάσεων με συγκεκριμένο επίπεδο αυτονομίας που περιλαμβάνει το συλλογισμό και τη μάθηση για ανάληψη δράσεων και εκτέλεση καθηκόντων, συμπεριλαμβανομένης της προσαρμογής και της αντίδρασης στις αλλαγές του περιβάλλοντος,

4. επίτευξη προκαθορισμένων στόχων που αποτελεί τον κύριο σκοπό των συστημάτων Τεχνητής Νοημοσύνης,

προχώρησε στον σχηματισμό ταξινομίας και αντιπροσωπευτικής επιλογής λέξεων-κλειδιών, αναγνωρίζοντας τρεις συμπληρωματικές προοπτικές προσέγγισης υπό τις οποίες εξετάζει την Τεχνητή Νοημοσύνη και αφορούν στην:

- ✓ πολιτική και θεσμική προοπτική (σε επίπεδο Ευρωπαϊκής Επιτροπής, σε εθνικό επίπεδο, σε επίπεδο διεθνών οργανισμών) που αντιμετωπίζει την Τεχνητή Νοημοσύνη ως μέσο ανάπτυξης και τεχνολογικής εξέλιξης και εστιάζει στην ανάπτυξη του κλάδου της Τεχνητής Νοημοσύνης, στην ερευνητική ικανότητα και στον αντίκτυπο των προηγμένων τεχνολογιών της στην κοινωνία,
- ✓ ερευνητική προοπτική που αφορά στην κατανόηση της Τεχνητής Νοημοσύνης ως ερευνητικού πεδίου και στην ανάπτυξή της, ως τεχνολογίας γενικού σκοπού,
- ✓ προοπτική της αγοράς που επικεντρώνεται στην βιομηχανική ανάπτυξη και αξιολόγηση της οικονομικής αξίας και στις μελλοντικές προοπτικές της αγοράς.

Η ταξινομία του AI-Watch είναι μια περιληπτική λίστα των βασικών τομέων και των σχετικών υποτομέων τους που χαρακτηρίζουν το πεδίο της Τεχνητής Νοημοσύνης. Χωρίζονται σε βασικούς και εγκάρσιους τομείς, με τους πρώτους να αναφέρονται στους κύριους θεωρητικούς κλάδους της Τεχνητής Νοημοσύνης που αποτελούν και τους θεμελιώδεις στόχους της και τους άλλους να αναφέρονται ως θέματα που σχετίζονται άμεσα με όλους τους βασικούς τομείς. Οι υποτομείς της Τεχνητής Νοημοσύνης αντιπροσωπεύονται από μια λίστα λέξεων-κλειδιών που αφορούν στις δραστηριότητες της Τεχνητής Νοημοσύνης που πραγματοποιούνται από οικονομικούς παράγοντες και επιτρέπουν την περαιτέρω ανάλυση του τοπίου της Τεχνητής Νοημοσύνης από τεχνικοοικονομική άποψη. Η ταξινομία του Παρατηρητηρίου, επομένως, αντιπροσωπεύει και διασυνδέει όλους τους υποτομείς της Τεχνητής Νοημοσύνης από πολιτική, ερευνητική και βιομηχανική άποψη με στόχο να καλύψει και να ταξινομήσει το τοπίο της Τεχνητής Νοημοσύνης, το οποίο αποτελείται από οικονομικούς παράγοντες με δραστηριότητες έρευνας και ανάπτυξης ή βιομηχανικές δραστηριότητες Τεχνητής Νοημοσύνης (European Commission, 2020).

2.2 Η εξελικτική πορεία της Τεχνητής Νοημοσύνης

Η έλλειψη ενός κοινά κατανοητού ορισμού της Τεχνητής Νοημοσύνης έχει ως αποτέλεσμα τη χρήση της Τεχνητής Νοημοσύνης ως γενικού όρου που αναφέρεται τακτικά σε οποιαδήποτε τεχνική που χρησιμοποιείται σε οποιοδήποτε πλαίσιο, πραγματικό ή φανταστικό, εφόσον με κάποιο τρόπο υποστηρίζεται ότι εμφανίζει χαρακτηριστικά που ορισμένοι περιγράφουν ως έξυπνα. Επιπλέον, μιλάμε τακτικά για Τεχνητή Νοημοσύνη που είναι ήδη σε ευρεία χρήση παράλληλα με την Τεχνητή Νοημοσύνη που βρίσκεται υπό ανάπτυξη, ακόμη και για Τεχνητή Νοημοσύνη που εικάζεται ότι θα υπάρχει πιθανώς στο μέλλον. Για παράδειγμα τα επιχειρήματα σχετικά με τα Έμπειρα Συστήματα (Expert Systems) που χρησιμοποιούνται για συμβουλευτικούς ρόλους πρέπει να διακρίνονται από αυτά που αφορούν σύνθετους αλγορίθμους που βασίζονται σε δεδομένα και εφαρμόζουν αυτοματοποιημένες αποφάσεις για τα άτομα. Ομοίως, είναι σημαντικό να διακρίνουμε τα επιχειρήματα σχετικά με τις θεωρητικές μελλοντικές εξελίξεις που μπορεί να μην προκύψουν ποτέ από εκείνα που αφορούν την τρέχουσα Τεχνητή Νοημοσύνη που επηρεάζει ήδη την κοινωνία σήμερα.

Σε μια προσπάθεια ανασκόπηση των βασικών τεχνολογιών της Τεχνητής Νοημοσύνης θα προχωρήσουμε στην ομαδοποίησή τους σε τρεις κατηγορίες, επιδιώκοντας να καταδείξουμε την εξελικτική πορεία της Τεχνητής Νοημοσύνης μέσω της ανάπτυξης διαφορετικών προσεγγίσεων:

2.2.1 Συμβολική Τεχνητή Νοημοσύνη (Symbolic AI)

Η Συμβολική Τεχνητή Νοημοσύνη αναφέρεται σε προσεγγίσεις που αναπτύσσουν ευφυείς μηχανές, κωδικοποιώντας τη γνώση και την εμπειρία των ειδικών σε σύνολα κανόνων που μπορούν να εκτελεστούν από μηχανές. Αναφέρεται ως «συμβολική» επειδή χρησιμοποιεί συμβολικό συλλογισμό (π.χ. αν X=Y και Y=Z τότε X=Z) για να αναπαραστήσει και να λύσει προβλήματα. Η Συμβολική Τεχνητή Νοημοσύνη κυριάρχησε σε εφαρμογές Τεχνητής Νοημοσύνης κατά τις δεκαετίες 1950 έως και 1990, αλλά, αν και άλλες προσεγγίσεις κυριαρχούν στο πεδίο σήμερα, εξακολουθεί να χρησιμοποιείται σε πολλά περιβάλλοντα, από θερμοστάτες έως προηγμένη ρομποτική. Ωστόσο, τα συστήματα Συμβολικής Τεχνητής Νοημοσύνης, ενώ μπορούν να εκτελούν εργασίες αυτόματα, το κάνουν μόνο βάσει των οδηγιών που τους δίνονται και μπορούν να βελτιωθούν μόνο με άμεση ανθρώπινη παρέμβαση, γεγονός που περιορίζει το βαθμό αυτονομίας τους. Αυτό καθιστά τη Συμβολική Τεχνητή Νοημοσύνη λιγότερο αποτελεσματική για πολύπλοκα προβλήματα, όπου τόσο οι

μεταβλητές όσο και οι κανόνες αλλάζουν σε πραγματικό χρόνο και για τα οποία χρειαζόμαστε μεγαλύτερη βοήθεια για την επίλυσή τους. Εκατομμύρια κανόνες "if-then-else" δε θα μπορούσαν να συλλάβουν την τεχνογνωσία ενός γιατρού, ούτε τη συνεχή εξέλιξή της με την πάροδο του χρόνου. Παρά τους περιορισμούς, η Συμβολική Τεχνητή Νοημοσύνη λειτουργεί αποτελεσματικά και αποτελεί χρήσιμο εργαλείο για την υποστήριξη ανθρώπων που εργάζονται σε περιβάλλον με επαναλαμβανόμενα προβλήματα σε σαφώς καθορισμένους τομείς, συμπεριλαμβανομένων συστημάτων ελέγχου μηχανών και υποστήριξης λήψης αποφάσεων. Για την αξιόπιστη απόδοσή της σε αυτούς τους τομείς αναφέρεται συχνά και με το προσφιλές παρωνύμιο «Good Old-Fashioned AI» (European Parliamentary Research Service, 2020).

2.2.2.1 Έμπειρα Συστήματα (Expert Systems) και Ασαφής Λογική (Fuzzy Logic)

Τα Έμπειρα Συστήματα και η Ασαφής Λογική αποτελούν τις πλέον δημοφιλείς προσεγγίσεις της Συμβολικής Τεχνητής Νοημοσύνης. Συγκεκριμένα:

- **Έμπειρα Συστήματα**

  Στα Έμπειρα Συστήματα, ένας άνθρωπος ειδικός στον τομέα εφαρμογής δημιουργεί κανόνες που μπορεί να ακολουθήσει ένας υπολογιστής, βήμα προς βήμα, για να αποφασίσει πώς να ανταποκριθεί έξυπνα σε μια δεδομένη κατάσταση. Αυτοί οι κανόνες, γνωστοί ως αλγόριθμοι, εκφράζονται συχνά ως κώδικας σε μορφή "if-then-else". Κάθε μεταβλητή έχει τιμή true ή false, δηλαδή μια ακολουθία 0 και 1 και το σύστημα πρέπει να γνωρίζει μια απόλυτη απάντηση για οποιαδήποτε ερώτηση. Στα εν λόγω συστήματα, οποιαδήποτε νοημοσύνη προέρχεται απευθείας από την ανθρώπινη τεχνογνωσία που καταγράφεται σε μηχαναγνώσιμη μορφή ώστε να μπορέσει να τη διαβάσει ο υπολογιστής και να λειτουργήσει ανάλογα. Υπό αυτήν την έννοια η διαδικασία λήψης αποφάσεων είναι στενά συνδεδεμένη με τον τρόπο με τον οποίο οι ειδικοί λαμβάνουν αποφάσεις. Οι ειδικοί μπορούν με ευκολία να εντοπίσουν λάθη ή να βελτιώσουν το πρόγραμμα, ενημερώνοντας τον κώδικα, ενώ οι άνθρωποι μπορούν με ευκολία να κατανοήσουν πώς τα εξειδικευμένα συστήματα λαμβάνουν συγκεκριμένες αποφάσεις που τους αφορούν. Τα Έμπειρα Συστήματα λειτουργούν καλύτερα σε περιορισμένα περιβάλλοντα που δεν τροποποιούνται ιδιαίτερα με την πάροδο του χρόνου και οι κανόνες είναι αυστηροί, ενώ οι μεταβλητές σαφείς και ποσοτικοποιημένες, όπως π.χ. ο υπολογισμός των φορολογικών δηλώσεων (European Parliamentary Research Service, 2020).

- **Ασαφής Λογική**

  Η Ασαφής Λογική αποτελεί μια άλλη προσέγγιση Εξειδικευμένων Συστημάτων που επιτρέπει στις μεταβλητές να έχουν μια «τιμή αληθείας» που είναι οπουδήποτε μεταξύ 0 και 1, η οποία καταγράφει τον βαθμό στον οποίο ταιριάζει μια κατηγορία. Για παράδειγμα, στην ερώτηση εάν ένας ασθενής έχει πυρετό ή όχι, η απάντηση βάσει της προσέγγισης των Εξειδικευμένων Συστημάτων, θα μπορούσε να περιοριστεί σε έναν απλό υπολογισμό μιας ένδειξης θερμοκρασίας πάνω από 37^{o} C. Ωστόσο, βάσει της Ασαφούς Λογικής, ο ασθενής λαμβάνει μια βαθμολογία για το πόσο καλά ταιριάζει στην κατηγορία του πυρετού. Η βαθμολογία μπορεί να εξαρτάται από την ένδειξη θερμοκρασίας του ασθενούς, αλλά και από άλλους σχετικούς παράγοντες, όπως η ηλικία, η ώρα της ημέρας κ.α. Επομένως, αυτή η ασαφής λογική είναι ιδιαίτερα χρήσιμη για την σύλληψη της διαισθητικής γνώσης, όπου οι ειδικοί παίρνουν σωστές αποφάσεις όταν αντιμετωπίζουν ευρείας κλίμακας αβέβαιες μεταβλητές που αλληλεπιδρούν μεταξύ τους. Η Ασαφής Λογική έχει χρησιμοποιηθεί για την ανάπτυξη συστημάτων ελέγχου για κάμερες που προσαρμόζουν αυτόματα τις ρυθμίσεις τους για να ταιριάζουν στις συνθήκες και για εφαρμογές χρηματιστηριακών συναλλαγών για τη θέσπιση κανόνων για αγορές και πωλήσεις υπό διαφορετικές συνθήκες αγοράς. Συμπερασματικά, το Ασαφές Σύστημα αξιολογεί συνεχώς δεκάδες μεταβλητές, ακολουθεί κανόνες που έχουν σχεδιαστεί από ειδικούς για την προσαρμογή των τιμών αληθείας και τις χρησιμοποιεί για να λαμβάνει αυτόματα αποφάσεις (European Parliamentary Research Service, 2020).

2.2.2 Μηχανική Μάθηση (Machine Learning ML) και Δεδομένα

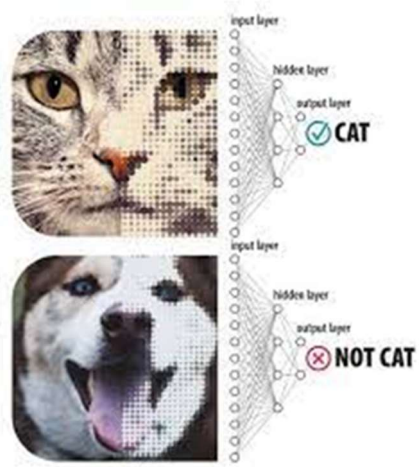
Η Μηχανική Μάθηση αναφέρεται σε ένα ευρύ φάσμα τεχνικών που αυτοματοποιούν τη διαδικασία εκμάθησης των αλγορίθμων. Στην Μηχανική Μάθηση ο αλγόριθμος βελτιώνεται εκπαιδεύοντας τον εαυτό του με δεδομένα, ενώ στη Συμβολική Τεχνητή Νοημοσύνη οι βελτιώσεις στην απόδοση επιτυγχάνονται μόνο από ειδικούς που προσαρμόζουν ή προσθέτουν στην τεχνογνωσία τους που κωδικοποιείται απευθείας στον αλγόριθμο. Ο βασικός, λοιπόν, παράγοντας εξέλιξης της Μηχανικής Μάθησης και η αλματώδης ανάπτυξη πρακτικών εφαρμογών της την τελευταία δεκαετία είναι η μαζική αύξηση της διαθεσιμότητας δεδομένων.

Οι αλγόριθμοι Μηχανικής Μάθησης βρίσκουν τους δικούς τους τρόπους αναγνώρισης προτύπων και εφαρμόζουν αυτά που μαθαίνουν για να κάνουν δηλώσεις σχετικά με

δεδομένα. Διαφορετικές προσεγγίσεις Μηχανικής Μάθησης είναι κατάλληλες για διαφορετικές εργασίες και καταστάσεις και έχουν διαφορετικές επιπτώσεις. Οι βασικές τεχνικές Μηχανικής Μάθησης, όπως τα Τεχνητά Νευρωνικά Δίκτυα και η Βαθιά Μάθηση, επιτρέπουν σε ένα σύστημα Τεχνητής Νοημοσύνης να μάθει πώς να επιλύει προβλήματα που δεν μπορούν να προσδιοριστούν με ακρίβεια ή των οποίων η μέθοδος επίλυσης δεν μπορεί να περιγραφεί με συμβολικούς συλλογιστικούς κανόνες (European Parliamentary Research Service, 2020).

2.2.2.1 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neutral Networks ANN) και Βαθιά Μάθηση (Deep Learning)

Σχήμα 1



Graphic by EPRS, produced by Samy Chahri; picture credits:
@ bedneyimages (freepik.com) and C. Brear (Unsplash).

Τα Τεχνητά Νευρωνικά Δίκτυα είναι εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου όπου τα σήματα μεταδίδονται μέσω ενός πολύπλοκου δικτύου νευρώνων και με αυτόν τον τρόπο, τόσο το σήμα όσο και η δομή του δικτύου μετασχηματίζονται. Στα Τεχνητά Νευρωνικά Δίκτυα, οι είσοδοι μετατρέπονται σε σήματα, τα οποία περνούν μέσω ενός δικτύου τεχνητών νευρώνων για να δημιουργήσουν εξόδους που μπορούν να ερμηνευθούν ως αποκρίσεις στις αρχικές εισόδους. Η διαδικασία μάθησης αναφέρεται στον μετασχηματισμό του δικτύου έτσι ώστε αυτές οι έξοδοι να είναι χρήσιμες – ή έξυπνες – αποκρίσεις στις εισόδους. Τα Τεχνητά Νευρωνικά Δίκτυα επεξεργάζονται δεδομένα που αποστέλλονται στο επίπεδο εισόδου και παράγουν μια απόκριση στο επίπεδο εξόδου. Ενδιάμεσα, υπάρχουν ένα ή περισσότερα «κρυμμένα επίπεδα», τα οποία χειρίζονται τα σήματα καθώς περνούν μέσα από αυτά. Η βασική δομή ενός ANN φαίνεται στο Σχήμα 1, με

ένα ενδεικτικό παράδειγμα ενός Τεχνητού Νευρωνικού Δικτύου που μπορεί να προβλέψει, εάν μια εικόνα απεικονίζει ή όχι μια γάτα. Αρχικά, η εικόνα χωρίζεται σε μεμονωμένα pixels, τα οποία αποστέλλονται στους νευρώνες στο στρώμα εισόδου του Τεχνητού Νευρωνικού Δικτύου. Από εκεί, αποστέλλονται ως σήμα στο πρώτο κρυφό επίπεδο.

Κάθε νευρώνας σε αυτό το κρυφό επίπεδο λαμβάνει πολλά σήματα, τα οποία συνδυάζονται και διαχειρίζονται ανάλογα, ώστε να δημιουργηθεί ένα μόνο σήμα εξόδου. Ενώ το Σχήμα 1 δείχνει μόνο ένα κρυφό επίπεδο, τα Τεχνητά Νευρωνικά Δίκτυα, συνήθως, περιέχουν πολλά διαδοχικά κρυφά επίπεδα. Σε αυτές τις περιπτώσεις, αυτό το βήμα επαναλαμβάνεται με σήματα που περνούν από κάθε κρυφό στρώμα μέχρι να φτάσουν στο τελικό επίπεδο εξόδου. Το σήμα που παράγεται στο επίπεδο εξόδου είναι η τελική έξοδος, η οποία ερμηνεύεται ως μια απόφαση σχετικά με το εάν η εικόνα απεικονίζει ή όχι μια γάτα.

Στο παράδειγμά μας, αναφερόμαστε σε ένα απλό Τεχνητό Νευρωνικό Δίκτυο, εμπνευσμένο από ένα απλοποιημένο μοντέλο του εγκεφάλου, το οποίο μπορεί να ανταποκριθεί σε μια συγκεκριμένη είσοδο με μια συγκεκριμένη έξοδο. Το Τεχνητό Νευρωνικό Δίκτυο δε γνωρίζει πραγματικά τι κάνει, ή ακόμα και τι είναι μια γάτα, αλλά αν του δώσουμε μια φωτογραφία, θα μας λέει πάντα αν «νομίζει» ότι περιέχει γάτα ή όχι (European Parliament Research Service, 2020).

Ωστόσο, ποια θα μπορούσαν να είναι τα κριτήρια ώστε ένα Τεχνητό Νευρωνικό Δίκτυο να λαμβάνει την σωστή απόφαση;

1. **Ένα Τεχνητό Νευρωνικό Δίκτυο θα πρέπει να έχει σωστή δομή.**

   Τα Τεχνητά Νευρωνικά Δίκτυα, για απλές εργασίες, είναι σε θέση να λειτουργήσουν καλά με μόλις δώδεκα (12) νευρώνες σε ένα και μόνο κρυφό επίπεδο. Η προσθήκη περισσότερων νευρώνων και επιπέδων επιτρέπει στα Τεχνητά Νευρωνικά Δίκτυα να αντιμετωπίζουν πιο περίπλοκα προβλήματα. Η Βαθιά Μάθηση αναφέρεται απλώς σε Τεχνητά Νευρωνικά Δίκτυα με τουλάχιστον δύο κρυφά επίπεδα, το καθένα από τα οποία περιέχει πολλούς νευρώνες. Η ύπαρξη περισσότερων επιπέδων επιτρέπει στα Τεχνητά Νευρωνικά Δίκτυα να αναπτύξουν πιο αφηρημένες εννοιολογήσεις προβλημάτων, χωρίζοντάς τα σε μικρότερα υποπροβλήματα και να παρέχουν, επίσης, πιο διαφοροποιημένες αποκρίσεις. Ενώ θεωρητικά τρία κρυφά επίπεδα μπορεί να είναι αρκετά για την επίλυση κάθε είδους προβλήματος, στην πράξη τα Τεχνητά Νευρωνικά Δίκτυα τείνουν να περιέχουν πολλά περισσότερα. Για παράδειγμα, οι ταξινομητές εικόνων της Google χρησιμοποιούν έως και τριάντα (30) κρυφά επίπεδα. Τα πρώτα

επίπεδα αναζητούν γραμμές που μπορούν να αναγνωρίσουν ως άκρα ή γωνίες, τα μεσαία επίπεδα προσπαθούν να αναγνωρίσουν σχήματα σε αυτές τις γραμμές και τα τελικά επίπεδα συναρμολογούν αυτά τα σχήματα για να ερμηνεύσουν την εικόνα (European Parliamentary Research Service, 2020).

2. **Ένα Τεχνητό Νευρωνικό Δίκτυο θα πρέπει να εκπαιδευτεί**

Και ενώ το «βαθύ» μέρος της Βαθιάς Μάθησης αφορά στην πολυπλοκότητα του Τεχνητού Νευρωνικού Δικτύου, η μάθηση αφορά στην εκπαίδευσή του. Έτσι λοιπόν, μόλις δημιουργηθεί η σωστή δομή του Τεχνητού Νευρωνικού Δικτύου, αυτό θα πρέπει να εκπαιδευτεί. Αν και η εκπαίδευση θεωρητικά θα μπορούσε να γίνει με το χέρι από έναν ειδικό που θα προσάρμοζε επιμελώς τους νευρώνες ώστε να αντικατοπτρίζουν τη δική τους τεχνογνωσία για το πώς να αναγνωρίζουν τις γάτες, αντίθετα, εφαρμόζεται ένας αλγόριθμος Μηχανικής Μάθησης για την αυτοματοποίηση της διαδικασίας.

Η διαδικασία μάθησης περιλαμβάνει δύο σημαντικές τεχνικές Μηχανικής Μάθησης. Η πρώτη που περιλαμβάνει την Αντίστροφή Διάδοση (Back Propagation) και την Κλιμακωτή Κάθοδο (Gradient Descent), εφαρμόζει μαθηματικές έννοιες, όπως ο λογισμός, για να κάνει σταδιακές βελτιώσεις σε μεμονωμένα Τεχνητά Νευρωνικά Δίκτυα. Η δεύτερη εφαρμόζει εξελικτικές αρχές εμπνευσμένες από την φύση για να επιφέρει σταδιακές βελτιώσεις σε μεγάλους πληθυσμούς Τεχνητών Νευρωνικών Δικτύων (European Parliamentary Research Service, 2020).

Αντίστροφη Διάδοση (Back Propagation) και Κλιμακωτή Κάθοδος (Gradient Descent)

Εάν συγκρίνουμε την πραγματική έξοδο ενός Τεχνητού Νευρωνικού Δικτύου με την επιθυμητή έξοδο, όπως προσδιορίζεται από τα επισημασμένα δεδομένα, η διαφορά μεταξύ των δύο περιγράφεται, ως σφάλμα. Οι αλγόριθμοι Μηχανικής Μάθησης, όπως η Αντίστροφη Διάδοση και η Κλιμακωτή Κάθοδος στοχεύουν να βελτιώσουν σταδιακά την απόδοση του Τεχνητού Νευρωνικού Δικτύου, ελαχιστοποιώντας αυτό το σφάλμα. Το επιτυγχάνουν προσαρμόζοντας το Τεχνητό Νευρωνικό Δίκτυο και ελέγχοντας εάν το σφάλμα έχει μειωθεί πριν από την εκ νέου ρύθμιση.

Αναλυτικότερα, η Αντίστροφη Διάδοση ασχολείται με την προσαρμογή των νευρώνων στο Τεχνητό Νευρωνικό Δίκτυο. Η διαδικασία ξεκινά όταν ένα σήμα εισόδου

αποστέλλεται στο Τεχνητό Νευρωνικό Δίκτυο περνά μέσα από τα κρυφά στρώματα στο επίπεδο εξόδου και παράγει ένα σήμα εξόδου. Στη συνέχεια, το σφάλμα υπολογίζεται συγκρίνοντας την έξοδο με αυτό που θα έπρεπε να είναι σύμφωνα με τα επισημασμένα δεδομένα. Τώρα, οι νευρώνες αλλάζουν για να μειωθεί το σφάλμα, επομένως, η έξοδος των Τεχνητών Νευρωνικών Συστημάτων να είναι πιο ακριβής. Αυτή η διαδικασία διόρθωσης ξεκινά από το επίπεδο εξόδου, το οποίο έχει ισχυρότερη επιρροή στα αποτελέσματα και στη συνέχεια κάνει αλλαγές προς τα πίσω μέσω των κρυφών επιπέδων. Ονομάζεται Αντίστροφη Διάδοση, επειδή η διόρθωση του σφάλματος διαδίδεται προς τα πίσω μέσω του Τεχνητού Νευρωνικού Συστήματος (European Parliament Research Service, 2020).

Στο πλαίσιο της Κλιμακωτής Καθόδου, αναζητάται το μικρότερο σφάλμα, ώστε να επιτευχθεί μεγαλύτερη ακρίβεια, δηλαδή ένα μέτρο του πόσο μεγάλο είναι το ποσοστό των σωστών απαντήσεων. Σύμφωνα με την εν λόγω τεχνική, δημιουργείται ένα Τεχνητό Νευρωνικό Σύστημα σε ένα τυχαίο σημείο στη περιοχή του σφάλματος. Υπολογίζεται το σφάλμα του, καθώς και το σφάλμα που προκύπτει από μερικά διαφορετικά είδη προσαρμογής που αντιστοιχούν σε κοντινές θέσεις στη περιοχή σφάλματος. Η προσαρμογή που προσφέρει την καλύτερη βελτίωση θεωρείται ότι είναι η πιο ακριβής, επομένως, οι αλλαγές υλοποιούνται και στη συνέχεια η διαδικασία επαναλαμβάνεται με ένα νέο σύνολο δοκιμών, δηλαδή το Τεχνητό Νευρωνικό Δίκτυο κάνει σταδιακές βελτιώσεις μέχρι να φτάσει στην καλύτερη λύση που μπορεί να εντοπίσει (European Parliament Research Service, 2020).

Σε πολλές περιπτώσεις ο αλγόριθμος θα μπορούσε να ικανοποιεί ένα «τοπικό βέλτιστο» και όχι την καλύτερη διαθέσιμη λύση, με αποτέλεσμα οι μικρές τροποποιήσεις που θα πραγματοποιούσε να χειροτέρευαν την κατάσταση προτού βελτιωθεί. Για το λόγο αυτό, η όλη διαδικασία επαναλαμβάνεται πολλές φορές, ξεκινώντας από διαφορετικά σημεία της περιοχής σφάλματος και χρησιμοποιώντας διαφορετικά προπονητικά δεδομένα.

Η Κλιμακωτή Κάθοδος και η Αντίστροφη Διάδοση χρησιμοποιούν επισημασμένα δεδομένα για τον υπολογισμό του σφάλματος. Προκειμένου να διασφαλιστεί ότι ο αλγόριθμος δεν θα απομνημονεύσει τα δεδομένα εκπαίδευσης χωρίς να έχει αποκτήσει κάποια χρήσιμη ικανότητα ώστε να ανταποκρίνεται σε νέα δεδομένα, ορισμένα από τα επισημασμένα δεδομένα δε χρησιμοποιούνται για την εκπαίδευση, αλλά

χρησιμοποιούνται μόνο για τη δοκιμή των αποτελεσμάτων (European Parliament Research Service, 2020).

Μέθοδοι εκπαίδευσης εμπνευσμένες από τη φύση

Οι εν λόγω μέθοδοι εκπαίδευσης Τεχνητών Νευρωνικών Δικτύων εμπνέονται από εξελικτικές έννοιες, όπως η επιβίωση του ισχυρότερου, η αναπαραγωγή και η μετάλλαξη. Ειδικότερα, δημιουργείται ένας πληθυσμός Τεχνητών Νευρωνικών Δικτύων που συναγωνίζονται μεταξύ τους και υπόκεινται σε τεχνητή επιλογή έτσι ώστε να απορριφθούν όσα έχουν κακή απόδοση και να επιβιώσουν στην επόμενη γενιά όσα έχουν καλή απόδοση. Για την αναπλήρωση του πληθυσμού, δημιουργούνται νέα Τεχνητά Νευρωνικά Δίκτυα μέσω της απάντησης του συστήματος Τεχνητής Νοημοσύνης στην αναπαραγωγή. Αυτά θα μπορούσαν να περιλαμβάνουν τον συνδυασμό διαφορετικών πτυχών ενός, δύο ή οποιουδήποτε αριθμού μητρικών Τεχνητών Νευρωνικών Δικτύων, μαζί με μια δόση τυχαίας μετάλλαξης.

Ας διερευνήσουμε πώς μπορεί να εφαρμοστεί μια εξελικτική προσέγγιση για την εκπαίδευση των Τεχνητών Νευρωνικών Δικτύων να παίζουν σκάκι. Θα μπορούσαμε να ξεκινήσουμε δημιουργώντας έναν πληθυσμό 100 τυχαίων «παικτών» Τεχνητών Νευρωνικών Δικτύων και να τους κάνουμε να παίζουν μεταξύ τους, ώστε να λαμβάνουν εναλλάξ εισόδους που περιγράφουν τη θέση των κομματιών και να δημιουργούν εξόδους που ερμηνεύονται ως κινήσεις. Αυτή η πρώτη γενιά μη εκπαιδευμένων παικτών δεν θα είναι πολύ καλοί στο παιχνίδι, αλλά μερικοί αναπόφευκτα θα είναι «λιγότερο κακοί» από άλλους και θα κερδίσουν μερικά παιχνίδια. Σε αυτό το σημείο, τα Τεχνητά Νευρωνικά Δίκτυα μπορούν να ταξινομηθούν. Σύμφωνα με την αρχή της επιβίωσης του ισχυρότερου, οι χειρότεροι παίκτες διαγράφονται. Καλύτεροι παίκτες επιβιώνουν και το «γενετικό υλικό» τους – με τη μορφή στρωμάτων Τεχνητών Νευρωνικών Δικτύων – συνδυάζεται και μεταλλάσσεται για να δημιουργηθεί μια νέα γενιά Τεχνητών Νευρωνικών Δικτύων που θα ενσωματωθούν στον επόμενο γύρο παιχνιδιών. Τα νέα Τεχνητά Νευρωνικά Δίκτυα θα ανταποκρίνονται διαφορετικά στα σήματα, επομένως κάποιοι μπορεί να παίζουν καλύτερα από τους «γονείς» τους και άλλοι χειρότερα (European Parliament Research Service, 2020).

Δεδομένου ότι μόνο οι καλύτεροι παίκτες επιβιώνουν και διαμορφώνουν τα χαρακτηριστικά των μελλοντικών Τεχνητών Νευρωνικών Δικτύων, το περιβάλλον ευνοεί μια σταθερή συνολική βελτίωση καθώς περνούν οι γενιές. Η προσέγγιση μπορεί

ακόμη και να αποφέρει πρωταθλητές παίκτες που νικούν τους περισσότερους ανθρώπους. Το ενδιαφέρον με τις εξελικτικές μεθόδους είναι ότι δίνουν αποτελέσματα χωρίς ανθρώπινη εξειδίκευση στο πρόβλημα, χωρίς επισημασμένα δεδομένα από προηγούμενα παιχνίδια, χωρίς καν να έχουν πρόσβαση στους κανόνες. Ωστόσο, όσοι τείνουν να παίζουν καλά, τείνουν να επιβιώνουν. Αυτό σημαίνει ότι τα Τεχνητά Νευρωνικά Δίκτυα μπορούν να αναπτύξουν ενδιαφέροντες τρόπους για να αποφασίσουν πώς να παίξουν καλά το παιχνίδι, συμπεριλαμβανομένων στρατηγικών που οι άνθρωποι δεν έχουν σκεφτεί ποτέ και μπορεί να δυσκολεύονται να εκτιμήσουν. Εάν ζητηθεί από έναν μηχανικό να εξηγήσει γιατί ένα τέτοιο Τεχνητό Νευρωνικό Δίκτυο έκανε μια κίνηση, τότε μπορεί να δείξει πώς η απόκρισή του καθορίστηκε μαθηματικά από τη δομή του, αλλά δεν μπορεί πάντα να εξηγήσει γιατί αυτή η δομή δημιουργεί καλές κινήσεις, γεγονός που οδηγεί στο μείζον πρόβλημα της αδιαφάνειας των αλγορίθμων (European Parliament Research Service, 2020).

Οι εξελικτικές προσεγγίσεις μπορούν να εφαρμοστούν σε προβλήματα βελτιστοποίησης, όπως η βελτίωση προγραμμάτων υπολογιστών ή χρονοδιαγραμμάτων μεταφοράς. Υπάρχουν, επίσης, πολλές άλλες ενδιαφέρουσες τεχνικές Τεχνητής Νοημοσύνης εμπνευσμένες από βιολογικούς και συμπεριφορικούς μηχανισμούς που παρατηρούνται στη φύση. Για παράδειγμα, η βελτιστοποίηση αποικίας μυρμηγκιών βασίζεται στο πώς τα μυρμήγκια χρησιμοποιούν τις φερομόνες ως σήματα για να βρουν και να τονίσουν την ταχύτερη διαδρομή μεταξύ δύο τοποθεσιών και μπορεί να χρησιμοποιηθεί για τη βελτιστοποίηση της πλοήγησης και των τηλεπικοινωνιακών δικτύων οχημάτων. Η αναζήτηση κυνηγιού είναι μια τεχνική αναζήτησης και βελτιστοποίησης που βασίζεται στο κυνήγι της αγέλης από λιοντάρια, λύκους και δελφίνια.

Η Ενισχυτική Μάθηση (Reinforcement Learning-RL) είναι ένας άλλος κλάδος της Μηχανικής Μάθησης που εστιάζει στην ανάπτυξη μιας πολιτικής για τη λήψη ακολουθιών αποφάσεων υπό διαφορετικές συνθήκες. Είναι ιδιαίτερα χρήσιμο όταν ένα σύστημα μπορεί να υπόκειται σε ένα ευρύ φάσμα συνθηκών, καθεμία με διαφορετικές επιπτώσεις για την κατάλληλη δράση. Πρώτα ο αλγόριθμος RL προσδιορίζει ορισμένα χαρακτηριστικά των συνθηκών και επιχειρεί ορισμένες ενέργειες, στη συνέχεια λαμβάνει ανατροφοδότηση σχετικά με την ποιότητα της απόκρισης, η οποία χρησιμοποιείται για τη διατήρηση ενός συνόλου βαθμολογιών για διαφορετικούς συνδυασμούς συνθηκών και ενεργειών. Το RL συγκρίνεται μερικές φορές με το πώς

μαθαίνουν τα παιδιά να περπατούν και οι τεχνικές δοκιμής και λάθους του έχουν αναπτυχθεί για την εκπαίδευση αυτοοδηγούμενων αυτοκινήτων. Από πολλές απόψεις, η διαδικασία είναι παρόμοια με την Αντίστροφη Διάδοση και την Κλιμακωτή Κάθοδο, αλλά, ενώ αυτές οι μέθοδοι απαιτούν μεγάλες ποσότητες επισημασμένων δεδομένων που έχουν προετοιμαστεί εκ των προτέρων, η RL επιτρέπει συνεχή προσαρμογή της συμπεριφοράς για μάθηση σε πραγματικό χρόνο (European Parliamentary Research Service, 2020).

3. **Ένα Τεχνητό Νευρωνικό Δίκτυο θα πρέπει να τροφοδοτείται από καλής ποιότητας δεδομένα**

Τα Τεχνητά Νευρωνικά Δίκτυα χρειάζονται πολλά δεδομένα καλής ποιότητας για να εκπαιδευτούν και να δοκιμάσουν τα αποτελέσματα που παράγουν. Έχουν αναπτυχθεί πολλές τεχνικές Μηχανικής Μάθησης σχετικές με τα δεδομένα, δεδομένου του κεντρικού ρόλου τους στην σύγχρονη ανάπτυξη της Τεχνητής Νοημοσύνης.

Η **Εποπτευόμενη Μάθηση (Supervised Learning)** αναφέρεται στη χρήση επισημασμένων δεδομένων – όπως εικόνες που λένε αν περιέχουν ή όχι γάτες – για την εκπαίδευση αλγορίθμων. Αυτές οι προσεγγίσεις επινοούν τις δικές τους μεθόδους για την πρόβλεψη του τρόπου με τον οποίο θα πρέπει να επισημαίνονται οι εικόνες. Η Μάθηση χωρίς Επίβλεψη (Unsupervised Learning) μπορεί να χρησιμοποιηθεί όταν δεν υπάρχουν διαθέσιμα επισημασμένα δεδομένα καλής ποιότητας. Διαπρέπουν στην εύρεση νέων συστάδων και συσχετισμών σε δεδομένα που διαφορετικά δεν θα μπορούσαν να είχαν εντοπιστεί ή επισημανθεί από τον άνθρωπο. Δεδομένου ότι οι ετικέτες είναι συχνά ελλιπείς ή ανακριβείς, πολλές εφαρμογές, όπως συστήματα συστάσεων περιεχομένου, συνδυάζουν προσεγγίσεις Μηχανικής Μάθησης Εποπτευόμενης και μη (European Parliament Research Service, 2020).

Η **Εξόρυξη Δεδομένων (Data Mining)**, είναι ένα πεδίο υπολογισμού που επικεντρώνεται στην αυτοματοποιημένη αναγνώριση προτύπων και ανωμαλιών σε σύνολα δεδομένων. Το σύνολο δεδομένων θα μπορούσε να είναι οτιδήποτε, από μετρήσεις υπόγειων γεωλογικών σχηματισμών μέχρι κείμενο που βρίσκεται στα μέσα κοινωνικής δικτύωσης και η διαδικασία εξόρυξης θα μπορούσε να αναπτύξει Τεχνητά Νευρωνικά Δίκτυα, στατιστικά και μοντελοποίηση για τον εντοπισμό χρήσιμων χαρακτηριστικών. Τα Μεγάλα Δεδομένα (Big Data) αναφέρονται σε σύνολα δεδομένων που είναι τόσο μεγάλα και πολύπλοκα – συμπεριλαμβανομένου περιεχομένου από

διαφορετικές πηγές, σε διαφορετικές μορφές και με διαφορετικούς βαθμούς γνησιότητας και ακρίβειας – που δεν μπορούν να αποθηκευτούν ή να υποβληθούν σε επεξεργασία με τον ίδιο τρόπο όπως μικρότερα σύνολα δεδομένων. Αυτό μας οδηγεί στα «δεδομένα στη φύση», συνήθως αναφερόμενα σε δεδομένα που παράγονται για έναν σκοπό, αλλά παραμένουν κατά κάποιο τρόπο προσβάσιμα και συλλέγονται ή χρησιμοποιούνται για κάποιον άλλο σκοπό. Ανάλογα με τις περιστάσεις, η χρήση αυτών των δεδομένων μπορεί να είναι αναξιόπιστη, ανήθικη, ακόμη και παράνομη (European Parliamentary Research Service, 2020).

4. **Ένα Τεχνητό Νευρωνικό Δίκτυο θα πρέπει να έχει σχεδιαστεί σωστά και να λαμβάνει ακριβείς οδηγίες από τον δημιουργό του για την βελτιστοποίηση του περιβάλλοντος στο οποίο μαθαίνει**

Η αποτελεσματικότητα ενός Τεχνητού Νευρωνικού Συστήματος εξαρτάται από την ικανότητα του δημιουργού του, του μηχανικού Τεχνητής Νοημοσύνης, ο οποίος αξιοποιεί τη δύναμη των εννοιών από μια σειρά επιστημονικών κλάδων, κυρίως υπολογιστών, λογικής, στατιστικής και λογισμού, εξισορροπώντας μια σειρά από σκέψεις σχετικά με το ίδιο το πρόβλημα και το πλαίσιο της επίλυσής του.

Αρχικά, ο μηχανικός Τεχνητής Νοημοσύνης πρέπει να αναζητήσει έναν καλό τρόπο κωδικοποίησης του ίδιου του προβλήματος. Εάν ο αλγόριθμος Μηχανικής Μάθησης χρησιμοποιεί δεδομένα εκπαίδευσης, ο μηχανικός Τεχνητής Νοημοσύνης πρέπει να εξετάσει ποια δεδομένα θα χρησιμοποιήσει και πώς. Όταν χρησιμοποιούνται «δεδομένα στη φύση», πρέπει να διασφαλίζει ότι είναι νόμιμα και ηθικά. Ακόμη και η ακούσια αποθήκευση και επεξεργασία ορισμένου περιεχομένου – όπως η τρομοκρατική προπαγάνδα και η παιδική πορνογραφία – μπορεί να είναι παράνομη. Άλλα δεδομένα ενδέχεται να υπόκεινται σε πνευματικά δικαιώματα ή να απαιτούν «ενημερωμένη συγκατάθεση» από τον ιδιοκτήτη προτού χρησιμοποιηθούν για έρευνα ή άλλους σκοπούς. Εάν τα δεδομένα περάσουν αυτές τις δοκιμές, ο μηχανικός πρέπει να προσδιορίσει εάν είναι αρκετά μεγάλα και αντιπροσωπευτικά, ώστε να είναι κατάλληλα για το υπό εξέταση πρόβλημα. Ο μηχανικός Τεχνητής Νοημοσύνης πρέπει να αποφασίσει πόσα δεδομένα θα χρησιμοποιήσει για εκπαίδευση και πόσα θα αφήσει για δοκιμές. Εάν το σύνολο δεδομένων εκπαίδευσης είναι πολύ μικρό, το Τεχνητό Νευρωνικό Δίκτυο μπορεί να το απομνημονεύσει, χωρίς να μάθει γενικούς κανόνες και

να έχει κακή απόδοση όταν δοκιμάζεται με νέα δεδομένα. Εάν το σύνολο δεδομένων δοκιμής είναι μικρό, υπάρχει μικρότερο περιθώριο αξιολόγησης της ποιότητας του αλγορίθμου (European Parliament Research Service, 2020).

Ο μηχανικός Τεχνητής Νοημοσύνης πρέπει επίσης να λάβει αρκετές σημαντικές αποφάσεις σχετικά με τη δομή του Τεχνητού Νευρωνικού Δικτύου και του αλγόριθμου Μηχανικής Μάθησης. Το Τεχνητό Νευρωνικό Δίκτυο χρειάζεται αρκετούς νευρώνες και στρώματα για να αντιμετωπίσει την πολυπλοκότητα του προβλήματος. Για την Κλιμακωτή Κάθοδο, πρέπει να καθορίσει πόσες αξιολογήσεις θα κάνει πριν αποφασίσει την κατεύθυνση που θα ακολουθηθεί, καθώς και πόσο μακριά θα οδηγηθεί προς την επιλεγμένη κατεύθυνση πριν την επαναξιολόγηση. Αυτό είναι γνωστό ως «ποσοστό μάθησης». Εάν είναι πιο αργό, ο αλγόριθμος παίρνει περισσότερο χρόνο, αλλά κάνει καλύτερες επιλογές. Ωστόσο, αν είναι πιο γρήγορο, προσαρμόζεται πιο γρήγορα, αλλά μπορεί να χάσει σημαντικά χαρακτηριστικά. Ο μηχανικός Τεχνητής Νοημοσύνης πρέπει να εξετάσει το πρόβλημα και να αποφασίσει πώς να εξισορροπήσει την ταχύτητα με την ακρίβεια (European Parliament Research Service, 2020).

Στις εξελικτικές προσεγγίσεις, ο μηχανικός Τεχνητής Νοημοσύνης πρέπει να αποφασίσει το μέγεθος του πληθυσμού και τον αριθμό των συσχετίσεων, εξισορροπώντας τη διεξοδική αξιολόγηση έναντι του φόρτου επεξεργασίας. Πρέπει, επίσης, να αποφασίσει πόσα Τεχνητά Νευρωνικά Δίκτυα θα διαγράψει ανά γενιά και πώς θα χρησιμοποιηθεί ο συνδυασμός και η μετάλλαξη για τη δημιουργία νέων γενεών. Η μετάλλαξη είναι σημαντική για την εμφάνιση νέων λύσεων, αλλά αν είναι πολύ ισχυρή τότε οι «απόγονοι» μπορεί να είναι τόσο διαφορετικοί από τους γονείς τους, ώστε να αποδίδουν τόσο άσχημα όσο και τυχαία δημιουργήθηκαν τα Τεχνητά Νευρωνικά Δίκτυα από την πρώτη γενιά της διαδικασίας. Ένα περαιτέρω ερώτημα τίθεται για να αποφασίσει πότε έχει βρεθεί μια λύση επαρκούς ποιότητας. Όπως συζητήθηκε στο πλαίσιο της Κλιμακωτής Καθόδου, ένας αλγόριθμος μπορεί να κολλήσει σε ένα «τοπικό βέλτιστο», το οποίο είναι σίγουρα η καλύτερη λύση στην περιοχή, αλλά όχι απαραίτητα η καλύτερη δυνατή λύση. Ομοίως, οι εξελικτικοί πληθυσμοί μπορούν να εξελιχθούν σε ένα τοπικό βέλτιστο, σύμφωνα με το οποίο τα μητρικά Τεχνητά Νευρωνικά Δίκτυα δεν μπορούν να παράγουν απογόνους που αποδίδουν καλύτερα από αυτούς, παρόλο που υπάρχουν καλύτερες λύσεις. Ο μηχανικός Τεχνητής Νοημοσύνης μπορεί να εξουδετερώσει τα τοπικά βέλτιστα προσαρμόζοντας τον ρυθμό μάθησης στην Κλιμακωτή Κάθοδο ή αλλάζοντας την προσέγγιση για την

αναπαραγωγή και τη μετάλλαξη στις εξελικτικές μεθόδους. Μπορεί, επίσης, να επαναλάβει τη διαδικασία εκπαίδευσης πολλές φορές από διαφορετικά σημεία εκκίνησης και σύνολα δεδομένων. Αυτό συχνά αξίζει τον κόπο γιατί, ενώ τα Τεχνητά Νευρωνικά Δίκτυα είναι δύσκολο να εκπαιδευτούν, μόλις βρεθεί μια καλή λύση μπορούν να εφαρμοστούν σε νέα δεδομένα πολύ γρήγορα.

Πολλές από αυτές τις αποφάσεις απαιτούν από τον μηχανικό Τεχνητής Νοημοσύνης να εξισορροπήσει τους περιορισμούς του προβλήματος, το πλαίσιο και τα δεδομένα και τους πόρους επεξεργασίας που έχει στη διάθεσή του. Ωστόσο, δεν υπάρχουν αντικειμενικά σωστές φόρμουλες, επομένως, αυτές οι αποφάσεις αποτελούν μέρος της εμπειρίας, της διαίσθησης, του πειραματισμού και της κοινής σοφίας για να λάβει αποτελεσματικές αποφάσεις. Ίσως δεν αποτελεί έκπληξη, ότι αποφάσεις σχετικά με τη δομή του Τεχνητού Νευρωνικού Δικτύου, το ποσοστό μετάλλαξης, το ποσοστό μάθησης και το μέγεθος του πληθυσμού ανατίθενται μερικές φορές στη Μηχανική Μάθηση. Ωστόσο, ο μηχανικός της Τεχνητής Νοημοσύνης χρειάζεται ακόμα να λάβει δύσκολες αποφάσεις και να δώσει ακριβείς οδηγίες σχετικά με το πότε και πόσο η Μηχανική Μάθηση μπορεί να προσαρμόσει κάθε μεταβλητή με αποτέλεσμα να διατηρεί έναν αναντικατάστατο ρόλο στο σχεδιασμό και τη βελτιστοποίηση του περιβάλλοντος στο οποίο μαθαίνουν οι μηχανές (European Parliamentary Research Service, 2020).

2.2.3 Μελλοντικές θεωρητικές προσεγγίσεις εξέλιξης Τεχνητής Νοημοσύνης

Οι προσεγγίσεις που αναλύσαμε ανωτέρω και αφορούν στη Συμβολική Τεχνητή Νοημοσύνη και τη Μηχανική Μάθηση αναφέρονται ως «αδύναμη» ή «στενή» Τεχνητή Νοημοσύνη. Η «ισχυρή» ή «γενική» Τεχνητή Νοημοσύνη (AGI), από την άλλη πλευρά, είναι πιο κοντά στην κατανόησή μας για την ανθρώπινη νοημοσύνη, καθώς αναφέρεται σε αλγόριθμους που μπορούν να επιδείξουν νοημοσύνη σε ένα ευρύ φάσμα πλαισίων και προβληματικών χώρων και όχι μόνο σε συγκεκριμένες θέσεις, όπως η αδύναμη ή στενή Τεχνητή Νοημοσύνη. Επίσης, θα μπορούσε να συνδυάσει πτυχές της αδύναμης Τεχνητής Νοημοσύνης με τρόπους που προσφέρουν εντελώς νέα λειτουργικότητα. Ένας άλλος βασικός όρος από το θεωρητικό τομέα είναι η Τεχνητή Υπερευφυΐα (ASI) που αναφέρεται σε υψηλότερα επίπεδα γενικής νοημοσύνης από αυτή των τυπικών ανθρώπων. Ωστόσο, καθώς δεν υπάρχουν ακόμη, ανήκουν στη σφαίρα της θεωρητικής Τεχνητής Νοημοσύνης. Παρόλα αυτά, υπάρχει κάποια συζήτηση σχετικά με το εάν αυτές οι θεωρητικές προσεγγίσεις Τεχνητής Νοημοσύνης θα μπορούσαν να επιτευχθούν με σταδιακή ανάπτυξη των υπαρχουσών τεχνολογιών και τεχνικών. Στη συζήτηση αυτή, ορισμένοι ειδικοί υποστηρίζουν ότι η ανάπτυξη

παραδειγμάτων θα μπορούσε τελικά να καταστήσει δυνατή - και ίσως ακόμη και αναπόφευκτη - τη Γενική Τεχνητή Νοημοσύνη, ενώ στην πλειοψηφία τους θεωρούν ότι οι εν λόγω προσεγγίσεις αντιμετωπίζουν σημαντικά τεχνικά εμπόδια στην εφαρμογή τους και ως εκ τούτου θα παραμείνουν, ως θεωρητικές μελλοντικές δυνατότητες χωρίς καμία εγγύηση υλοποίησης. Παρακάτω παραθέτουμε κάποιες προσεγγίσεις της θεωρητικής Τεχνητής Νοημοσύνης, μερικές εκ των οποίων υπόκεινται σε ενεργό έρευνα κι ανάπτυξη και προτείνονται ως διαδρομές προς την Τεχνητή Υπερευφυΐα, ενώ άλλες έχουν πιο μετριοπαθείς και πρακτικούς στόχους (European Parliamentary Research Service, 2020).

2.2.3.1 Ρομποτική Τεχνητή Νοημοσύνη

Η ρομποτική δεν είναι πραγματικά μια διαδρομή προς την Τεχνητή Νοημοσύνη από μόνη της, αλλά είναι ένας συμπληρωματικός τομέας με πιθανές συνέργειες που θα μπορούσαν να οδηγήσουν σε ουσιαστική πρόοδο σε ορισμένους τομείς εφαρμογής. Για παράδειγμα, η Μηχανική Μάθηση μπορεί να εφαρμοστεί για να βοηθήσει τα ρομπότ να χειρίζονται φυσικά αντικείμενα με μεγαλύτερη αυτονομία, ευελιξία και επιδεξιότητα, γεγονός που θα μπορούσε να κάνει την αυτοματοποιημένη παραγωγή και διανομή πιο αποτελεσματική. Η Τεχνητή Νοημοσύνη και η Ρομποτική θα μπορούσαν να συνδυαστούν για το σχεδιασμό, την κατασκευή και τον έλεγχο νέου υλικού υπολογιστών και ρομποτικής που ενισχύουν και τους δύο τομείς, ξεκινώντας έναν κύκλο συνεχούς συνεργατικής ανάπτυξης και προόδου. Εδώ, αξίζει επίσης να αναφέρουμε ότι ο συνδυασμός της Τεχνητής Νοημοσύνης και της Ρομποτικής είναι ένας σημαντικός τομέας ανάπτυξης για τις στρατιωτικές τεχνολογίες, ιδίως στα αυτόνομα οπλικά συστήματα. Προς το παρόν, τα μη επανδρωμένα αεροσκάφη οδηγούνται εξ αποστάσεως από ανθρώπους, αλλά αυτό επιφέρει αρκετές αδυναμίες, συμπεριλαμβανομένων των καναλιών επικοινωνίας που είναι ευάλωτα σε ανίχνευση και επίθεση, καθώς και πολύ πιο αργούς χρόνους ανθρώπινης απόφασης και απόκρισης σε σχέση με τα αυτοματοποιημένα συστήματα ελέγχου. Η πλήρης εντολή Τεχνητής Νοημοσύνης επιλύει και τα δύο ζητήματα, ενώ ανοίγει νέες ευκαιρίες, όπως οι δυνατότητες «σμήνους», δηλαδή χρήση πολλαπλών π.χ. drones ταυτόχρονα, που είναι πέρα από τις ανθρώπινες δυνατότητες. Τέτοια συστήματα δεν ξεπερνούν τις σημερινές τεχνικές δυνατότητες, αλλά το πεδίο είναι αμφιλεγόμενο αφού ενσκήπτουν πολιτικά, κοινωνικά και δεοντολογικά ζητήματα (European Parliamentary Research Service, 2020).

2.2.3.2 Κβαντική Τεχνητή Νοημοσύνη

Οι κβαντικοί υπολογιστές αξιοποιούν τη δύναμη του ταυτόχρονου για να βρίσκουν γρήγορα λύσεις σε πολύ περίπλοκα προβλήματα, υποσχόμενοι μια επαναστατική αύξηση της

υπολογιστικής ισχύος. Εάν το πρόβλημα είναι να βρεθεί ένας συνδυασμός ενός στο τρισεκατομμύριο που να λειτουργεί ως λύση, ένας κανονικός υπολογιστής θα πρέπει να ελέγξει κάθε πιθανότητα μία προς μία, ενώ ένας κβαντικός υπολογιστής μπορεί να τις ελέγξει όλες ταυτόχρονα, με μία μόνο λειτουργία. Αυτό σημαίνει ότι είναι ιδιαίτερα κατάλληλοι για προβλήματα, όπως η προσομοίωση περιβαλλόντων, η εύρεση λύσεων και η βελτιστοποίησή τους. Δεδομένου ότι αυτού του είδους τα προβλήματα είναι κεντρικά για την Τεχνητή Νοημοσύνη, οι εξελίξεις στον κβαντικό υπολογισμό θα μπορούσαν να επιτρέψουν σημαντικές προόδους στον τομέα. Ενώ υπήρξαν μερικές ελπιδοφόρες πρόσφατες ανακαλύψεις στον κβαντικό υπολογισμό,[5] οι λεπτομέρειές τους συχνά χρησιμεύουν για να καταδείξουν πόσο απέχει η τεχνολογία από την πραγματικότητα στην αγορά. Για παράδειγμα, στα τέλη του 2017, η μηχανή 50 qubit της IBM έσπασε τα ρεκόρ του κλάδου, παραμένοντας σταθερή για 0,00009 δευτερόλεπτα. Δύο χρόνια αργότερα, η Google ισχυρίστηκε την κβαντική της υπεροχή όταν η μηχανή της 54-qubit ολοκλήρωσε έναν υπολογισμό σε 200 δευτερόλεπτα ενώ ένας μη κβαντικός υπερυπολογιστής θα χρειαζόταν έως και 10.000 χρόνια για να τον ολοκληρώσει. Ωστόσο, ενώ το μηχάνημα είναι μια εντυπωσιακή απόδειξη της ιδέας, δεν είναι ακόμη σε θέση να εκτελέσει υπολογισμούς με συγκεκριμένες πρακτικές χρήσεις. Ένας κβαντικός υπολογιστής γενικής χρήσης θα απαιτούσε περισσότερο από 1 εκατομμύριο qubits που λειτουργούν κοντά στο απόλυτο μηδέν (-273 °C). Ως εκ τούτου, φαίνεται ότι οι αξιόπιστοι και χρήσιμοι κβαντικοί υπολογιστές θα παραμείνουν πιθανώς μη διαθέσιμοι για τουλάχιστον την επόμενη δεκαετία. Κάποιοι προτείνουν ότι είναι ένας κινούμενος στόχος που θα παραμένει πάντα δελεαστικά απρόσιτος. Εδώ, αρκεί να σημειωθεί ότι ο κβαντικός υπολογισμός είναι μια θεωρητική εξέλιξη που, εάν επιτευχθεί, θα μπορούσε να επιτρέψει την εμφάνιση της μελλοντικής εξέλιξης της Τεχνητής Νοημοσύνης, είτε εφαρμόζοντας τις τρέχουσες μεθόδους πιο αποτελεσματικά είτε επιτρέποντας την ανάπτυξη εντελώς νέων προσεγγίσεων (European Parliamentary Research Service, 2020).

[5] Τα μεμονωμένα bit δεδομένων σε κανονικούς υπολογιστές υπάρχουν σε μία κατάσταση, είτε 0 είτε 1. Τα μεμονωμένα bit σε έναν κβαντικό υπολογιστή, γνωστά ως «qubits» μπορούν να υπάρχουν και στις δύο καταστάσεις ταυτόχρονα. Εάν κάθε qubit μπορεί να είναι ταυτόχρονα και 0 και 1, τότε τέσσερα qubits μαζί θα μπορούσαν να βρίσκονται ταυτόχρονα σε 16 διαφορετικές καταστάσεις (0000, 0001, 0010, κ.λπ.). Μικρές αυξήσεις στον αριθμό των qubit οδηγούν σε μαζικές αυξήσεις (2n) στον αριθμό των ταυτόχρονων καταστάσεων. Έτσι, 50 qubits μαζί μπορούν να βρίσκονται σε πάνω από ένα τρισεκατομμύριο διαφορετικές καταστάσεις ταυτόχρονα. Ο κβαντικός υπολογισμός λειτουργεί εκμεταλλευόμενος αυτήν την ταυτόχρονη εύρεση λύσεων σε πολύπλοκα προβλήματα πολύ γρήγορα.

2.2.3.3 Η εξελισσόμενη Τεχνητή Υπερευφυΐα

Ένας προτεινόμενος δρόμος προς την Τεχνητή Υπερευφυΐα (Artificial Super Intelligence - ASI) είναι η ανάπτυξη ολοένα και πιο εξελιγμένων Τεχνητών Νευρωνικών Δικτύων μέσω καλύτερων εξελικτικών μεθόδων που εκτελούνται σε πιο ισχυρούς υπολογιστές. Η εξελισσόμενη «υπερευφυΐα» θα μπορούσε να ξεκινήσει με τον σχεδιασμό ενός αλγορίθμου για τη δημιουργία τεράστιων πληθυσμών πολλαπλών ειδών Τεχνητών Νευρωνικών Δικτύων σε ένα τεράστιο προσομοιωμένο εξελικτικό περιβάλλον. Χρειάστηκαν εκατομμύρια χρόνια από την εμφάνιση των πρώτων βιολογικών νευρώνων μέχρι την εξέλιξη των ευφυών ανθρώπων και, κατά τη διάρκεια αυτής της περιόδου, μια τεράστια γκάμα μορφών ζωής κατέλαβε το πιο άγνωστο και περίπλοκο περιβάλλον που γνώριζε η ανθρωπότητα: τη Γη. Οι τρέχουσες δυνατότητες επεξεργασίας δε θα μπορούσαν να ανταποκριθούν στα εξελικτικά περιβάλλοντα αυτής της κλίμακας. Ωστόσο, ορισμένες συντομεύσεις για την υπερευφυΐα ενδέχεται να είναι δυνατές σε περιβάλλοντα υπολογιστή. Η βιολογική εξέλιξη μέχρι σήμερα ασχολήθηκε όχι μόνο με τη νοημοσύνη, αλλά και με την ανάπτυξη πολύπλοκων οργάνων και φυσικών αμυνών. Η βιολογική ανάπτυξη μπορεί επίσης να κολλήσει σε μη παραγωγικές εξελικτικές απολήξεις, όπως οι αλγόριθμοι ML κολλάνε στα τοπικά βέλτιστα. Ένας προσομοιωτής Τεχνητής Νοημοσύνης θα μπορούσε να παρακάμψει αυτές και πολλές άλλες χρονοβόρες βιολογικές διεργασίες, όπως η ωρίμανση και η γήρανση, αφαιρώντας τους αναδυόμενους πληθυσμούς από εξελικτικά αδιέξοδα και να παρακάμψει τους περιττούς περισπασμούς που σχετίζονται με τη σωματική επιβίωση και αναπαραγωγή. Επιλέγοντας αποκλειστικά για γενική νοημοσύνη, ίσως θα μπορούσε να αναπτυχθεί πιο γρήγορα για υπολογιστές από ό,τι για την ανθρωπότητα (European Parliamentary Research Service, 2020).

Άραγε, θα μπορούσε αυτό να οδηγήσει στην εμφάνιση του AGI και, με την εκτέλεση της προσομοίωσης ακόμη περισσότερο, στην εμφάνιση του ASI;

Η απάντηση δεν είναι ξεκάθαρη. Λόγω των απαιτήσεων του περιβάλλοντος στο οποίο εξελίχθηκαν, οι άνθρωποι έγιναν καλοί στο να αναγνωρίζουν τα ζώα και να κατανοούν τις κινήσεις τους, αλλά όχι στο να κάνουν γρήγορους και περίπλοκους μαθηματικούς υπολογισμούς. Ομοίως, το είδος των δυνατοτήτων που αναπτύσσουν οι προσομοιωμένοι πληθυσμοί θα εξαρτηθεί από το είδος των προκλήσεων που αντιμετωπίζουν και τους πόρους τους οποίους μπορούν να αξιοποιήσουν για την ανάπτυξη λύσεων. Ενδέχεται, λοιπόν, να μην αντιμετωπίσουν ποτέ τα ανθρώπινα προβλήματα που θα τους οδηγήσουν να αναπτύξουν λύσεις που να μοιάζουν με αυτές του ανθρώπου. Αυτό δεν σημαίνει ότι δεν θα μπορούσαν

να αναπτύξουν εκπληκτικές λύσεις σε ενδιαφέροντα προβλήματα, είναι, ωστόσο, περισσότερο ζήτημα εάν οι άνθρωποι θα μπορούσαν να συσχετιστούν με αυτά τα προβλήματα και λύσεις με χρήσιμο ή ουσιαστικό τρόπο (Collins, 2018).

2.2.3.4 Εξομοίωση εγκεφάλου και τεχνητή συνείδηση

Μια άλλη προτεινόμενη προσέγγιση για την ανάπτυξη της Γενικής Τεχνητής Νοημοσύνης θα μπορούσε να διατηρήσει την ευθυγράμμιση με την ανθρώπινη νοημοσύνη, παράγοντας ένα πολύ λεπτομερές ψηφιακό αντίγραφο του ανθρώπινου εγκεφάλου, συμπεριλαμβανομένων όλων των νευρώνων και των συνδέσεων διαφόρων δυνατοτήτων. Εάν υποθέσουμε ότι μπορεί να έχουμε την τεχνική ικανότητα να το κάνουμε αυτό, αλλά και μια αρκετά ακριβή και πλήρη κατανόηση του εγκεφάλου, το αποτέλεσμα μπορεί να είναι μια πλήρης ψηφιακή εξομοίωση ενός ανθρώπινου μυαλού, με την ικανότητα να επεξεργάζεται αισθητηριακές εισροές, να θυμάται, να μαθαίνει και να εφαρμόσει τη γενική νοημοσύνη.

Δεδομένου ότι το Τεχνητό Νευρωνικό Σύστημα θα χρειαστεί να προσομοιώσει περίπου 86 δισεκατομμύρια νευρώνες και περίπου 150 τρισεκατομμύρια συνδέσεις σε πραγματικό χρόνο, η πλήρης εξομοίωση του εγκεφάλου παραμένει σταθερά στον τομέα της εικασίας. Ωστόσο, σοβαρά ερευνητικά προγράμματα ασχολούνται στον εν λόγω τομέα, συμπεριλαμβανομένου του προγράμματος Human Brain που χρηματοδοτείται από δισεκατομμύρια ευρώ, το οποίο έχει σημειώσει κάποια πρόοδο στη χαρτογράφηση εγκεφάλου ποντικιών, αν και τα μοντέλα είναι ημιτελή και λειτουργούν πολύ πιο αργά από τον πραγματικό χρόνο. Δεν είναι ξεκάθαρο αν ένας τέτοιος μιμούμενος εγκέφαλος θα χρειαζόταν ύπνο, πόσο περιορισμένη μπορεί να είναι η τελική του ικανότητα για μνήμη ή γνώση, ούτε εάν θα βιώσει πόνο, θλίψη, υπαρξιακή τρομοκρατία ή συνείδηση (European Parliamentary Research Service, 2020).

2.2.3.5 Wetware και βιολογικά συστήματα

Το εκκολαπτόμενο πεδίο της Τεχνητής Ζωής (Alife) διαφέρει από την Τεχνητή Νοημοσύνη στο ότι οι ιδέες και οι τεχνικές του βασίζονται σε θεμελιώδεις βιολογικές διεργασίες και όχι σε νοημοσύνη ή τεχνογνωσία. Ωστόσο, έχει κάποιες διασταυρώσεις με την Τεχνητή Νοημοσύνη, ιδιαίτερα στο πλαίσιο των προσεγγίσεων εξελικτικής μάθησης και άλλων μεθόδων εμπνευσμένων από τη φύση. Όπως και η Τεχνητή Νοημοσύνη, το Alife αναπτύσσεται, κυρίως, μέσω λογισμικού (κώδικας και δεδομένα) και υλικού (φυσικά στοιχεία), αλλά μπορεί, επίσης, να περιλαμβάνει «wetware», το οποίο αναφέρεται στη χρήση βιολογικών υλικών ως εξαρτημάτων του συστήματος που αναπτύσσεται. Μεταξύ τομέων,

όπως η επεξεργασία γονιδίων (η τροποποίηση του DNA) και η συνθετική βιολογία (η δημιουργία τεχνητών βιολογικών συστημάτων), το wetware κατέχει το επίκεντρο. Τόσο η συνθετική βιολογία όσο και η γονιδιακή επεξεργασία μπορεί να ωφεληθούν από τις γνώσεις της Τεχνητής Νοημοσύνης και προσδιορίζονται ως μια πιθανή οδός προς κάποια μελλοντική Bio-AI, αν και αυτό παραμένει μια μακρινή εικασία (European Parliament Research Service, 2020).

3. Τεχνητή Νοημοσύνη και Δημόσιος Τομέας

Η υιοθέτηση της Τεχνητής Νοημοσύνης από τις κυβερνήσεις αυξάνεται ραγδαία, όπως και οι υποχρεώσεις του δημόσιου τομέα να διασφαλίσει ότι η τεχνολογία χρησιμοποιείται με υπεύθυνο, αξιόπιστο και ηθικό τρόπο. Η εφαρμογή της Τεχνητής Νοημοσύνης στο δημόσιο τομέα απαιτεί μια στοχαστική και στρατηγική πορεία δράσης για την αξιοποίηση των μεγάλων ευκαιριών που υπόσχεται και τελικά τη δημιουργία αξίας από αυτές (Mehr, 2017).

Οι κυβερνήσεις διαδραματίζουν βασικό ρόλο στο να κάνουν την Τεχνητή Νοημοσύνη να λειτουργεί για την καινοτομία, την ανάπτυξη και τη δημόσια αξία. Πέραν του ρόλου τους ως ρυθμιστών του θεσμικού πλαισίου λειτουργίας της τεχνολογίας, οι κυβερνήσεις αποτελούν και κύριους χρήστες της, βελτιώνοντας τη λειτουργία του κυβερνητικού μηχανισμού και τη παροχή δημόσιας υπηρεσίας με σκοπό την αύξηση της αποδοτικότητας, μέσω της αυτοματοποίησης των φυσικών και ψηφιακών εργασιών, την βελτίωση της αποτελεσματικότητας, μέσω της λήψης καλύτερων πολιτικών αποφάσεων και καλύτερων αποτελεσμάτων ως απόρροια βελτιωμένων δυνατοτήτων πρόβλεψης και επίσης, την αμεσότερη ανταπόκριση στις ανάγκες των χρηστών, μέσω εξατομικευμένων και ανθρωποκεντρικών υπηρεσιών με βάση τον σχεδιασμό ανθρωποκεντρικών διεπαφών (Santiso, 2022).

Περισσότερες από εξήντα χώρες σε παγκόσμιο επίπεδο και είκοσι τέσσερις σε ευρωπαϊκό έχουν ήδη αναπτύξει εθνικές στρατηγικές για την Τεχνητή Νοημοσύνη, με ιδιαίτερη έμφαση στον τρόπο με τον οποίο αυτές οι στρατηγικές αντιμετωπίζουν τις προκλήσεις ανάπτυξης και χρήσης της Τεχνητής Νοημοσύνης που σχετίζονται με το δημόσιο τομέα. Στην έκθεσή του το Παρατηρητήριο για την Τεχνητή Νοημοσύνης (AI-Watch) παρουσιάζει την ανάλυση των εθνικών στρατηγικών των κρατών μελών της Ευρωπαϊκής Ένωσης, τονίζοντας τον τρόπο με τον οποίο τα κράτη μέλη στοχεύουν να ενισχύσουν τη χρήση της Τεχνητής Νοημοσύνης στον δικό τους δημόσιο τομέα (Tangi et al., 2022).

Τα συμπεράσματα της ανάλυσης αποτυπώνουν μια ομαδοποίηση των εθνικών στρατηγικών με βάση την εστίαση τους σε συγκεκριμένες πολιτικές υιοθέτησης της τεχνολογίας στο δημόσιο τομέα, η οποία σχετίζεται για κάποιες με τον εξωτερικό ή τον εσωτερικό τους προσανατολισμό ενώ για κάποιες άλλες με τα δεδομένα (Tangi et al., 2022).

Εθνικές στρατηγικές Τεχνητής Νοημοσύνης με εξωτερικό προσανατολισμό

Οι στρατηγικές αναγνωρίζουν ότι λόγω της πολυπλοκότητας των συστημάτων Τεχνητής Νοημοσύνης και των περιορισμένων δυνατοτήτων και τεχνογνωσίας των δημοσίων διοικήσεων να αντιμετωπίσουν αυτόνομα την Τεχνητή Νοημοσύνη, η ενίσχυση των σχέσεων και η συνεργασία δημόσιου και ιδιωτικού τομέα θα τονώσει και θα προωθήσει την ανάπτυξη και υιοθέτηση της τεχνολογίας στο δημόσιο τομέα. Δίνουν, λοιπόν, έμφαση στον τρόπο ενίσχυσης της διαδικασίας σύναψης συμβάσεων.

Ειδικότερα, οι στρατηγικές αυτές:

- Βοηθούν νεοφυείς και άλλες εταιρείες να αναδυθούν και να αναπτύξουν Τεχνητή Νοημοσύνη για χρήση στον κυβερνητικό τομέα.
- Αναγνωρίζουν ότι οι υφιστάμενες διαδικασίες προμηθειών περιορίζουν τις προμήθειες καινοτόμων τεχνολογιών εντός του δημόσιου τομέα και, ως εκ τούτου, λαμβάνουν μέτρα για τη βελτίωση της συνεργασίας μεταξύ του δημόσιου και του ιδιωτικού τομέα.

Εθνικές στρατηγικές Τεχνητής Νοημοσύνης με προσανατολισμό στα δεδομένα

Οι στρατηγικές αναγνωρίζουν τον καίριο ρόλο των δεδομένων και στοχεύουν στην διευκόλυνση τόσο της διαθεσιμότητας και της διασφάλισης της ποιότητας των δεδομένων όσο και στη βελτίωση της τεχνικής υποδομής για την υποστήριξη του γενικού οικοσυστήματος της χώρας για την ανάπτυξη της Τεχνητής Νοημοσύνης είτε από τον ιδιωτικό είτε από τον ίδιο τον δημόσιο τομέα.

Ειδικότερα οι στρατηγικές αυτές στοχεύουν στη:

- Διαθεσιμότητα περισσότερων δημοσίων συνόλων δεδομένων για την ανάπτυξη Τεχνητής Νοημοσύνης και διευκόλυνση της ανταλλαγής δεδομένων μεταξύ των δημόσιων υπηρεσιών.
- Βελτίωση της διακυβέρνησης δεδομένων, των προτύπων δεδομένων και των πρακτικών συλλογής δεδομένων για την ύπαρξη περισσότερων διαθέσιμων δεδομένων.
- Διασφάλιση ότι η συνολική συνδεσιμότητα και η υπολογιστική ισχύς υψηλής απόδοσης διατίθενται για την ανάπτυξη λύσεων Τεχνητής Νοημοσύνης.

Εθνικές στρατηγικές Τεχνητής Νοημοσύνης με εσωτερικό προσανατολισμό

Οι στρατηγικές επικεντρώνονται στη βελτίωση της εσωτερικής ικανότητας των δημοσίων διοικήσεων και των δημοσίων υπαλλήλων ως βασικό μέσο για την τόνωση της υιοθέτησης της Τεχνητής Νοημοσύνης στο δημόσιο τομέα.

Ειδικότερα, οι στρατηγικές αυτές περιγράφουν διάφορες δραστηριότητες, όπως:

- Δημιουργία νέων δημόσιων φορέων ή μονάδων/τμημάτων για την αντιμετώπιση της Τεχνητής Νοημοσύνης.
- Εκδηλώσεις εκπαίδευσης και ευαισθητοποίησης για την αύξηση της γνώσης και των δεξιοτήτων των δημοσίων διοικήσεων και των δημοσίων υπαλλήλων σχετικά με την Τεχνητή Νοημοσύνη.
- Απόκτηση περισσότερων τεχνικών γνώσεων μέσω προγραμμάτων προσλήψεων ή εξειδικευμένης εκπαίδευσης.

Τα συμπεράσματα της ανάλυσης των εθνικών στρατηγικών κρατών μελών της Ευρωπαϊκής Ένωσης αναδεικνύουν τον σχεδιασμό πολιτικών και δράσεων των εθνικών κυβερνήσεων για την ανάπτυξη, υιοθέτηση και χρήση της Τεχνητής Νοημοσύνης σε όλους τους τομείς του δημόσιου τομέα. Οι δημόσιοι οργανισμοί θα πρέπει να αρχίσουν να θεωρούν την Τεχνητή Νοημοσύνη όχι μόνο ως τομέα έρευνας και καινοτομίας, αλλά και ως ένα σύνολο διαθέσιμων τεχνολογιών για τη βελτίωση της διοικητικής μηχανής. Αν και υπάρχει μεγάλος προβληματισμός σχετικά με τις δεξιότητες που απαιτούνται για την ανάπτυξη ενός συστήματος Τεχνητής Νοημοσύνης και τον τρόπο απόκτησής τους, οι δημόσιοι οργανισμοί έχουν εντοπίσει διάφορους τρόπους για την ανάπτυξη ενός έργου από τεχνική άποψη, από την ύπαρξη εσωτερικής ομάδας έως την ανάθεση της ανάπτυξης σε εξωτερικούς συνεργάτες. Ωστόσο, οι δημόσιοι οργανισμοί δεν μπορούν να βασίζονται πλήρως σε εξωτερικές ικανότητες γιατί αποτελεί ανάγκη να υπάρχει κάποιος μέσα στον οργανισμό που να είναι σε θέση να κατανοήσει και να κατευθύνει την ανάπτυξη ενός έργου Τεχνητής Νοημοσύνης ή να το προσαρμόσει στις ανάγκες του, όπως επίσης και να παρακολουθεί τη διαγωνιστική διαδικασία, διασφαλίζοντας ότι ο ανάδοχος αναπτύσσει σωστά την απαιτούμενη λύση, τόσο από πλευράς λειτουργικότητας όσο και από πλευράς τήρησης των ηθικών απαιτήσεων και της διαφάνειας. Συνεπώς, οι δημόσιοι οργανισμοί θα πρέπει να διασφαλίζουν την παρουσία εσωτερικής γνώσης για την Τεχνητή Νοημοσύνη, για τη μερική ή πλήρη εσωτερική ανάπτυξη της λύσης, για την κατεύθυνση και την προσαρμογή του συστήματος που

αναπτύσσεται από εξωτερικούς προμηθευτές ή/και για τη διασφάλιση της σωστής διαχείρισης των διαδικασιών προμηθειών (Tangi et al., 2022).

Οι δημόσιοι υπάλληλοι θα πρέπει να έχουν μια γενική κατανόηση της έννοιας της Τεχνητής Νοημοσύνης ώστε να είναι σε θέση να αμφισβητούν το σύστημα, όταν οι αλγόριθμοι είναι προκατειλημμένοι και κάνουν λανθασμένες προτάσεις ή παίρνουν λανθασμένες αποφάσεις. Ως εκ τούτου, οι δημόσιοι οργανισμοί θα πρέπει να θεωρούν την Τεχνητή Νοημοσύνη ως μια τεχνολογία που θα επηρεάσει τις καθημερινές ρουτίνες των περισσότερων εργαζομένων και επομένως θα πρέπει να αρχίσουν να σκέφτονται την ευρεία διάχυση της βασικής γνώσης σχετικά με το πώς λειτουργεί ο αλγόριθμος και πώς να αντιμετωπίζουν συστήματα που χρησιμοποιούν τεχνικές Τεχνητής Νοημοσύνης. Οι συνεργασίες του δημόσιου τομέα με εξωτερικούς εταίρους, όπως πανεπιστήμια, νεοφυείς επιχειρήσεις και εταιρίες του ιδιωτικού τομέα παρέχουν πρόσβαση σε δυνατότητες και δεξιότητες που δεν είναι διαθέσιμες εσωτερικά. Δεδομένου ότι πιθανότατα οι δημόσιοι οργανισμοί θα χρειάζονταν υποστήριξη για την ανάπτυξη ενός συστήματος Τεχνητής Νοημοσύνης, θα πρέπει να επιλέξουν προσεκτικά τους κατάλληλους συνεργάτες ή/και προμηθευτές και να εξισορροπήσουν την εσωτερική και την εξωτερική ανάπτυξη (Tangi et al., 2022).

Ένα μεγάλο μέρος της συζήτησης γύρω από την Τεχνητή Νοημοσύνη και τις ιδιαιτερότητες που σχετίζονται με την εφαρμογή της εμπεριέχει τους κινδύνους της. Αυτό γίνεται κρίσιμο στον δημόσιο τομέα όπου, λόγω της φύσης του, οι άδικες συμπεριφορές είναι απαράδεκτες. Ως εκ τούτου, πριν ξεκινήσει η ανάπτυξη ενός έργου Τεχνητής Νοημοσύνης, είναι απαραίτητη η αξιολόγηση του έργου ως προς την πλήρη συμμόρφωσή του με τις δημόσιες αξίες της Ευρωπαϊκής Ένωσης όπως η ισότητα, η μη διάκριση και η διαφάνεια. Αυτές οι απαιτήσεις είναι ανεξάρτητες από οποιαδήποτε νομοθεσία και πρέπει να εφαρμόζονται σε οποιοδήποτε σύστημα Τεχνητής Νοημοσύνης. Η δημόσια διοίκηση αναμένεται να είναι στην πρώτη γραμμή όσον αφορά την αξιόπιστη και ανθρωποκεντρική χρήση της Τεχνητής Νοημοσύνης και ο δημόσιος τομέας θα πρέπει να αρχίσει να δομεί μια καλά καθορισμένη διαδικασία για τη διασφάλιση της δίκαιης και χωρίς διακρίσεις χρήσης της. Οι κίνδυνοι θα πρέπει να αξιολογούνται συστηματικά με μια δομημένη και καλά καθορισμένη διαδικασία, αποφεύγοντας κάθε μορφή μεροληπτικής και αθέμιτης χρήσης του συστήματος Τεχνητής Νοημοσύνης και διασφαλίζοντας την ανθρωποκεντρική χρήση της (Tangi et al., 2022).

Δεδομένων των κινδύνων, θα πρέπει να ληφθούν και τα κατάλληλα μέτρα μετριασμού τους. Η διαφάνεια, η επεξηγησιμότητα των αλγορίθμων και η ανθρώπινη επίβλεψη είναι πιθανώς τα πιο συζητημένα και εφαρμοσμένα μέτρα μετριασμού, παρόλο που το φάσμα των απαραίτητων μέτρων είναι ευρύτερο. Για παράδειγμα, μια εξαιρετικά σχετική συζήτηση που αποτελεί πλέον κύριο μέλημα των δημόσιων οργανισμών που ασχολούνται με την Τεχνητή Νοημοσύνη, αφορά στη συλλογή κατάλληλων και ποιοτικών δεδομένων και ποια από αυτά μπορούν να επαναχρησιμοποιηθούν για την εκπαίδευση των αλγορίθμων (Tangi et al., 2022).

3.1 Περιπτώσεις χρήσης Τεχνητής Νοημοσύνης στο Δημόσιο Τομέα

Η επικινδυνότητα της Τεχνητής Νοημοσύνης δεν θα πρέπει να θεωρείται αποτρεπτικός παράγοντας για τη χρήση της στη διαχείριση και την παροχή δημόσιων υπηρεσιών. Οι δημόσιες διοικήσεις έχουν ήδη αρχίσει να υιοθετούν την Τεχνητή Νοημοσύνη σε διάφορους και διαφορετικούς τομείς του δημόσιου τομέα και δεν διερευνούν απλώς τις δυνατότητές της με πιλοτικές λύσεις ή περιβάλλοντα δοκιμών, αλλά αρκετές λύσεις Τεχνητής Νοημοσύνης έχουν ήδη αναπτυχθεί και χρησιμοποιούνται σε καθημερινές λειτουργίες (Misuraca & van Noordt, 2020. Molinari et a., 2021).

Στη συνέχεια θα παραθέσουμε το παράδειγμα της Φινλανδίας που επιδιώκει να δημιουργήσει ένα πρόγραμμα για την Τεχνητή Νοημοσύνη στο δημόσιο τομέα με ανθρωποκεντρική εστίαση και περιπτώσεις χρήσης Τεχνητής Νοημοσύνης που έχουν ήδη υιοθετηθεί σε διάφορες χώρες του κόσμου για την αντιμετώπιση συγκεκριμένων προκλήσεων και ζητημάτων του δημόσιου τομέα. Οι περιπτώσεις χρήσης παρέχουν μια σύνοψη πρωτοβουλιών Τεχνητής Νοημοσύνης για την αντιμετώπιση του κινδύνου στις μεταφορές στον Καναδά, για την αντιμετώπιση της διαφθοράς στην Κίνα και τη Βραζιλία, για την συμμετοχή πολιτών στο Βέλγιο και στη Νιγηρία, για τη βελτίωση της αποτελεσματικότητας και της συμμόρφωσης στην τελωνειακή διοίκηση των Ηνωμένων Πολιτειών, για την αντιμετώπιση της υγειονομικής κρίσης του COVID-19 στη Σιγκαπούρη, για τη βελτίωση των δημόσιων προμηθειών στη Νότια Κορέα και στις Ηνωμένες Πολιτείες, για τη βελτίωση της αποτελεσματικότητας του τομέα της δικαιοσύνης στην Κίνα και στο Ηνωμένο Βασίλειο και για τη βελτίωση της φορολογικής συμμόρφωσης στην Αρμενία και στις Ηνωμένες Πολιτείες.

3.1.1 Εθνική Στρατηγική Τεχνητής Νοημοσύνης – Το παράδειγμα της Φινλανδίας

Πρόθεση της Φινλανδίας είναι να γίνει παγκόσμιος ηγέτης στην εφαρμογή της Τεχνητής Νοημοσύνης. Στον στρατηγικό της σχεδιασμό υπάρχει μια ισχυρή επιμέρους εστίαση στον δημόσιο τομέα, για τον οποίο οραματίζεται μια κυβέρνηση που παρέχει εξατομικευμένες υπηρεσίες σε όλους τους πολίτες και για όλα τα στάδια της ζωής τους προκειμένου να υποστηρίξει μια κοινωνία που λειτουργεί με το βέλτιστο δυνατό τρόπο. Με μοναδικό τρόπο, σε σύγκριση με άλλες εθνικές στρατηγικές, η προσέγγιση της Φινλανδίας τοποθετεί την αποτελεσματικότητα του δημόσιου τομέα και την αποτελεσματικότητα των υπηρεσιών του στο ίδιο επίπεδο με την οικονομική ανάπτυξη (Berryhill et al., 2019).

Οι στόχοι που σχετίζονται άμεσα με την καινοτομία και τον μετασχηματισμό του δημόσιου τομέα περιλαμβάνουν τα ακόλουθα:

- Ανάπτυξη νέων μοντέλων λειτουργίας για τη μετάβαση από τις δραστηριότητες που βασίζονται στον οργανισμό σε προσεγγίσεις που αφορούν όλο το σύστημα.
- Προσαρμογή του ρόλου της κυβέρνησης για να διασφαλίσει ότι οι πολίτες έχουν το δικαίωμα να καθορίζουν ανεξάρτητα τον τρόπο χρήσης των δεδομένων τους, προστατεύοντας παράλληλα το απόρρητο των πολιτών.
- Βελτίωση της διαλειτουργικότητας των κρατικών δεδομένων και άνοιγμα αυτών των δεδομένων για την τροφοδοσία της καινοτομίας σε όλους τους τομείς.
- Δημιουργία Κέντρου Αριστείας για την Τεχνητή Νοημοσύνη, ένα εικονικό πανεπιστήμιο Τεχνητής Νοημοσύνης και ένα πρόγραμμα μεταπτυχιακών σπουδών στην Τεχνητή Νοημοσύνη για την ενίσχυση της δεξαμενής ταλέντων τόσο για τον ιδιωτικό όσο και για το δημόσιο τομέα.
- Επιδίωξη και οικοδόμηση ενός δικτύου για συμπράξεις δημόσιου και ιδιωτικού τομέα για να επιτραπούν συνεργατικές πρωτοβουλίες, ανταλλαγή γνώσεων και καλύτερη υιοθέτηση της πολυδιάστατης σκέψης.
- Διοργάνωση δημόσιας συζήτησης σχετικά με την ηθική της Τεχνητής Νοημοσύνης μέσω προσωπικών εκδηλώσεων και διαδικτύου.
- Καταστροφή σιλό εντός και μεταξύ επιχειρήσεων και δημόσιων υπηρεσιών.
- Αναθεώρηση της νομοθεσίας περί δημοσίων συμβάσεων για να καταστεί δυνατή αποτελεσματική η από κοινού ανάπτυξη δημόσιου και ιδιωτικού τομέα.

Κρίσιμης σημασίας για τον δημόσιο τομέα είναι η δημιουργία του εθνικού προγράμματος για την Τεχνητή Νοημοσύνη *AuroraAI*. Το *AuroraAI* επιδιώκει να παρέχει ένα ολιστικό σύνολο εξατομικευμένων υπηρεσιών με γνώμονα την Τεχνητή Νοημοσύνη για πολίτες και επιχειρήσεις με τρόπο ανθρωποκεντρικό και λειτουργεί με απώτερο στόχο την ευημερία τους. Το *AuroraAI*, ως ευρύτερη έννοια, έχει ως στόχο να δώσει στους πολίτες τη δυνατότητα να έχουν πρόσβαση στο ευρύ φάσμα υπηρεσιών που διατίθενται από διάφορους κρατικούς και διατομεακούς παρόχους υπηρεσιών με απρόσκοπτο τρόπο (AuroraAI, 2019).

Το πρόγραμμα *AuroraAI* έχει ανθρωποκεντρική εστίαση. Επιδιώκει να προσανατολίσει την παροχή υπηρεσιών γύρω από πολίτες και επιχειρήσεις, συνδυάζοντας δεδομένα από πολλαπλούς τομείς και δημιουργώντας ένα δίκτυο εφαρμογών με επίκεντρο την Τεχνητή Νοημοσύνη των πολιτών που παρέχουν υπηρεσίες όταν χρειάζονται – γύρω από διάφορες επιχειρηματικές δραστηριότητες ή στάδια ζωής και συμβάντα, όπως ο τοκετός, η αγορά σπιτιού ή η συνταξιοδότηση. Με τη συγκέντρωση δεδομένων για τη δημιουργία ανθρωποκεντρικών υπηρεσιών, «η επίγνωση της κατάστασης που βασίζεται σε δεδομένα διευκολύνει τη στόχευση αποτελεσματικών υπηρεσιών με βάση τις πραγματικές ανάγκες των ατόμων και δίνει τη δυνατότητα στους ανθρώπους να διαχειρίζονται τη ζωή τους πιο αποτελεσματικά σε διάφορες συνθήκες ζωής» (AuroraAI, 2019). Αυτό διευκολύνεται από τη χρήση της Ενισχυτικής Μάθησης (RL), μέσω της οποίας οι εφαρμογές δικτύου βελτιώνονται με βάση τα σχόλια των χρηστών. Το *AuroraAI* έχει σχεδιαστεί για να περιλαμβάνει όχι μόνο δημόσιες, αλλά και ιδιωτικές υπηρεσίες και υπηρεσίες της κοινωνίας των πολιτών.

Προς το παρόν, το *AuroraAI* λειτουργεί πιλοτικά, εστιάζοντας σε τρία γεγονότα ζωής:

- μετακίνηση για σπουδές
- παραμονή στην αγορά εργασίας μέσω της δια βίου μάθησης
- διασφάλιση της οικογενειακής ευημερίας μετά το διαζύγιο

Στο δίκτυο *AuroraAI*, το *DigiMe* λειτουργεί ως η ψηφιακή «περσόνα» του χρήστη και αναφέρεται στον τρόπο με τον οποίο ένας χρήστης μπορεί να χρησιμοποιήσει τα προσωπικά του δεδομένα στο δίκτυο. Το βασικό χαρακτηριστικό είναι ότι ο χρήστης μπορεί να διαχειρίζεται τα δικά του δεδομένα και να τα επεξεργάζεται σε περιστασιακά, προσωρινά ψηφιακά προφίλ, προκειμένου να έχει πρόσβαση σε μια εξατομικευμένη προσφορά υπηρεσίας σε πραγματικό χρόνο.

Το δίκτυο *AuroraAI* στοχεύει να προβλέψει τις ανάγκες του χρήστη και να βελτιώσει την εμπειρία της υπηρεσίας διαμορφώνοντας μια συνεκτική συνολική άποψη για κάθε χρήστη. Αυτό γίνεται συγκρίνοντας τα χαρακτηριστικά του χρήστη, δηλαδή τα δεδομένα που μοιράζεται ο χρήστης σε μια δεδομένη κατάσταση, με άλλους χρήστες προκειμένου να εντοπιστούν ομοιότητες, διαφορές και μοτίβα. Ο μόνος τρόπος για να κατανοήσουμε αντικειμενικά ένα άτομο είναι συγκρίνοντάς το με ένα σύνολο άλλων ατόμων. Ωστόσο, πρέπει να σημειωθεί ότι ο χρήστης δεν πρέπει να είναι αναγνωρίσιμος σε κανένα στάδιο, εάν η σύγκριση πρόκειται να γίνει σε ανώνυμη βάση. Η ιδέα του *DigiMe* μπορεί να αναπτυχθεί σε περιπτώσεις όπου η σύνδεση μεταξύ του πραγματικού ατόμου και της ψηφιακής του προσωπικότητας πρέπει να γίνει αόρατη. Ο χρήστης συγκεντρώνει τα προσωπικά του δεδομένα για να δημιουργήσει μια συλλογή ή περίληψη που μπορεί να υποβληθεί σε επεξεργασία από το δίκτυο χωρίς το δίκτυο να τα συνδέσει με τα δεδομένα πηγής του χρήστη. Η ανάπτυξη και η ελεγχόμενη δοκιμή μιας τέτοιας ιδέας είναι σημαντικές μέθοδοι για τη μετάβαση προς μια ανθρωποκεντρική κοινωνία στην οποία οι χρήστες μπορούν να εμπιστεύονται το απόρρητο των δεδομένων (AuroraAI, 2019).

Είναι σημαντικό ότι η εθνική στρατηγική Τεχνητής Νοημοσύνης της Φινλανδίας όσον αφορά τους πολίτες ακολουθεί τις αρχές του *«MyData»*, σύμφωνα με τις οποίες ο πολίτης και κανείς άλλος είναι ο κάτοχος των δικών του δεδομένων. Ως ιδιοκτήτης, ένας πολίτης έχει τον πλήρη έλεγχο των δεδομένων του. Έχει την εξουσία να επιλέγει και να αποκλείει υπηρεσίες και να λαμβάνει αποφάσεις σχετικά με ποιον μοιράζεται τα δεδομένα του (AuroraAI, 2019).

3.1.2 Τεχνητή Νοημοσύνη και Μεταφορές - Το παράδειγμα του Καναδά “bomb-in-the-box”

Το πρόβλημα

Η Transport Canada είναι η αρμόδια υπηρεσία για τις πολιτικές και τα προγράμματα μεταφορών της κυβέρνησης του Καναδά. Εργάζεται για την προώθηση ασφαλών, αποτελεσματικών και περιβαλλοντικά υπεύθυνων μεταφορών.

Κάθε χρόνο, η ομάδα Προφόρτωσης Αεροπορικού Φορτίου (Pre-load Air Cargo Targeting - PACT) της Transport Canada λαμβάνει σχεδόν ένα εκατομμύριο αρχεία προφόρτωσης αεροπορικού φορτίου ετησίως που περιέχουν πληροφορίες, όπως όνομα και διεύθυνση αποστολέα, όνομα και διεύθυνση παραλήπτη, βάρος και αριθμό τεμαχίων. Κάθε εγγραφή μπορεί να περιλαμβάνει από 10 έως 100 πεδία, ανάλογα με τον αερομεταφορέα και το

επιχειρηματικό μοντέλο του αποστολέα, οπότε είναι αδύνατον για έναν εργαζόμενο να ελέγξει πάνω από το 10% των αρχείων προφόρτωσης. Μέχρι σήμερα, πολύ λίγες κυβερνήσεις διαθέτουν τους ειδικούς πόρους για τη σάρωση αρχείων αεροπορικού φορτίου για ανίχνευση κίνδυνου πριν από τη φόρτωση και από αυτές που έχουν, καμία δεν χρησιμοποιεί Τεχνητή Νοημοσύνη (Berryhill et al., 2019).

Η λύση

Η Transport Canada αποφάσισε να υιοθετήσει την Τεχνητή Νοημοσύνη για να βελτιώσει τις διαδικασίες, να απελευθερώσει ανθρώπινους πόρους και να ενισχύσει την ασφάλεια των αεροπορικών μεταφορών εμπορευμάτων. Σε πρώτη φάση ξεκίνησε με τη διερεύνηση της χρήσης της Τεχνητής Νοημοσύνης σε αξιολογήσεις αεροπορικών φορτίων υψηλού κινδύνου. Για να επιτευχθεί αυτό, συγκρότησε μια διεπιστημονική εσωτερική ομάδα έργου υποστηριζόμενη από μια εξωτερική εταιρεία πληροφορικής με εξειδίκευση στην Τεχνητή Νοημοσύνη. Η Transport Canada προσπάθησε να απαντήσει σε δύο ερωτήσεις που σχετίζονται με την απόδοσή της:

- Μπορεί η Τεχνητή Νοημοσύνη να βελτιώσει την ικανότητά μας να ασκούμε έλεγχο βάσει κινδύνου;
- Πώς μπορούμε να βελτιώσουμε την αποτελεσματικότητα και την αποδοτικότητα κατά την αξιολόγηση του κινδύνου στις αεροπορικές αποστολές φορτίου;

Για να απαντήσει σε αυτά τα ερωτήματα, η ομάδα καινοτομίας ανέπτυξε και εφάρμοσε μια προσέγγιση δύο βημάτων το 2018. Ως πρώτο βήμα, χρησιμοποίησε δεδομένα από προηγούμενα αρχεία αεροπορικών φορτίων και μη αυτόματες αξιολογήσεις κινδύνου για να διερευνήσει μη εποπτευόμενες και εποπτευόμενες προσεγγίσεις Τεχνητής Νοημοσύνης. Χρησιμοποιώντας την εποπτευόμενη προσέγγιση, η ομάδα προσπάθησε να κατανοήσει τη σχέση μεταξύ των εισροών (αρχεία φορτίου) και του αποτελέσματος (δηλαδή, αυτό το αρχείο φορτίου έδειξε μεγαλύτερο επίπεδο κινδύνου, όπως με βάση προηγούμενες μη αυτόματες εκτιμήσεις κινδύνου;). Χρησιμοποιώντας μάθηση χωρίς επίβλεψη, η ομάδα προσπάθησε να κατανοήσει τις σχέσεις μεταξύ όλων των εισροών φορτίου προκειμένου να εντοπίσει σπάνιες ή ασυνήθιστες αποστολές, οι οποίες θα μπορούσαν να είναι ενδεικτικές του κινδύνου (Berryhill et al., 2019).

Δεύτερον, η ομάδα ανέπτυξε μια απόδειξη της ιδέας για να δοκιμάσει την επεξεργασία φυσικής γλώσσας (NLP) σε ένα διαφορετικό υποσύνολο δεδομένων. Ο στόχος ήταν η δυνατότητα επεξεργασίας αρχείων αεροπορικού φορτίου και η αυτόματη προσθήκη

ετικετών σε ένα αρχείο φορτίου με δείκτη κινδύνου με βάση τα περιεχόμενα των πεδίων «ελεύθερου κειμένου» στα αρχεία αεροπορικού φορτίου και σε άλλα δομημένα πεδία. Αυτό ολοκληρώθηκε το πρώτο τρίμηνο του 2018 και έδειξε ότι η NLP μπορούσε να ταξινομήσει με επιτυχία τα δεδομένα φορτίου σε σημαντικές κατηγορίες σε πραγματικό χρόνο.

Και τα δύο βήματα οδήγησαν σε νέες πληροφορίες σχετικά με τα κρυφά μοτίβα που μπορεί να υποδηλώνουν κίνδυνο. Ως αποτέλεσμα, η ομάδα μπόρεσε να χρησιμοποιήσει την Τεχνητή Νοημοσύνη για να δημιουργήσει αυτόματα ακριβείς δείκτες κινδύνου. Μέσω αυτού του πιλότου, η Transport Canada έμαθε ότι η Τεχνητή Νοημοσύνη ήταν πράγματι μια βιώσιμη λύση για την αντιμετώπιση των βασικών της ερωτημάτων (Berryhill et al., 2019).

Το αποτέλεσμα

Πριν από την εισαγωγή της Τεχνητής Νοημοσύνης, η διεξαγωγή αξιολογήσεων κινδύνου ήταν επαχθής και χρονοβόρα. Χρειάστηκαν χιλιάδες ώρες ετησίως για την εισαγωγή, τον καθαρισμό και την αρχειοθέτηση δεδομένων. Με την εισαγωγή της Τεχνητής Νοημοσύνης, μεγάλο μέρος αυτής της διαδικασίας έχει αυτοματοποιηθεί και οι αξιολογήσεις κινδύνου διεξάγονται σε πραγματικό χρόνο. Η Τεχνητή Νοημοσύνη βοηθά το PACT στην επίτευξη ασφαλών αποτελεσμάτων και του δίνει τη δυνατότητα να σαρώσει περισσότερα φορτωτικά έγγραφα από περισσότερους μεταφορείς από ποτέ.

Τα αποτελέσματα του συγκεκριμένου μοντέλου είναι πολλά υποσχόμενα. Γίνονται προκαταρκτικές συζητήσεις εντός της κυβέρνησης σχετικά με την προσαρμογή του και σε άλλους τρόπους μεταφοράς (π.χ. θαλάσσιες, σιδηροδρομικές, οδικές κ.λπ.) ή ακόμη και για την επέκτασή του για την υποστήριξη της υπηρεσίας τελωνείων του Καναδά (Berryhill et al., 2019).

3.1.3 Τεχνητή Νοημοσύνη και Αντιμετώπιση Διαφθοράς – Το παράδειγμα της Βραζιλίας

Το πρόβλημα

Εκτιμάται ότι η Βραζιλία χάνει 3-5% του ΑΕΠ, ετησίως, λόγω της διαφθοράς. Οι δημόσιες συμβάσεις αποτελούν χώρο υψηλού κινδύνου για τις δημόσιες δαπάνες. Πάνω από 48.000 εταιρείες συμμετείχαν σε δημόσιους διαγωνισμούς μεταξύ 2016-2018 μόνο στην Πολιτεία του Σάο Πάολο.

Οι κρατικές υπηρεσίες δεν διαθέτουν τα εργαλεία ή την ικανότητα να διεξάγουν συστηματικές αξιολογήσεις κινδύνου απάτης. Η τρέχουσα προσέγγιση, η οποία εξαρτάται

σε μεγάλο βαθμό από χειροκίνητη εισαγωγή στοιχείων, είναι χρονοβόρα και αναποτελεσματική ενώ η ερμηνεία των στοιχείων που αφορούν εταιρίες υψηλού κινδύνου για διαφθορά γίνεται από χρήστες περιορισμένων δυνατοτήτων (World Bank, 2020).

Η λύση

Η ομάδα της Παγκόσμιας Τράπεζας στη Βραζιλίας ανέπτυξε ένα σύστημα Τεχνητής Νοημοσύνης που εντοπίζει 225 κόκκινες σημαίες πιθανής απάτης στις διαδικασίες δημοσίων συμβάσεων και μπορεί να βοηθήσει στη βελτίωση των κρατικών δαπανών. Η Παγκόσμια Τράπεζα συνεργάστηκε με την πόλη του Σάο Πάολο, τις πολιτείες του Ρίο ντε Τζανέιρο και του Μάτο Γκρόσο και με το Ομοσπονδιακό Υπουργείο Υγείας για να αξιοποιήσει τις τεράστιες ποσότητες αχρησιμοποίητων δεδομένων (βάσεις δεδομένων δαπανών, εκλογικές βάσεις δεδομένων, βάσεις δεδομένων δικαιούχων κοινωνικών προγραμμάτων, βάσεις δεδομένων εταιρειών που περιλαμβάνονται σε μαύρες λίστες και ηλεκτρονικά τιμολόγια) για να δημιουργήσει μια από τις μεγαλύτερες δεξαμενές δεδομένων στον κόσμο, η οποία περιλαμβάνει επί του παρόντος 27 σύνολα δεδομένων με περισσότερα από 250 εκατομμύρια σημεία δεδομένων και περισσότερα από 500 δισεκατομμύρια R$ σε δημόσια δαπάνη (περίπου 100 δισεκατομμύρια δολάρια ΗΠΑ) (World Bank, 2020).

Συνολικά, το σύστημα βασίζεται σε:

• Ανάλυση άνω των 500 δις. R$ δημοσίων συμβάσεων στη Βραζιλία από 12 Πολιτείες και σε Ομοσπονδιακό Επίπεδο.

• Ανάλυση άνω των 15 εκατομμυρίων ηλεκτρονικών τιμολογίων.

• Ανάλυση και γεωγραφική αναφορά σε περισσότερες από 750.000 εταιρείες και ένα σύνολο δεδομένων Δημόσιου Μητρώου που περιέχει λεπτομέρειες για 30 εκατομμύρια εταιρείες, όπως διεύθυνση κεντρικών γραφείων, συνεργάτες, δεδομένα σύστασης, οικονομικός τομέας.

• Ενσωμάτωση πάνω από 30.000 ειδήσεων σχετικά με τη διαφθορά.

• Στοιχεία 20 εκατομμυρίων δικαιούχων κοινωνικών προγραμμάτων.

• Στοιχεία 30.000 εταιρειών που περιλαμβάνονται σε μαύρες λίστες.

• Στοιχεία 20 εκατομμυρίων πολιτικών και 800.000 πολιτικών δωρεών.

Το αποτέλεσμα

Το σύστημα μέσω χρήσης τεχνολογιών και μεθόδων Τεχνητής Νοημοσύνης βελτιστοποιεί τη διαδικασία ανίχνευσης απάτης σε δημόσιες δαπάνες, εξοικονομώντας πολύτιμους πόρους (χρόνο και χρήμα) και αυξάνοντας την αποτελεσματικότητα των ελέγχων και των ερευνών. Το σύστημα έχει οδηγήσει στην αποκάλυψη πολυάριθμων περιπτώσεων υψηλού κινδύνου, όπως:

• Προσδιόρισε περισσότερες από 420 εταιρείες που κέρδισαν διαγωνισμούς έναντι εταιρειών που είχαν μεγάλη πιθανότητα να είναι εταιρείες-βιτρίνα, αποκαλύπτοντας πιθανή νοθεία προσφορών. Οι νικήτριες εταιρείες έχουν πάνω από 600 εκατομμύρια R$ σε δημόσιες συμβάσεις.

• Προσδιόρισε 857 εταιρείες που κέρδισαν διαγωνισμούς έναντι εταιρειών που μοιράζονταν μαζί τους τουλάχιστον έναν κοινό εταίρο. Αυτές οι εταιρείες έχουν εκτελέσει τουλάχιστον 800 εκατομμύρια R$ σε δημόσιες συμβάσεις.

• Προσδιόρισε 450 εταιρείες των οποίων οι εταίροι είναι δικαιούχοι του προγράμματος μεταφοράς μετρητών υπό όρους, *Bolsa Família*, το οποίο δείχνει ότι αυτά τα άτομα είναι δυνητικά «αχυράνθρωποι». Αυτές οι εταιρείες έχουν περισσότερα από 600 εκατομμύρια R$ σε δημόσιες συμβάσεις.

• Προσδιόρισε περισσότερες από 500 εταιρείες που ανήκουν σε δημόσιους υπαλλήλους που εργάζονται στην ίδια κρατική υπηρεσία που έχει εκτελέσει τη σύμβαση. Αυτές οι υποθέσεις ανέρχονται σε πάνω από 4,5 δισεκατομμύρια R$ σε δημόσιες συμβάσεις.

3.1.4 Τεχνητή Νοημοσύνη και Αντιμετώπιση Διαφθοράς - Το παράδειγμα της Κίνας

Το πρόβλημα

Η έκταση της λειτουργικής διαφθοράς μεταξύ των δημοσίων υπαλλήλων της Κίνας που ανέρχονται σε 50 εκατομμύρια και η αντιμετώπισή της από την κινεζική κυβέρνηση.

Η λύση

Η Κινεζική Ακαδημία Επιστημών και τα όργανα Εσωτερικού Ελέγχου του Κινεζικού Κομμουνιστικού Κόμματος ανέπτυξαν το *Zero Trust*, ένα σύστημα Τεχνητής Νοημοσύνης που περιλαμβάνει Επεξεργασία Φυσικής Γλώσσας (NLP), Μεγάλα Δεδομένα (Big Data), εξόρυξη δεδομένων (data-mining), ανίχνευση ανωμαλιών (anomaly detection), για την παρακολούθηση, την αξιολόγηση και τον έλεγχο της εργασίας και της ζωής των δημοσίων

υπαλλήλων. Το *Zero Trust* μπορεί να διασταυρώσει περισσότερες από 150 βάσεις δεδομένων σε συστήματα κεντρικής διοίκησης και τοπικής αυτοδιοίκησης. Το σύστημα ανιχνεύει μεταβιβάσεις ακινήτων, υποδομές, κατασκευές, αγορές γης και κατεδαφίσεις σπιτιών ενός ατόμου. Το *Zero Trust* εντοπίζει, επίσης, ασυνήθιστες αυξήσεις στις τραπεζικές αποταμιεύσεις ενός δημοσίου υπαλλήλου, αγορές νέων αυτοκινήτων και εάν ένας υπάλληλος υποβάλλει προσφορές για κρατικές συμβάσεις ή το κάνει υπό το όνομα μελών της οικογένειας ή φίλων του. Στη συνέχεια, το σύστημα υπολογίζει την πιθανότητα αυτές οι ενέργειες να είναι διεφθαρμένες και ειδοποιεί τους αξιωματούχους για πολύ πιθανές περιπτώσεις διαφθοράς (World Bank, 2020).

Το αποτέλεσμα

Το *Zero Trust* κυκλοφόρησε σε 30 κομητείες και πόλεις της Κίνας και εντόπισε 8.721 κυβερνητικούς αξιωματούχους που ήταν ύποπτοι για υπεξαίρεση, κατάχρηση εξουσίας, κατάχρηση κρατικών πόρων και νεποτισμό. Ορισμένες από αυτές τις περιπτώσεις οδήγησαν σε ποινή φυλάκισης ενώ στις περισσότερες περιπτώσεις οι υπόλογοι κινέζοι δημόσιοι υπάλληλοι είχαν τη δυνατότητα να διατηρήσουν τη θέση τους, αφού προηγουμένως είχαν λάβει μια προειδοποίηση ή μια μικρή τιμωρία (Chen, 2019). Ωστόσο, το μέλλον του *Zero Trust* είναι αβέβαιο. Παρά την πολιτική βούληση του Προέδρου της Κίνας Xi Jinping για διοικητική μεταρρύθμιση και εξυγίανση του δημόσιου τομέα μέσω προώθησης καινοτόμων τεχνολογιών όπως τα Μεγάλα Δεδομένα και η Τεχνητή Νοημοσύνη, το *Zero Trust* αντιμετωπίζει αντιδράσεις από δημόσιους λειτουργούς και μπορεί να παροπλιστεί (Chen, 2019).

3.1.5 Τεχνητή Νοημοσύνη και Εμπλοκή Πολιτών - Το παράδειγμα του Βελγίου

Το πρόβλημα

Οι κυβερνήσεις εργάζονται όλο και περισσότερο για την ανάπτυξη πολιτικών και υπηρεσιών με επίκεντρο τους πολίτες. Εξ ορισμού, αυτό απαιτεί εκτεταμένη ενασχόληση με τους πολίτες προκειμένου να κατανοήσουν τις απόψεις και τις ανάγκες τους. Οι πλατφόρμες ψηφιακής συμμετοχής είναι σημαντικά εργαλεία για την επίτευξη αυτού του στόχου και τη βελτίωση της κυβερνητικής ανταπόκρισης. Ωστόσο, η ανάλυση των υψηλών όγκων εισροών πολιτών που συλλέγονται σε αυτές τις πλατφόρμες είναι εξαιρετικά χρονοβόρα και περίπλοκη για τους κυβερνητικούς αξιωματούχους και τους εμποδίζει να αποκαλύψουν πολύτιμες εισροές. Η δημιουργία μιας πλατφόρμας ψηφιακής συμμετοχής, επομένως, δεν αρκεί. Αυτό που πρέπει να αντιμετωπιστεί είναι η διαδικασία ανάλυσης δεδομένων που

πρέπει να είναι πιο προσιτή ώστε να μπορούν οι δημόσιοι υπάλληλοι να αξιοποιούν τη συλλογική νοημοσύνη και να λαμβάνουν καλύτερα ενημερωμένες αποφάσεις (Berryhill et al., 2019).

Η λύση

Η CitizenLab[6] του Βελγίου είναι μια νεοφυής εταιρεία πολιτικής τεχνολογίας που στοχεύει να ενδυναμώσει τους δημοσίους υπαλλήλους και να τους παρέχει επαυξημένες διαδικασίες Μηχανικής Μάθησης που θα τους βοηθήσουν να αναλύουν τη συμβολή των πολιτών, να λαμβάνουν καλύτερες αποφάσεις και να συνεργάζονται πιο αποτελεσματικά εσωτερικά. Η CitizenLab έχει αναπτύξει μια δημόσια πλατφόρμα συμμετοχής που χρησιμοποιεί αλγόριθμους Μηχανικής Μάθησης για να βοηθά τους δημόσιους υπαλλήλους να επεξεργάζονται εύκολα χιλιάδες συνεισφορές πολιτών και να χρησιμοποιούν αυτές τις πληροφορίες αποτελεσματικά στη λήψη αποφάσεων. Οι πίνακες εργαλείων στην πλατφόρμα μπορούν να ταξινομούν ιδέες, να επισημαίνουν αναδυόμενα θέματα, να συνοψίζουν τάσεις και να συγκεντρώνουν παρόμοιες συνεισφορές ανά θέμα, δημογραφικό χαρακτηριστικό ή τοποθεσία (Berryhill et al., 2019).

Η πλατφόρμα του *CitizenLab* χρησιμοποιεί τεχνικές Επεξεργασίας Φυσικής Γλώσσας (NLP) και Μηχανικής Μάθησης (ML) για την αυτόματη ταξινόμηση και ανάλυση χιλιάδων συνεισφορών που συλλέγονται σε πλατφόρμες συμμετοχής πολιτών. Οι αλγόριθμοι προσδιορίζουν τα κύρια θέματα και ομαδοποιούν παρόμοιες ιδέες σε ομάδες, οι οποίες στη συνέχεια μπορούν να αναλυθούν ανά δημογραφικό χαρακτηριστικό ή γεωγραφική θέση.

Οι δημόσιοι υπάλληλοι που διαχειρίζονται αυτές τις πλατφόρμες συμμετοχής πολιτών μπορούν να έχουν πρόσβαση σε αυτές τις πληροφορίες με μια ματιά μέσω έξυπνων πινάκων ελέγχου σε πραγματικό χρόνο. Η λειτουργία «Μοντελοποίηση θεμάτων» τους επιτρέπει να εντοπίζουν εύκολα τις προτεραιότητες των πολιτών και να λαμβάνουν ανάλογες αποφάσεις. Η πλατφόρμα επιτρέπει στους δημοσίους υπαλλήλους να αναλύουν τα αποτελέσματα κατά δημογραφικές ομάδες και τοποθεσία, γεγονός που τους δίνει μια καλύτερη επισκόπηση της διακύμανσης των προτεραιοτήτων (Berryhill et al., 2019).

Σε ένα σχετικό παράδειγμα από τις αρχές του 2019, ένας αυξανόμενος αριθμός νέων Βέλγων διαμαρτυρόταν για την αδράνεια κατά της κλιματικής αλλαγής, ένα κίνημα που εξελίχθηκε

[6] www.citizenlab.co

στο Youth for Climate Belgium. Σε απάντηση, το CitizenLab δημιούργησε μια πλατφόρμα συμμετοχής για το θέμα με τίτλο Youth4Climate και κάλεσε τους χρήστες να υποβάλουν ιδέες για την αντιμετώπιση της κλιματικής αλλαγής.[7]

Σε διάστημα τριών μηνών, οι χρήστες υπέβαλαν 1.700 ιδέες, 2.600 σχόλια και 32.000 ψήφους για πρωτοβουλίες που ήθελαν να υποστηρίξουν. Το σύστημα Τεχνητής Νοημοσύνης ανέλυσε αυτά τα στοιχεία και παρουσίασε και συγκέντρωσε τις πιο σημαντικές και υποστηριζόμενες προτάσεις σε μια έκθεση με 16 συστάσεις πολιτικής την οποία προώθησε σε εκλεγμένους αξιωματούχους για ενέργεια (Berryhill et al., 2019).

Το αποτέλεσμα

Το αποτέλεσμα της χρήσης της πλατφόρμας *CitizenLab* είναι ένας πραγματικός διάλογος με τους πολίτες και όχι μια απρόσωπη πρωτοβουλία από πάνω προς τα κάτω. Αυτοματοποιώντας το χρονοβόρο έργο της ανάλυσης δεδομένων, η πλατφόρμα αφήνει χρόνο στις διοικήσεις για ουσιαστική συνεργασία με τους πολίτες. Παρέχει καλύτερη κατανόηση των αναγκών και των προτεραιοτήτων των πολιτών, κάτι που με τη σειρά του οδηγεί σε καλύτερα ενημερωμένες αποφάσεις. Από την οπτική γωνία των πολιτών, αυτή η ανοιχτή και διαφανής διαδικασία συμβάλλει στην ενίσχυση της εμπιστοσύνης και στην αύξηση της υποστήριξης για αποφάσεις πολιτικής (Berryhill et al., 2019).

3.1.6 Τεχνητή Νοημοσύνη και Εμπλοκή Πολιτών - Το παράδειγμα της Νιγηρίας

Το πρόβλημα

Η περιορισμένη ικανότητα των φορέων του δημόσιου τομέα να λαμβάνουν, να αναλύουν και να ανταποκρίνονται στα σχόλια των πολιτών για την αξιολόγηση πολιτικών, σχεδιασμού έργων και εφαρμογών τους, ώστε να ενισχύονται οι επίσημοι μηχανισμοί λογοδοσίας σε σχέση με τις δημόσιες δαπάνες.

Η λύση

Μια ομάδα της Παγκόσμιας Τράπεζας συνεργάστηκε με το Data Science Nigeria (DSN) στην Πολιτεία Έντο της Νιγηρίας για να πιλοτάρει μια λύση Τεχνητής Νοημοσύνης που αφορά στην αξιοποίηση της ανατροφοδότησης των πολιτών για την παρακολούθηση της προόδου του έργου Κρατικής Απασχόλησης και Δαπάνης για Αποτελέσματα (State Employment, and Expenditure for Results - SEEFOR) της Ευρωπαϊκής Ένωσης σε

[7] www.citizenlab.co/blog/civic-engagement/youth-for-climate-case-study.

δειγματοληπτικές τοποθεσίες. Αφορά το *DataCrowd,* μια εφαρμογή για κινητά βασισμένη στην Τεχνητή Νοημοσύνη. Το πιλοτικό πρόγραμμα πραγματοποιήθηκε σε διάστημα τεσσάρων εβδομάδων τον Μάιο του 2020. Το πεδίο εφαρμογής του κάλυψε 77 τοποθεσίες στην πολιτεία Έντο και συγκέντρωσε τα σχόλια των πολιτών για το έργο SEEFOR. Η λύση Τεχνητής Νοημοσύνης *DataCrowd* έχει πολλά χαρακτηριστικά.

Τα παρακάτω περιλαμβάνονται στον πιλότο του Έντο:

Σύννεφο ετικετών Τεχνητής Νοημοσύνης (AI-powered tag cloud): Το *DataCrowd* μπορεί να συνοψίσει κείμενο και προτάσεις, όπως τα σχόλια των πολιτών μέσω κινητών τηλεφώνων, εμφανίζοντας αμέσως τις λέξεις-κλειδιά και τη συνάφειά τους.

Γεωφράκτης με Τεχνητή Νοημοσύνη (AI-powered geofencing): Αυτή η δυνατότητα απορρίπτει αμέσως μια υποβολή σχολίου που έγινε εκτός μιας τοποθεσίας με γεωγραφική προστασία.

Ταξινομητής εικόνας με Τεχνητή Νοημοσύνη (AI-powered image classifier): Αυτή η δυνατότητα μπορεί να ταξινομήσει τα περιεχόμενα μιας εικόνας. Για παράδειγμα, εάν τραβήξετε μια φωτογραφία ενός ατόμου, το μοντέλο AI μπορεί να πει αν είναι άνδρας ή γυναίκα. Σε αντίθεση με το σύννεφο ετικετών και τις λειτουργίες του αναλυτή συναισθήματος, η δυνατότητα ταξινόμησης εικόνων εκπαιδεύεται προσαρμοσμένα με βάση τα δεδομένα εικόνας που συλλέγονται για ένα συγκεκριμένο έργο, το οποίο απαιτεί πάντα πολλές εικόνες για την εκπαίδευση. Στην περίπτωση του SEEFOR, η Ομάδα Έρευνας και Ανάπτυξης εργάζεται για την εκπαίδευση του μοντέλου ταξινομητή εικόνων για την ταξινόμηση ορισμένων εικόνων του SEEFOR, ειδικά στην κατηγορία των δημοσίων έργων. Αυτή η δυνατότητα είναι ιδιαίτερα χρήσιμη για ελέγχους διασφάλισης ποιότητας και όταν συλλέγονται πολλές εικόνες.

Αντιστοίχιση εικόνας με Τεχνητή Νοημοσύνη (AI-powered image matching): Αυτή η δυνατότητα θα επιτρέψει στο *DataCrowd* να αντιστοιχίσει αμέσως μια υπάρχουσα εικόνα με μια νέα εικόνα και να αναφέρει εάν είναι ίδια ή όχι. Αυτή η δυνατότητα είναι υπό ανάπτυξη. Αναμένεται να είναι χρήσιμο ως επαλήθευση και επικύρωση δεδομένων πρώτου επιπέδου όταν υποβάλλονται πολλές εικόνες από συλλέκτες δεδομένων.

Αναλυτής εξόρυξης γνώμης και συναισθημάτων με Τεχνητή Νοημοσύνη (AI-powered opinion mining and sentiment analyzer): Η λειτουργία του αναλυτή συναισθήματος μπορεί να μετρήσει τον παλμό συναισθήματος των δεδομένων κειμένου, όπως τα σχόλια των

πολιτών, και να κατηγοριοποιήσει τις προτάσεις σε αρνητικά, ουδέτερα και θετικά συναισθήματα. Αν και αυτή η δυνατότητα υπάρχει στο *DataCrowd*, δεν συμπεριλήφθηκε στο πιλοτικό πρόγραμμα. Είναι χρήσιμο για την κατανόηση του συναισθήματος που εκφράζεται σε όλα τα σχόλια των πολιτών και θα μπορούσε να χρησιμοποιηθεί δυνητικά για κλιμάκωση.

Το αποτέλεσμα

Στο πιλοτικό πρόγραμμα, οι αρχές του έργου SEEFOR μπόρεσαν να λάβουν σχόλια από τους πολίτες σχετικά με την πορεία εφαρμογής του έργου και να επιβεβαιώσουν διάφορες πτυχές της προόδου υλοποίησης, συμπεριλαμβανομένης της τοποθεσίας, της απόδοσης, της ποιότητας και της ολοκλήρωσης. Μετά τα αρχικά θετικά αποτελέσματα κατά την πιλοτική φάση, το *DataCrowd* σχεδιάζεται να επεκταθεί ώστε να καλύψει τρεις ακόμη πολιτείες της Νιγηρίας και περίπου 350 επιπλέον τοποθεσίες (World Bank, 2020).

3.1.7 Τεχνητή Νοημοσύνη και Τελωνεία - Το παράδειγμα των Ηνωμένων Πολιτειών

Το πρόβλημα

Οι παράνομες δραστηριότητες που πραγματοποιούνται στα βόρεια σύνορα των Ηνωμένων Πολιτειών με τον Καναδά, όπως το παράνομο εμπόριο, συμπεριλαμβανομένου του λαθρεμπορίου ναρκωτικών και της εμπορίας ανθρώπων και της εισόδου όπλων, τρομοκρατών και λαθρομεταναστών. Υπάρχουν 300 λιμάνια εισόδου στις Ηνωμένες Πολιτείες που πρέπει να ασφαλιστούν χωρίς να διακοπεί το εμπόριο και η διαμετακόμιση.

Η λύση

Η Υπηρεσία Τελωνείων και Προστασίας Συνόρων των ΗΠΑ (US Customs and Border Protection Agency) χρησιμοποιεί το Σύστημα Απομακρυσμένης Βιντεοεπιτήρησης των Βορείων Συνόρων (Northern Border Remote Video Surveillance System – NBRVSS) για να ενισχύσει τη συνοριακή επιτήρηση, διευκολύνοντας τις συνοριακές περιπολίες. Είναι ένα σύστημα Τεχνητής Νοημοσύνης που περιλαμβάνει Συνελικτικό Νευρωνικό Δίκτυο (Convolutional Neural Network), όραση υπολογιστή (Computer Vision), αντιστοίχιση προτύπων (pattern matching), ανίχνευση ανωμαλιών (anomaly detection), πρόβλεψη (prediction) και μπορεί να ανιχνεύει και να παρακολουθεί πλοία από μίλια μακριά και να ειδοποιεί τις αρχές όταν αναγνωρίζει ασυνήθιστες κινήσεις. Ξεκίνησε πριν από το 2016 και χρησιμοποιεί πολλούς πύργους επιτήρησης ύψους περίπου 50 μέτρων με κάμερες υψηλής ανάλυσης και ραντάρ, εξοπλισμένους με όραση υπολογιστή που εντοπίζουν ανωμαλίες στη

συμπεριφορά των σκαφών και επιτρέπουν σε πράκτορες στο έδαφος να αναχαιτίζουν πιθανές πηγές λαθρεμπορίου που εισέρχονται στις Ηνωμένες Πολιτείες από τα καναδικά σύνορα (World Bank, 2020).

Το αποτέλεσμα

Το Σύστημα Απομακρυσμένης Βιντεοεπιτήρησης των Βορείων Συνόρων με χρήση Τεχνητής Νοημοσύνης βοηθά στην αντιμετώπιση της παράνομης δραστηριότητας στα σύνορα ΗΠΑ-Καναδά. Η εγκατάσταση των πύργων επιτήρησης οδήγησε στην καλύτερη ανάπτυξη των συνοριακών πρακτόρων με επακόλουθη αύξηση της επίγνωσης της κατάστασης, της ακρίβειας και της ασφάλειας (World Bank, 2020).

3.1.8 Τεχνητή Νοημοσύνη και Υγεία - Η περίπτωση της Σιγκαπούρης

Το πρόβλημα

Οι υγειονομικές εγκαταστάσεις και το υγειονομικό προσωπικό βρίσκονται υπό τεράστια πίεση προκειμένου να ανταποκριθούν στον πρωτοφανή μεγάλο όγκο ασθενών με COVID-19. Οι γιατροί και οι εργαζόμενοι στον τομέα της υγείας πρώτης γραμμής χρειάζονται έγκαιρη ενημέρωση σχετικά με τα πιο πρόσφατα πρωτόκολλα υγείας, τους καταλόγους προσωπικού, τις επιχειρησιακές οδηγίες και τις δοσολογίες φαρμάκων για να διαχειριστούν αποτελεσματικά την πανδημία COVID-19.

Η λύση

Το *Bot MD* είναι μια εφαρμογή AI Chatbot για κινητά που αναπτύχθηκε το 2018 από το Νοσοκομείο Tan Tock Seng (TTSH) και την ομάδα IT του Υπουργείου Υγείας της Σιγκαπούρης. Το *Bot MD* λειτουργεί σαν «google» για νοσοκομειακές και κλινικές πληροφορίες σχετικά με το COVID-19 για γιατρούς και εργαζόμενους στον τομέα υγείας πρώτης γραμμής. Οι γιατροί και οι εργαζόμενοι στον τομέα της υγείας πρώτης γραμμής όπως και οι αξιωματούχοι του Υπουργείου Υγείας (MoH) μπορούν να πληκτρολογήσουν μια ερώτηση και η εφαρμογή μπορεί να παρέχει πληροφορίες σχετικά με τους καταλόγους προσωπικού, τα πρωτόκολλα υγείας, τη συνταγογράφηση φαρμάκων, τις οδηγίες για τις ασθένειες, τις επιχειρησιακές οδηγίες και τις τελευταίες εγκυκλίους του Υπουργείου Υγείας. Το σύστημα χρησιμοποιεί Τεχνητή Νοημοσύνη για να προβλέψει καταστάσεις πριν αυτές συμβούν και παρέχει πληροφορίες για τη λήψη αποφάσεων σχετικά με την κατανομή πόρων για την αντιμετώπιση των πιέσεων. Αυτοί οι πόροι θα μπορούσαν να περιλαμβάνουν ανθρώπινο δυναμικό, εξοπλισμό, προμήθειες, φάρμακα, νοσοκομειακά κρεβάτια, κέντρα πρόσληψης κ.λπ. (World Bank, 2020).

Το αποτέλεσμα

Περισσότεροι από 13.000 γιατροί σε 52 χώρες χρησιμοποιούν τώρα την εφαρμογή *Bot MD* λόγω της χρησιμότητας και της αποτελεσματικότητάς της σε καταστάσεις υγειονομικής κρίσης και πίεσης του υγειονομικού συστήματος.

3.1.9 Τεχνητή Νοημοσύνη και Δικαστικός Τομέας - Το παράδειγμα της Κίνας

Το πρόβλημα

Η ασυνεπής εφαρμογή του νόμου κατά τις δικαστικές αποφάσεις.

Η λύση

Το Ανώτατο Λαϊκό Δικαστήριο της Κίνας προωθεί την πολιτική των «Παρόμοιων Αποφάσεων σε Παρόμοιες Υποθέσεις» με την καθιέρωση της αρχής της αυτολογοδοσίας και της ανεξαρτησίας, βάσει της οποίας η τελική απόφαση εκδίδεται από τον ενδιαφερόμενο δικαστή χωρίς έγκριση ανώτερου επιπέδου, για να υποστηρίξει τη συνέπεια στις δικαστικές αποφάσεις.

Οι πολιτικές του Ανώτατου Λαϊκού Δικαστηρίου απαιτούν πλέον από τους δικαστές να ερευνούν παρόμοιες υποθέσεις και να αναφέρουν αυτές τις περιπτώσεις στις αποφάσεις για να διασφαλιστεί η συνέπεια. Για να υποστηρίξει αυτή την έρευνα, το κινεζικό δικαστικό σώμα εφαρμόζει πιλοτικά την Τεχνητή Νοημοσύνη σε ορισμένες επαρχίες για να βελτιώσει τη συνοχή. Στο πλαίσιο αυτής της εφαρμογής, όλες οι προηγούμενες κρίσεις ψηφιοποιήθηκαν και αποθηκεύτηκαν σε μια βάση δεδομένων. Στη συνέχεια, το Ανώτατο Λαϊκό Δικαστήριο ανέπτυξε δυνατότητες NLP AI, μέσω του *Similar Cases Push System*, για να αντιστοιχίσει το βασικό κείμενο που σχετίζεται με εκκρεμείς υποθέσεις που χρησιμοποιούν τη βάση δεδομένων. Το σύστημα παρουσιάζει σχετικές κρίσεις ενώπιον ενός δικαστή, χρησιμοποιώντας ένα προσυμπληρωμένο πρότυπο κρίσης που ο δικαστής εξετάζει και επεξεργάζεται.

Επίσης, ένα πιλοτικό πρόγραμμα Τεχνητής Νοημοσύνης καταγράφει τις δικαστικές διαδικασίες. Ορισμένα δικαστήρια στην Κίνα χρησιμοποιούν πλέον προϊόντα αναγνώρισης ομιλίας AI για να μεταφράζουν απευθείας τις ηχογραφήσεις της ακροαματικής διαδικασίας σε κείμενα σε πραγματικό χρόνο και να τις μετατρέπουν σε γραπτές δικαστικές διαδικασίες, χρησιμοποιώντας μεθόδους Speech-to-Text NLP (World Bank, 2020)..

Το αποτέλεσμα

Η αξιοποίηση και χρήση της Τεχνητής Νοημοσύνης στην κινεζική δικαιοσύνη περιόρισε το διοικητικό κόστος, τις καθυστερήσεις έκδοσης αποφάσεων λόγω υπερβολικού φόρτου εργασίας και τον κίνδυνο ασυνεπών κρίσεων μεταξύ διαφορετικών δικαιοδοσιών. Το *Similar Cases Push System* μείωσε τον χρόνο που απαιτείται για τη διαμόρφωση μιας γραπτής απόφασης και όλων των νομικών διαδικαστικών εγγράφων κατά 70 τοις εκατό και 90 τοις εκατό, αντίστοιχα (World Bank, 2020).

3.1.10 Τεχνητή Νοημοσύνης και Δικαστικός Τομέας – Το παράδειγμα του Ηνωμένου Βασιλείου

Το πρόβλημα

Προσπάθεια βελτιστοποίησης στην ανάλυση νομικών συμβάσεων, υποστήριξη πολιτών στη νομική γραφειοκρατία, σε νομικά θέματα και διεκδικήσεις.

Η λύση

Οι αυτοματοποιημένοι νομικοί βοηθοί και δικηγόροι Τεχνητής Νοημοσύνης σε αρκετές περιπτώσεις ξεπερνούν σε ακρίβεια και αποτελεσματικότητα αυτή των ειδικών του χώρου. Το AI bot *Case Cruncher* είχε καλύτερες επιδόσεις στην πρόβλεψη αποτελέσματος αξιώσεων σε σχέση με 100 δικηγόρους κορυφαίων δικηγορικών γραφείων του Λονδίνου σε διαγωνισμό που διεξήχθη από τον Οικονομικό Διαμεσολαβητή για την επαλήθευση της ακρίβειας των προβλέψεων. Από τις 775 συνολικά προβλέψεις, το AI *Case Cruncher* αναδείχθηκε στην κορυφή με ποσοστό ακρίβειας 86,6% σε σύγκριση με 66,3% μεταξύ των 100 δικηγόρων (World Bank, 2020).

Το *DoNotPay* που διαφημίζεται ως το πρώτο ρομπότ-δικηγόρος στον κόσμο, ξεκίνησε, βοηθώντας τους χρήστες του να αμφισβητήσουν τις κλήσεις για παράνομη στάθμευση. Σε ένα μήνα μετά την κυκλοφορία του, 160.000 από τις 250.000 κλήσεις ανατράπηκαν με το ποσοστό επιτυχίας να φτάνει στο 64% (King, n.d.). Το *DoNotPay* έχει επεκτείνει τις δραστηριότητές του με λειτουργίες που βοηθούν τους χρήστες του να λαμβάνουν επιστροφές χρημάτων από κρατήσεις αεροπορικών εισιτηρίων και κρατήσεων ξενοδοχείων αλλά και για παροχή νομικών υπηρεσιών για διάφορα κοινωνικά θέματα όπως αιτήσεις ασύλου, στέγασης, αυτοματοποιημένες υπηρεσίες σε χρήστες που επιθυμούν να λάβουν άδεια παραμονής ή πράσινη κάρτα κ.α. (World Bank, 2020).

Το αποτέλεσμα

AI bot *Case Cruncher* προσφέρει μεγάλες ευκαιρίες αφού μπορεί να λειτουργήσει σαν νομικό εργαλείο που ξεπερνά τις ανθρώπινες δυνατότητες σε επίπεδο προβλέψεων αποτελέσματος.

Το *DoNotPay* χρησιμοποιείται τώρα και στις 50 πολιτείες των ΗΠΑ παρέχοντας αυτοματοποιημένα δωρεάν νομικές συμβουλές στους χρήστες του.

3.1.11 Τεχνητή Νοημοσύνη και Δημόσιες Προμήθειες – Το παράδειγμα των Ηνωμένων Πολιτειών

Το πρόβλημα

Οι κεντρικές υπηρεσίες προμηθειών δεν μπορούν να εξασφαλίσουν συμμόρφωση με τις απαιτήσεις του άρθρου 508 του νόμου περί Αποκατάστασης γιατί οι αρμόδιες υπηρεσίες εφαρμόζουν τις δικές τους διαδικασίες για τη σύνταξη προσκλήσεων προμηθειών ΤΠΕ, παραβλέποντας τον κανόνα.

Το άρθρο 508 του νόμου περί Αποκατάστασης απαιτεί οι ΤΠΕ που προμηθεύεται η ομοσπονδιακή κυβέρνηση να είναι προσβάσιμες στα περίπου 60 εκατομμύρια άτομα με αναπηρίες στις ΗΠΑ. Όταν ο κανόνας παραβλέπεται στις προσκλήσεις, υπηρεσίες, εφαρμογές και προϊόντα ΤΠΕ δεν είναι προσβάσιμα για παράδειγμα σε ένα άτομο με τύφλωση.

Η λύση

Η κυβέρνηση των ΗΠΑ αξιοποιεί τη δύναμη της Τεχνητής Νοημοσύνης για να ενισχύσει τη συμμόρφωση στις προμήθειες. Το Γραφείο Κυβερνητικής Πολιτικής της Διοίκησης Γενικών Υπηρεσιών των ΗΠΑ ανέπτυξε το *Solicitation Review Tool (SRT),* ένα εργαλείο που χρησιμοποιεί Τεχνητή Νοημοσύνη για να βοηθήσει τις ομοσπονδιακές υπηρεσίες να αξιολογήσουν και να βελτιώσουν τη συνολική συμμόρφωση των προσκλήσεων έργων ΤΠΕ σε σχέση με τη προσβασιμότητα.

Η πλατφόρμα *SRT AI* χρησιμοποιεί αλγόριθμους Επεξεργασίας Φυσικής Γλώσσας (NLP), εξόρυξης κειμένου (text mining) και Μηχανικής Μάθησης (ML) για να σαρώσει και να ελέγξει αυτόματα εάν οι ομοσπονδιακές προσκλήσεις που δημοσιεύονται στο fbo.gov συμμορφώνονται με άρθρο 508 του νόμου περί Αποκατάστασης. Προειδοποιεί τα αρμόδια μέρη για μη συμμόρφωση και επισημαίνει την ανάγκη για διορθωτικές ενέργειες. Μέσω της ανεξάρτητης αξιολόγησης, οι προβλέψεις έχουν ακρίβεια 95 τοις εκατό (World Bank, 2020)..

Το αποτέλεσμα

Το *SRT AI* διευκολύνει την εύρεση εκείνων των προσκλήσεων που έχουν αναγνωριστεί ως συμμορφούμενες ή μη με αποτέλεσμα οι αρμόδιες ομοσπονδιακές υπηρεσίες να έχουν τις πληροφορίες για προσκλήσεις που δικαιολογούν πρόσθετες απαιτήσεις και αυτές που δικαιολογούν την τροποποίηση των προσκλήσεων πριν ληφθούν οι αποφάσεις επιλογής αναδόχου. Το *SRT AI* μετριάζει σημαντικά τους ανθρώπινους πόρους που απαιτούνται για τον εντοπισμό, τον έλεγχο και την επιβολή της συμμόρφωσης. Προωθεί μια ενιαία πολιτική κρατικών προμηθειών και έχει επεκταθεί για να προβλέψει εάν οι προσκλήσεις για προμήθειες ΤΠΕ συμμορφώνονται και με άλλες ομοσπονδιακές κανονιστικές απαιτήσεις, όπως η ασφάλεια στον κυβερνοχώρο ή η βιωσιμότητα (World Bank, 2020).

3.1.12 Τεχνητή Νοημοσύνη και Δημόσιες Προμήθειες – Το παράδειγμα της Νότιας Κορέας

Το πρόβλημα

Αθέμιτες πρακτικές νόθευσης προσφορών στις κρατικές προμήθειες ώστε να νικηθεί ο ανταγωνισμός.

Η λύση

Η Νότια Κορέα καταπολεμά τη νοθεία προσφορών μέσω της χρήσης Τεχνητής Νοημοσύνης. Η Κορεάτικη Επιτροπή Δίκαιου Εμπορίου αξιοποιεί το Σύστημα Ανάλυσης Δεικτών Προσφορών (*Bid Rigging Indicator Analysis System - BRIAS*), μια πλατφόρμα Τεχνητής Νοημοσύνης και ανάλυσης στοιχείων, για την καταπολέμηση πρακτικών διαφθοράς στις κρατικές προμήθειες. Η Κορεατική Επιτροπή Δίκαιου Εμπορίου συλλέγει και αναλύει ηλεκτρονικά πληροφορίες που σχετίζονται με τις προσφορές που υποβάλλονται σε μεγάλους δημόσιους οργανισμούς και επισημαίνει περιπτώσεις ύποπτων δραστηριοτήτων.

Συνολικά, 322 δημόσιοι οργανισμοί πρέπει να αναφέρουν τις προσφορές τους στη Κορεάτικη Επιτροπή Δίκαιου Εμπορίου. Τα κατασκευαστικά έργα άνω των 5 δισεκατομμυρίων ₩ και οι προσφορές για προμήθειες αγαθών και υπηρεσιών άνω των 500 εκατομμυρίων ₩ πρέπει να υποβάλλονται στο KFTC. Οι επηρεαζόμενοι δημόσιοι οργανισμοί πρέπει να αναφέρουν σχετικά δεδομένα στο *BRIAS* εντός 30 ημερών από την επιλογή ενός πλειοδότη. Οι οργανισμοί που χρησιμοποιούν εσωτερικά συστήματα υποβολής προσφορών μπορούν να μεταδίδουν δεδομένα προσφορών στη Κορεατική Επιτροπή Δίκαιου Εμπορίου σε πραγματικό χρόνο, χρησιμοποιώντας BRIAS API. Οι υπόλοιποι πρέπει να αναφέρουν πληροφορίες προσφοράς στην πύλη της Κορεάτικη Επιτροπή Δίκαιου Εμπορίου.

Η Κορεάτικη Επιτροπή Δίκαιου Εμπορίου σταθμίζει τα χαρακτηριστικά των δημόσιων διαγωνισμών και συμβάσεων προμήθειας σύμφωνα με έναν προκαθορισμένο τύπο και χρησιμοποιεί τα δεδομένα για να αναλύσει ποσοτικά την πιθανότητα διόρθωσης. Ένα αυτοματοποιημένο σύστημα υπολογίζει και αποδίδει βαθμολογία μεταξύ 0 και 100 στο αντικείμενο ή στη σύμβαση προμήθειας. Όσο υψηλότερη είναι η βαθμολογία, τόσο πιο πιθανό είναι να υπάρχει νοθεία στη σχετική προσφορά. Η Επιτροπή στέλνει προσφορές με σημαία σε εξωτερικά τμήματα για περαιτέρω έρευνα. Σε ένα παράδειγμα που αφορούσε 12 κατασκευαστικές εταιρείες για το μετρό της Σεούλ, η Επιτροπή εντόπισε νόθευση προσφορών και η κυβέρνηση επέβαλε το ποσό των 5,108 δισεκατομμυρίων ₩, ως πρόσθετη επιβάρυνση (World Bank, 2020).

Το αποτέλεσμα

Το Σύστημα Ανάλυσης Δεικτών Προσφορών (BRIAS) αύξησε σημαντικά την ταχύτητα και την αποτελεσματικότητα εντοπισμού υποθέσεων νοθείας προσφορών, διασφαλίζοντας τον θεμιτό ανταγωνισμό στις κρατικές προμήθειες.

3.1.13 Τεχνητή Νοημοσύνη και Φορολογική Συμμόρφωση – Το παράδειγμα της Αρμενίας

Το πρόβλημα

Φοροδιαφυγή επιχειρήσεων και ιδιωτών. Οι πρακτικές φοροδιαφυγής είναι δύσκολο να αποκαλυφθούν καθώς δεν μπορούν να διασταυρωθούν με τα δημοσιονομικά αρχεία που μπορεί να αποκαλύψουν συσχετίσεις με αποτέλεσμα τον εντοπισμό ανωμαλιών φορολογικής δήλωσης.

Η λύση

Οι φορολογικές αρχές της Αρμενίας χειρίστηκαν το πρόβλημα της φοροδιαφυγής χρησιμοποιώντας διάφορες τεχνικές:

Ενιαίο διοικητικό έγγραφο (ΕΔΕ): Η δημιουργία ενός ΕΔΕ για τους εισαγωγείς αγαθών είναι ένας τρόπος για να υπάρχει μια πληρέστερη εικόνα της κατάστασης. Τα Analytics εντοπίζουν εάν ένας φορολογούμενος εισάγει πάντα τα ίδια αγαθά από την ίδια χώρα και από τις ίδιες επιχειρήσεις. Επιπλέον, τα ηλεκτρονικά τιμολόγια βοηθούν στον εντοπισμό ομάδων φορολογουμένων που χρησιμοποιούν την ίδια αποθήκευση για εισαγόμενα αγαθά. Η φορολογική διοίκηση διερευνά τις ανωμαλίες.

Διασταύρωση πωλήσεων και τιμολογίων: Η διασταύρωση δεδομένων πωλήσεων και τιμολογίων παρέχει σημαντικές πληροφορίες για τα έσοδα διαφόρων πωλητών. Η φορολογική αρχή της Αρμενίας συλλέγει δεδομένα από τη βάση δεδομένων καταχώρισης — φορολογικές μηχανές νέας γενιάς ή ταμειακές μηχανές συνδεδεμένες στους διακομιστές του οργανισμού — και βάσεις δεδομένων τιμολογίων. Η βάση δεδομένων τιμολογίων ανιχνεύει πότε μια ποικιλία οντοτήτων πωλούν αγαθά από την ίδια αποθήκη. Τα δεδομένα των φορολογικών μηχανών νέας γενιάς αποκαλύπτουν πότε μια ομάδα φορολογουμένων χρησιμοποιεί μια ποικιλία φορολογικών μηχανών στην ίδια τοποθεσία. Η βάση δεδομένων εγγραφής αποκαλύπτει πότε διαφορετικές επιχειρήσεις έχουν τους ίδιους ιδρυτές. Τέτοιες ύποπτες ανωμαλίες υπόκεινται σε λεπτομερείς ελέγχους, οι οποίοι μπορεί να μην αποτελούν απαραίτητα φορολογική απάτη, αλλά χρειάζονται βαθύτερο έλεγχο.

Οι υπάλληλοι των φορολογουμένων: Η Τεχνητή Νοημοσύνη και τα analytics μπορούν να ανιχνεύσουν ύποπτες περιπτώσεις στις οποίες διαφορετικοί εργοδότες δηλώνουν πανομοιότυπες ομάδες εργαζομένων στις δηλώσεις φόρου εισοδήματος ή όταν μια επιχείρηση που έχει κλείσει και άνοιξε ξανά προσλαμβάνει τους ίδιους υπαλλήλους. Οι πληροφορίες των εργαζομένων λαμβάνονται συνδέοντας έναν αριθμό κοινωνικής ασφάλισης και την ταυτότητα του προσώπου με μια βάση δεδομένων χρησιμοποιώντας μια υποδομή Big Data.

AI και αναλυτικά στοιχεία για τις πωλήσεις από τις ταμειακές μηχανές: Το Κέντρο Παρακολούθησης αξιοποιεί τις πληροφορίες που λαμβάνονται από τις φορολογικές μηχανές νέας γενιάς. Ορισμένα πρότυπα απαιτούν περαιτέρω έλεγχο. Για παράδειγμα, εάν ένα δημοσιονομικό μηχάνημα δεν λειτουργεί όλη την ημέρα και ο φορολογούμενος εκτυπώσει 100 ή 200 αποδείξεις εντός μιας ή δύο ωρών, αυτό επισημαίνει ένα πλαστό φορολογικό ποσό με πλαστές αποδείξεις χωρίς πραγματικές πωλήσεις. Ορισμένοι φορολογούμενοι εκτυπώνουν στο τέλος της ημέρας μια μόνο απόδειξη με μη ρεαλιστικό ποσό. Όλες αυτές οι περιπτώσεις είναι υπό έλεγχο αφού το Παρατηρητήριο στέλνει αυτόματα ειδοποιήσεις και απαιτεί εξηγήσεις. Εάν δεν δοθεί εύλογη εξήγηση, η υπόθεση περνά σε έλεγχο.

Σύγκριση δεδομένων από παρόχους υπηρεσιών κοινής ωφέλειας: Τα δεδομένα από το νερό, την ηλεκτρική ενέργεια και το φυσικό αέριο, για παράδειγμα, αποκαλύπτουν τα έξοδα της επιχείρησης, τα οποία καταδεικνύουν μια λογική συσχέτιση με το συνολικό ποσό των αναφερόμενων πωλήσεων για μια συγκεκριμένη επιχειρηματική δραστηριότητα. Και πάλι,

αυτή η διασταύρωση και ο συσχετισμός αποκαλύπτει πολύτιμες πληροφορίες (World Bank, 2020).

Το αποτέλεσμα

Η φορολογική διοίκηση μείωσε τον αριθμό των υποθέσεων ελέγχου κατά περίπου 2,5 φορές τα τελευταία χρόνια. Η αποτελεσματικότητα, μετρούμενη με το μέσο ποσό πρόσθετου φόρου ανά έλεγχο, αυξανόταν συνεχώς τα τελευταία χρόνια. Η Αρμενία πέτυχε σημαντική εξοικονόμηση διοικητικού κόστους λόγω της ταχείας μείωσης του αριθμού των τοπικών φορολογικών υπηρεσιών (από 52 το 2009 σε μόνο δύο γραφεία το 2017) (World Bank, 2020).

3.1.14 Τεχνητή Νοημοσύνης και Φορολογική Συμμόρφωση – Το παράδειγμα των Ηνωμένων Πολιτειών

Το πρόβλημα

Η φοροδιαφυγή, η φορολογική απάτη, η κλοπή ταυτότητας φορολογούμενου και η μη φορολογική συμμόρφωση.

Η λύση

Το 2011, η Υπηρεσία Εσωτερικών Εσόδων δημιούργησε το Γραφείο Αναλύσεων Συμμόρφωσης για την ανάπτυξη αναλυτικών προγραμμάτων που θα μπορούσαν να εντοπίζουν πιθανές απάτες επιστροφής χρημάτων, κλοπής ταυτότητας φορολογούμενου και θα χειρίζονται αποτελεσματικά ζητήματα μη συμμόρφωσης. Το Γραφείο Αναλύσεων Συμμόρφωσης αξιοποιεί ένα προηγμένο πρόγραμμα ανάλυσης που βασίζεται στη χρήση Big Data και προγνωστικών αλγορίθμων για τη μείωση της φορολογικής απάτης. Το 2016 το Γραφείο Αναλύσεων Συμμόρφωσης συγχωνεύτηκε με το Γραφείο Έρευνας, Ανάλυσης και Στατιστικών για να δημιουργήσουν Τμήμα Έρευνα Εφαρμοσμένων Αναλύσεων και Στατιστικών, το οποίο επεξεργάζεται δεδομένα μέσω καινοτόμου και στρατηγικής έρευνας, ανάλυσης, στατιστικών και τεχνολογικών υπηρεσιών σε συνεργασία με εσωτερικούς και εξωτερικούς ενδιαφερόμενους φορείς και εξάγει αξία, αξιοποιώντας τεράστιες ποσότητες ιδιόκτητων δεδομένων που είναι αποθηκευμένα στους υπολογιστές παλαιού τύπου της Υπηρεσίας Εσωτερικών Εσόδων (World Bank, 2020).

Η Υπηρεσία Εσωτερικών Εσόδων χρησιμοποιεί την πλατφόρμα *Palantir Gotham* για την εκτέλεση της υπηρεσίας Lead and Case Analytics (LCA). Ειδικοί πράκτορες και αναλυτές ερευνών του Τμήματος Έρευνας Οικονομικού Εγκλήματος χρησιμοποιούν την LCA για τη

δημιουργία δυνητικών πελατών, τον εντοπισμό σχημάτων δεδομένων, την αποκάλυψη φορολογικής απάτης και τη διεξαγωγή ερευνητικών δραστηριοτήτων για ξέπλυμα χρήματος και κατάσχεση. Το 2016, η εφαρμογή εξόρυξης δεδομένων Return Review Program (RRP) δημιούργησε περισσότερους από 693.000 δυνητικούς πελάτες κλοπής φορολογικής ταυτότητας, με ποσοστό ακρίβειας 62% και περισσότερους από 103.000 άλλους δυνητικούς πελάτες απάτης με ποσοστό ακρίβειας 49% (World Bank, 2020).

Το αποτέλεσμα

Η Υπηρεσία Εσωτερικών Εσόδων χρησιμοποιεί την Τεχνητή Νοημοσύνη για να επιλέξει τις καλύτερες υποθέσεις για φορολογικό έλεγχο, βελτιώνοντας την αποτελεσματικότητα της Υπηρεσίας και συμβάλλοντας στη φορολογική συμμόρφωση και στη δημοσιονομική βιωσιμότητα (World Bank, 2020).

4. Λήψη αποφάσεων με χρήση αλγορίθμων

Η αλγοριθμική λήψη αποφάσεων αναφέρεται στη χρήση αλγορίθμων για την εκτέλεση ή την ενημέρωση αποφάσεων και η οποία μπορεί να ποικίλλει σε μεγάλο βαθμό ως προς την απλότητα και την πολυπλοκότητά της (Yeung, 2017).

4.1 Αλγόριθμοι

Με την ευρεία έννοια, οι αλγόριθμοι είναι *«κωδικοποιημένες διαδικασίες για τη μετατροπή των δεδομένων εισόδου σε επιθυμητό αποτέλεσμα με βάση συγκεκριμένους υπολογισμούς»* (Gillespie, 2013). Αν και οι αλγόριθμοι δε χρειάζεται να εφαρμοστούν σε λογισμικό, οι υπολογιστές εφαρμόζουν αλγορίθμους – μαθηματικές οδηγίες ή κανόνες που εφαρμόζονται σε δεδομένα – που τους βοηθούν να υπολογίσουν τις απαντήσεις σε ένα πρόβλημα. Οι υπολογιστές είναι βασικά μηχανές αλγορίθμων, σχεδιασμένες να αποθηκεύουν και να διαβάζουν δεδομένα, να εφαρμόζουν μαθηματικές διαδικασίες στα δεδομένα με ελεγχόμενο τρόπο και να προσφέρουν νέες πληροφορίες ως έξοδο. Ακόμη και όταν περιορίζεται σε λογισμικό, ο όρος «αλγόριθμος» μπορεί να κατανοηθεί ποικιλοτρόπως. Οι μηχανικοί λογισμικού είναι πιθανό να υιοθετήσουν μια τεχνική κατανόηση των αλγορίθμων, αναφερόμενοι στη λογική σειρά βημάτων για την οργάνωση και δράση σε ένα σύνολο δεδομένων για να επιτευχθεί γρήγορα ένα επιθυμητό αποτέλεσμα που συμβαίνει μετά τη δημιουργία ενός «μοντέλου», δηλαδή την επισημοποίηση του προβλήματος και του στόχου με υπολογιστικούς όρους (Gillespie, 2013. Dourish, 2016). Αλλά οι κοινωνικοί επιστήμονες, συνήθως, χρησιμοποιούν τον όρο, ως επίθετο, για να περιγράψουν το κοινωνικοτεχνικό σύνολο που περιλαμβάνει, όχι μόνο αλγόριθμους, αλλά και τα υπολογιστικά δίκτυα στα οποία λειτουργούν, τους ανθρώπους που τα σχεδιάζουν και τα λειτουργούν, τα δεδομένα και τους χρήστες στα οποία ενεργούν και τους οργανισμούς που παρέχουν αυτές τις υπηρεσίες, όλα συνδεδεμένα με μια ευρύτερη κοινωνική προσπάθεια, αποτελώντας μέρος μιας οικογένειας έγκυρων συστημάτων παραγωγής γνώσης. Αντίστοιχα, ο Gillespie αναφέρει ότι, όταν περιγράφουμε κάτι ως «αλγοριθμικό», το ενδιαφέρον μας είναι η εισαγωγή διαδικασίας που παράγεται ή σχετίζεται με ένα κοινωνικο-τεχνικό πληροφοριακό σύστημα που προορίζεται από τους σχεδιαστές του να είναι λειτουργικά και ιδεολογικά αφοσιωμένο στην υπολογιστική παραγωγή γνώσης (Gillespie, 2014).

Αν και οι υπολογιστικοί αλγόριθμοι περιλαμβάνουν αυτούς που κωδικοποιούν απλές μαθηματικές συναρτήσεις, ο ενθουσιασμός που περιβάλλει τα Μεγάλα Δεδομένα αποδίδεται σε μεγάλο βαθμό σε εξελιγμένους αλγορίθμους Μηχανικής Μάθησης, που τροφοδοτούνται

από ογκώδη και συχνά αδόμητα σύνολα δεδομένων, που λειτουργούν υπολογιστικά και απομακρύνονται από τις παραδοσιακές τεχνικές στατιστικής μοντελοποίησης (Dourish, 2016). Η εν λόγω μεθοδολογική προσέγγιση μερικές φορές περιγράφεται και ως «αφήνοντας τα δεδομένα να μιλήσουν» (Mayer-Schonberger et al., 2013).

Ωστόσο, με τη συνεχιζόμενη επανάσταση των δεδομένων και τη μετάβαση προς το λεγόμενο «Διαδίκτυο των Πραγμάτων», η εξάρτηση από τους αλγορίθμους αυξάνεται και η τάση για την αλγοριθμική λήψη αποφάσεων μπορεί μόνο να αναπτυχθεί (Kitchin, 2014a. Rifkin, 2014). Οι αλγόριθμοι έχουν φτάσει σε τέτοιο επίπεδο ώστε να επηρεάζουν όλο και περισσότερο τα άτομα, τους οργανισμούς και την κοινωνία (Pasquale, 2015). Ζούμε σε μια εποχή όπου οι αλγόριθμοι μας μετατρέπουν σε «αντικείμενα κατάταξης και βαθμολόγησής», που καθορίζουν τους κινδύνους της πιστοληπτικής μας ικανότητας, αν θα βρούμε δουλειά και ακόμη και αν θα μας χορηγηθεί εγγύηση ή αν θα θεωρηθούμε «κίνδυνος» για την κοινωνία και θα μας δοθεί μια μακρά τιμωρία (Citron et al., 2014. Pasquale, 2015. Buranyi, 2018). Καθοδηγούν προσωπικές συστάσεις που επηρεάζουν τι αγοράζουμε, τις ταινίες που βλέπουμε και τα δίκτυα κοινωνικής δικτύωσης που σχηματίζουμε. Άλλοι αλγόριθμοι επηρεάζουν τις αποφάσεις για τη ζωή μας, όπως αν θα λάβουμε δάνειο αυτοκινήτου, αν θα πάρουμε μέρος σε συνέντευξη για μια δουλειά και ακόμη αν θα λάβουμε ποινή φυλάκισης αντί για αναστολή. Ένα άρθρο της Wall Street Journal αναφέρει ότι ένας αλγόριθμος μπορεί να είναι ο επόμενος εργοδότης μας και θα προγραμματίζει διακοπές, θα καθορίζει ομάδες εργασίας και θα αναθέτει καθημερινές εργασίες (Schechner, 2017).

4.2 Αλγοκρατία

Καθώς η διαχείριση του ανθρώπου γίνεται όλο και περισσότερο μέσω πληροφοριών, λογισμικού και αλγοριθμικών αποφάσεων, ο κανόνας από τους δημόσιους αξιωματούχους θα αντικατασταθεί από αλγόριθμους με κώδικα υπολογιστή. Υπό αυτήν την έννοια το συγκεκριμένο σύστημα διακυβέρνησης μπορεί να οριστεί ως «αλγοκρατία», το οποίο είναι οργανωμένο και δομημένο βάσει αλγορίθμων προγραμματισμένων από υπολογιστή και μέσω των οποίων λαμβάνονται αποφάσεις, οι οποίες δομούν, υποκινούν, επηρεάζουν, περιορίζουν και ελέγχουν τη συμπεριφορά των ανθρώπων που απευθύνονται (Danaher, 2016). Ως σύστημα διακυβέρνησης, η «αλγοκρατία» αποτυπώνει την εξουσία που δίνεται στις αλγοριθμικά κωδικοποιημένες αρχιτεκτονικές στην σύγχρονη ζωή. Κάθε φορά που μας αρνούνται ένα δάνειο ύστερα από μια πιστοληπτική αλγοριθμική αξιολόγηση, όποτε μας

λένε με ποιον τρόπο να οδηγήσουμε με αλγόριθμο δρομολόγησης, οποτεδήποτε μας ζητείται από μια εφαρμογή υγείας και φυσικής κατάστασης να ασκηθούμε με ένα συγκεκριμένο τρόπο ή να φάμε ένα συγκεκριμένο φαγητό, ζούμε σε ένα αλγοκρατικό σύστημα (Danaher, 2018).

Ο όρος «αλγοκρατία» επινοήθηκε από τον κοινωνιολόγο Aneesh. Το 2009 ο Aneesh δημοσίευσε τα ευρήματα της εθνογραφικής του μελέτης για τους Ινδούς εργαζομένους που παρείχαν υπηρεσίες πληροφορικής και επικοινωνιών σε εταιρείες των ΗΠΑ, για να διερευνήσει πώς οργανώνεται η εργασιακή πρακτική του «off-shoring». Ο Aneesh χρησιμοποίησε την ιδέα για να καταλάβει πώς οι εργαζόμενοι συμμετείχαν σε μια παγκοσμιοποιημένη οικονομία. Προσδιόρισε τα χρονοδιαγράμματα προγραμματισμού λογισμικού ως κρίσιμα για την οργάνωση της κατανεμημένης σε όλο τον κόσμο εργασίας μέσω διακομιστών δεδομένων, προσδιορίζοντας ένα σύστημα διακυβέρνησης, το οποίο ονόμασε «αλγοκρατία» με βάση τον «κανόνα του αλγορίθμου» ή τον «κανόνα του κώδικα», που διαφέρει τόσο από τα γραφειοκρατικά, όσο και από αυτά των αγορών. Για τον Aneesh (2009) *οι αγορές* είναι ένα σύστημα στο οποίο *οι τιμές* δομούν και περιορίζουν τους τρόπους με τους οποίους ενεργούν οι άνθρωποι. Η *γραφειοκρατία* είναι ένα σύστημα στο οποίο *οι νόμοι και οι κανονισμοί* δομούν και περιορίζουν τους τρόπους με τους οποίους ενεργούν οι άνθρωποι και *η αλγοκρατία* είναι ένα σύστημα στο οποίο *οι αλγόριθμοι* δομούν και περιορίζουν τους τρόπους με τους οποίους ενεργούν οι άνθρωποι (Danaher, 2016).

4.2.1 Η απειλή της Αλγοκρατίας

Τα όρια μεταξύ τέτοιων συστημάτων δεν είναι ακριβή, με αποτέλεσμα συχνά να ενσωματώνονται και να επικαλύπτονται το ένα με το άλλο, όπως τα αλγοκρατικά συστήματα λήψης αποφάσεων μπορούν να ενσωματωθούν σε προϋπάρχοντα γραφειοκρατικά συστήματα λήψης αποφάσεων. Ωστόσο, ως δημόσιες διαδικασίες λήψης αποφάσεων που εκδίδουν υποχρεωτικούς κανόνες και κρίσεις, είναι ευρέως αποδεκτό ότι τέτοιες διαδικασίες πρέπει να είναι ηθικά και πολιτικά νόμιμες (Peter, 2014). Η νομιμότητα είναι η ιδιότητα που πρέπει να διαθέτουν οι δημόσιες διαδικασίες λήψης αποφάσεων, εάν θέλουν να ασκήσουν σωστά την απαιτούμενη εξουσία στις ζωές των ανθρώπων. Ωστόσο, η νομιμότητα απειλείται από την αλγοκρατία. Σύμφωνα με τον Danaher (2016), η απειλή της αλγοκρατίας υφίσταται και σχετίζεται, κυρίως, με την αδιαφάνεια ορισμένων αλγοκρατικών συστημάτων διακυβέρνησης, τα οποία λειτουργούν με τρόπους που είναι απροσπέλαστοι ή αδιαφανείς για την ανθρώπινη λογική και κατανόηση. Οι τεχνολογίες που καθιστούν δυνατή την αλγοκρατία γίνονται λιγότερο αισθητές και πιο πανταχού παρούσες. Μπορεί αρχικά να

ευνοούμε τα αλγοκρατικά συστήματα διακυβέρνησης για κατάλληλους εργαλειακούς λόγους, εντυπωσιασμένοι από τη μεγαλύτερη ταχύτητα, ακρίβεια και διορατικότητα (σε σύγκριση με παρόμοια ανθρώπινα συστήματα) και μπορεί να επιθυμούμε να εκμεταλλευτούμε τα εντυπωσιακά τους αποτελέσματα. Αλλά, ευνοώντας τέτοιου είδους συστήματα, μπορεί να καταλήξουμε με συστήματα που είναι όλο και πιο αδιαφανή και ανεξερεύνητα. Ο Μορόζοφ (2013) αναφέρει εύστοχα για το θέμα:

«Χάρη στα smartphones ή στο Google Glass, μπορούν τώρα να μας κάνουν «ping» κάθε φορά που πρόκειται να κάνουμε κάτι ανόητο, ανθυγιεινό ή κακόβουλο. Δε χρειάζεται απαραίτητα να γνωρίζουμε γιατί η ενέργεια θα ήταν λάθος: οι αλγόριθμοι του συστήματος κάνουν τον ηθικό λογισμό. Οι πολίτες αναλαμβάνουν το ρόλο των μηχανών πληροφόρησης που τροφοδοτούν το τεχνογραφειοκρατικό σύμπλεγμα με τα δεδομένα τους. Και γιατί να μην το κάνουμε, εάν μας υποσχεθούν πιο λεπτή μέση, καθαρότερο αέρα ή μεγαλύτερη και ασφαλέστερη ζωή σε αντάλλαγμα;»

Στη συνέχεια παγιδευόμαστε, όπως το θέτει ο Morozov (2013), σε έναν ιστό από «αόρατα συρματοπλέγματα». Είμαστε πεπεισμένοι ότι τα συστήματα αλγοριθμικού ελέγχου ενισχύουν την αυτονομία μας, αυξάνουν την υγεία και την ευημερία μας και βελτιώνουν τα κοινωνικά αποτελέσματα, αλλά δεν έχουμε ξεκάθαρη αίσθηση του πώς ακριβώς το καταφέρνουν. Το αποτέλεσμα είναι κοινωνικοί χώροι αδιαφανείς για την ανθρώπινη λογική (Pentland, 2014).

Μια απλή απεικόνιση της ανωτέρω διαπίστωσης του Morozov αποτελεί το σύστημα αποθήκευσης των μεγάλων αποθηκών της Amazon, το οποίο βασίζεται σε έναν «chaotic» χαοτικό αλγόριθμο αποθήκευσης (Greenfield, 2012. Bumbulsky, 2013). Από τη μια, οι άνθρωποι, αιώνες τώρα, διαθέτουν αποθήκες και παρόμοιες εγκαταστάσεις αποθήκευσης, ακολουθώντας τους δικούς τους «αλγόριθμους». Για παράδειγμα, αποθηκεύουν, ομαδοποιώντας παρόμοια αντικείμενα (π.χ. βιβλία, DVD, έπιπλα σπιτιού, συσκευές κ.λ.π.) και στη συνέχεια, υποδιαιρώντας αυτές τις ομάδες αντικειμένων σε διάφορες κατηγορίες (π.χ. αλφαβητική σειρά, συγγραφέας, τύπος επίπλου ή συσκευής). Το σκεπτικό πίσω από αυτά τα συστήματα αποθήκευσης έχει νόημα και είναι σαφώς κατανοητό από τους απλούς ανθρώπους. Επιπλέον, η διαδικασία αναγνώρισης αντικειμένων και διεκπεραίωσης παραγγελιών είναι αυτή που οι άνθρωποι μπορούν να κατανοήσουν πλήρως και να συμμετάσχουν σε αυτήν. Από την άλλη, το σύστημα του χαοτικού αλγορίθμου αποθήκευσης είναι μάλλον διαφορετικό. Το σύστημα λειτουργεί, επισημαίνοντας κάθε είδος που

εισέρχεται στην αποθήκη με έναν γραμμωτό κώδικα και στη συνέχεια, εκχωρώντας το σε μια τοποθεσία στην αποθήκη με βάση τον διαθέσιμο χώρο στο ράφι. Αυτό γίνεται με αλγόριθμο. Το αποτέλεσμα είναι ένα σύστημα που είναι προφανώς πολύ πιο αποτελεσματικό (λιγότερη σπατάλη προϊόντος, ταχύτερος κύκλος εργασιών αποθεμάτων) και στο οποίο πολύ διαφορετικά προϊόντα βρίσκονται δίπλα-δίπλα στα ράφια. Όταν έρθει η ώρα να εκπληρώσει μια παραγγελία, ένας εργαζόμενος πρέπει να βασιστεί σε έναν αλγόριθμο για να σχεδιάσει μια πορεία μέσα στην αποθήκη ώστε να παραλάβει τα διάφορα είδη. Αυτό δημιουργεί ένα πολύ ενδιαφέρον φυσικό περιβάλλον εργασίας. Είναι ένα μέρος στο οποίο οι άνθρωποι βρίσκονται «στο βρόχο», αλλά του οποίου η οργάνωση καθορίζεται από τους αλγόριθμους και του οποίου ο φυσικός χώρος δεν μπορεί να πλοηγηθεί από τους ανθρώπους χωρίς αλγοριθμική βοήθεια. Υπάρχει, συνεπώς, σεβασμός στη γνωσιολογική υπεροχή του συγκεκριμένου αλγοκρατικού συστήματος. Ωστόσο, το χαοτικό σύστημα αποθήκευσης δεν είναι εντελώς αδιαφανές για την ανθρώπινη λογική. Έχει έναν υποκείμενο σκοπό που μπορεί να ακολουθηθεί από τα ανθρώπινα όντα, δηλαδή, η ανάθεση βάσει του χώρου στο ράφι οδηγεί σε μεγαλύτερη αποτελεσματικότητα. Αυτός ο σκοπός είναι ελκυστικός, ακόμη και για τους ανθρώπους που τον δημιούργησαν. Αλλά, αν και λειτουργεί αποτελεσματικά το σύστημα αποθήκευσης, η πραγματική μηχανική του αλγορίθμου είναι πολύ περίπλοκη για να την ακολουθήσει κάποιος άνθρωπος. Ένας άνθρωπος δε θα μπορούσε να παρακολουθήσει τους γραμμωτούς κώδικες, ούτε τον διαθέσιμο χώρο στο ράφι. Πρέπει να αναθέσουν όλη αυτή την κατανόηση σε υπολογιστή (Danaher, 2016). Το αποτέλεσμα είναι οι άνθρωποι να αρχίσουν να φυλακίζονται στα «αόρατα συρματοπλέγματα», όπως ήδη έχει αναφέρει ο Morozov (2013).

Αυτό που συμβαίνει στις αποθήκες της Amazon μπορεί να συμβεί σε πολύ μεγαλύτερη και πιο επιθετική κλίμακα στις δημόσιες διαδικασίες λήψης αποφάσεων. Θα μπορούσαμε να εισαγάγουμε και να μεταβούμε σε ολοένα και περισσότερα αλγοκρατικά συστήματα, ξεκινώντας από αυτά που είναι σχετικά εύκολο να ακολουθηθούν, αλλά που μεταμορφώνονται σε συστήματα πολύ πιο περίπλοκα και έξω από τα ανώτερα όρια της ανθρώπινης λογικής. Εδώ, θα είναι πολύ πιο δύσκολο να υποχωρήσουμε στην ανάγκη για συμμετοχή και κατανόηση, επειδή το πεδίο για γνήσια ανθρώπινη συμμετοχή θα είναι πολύ πιο περιορισμένο και αυτό γιατί οι αλγόριθμοι θα οργανώνουν και θα χειρίζονται τεράστιες ροές δεδομένων και θα υιοθετούν ένα ολοένα και περισσότερο πολύπλοκο οικοσύστημα άλλων αλγορίθμων (Danaher, 2016).

Εάν, για παράδειγμα, είχαμε έναν περίπλοκο αλγόριθμο παρακολούθησης της φοροδιαφυγής, άραγε δε θα ήταν αρκετό για τους ανθρώπους να γνωρίζουν, απλώς, ότι το σύστημα λειτουργεί εντοπίζοντας εκείνους που είναι πιο πιθανό να είναι φοροφυγάδες, όπως ακριβώς οι εργαζόμενοι της Amazon γνωρίζουν, κατά προσέγγιση, πώς λειτουργεί το σύστημα και ποιος είναι ο σκοπός του; Χρειάζεται πραγματικά να γνωρίζουν με ακρίβεια ποιοι παράγοντες ενεργοποιούν το σύστημα; Με άλλα λόγια, δεν αρκεί μια αδρομερής περιγραφή της ορθολογικής βάσης για το σύστημα; Η απάντηση είναι όχι, αυτό δεν πρέπει να είναι αρκετό. Εάν θέλουμε να σεβαστούμε την ηθική ισότητα των μεμονωμένων πολιτών, δεν μπορούμε να ασκήσουμε νόμιμα εξουσία πάνω τους με τέτοιο τρόπο. Δεν αρκεί να γνωρίζουν απλώς, ότι το σύστημα είναι πιο πιθανό να φτάσει σε προτιμώμενα αποτελέσματα, αλλά πρέπει να είναι σε θέση να ελέγχουν και να εμπλέκονται κριτικά με εκείνους ακριβώς τους παράγοντες που επιτρέπουν στο σύστημα να το κάνει (Danaher, 2016). Τα αλγοκρατικά συστήματα δεν πρέπει να λειτουργούν με τεχνικές και μεθόδους «μαύρου κουτιού (black box)», ώστε να είναι πιο κατανοητά στους δημιουργούς τους και υπόλογα στους χρήστες τους (Knightarchive, 2017).

Αυτό δεν σημαίνει ότι απαιτείται μια εξαιρετικά λεπτομερής κατανόηση του αλγοκρατικού συστήματος, αλλά χρειαζόμαστε περισσότερα από τη γενική λογική που, ωστόσο, είναι εξαιρετικά δύσκολο να επιτευχθεί, αν σκεφτούμε ότι:

- Πολλά αλγοριθμικά συστήματα προστατεύονται από νόμους περί απορρήτου, είτε επειδή βασίζονται σε «εμπορικά μυστικά» και συναφή εμπορικά συμφέροντα, είτε επειδή χρησιμοποιούνται από κυβερνητικές υπηρεσίες και υπάρχουν κυβερνητικά συμφέροντα που εμποδίζουν τους ανθρώπους να «παίζουν» ή να παραβιάζουν αυτά τα συστήματα (Pasquale, 2015),
- Τα σύγχρονα συστήματα εξόρυξης δεδομένων βασίζονται όλο και περισσότερο σε αλγόριθμους Μηχανικής Μάθησης. Αυτό οφείλεται, εν μέρει, στην αύξηση του μεγέθους των συνόλων δεδομένων που πρέπει να εξορυχθούν για χρήσιμες πληροφορίες. Το μοναδικό με τέτοιους αλγόριθμους είναι ότι οι άνθρωποι δεν χρειάζεται να επιλέγουν εκ των προτέρων ή να προκαθορίζουν τους κανόνες ή τις αρχές που χρησιμοποιούν οι αλγόριθμοι για να εκτελούν τα καθήκοντά τους. Αντίθετα, οι αλγόριθμοι μπορούν να εκπαιδευτούν σε μεγάλα σύνολα δεδομένων για να δημιουργήσουν τους δικούς τους κανόνες και αρχές. Το πρόβλημα είναι η ερμηνευσιμότητα των εξόδων τέτοιων αλγορίθμων, οι οποίοι, συχνά, δεν είναι σε

θέση να εξηγήσουν με επάρκεια στους προγραμματιστές γιατί παράγουν τα αποτελέσματα που παράγουν (Danaher, 2016),

- Οι αλγόριθμοι δεν είναι μοναδικά φαινόμενα. Οποιοσδήποτε νέος αλγόριθμος είναι πιθανό να δομηθεί πάνω από άλλους, να συνταχθεί συλλογικά από ομάδες κωδικοποιητών που χρησιμοποιούν προϋπάρχουσες κωδικοποιημένες αρχιτεκτονικές και στη συνέχεια να υφανθεί σε όλο και πιο περίπλοκα αλγοριθμικά οικοσυστήματα (Seaver, 2013. Kitchin 2014a; 2014b). Είναι η αλληλεπίδραση μεταξύ όλων των μελών αυτού του αλγοριθμικού οικοσυστήματος που παράγει το χρήσιμο αποτέλεσμα, όχι η λειτουργία του μεμονωμένου νέου αλγορίθμου.

Αλλά όταν έχουμε ένα τόσο περίπλοκο οικοσύστημα, τα περιθώρια ατομικής συμμετοχής και κατανόησης είναι περιορισμένα. Ωστόσο, υπάρχει μια αντιστάθμιση αξιών που μπορεί να καταστήσει τη διατήρηση των αλγοκρατικών συστημάτων πιο κατάλληλη από το μοντέλο πολιτικής αντίστασης που θέτει ο Morozov (2013). Η αδιαφάνεια των αλγοκρατικών συστημάτων θα πρέπει να σταθμιστεί παράλληλα με άλλες ηθικές ανησυχίες, όπως ο αντίκτυπος στο απόρρητο και παράλληλα με άλλα οφέλη. Είναι σημαντικό να μην αγνοήσουμε τα οφέλη. Υπάρχουν, συχνά, ισχυρά οργανικά οφέλη που σχετίζονται με την κατασκευή και τη χρήση αλγοκρατικών συστημάτων (Mayer-Schonberger et al., 2013). Συλλέγουμε και συγκεντρώνουμε ολοένα και μεγαλύτερα σύνολα δεδομένων και οι αλγοκρατικές τεχνολογίες μας δίνουν κάποια ελπίδα να αξιοποιήσουμε αυτά τα σύνολα δεδομένων αποτελεσματικά. Αυτό ισχύει τόσο για τις κοινωνικές αρχές, όσο και για το ευρύτερο κοινό. Για να δώσουμε ένα απλό παράδειγμα, τα έξυπνα δίκτυα ηλεκτρικής ενέργειας, τα οποία βασίζονται σε μεγάλο βαθμό σε τεχνολογίες παρακολούθησης και εξόρυξης δεδομένων, μπορούν να βοηθήσουν στην ενίσχυση της αποτελεσματικότητας και της αποδοτικότητας των ανανεώσιμων πηγών ενέργειας (Rifkin, 2014). Αυτό είναι ιδιαίτερα επιθυμητό σε μια εποχή κλιματικής κρίσης και ενεργειακής ανασφάλειας.

Οι χαοτικοί αλγόριθμοι αποθήκευσης της Amazon, ανεξάρτητα με το τι πιστεύουμε για την ίδια την εταιρεία και τις ευρύτερες εργασιακές της πρακτικές, συμβάλλουν στη μείωση της σπατάλης και της αναποτελεσματικότητας και στην αύξηση της κερδοφορίας. Και ακόμη και οι εφαρμογές αυτο-παρακολούθησης, όπως αυτές που χρησιμοποιούμε στα τηλέφωνά μας καθημερινά, μπορούν να βοηθήσουν στη βελτίωση της ατομικής παραγωγικότητας, της υγείας και της ευημερίας, κυρίως, βοηθώντας μας με τον καθορισμό στόχων, τον αυτο-πειραματισμό και τη δημιουργία συνήθειας (Danaher, 2016).

Το ίδιο ισχύει και όταν εξετάζουμε τη δημόσια σφαίρα. Για να δώσουμε ένα παράδειγμα, η φοροδιαφυγή είναι ένα σημαντικό πρόβλημα. Η αποτυχία είσπραξης επαρκών φόρων υπονομεύει παροχές και πολύτιμες δημόσιες υπηρεσίες. Οι κρατικοί οργανισμοί εσόδων, ιδιαίτερα στον απόηχο της μεγάλης ύφεσης μετά το 2008, είναι συχνά υποστελεχωμένοι και δε διαθέτουν επαρκείς πόρους. Επιπλέον, οι υπάλληλοι σε αυτούς τους οργανισμούς δεν είναι πάντα ικανοί να εκμεταλλευτούν και να δουν συνδέσεις μεταξύ διαφορετικών δεξαμενών οικονομικών δεδομένων. Οι αλγόριθμοι μπορούν να βοηθήσουν. Μπορούν να εξορύξουν τις σχετικές δεξαμενές δεδομένων για χρήσιμα μοτίβα, να το κάνουν ακούραστα και αποτελεσματικά και να κάνουν συστάσεις για ελέγχους. Αυτό θα μπορούσε να είναι ένα μεγάλο όφελος για τη συλλογή φόρων. Τα οφέλη δεν είναι ούτε υποθετικά. Έχει ήδη αποδειχθεί ότι τα αλγοριθμικά συστήματα είναι καλύτερα στο να κάνουν προβλέψεις από ό,τι οι ειδικοί σε ορισμένα πεδία (Bishop et al., 2002). Έτσι, σε πολλές περιπτώσεις μπορεί να αποδειχθεί αληθές ότι αν θέλουμε να επιτύχουμε καλύτερα αποτελέσματα, καλό θα ήταν να απευθυνθούμε σε ένα αλγοκρατικό σύστημα (Danaher, 2016).

Και δεν είναι μόνο τα αποτελέσματα. Μπορεί, επίσης, να υπάρχουν διαδικαστικά οφέλη στα αλγοκρατικά συστήματα. Ο Zarsky (2011; 2012) υποστηρίζει ότι μια σημαντική διαδικαστική ανεπάρκεια με τα ανθρωποκεντρικά συστήματα λήψης αποφάσεων είναι η ευαισθησία τους σε έμμεση μεροληψία. Ας λάβουμε υπόψη τη συζήτηση για την κατάρτιση προφίλ σε σχέση με την καταπολέμηση της τρομοκρατίας και την πρόληψη του εγκλήματος. Μια ανησυχία σχετικά με τη δημιουργία προφίλ είναι ότι μπορεί αυθαίρετα να στοχεύει και να κάνει διακρίσεις εναντίον ορισμένων φυλετικών και εθνοτικών μειονοτήτων. Αυτό είναι κάτι που θα μπορούσαμε να κάνουμε και χωρίς τη χρήση αλγοριθμικών συστημάτων. Εάν οι άνθρωποι πρόκειται να στοχοποιηθούν από τέτοια μέτρα, πρέπει να στοχοποιηθούν για νόμιμους λόγους, δηλαδή επειδή είναι πραγματικά πιο πιθανό να είναι τρομοκράτες ή να διαπράξουν εγκλήματα. Το πρόβλημα είναι ότι, λόγω σιωπηρών προκαταλήψεων, οι ανθρώπινες αρχές μπορεί να μην είναι σε θέση να το κάνουν. Τα αυτοματοποιημένα αλγοκρατικά συστήματα θα μπορούσαν να κατασκευαστούν με τέτοιο τρόπο ώστε να μην είναι επιρρεπή στις ίδιες σιωπηρές προκαταλήψεις. Ως εκ τούτου, μπορεί να είναι διαδικαστικά προτιμότερα από τα συστήματα που βασίζονται στον άνθρωπο.

Όπως το θέτει ο Zarsky (2012):

«Ο αυτοματισμός εισάγει ένα εκπληκτικό όφελος. Περιορίζοντας τον ρόλο της ανθρώπινης διακριτικότητας και διαίσθησης και βασιζόμενοι σε αποφάσεις που βασίζονται σε υπολογιστή, αυτή η διαδικασία προστατεύει τις μειονότητες και άλλες ασθενέστερες ομάδες.»

Ωστόσο, όπως ο Zarsky (2012), έτσι και άλλοι έχουν επισημάνει, λόγους για να πιστεύουμε ότι τα αυτοματοποιημένα συστήματα θα μπορούσαν να αναπαράγουν τις προκαταλήψεις των ανθρώπων (Citron et al., 2014). Η κατασκευή αλγορίθμου είναι μια διαδικασία μετάφρασης (Kitchin, 2014b), υπό την έννοια ότι ένα πρόβλημα ή μια εργασία πρέπει να μετατραπεί σε ένα σύνολο από οδηγίες βήμα προς βήμα, οι οποίες με τη σειρά τους πρέπει να μεταφραστούν σε κώδικα υπολογιστή. Υπάρχει αρκετός χώρος σε αυτή τη μεταφραστική διαδικασία για να παίξουν κάποιο ρόλο σιωπηρές ή ακόμα και ρητές προκαταλήψεις. Αλλά, αν είμαστε ευσυνείδητοι σχετικά με αυτήν την πιθανότητα, μπορεί να είμαστε σε θέση να φιλτράρουμε ή να μειώσουμε την πιθανότητα μεροληψίας. Το επιχείρημα του Zarsky (2012) υποδηλώνει ότι εκτός από την εξασφάλιση καλύτερων αποτελεσμάτων, τα αλγοκρατικά συστήματα θα μπορούσαν να είναι διαδικαστικά πιο δίκαια σε όσους επηρεάζονται από αυτά. Έτσι, κατά την αξιολόγηση του τρόπου αντιμετώπισης της απειλής της αλγοκρατίας, θα χρειαστεί να εξισορροπήσουμε την απώλεια κατανόησης και συμμετοχής έναντι των πιθανών κερδών στα αποτελέσματα και στη διαδικαστική δικαιοσύνη, επιδιώκοντας τη συμμετοχή μας με κάποιον άλλον τρόπο.

4.3 Γιατί ανησυχούμε για την αλγοριθμική λήψη αποφάσεων;

Η εστίαση στη λήψη αλγοριθμικών αποφάσεων ή αλλιώς στη λήψη αποφάσεων από μηχανή δε συνεπάγεται ότι τα ανθρώπινα συστήματα λήψης αποφάσεων είναι απαραίτητα ανώτερα από ηθική ή νομική άποψη. Αντίθετα, είναι ευρέως αποδεκτό ότι οι ανθρώπινες διαδικασίες λήψης αποφάσεων και τα αποτελέσματα τους απέχουν πολύ από το να είναι τέλεια και συχνά είναι ελαττωματικά. Όμως, τα σύγχρονα νομικά συστήματα έχουν πλούσια εμπειρία στην αντιμετώπιση ελαττωμάτων στη λήψη αποφάσεων από τον άνθρωπο, τουλάχιστον στο πλαίσιο των αποφάσεων των κυβερνητικών αρχών και επίσης, το καθήκον του εντοπισμού, της ανταπόκρισης και της αποκατάστασης αυτών των ελαττωμάτων είναι η βασική συνιστώσα του σύγχρονου διοικητικού δικαίου (Oswald, 2018). Μεγάλο μέρος αυτού του εκτεταμένου νομικού σώματος στοχεύει στον εντοπισμό πότε οι αποφάσεις των δημοσίων αρχών μπορούν να προσβληθούν με δικαστικό έλεγχο, επιτρέποντας σε ένα δικαστήριο να ακυρώσει την απόφαση για μη συμμόρφωση με τις απαιτήσεις της νόμιμης διοικητικής λήψης αποφάσεων. Επομένως, αν και έχουμε μια καλά αναπτυγμένη κατανόηση της

ελλαττωματικής φύσης της ανθρώπινης λήψης αποφάσεων, όπως και ένα σύνολο νομικών και θεσμικών μηχανισμών που αποσκοπούν στην παροχή διασφαλίσεων έναντι των χειρότερων υπερβολών του, όσο ατελείς κι αν είναι, δεν έχουμε ακόμη μια ισοδύναμη ολοκληρωμένη και συστηματική αναφορά και εμπειρία των πιθανών ελαττωμάτων και μειονεκτημάτων που σχετίζονται με τη λήψη αποφάσεων από μηχανές και συστηματικών και αποτελεσματικών θεσμικών μηχανισμών για την προστασία από αυτά (Yeung, 2019).

Η Yeung (2019), υποθέτοντας ότι τα αλγοριθμικά συστήματα λήψης αποφάσεων χρησιμοποιούνται για νόμιμους κοινωνικούς σκοπούς, συνήθως, με στόχο τη βελτίωση της ποιότητας, της ακρίβειας, της αποτελεσματικότητας και της επικαιροποιημένης λήψης αποφάσεων, προσφέρει ένα εννοιολογικό πλαίσιο που προσδιορίζει τρεις (3) ευρείες πηγές κανονιστικού άγχους που προκύπτουν από τη χρήση αλγοριθμικών συστημάτων λήψης αποφάσεων:

1. ανησυχίες σχετικές με τη διαδικασία λήψης αποφάσεων,
2. ανησυχίες σχετικές με τα αποτελέσματα που δημιουργούνται και
3. ανησυχίες σχετικές με τη χρήση τέτοιων συστημάτων για την πρόβλεψη και την εξατομίκευση των υπηρεσιών που προσφέρονται σε άτομα.

4.3.1 Ανησυχίες που σχετίζονται με τις διαδικασίες

Οι ακόλουθες ανησυχίες είναι «διαδικαστικές» και προκύπτουν, κυρίως, σε σχέση με τη διαδικασία λήψης απόφασης και όχι με το ουσιαστικό περιεχόμενο ή το αποτέλεσμα της απόφασης.

Ανυπαρξία ανθρώπινου παράγοντα που να μπορεί να αναλάβει την ευθύνη για τη λήψη αποφάσεων

Πολλές ανησυχίες σχετικά με τα αυτοματοποιημένα συστήματα λήψης αποφάσεων βασίζονται σε μια αντιληπτή, ηθικά σημαντική διαφορά μεταξύ αυτοματοποιημένων και ανθρώπινων διαδικασιών λήψης αποφάσεων. Ειδικότερα, υπάρχουν ανησυχίες ότι ενδέχεται να προκύψουν λάθη στη διαδικασία λήψης αποφάσεων, συμπεριλαμβανομένων περιπτώσεων που συνεπάγονται την ορθή εφαρμογή των κριτηρίων της απόφασης με τη στενή έννοια, αλλά οι οποίες, στις ιδιαίτερες περιστάσεις της υπόθεσης, μπορεί να είναι ακατάλληλες. Αυτές οι ανησυχίες προκύπτουν σε σχέση με τα έννομα συμφέροντα των ατόμων, τα οποία θα πρέπει να είναι σε θέση να προσδιορίσουν το αρμόδιο ανθρώπινο πρόσωπο στο οποίο μπορούν να προσφύγουν για να αμφισβητήσουν την απόφαση και το

οποίο θα μπορεί να διερευνήσει και, εάν χρειαστεί, να παρακάμψει, τις αυτοματοποιημένες αποφάσεις. Η φύση αυτής της ανησυχίας και ο βαθμός στον οποίο θα μπορούσε να ξεπεραστεί ικανοποιητικά, μπορεί να εξαρτηθεί από το εάν σχετίζεται με:

- την αναλυτική διαδικασία ή τη συλλογιστική διαδικασία που αποσκοπεί στην ενημέρωση της διαδικασίας λήψης αποφάσεων, παρέχοντας πληροφορίες που βασίζονται σε δεδομένα μέσω της εφαρμογής υπολογιστικών αλγορίθμων και/ή
- το έργο της λήψης αποφάσεων.

Εάν η υποκείμενη διαδικασία συλλογιστικής βασίζεται σε απλούς αλγόριθμους βασισμένους σε κανόνες και το έργο της λήψης αποφάσεων είναι αυτοματοποιημένο, τότε η τοποθέτηση ενός ανθρώπου στον βρόχο μπορεί να βοηθήσει στην αντιμετώπιση αυτών των ανησυχιών. Για παράδειγμα, οι αυτοματοποιημένες πύλες εισιτηρίων που ανοίγουν τώρα κατά την είσοδο και την έξοδο στις αποβάθρες τρένων σε όλο το σιδηροδρομικό δίκτυο του Ηνωμένου Βασιλείου, συνήθως, επιβλέπονται από έναν άνθρωπο υπεύθυνο για τη λειτουργίας τους, ο οποίος μπορεί να βοηθήσει άτομα με κινητικές δυσκολίες και να εντοπίσει εάν η αυτοματοποιημένη συλλογιστική έχει δυσλειτουργήσει και, αν ναι, μπορεί να παρακάμψει την άρνηση πρόσβασης του μηχανήματος (Yeung, 2019).

Αντίθετα, εάν η λογική στην οποία βασίζεται μια απόφαση που δημιουργείται από μηχανή βασίζεται σε διαδικασίες δυναμικής μάθησης που χρησιμοποιούνται από διάφορες μορφές αλγορίθμων Μηχανικής Μάθησης, τότε η ουσιαστική ανθρώπινη επίβλεψη και παρέμβαση μπορεί να είναι αδύνατη, επειδή η μηχανή έχει σημαντικά πλεονεκτήματα πληροφοριών έναντι ενός ανθρώπινου χειριστή λόγω της ικανότητας επεξεργασίας χιλιάδων επιχειρησιακών μεταβλητών/χαρακτηριστικών με πολύ υψηλή ταχύτητα, έτσι ώστε να είναι πέρα από την ικανότητα ενός ανθρώπου ουσιαστικά να παρακολουθεί την ακρίβεια και την ποιότητα των εξόδων του συστήματος σε πραγματικό χρόνο. Συνεπώς, στο σημείο εφαρμογής αυτών των αποφάσεων, μπορεί να μην υπάρχει πρακτικός τρόπος για να αποδειχθεί η αξιοπιστία των αποτελεσμάτων (Matthias, 2004). Η ανθρώπινη εποπτεία μπορεί, ωστόσο, να αξίζει τον κόπο, τουλάχιστον στο βαθμό που έχει ανατεθεί σε ένα φυσικό πρόσωπο, αφού μπορεί να θεωρηθεί ότι αναλαμβάνει την ευθύνη για την τελική απόφαση και στον οποίο μπορεί να προσφύγει το θιγόμενο άτομο. Το εάν μια τέτοια εποπτεία έχει νόημα ή όχι πρέπει να αξιολογείται για το συγκεκριμένο πλαίσιο, ιδίως υπό το φως του καλά τεκμηριωμένου προβλήματος της μεροληψίας του αυτοματισμού (Skitka et al., 2000). Με άλλα λόγια, η ανθρώπινη εποπτεία στο σημείο λήψης απόφασης είναι στην καλύτερη

περίπτωση ένας μερικός και ημιτελής μηχανισμός για την αντιμετώπιση αυτών των ανησυχιών, υποδηλώνοντας ότι κάποιοι εκ των υστέρων μηχανισμοί για τη διασφάλιση της λογοδοσίας στη λήψη αποφάσεων είναι ιδιαίτερα σημαντικοί. (Yeung, 2019).

Έλλειψη συμμετοχής, δέουσας διαδικασίας ή ευκαιριών για αμφισβήτηση

Άμεσα σχετιζόμενες με ανησυχίες που αφορούν στην αδυναμία των μηχανών να «αναλάβουν την ευθύνη» για τα αποτελέσματά τους είναι ανησυχίες όπου τα αυτοματοποιημένα συστήματα λήψης αποφάσεων αρνούνται σε αυτούς που επηρεάζονται από αυτά την ευκαιρία να ενημερώνονται, να αμφισβητούν, ή με άλλον τρόπο να συμμετέχουν στην απόφαση (Veale et al., 2018). Αυτή η ανησυχία επιδεινώνεται όταν αυτές οι αποφάσεις βασίζονται σε ανάλυση που εκτελείται από αλγόριθμους Μηχανικής Μάθησης που δεν μπορούν να εξηγηθούν με όρους που είναι αντιληπτοί και κατανοητοί στο άτομο, καθιστώντας την προκύπτουσα απόφαση πολύ δύσκολη προς αμφισβήτηση (Hildebrandt, 2017). Με την άρνηση στο θιγόμενο άτομο του «δικαιώματος ακρόασης» και αμφισβήτησης της απόφασης υπονομεύεται το βασικό δικαίωμα του ατόμου να αναγνωρίζεται και να αντιμετωπίζεται ως ηθικός παράγοντας, δηλαδή άτομο με δικές του απόψεις και ικανό να ενεργεί με βάση τους δικούς του λόγους και επομένως το δικαίωμα του για αξιοπρέπεια και σεβασμό. (Gardner, 2006). Ταυτόχρονα, η άρνηση της δυνατότητας ακρόασης στο επηρεαζόμενο άτομο μπορεί να εμποδίσει τον λήπτη να αποκτήσει σχετικές πληροφορίες που μπορεί να είναι συναφείς με την απόφαση, διακινδυνεύοντας με αυτόν τον τρόπο να υπονομευτεί η ακρίβεια και η ποιότητα της προκύπτουσας απόφασης (Galligan, 1997).

Το λεγόμενο «δικαίωμα στην ακρόαση» αποτελεί μία από τις δύο βασικές αρχές που διέπουν το διοικητικό δίκαιο στο Ηνωμένο Βασίλειο και αλλού. Το «δικαίωμα στην ακρόαση» προβλέπει ότι τα άτομα των οποίων τα δικαιώματα και τα συμφέροντα επηρεάζονται αρνητικά από μια απόφαση θα πρέπει να έχουν δικαίωμα και σε α) δίκαιη ακρόαση, συμπεριλαμβανομένου του δικαιώματος συμμετοχής στη λήψη αυτής της απόφασης που περιλαμβάνει το δικαίωμα ενημέρωσης για την υπόθεση εναντίον του, και β) ο λήπτης της απόφασης θα πρέπει να είναι αμερόληπτος (Endicott, 2015). Και οι δύο αυτές αρχές εμπλέκονται σε αλγοριθμικά συστήματα λήψης αποφάσεων και προκαλούν ανησυχίες. Ενώ η απαίτηση για δίκαιη ακρόαση είναι πιο έντονη σε σχέση με αποφάσεις που παρεμβαίνουν στα θεμελιώδη δικαιώματα ενός ατόμου (όπως ένα άτομο που κατηγορείται για σοβαρό ποινικό αδίκημα που, εάν καταδικαζόταν, θα στερούνταν την ελευθερία του), εφαρμόζεται όποτε εκείνοι των οποίων τα δικαιώματα, τα συμφέροντα ή οι θεμιτές προσδοκίες ενδέχεται

να επηρεαστούν αρνητικά από μια απόφαση δημόσιας αρχής. Έτσι, για παράδειγμα, η νομοθεσία της Ευρωπαϊκής Ένωσης θεσμοθετεί εκτελεστά «δικαιώματα» συμμετοχής στη διακυβέρνηση των νέων τεχνολογιών, αν και ορισμένοι επικρίνουν αυτά τα δικαιώματα, ως αδικαιολόγητα στενά, ως προς το πεδίο και το περιεχόμενό τους (Lee, 2017).

Ειδικότερα, η νομικός Sheila Jasanoff (2016) έχει επανειλημμένα επισημάνει ότι το τεχνολογικό μας περιβάλλον και τα τεχνολογικά συστήματα που χρησιμοποιούμε κατανέμουν την εξουσία και έχουν «διαφορετικές επιπτώσεις» σε ομάδες και άτομα, αλλά για τα οποία η υπεύθυνη διακυβέρνηση είναι ζωτικής σημασίας, αν και απουσιάζει σοβαρά και εντυπωσιακά από τις διαδικασίες μέσω των οποίων αναπτύσσονται και εφαρμόζονται οι τεχνολογικές καινοτομίες στην κοινωνία. Ως εκ τούτου, υποστηρίζει την επείγουσα ανάγκη για μηχανισμούς συμμετοχής των ατόμων, ώστε να διαδραματίσουν πιο ενεργούς ρόλους στο σχεδιασμό και τη διαχείριση του τεχνολογικού μέλλοντός τους (Jasanoff, 2016).

Αθέμιτες ή παράνομες διακρίσεις

Το δεύτερο σκέλος του δικαιώματος στη δέουσα διαδικασία που απαιτεί αμερόληπτη λήψη αποφάσεων, εγείρει, επίσης, προβληματισμό, κυρίως σε σχέση με τη δυνατότητα των αλγοριθμικών διαδικασιών να εισαγάγουν άδικες ή παράνομες διακρίσεις.

Οι διακρίσεις είναι αναπόφευκτες κατά τη λήψη αποφάσεων επιλογής ατόμων, με αποτέλεσμα όσοι δεν επιλέγονται για μια εργασία, για παράδειγμα, να πληγούν τα συμφέροντά τους. Ωστόσο, οι διακρίσεις δεν είναι ηθικά ή νομικά προβληματικές, εκτός εάν είναι άδικες ή παράνομες και αυτό εξαρτάται σε μεγάλο βαθμό από το εάν τα κριτήρια βάσει των οποίων έγινε η αξιολόγηση είναι σχετικά με το θέμα που θα αποφασιστεί. Το παράδειγμα της δασκάλας που απολύθηκε επειδή είχε κόκκινα μαλλιά, αποτελεί ένα περίφημο παράδειγμα απόφασης τόσο παράλογης που κανένας λογικός άνθρωπος δεν θα μπορούσε να πιστέψει ότι δεν θα ακυρωνόταν από ένα δικαστήριο. Μια τέτοια απόφαση είναι άδικη επειδή, συνήθως, θεωρούμε το χρώμα των μαλλιών δεν έχει αιτιώδη σχέση με την επαγγελματική ικανότητα ενός ατόμου, ως δασκάλου. Μια δασκάλα που απολύθηκε αποκλειστικά λόγω του χρώματος των μαλλιών της υφίσταται όχι μόνο άμεση βλάβη που σχετίζεται με την απώλεια της εργασίας της, αλλά και ηθική βλάβη, επειδή η απόλυσή της ήταν αυθαίρετη και επομένως άδικη (Endicot, 2005).

Σύμφωνα με το αγγλικό δίκαιο, μόνο οι δημόσιες αρχές υποχρεούνται από το διοικητικό δίκαιο να λαμβάνουν υπόψη τους σχετικούς λόγους και να αποφεύγουν να λαμβάνουν υπόψη άσχετους παράγοντες κατά την άσκηση της εξουσίας λήψης αποφάσεων (Endicott, 2005).

Αντίθετα, ο ιδιωτικός τομέας, συνήθως, απολαμβάνει σημαντική ελευθερία λήψης αποφάσεων υπό τον όρο ότι συμμορφώνεται με γενικούς νόμους, όπως νομικές υποχρεώσεις που απορρέουν από τη νομοθεσία περί ισότητας. Η χρήση προγνωστικών αναλυτικών στοιχείων είναι ιδιαίτερα προβληματική επειδή μεταβλητές που συνήθως θεωρούνται ηθικά ή/και αιτιωδώς άσχετες μπορεί να έχουν πολύ υψηλό βαθμό προγνωστικής αξίας (στατιστική συνάφεια). Ωστόσο, είναι δεοντολογικά αποδεκτό να βασίζεται μια απόφαση σε μια υποκείμενη λογική που αφορά τον εντοπισμό αξιόπιστων προγνωστικών παραγόντων της απόδοσης της εργασίας, ανεξάρτητα από το εάν αυτοί οι προγνωστικοί παράγοντες έχουν κάποια αιτιώδη συνάφεια με την ίδια την εργασία;

Από τη μια πλευρά, το θιγόμενο άτομο έχει έννομο συμφέρον να μην αξιολογηθεί βάσει εκτιμήσεων που δεν σχετίζονται αιτιωδώς με την απόφαση. Από την άλλη πλευρά, ένας υποψήφιος εργοδότης έχει έννομο συμφέρον να υιοθετήσει τις πιο αξιόπιστες μεθόδους για την αξιολόγηση της καταλληλότητας και της πιθανής επιτυχίας του υποψηφίου για τους σκοπούς της επιχείρησής του.

Η επίλυση αυτής της σύγκρουσης εξαρτάται καταρχάς από το εάν το θιγόμενο άτομο έχει θεμελιώδες και νόμιμο δικαίωμα να μην υφίσταται διακρίσεις με βάση την εν λόγω μεταβλητή. Σε αυτήν την περίπτωση, επειδή δεν υπάρχει δικαίωμα να απαλλαγούμε από διακρίσεις με βάση το χρώμα των μαλλιών, μπορεί να είναι νομικά αποδεκτό να λαμβάνεται υπόψη το χρώμα των μαλλιών, αν και μπορεί παρόλα αυτά να είναι ηθικά ψευδές - και εγείρει επίσης ερωτήματα σχετικά με την επεξήγηση της απόφασης και το συνακόλουθο καθήκον του υπεύθυνου λήψης αποφάσεων να αιτιολογεί. Αλλά αυτό δεν τελειώνει, επειδή το χρώμα των μαλλιών (μια νομικά επιτρεπτή βάση για διακρίσεις) μπορεί να σχετίζεται άμεσα με τη φυλή (οι μαύροι σπάνια έχουν ξανθά μαλλιά, ενώ όσοι έχουν σκούρα μαλλιά είναι πιο πιθανό να είναι μαύροι). Εάν ναι, τότε μπορεί να είναι νομικά και ηθικά απαραίτητο να χρησιμοποιηθούν τεχνικές για την εξόρυξη δεδομένων χωρίς διακρίσεις, προκειμένου να αποφευχθεί η παραβίαση του δικαιώματος των ατόμων να μην υπόκεινται σε παράνομες διακρίσεις και η επακόλουθη βλάβη που σχετίζεται με μια τέτοια μεταχείριση (Kamiran, 2012).

Διαφάνεια, Επεξηγησιμότητα και Αιτιολογία

Τα αλγοριθμικά συστήματα έχουν προσελκύσει ανησυχίες σχετικά με την αδιαφάνειά τους, η οποία αντικατοπτρίζεται στις απεικονίσεις των αλγορίθμων, ως «μαύρα κουτιά» (Pasquale 2015). Οι ανησυχίες σχετικά με τη διαφάνεια είναι ιδιαίτερα έντονες όταν οι αποφάσεις που

προκύπτουν λαμβάνονται βάσει αλγοριθμικής ανάλυσης προσωπικών δεδομένων προκειμένου να καθοριστεί ή να γίνουν συστάσεις σχετικά με την πρόσβαση ενός ατόμου σε οφέλη ή άλλες ευκαιρίες, αλλά προκύπτουν και σε σχέση με τα ίδια τα συστήματα γενικότερα (Yeung, 2018a). Όχι μόνο οι αλγόριθμοι που χρησιμοποιούνται στις διαδικασίες λήψης αποφάσεων, συνήθως, αποκρύπτονται από το κοινό, αφού συχνά προστατεύονται νομικά, ως εμπορικά μυστικά (Pasquale 2015), αλλά ακόμη και αν οι αλγόριθμοι αποκαλύπτονταν, θα ήταν άνευ σημασίας για όλους εκτός από εκείνους με την απαιτούμενη εξειδικευμένη τεχνική γνώση και εμπειρία για την ερμηνεία τους. Αυτό αφορά ιδιαίτερα τους αλγόριθμους εξόρυξης δεδομένων που έχουν διαμορφωθεί για να εντοπίζουν «κρυφά» μοτίβα και συσχετίσεις σε ογκώδη και συχνά πολλαπλά σύνολα δεδομένων. Οι οργανισμοί που χρησιμοποιούν αυτά τα συστήματα μπορεί να είναι απρόθυμοι να αποκαλύψουν αυτές τις προγνωστικές μεταβλητές, τόσο για να μειώσουν την προοπτική «παιχνιδιού» από άτομα, όσο και επειδή η απουσία οποιασδήποτε αναγκαίας αιτιώδους σχέσης μεταξύ αυτών των μεταβλητών και του φαινομένου ενδιαφέροντος μπορεί να προκαλέσει δυσπιστία και ανησυχία μεταξύ των «επιλεγμένων ατόμων» για τους οποίους η συλλογιστική μπορεί να φαίνεται πολύ ψεύτικη (όπως φαίνεται από το παράδειγμα που αναφέρεται παραπάνω στο οποίο το χρώμα των μαλλιών μπορεί να αναγνωριστεί ως ένας αξιόπιστος προγνωστικός παράγοντας της διδακτικής ικανότητας). Αν και υπάρχουν εύλογες ανησυχίες σχετικά με το «παιχνίδι» σε ορισμένες περιπτώσεις (Weller, 2017), η έλλειψη οποιασδήποτε αιτιώδους συνάφειας ή «κοινής λογικής» σχέσης μεταξύ της μεταβλητής και του αποτελέσματος που υποστηρίζεται ότι προβλέπει μπορεί να μην δικαιολογεί την απόκρυψη επεξήγησης της βάσης για την απόφαση: η δασκάλα που απολύθηκε λόγω του χρώματος των μαλλιών της θα πρέπει να ενημερωθεί για τους λόγους της απόλυσής της.

Ωστόσο, οι δυσκολίες που σχετίζονται με την επεξήγηση της βάσης μιας απόφασης και συγκεκριμένα στην περίπτωση ορισμένων μορφών αλγορίθμων Μηχανικής Μάθησης όπου δεν μπορούν εύκολα να εξηγηθούν, ακόμη και από τους προγραμματιστές του αλγορίθμου, μπορεί να οδηγήσουν σε αδυναμία να ληφθούν επαρκώς υπόψη τα έννομα συμφέροντα όσων επηρεάζονται αρνητικά από αλγοριθμικές αποφάσεις, ώστε να γνωρίζουν και να κατανοούν πραγματικά τους λόγους της αρνητικής απόφασης με όρους κατανοητούς σε αυτούς. Αντίθετα, στις συμβατικές ανθρώπινες διαδικασίες λήψης αποφάσεων, οι άνθρωποι μπορούν, τουλάχιστον σε πρώτη φάση, να διατυπώσουν τους λόγους για την απόφασή τους όταν ερωτηθούν, αν και δεν υπάρχει καμία εγγύηση ότι οι λόγοι που προσφέρονται θα είναι

μια αληθινή και πιστή αναπαράσταση της «πραγματικής» βάσης στην οποία ελήφθη η απόφαση (Weller, 2017).

Λήψη Αποφάσεων χωρίς ανθρώπινη συνεισφορά – Dehumananization.

Τα πλήρως αυτοματοποιημένα συστήματα λήψης αποφάσεων ενδέχεται, επίσης, να αποκλείουν την ενσωμάτωση σημαντικών ηθικών αξιών στη διαδικασία λήψης αποφάσεων (Roth, 2016). Αν και η αφαίρεση της ανθρώπινης διακριτικής ευχέρειας από τις διαδικασίες λήψης αποφάσεων μπορεί να μειώσει την πιθανότητα αυθαιρεσίας και συνειδητής και υποσυνείδητης μεροληψίας και προκατάληψης που μπορεί να επηρεάσει την ανθρώπινη λήψη αποφάσεων, εξαλείφει, επίσης, τις αρετές της ανθρώπινης διακριτικότητας, κρίσης και δικαιοσύνης, οι οποίες έχουν από καιρό αναγνωριστεί στην κοινωνικο-νομική επιστήμη, ως ζωτικής σημασίας για την αντιμετώπιση της αναπόφευκτης ατέλειας που σχετίζεται με τους νομικούς κανόνες (Black, 1997). Ο εν λόγω προβληματισμός μπορεί να εξεταστεί από τουλάχιστον τρεις οπτικές γωνίες και συγκεκριμένα από την οπτική γωνία (α) του λήπτη αποφάσεων, (β) του ατόμου για το οποίο λαμβάνεται η απόφαση και (γ) των σχεσιακών διαστάσεων της ανθρώπινης λήψης αποφάσεων.

(α) Ο λήπτης της απόφασης

Από τη σκοπιά του υπεύθυνου λήψης αποφάσεων, η διακριτικότητα στη λήψη αποφάσεων επιτρέπει να μετριαστεί η σκληρότητα των κανόνων από εκτιμήσεις συμπόνιας, συμπάθειας, ηθικής και ευσπλαχνίας σε συγκεκριμένες περιπτώσεις. Αντίθετα, οι αποφάσεις που λαμβάνονται από μηχανή βασίζονται σε μια ψυχρή, υπολογιστική λογική που εφαρμόζεται με αμείλικτη συνέπεια. Ενώ η συνέπεια είναι μια ηθικά επιθυμητή ιδιότητα των συστημάτων λήψης αποφάσεων και μια από τις αρετές που συνδέεται με τη λήψη αποφάσεων βάσει κανόνων, μπορεί, ωστόσο, να υπάρχουν περιπτώσεις όπου μπορεί να είναι θεμιτό να απομακρυνθεί κανείς από την αυστηρή εφαρμογή ενός κανόνα. Για παράδειγμα, ο Roth (2016) παρατηρεί ότι σε ορισμένες αμερικάνικες σχολικές περιφέρειες, τα αυτοματοποιημένα συστήματα παρακολούθησης απουσιών χρησιμοποιούνται για την αυτόματη παραπομπή των μαθητών στο δικαστήριο απουσιών, όταν ο μαθητής έχει συγκεντρώσει συγκεκριμένο αριθμό απουσιών, ένα σύστημα που έχει επικριθεί από μια ομάδα δικαιωμάτων για άτομα με αναπηρία με βάση το ότι «δεν αφήνει περιθώρια διόρθωσης εάν μια απουσία δικαιολογείται επειδή σχετίζεται με την αναπηρία ενός μαθητή» (Roth 2016). Αν και οι επιστήμονες δεδομένων μπορεί να απαντήσουν ότι θα μπορούσαμε να εντοπίσουμε αυτούς τους

παράγοντες και στη συνέχεια να τους προσθέσουμε στο μοντέλο για να βελτιώσουμε την απόδοσή του, είναι αμφίβολο εάν τέτοιοι ακατάστατοι παράγοντες «πραγματικής ζωής» μπορούν να μεταφραστούν εύκολα και πιστά σε αναγνώσιμα από μηχανή δεδομένα και προγράμματα (Oswald, 2018).

Σε μια διαφορετική περίπτωση, αυξάνεται το άγχος ότι η αυτοματοποίηση σημαντικών αποφάσεων, ιδιαίτερα, όταν είναι κρίσιμες για τη ζωή, όπως η λήψη της απόφασης για πυροδότηση θανατηφόρων όπλων, μπορεί να εξαλείψει την ικανότητα του λήπτη αποφάσεων να βασίζεται στη δική του ηθική αίσθηση του σωστού και του λάθους για τη λήψη της απόφασης. Για μια απόφαση που μπορεί να έχει πολύ σημαντικές αρνητικές επιπτώσεις σε ένα συγκεκριμένο άτομο, η προσωπική ηθική και η επιθυμία του λήπτη της απόφασης να κάνει αυτό που είναι σωστό, μπορεί να θεωρηθεί σημαντική διασφάλιση για το εν λόγω άτομο και να λειτουργήσει ως διαβεβαίωση ότι τέτοιες αποφάσεις είναι ηθικές και κατάλληλες. Ως εκ τούτου, όταν αποφασίζει εάν, για παράδειγμα, θα στοχεύσει ένα φονικό όπλο εναντίον άλλου, ο ανθρώπινος στρατιώτης μπορεί να αναμένεται να αναγνωρίσει τον στόχο ως άνθρωπο με τον οποίο μοιράζεται τον κοινό δεσμό της ανθρωπότητας και πρέπει να αντέχει για πάντα τις ηθικές και συναισθηματικές συνέπειες της απόφασής του. Αυτό, ένα πλήρως αυτοματοποιημένο φονικό όπλο δεν μπορεί και δεν θα το κάνει (Yeung, 2019).

(β) Το επηρεαζόμενο άτομο

Ενώ οι ανησυχίες σχετικά με την ικανότητα των συστημάτων αλγοριθμικής λήψης αποφάσεων να κάνουν διακρίσεις κατά ομάδων που ιστορικά μειονεκτούσαν είναι εμφανείς στις σύγχρονες συζητήσεις, συχνά, παραβλέπεται η ικανότητά τους να κάνουν διακρίσεις εναντίον ενός ατόμου. Επειδή τα μοντέλα στα οποία στηρίζονται αυτές οι τεχνικές βασίζονται στην υπόθεση ότι οι προηγούμενες συμπεριφορές είναι οι πιο αξιόπιστοι προγνωστικοί παράγοντες μελλοντικών συμπεριφορών, αποτυγχάνουν να λάβουν υπόψη τη φύση των ατόμων ως ηθικών παραγόντων, ως ατόμων με τη δική τους θέληση και την ικανότητα να απελευθερωθούν από προηγούμενες συνήθειες, συμπεριφορές και προτιμήσεις. Παρόλο που εμείς, ως άνθρωποι, μπορεί να είμαστε προβλέψιμοι και οι συμπεριφορές μας να ακολουθούν συγκεκριμένα μοτίβα, εντούτοις παραμένουμε αυτοαντιδραστικοί ηθικοί παράγοντες, ικανοί να κάνουμε ενεργές επιλογές και να ασκούμε αυτοέλεγχο ώστε να μπορούμε να αντισταθούμε στον πειρασμό και να χαράξουμε ένα νέο μονοπάτι μέσω του οποίου μπορούμε να επιδιώξουμε να αλλάξουμε τον εαυτό μας και το προσδοκώμενο μέλλον μας, υπό την

προϋπόθεση ότι έχουμε αρκετό θάρρος και αποφασιστικότητα να το κάνουμε (Hildebrandt, 2017).

Εάν οι κρίσιμες αποφάσεις για εμάς αφεθούν εξ ολοκλήρου στην αλγοριθμική αξιολόγηση, τότε κινδυνεύουμε να παγιδευτούμε σε αλγοριθμικές φυλακές που είναι, κατά μία έννοια, δική μας κατασκευή (Davidow, 2014). Δεν έχουμε καμία ελπίδα λύτρωσης ή επανεφεύρεσης μέσω της άσκησης της δικής μας ηθικής αποφασιστικότητας και θάρρους να ξεφύγουμε από τις προηγούμενες συνήθειες και συμπεριφορές μας. Εν ολίγοις, τα αλγοριθμικά συστήματα λήψης αποφάσεων που χρησιμοποιούν προφίλ συμπεριφοράς ενδέχεται να υπονομεύσουν το δικαίωμά μας να μας φέρονται με αξιοπρέπεια και σεβασμό, διακινδυνεύοντας να διαβρώσουν την ικανότητά μας για αυτονομία και αυτοδιάθεση.

(γ) Η Σχεσιακή Διάσταση της Ανθρώπινης Λήψης Αποφάσεων

Σχετικές αλλά διαφορετικές και από τις δύο αυτές οπτικές γωνίες είναι οι ανησυχίες ότι η εξάρτηση από αλγοριθμικά συστήματα λήψης αποφάσεων μπορεί να εξαλείψει τον σχεσιακό, επικοινωνιακό χαρακτήρα των ανθρώπινων διαδικασιών λήψης αποφάσεων που παίζει ζωτικό ρόλο στη συλλογική μας ζωή. Ας δούμε, για παράδειγμα, την αξία του λεγόμενου «δικαιώματος στην ακρόαση», το οποίο αναγνωρίζεται ως βασικό στοιχείο του δικαιώματος της δέουσας διαδικασίας. Αν και η πρωταρχική του αξία έγκειται στο να δίνει στο θιγόμενο άτομο την ευκαιρία να εκφράσει τις δικές του απόψεις σχετικά με το θέμα, κάτι που μπορεί να παρέχει στον λήπτη της απόφασης πληροφορίες που σχετίζονται με την εκάστοτε απόφαση, η αξία έγκειται στο να δώσει στον θιγόμενο την υποκειμενική εμπειρία του να τον ακούει ένας συνάνθρωπός του. Οι ανθρώπινες διαδικασίες λήψης αποφάσεων παρέχουν στον λήπτη της απόφασης την ευκαιρία να αναγνωρίσει και να συμπονέσει την υποκειμενική πραγματικότητα της εμπειρίας αυτού του ατόμου. Κατά τη λήψη αποφάσεων που συνδέονται με δυστυχία, θλίψη και αγωνία, όπως για παράδειγμα αποφάσεις που αφορούν οικογενειακά θέματα, μπορεί να είναι ιδιαίτερα σημαντικό όσοι επηρεάζονται άμεσα από μια απόφαση να αισθάνονται ότι εισακούστηκαν από τους αρμόδιους φορείς λήψης αποφάσεων και ότι η υποκειμενική τους εμπειρία έχει κατανοηθεί σωστά. Αυτό απαιτεί αναγνώριση του συναισθηματικού κόστους και του νοήματος των εν λόγω γεγονότων για τα οποία η ενσυναίσθηση και η συμπόνια είναι η ηθικά κατάλληλη απάντηση (Jack, 2018), ανεξάρτητα από το αποτέλεσμα της διαδικασίας λήψης αποφάσεων (Parens, 2010). Ενώ τα συστήματα Τεχνητής Νοημοσύνης είναι ολοένα και πιο ικανά να προσομοιώνουν

ανθρώπινα συναισθήματα και αποκρίσεις, αποτελούν τεχνητά και κατώτερα υποκατάστατα της αυθεντικής ενσυναίσθησης, της συμπόνιας και του ενδιαφέροντος εκείνων με τους οποίους μοιραζόμαστε τους κοινούς δεσμούς της ανθρώπινης εμπειρίας (Yeung, 2019).

4.3.2 Ανησυχίες με βάση το αποτέλεσμα

Εκτός από αυτές τις διαδικαστικές ανησυχίες, υπάρχουν ανησυχίες σχετικά με το ουσιαστικό περιεχόμενο ή τα αποτελέσματα που δημιουργούνται από αλγοριθμικά συστήματα λήψης αποφάσεων, τις οποίες αναλύουμε παρακάτω.

Λανθασμένες και Ανακριβείς Αποφάσεις

Προκύπτουν κατανοητές ανησυχίες σχετικά με την πιθανότητα οι αποφάσεις που παράγονται από αλγοριθμικά συστήματα λήψης αποφάσεων να είναι λανθασμένες ή ανακριβείς. Τα σφάλματα μπορεί να έχουν σοβαρές αρνητικές συνέπειες, ανάλογα με την εφαρμογή στην οποία είναι ενσωματωμένα αυτά τα συστήματα. Έτσι, για παράδειγμα, εάν είναι ενσωματωμένα σε κρίσιμα συστήματα ασφαλείας, όπως αυτά που ενημερώνουν τη λειτουργία αυτόνομων οχημάτων, αυτά τα σφάλματα μπορεί να έχουν θανατηφόρες συνέπειες. Άλλα είδη σφαλμάτων μπορεί να φαίνονται σχετικά ασήμαντα, για παράδειγμα, αλγοριθμικά συστήματα συστάσεων λήψης αποφάσεων για καταναλωτικά προϊόντα. Ωστόσο, ακόμη και τότε, όπως δείχνει το σκάνδαλο της Cambridge Analytica, τα αλγοριθμικά συστήματα που παρέχουν συστάσεις περιεχομένου πολυμέσων θα μπορούσαν να έχουν εξαιρετικά αποτελέσματα στο να παραμορφώσουν τα αποτελέσματα δημοκρατικών εκλογών. Για τα άτομα, τα λάθη που προκύπτουν από τη χρήση προσωπικών προφίλ βάσει δεδομένων μπορεί να είναι εξαιρετικά σημαντικά, ιδιαίτερα όταν χρησιμοποιούνται για την ενημέρωση κρίσεων σχετικά με άτομα βάσει ομαδικών χαρακτηριστικών που μπορεί να μην αντικατοπτρίζουν με ακρίβεια τα ιδιαίτερα χαρακτηριστικά ή τις περιστάσεις του ατόμου και ως εκ τούτου να οδηγούν σε λανθασμένα αποφάσεις για αυτό το άτομο. Σε τέτοιες περιπτώσεις, η άμεση βλάβη στο άτομο έγκειται στην ανακριβή ή εσφαλμένη απόφαση. Η σχετική σοβαρότητα της ζημίας θα εξαρτηθεί από τις υλικές και άυλες συνέπειες του λάθους, γιατί ενώ η στέρηση των θεμελιωδών δικαιωμάτων και ελευθεριών του ατόμου βρίσκεται στο πιο σοβαρό άκρο της κλίμακας, αποφάσεις που καταλήγουν σε στέρηση νομικών δικαιωμάτων ή μείωση ή άρνηση των ευκαιριών ζωής και των οφελών τους και άλλες μορφές δυσμενούς μεταχείρισης που

σχετίζονται με εσφαλμένη ταξινόμηση (στίγμα, στερεότυπα) μπορεί, επίσης, να έχουν σημαντικές δυσμενείς συνέπειες για όσους επηρεάζονται άμεσα (Davidow, 2014).

Προκατειλημμένα ή μεροληπτικά αποτελέσματα που δημιουργούν βλάβη ή αδικία

Η παράνομη διάκριση είναι ένα από τα πολλά προβλήματα μεροληψίας που μπορεί να επηρεάσει την αλγοριθμική λήψη αποφάσεων. Προκατειλημμένες αποφάσεις μπορούν, επίσης, να προκύψουν, εάν οι τεχνικές εξόρυξης δεδομένων που χρησιμοποιούνται για την ενημέρωση της λήψης αποφάσεων κωδικοποιούν προκαταλήψεις λόγω κρυφής προκατάληψης στον σχεδιασμό των αλγορίθμων (Roth, 2016). Εάν τα δεδομένα εκπαίδευσης είναι προκατειλημμένα ή εάν τα δεδομένα που τροφοδοτούνται στο σύστημα είναι προκατειλημμένα, τότε θα προκύπτουν μεροληπτικά αποτελέσματα. Όταν αυτά τα αποτελέσματα έχουν άμεσο αντίκτυπο σε άτομα, αυτό μπορεί να προκαλέσει βλάβη και αδικία, ανάλογα με τις συνέπειες της μεροληπτικής απόφασης για το θιγόμενο άτομο (Yeung, 2019).

Σχετικοί αλλά διακριτοί κίνδυνοι μπορεί να προκύψουν όταν τα υποκείμενα δεδομένα είναι προκατειλημμένα ή ο τρόπος με τον οποίο ο ίδιος ο αλγόριθμος λειτουργεί είναι προκατειλημμένος, με αποτέλεσμα μεροληπτικά και ανακριβή αποτελέσματα. Αυτά τα προβλήματα είναι γενικής φύσεως και ενδέχεται να προκύψουν ανεξάρτητα από το αν τα δεδομένα είναι προσωπικά δεδομένα ή όχι. Για παράδειγμα, ένας αλγόριθμος ταξινόμησης εικόνων μπορεί να είναι προκατειλημμένος υπέρ της έμφασης ιδιαίτερων χαρακτηριστικών ενός τοπίου, όπως π.χ. να τείνει να προσδιορίζει υπερβολικά τα ανθρωπογενή αντικείμενα ενώ υπογραμμίζει τα χαρακτηριστικά του φυσικού τοπίου, με αποτέλεσμα οι έξοδοί του να παραμορφώνονται υπέρ των ανθρωπογενών αντικειμένων. Αυτό μπορεί να μην προκαλέσει καμία βλάβη ή άλλες δυσμενείς συνέπειες από μόνο του, αλλά μπορεί να το κάνει εάν αυτό το προϊόν χρησιμοποιηθεί με τρόπους που επηρεάζουν άμεσα και διαφοροποιημένα άτομα ή/και κοινότητες. Για αλγοριθμικά συστήματα που επεξεργάζονται προσωπικά δεδομένα για να ενημερώσουν και να αυτοματοποιήσουν αποφάσεις που επηρεάζουν άμεσα άτομα, αυτές οι διακρίσεις αποτελούν σημαντική ανησυχία λόγω των μεροληπτικών και ανακριβών αποτελεσμάτων που δημιουργούν και μπορούν να προκαλέσουν ουσιαστική αδικία (O'Neil, 2016).

Ένα διαφορετικό πρόβλημα προκύπτει εάν συγκεκριμένες ομάδες (π.χ. με βάση τη φυλή, τη θρησκεία, την εθνικότητα κ.λπ.) βρίσκονται ιστορικά σε μειονεκτική θέση. Υπό αυτές τις συνθήκες, ακόμη και όταν οι αλγόριθμοι και τα υποκείμενα δεδομένα είναι ακριβή και ο

ίδιος ο αλγόριθμος δε μεροληπτεί, οι τεχνικές δημιουργίας προφίλ μπορούν άθελά τους να δημιουργήσουν μια βάση αποδεικτικών στοιχείων που οδηγεί σε διακρίσεις, διαιωνίζοντας και εμβαθύνοντας προηγούμενα μοτίβα διακρίσεων εναντίον αυτών των ομάδων και δημιουργώντας σχετικές βλάβες. Για παράδειγμα, πειράματα σε προφίλ χρηστών της Google που χρησιμοποιούσαν αλγοριθμικό μοντέλο λήψης αποφάσεων βασισμένο στο σύστημα Adfisher είχαν ως αποτέλεσμα στους άνδρες χρήστες να εμφανίζονται υψηλά αμειβόμενες διαφημίσεις εργασίας έξι φορές συχνότερα από ό,τι εμφανίζονταν σε γυναίκες χρήστες (Datta, 2015). Αυτό το πρόβλημα δεν αποδίδεται στον αλγόριθμο καθαυτό, αλλά στον τρόπο με τον οποίο αναπαράγει και ενισχύει τις ιστορικές διακρίσεις, διαιωνίζοντας έτσι την αδικία, τα σχετικά στερεότυπα και τον στιγματισμό έναντι των μειονεκτούν ομάδων.

Μίμηση ανθρώπινων χαρακτηριστικών και συναισθηματικές αντιδράσεις

Οι μεροληπτικές επιδράσεις των αλγοριθμικών συστημάτων λήψης αποφάσεων είναι συχνά ακούσιες. Μάλλον, διαφορετικές ανησυχίες προκύπτουν όταν τα συστήματα έχουν σχεδιαστεί σκόπιμα για να μιμούνται τις ενεργητικές αποκρίσεις άλλων με τρόπους που μπορεί να είναι παραπλανητικοί ή να εκμεταλλεύονται την ευπάθεια των χρηστών (Boden et al., 2011). Ανησυχίες αυτού του είδους έχουν εκφραστεί, κυρίως, σε σχέση με τα ρομπότ, ιδιαίτερα όταν έχουν σχεδιαστεί για να παρέχουν φροντίδα ή συντροφιά σε ευάλωτα άτομα. Έχουν εκφραστεί φόβοι ότι εάν η εμφάνιση και η συμπεριφορά ενός ρομπότ είναι σχεδιασμένη να προσομοιώνει αυτή ενός ζωντανού πλάσματος, όπως ανθρώπου ή ζώου, προκειμένου να αυξηθεί η αποδοχή και η προσκόλληση του χρήστη, αυτό μπορεί να θεωρηθεί παραπλανητικό και ως εκ τούτου ηθικά αμφίβολο (Boden et al., 2011). Αν και αυτές οι ανησυχίες μπορεί να φαίνονται, αρχικά, ότι είναι απίθανο να συνδέονται με αλγοριθμικά συστήματα λήψης αποφάσεων, πιο πρόσφατη έρευνα αναγνωρίζει πώς μπορούν αυτά τα συστήματα να χρησιμοποιηθούν για την προσομοίωση της φωνής ενός συγκεκριμένου ατόμου. Τα συστήματα αυτά θα μπορούσαν στη συνέχεια να χρησιμοποιηθούν για να εξαπατήσουν άτομα, τα οποία οδηγούνται να πιστεύουν ότι μιλούν απευθείας με το άτομο του οποίου η φωνή προσομοιώνεται, συνήθως ένα άτομο γνωστό σε αυτούς, όπως ένας συγκεκριμένος φίλος ή μέλος της οικογένειας (Brundage et al., 2018). Ακόμη και αν παραμερίσουμε τις ανησυχίες σχετικά με τη δυνατότητα εκμετάλλευσης αυτών των εφαρμογών από κακοήθεις παράγοντες, υπάρχουν ανησυχίες ότι ακόμη και η καλοπροαίρετη χρήση αυτών των συστημάτων για μίμηση της συμπεριφοράς ανθρώπων ή ζώων μπορεί να προκαλέσει ψευδείς πεποιθήσεις στο μυαλό των χρηστών, αποτελώντας μια μορφή εξαπάτησης και άρα ηθικά προβληματικής (Boden et al., 2011).

4.3.3 Πρόβλεψη βάσει δεδομένων και εξατομικευμένες υπηρεσίες πληροφόρησης

Οι ακόλουθες ανησυχίες προκύπτουν, κυρίως, σε σχέση με τη διαδικασία δημιουργίας προφίλ που μπορούν να διαμορφωθούν ώστε να προσφέρουν εξατομικευμένες υπηρεσίες πληροφόρησης.

Προγνωστική εξατομίκευση και «Hypernudging» με γνώμονα τα δεδομένα

Το πρόσθετο ηθικό άγχος προκύπτει από την αυξανόμενη χρήση αλγοριθμικών συστημάτων λήψης αποφάσεων που βασίζονται στη συλλογή και εξόρυξη προσωπικών δεδομένων σε πραγματικό χρόνο, τα οποία συλλέγονται από τα ψηφιακά ίχνη καθημερινών συμπεριφορών ατόμων σε έναν πληθυσμό για την παροχή «εξατομικευμένων» υπηρεσιών σε μεμονωμένους χρήστες, συνήθως σε προγνωστική βάση. Γνωστά παραδείγματα περιλαμβάνουν αλγοριθμικές μηχανές σύστασης καταναλωτικών προϊόντων, όπως αυτές που παρέχονται από την Amazon, το σύστημα News Feed που χρησιμοποιείται από την Facebook/Meta για την προώθηση περιεχομένου πολυμέσων στους χρήστες και το σύστημα Up Next που χρησιμοποιεί αυτόματα το YouTube για την αναγνώριση και αναπαραγωγή περιεχομένου βίντεο αλγοριθμικά ταυτοποιημένου, ως σχετικό με το χρήστη. Αυτά είναι μόνο μερικά παραδείγματα συστημάτων Μηχανικής Μάθησης που χρησιμοποιούνται σε ένα πολύ ευρύ και ποικίλο φάσμα εμπορικών εφαρμογών, φαινομενικά σε βάση «δωρεάν υπηρεσιών», δηλαδή χωρίς να απαιτείται η καταβολή χρηματικής αμοιβής. Αντίθετα, προσφέρουν την υπηρεσία με αντάλλαγμα το δικαίωμα πρόσβασης, επεξεργασίας και επαναχρησιμοποίησης των προσωπικών δεδομένων του χρήστη (Acquisti et al., 2015). Αυτά τα συστήματα βασίζονται σε προηγμένες τεχνικές εξόρυξης δεδομένων για τη δημιουργία προφίλ ατόμων με βάση ομαδικά γνωρίσματα ή χαρακτηριστικά που προσδιορίζονται μέσω της αλγοριθμικής εξόρυξης δεδομένων από τα ψηφιακά τους ίχνη, τα οποία αποκτώνται με συνεχή και σε πραγματικό χρόνο παρακολούθηση ενός πολύ μεγάλου αριθμού ατόμων (Hildebrandt, 2010). Μπορούν να διαμορφωθούν ώστε να προσφέρουν εξατομικευμένες υπηρεσίες, οι οποίες να ταιριάζουν στα συναγόμενα ενδιαφέροντα και συνήθειες των χρηστών, παρέχοντας αποτελεσματικότητα και ευκολία για απρόσκοπτη πλοήγηση στον τεράστιο όγκο των ψηφιακών πληροφοριών και οι οποίες θα ήταν ουσιαστικά μη διαχειρίσιμες χωρίς το είδος των ισχυρών αλγοριθμικών εργαλείων που είναι τώρα διαθέσιμα (Yeung, 2019). Αν και απεικονίζεται από τους παρόχους ως προσφορά προς στους χρήστες μιας πιο «ουσιαστικής» εμπειρίας, η αλγοριθμική εξατομίκευση διαφέρει από την παραδοσιακή, καθώς η αλγοριθμική εξατομίκευση βασίζεται, συνήθως, στις προτιμήσεις και τα ενδιαφέροντα που έχει συναγάγει ο πάροχος υπηρεσιών για το άτομο, με τις υπηρεσίες

να προσφέρονται προαιρετικά, χωρίς αίτημα εξυπηρέτησης από το άτομο. Επειδή το άτομο δεν έχει δηλώσει ρητά τις προτιμήσεις και τα ενδιαφέροντά του σχετικά με την υπηρεσία (πράγματι, μπορεί να μην θέλει καθόλου την υπηρεσία), οι ηθικές ανησυχίες που σχετίζονται με την προγνωστική εξατομίκευση γίνονται εμφανείς μόλις παρακολουθήσουμε τον υποκείμενο στόχο αυτών των συστημάτων. Τι ακριβώς επιδιώκουν να βελτιστοποιήσουν αυτά τα συστήματα και ποιος έχει τη δύναμη να προσδιορίσει αυτόν τον πρωταρχικό στόχο; Αυτά τα συστήματα προορίζονται να διοχετεύουν τη συμπεριφορά και τις αποφάσεις του χρήστη προς την προτιμώμενη κατεύθυνση του κατόχου του συστήματος, και έτσι τα εμπορικά συστήματα διαμορφώνονται σκόπιμα για να βελτιστοποιούν όποιες μεταβλητές θα παράγουν τις μέγιστες εμπορικές αποδόσεις για τον κάτοχό τους. Επειδή αυτά τα συστήματα στοχεύουν, κυρίως, στη βελτιστοποίηση των μακροπρόθεσμων συμφερόντων του κατόχου του συστήματος, δεν υπάρχει καμία εγγύηση ότι ευθυγραμμίζονται με τα μακροπρόθεσμα συμφέροντα και την ευημερία των χρηστών των οποίων τις αποφάσεις και τις συμπεριφορές επιδιώκουν να επηρεάσουν (Yeung, 2012).

Παρόλο που τα άτομα διατηρούν επίσημα την ελευθερία να αποφασίσουν εάν θα καταναλώσουν αυτές τις υπηρεσίες, είναι σημαντικό να αναγνωρίσουμε τον ισχυρό, λεπτό και τυπικά υποσυνείδητο τρόπο με τον οποίο λειτουργούν αυτά τα συστήματα μέσω της εξάρτησης από τις τεχνικές προώθησης «nudging» (Thaler et al., 2008), οι οποίες επιδιώκουν να εκμεταλλεύονται τη συστηματική τάση των ατόμων να βασίζονται σε γνωστικές ευρετικές ή νοητικές συντομεύσεις στη λήψη αποφάσεων, αντί να καταλήγουν σε αυτές μέσω συνειδητής και αναστοχαστικής σκέψης. Όχι μόνο πολλές τεχνικές προώθησης είναι ηθικά προβληματικές επειδή μπορούν να θεωρηθούν ως χειραγωγικές και αδιαφανείς (Yeung, 2017), αλλά και λόγω της ικανότητάς τους να αναδιαμορφώνουν συνεχώς τις διαδικτυακές υπηρεσίες υπό το φως της ανατροφοδότησης σε πραγματικό χρόνο από άτομα που παρακολουθούνται ταυτόχρονα σε έναν πληθυσμό και σε ευρεία βάση, αυτό ενισχύει ποιοτικά τη χειραγωγική τους δύναμη, την οποία η Yeung ορίζει ως «hypernudging» «υπερπροώθηση» (Yeung, 2017). Λόγω της συντριπτικής κυριαρχίας του επιχειρηματικού μοντέλου «δωρεάν υπηρεσίες (free services)» για την παροχή διαδικτυακών υπηρεσιών στις οποίες οι πάροχοι κερδίζουν έσοδα από διαφημίσεις και εκμεταλλευόμενοι την αξία των δεδομένων των πελατών τους, αυτά τα αλγοριθμικά συστήματα, συνήθως, διαμορφώνονται για να μεγιστοποιούν τον χρόνο των χρηστών στον ιστότοπο, προκειμένου να μεγιστοποιηθούν οι αποδόσεις. Όσο περισσότερο καθυστερούν οι χρήστες, τόσο μεγαλύτερη είναι η ευκαιρία που έχει ο χρήστης να δει και να κάνει κλικ σε διαφημίσεις και τόσο πιο

εκτεταμένες είναι οι ευκαιρίες για τον πάροχο υπηρεσιών να συλλέξει και να αναλύσει περισσότερα δεδομένα από τη συνεχή παρακολούθηση της συμπεριφοράς του χρήστη. Πολλοί σχολιαστές έχουν επισημάνει, λοιπόν, πώς αυτά τα συστήματα ενθαρρύνουν εθιστικές συμπεριφορές, σχεδιασμένες να βελτιστοποιούν την αφοσίωση των χρηστών με τρόπους που συνήθως δεν είναι προς το μακροπρόθεσμο συμφέρον των ίδιων των χρηστών, αποσπώντας την προσοχή τους από τους δικούς τους προσωπικούς στόχους και έργα (Williams, 2018). Συνολικά, υπάρχει μια σημαντική ανησυχία ότι τα ολοένα και πιο έξυπνα περιβάλλοντά μας, τα οποία βασίζονται σε προγνωστικούς αλγορίθμους για τη διαμόρφωση προφίλ, εξατομικεύοντας το περιβάλλον μας και τις ευκαιρίες και τα εμπόδια που αναδύονται από τις ψηφιακές μας επαφές, ενδέχεται να μη σέβονται και να πλήττουν σημαντικά την ατομική μας ελευθερία και αυτονομία (Yeung, 2017. Zuboff, 2019).

Πληθυσμιακή παρακολούθηση δεδομένων

Όχι μόνο η συνεχής, συστηματική φύση και η εξαιρετικά επεμβατική ψηφιακή επιτήρηση («dataveillance») στην οποία βασίζεται το προφίλ συμπεριφοράς και ό,τι αυτές οι διαδικασίες συνήθως συνεπάγονται, εγείρουν ηθικούς προβληματισμούς, αλλά προκαλούν, επίσης, εξέταση μιας περαιτέρω σειράς ερωτημάτων, όπως: ποιος έχει ικανότητα συμμετοχής στην αλγοριθμική λήψη αποφάσεων; Πώς ασκείται αυτή η εξουσία, σε σχέση με ποιον και για ποιους σκοπούς; Ενώ οι συμβατικές αντιρρήσεις για την παρακολούθηση σε όλο τον πληθυσμό, όποια και αν είναι η τεχνολογική της μορφή, εκφράζονται συνήθως ως προς τον αντίκτυπό της στο δικαίωμα στην ιδιωτική ζωή, είναι αμφίβολο ότι οι ανησυχίες για την προστασία της ιδιωτικής ζωής μπορούν να ξεπεραστούν ικανοποιητικά στη βάση του ότι τα άτομα συναινούν στην ψηφιακή παρακολούθηση προκειμένου να επωφεληθούν «δωρεάν» ψηφιακές υπηρεσίες που καθιστά δυνατή την παροχή πρόσβασης στα προσωπικά τους δεδομένα. Όχι μόνο είναι γνωστές οι δυσεπίλυτες προκλήσεις που σχετίζονται με την απόκτηση ουσιαστικής συναίνεσης για την κοινή χρήση δεδομένων και το σχετικό «παράδοξο διαφάνειας» (Nissenbaum, 2011, Cranor et al., 2013), αλλά δεδομένου ότι η ανάπτυξη και εφαρμογή αυτών των συστημάτων βρίσκεται σχεδόν αποκλειστικά στα χέρια μεγάλων, πλούσιων και ισχυρών οργανισμών, αυτό δημιουργεί χρόνιες ασυμμετρίες που προκύπτουν μεταξύ των μεγάλων οργανισμών (τόσο εμπορικών όσο και κυβερνητικών) που χρησιμοποιούν αυτές τις τεχνολογίες για να καθοδηγούν, να ενημερώνουν και να εκτελούν αποφάσεις και των ατόμων για τα οποία λαμβάνονται οι αποφάσεις. Ακόμη και τα πιο αβλαβή δεδομένα σχετικά με τις καθημερινές μας δραστηριότητες, όταν συγκεντρώνονται και εξορύσσονται αλγοριθμικά, μπορεί να αποκαλύπτουν με υψηλούς βαθμούς ακρίβειας

πολύ προσωπικές πληροφορίες για εμάς, συμπεριλαμβανομένου του φύλου, του σεξουαλικού μας προσανατολισμού, θρησκευτικών και πολιτικών πεποιθήσεων, εθνικότητας και ούτω καθεξής. Στις εν λόγω προσωπικές πληροφορίες έχουν πρόσβαση μόνο μεγάλοι και ισχυροί οργανισμοί, γεγονός που δημιουργεί σοβαρούς κινδύνους αυτές οι πληροφορίες να χρησιμοποιηθούν εναντίον μας με τρόπους που μπορεί να είναι δύσκολο, αν όχι αδύνατο, να εντοπιστούν και τους οποίους αγνοούμε παντελώς (Kosinski et al., 2013). Αυτή η ικανότητα να βγάζουν πιθανολογικά, αλλά εύλογα και αξιόπιστα συμπεράσματα σχετικά με τις οικείες πτυχές του εαυτού μας, τα οποία επηρεάζουν τις αποφάσεις που λαμβάνονται για εμάς σε ένα ευρύ φάσμα τομέων, όχι μόνο συνιστά ξεκάθαρη παραβίαση του ατομικού μας απορρήτου, αλλά είναι πιθανό να αλλάξει και την αίσθηση της έννοιας της ελευθερίας μας που περιλαμβάνει το να πειραματιζόμαστε και να αλλάζουμε αυτό που είμαστε. Με άλλα λόγια, η διάχυτη επιτήρηση στην οποία βασίζεται το προφίλ συμπεριφοράς μπορεί να έχει διαβρωτική επίδραση στην ικανότητά μας να ασκούμε τα ανθρώπινα δικαιώματα και τις θεμελιώδεις ελευθερίες μας, ιδιαίτερα την ικανότητά μας να συμμετέχουμε σε δραστηριότητες μέσω των οποίων μπορούμε να αναπτύξουμε την αίσθηση του εαυτού μας μέσω των δημιουργικών μορφών δράσης, έκφρασης και πειραματισμού μέσα σε ένα περιβάλλον στο οποίο έχουμε εμπιστοσύνη ότι δεν θα παρατηρηθεί από άλλους με τρόπους που θα μπορούσαν να αυξήσουν την έκθεσή μας στις πιθανές αρνητικές επιπτώσεις της αλγοριθμικής λήψης αποφάσεων. (Richards, 2013).

Όχι μόνο η διάχυτη παρακολούθηση και η πρόβλεψη συμπεριφοράς απειλούν την αυτονομία και το δικαίωμα αυτοδιάθεσής μας (Hildebrandt, 2015), αλλά όταν η εξατομικευμένη πρόγνωση κλιμακώνεται και εφαρμόζεται σε ευρεία πληθυσμιακή βάση, μπορεί να θεωρηθεί αντίθετη προς την αρχή της καθολικής και ίσης μεταχείρισης που έχουμε συνδέσει με το σύγχρονο ιδεώδες του κράτους δικαίου, στο οποίο εφαρμόζονται τα ίδια γενικά νομικά πρότυπα σε όλους τους νομικά ικανούς ενήλικες ως βάση για την αξιολόγηση της προηγούμενης συμπεριφοράς, ανεξαρτήτως θέσης, πλούτου ή άλλων ιδιοσυγκρασιακών πεποιθήσεων ή τάσεων. Επομένως, η χρήση τεχνικών αυτοματοποιημένης εξατομίκευσης εγείρει ακανθώδη ερωτήματα σχετικά με το εάν, με ποιους τρόπους και υπό ποιες συνθήκες, η εξατομίκευση μπορεί να θεωρηθεί συμβατή ή ανταγωνιστική με το βασικό δικαίωμα όλων των ανθρώπινων όντων να αντιμετωπίζονται ως ίσοι, με την ίδια μέριμνα και τον ίδιο σεβασμό (Yeung, 2012, 2018b). Ο εντοπισμός των ορίων της αποδεκτής εξατομίκευσης θα μπορούσε να μετριάσει τις δυσμενείς επιπτώσεις που σχετίζονται με την προγνωστική εξατομίκευση του ψηφιακού περιβάλλοντος των ατόμων. Τα όρια είναι ανάλογα των

επιπτώσεων που επιφέρουν. Κάποιες επιπτώσεις, όπως αυτές επισημάνθηκαν πρόσφατα από τη χρήση «ψευδών ειδήσεων» και άλλης παραπληροφόρησης από μέσα κοινωνικής δικτύωσης για το δημοψήφισμα του Ηνωμένου Βασιλείου σχετικά με το Brexit και για τις προεδρικές εκλογές του 2016 στις ΗΠΑ (Wardle et al., 2017) είναι πιθανό να διαφέρουν αρκετά από αυτές που σχετίζονται με τη χρήση προγνωστικής αστυνόμευσης από τις αρχές επιβολής του νόμου (Ferguson, 2017), την εξατομικευμένη τιμολόγηση που χρησιμοποιείται από διαδικτυακούς λιανοπωλητές (Townley et al., 2017) ή την παροχή αυτοματοποιημένων νομικών συμβουλών σε εξατομικευμένη βάση (Casey et al., 2016). Η ανησυχία, ωστόσο, είναι εντονότερη όταν οι τεχνικές αυτοματοποιημένης εξατομίκευσης χρησιμοποιούνται συστηματικά σε πληθυσμιακή βάση, έχοντας τη δυνατότητα να διαβρώσουν τα κοινωνικά θεμέλια της ατομικής αυτονομίας, της κοινωνικής αλληλεγγύης και των δημοκρατικών θεσμών (Yeung, 2018b).

5. Ηθική και Δεοντολογία στο Δημόσιο Τομέα

Το προηγούμενο Κεφάλαιο έχει επισημάνει ένα ευρύ φάσμα ηθικών ανησυχιών που σχετίζονται με τη χρήση της αλγοριθμικής λήψης αποφάσεων. Προκειμένου να ανταποκριθούν αποτελεσματικά σε αυτές της ανησυχίες οι υπεύθυνοι χάραξης πολιτικής έχουν στραφεί σε ρυθμιστικά εργαλεία και εργαλεία πολιτικής, ελπίζοντας να εξασφαλίσουν ηθική και αξιόπιστη Τεχνητή Νοημοσύνη σε όλες τις χώρες και τα πλαίσια. Αυτές οι απαντήσεις εξελίσσονται γρήγορα και ποικίλλουν σε μεγάλο βαθμό ως προς τη μορφή και την ουσία, από νομικά δεσμευτικές δεσμεύσεις έως αρχές και κατευθυντήριες γραμμές υψηλού επιπέδου. Οι πολιτικές που διατυπώνουν αρχές ή κατευθυντήριες γραμμές υψηλού επιπέδου γενικά δεν προορίζονται να είναι δεσμευτικές και εκδίδονται ως κανονιστικά πρότυπα, βάσει των οποίων οι φορείς μπορούν να αξιολογήσουν τη δικής τους χρήση αλγοριθμικών συστημάτων. Συχνά, ως μη δεσμευτικές και αυτόνομες αρχές, δε δημιουργούν εκτελεστές υποχρεώσεις, αλλά μπορούν να παρέχουν χρήσιμα βοηθήματα και καθοδήγηση σε δημόσιους φορείς που προβληματίζονται σχετικά με την κατάλληλη χρήση αλγοριθμικών συστημάτων. Επίσης, μπορούν να χρησιμεύσουν ως δηλώσεις προθέσεων σχετικά με ευρύτερους στόχους των διοικήσεων στην ανάπτυξη δημόσιας πολιτικής για αλγοριθμικά συστήματα. Οι «Κατευθυντήριες Γραμμές Δεοντολογίας για Αξιόπιστη Τεχνητή Νοημοσύνη» της Ευρωπαϊκής Ένωσης και η «Σύσταση για την ηθική της Τεχνητής Νοημοσύνης» της UNESCO, αν και δεν είναι δεσμευτικές, προσπαθούν να αντιμετωπίσουν τις ανησυχίες που προκύπτουν από τη χρήση αλγοριθμικών συστημάτων, δίνοντας κατευθύνσεις που μπορούν να συμβάλλουν ακριβώς στην επίτευξη ηθικής και αξιόπιστης Τεχνητής Νοημοσύνης.

5.1 «Κατευθυντήριες Γραμμές Δεοντολογίας για Αξιόπιστη Τεχνητή Νοημοσύνη» - Ευρωπαϊκή Επιτροπή

Στις 8 Νοεμβρίου 2019, ανακοινώθηκαν από την Ευρωπαϊκή Επιτροπή οι «Κατευθυντήριες Γραμμές Δεοντολογίας για Αξιόπιστη Τεχνητή Νοημοσύνη», οι οποίες εκπονήθηκαν από την ανεξάρτητη Ομάδα Εμπειρογνωμόνων Υψηλού Επιπέδου για την Τεχνητή Νοημοσύνη.

Οι Κατευθυντήριες Γραμμές επιδιώκουν να συμβάλλουν στην επίτευξη του οράματος της Ευρωπαϊκής Επιτροπής για μια «δεοντολογική, ασφαλή και προηγμένη Τεχνητή Νοημοσύνη με τη σφραγίδα της Ευρώπης»,[8] εστιασμένης στον άνθρωπο και προσανατολισμένης να λειτουργήσει ως μέσο που μπορεί να προάγει την ατομική και κοινωνική ευημερία, το κοινό

[8] COM (2018) 237, COM (2018) 795

καλό και να συντελέσει στην πρόοδο και την ευημερία, μεγιστοποιώντας τα οφέλη των συστημάτων Τεχνητής Νοημοσύνης και προλαμβάνοντας και ελαχιστοποιώντας τους κινδύνους τους. Οι Κατευθυντήριες Γραμμές δεν είναι δεσμευτικές. Δεν γεννούν νομικά δικαιώματα ούτε επιβάλλουν νομικές υποχρεώσεις σε τρίτους. Η συμβολή τους έγκειται στη θέσπιση ενός πλαισίου για την επίτευξη αξιόπιστης Τεχνητής Νοημοσύνης βασισμένο σε τρεις (3) συνιστώσες, οι οποίες θα πρέπει να πληρούνται σε ολόκληρο τον κύκλο ζωής του συστήματος. Αξιόπιστη Τεχνητή Νοημοσύνη νοείται αυτή που είναι: (α) σύννομη, δηλαδή, τηρεί τις νομοθετικές και κανονιστικές διατάξεις, (β) δεοντολογική, δηλαδή, διασφαλίζει τη συμμόρφωση με δεοντολογικές αρχές και αξίες και (γ) στιβαρή, δηλαδή, κατάλληλη τεχνικά να ανταποκριθεί σε οποιαδήποτε φάση του κύκλου ζωή του συστήματος και ικανή να αντιλαμβάνεται το κοινωνικό πλαίσιο και το περιβάλλον εντός του οποίου λειτουργεί το σύστημα, αποφεύγοντας έστω και ακούσια να προκαλέσει βλάβη (High Level Expert Group on AI, 2019).

Η αξιοπιστία αποτελεί προϋπόθεση για την ανάπτυξη, εγκατάσταση και χρήση συστημάτων Τεχνητής Νοημοσύνης από τον άνθρωπο και τις κοινωνίες. Η αξιοπιστία δεν αφορά μόνον την αξιοπιστία των ίδιων των συστημάτων, αλλά και την αξιοπιστία όλων των παραγόντων και των διεργασιών που αποτελούν μέρος του κοινωνικοτεχνικού περιβάλλοντος των συστημάτων καθόλη τη διάρκεια του κύκλου ζωής τους. Εάν, λοιπόν, τα συστήματα Τεχνητής Νοημοσύνης δεν είναι σε θέση να αποδείξουν ότι είναι άξια εμπιστοσύνης, μπορεί να αμφισβητηθούν και να μη γίνουν αποδεκτά, με αποτέλεσμα να μην καταστεί δυνατή η αξιοποίηση των δυνητικά τεράστιων κοινωνικών και οικονομικών οφελών που μπορούν να επιφέρουν. Η Ευρώπη, θέλοντας να αξιοποιήσει αυτά τα οφέλη και να προωθήσει την υπεύθυνη και βιώσιμη καινοτομία στον τομέα της Τεχνητής Νοημοσύνης, θέτει τη δεοντολογία ως βασικό πυλώνα διασφάλισης της αξιοπιστίας της Τεχνητής Νοημοσύνης. Οι Κατευθυντήριες Γραμμές επιδιώκουν την ανάπτυξη μιας ενιαίας προσέγγισης της Τεχνητής Νοημοσύνης με εστίαση στην δεοντολογία, η οποία θα επιχειρεί να ωφελεί, να ενισχύει και να προστατεύει τόσο την ατομική ευημερία του ανθρώπου όσο και το κοινό καλό της κοινωνίας (High Level Expert Group on AI, 2019).

Πλαίσιο για αξιόπιστη Τεχνητή Νοημοσύνη

Όπως προαναφέραμε, το πλαίσιο για την επίτευξη αξιόπιστης Τεχνητής Νοημοσύνης στηρίζεται σε τρεις συνιστώσες, οι οποίες θα πρέπει να συνεργάζονται και να αλληλεπιδρούν προκειμένου να προάγουν και να υπερασπίζονται τα θεμελιώδη δικαιώματα, όπως αυτά

κατοχυρώνονται στον Χάρτη των Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης και στο διεθνές δίκαιο των ανθρωπίνων δικαιωμάτων.

1η Συνιστώσα: Σύννομη Τεχνητή Νοημοσύνη

Τα συστήματα Τεχνητής Νοημοσύνης λειτουργούν βάσει υφιστάμενης νομοθεσίας που προβλέπει τόσο θετικές όσο και αρνητικές υποχρεώσεις, υπό την έννοια ότι απαγορεύει συγκεκριμένες ενέργειες, ενώ επιτρέπει κάποιες άλλες. Διέπονται από μια σειρά νομικά δεσμευτικούς κανόνες σε ευρωπαϊκό, εθνικό και διεθνές επίπεδο και εφαρμόζονται σε όλες της φάσεις του κύκλου ζωής τους, όπως οι Συνθήκες της Ευρωπαϊκής Ένωσης και ο Χάρτης Θεμελιωδών Δικαιωμάτων, ο Γενικός Κανονισμός για την Προστασία Προσωπικών Δεδομένων, οι Συνθήκες του ΟΗΕ για τα ανθρώπινα δικαιώματα, η Ευρωπαϊκή Σύμβαση Ανθρωπίνων Δικαιωμάτων, μια σειρά νόμων των κρατών μελών της Ευρωπαϊκής Ένωσης και εξειδικευμένοι κανόνες για συγκεκριμένες εφαρμογές Τεχνητής Νοημοσύνης. Ωστόσο, η συγκεκριμένη συνιστώσα δεν αναλύεται. Για τις Κατευθυντήριες Γραμμές θεωρείται δεδομένη η τήρηση της νομοθεσίας σχετικά με την ανάπτυξη, εγκατάσταση και χρήση συστημάτων Τεχνητής Νοημοσύνης γιατί στόχος των Κατευθυντήριων Γραμμών δεν είναι η παροχή νομικών συμβουλών ή κατευθύνσεων σχετικά με πώς μπορεί να επιτυγχάνεται η συμμόρφωση με εφαρμοστέους νομικούς κανόνες και απαιτήσεις, αλλά η παροχή κατευθύνσεων όσον αφορά στην προαγωγή και στη διασφάλιση της δεοντολογικής και στιβαρής Τεχνητής Νοημοσύνης (High Level Expert Group on AI, 2019).

2η Συνιστώσα: Δεοντολογική Τεχνητή Νοημοσύνη

Για την επίτευξη αξιόπιστης Τεχνητής Νοημοσύνης δεν αρκεί η τήρηση της νομοθεσίας, καθώς αποτελεί τη μία μόνο από τις τρεις συνιστώσες της. Η νομοθεσία δεν συμβαδίζει πάντοτε με τις τεχνολογικές εξελίξεις, ενίοτε δεν είναι ευθυγραμμισμένη με τους κανόνες δεοντολογίας ή, απλώς, δεν είναι κατάλληλη για την αντιμετώπιση ορισμένων ζητημάτων. Επομένως, για να είναι αξιόπιστα, τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να είναι δεοντολογικά, ώστε να διασφαλίζεται η ευθυγράμμισή τους με τους κανόνες δεοντολογίας (High Level Expert Group on AI, 2019).

3η Συνιστώσα: Στιβαρή Τεχνητή Νοημοσύνη

Ακόμη κι αν διασφαλίζεται η ύπαρξη δεοντολογικού σκοπού, τα άτομα και η κοινωνία θα πρέπει να έχουν και τη βεβαιότητα ότι τα συστήματα Τεχνητής Νοημοσύνης δεν θα προκαλέσουν ακούσια βλάβη. Τα συστήματα αυτά θα πρέπει να λειτουργούν με ασφαλή και

αξιόπιστο τρόπο, και θα πρέπει να προβλέπονται ασφαλιστικές δικλείδες που θα προλαμβάνουν τις όποιες ακούσιες δυσμενείς επιπτώσεις. Επομένως, είναι σημαντικό να διασφαλίζεται ότι τα συστήματα Τεχνητής Νοημοσύνης είναι στιβαρά. Αυτό είναι αναγκαίο τόσο από τεχνικής άποψης (ώστε να διασφαλίζεται η τεχνική στιβαρότητα του συστήματος με τον ενδεδειγμένο τρόπο σε δεδομένο πλαίσιο, όπως ο τομέας εφαρμογής ή η φάση του κύκλου ζωής), όσο και από κοινωνικής άποψης (λαμβανομένου δεόντως υπόψη του πλαισίου και του περιβάλλοντος στο οποίο λειτουργεί το σύστημα). Η δεοντολογική και η στιβαρή Τεχνητής Νοημοσύνης είναι, επομένως, στενά συνυφασμένες και αλληλοσυμπληρούμενες (High Level Expert Group on AI, 2019).

Διάρθρωση Πλαισίου

Το πλαίσιο διαρθρώνεται σε τρία (3) επίπεδα καθένα από τα οποία παρέχει κατευθύνσεις για την επίτευξη αξιόπιστης Τεχνητής Νοημοσύνης. Το πρώτο επίπεδο αφορά στις βάσεις της αξιόπιστης Τεχνητής Νοημοσύνης, το δεύτερο αναφέρεται στην πραγμάτωσή της και το τρίτο στην αξιολόγησή της, τα οποία και θα αναλυθούν περαιτέρω.

1ο Επίπεδο Πλαισίου – Οι βάσεις της αξιόπιστης Τεχνητής Νοημοσύνης

Οι βάσεις της αξιόπιστης Τεχνητής Νοημοσύνης εδράζονται στα θεμελιώδη δικαιώματα και αποτυπώνονται με τέσσερις δεοντολογικές αρχές που θα πρέπει να τηρούνται προκειμένου να διασφαλιστεί η δεοντολογική και στιβαρή Τεχνητή Νοημοσύνη. Κύριο μέλημα της δεοντολογίας της Τεχνητής Νοημοσύνης είναι να προσδιορίσει τον τρόπο με τον οποίο η Τεχνητή Νοημοσύνη μπορεί να προωθήσει ή να εγείρει προβληματισμούς για την ευημερία των ατόμων, είτε από την άποψη της ποιότητας ζωής είτε της ανθρώπινης αυτονομίας και ελευθερίας που είναι απαραίτητες για μια δημοκρατική κοινωνία. Όπως συμβαίνει με κάθε ισχυρή τεχνολογία, η χρήση συστημάτων Τεχνητής Νοημοσύνης στην κοινωνία μας θέτει αρκετές δεοντολογικές προκλήσεις, παραδείγματος χάρη σχετικά με τις επιπτώσεις τους στον άνθρωπο και την κοινωνία, τις δυνατότητες λήψης αποφάσεων και την ασφάλεια. Εάν πρόκειται να χρησιμοποιούμε ολοένα και περισσότερο τη βοήθεια των συστημάτων Τεχνητής Νοημοσύνης ή να μεταβιβάζουμε τη λήψη αποφάσεων σε αυτά, θα πρέπει να διασφαλίζουμε ότι τα συγκεκριμένα συστήματα έχουν δίκαιες επιπτώσεις στη ζωή των ανθρώπων, ότι συμμορφώνονται με μη διαπραγματεύσιμες αξίες και ότι είναι σε θέση να ενεργήσουν ανάλογα, καθώς και ότι κατάλληλες διαδικασίες λογοδοσίας μπορούν να το διασφαλίζουν αυτό. Επομένως, μια προσέγγιση της δεοντολογίας της Τεχνητής Νοημοσύνης βασισμένη στο σεβασμό των θεμελιωδών δικαιωμάτων, μέσα σε ένα πλαίσιο δημοκρατίας

και κράτους δικαίου, προσφέρει τις πιο ελπιδοφόρες βάσεις για τον προσδιορισμό δεοντολογικών αρχών και αξιών που είναι δυνατό να υλοποιηθούν επιχειρησιακά στο πλαίσιο της Τεχνητής Νοημοσύνης. Τα συγκεκριμένα δικαιώματα περιγράφονται στον Χάρτη της ΕΕ και αφορούν στην αξιοπρέπεια, στις ελευθερίες, στην ισότητα και στην αλληλεγγύη, στα δικαιώματα των πολιτών και στη δικαιοσύνη. Η κοινή βάση που συνενώνει τα εν λόγω δικαιώματα μπορεί να θεωρηθεί ότι εδράζεται στον σεβασμό για την ανθρώπινη αξιοπρέπεια, αποτυπώνοντας αυτό που περιγράφεται ως «ανθρωποκεντρική προσέγγιση», σύμφωνα με την οποία ο άνθρωπος απολαμβάνει τη μοναδική και αναφαίρετη ηθική υπόσταση υπεροχής στους τομείς των ατομικών, πολιτικών, οικονομικών και κοινωνικών δικαιωμάτων. Ως νομικώς ισχυρά δικαιώματα, τα θεμελιώδη δικαιώματα εμπίπτουν στην πρώτη συνιστώσα της αξιόπιστης Τεχνητής Νοημοσύνης (σύννομη Τεχνητή Νοημοσύνη), η οποία διασφαλίζει τη συμμόρφωση με τον νόμο. Ως δικαιώματα με καθολικό χαρακτήρα, που έχουν τις ρίζες τους στην εγγενή ηθική υπόσταση των ανθρώπων, τα εν λόγω δικαιώματα στηρίζουν επίσης τη δεύτερη συνιστώσα της αξιόπιστης Τεχνητής Νοημοσύνης (δεοντολογική Τεχνητή Νοημοσύνη) και αφορούν τους δεοντολογικούς κανόνες που δεν είναι απαραίτητα νομικώς δεσμευτικοί, αλλά έχουν καίρια σημασία για την εξασφάλιση της αξιοπιστίας (High Level Expert Group on AI, 2019).

Θεμελιώδη δικαιώματα

Από τη συνολική δέσμη δικαιωμάτων που ορίζονται στο διεθνές δίκαιο των ανθρωπίνων δικαιωμάτων, στις Συνθήκες της ΕΕ και στον Χάρτη της ΕΕ, οι ακόλουθες οικογένειες θεμελιωδών δικαιωμάτων είναι ιδιαιτέρως πρόσφορες για να καλύψουν τα συστήματα Τεχνητής Νοημοσύνης.

- **Σεβασμός της ανθρώπινης αξιοπρέπειας**

  Στο πλαίσιο της Τεχνητής Νοημοσύνης, ο σεβασμός της ανθρώπινης αξιοπρέπειας συνεπάγεται ότι όλοι οι άνθρωποι αντιμετωπίζονται με σεβασμό ως ηθικά υποκείμενα, και όχι απλώς αντικείμενα που υφίστανται διαλογή, ταξινόμηση, χαρακτηρισμό, αγελοποίηση, κοινωνικό προγραμματισμό ή χειραγώγηση. Επομένως, τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να αναπτυχθούν κατά τέτοιον τρόπο ώστε να σέβονται, να εξυπηρετούν και να προστατεύουν τη σωματική και πνευματική ακεραιότητα του ανθρώπου, την προσωπική και πολιτισμική αίσθηση της ταυτότητας και την ικανοποίηση των βασικών αναγκών του (High Level Expert Group on AI, 2019).

- **Ελευθερία του ατόμου**

  Στο πλαίσιο της Τεχνητής Νοημοσύνης, η ελευθερία του ατόμου απαιτεί μετριασμό του άμεσου και έμμεσου παράνομου εξαναγκασμού, των απειλών στην ψυχική αυτονομία και ψυχική υγεία, της αδικαιολόγητης επιτήρησης, της εξαπάτησης και της αθέμιτης χειραγώγησης. Στην πραγματικότητα, η ελευθερία του ατόμου συνεπάγεται τη δέσμευση να δοθεί στα άτομα η δυνατότητα να ασκούν ακόμη μεγαλύτερο έλεγχο στη ζωή τους, συμπεριλαμβανομένης της προστασίας (μεταξύ άλλων δικαιωμάτων) της επιχειρηματικής ελευθερίας, της ελευθερίας της τέχνης και της επιστήμης, της ελευθερίας έκφρασης, της ιδιωτικής ζωής και της προστασίας προσωπικών δεδομένων, καθώς και της ελευθερίας του συνέρχεσθαι και του συνεταιρίζεσθαι (European Commission, 2019).

- **Σεβασμός της δημοκρατίας, της δικαιοσύνης και του κράτους δικαίου**

  Τα συστήματα Τεχνητής Νοημοσύνης δεν θα πρέπει να υπονομεύουν τις δημοκρατικές διαδικασίες, τις ανθρώπινες διαβουλεύσεις ή τα συστήματα δημοκρατικής ψηφοφορίας, αλλά θα πρέπει να ενσωματώνουν τη δέσμευση να διασφαλίσουν ότι δεν λειτουργούν με τρόπους που υπονομεύουν τις θεμελιώδεις δεσμεύσεις στις οποίες στηρίζεται το κράτος δικαίου, τους υποχρεωτικούς νόμους και κανονισμούς, καθώς επίσης και να διασφαλίσουν τις δέουσες διαδικασίες και την ισότητα ενώπιον του νόμου (High Level Expert Group on AI, 2019).

- **Ισότητα, απαγόρευση των διακρίσεων και αλληλεγγύη**

  Στο πλαίσιο της Τεχνητής Νοημοσύνης, θα πρέπει να διασφαλίζεται ο ισότιμος σεβασμός της ηθικής αξίας και της αξιοπρέπειας όλων των ανθρώπων. Η ισότητα συνεπάγεται ότι οι λειτουργίες του συστήματος δε θα οδηγούν σε αποτελέσματα που χαρακτηρίζονται από αθέμιτη μεροληψία (π.χ. τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση των συστημάτων Τεχνητής Νοημοσύνης θα πρέπει να είναι όσο το δυνατόν πιο συμπεριληπτικά, αντιπροσωπεύοντας διαφορετικές ομάδες πληθυσμού). Το γεγονός αυτό απαιτεί επίσης τον κατάλληλο σεβασμό των δυνητικά ευάλωτων ατόμων και ομάδων, όπως οι εργαζόμενοι, οι γυναίκες, τα άτομα με αναπηρίες, οι εθνοτικές μειονότητες, τα παιδιά, οι καταναλωτές ή άλλα άτομα που αντιμετωπίζουν κίνδυνο αποκλεισμού (High Level Expert Group on AI, 2019).

- **Δικαιώματα πολιτών**

  Οι πολίτες επωφελούνται από ένα ευρύ φάσμα δικαιωμάτων, μεταξύ των οποίων το δικαίωμα ψήφου, το δικαίωμα στη χρηστή διοίκηση ή την πρόσβαση σε δημόσια έγγραφα και το δικαίωμα υποβολής αναφοράς στη διοίκηση. Αν και τα συστήματα

Τεχνητής Νοημοσύνης προσφέρουν σημαντικές δυνατότητες βελτίωσης της κλίμακας και της αποδοτικότητας της διακυβέρνησης όσον αφορά την παροχή δημόσιων αγαθών και υπηρεσιών προς την κοινωνία, θα πρέπει να διασφαλίζονται τα δικαιώματα των πολιτών που είναι δυνατόν να επηρεαστούν αρνητικά από την εφαρμογή τους (High Level Expert Group on AI, 2019).

Δεοντολογικές αρχές με βάση τα θεμελιώδη δικαιώματα

Τα προαναφερόμενα θεμελιώδη δικαιώματα πλαισιώνονται από τέσσερεις (4) δεοντολογικές αρχές που πρέπει να τηρούνται προκειμένου να διασφαλίζεται ότι τα συστήματα Τεχνητής Νοημοσύνης αναπτύσσονται, εγκαθίστανται και χρησιμοποιούνται με αξιόπιστο τρόπο.

- **Η αρχή του σεβασμού της ανθρώπινης αυτονομίας**

  Τα θεμελιώδη δικαιώματα στα οποία βασίζεται η Ευρωπαϊκή Ένωση έχουν γνώμονα τον σεβασμό της ελευθερίας και της αυτονομίας των ανθρώπων. Οι άνθρωποι που αλληλεπιδρούν με τα συστήματα Τεχνητής Νοημοσύνης πρέπει να είναι σε θέση να διατηρούν πλήρη και αποτελεσματική αυτοδιάθεση και να είναι σε θέση να συμμετέχουν στη δημοκρατική διαδικασία. Τα συστήματα Τεχνητής Νοημοσύνης δε θα πρέπει αδικαιολόγητα να προβαίνουν σε εξαναγκασμό, εξαπάτηση, χειραγώγηση ή αγελοποίηση των ανθρώπων, αλλά θα πρέπει να σχεδιάζονται έτσι ώστε να αυξάνουν, να συμπληρώνουν και να ενισχύουν τις ανθρώπινες γνωστικές, κοινωνικές και πολιτισμικές δεξιότητες. Η κατανομή των λειτουργιών μεταξύ των ανθρώπων και των συστημάτων Τεχνητής Νοημοσύνης θα πρέπει να ακολουθεί τις αρχές του ανθρωποκεντρικού σχεδιασμού και να προσφέρει ουσιαστικές ευκαιρίες για ανθρώπινη επιλογή. Αυτό σημαίνει εξασφάλιση της ανθρώπινης εποπτείας και ελέγχου των διαδικασιών εργασίας στα συστήματα Τεχνητής Νοημοσύνης. Επίσης, τα συστήματα Τεχνητής Νοημοσύνης ενδέχεται να μετασχηματίσουν ουσιαστικά τον τομέα της εργασίας. Θα πρέπει να στηρίζουν τους ανθρώπους στο εργασιακό περιβάλλον και να αποσκοπούν στη δημιουργία ουσιαστικού έργου (High Level Expert Group on AI, 2019).

- **Η αρχή της πρόληψης βλάβης**

  Τα συστήματα Τεχνητής Νοημοσύνης δεν θα πρέπει ούτε να προκαλούν ούτε να επιδεινώνουν τυχόν βλάβες ή να επηρεάζουν με άλλο τρόπο αρνητικά τον άνθρωπο. Αυτό συνεπάγεται την προστασία της ανθρώπινης αξιοπρέπειας, καθώς και της πνευματικής και σωματικής ακεραιότητας. Τα συστήματα Τεχνητής Νοημοσύνης και

τα περιβάλλοντα στα οποία λειτουργούν θα πρέπει να είναι ασφαλή και προστατευμένα. Θα πρέπει να είναι στιβαρά από τεχνικής άποψης, ενώ θα πρέπει να διασφαλίζεται ότι δεν είναι ανοικτά σε κακόβουλη χρήση. Θα πρέπει να δοθεί περισσότερη προσοχή στα ευάλωτα άτομα και να συμπεριληφθούν στην ανάπτυξη και την εγκατάσταση συστημάτων Τεχνητής Νοημοσύνης. Επίσης, ιδιαίτερη προσοχή θα πρέπει να δοθεί σε καταστάσεις όπου τα συστήματα Τεχνητής Νοημοσύνης είναι δυνατό να προκαλέσουν ή να επιδεινώσουν αρνητικές επιπτώσεις λόγω ασυμμετρίας εξουσίας ή πληροφόρησης, όπως μεταξύ εργοδοτών και εργαζομένων, επιχειρήσεων και καταναλωτών ή κυβερνήσεων και πολιτών. Επιπλέον, η πρόληψη βλάβης συνεπάγεται την εξέταση του φυσικού περιβάλλοντος και όλων των έμβιων όντων (High Level Expert Group on AI, 2019).

- **Η αρχή της δικαιοσύνης**

  Η ανάπτυξη, η εγκατάσταση και η χρήση συστημάτων Τεχνητής Νοημοσύνης θα πρέπει να γίνεται με τρόπο δίκαιο. Η δικαιοσύνη έχει τόσο ουσιαστική όσο και διαδικαστική διάσταση. Η ουσιαστική διάσταση συνεπάγεται δέσμευση για εξασφάλιση ισότιμης και δίκαιης κατανομής τόσο των ωφελειών όσο και του κόστους, καθώς και εξασφάλιση ότι τα άτομα και οι ομάδες δεν υφίστανται αθέμιτη μεροληψία, διακρίσεις και στιγματισμό. Εάν καταστεί δυνατό να αποφευχθεί η αθέμιτη μεροληψία, τα συστήματα Τεχνητής Νοημοσύνης μπορεί ακόμη και να αυξήσουν την κοινωνική δικαιοσύνη. Επίσης, θα πρέπει να προωθηθούν οι ίσες ευκαιρίες όσον αφορά την πρόσβαση στην εκπαίδευση, τα αγαθά, τις υπηρεσίες και την τεχνολογία. Επιπλέον, η χρήση συστημάτων Τεχνητής Νοημοσύνης δεν θα πρέπει ποτέ να οδηγεί στην εξαπάτηση των χρηστών ή τον περιορισμό της ελευθερίας επιλογής τους. Επιπλέον, η δικαιοσύνη συνεπάγεται ότι οι επαγγελματίες του τομέα της Τεχνητής Νοημοσύνης θα πρέπει να σέβονται την αρχή της αναλογικότητας μεταξύ μέσων και σκοπών και να εξετάζουν προσεκτικά τον τρόπο εξισορρόπησης ανταγωνιστικών συμφερόντων και στόχων. Η διαδικαστική διάσταση της δικαιοσύνης συνεπάγεται τη δυνατότητα αμφισβήτησης και αποτελεσματικής έννομης προστασίας έναντι αποφάσεων που λαμβάνονται από συστήματα Τεχνητής Νοημοσύνης και από τους ανθρώπους που τα χειρίζονται. Προκειμένου να γίνει αυτό, η οντότητα που είναι υπεύθυνη για την απόφαση θα πρέπει να είναι αναγνωρίσιμη, ενώ θα πρέπει να επεξηγούνται οι διαδικασίες λήψης αποφάσεων (High Level Expert Group on AI, 2019).

- **Η αρχή της επεξηγησιμότητας**

  Η επεξηγησιμότητα είναι καθοριστική για την οικοδόμηση και διατήρηση της εμπιστοσύνης των χρηστών έναντι των συστημάτων Τεχνητής Νοημοσύνης. Αυτό σημαίνει ότι οι διαδικασίες πρέπει να είναι διαφανείς, οι δυνατότητες και ο σκοπός των συστημάτων Τεχνητής Νοημοσύνης να κοινοποιούνται ανοικτά και οι αποφάσεις, στο μέτρο του δυνατού, να επεξηγούνται στους άμεσα και έμμεσα επηρεαζόμενους. Χωρίς αυτές τις πληροφορίες, μια απόφαση δεν είναι δυνατό να αμφισβητηθεί δεόντως. Δεν είναι πάντα εφικτή η επεξήγηση των λόγων για τους οποίους ένα μοντέλο έχει οδηγήσει σε ένα συγκεκριμένο αποτέλεσμα ή απόφαση. Αυτές οι περιπτώσεις αναφέρονται ως αλγόριθμοι «μαύρου κουτιού» και απαιτούν ιδιαίτερη προσοχή. Υπό αυτές τις συνθήκες, ενδέχεται να απαιτούνται άλλα μέτρα επεξηγησιμότητας (π.χ. ιχνηλασιμότητα, ελεγξιμότητα και διαφανής επικοινωνία σχετικά με τις δυνατότητες του συστήματος), υπό την προϋπόθεση ότι το σύστημα στο σύνολό του σέβεται τα θεμελιώδη δικαιώματα. Ο βαθμός στον οποίο απαιτείται επεξηγησιμότητα εξαρτάται σημαντικά από το πλαίσιο και τη σοβαρότητα των συνεπειών, στην περίπτωση που το συγκεκριμένο αποτέλεσμα είναι εσφαλμένο ή κατ' άλλον τρόπο ανακριβές (High Level Expert Group on AI, 2019).

2ο Επίπεδο Πλαισίου – Η πραγμάτωση της αξιόπιστης Τεχνητής Νοημοσύνης

Η εφαρμογή και η πραγμάτωση της αξιόπιστης Τεχνητής Νοημοσύνης επιτυγχάνεται μέσω ενός καταλόγου επτά (7) απαιτήσεων που πρέπει να τηρούνται βάσει των τεσσάρων (4) αρχών που προαναφέρθηκαν. Οι εν λόγω αρχές θα πρέπει να κωδικοποιηθούν σε απαιτήσεις και να ισχύουν για όλους τους εμπλεκόμενους στον κύκλο ζωής των συστημάτων Τεχνητής Νοημοσύνης και αφορούν προγραμματιστές, εγκαταστάτες, τελικούς χρήστες και ευρύτερη κοινωνία, δηλαδή όλους όσους επηρεάζονται άμεσα ή έμμεσα από τα συγκεκριμένα συστήματα Τεχνητής Νοημοσύνης. Αν και οι περισσότερες απαιτήσεις ισχύουν για όλα τα συστήματα Τεχνητής Νοημοσύνης, ιδιαίτερη προσοχή δίνεται σε εκείνες που επηρεάζουν άμεσα ή έμμεσα τα άτομα (High Level Expert Group on AI, 2019).

Ακολούθως αποτυπώνονται οι απαιτήσεις, οι οποίες θα πρέπει να διασφαλίζονται από τους εμπλεκόμενους ανάλογα με τον ρόλο τους στον κύκλο ζωής του συστήματος Τεχνητής Νοημοσύνης:

Ανθρώπινη παρέμβαση και εποπτεία

Τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να στηρίζουν την ανθρώπινη αυτονομία και τη λήψη αποφάσεων, όπως προβλέπεται στην αρχή του σεβασμού της ανθρώπινης αυτονομίας. Η απαίτηση αυτή προϋποθέτει ότι τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει και να λειτουργούν ως εργαλεία ανάπτυξης μιας δημοκρατικής, ευημερούσας και ισότιμης κοινωνίας, υποστηρίζοντας την παρέμβαση του χρήστη, αλλά και να προάγουν τα θεμελιώδη δικαιώματα, καθώς και να αφήνουν περιθώριο για ανθρώπινη εποπτεία. Σε περιπτώσεις που τα συστήματα Τεχνητής νοημοσύνης μπορούν να επηρεάσουν αρνητικά τα θεμελιώδη δικαιώματα, θα πρέπει να διενεργείται εκτίμηση επιπτώσεων (impact assessment) για τα θεμελιώδη δικαιώματα πριν από την ανάπτυξή τους και αξιολόγηση του κατά πόσο οι συγκεκριμένοι κίνδυνοι είναι δυνατό να μειωθούν ή να δικαιολογηθούν ως απαραίτητοι. Σε άλλες περιπτώσεις που ενδέχεται να αποτελούν απειλή για την ατομική αυτονομία, ενισχύεται το δικαίωμα των χρηστών να μην υπόκεινται σε απόφαση που λαμβάνεται αποκλειστικά βάσει αυτοματοποιημένης επεξεργασίας, εάν η εν λόγω απόφαση παράγει έννομα αποτελέσματα για τους χρήστες ή τους επηρεάζει σημαντικά με παρόμοιο τρόπο. Επίσης, η ατομική αυτονομία διασφαλίζεται και με την ανθρώπινη εποπτεία, η οποία επιτυγχάνεται μέσω μηχανισμών διακυβέρνησης, όπως η προσέγγιση στην οποία ο άνθρωπος είναι εκείνος που παρεμβαίνει (human-in-the-loop - HITL), εκείνος που εποπτεύει (human-on-the-loop - HOTL) ή εκείνος που ελέγχει (human-in-command -HIC) (High Level Expert Group on AI, 2019).

Τεχνική στιβαρότητα και ασφάλεια

Σημαντική συνιστώσα για την επίτευξη αξιόπιστης Τεχνητής Νοημοσύνης αποτελεί η τεχνική στιβαρότητα, η οποία συνδέεται στενά με την αρχή της πρόληψης βλάβης. Η τεχνική στιβαρότητα προϋποθέτει την ανάπτυξη συστημάτων Τεχνητή Νοημοσύνη με προληπτική προσέγγιση των κινδύνων και με τέτοιο τρόπο ώστε να συμπεριφέρονται αξιόπιστα με τον αναμενόμενο τρόπο, ενώ παράλληλα ελαχιστοποιείται η ακούσια και μη αναμενόμενη βλάβη και αποτρέπεται η απαράδεκτη. Το ίδιο θα πρέπει να ισχύει επίσης για πιθανές αλλαγές στο λειτουργικό τους περιβάλλον ή για την παρουσία άλλων πρακτόρων (ανθρώπινων και τεχνητών) που ενδέχεται να αλληλεπιδρούν κατ' αντιπαράθεση με το σύστημα. Επιπλέον, θα πρέπει να διασφαλίζεται η σωματική και διανοητική ακεραιότητα των ανθρώπων. Τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να προστατεύονται από ευπάθειες και να λαμβάνονται μέτρα για την πρόληψη και τον μετριασμό τους, ώστε να μην είναι ευάλωτα σε

κακόβουλες επιθέσεις. Οι ανεπαρκείς διαδικασίες προστασίας είναι επίσης δυνατό να οδηγήσουν σε εσφαλμένες αποφάσεις ή ακόμα και σε σωματική βλάβη. Τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να διαθέτουν εγγυήσεις για την εφαρμογή εφεδρικού σχεδίου σε περίπτωση που προκύψουν προβλήματα. Σε περιπτώσεις συστημάτων υψηλού κινδύνου είναι σημαντικό να αναπτύσσονται και να δοκιμάζονται προληπτικά τα μέτρα ασφαλείας. Είναι σημαντικό και ιδιαίτερα σε καταστάσεις όπου τα συστήματα Τεχνητής Νοημοσύνης επηρεάζουν άμεσα τις ανθρώπινες ζωές, να διασφαλίζεται η ακρίβεια τους, η οποία σχετίζεται με την ικανότητα τους να διατυπώνουν σωστές κρίσεις, όπως για παράδειγμα για να ταξινομούν σωστά τις πληροφορίες στις κατάλληλες κατηγορίες ή την ικανότητά τους να κάνουν σωστές προβλέψεις, συστάσεις ή να λαμβάνουν αποφάσεις που βασίζονται σε δεδομένα ή μοντέλα. Επίσης, σημαντικό είναι τα αποτελέσματα των συστημάτων Τεχνητής Νοημοσύνης να είναι αξιόπιστα και να μπορούν να αναπαραχθούν δηλαδή να μπορούν να λειτουργούν σωστά σε μια σειρά εισροών και σε μια σειρά καταστάσεων και να παρουσιάζουν την ίδια συμπεριφορά όταν επαναλαμβάνεται υπό τις ίδιες συνθήκες ώστε να δίνει τη δυνατότητα στους επιστήμονες και στους υπεύθυνους χάραξης πολιτικής να περιγράφουν με ακρίβεια τις ενέργειές τους (High Level Expert Group on AI, 2019).

Ιδιωτική ζωή και διακυβέρνηση των δεδομένων

Η ιδιωτική ζωή, η οποία αποτελεί θεμελιώδες δικαίωμα το οποίο επηρεάζεται σε μεγάλο βαθμό από τα συστήματα Τεχνητής Νοημοσύνης, είναι στενά συνδεδεμένη με την αρχή της πρόληψης βλάβης. Η πρόληψη βλάβης στην ιδιωτική ζωή απαιτεί επίσης επαρκή διακυβέρνηση δεδομένων που καλύπτει την ποιότητα και την ακεραιότητα των χρησιμοποιούμενων δεδομένων, τη συνάφειά τους σε σχέση με τον τομέα στον οποίο τα συστήματα Τεχνητής Νοημοσύνης θα εγκαθίστανται, τα πρωτόκολλα πρόσβασής τους και την ικανότητα επεξεργασίας δεδομένων κατά τρόπο που προστατεύει την ιδιωτική ζωή. Τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να εγγυώνται την προστασία της ιδιωτικής ζωής και των δεδομένων καθόλη τη διάρκεια του κύκλου ζωής ενός συστήματος. Αυτό περιλαμβάνει τις πληροφορίες που παρέχονται αρχικά από τον χρήστη, καθώς και τις πληροφορίες που δημιουργούνται σχετικά με τον χρήστη κατά τη διάρκεια της αλληλεπίδρασής του με το σύστημα (π.χ. τα αποτελέσματα που παράγει το σύστημα Τεχνητής Νοημοσύνης για συγκεκριμένους χρήστες ή τον τρόπο με τον οποίο οι χρήστες ανταποκρίθηκαν σε συγκεκριμένες συστάσεις). Η ψηφιακή καταγραφή της ανθρώπινης συμπεριφοράς μπορεί να επιτρέψει στα συστήματα Τεχνητής Νοημοσύνης να συνάγουν όχι

μόνο τις προτιμήσεις του ατόμου, αλλά και τον σεξουαλικό προσανατολισμό, την ηλικία, το φύλο, τις θρησκευτικές ή πολιτικές απόψεις του. Προκειμένου η διαδικασία συλλογής δεδομένων να εμπνέει εμπιστοσύνη στους χρήστες, θα πρέπει να διασφαλίζεται ότι τα δεδομένα που συλλέγονται σχετικά με αυτούς δεν θα χρησιμοποιηθούν για αθέμιτες ή παράνομες διακρίσεις απέναντί τους. Ένα άλλο σημαντικό θέμα, αφορά στην ποιότητα των χρησιμοποιούμενων συνόλων δεδομένων που έχει άμεσο αντίκτυπο στις επιδόσεις των συστημάτων Τεχνητής Νοημοσύνης. Τα δεδομένα που συλλέγονται μπορεί να περιέχουν κοινωνικά κατασκευασμένες προκαταλήψεις ή ανακρίβειες, σφάλματα και λάθη. Ο κίνδυνος αυτός πρέπει να αντιμετωπίζεται πριν από την εκπαίδευση με οποιοδήποτε σύνολο δεδομένων. Επιπλέον, πρέπει να διασφαλίζεται η ακεραιότητα των δεδομένων. Η τροφοδοσία ενός συστήματος Τεχνητής Νοημοσύνης με κακόβουλα δεδομένα μπορεί να αλλοιώσει τη συμπεριφορά του, ιδίως με αυτοεκπαιδευόμενα συστήματα. Οι διεργασίες και τα σύνολα δεδομένων που χρησιμοποιούνται πρέπει να υποβάλλονται σε δοκιμές και να τεκμηριώνονται σε κάθε στάδιο, όπως για παράδειγμα στα στάδια του σχεδιασμού, της εκπαίδευσης, των δοκιμών και της εγκατάστασης. Το ίδιο θα πρέπει να ισχύει και για τα συστήματα Τεχνητής Νοημοσύνης που δεν αναπτύσσονται στο εσωτερικό ενός οργανισμού, αλλά αγοράζονται από τρίτους. Επίσης, οποιοσδήποτε οργανισμός χειρίζεται προσωπικά δεδομένα θα πρέπει να εφαρμόζει πρωτόκολλα δεδομένων που διέπουν την πρόσβαση σε δεδομένα. Θα πρέπει να επιτρέπεται η πρόσβαση στα δεδομένα ενός ατόμου μόνο σε δεόντως ειδικευμένο προσωπικό με τη σχετική αρμοδιότητα και ανάγκη (High Level Expert Group on AI, 2019).

Διαφάνεια

Η διαφάνεια είναι στενά συνδεδεμένη με την αρχής της επεξηγησιμότητας και αφορά στη διαφάνεια των στοιχείων που σχετίζονται με ένα σύστημα Τεχνητής Νοημοσύνης και συγκεκριμένα τα δεδομένα, το σύστημα και τα επιχειρηματικά μοντέλα. Τα σύνολα δεδομένων, οι αλγόριθμοι που χρησιμοποιούνται και οι διαδικασίες που οδηγούν στην απόφαση του συστήματος Τεχνητής Νοημοσύνης θα πρέπει να τεκμηριώνονται στο βέλτιστο δυνατό επίπεδο ώστε να καθίσταται δυνατή η ιχνηλασιμότητα και η αύξηση της διαφάνειας. Με αυτόν τον τρόπο καθίσταται δυνατός ο εντοπισμός των λόγων για τους οποίους μια απόφαση ενός συστήματος Τεχνητής Νοημοσύνης ήταν εσφαλμένη, κάτι που με τη σειρά του θα μπορούσε να συμβάλλει στην αποφυγή μελλοντικών λαθών. Επομένως, η ιχνηλασιμότητα διευκολύνει την ελεγξιμότητα και την επεξηγησιμότητα. Η επεξηγησιμότητα αφορά την δυνατότητα επεξήγησης τόσο των τεχνικών διεργασιών ενός

συστήματος Τεχνητής Νοημοσύνης όσο και των σχετικών ανθρώπινων αποφάσεων. Η τεχνική επεξηγησιμότητα προϋποθέτει τη δυνατότητα κατανόησης και ανίχνευσης, από ανθρώπους, των αποφάσεων που λαμβάνονται από ένα σύστημα Τεχνητής Νοημοσύνης. Σε κάθε περίπτωση που ένα σύστημα Τεχνητής Νοημοσύνης έχει σημαντικό αντίκτυπο στη ζωή των ανθρώπων, θα πρέπει να είναι δυνατό να αξιώνεται η παροχή κατάλληλων επεξηγήσεων σχετικά με τη διαδικασία λήψης αποφάσεων του συστήματος Τεχνητής Νοημοσύνης. Οι επεξηγήσεις αυτές θα πρέπει να παρέχονται εγκαίρως και να προσαρμόζονται στην εμπειρογνωμοσύνη του ενδιαφερόμενου μέρους (π.χ. μη ειδήμων, ρυθμιστικός φορέας ή ερευνητής). Επιπλέον, θα πρέπει να παρέχονται επεξηγήσεις σχετικά με τον βαθμό στον οποίο ένα σύστημα Τεχνητής Νοημοσύνης επηρεάζει και διαμορφώνει τη διαδικασία λήψης αποφάσεων ενός οργανισμού, τις σχεδιαστικές επιλογές του συστήματος, καθώς και το σκεπτικό της εγκατάστασής του, ώστε να διασφαλίζεται η διαφάνεια του επιχειρηματικού μοντέλου. Επίσης, τα συστήματα Τεχνητής Νοημοσύνης δεν πρέπει να παρουσιάζονται ως άνθρωποι στον χρήστη, αλλά να τον ενημερώνουν ότι αλληλεπιδρά με ένα σύστημα. Τούτο συνεπάγεται ότι τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να είναι αναγνωρίσιμα, ως τέτοια. Θα πρέπει να παρέχεται η δυνατότητα να παρακαμφθεί η συγκεκριμένη αλληλεπίδραση προς όφελος της ανθρώπινης αλληλεπίδρασης, όπου αυτό είναι αναγκαίο, προκειμένου να διασφαλίζεται η τήρηση των θεμελιωδών δικαιωμάτων (High Level Expert Group on AI, 2019).

Πολυμορφία, απαγόρευση των διακρίσεων και δικαιοσύνη

Για την επίτευξη αξιόπιστης Τεχνητής Νοημοσύνης, ολόκληρος ο κύκλος ζωής των συστημάτων της θα πρέπει να χαρακτηρίζεται από συμπεριληπτικότητα και πολυμορφία. Τούτο συνεπάγεται τόσο την εξέταση και συμμετοχή όλων των επηρεαζόμενων μερών καθόλη τη διάρκεια της διαδικασίας, όσο και την εξασφάλιση ισότιμης πρόσβασης μέσω συμπεριληπτικών διαδικασιών σχεδιασμού και ίσης μεταχείρισης. Η εν λόγω απαίτηση συνδέεται στενά με την αρχή της δικαιοσύνης. Τα σύνολα δεδομένων που χρησιμοποιούνται από συστήματα Τεχνητής Νοημοσύνης, τόσο για την εκπαίδευση όσο και για τη λειτουργία τους, ενδέχεται να είναι μεροληπτικά, ελλιπή και να βασίζονται σε ακατάλληλα μοντέλα διακυβέρνησης. Η συνέχιση των μεροληψιών αυτού του είδους θα μπορούσε να οδηγήσει σε ακούσιες άμεσες και έμμεσες προκαταλήψεις και διακρίσεις κατά ορισμένων ομάδων ή ανθρώπων, επιδεινώνοντας ενδεχομένως τις υφιστάμενες προκαταλήψεις και την περιθωριοποίηση. Βλάβη μπορεί επίσης να προκύψει από την εκούσια εκμετάλλευση της μεροληψίας (π.χ. των καταναλωτών) ή από την άσκηση αθέμιτου ανταγωνισμού, όπως η

ομογενοποίηση των τιμών μέσω συμπράξεων ή μη διαφάνειας της αγοράς. Η αναγνωρίσιμη και διακριτική μεροληψία θα πρέπει να παρακάμπτεται κατά τη φάση συλλογής, όπου είναι δυνατό. Αθέμιτη μεροληψία ενδέχεται να παρατηρείται επίσης στον τρόπο ανάπτυξης των συστημάτων Τεχνητής Νοημοσύνης (π.χ. προγραμματισμός αλγορίθμων). Η μεροληψία αυτή θα μπορούσε να αντιμετωπιστεί με την εφαρμογή διαδικασιών εποπτείας για την ανάλυση και την αντιμετώπιση του σκοπού, των περιορισμών, των απαιτήσεων και των αποφάσεων του συστήματος κατά τρόπο σαφή και διαφανή. Επιπλέον, η πρόσληψη ατόμων με διαφορετικές καταβολές, διαφορετικούς πολιτισμούς και από διαφορετικά επιστημονικά πεδία μπορεί να εξασφαλίσει πολυφωνία απόψεων και θα πρέπει να ενθαρρύνεται. Ιδίως στις σχέσεις επιχείρησης με καταναλωτή, τα συστήματα θα πρέπει να είναι προσανατολισμένα στον χρήστη και να σχεδιάζονται με τρόπο που να επιτρέπει σε όλους τους ανθρώπους να χρησιμοποιούν προϊόντα ή υπηρεσίες Τεχνητής Νοημοσύνης, ανεξάρτητα από ηλικία, φύλο, ικανότητες ή τα χαρακτηριστικά τους. Η προσβασιμότητα στη συγκεκριμένη τεχνολογία για τα άτομα με αναπηρίες, όλων των κοινωνικών ομάδων, έχει ιδιαίτερη σημασία. Τα συστήματα Τεχνητής Νοημοσύνης δεν θα πρέπει να έχουν ενιαία προσέγγιση και θα πρέπει να λαμβάνουν υπόψη τις αρχές του καθολικού σχεδιασμού που απευθύνεται σε όσο το δυνατόν ευρύτερο φάσμα χρηστών, σύμφωνα με τα σχετικά πρότυπα προσβασιμότητας. Κάτι τέτοιο θα καταστήσει δυνατή την ισότιμη πρόσβαση και την ενεργή συμμετοχή όλων των ανθρώπων στις υφιστάμενες και αναδυόμενες ανθρώπινες δραστηριότητες μέσω υπολογιστών και σε σχέση με τις τεχνολογίες υποβοήθησης. Παράλληλα, κρίνεται επωφελές να πραγματοποιούνται διαβουλεύσεις με τα ενδιαφερόμενα μέρη που ενδέχεται να επηρεάζονται άμεσα ή έμμεσα από το σύστημα καθόλη τη διάρκεια του κύκλου ζωής του και να ζητείται τακτική αναπληροφόρηση ακόμη και μετά την εγκατάστασή του (High Level Expert Group on AI, 2019).

Κοινωνική και περιβαλλοντική ευημερία

Σύμφωνα με τις αρχές της δικαιοσύνης και της πρόληψης βλάβης, η ευρύτερη κοινωνία, τα άλλα ευαίσθητα όντα και το περιβάλλον θα πρέπει επίσης να θεωρούνται ενδιαφερόμενα μέρη σε ολόκληρο τον κύκλο ζωής των συστημάτων Τεχνητής Νοημοσύνης. Θα πρέπει να ενθαρρύνεται η βιωσιμότητα και η οικολογική υπευθυνότητα των συστημάτων Τεχνητής Νοημοσύνης και να ενισχύεται η έρευνα σε λύσεις Τεχνητής Νοημοσύνης που να καλύπτουν τομείς παγκόσμιου ενδιαφέροντος, όπως για παράδειγμα οι στόχοι βιώσιμης ανάπτυξης. Ιδανικά, η Τεχνητή Νοημοσύνη θα πρέπει να χρησιμοποιείται προς όφελος όλων των ανθρώπων, συμπεριλαμβανομένων των μελλοντικών γενεών. Η διαδικασία ανάπτυξης,

εγκατάστασης και χρήσης του συστήματος, καθώς και ολόκληρη η αλυσίδα εφοδιασμού του, θα πρέπει να αξιολογείται σε αυτό το πλαίσιο, π.χ. μέσω κριτικής εξέτασης της χρήσης πόρων και της κατανάλωσης ενέργειας κατά τη διάρκεια της εκπαίδευσης, και να επιλέγονται λιγότερο επιβλαβείς επιλογές. Θα πρέπει να ενθαρρύνονται μέτρα που εξασφαλίζουν την φιλικότητα προς το περιβάλλον ολόκληρης της αλυσίδας εφοδιασμού του συστήματος Τεχνητής Νοημοσύνης. Επίσης, η γενικευμένη έκθεση σε κοινωνικά συστήματα Τεχνητής Νοημοσύνης σε όλους τους τομείς της ζωής μας, όπως στην εκπαίδευση, στην εργασία, στη φροντίδα ή στην ψυχαγωγία) μπορεί να αλλοιώσει την αντίληψή μας για την κοινωνική παρέμβαση ή να επηρεάσει τις κοινωνικές σχέσεις και τις συνδέσεις μας με αποτέλεσμα να συμβάλουν στην υποβάθμισή τους. Κάτι τέτοιο θα μπορούσε να επηρεάσει επίσης τη σωματική και πνευματική ευημερία των ανθρώπων. Πέρα από την αξιολόγηση των επιπτώσεων της ανάπτυξης, εγκατάστασης και χρήσης των συστημάτων Τεχνητής Νοημοσύνης σε άτομα, οι συγκεκριμένες επιπτώσεις θα πρέπει επίσης να αξιολογούνται από κοινωνιακή άποψη, λαμβανομένης υπόψη της επίδρασής τους στα ιδρύματα, τη δημοκρατία και ευρύτερα την κοινωνία. Η χρήση συστημάτων Τεχνητής Νοημοσύνης θα πρέπει να εξεταστεί προσεκτικά, ιδίως σε περιπτώσεις που αφορούν τη δημοκρατική διαδικασία, όχι μόνο σχετικά με τη λήψη πολιτικών αποφάσεων, αλλά και με το εκλογικό περιβάλλον (High Level Expert Group on AI, 2019).

Λογοδοσία

Η απαίτηση της λογοδοσίας συμπληρώνει τις ανωτέρω απαιτήσεις που συνδέονται στενά με την αρχή της δικαιοσύνης. Προϋποθέτει τη δημιουργία μηχανισμών μέσω των οποίων θα διασφαλίζεται η υπευθυνότητα και η λογοδοσία για τα συστήματα Τεχνητής Νοημοσύνης και τα αποτελέσματά τους, τόσο πριν όσο και μετά την υλοποίησή τους. Η ελεγξιμότητα συνεπάγεται τη δυνατότητα αξιολόγησης των αλγορίθμων, των δεδομένων και των διαδικασιών σχεδιασμού. Η αξιολόγηση από εσωτερικούς και εξωτερικούς ελεγκτές και η διαθεσιμότητα εκθέσεων αξιολόγησης αυτού του είδους μπορούν να συνεισφέρουν στην αξιοπιστία της τεχνολογίας. Στις εφαρμογές που επηρεάζουν τα θεμελιώδη δικαιώματα, συμπεριλαμβανομένων των εφαρμογών που είναι καίριας σημασίας για την ασφάλεια, τα συστήματα Τεχνητής Νοημοσύνης θα πρέπει να είναι δυνατό να ελέγχονται ανεξάρτητα. Θα πρέπει, επίσης, να διασφαλίζεται η ικανότητα τόσο γνωστοποίησης για δράσεις ή αποφάσεις που συμβάλλουν σε ένα συγκεκριμένο αποτέλεσμα του συστήματος, όσο και αντιμετώπισης των συνεπειών ενός τέτοιου αποτελέσματος. Ο εντοπισμός, η αξιολόγηση, η γνωστοποίηση και η ελαχιστοποίηση των δυνητικών αρνητικών επιπτώσεων των συστημάτων Τεχνητής

Νοημοσύνης είναι ιδιαιτέρως σημαντικές για εκείνους που επηρεάζονται άμεσα ή έμμεσα. Η χρήση εκτιμήσεων επιπτώσεων τόσο πριν όσο και κατά τη διάρκεια της ανάπτυξης, εγκατάστασης και χρήσης συστημάτων Τεχνητής Νοημοσύνης μπορεί να αποδειχθεί χρήσιμη για την ελαχιστοποίηση των αρνητικών επιπτώσεων. Οι εκτιμήσεις αυτές πρέπει να είναι ανάλογες με τους κινδύνους που ενέχουν τα συστήματα Τεχνητής Νοημοσύνης. Κατά την εφαρμογή των ανωτέρω απαιτήσεων, ενδέχεται να προκύψουν τριβές μεταξύ τους, οι οποίες ενδέχεται να οδηγήσουν σε αναπόφευκτες αντισταθμιστικές ρυθμίσεις. Σε περιπτώσεις που δεν είναι δυνατό να προσδιοριστούν δεοντολογικά αποδεκτές αντισταθμιστικές ρυθμίσεις, η ανάπτυξη, η εγκατάσταση και η χρήση του συστήματος Τεχνητής Νοημοσύνης δεν θα πρέπει να προχωρά με τη συγκεκριμένη μορφή. Οποιαδήποτε απόφαση σχετικά με την αντισταθμιστική ρύθμιση που πρέπει να γίνει θα πρέπει να αιτιολογείται και να τεκμηριώνεται δεόντως. Ο υπεύθυνος για τη λήψη αποφάσεων θα πρέπει να λογοδοτεί για τον τρόπο με τον οποίο γίνεται η κατάλληλη αντισταθμιστική ρύθμιση και θα πρέπει να εξετάζει συνεχώς την καταλληλότητα της απόφασης που προκύπτει, προκειμένου να διασφαλίζει ότι μπορούν να γίνουν οι απαραίτητες αλλαγές στο σύστημα όπου αυτό είναι απαραίτητο. Θα πρέπει να προβλέπονται προσβάσιμοι μηχανισμοί που να εξασφαλίζουν κατάλληλη έννομη προστασία σε περιπτώσεις άδικων, δυσμενών επιπτώσεων. Η επίγνωση ότι προβλέπεται έννομη προστασία σε περίπτωση που προκύψει κάποιο πρόβλημα είναι πολύ σημαντική για την εξασφάλιση της εμπιστοσύνης. Ιδιαίτερη προσοχή θα πρέπει να δίνεται σε ευάλωτα άτομα ή ομάδες (High Level Expert Group on AI, 2019).

Υλοποίηση απαιτήσεων με τεχνικές και μη τεχνικές μεθόδους

Η υλοποίηση των απαιτήσεων για την πραγμάτωση της αξιόπιστης Τεχνητής Νοημοσύνης αφορούν τόσο τεχνικές όσο και μη τεχνικές μεθόδους και περιλαμβάνουν όλα τα στάδια του κύκλου ζωής ενός συστήματος Τεχνητής Νοημοσύνης. Οι ακόλουθες μέθοδοι που χρησιμοποιούνται για την εφαρμογή αξιόπιστης Τεχνητής Νοημοσύνης και οι οποίες λειτουργούν συμπληρωματικά ή εναλλακτικά μεταξύ τους, μπορούν να ενσωματωθούν στις φάσεις σχεδιασμού, ανάπτυξης και χρήσης ενός συστήματος Τεχνητής Νοημοσύνης και θα πρέπει να αξιολογούνται σε διαρκή βάση, καθώς τα συστήματα Τεχνικής Νοημοσύνης εξελίσσονται συνεχώς και λειτουργούν σε δυναμικό περιβάλλον.

Οι τεχνικές μέθοδοι εφαρμογής αξιόπιστης Τεχνητής Νοημοσύνης αφορούν μεθόδους που σχετίζονται με:

- Αρχιτεκτονικές για αξιόπιστη Τεχνητή Νοημοσύνη
- Δεοντολογία και κράτος δικαίου ήδη από τον σχεδιασμό
- Μέθοδοι επεξήγησης
- Δοκιμή και επικύρωση
- Δείκτες ποιότητας υπηρεσίας

Οι μη τεχνικές μέθοδοι διασφάλισης και διατήρησης αξιόπιστης Τεχνητής Νοημοσύνης αφορούν μεθόδους που σχετίζονται με:

- Κανονιστική ρύθμιση
- Κώδικες δεοντολογίας
- Τυποποίηση
- Πιστοποίηση
- Λογοδοσία μέσω πλαισίων διακυβέρνησης
- Εκπαίδευση και ευαισθητοποίηση για την προώθηση πνεύματος δεοντολογίας
- Συμμετοχή των ενδιαφερόμενων μερών και κοινωνικός διάλογος
- Πολυμορφία και συμπεριληπτικές ομάδες σχεδιασμού

3ο Επίπεδο Πλαισίου – Η αξιολόγηση της αξιοπιστίας Τεχνητής Νοημοσύνης

Οι Κατευθυντήριες Γραμμές περιέχουν έναν μη εξαντλητικό κατάλογο αξιολόγησης της αξιοπιστίας της Τεχνητής Νοημοσύνης βασισμένο στις επτά (7) απαιτήσεις που προσδιορίστηκαν στο 2ο Επίπεδο του Πλαισίου. Το 2019 ο κατάλογος αξιολόγησης για αξιόπιστη Τεχνητή Νοημοσύνη έθεσε τις εν λόγω απαιτήσεις σε εφαρμογή στο πλαίσιο πιλοτικής διαδικασίας με περισσότερους από 350 οργανισμούς, ενώ το 2020 οριστικοποιήθηκε η αναθεωρημένη του έκδοση από την Ομάδα Εμπειρογνωμόνων Υψηλού Επιπέδου.

Αυτός ο αναθεωρημένος κατάλογος αξιολόγησης για αξιόπιστη Τεχνητή Νοημοσύνη (ALTAI) προορίζεται για ευέλικτη χρήση, γιατί οι οργανισμοί μπορούν να αντλήσουν στοιχεία σχετικά με το συγκεκριμένο σύστημα Τεχνητής Νοημοσύνης από τον κατάλογο ή να προσθέσουν στοιχεία σε αυτόν, λαμβάνοντας υπόψη τον τομέα στον οποίο δραστηριοποιούνται. Βοηθά τους οργανισμούς, μέσω δέσμης συγκεκριμένων ερωτήσεων, να κατανοήσουν τί είναι η αξιόπιστη Τεχνητή Νοημοσύνη, ιδιαίτερα τους κινδύνους που μπορεί να δημιουργήσει ένα σύστημα Τεχνητής Νοημοσύνης και πώς να ελαχιστοποιήσουν αυτούς τους κινδύνους, μεγιστοποιώντας παράλληλα τα οφέλη της Τεχνητής Νοημοσύνης. Σκοπός του είναι να βοηθήσει τους οργανισμούς να εντοπίσουν πώς τα προτεινόμενα

συστήματα Τεχνητής Νοημοσύνης ενδέχεται να δημιουργήσουν κινδύνους και να προσδιορίσουν εάν και τι είδους ενεργά μέτρα μπορεί να χρειαστεί να ληφθούν για την αποφυγή και την ελαχιστοποίηση αυτών των κινδύνων, διασφαλίζοντας την αξιοπιστία της Τεχνητής Νοημοσύνης σε όλα τα επίπεδα (High Level Expert Group on AI, 2019; 2020).

5.2 «Σύσταση για την Ηθική της Τεχνητής Νοημοσύνης» - UNESCO

Στις 24 Νοεμβρίου 2021, η Σύσταση για την Ηθική της Τεχνητής Νοημοσύνης εγκρίθηκε από τη Γενική Διάσκεψη του Επιστημονικού και Πολιτιστικού Οργανισμού των Ηνωμένων Εθνών (UNESCO) στην 41η σύνοδό της, ύστερα από μια διετή διαδικασία για την εκπόνηση αυτού του πρώτου παγκόσμιου μέσου καθορισμού προτύπων σχετικά με την ηθική της Τεχνητής Νοημοσύνης με τη μορφή σύστασης.

Το σχέδιο της Σύστασης επεξεργάστηκε και προετοίμασε μια Ad Hoc Ομάδα Εμπειρογνωμόνων (AHEG) ύστερα από διεπιστημονικές και διακυβερνητικές διαβουλεύσεις με ένα ευρύ φάσμα ενδιαφερομένων, βασισμένη στην προκαταρκτική μελέτη για τη δεοντολογία της Τεχνητής Νοημοσύνης της Παγκόσμιας Επιτροπής της UNESCO για την Ηθική της Επιστημονικής Γνώσης και Τεχνολογίας (COMEST).[9]

Αυτή η Σύσταση αντιμετωπίζει ηθικά ζητήματα που σχετίζονται με τον τομέα της Τεχνητής Νοημοσύνης. Προσεγγίζει την ηθική της Τεχνητής Νοημοσύνης ως ένα συστηματικό κανονιστικό προβληματισμό που βασίζεται σε ένα ολιστικό, περιεκτικό, πολυπολιτισμικό και εξελισσόμενο πλαίσιο αλληλεξαρτώμενων αξιών, αρχών και ενεργειών που μπορεί να καθοδηγήσει τις κοινωνίες στην υπεύθυνη αντιμετώπιση των γνωστών και άγνωστων επιπτώσεων των τεχνολογιών Τεχνητής Νοημοσύνης στους ανθρώπους, τις κοινωνίες, το περιβάλλον και τα οικοσυστήματα και τους προσφέρει μια βάση για να αποδεχτούν ή να απορρίψουν τις τεχνολογίες της. Θεωρεί την ηθική, ως μια δυναμική βάση για την κανονιστική αξιολόγηση και καθοδήγηση των τεχνολογιών Τεχνητής Νοημοσύνης που αναφέρεται στην ανθρώπινη αξιοπρέπεια, την ευημερία και την πρόληψη της βλάβης και λειτουργεί ως πυξίδα για την ηθική της επιστήμης και της τεχνολογίας. Τα ηθικά ερωτήματα σχετικά με τα συστήματα Τεχνητής Νοημοσύνης αφορούν όλα τα στάδια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης και οι φορείς Τεχνητής Νοημοσύνης περιλαμβάνουν οποιουσδήποτε παράγοντες εμπλέκονται σε τουλάχιστον ένα στάδιο του κύκλου ζωής του

[9] Αυτή η μελέτη τονίζει ότι επί του παρόντος κανένα παγκόσμιο μέσο δεν καλύπτει όλους τους τομείς που καθοδηγούν την ανάπτυξη και την εφαρμογή της τεχνητής νοημοσύνης σε μια ανθρωποκεντρική προσέγγιση.

συστήματος Τεχνητής Νοημοσύνης και μπορεί να είναι τόσο φυσικά όσο και νομικά πρόσωπα (UNESCO, 2021).

Αυτή η Σύσταση δίνει ιδιαίτερη προσοχή στις ευρύτερες ηθικές επιπτώσεις των συστημάτων Τεχνητής Νοημοσύνης σε σχέση με τους κεντρικούς τομείς της UNESCO, όπως εκπαίδευση, επιστήμη, πολιτισμός, επικοινωνία και πληροφορίες, οι οποίες διερευνήθηκαν στην Προκαταρκτική Μελέτη του 2019 για την Ηθική της Τεχνητής Νοημοσύνης από την Παγκόσμια Επιτροπή της UNESCO για την Ηθική της Επιστημονικής Γνώσης και Τεχνολογίας (COMEST):

(α) Εκπαίδευση, επειδή η ζωή σε κοινωνίες που ψηφιοποιούνται απαιτεί νέες εκπαιδευτικές πρακτικές, ηθικό προβληματισμό, κριτική σκέψη, υπεύθυνες πρακτικές σχεδιασμού και νέες δεξιότητες, δεδομένων, των επιπτώσεων για την αγορά εργασίας, την απασχολησιμότητα και τη συμμετοχή των πολιτών.

(β) Επιστήμη, με την ευρεία της έννοια που περιλαμβάνει όλα τα ακαδημαϊκά πεδία από τις φυσικές και ιατρικές επιστήμες έως τις κοινωνικές και ανθρωπιστικές επιστήμες, καθώς οι τεχνολογίες Τεχνητής Νοημοσύνης φέρνουν νέες ερευνητικές ικανότητες και προσεγγίσεις, έχουν επιπτώσεις στις έννοιες της επιστημονικής κατανόησης και εξήγησης και έχουν δημιουργήσει μια νέα βάση για τη λήψη αποφάσεων.

(γ) Πολιτιστική ταυτότητα και ποικιλομορφία, καθώς οι τεχνολογίες Τεχνητής Νοημοσύνης μπορούν να εμπλουτίσουν πολιτιστικές και δημιουργικές βιομηχανίες, αλλά μπορούν επίσης να οδηγήσουν σε αυξημένη συγκέντρωση της προσφοράς πολιτιστικού περιεχομένου, δεδομένων, αγορών και εισοδήματος στα χέρια λίγων μόνο παραγόντων, με πιθανές αρνητικές επιπτώσεις για την πολυμορφία και τον πλουραλισμό των γλωσσών, των μέσων ενημέρωσης, των πολιτιστικών εκφράσεων, της συμμετοχής και της ισότητας.

(δ) Επικοινωνία και πληροφορίες, καθώς οι τεχνολογίες Τεχνητής Νοημοσύνης διαδραματίζουν ολοένα και σημαντικότερο ρόλο στην επεξεργασία, τη δομή και την παροχή πληροφοριών, τα ζητήματα της αυτοματοποιημένης δημοσιογραφίας και της αλγοριθμικής παροχής ειδήσεων και της εποπτείας και επιμέλειας περιεχομένου στα μέσα κοινωνικής δικτύωσης και στις μηχανές αναζήτησης είναι μερικά μόνο παραδείγματα μεταξύ άλλων που εγείρουν ζητήματα σχετικά με την πρόσβαση σε πληροφορίες, την παραπληροφόρηση, τον μισαλλόδοξο λόγο, τις διακρίσεις, την ελευθερία έκφρασης και την ιδιωτικότητα.

Η παρούσα Σύσταση απευθύνεται στα κράτη μέλη, τόσο ως φορέων Τεχνητής Νοημοσύνης, όσο και ως αρχών υπεύθυνων για την ανάπτυξη νομικών και ρυθμιστικών πλαισίων καθ' όλη τη διάρκεια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης και για την προώθηση της επιχειρηματικής ευθύνης. Παρέχει, επίσης, ηθική καθοδήγηση σε όλους τους παράγοντες της Τεχνητής Νοημοσύνης, συμπεριλαμβανομένου του δημόσιου και του ιδιωτικού τομέα, παρέχοντας μια βάση για μια αξιολόγηση ηθικού αντίκτυπου των συστημάτων Τεχνητής Νοημοσύνης καθ' όλη τη διάρκεια του κύκλου ζωής τους.

Προκειμένου η ανάπτυξη και η χρήση των τεχνολογιών της Τεχνητής Νοημοσύνης να καθοδηγείται από υγιή επιστημονική έρευνα και ηθική ανάλυση και αξιολόγηση, η UNESCO συνιστά στα κράτη μέλη της να εφαρμόσουν σε εθελοντική βάση τις αρχές και τους κανόνες της παρούσας Σύστασης και να δεσμεύσουν όλα τα ενδιαφερόμενα μέρη, συμπεριλαμβανομένων των δημόσιων αρχών και φορέων, των επιχειρήσεων, ερευνητικών και ακαδημαϊκών οργανισμών, ιδρυμάτων και οργανισμών του δημόσιου και ιδιωτικού τομέα και της κοινωνίας των πολιτών που εμπλέκονται στις τεχνολογίες Τεχνητής Νοημοσύνης, ότι διαδραματίζουν τον αντίστοιχο ρόλο τους στην εφαρμογή της παρούσας σύστασης. (UNESCO, 2021)

Σκοποί και Στόχοι

Αυτή η σύσταση στοχεύει να παρέχει μια βάση ώστε τα συστήματα Τεχνητής Νοημοσύνης να λειτουργούν για το καλό της ανθρωπότητας, των ατόμων, των κοινωνιών, του περιβάλλοντος και των οικοσυστημάτων, για την πρόληψη των ζημιών, καθώς και για την τόνωση της ειρηνικής χρήσης συστημάτων Τεχνητής Νοημοσύνης. Εκτός από τα υφιστάμενα δεοντολογικά πλαίσια σχετικά με την Τεχνητή Νοημοσύνη σε όλο τον κόσμο, αυτή η Σύσταση στοχεύει στη δημιουργία ενός διεθνώς αποδεκτού κανονιστικού μέσου που θα εστιάζει όχι μόνο στη άρθρωση αξιών και αρχών, αλλά και στην πρακτική τους υλοποίηση, μέσω συγκεκριμένων συστάσεων πολιτικής, με ισχυρή έμφαση σε θέματα ισότητας των φύλων και προστασίας του περιβάλλοντος και των οικοσυστημάτων (UNESCO, 2021).

Επειδή η πολυπλοκότητα των ηθικών θεμάτων που περιβάλλουν την Τεχνητή Νοημοσύνη απαιτεί τη συνεργασία πολλών ενδιαφερομένων σε διάφορα επίπεδα και τομείς διεθνών, περιφερειακών και εθνικών κοινοτήτων, αυτή η σύσταση έχει ως στόχο να επιτρέψει στους

ενδιαφερόμενους να αναλάβουν κοινή ευθύνη βάσει ενός παγκόσμιου και διαπολιτισμικού διαλόγου.

Οι στόχοι της Σύστασης αυτής είναι:

- να παρέχει ένα οικουμενικό πλαίσιο αξιών, αρχών και ενεργειών που θα καθοδηγούν τα κράτη στη διαμόρφωση της νομοθεσίας, των πολιτικών ή άλλων μέσων τους σχετικά με την Τεχνητή Νοημοσύνη, σύμφωνα με το διεθνές δίκαιο,
- να καθοδηγεί τις ενέργειες ατόμων, ομάδων, κοινοτήτων, ιδρυμάτων και εταιρειών του ιδιωτικού τομέα για τη διασφάλιση της ενσωμάτωσης της δεοντολογίας σε όλα τα στάδια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης,
- να προστατεύει, να προωθεί και να σέβεται τα ανθρώπινα δικαιώματα και τις θεμελιώδεις ελευθερίες, την ανθρώπινη αξιοπρέπεια και ισότητα, συμπεριλαμβανομένης της ισότητας των φύλων με σκοπό την προστασία των συμφερόντων των σημερινών και των μελλοντικών γενεών, τη διατήρηση του περιβάλλοντος, της βιοποικιλότητας και των οικοσυστημάτων και το σεβασμό της πολιτισμικής ποικιλομορφίας σε όλα τα στάδια του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης,
- να προωθεί τον πολυμερή, πολυεπιστημονικό και πλουραλιστικό διάλογο και την οικοδόμηση συναίνεσης σχετικά με ηθικά ζητήματα που σχετίζονται με συστήματα Τεχνητής Νοημοσύνης
- να προωθεί τη δίκαιη πρόσβαση στις εξελίξεις και τις γνώσεις στον τομέα της Τεχνητής Νοημοσύνης και την δίκαιη κατανομή των οφελών από τη χρήση των συστημάτων Τεχνητής Νοημοσύνης

Αξίες και Αρχές Σύστασης

Η Σύσταση καθορίζει τις κοινές αξίες και αρχές που θα πρέπει να γίνονται σεβαστές από όλους τους παράγοντες του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης και να θεωρούνται ως συμπληρωματικές και αλληλένδετες. Η αξιοπιστία και η ακεραιότητα του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης είναι ουσιαστικής σημασίας προκειμένου να διασφαλίζεται ότι οι τεχνολογίες Τεχνητής Νοημοσύνης θα λειτουργούν για το καλό της ανθρωπότητας, των ατόμων, των κοινωνιών, του περιβάλλοντος και των οικοσυστημάτων και θα ενσωματώνουν τις αξίες και τις αρχές που ορίζονται στην εν λόγω Σύσταση. Καθώς η αξιοπιστία είναι αποτέλεσμα της λειτουργικότητας των αρχών, οι δράσεις πολιτικής που προτείνονται σε αυτήν τη Σύσταση στοχεύουν όλες στην προώθηση

της αξιοπιστίας σε όλα τα στάδια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης (UNESCO, 2021).

Αξίες

Σεβασμός, προστασία και προαγωγή των ανθρωπίνων δικαιωμάτων, των θεμελιωδών ελευθεριών και της ανθρώπινης αξιοπρέπειας

Ο σεβασμός, η προστασία και η προώθηση της ανθρώπινης αξιοπρέπειας και των δικαιωμάτων όπως καθορίζονται από το Διεθνές Δίκαιο, συμπεριλαμβανομένου του Διεθνούς Δικαίου για τα ανθρώπινα δικαιώματα, είναι ουσιαστικής σημασίας καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης. Η ανθρώπινη αξιοπρέπεια σχετίζεται με την αναγνώριση της εγγενούς και ίσης αξίας κάθε ανθρώπου ξεχωριστά, ανεξαρτήτως φυλής, χρώματος, καταγωγής, φύλου, ηλικίας, γλώσσας, θρησκείας, πολιτικής γνώμης, εθνικής καταγωγής, εθνικής καταγωγής, κοινωνικής καταγωγής, οικονομικής ή κοινωνικής κατάστασης γέννησης ή αναπηρίας και κάθε άλλου λόγου. Κανένα ανθρώπινο ον ή ανθρώπινη κοινότητα δεν πρέπει να βλάπτεται ή να υποτάσσεται σε οποιαδήποτε φάση του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης, ενώ θα πρέπει να προωθείται η ενεργή τους συμμετοχή χωρίς προκαταλήψεις και διαχωρισμούς. Τα άτομα μπορούν να αλληλεπιδρούν με συστήματα Τεχνητής Νοημοσύνης καθ' όλη τη διάρκεια του κύκλου ζωής τους και να λαμβάνουν βοήθεια από αυτά, χωρίς να «αντικειμενοποιούνται», ούτε να υπονομεύεται με άλλο τρόπο η αξιοπρέπειά τους ή να παραβιάζονται ή να καταχρώνται τα ανθρώπινα δικαιώματα και οι θεμελιώδεις ελευθερίες. Καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης, η ποιότητα ζωής των ανθρώπων θα πρέπει να βελτιώνεται (UNESCO, 2021).

Τα ανθρώπινα δικαιώματα και οι θεμελιώδεις ελευθερίες πρέπει να γίνονται σεβαστά, να προστατεύονται και να προωθούνται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων τεχνητής νοημοσύνης. Οι κυβερνήσεις, ο ιδιωτικός τομέας, η κοινωνία των πολιτών, οι διεθνείς οργανισμοί, οι τεχνικές κοινότητες και η ακαδημαϊκή κοινότητα πρέπει να σέβονται τα μέσα και τα πλαίσια ανθρωπίνων δικαιωμάτων στις παρεμβάσεις τους στις διαδικασίες που περιβάλλουν τον κύκλο ζωής των συστημάτων Τεχνητής Νοημοσύνης, ενώ οι νέες τεχνολογίες πρέπει να παρέχουν νέα μέσα για την υπεράσπιση και την άσκηση των ανθρωπίνων δικαιωμάτων και όχι για την παραβίασή τους (UNESCO, 2021).

Άνθηση περιβάλλοντος και οικοσυστημάτων

Η άνθηση του περιβάλλοντος και των οικοσυστημάτων θα πρέπει να αναγνωρίζεται, να προστατεύεται και να προωθείται μέσω του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης. Όλοι οι φορείς που εμπλέκονται στον κύκλο ζωής των συστημάτων Τεχνητής Νοημοσύνης πρέπει να συμμορφώνονται με το ισχύον Διεθνές Δίκαιο, νομοθεσία και τις πολιτικές για βιώσιμη ανάπτυξη. Θα πρέπει να μειώσουν τον περιβαλλοντικό αντίκτυπο των συστημάτων Τεχνητής Νοημοσύνης, συμπεριλαμβανομένου, ενδεικτικά, του αποτυπώματος άνθρακα, για να εξασφαλίσουν την ελαχιστοποίηση της κλιματικής κρίσης και των περιβαλλοντικών παραγόντων κινδύνου και να αποτρέψουν τη μη βιώσιμη εκμετάλλευση, χρήση και μετασχηματισμό των φυσικών πόρων που συμβάλλουν στην υποβάθμιση του περιβάλλοντος και την υποβάθμιση των οικοσυστημάτων (UNESCO, 2021).

Διασφάλιση της διαφορετικότητας και της ένταξης

Ο σεβασμός, η προστασία και η προώθηση της διαφορετικότητας και της ένταξης θα πρέπει να διασφαλίζονται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης, σύμφωνα με το Διεθνές Δίκαιο, συμπεριλαμβανομένου του Δικαίου των Ανθρωπίνων Δικαιωμάτων, με την προώθηση της ενεργούς συμμετοχής όλων των ατόμων ή ομάδων, χωρίς προκαταλήψεις και διαχωρισμούς (UNESCO, 2021).

Ζώντας σε ειρηνικές, δίκαιες και αλληλένδετες κοινωνίες

Οι φορείς της Τεχνητής Νοημοσύνης θα πρέπει να διαδραματίζουν συμμετοχικό και βοηθητικό ρόλο για τη διασφάλιση ειρηνικών και δίκαιων κοινωνιών που βασίζεται σε ένα διασυνδεδεμένο μέλλον προς όφελος όλων, συνεπές με τα ανθρώπινα δικαιώματα και τις θεμελιώδεις ελευθερίες. Η αξία της ζωής σε ειρηνικές και δίκαιες κοινωνίες δείχνει τη δυνατότητα των συστημάτων Τεχνητής Νοημοσύνης να συμβάλλουν καθ' όλη τη διάρκεια του κύκλου ζωής τους στη διασύνδεση όλων των ζωντανών πλασμάτων μεταξύ τους αλλά και με το φυσικό περιβάλλον (UNESCO, 2021).

Αρχές

Αναλογικότητα και μη Βλάβη

Θα πρέπει να αναγνωριστεί ότι οι τεχνολογίες Τεχνητής Νοημοσύνης δεν διασφαλίζουν απαραιτήτως, αυτές καθαυτές, την άνθηση του ανθρώπου, του περιβάλλοντος και των οικοσυστημάτων. Επιπλέον, καμία από τις διαδικασίες που σχετίζονται με τον κύκλο ζωής

ενός συστήματος Τεχνητής Νοημοσύνης δεν πρέπει να υπερβαίνει αυτό που είναι απαραίτητο για την επίτευξη θεμιτών στόχων και θα πρέπει να είναι κατάλληλη για το πλαίσιο. Σε περίπτωση πιθανής πρόκλησης βλάβης στον άνθρωπο, τα ανθρώπινα δικαιώματα και τις θεμελιώδεις ελευθερίες, τις κοινότητες και την κοινωνία γενικότερα ή το περιβάλλον και τα οικοσυστήματα, θα πρέπει να διασφαλίζεται η εφαρμογή διαδικασιών για την εκτίμηση του κινδύνου και η υιοθέτηση μέτρων για την αποτροπή της εμφάνισης βλάβης (UNESCO, 2021).

Η επιλογή χρήσης συστημάτων Τεχνητής Νοημοσύνης και η χρήση μεθόδου Τεχνητής Νοημοσύνης θα πρέπει να αιτιολογούνται με τους ακόλουθους τρόπους:

(α) η επιλεγμένη μέθοδος Τεχνητής Νοημοσύνης πρέπει να είναι κατάλληλη και ανάλογη για την επίτευξη ενός δεδομένου θεμιτού στόχου,

(β) η επιλεγμένη μέθοδος Τεχνητής Νοημοσύνης δεν θα πρέπει να παραβιάζει τις θεμελιώδεις αξίες που προαναφέρθηκαν, ιδίως, η χρήση της δεν πρέπει να παραβιάζει ή να καταστρατηγεί τα ανθρώπινα δικαιώματα· και

(γ) η μέθοδος Τεχνητής Νοημοσύνης θα πρέπει να είναι κατάλληλη για το πλαίσιο και να βασίζεται σε αυστηρά επιστημονικά θεμέλια.

Σε σενάρια όπου οι αποφάσεις θεωρείται ότι έχουν αντίκτυπο που είναι μη αναστρέψιμος ή δύσκολο να ανατραπεί ή μπορεί να περιλαμβάνει αποφάσεις ζωής και θανάτου, θα πρέπει να ισχύει η τελική ανθρώπινη απόφαση. Ειδικότερα, τα συστήματα Τεχνητής Νοημοσύνης δεν θα πρέπει να χρησιμοποιούνται για σκοπούς κοινωνικής βαθμολόγησης ή μαζικής επιτήρησης (UNESCO, 2021).

Ασφάλεια και προστασία

Οι ανεπιθύμητες βλάβες (safety risks), καθώς και οι διακινδυνεύσεις ασφάλειας (secure risks) θα πρέπει να αποφεύγονται και θα πρέπει να αντιμετωπίζονται, να προλαμβάνονται και να εξαλείφονται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης, ώστε να διασφαλίζεται η ασφάλεια του ανθρώπου, του περιβάλλοντος και του οικοσυστήματος. Η ασφαλής Τεχνητή Νοημοσύνη θα καταστεί δυνατή με την ανάπτυξη βιώσιμων πλαισίων πρόσβασης δεδομένων που προστατεύουν το απόρρητο, που ενισχύουν την καλύτερη εκπαίδευση και επικύρωση μοντέλων Τεχνητής Νοημοσύνης, αξιοποιώντας ποιοτικά δεδομένα (UNESCO, 2021).

Δικαιοσύνη και μη διάκριση

Οι φορείς της Τεχνητής Νοημοσύνης θα πρέπει να προάγουν την κοινωνική δικαιοσύνη και να προστατεύουν τη δικαιοσύνη και τη μη διάκριση οποιουδήποτε είδους, σύμφωνα με το διεθνές δίκαιο. Αυτό συνεπάγεται μια προσέγγιση χωρίς αποκλεισμούς ώστε τα οφέλη των τεχνολογιών Τεχνητής Νοημοσύνης να είναι διαθέσιμα και προσβάσιμα σε όλους. Τα κράτη μέλη θα πρέπει να εργαστούν για την προώθηση της πρόσβασης χωρίς αποκλεισμούς για όλους, για την αντιμετώπιση του ψηφιακού χάσματος και τη για τη συμμετοχή στην ανάπτυξη της Τεχνητής Νοημοσύνης. Σε διεθνές επίπεδο, οι πιο προηγμένες τεχνολογικά χώρες έχουν την ευθύνη αλληλεγγύης με τις λιγότερο προηγμένες για να διασφαλίσουν ότι τα οφέλη των τεχνολογιών Τεχνητής Νοημοσύνης κοινοποιούνται έτσι ώστε η πρόσβαση και η συμμετοχή στον κύκλο ζωής του συστήματος Τεχνητής Νοημοσύνης για τους τελευταίους να συμβάλλει σε μια δικαιότερη παγκόσμια τάξη πραγμάτων όσον αφορά την πληροφόρηση, την επικοινωνία, τον πολιτισμό, την εκπαίδευση, την έρευνα και την κοινωνικοοικονομική και πολιτική σταθερότητα (UNESCO, 2021).

Οι φορείς της Τεχνητής Νοημοσύνης θα πρέπει να καταβάλουν τις εύλογες προσπάθειες για να ελαχιστοποιήσουν και να αποφύγουν την ενίσχυση ή τη διαιώνιση των μεροληπτικών εφαρμογών και αποτελεσμάτων καθ' όλη τη διάρκεια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης, προκειμένου να διασφαλιστεί η δικαιοσύνη τέτοιων συστημάτων. Θα πρέπει να υπάρχει αποτελεσματική θεραπεία κατά των διακρίσεων και του μεροληπτικού αλγοριθμικού προσδιορισμού (UNESCO, 2021).

Βιωσιμότητα

Η έλευση των τεχνολογιών Τεχνητής Νοημοσύνης μπορεί είτε να ωφελήσει τους στόχους βιωσιμότητας είτε να εμποδίσει την υλοποίησή τους, ανάλογα με τον τρόπο εφαρμογής τους σε χώρες με διαφορετικά επίπεδα ανάπτυξης. Η συνεχής αξιολόγηση του ανθρώπινου, κοινωνικού, πολιτιστικού, οικονομικού και περιβαλλοντικού αντίκτυπου των τεχνολογιών τεχνητής νοημοσύνης θα πρέπει να πραγματοποιείται, έχοντας πλήρη επίγνωση των επιπτώσεων των τεχνολογιών Τεχνητής Νοημοσύνης για τη βιωσιμότητα ως ένα σύνολο συνεχώς εξελισσόμενων στόχων σε μια σειρά διαστάσεων, όπως αυτή τη στιγμή προσδιορίζονται στους Στόχους Βιώσιμης Ανάπτυξης των Ηνωμένων Εθνών (UNESCO, 2021).

Δικαίωμα απορρήτου και Προστασία Δεδομένων

Το απόρρητο, ένα δικαίωμα απαραίτητο για την προστασία της ανθρώπινης αξιοπρέπειας, αυτονομίας και δράσης, πρέπει να γίνεται σεβαστό, να προστατεύεται και να προωθείται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης. Είναι σημαντικό τα δεδομένα για συστήματα Τεχνητής Νοημοσύνης να συλλέγονται, να χρησιμοποιούνται, να μοιράζονται, να αρχειοθετούνται και να διαγράφονται με τρόπους που συνάδουν με το Διεθνές Δίκαιο και σύμφωνα με τις αξίες και τις αρχές που ορίζονται στην παρούσα Σύσταση, με ταυτόχρονη τήρηση των σχετικών εθνικών, περιφερειακών και διεθνών νομικών πλαισίων .

Τα αλγοριθμικά συστήματα απαιτούν επαρκείς αξιολογήσεις επιπτώσεων στο απόρρητο, οι οποίες περιλαμβάνουν επίσης κοινωνικούς και ηθικούς λόγους για τη χρήση τους και μια καινοτόμο χρήση της προσέγγισης της ιδιωτικής ζωής βάσει σχεδιασμού. Οι φορείς της Τεχνητής Νοημοσύνης πρέπει να διασφαλίσουν ότι είναι υπεύθυνοι για τον σχεδιασμό και την εφαρμογή συστημάτων Τεχνητής Νοημοσύνης με τέτοιο τρόπο ώστε να διασφαλίζεται ότι οι προσωπικές πληροφορίες προστατεύονται καθ' όλη τη διάρκεια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης (UNESCO, 2021).

Ανθρώπινη εποπτεία και αποφασιστικότητα

Τα κράτη μέλη θα πρέπει να διασφαλίζουν ότι είναι πάντα δυνατό να αποδίδεται ηθική και νομική ευθύνη για οποιοδήποτε στάδιο του κύκλου ζωής των συστημάτων Τεχνητής Νοημοσύνης, καθώς και σε περιπτώσεις αποκατάστασης που σχετίζονται με συστήματα Τεχνητής Νοημοσύνης, σε φυσικά πρόσωπα ή σε υφιστάμενες νομικές οντότητες. Η ανθρώπινη εποπτεία αναφέρεται επομένως όχι μόνο στην ατομική ανθρώπινη εποπτεία, αλλά στη δημόσια εποπτεία χωρίς αποκλεισμούς, ανάλογα με την περίπτωση. Αν και οι άνθρωποι μπορεί μερικές φορές να επιλέγουν να βασίζονται σε συστήματα τεχνητής νοημοσύνης για λόγους αποτελεσματικότητας, όσον αφορά στη λήψη αποφάσεων και δράση, ένα σύστημα Τεχνητής Νοημοσύνης δεν μπορεί ποτέ να αντικαταστήσει την τελική ανθρώπινη ευθύνη και υπευθυνότητα. Κατά κανόνα, οι αποφάσεις ζωής και θανάτου δεν πρέπει να εκχωρούνται στα συστήματα Τεχνητής Νοημοσύνης (UNESCO, 2021).

Διαφάνεια και επεξήγηση

Η διαφάνεια και η επεξήγηση των συστημάτων Τεχνητής Νοημοσύνης αποτελούν συχνά βασικές προϋποθέσεις για τη διασφάλιση του σεβασμού, της προστασίας και της προώθησης

των ανθρωπίνων δικαιωμάτων, των θεμελιωδών ελευθεριών και των ηθικών αρχών. Η διαφάνεια είναι απαραίτητη για να λειτουργούν αποτελεσματικά τα σχετικά εθνικά και διεθνή καθεστώτα ευθύνης. Η έλλειψη διαφάνειας θα μπορούσε επίσης να υπονομεύσει τη δυνατότητα αποτελεσματικής αμφισβήτησης αποφάσεων με βάση τα αποτελέσματα που παράγονται από συστήματα Τεχνητής Νοημοσύνης και, ως εκ τούτου, μπορεί να παραβιάσει το δικαίωμα σε δίκαιη δίκη και αποτελεσματική ένδικη προστασία και να περιορίσει τους τομείς στους οποίους μπορούν να χρησιμοποιηθούν νόμιμα αυτά τα συστήματα (UNESCO, 2021).

Επιπλέον, οι άνθρωποι θα πρέπει να ενημερώνονται πλήρως όταν μια απόφαση ενημερώνεται από ή λαμβάνεται με βάση αλγόριθμους Τεχνητής Νοημοσύνης, συμπεριλαμβανομένου του γεγονότος ότι επηρεάζει την ασφάλειά τους ή τα ανθρώπινα δικαιώματά τους, και σε αυτές τις περιπτώσεις θα πρέπει να έχουν την ευκαιρία να ζητούν επεξηγηματικές πληροφορίες. Τα άτομα θα πρέπει να έχουν πρόσβαση στους λόγους μιας απόφασης που επηρεάζει τα δικαιώματα και τις ελευθερίες τους και να έχουν τη δυνατότητα να υποβάλλουν τις αντιρρήσεις τους σε ένα εξουσιοδοτημένο μέλος του προσωπικού της εταιρείας του ιδιωτικού τομέα ή του δημόσιου τομέα ικανό να επανεξετάσει και να διορθώσει την απόφαση. Οι φορείς Τεχνητής Νοημοσύνης θα πρέπει να ενημερώνουν τους χρήστες όταν ένα προϊόν ή υπηρεσία παρέχεται απευθείας ή με τη βοήθεια συστημάτων Τεχνητής Νοημοσύνης έγκαιρα και με τον κατάλληλο τρόπο (UNESCO, 2021).

Η επεξηγησιμότητα συνδέεται στενά με τη διαφάνεια, καθώς τα αποτελέσματα και οι διαδικασίες που οδηγούν σε αποτελέσματα θα πρέπει να στοχεύουν στο να είναι κατανοητά και ανιχνεύσιμα, ανάλογα με το πλαίσιο. Οι φορείς της Τεχνητής Νοημοσύνης θα πρέπει να διασφαλίζουν ότι οι αλγόριθμοι που αναπτύσσονται είναι επεξηγήσιμοι. Στην περίπτωση εφαρμογών Τεχνητής Νοημοσύνης που επηρεάζουν τον τελικό χρήστη με τρόπο που δεν είναι προσωρινός, εύκολα αναστρέψιμος ή με άλλο τρόπο χαμηλού κινδύνου, θα πρέπει να διασφαλίζεται η παροχή ουσιαστικής εξήγησης, ώστε η απόφαση να θεωρείται διαφανής. Η διαφάνεια και η επεξηγησιμότητα συνδέονται στενά με την ευθύνη και τη λογοδοσία, καθώς και με την αξιοπιστία των συστημάτων Τεχνητής Νοημοσύνης (UNESCO, 2021).

Υπευθυνότητα και Λογοδοσία

Οι φορείς της Τεχνητής Νοημοσύνης και τα κράτη μέλη θα πρέπει να σέβονται, να προστατεύουν και να προωθούν τα ανθρώπινα δικαιώματα και τις θεμελιώδεις ελευθερίες και θα πρέπει επίσης να προωθούν την προστασία του περιβάλλοντος και των

οικοσυστημάτων, αναλαμβάνοντας την αντίστοιχη ηθική και νομική ευθύνη τους. Η ηθική ευθύνη και η ευθύνη για τις αποφάσεις και τις ενέργειες που βασίζονται με οποιονδήποτε τρόπο σε ένα σύστημα Τεχνητής Νοημοσύνης θα πρέπει πάντα να αποδίδεται τελικά στους φορείς της Τεχνητής Νοημοσύνης ανάλογα με το ρόλο τους στον κύκλο ζωής του συστήματος Τεχνητής Νοημοσύνης.

Θα πρέπει να αναπτυχθούν κατάλληλοι μηχανισμοί εποπτείας, εκτίμησης επιπτώσεων, ελέγχου και δέουσας επιμέλειας, συμπεριλαμβανομένης της προστασίας των καταγγελιών, για να διασφαλιστεί η λογοδοσία για τα συστήματα Τεχνητής Νοημοσύνης και τον αντίκτυπό τους σε όλο τον κύκλο ζωής τους. Τόσο οι τεχνικοί όσο και οι θεσμικοί σχεδιασμοί θα πρέπει να διασφαλίζουν τη δυνατότητα ελέγχου και ιχνηλασιμότητας (λειτουργίας) συστημάτων Τεχνητής Νοημοσύνης, ιδίως, για την αντιμετώπιση τυχόν σύγκρουσης με τους κανόνες και τα πρότυπα των ανθρωπίνων δικαιωμάτων και απειλών για την ευημερία του περιβάλλοντος και του οικοσυστήματος (UNESCO, 2021).

Ευαισθητοποίηση και Αλφαβητισμός

Η ευαισθητοποίηση του κοινού και η κατανόηση των τεχνολογιών Τεχνητής Νοημοσύνης και της αξίας των δεδομένων θα πρέπει να προωθούνται μέσω της ανοιχτής και προσβάσιμης εκπαίδευσης, της συμμετοχής στα κοινά, της εκπαίδευσης σε ψηφιακές δεξιότητες και στη δεοντολογία της Τεχνητής Νοημοσύνης, με επικεφαλής κυβερνήσεις, διακυβερνητικούς οργανισμούς, την κοινωνία των πολιτών, τους ακαδημαϊκούς, τα μέσα ενημέρωσης, τον ιδιωτικό τομέα, ώστε να διασφαλίζεται η αποτελεσματική συμμετοχή του κοινού και να μπορούν όλα τα μέλη της κοινωνίας να λαμβάνουν ενημερωμένες αποφάσεις σχετικά με τη χρήση των συστημάτων Τεχνητής Νοημοσύνης και να προστατεύονται από αθέμιτη επιρροή. Η προσέγγιση και η κατανόηση των συστημάτων Τεχνητής Νοημοσύνης θα πρέπει να βασίζεται στον αντίκτυπό τους στα ανθρώπινα δικαιώματα και στην πρόσβαση στα δικαιώματα, καθώς και στο περιβάλλον και στα οικοσυστήματα (UNESCO, 2021).

Πολυμετοχική και προσαρμοστική διακυβέρνηση και συνεργασία

Το διεθνές δίκαιο και η εθνική κυριαρχία πρέπει να γίνονται σεβαστά κατά τη χρήση των δεδομένων. Αυτό σημαίνει ότι τα κράτη, σύμφωνα με το διεθνές δίκαιο, μπορούν να ρυθμίζουν τα δεδομένα που παράγονται εντός ή που διέρχονται από την επικράτειά τους και να λαμβάνουν μέτρα για την προστασία των δεδομένων, με βάση το σεβασμό του δικαιώματος της ιδιωτικής ζωής σύμφωνα με το διεθνές δίκαιο και άλλους κανόνες και πρότυπα για τα ανθρώπινα δικαιώματα.

Η συμμετοχή διαφορετικών ενδιαφερομένων μερών καθ' όλη τη διάρκεια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης είναι απαραίτητη για περιεκτικές προσεγγίσεις στη διακυβέρνηση της Τεχνητής Νοημοσύνης, επιτρέποντας τα οφέλη να τα μοιράζονται όλοι και να συμβάλλουν στη βιώσιμη ανάπτυξη. Η υιοθέτηση ανοιχτών προτύπων και διαλειτουργικότητας για τη διευκόλυνση της συνεργασίας θα πρέπει να είναι σε ισχύ. Θα πρέπει να ληφθούν μέτρα για να ληφθούν υπόψη οι αλλαγές στις τεχνολογίες, η εμφάνιση νέων ομάδων ενδιαφερομένων και να επιτραπεί η ουσιαστική συμμετοχή περιθωριοποιημένων ομάδων, κοινοτήτων και ατόμων και αυτόχθονων πληθυσμών, με σεβασμό στη διαχείριση των δεδομένων τους (UNESCO, 2021).

Τομείς δράσης πολιτικής

Οι δράσεις πολιτικής που περιγράφονται στους ακόλουθους τομείς πολιτικής υλοποιούν τις αξίες και τις αρχές της Σύστασης που αναλύθηκαν ανωτέρω. Η κύρια δράση τους είναι να θεσπίσουν τα κράτη μέλη αποτελεσματικά μέτρα, συμπεριλαμβανομένων, για παράδειγμα, πλαισίων ή μηχανισμών πολιτικής και να διασφαλίσουν ότι άλλοι ενδιαφερόμενοι φορείς, όπως εταιρείες του ιδιωτικού τομέα, ακαδημαϊκά και ερευνητικά ιδρύματα και η κοινωνία των πολιτών, τα τηρούν. Τα κράτη μέλη οφείλουν να ενθαρρύνουν τους ενδιαφερόμενους φορείς να αναπτύξουν εργαλεία αξιολόγησης σεβόμενοι τα ανθρώπινα δικαιώματα, το κράτος δικαίου και τη δημοκρατία και πάντα με την καθοδήγηση των Κατευθυντήριων Αρχών των Ηνωμένων Εθνών για τις Επιχειρήσεις και τα Ανθρώπινα Δικαιώματα, ώστε να μειώσουν τις ηθικές επιπτώσεις. Η διαδικασία για την ανάπτυξη τέτοιων πολιτικών ή μηχανισμών θα πρέπει να περιλαμβάνει όλους τους ενδιαφερομένους και θα πρέπει να λαμβάνει υπόψη τις περιστάσεις και τις προτεραιότητες κάθε κράτους μέλους (UNESCO, 2021).

Συγκεκριμένα, οι δράσεις πολιτικής αναφέρονται στους κάτωθι έντεκα (11) τομείς πολιτικής δράσης:

- **Αξιολόγηση ηθικών επιπτώσεων**

  Τα κράτη μέλη θα πρέπει να εισαγάγουν πλαίσια για τις εκτιμήσεις επιπτώσεων, όπως η ηθική αξιολόγηση επιπτώσεων, για τον εντοπισμό και την αξιολόγηση των οφελών, των ανησυχιών και των κινδύνων των συστημάτων Τεχνητής Νοημοσύνης, καθώς και κατάλληλα μέτρα πρόληψης, μετριασμού και παρακολούθησης κινδύνων, μεταξύ άλλων μηχανισμών διασφάλισης. Τέτοιες εκτιμήσεις επιπτώσεων θα πρέπει να προσδιορίζουν τις ηθικές και κοινωνικές επιπτώσεις στα ανθρώπινα δικαιώματα

και στις θεμελιώδεις ελευθερίες, στα εργασιακά δικαιώματα, στο περιβάλλον και στα οικοσυστήματα (UNESCO, 2021).

- **Ηθική διακυβέρνηση και εποπτεία**

  Τα κράτη μέλη θα πρέπει να διασφαλίζουν ότι οι μηχανισμοί διακυβέρνησης της Τεχνητής Νοημοσύνης είναι περιεκτικοί, διαφανείς, πολυεπιστημονικοί, πολυμερείς (αυτό περιλαμβάνει τη δυνατότητα μετριασμού και αποκατάστασης της ζημίας διασυνοριακά) και πολυμετοχικοί. Ειδικότερα, η διακυβέρνηση θα πρέπει να περιλαμβάνει πτυχές πρόβλεψης και αποτελεσματικής προστασίας, παρακολούθησης των επιπτώσεων, επιβολής και αποκατάστασης (UNESCO, 2021).

- **Πολιτική δεδομένων**

  Τα κράτη μέλη θα πρέπει να εργαστούν για να αναπτύξουν στρατηγικές διακυβέρνησης δεδομένων που να διασφαλίζουν τη συνεχή αξιολόγηση της ποιότητας των δεδομένων εκπαίδευσης για συστήματα Τεχνητής Νοημοσύνης, συμπεριλαμβανομένης της επάρκειας των διαδικασιών συλλογής και επιλογής δεδομένων, κατάλληλων μέτρων ασφάλειας και προστασίας δεδομένων, καθώς και μηχανισμών ανατροφοδότησης για μάθηση από λάθη και επικοινωνία των βέλτιστων πρακτικών μεταξύ όλων των παραγόντων της Τεχνητής Νοημοσύνης (UNESCO, 2021).

- **Ανάπτυξη και διεθνής συνεργασία**

  Τα κράτη μέλη και οι διεθνικές εταιρείες θα πρέπει να δώσουν προτεραιότητα στην ηθική της Τεχνητής Νοημοσύνης συμπεριλαμβάνοντας συζητήσεις για ηθικά ζητήματα που σχετίζονται με την Τεχνητή Νοημοσύνη σε σχετικά διεθνή, διακυβερνητικά φόρουμ και φόρουμ πολλών ενδιαφερομένων (UNESCO, 2021).

- **Περιβάλλον και οικοσυστήματα**

  Τα κράτη μέλη και οι επιχειρήσεις θα πρέπει να αξιολογούν τις άμεσες και έμμεσες περιβαλλοντικές επιπτώσεις σε όλο τον κύκλο ζωής του συστήματος Τεχνητής Νοημοσύνης, συμπεριλαμβανομένου, ενδεικτικά, του αποτυπώματος άνθρακα, της κατανάλωσης ενέργειας και τον περιβαλλοντικό αντίκτυπο της εξόρυξης πρώτων υλών για την υποστήριξη της κατασκευής τεχνολογιών Τεχνητής Νοημοσύνης και τη μείωση των περιβαλλοντικών επιπτώσεων των συστημάτων Τεχνητής Νοημοσύνης και των υποδομών δεδομένων. Τα κράτη μέλη θα πρέπει να διασφαλίζουν τη συμμόρφωση όλων των παραγόντων της Τεχνητής Νοημοσύνης με την περιβαλλοντική νομοθεσία, τις πολιτικές και τις πρακτικές στον εν λόγω τομέα (UNESCO, 2021).

- **Φύλο**

  Τα κράτη μέλη θα πρέπει να διασφαλίσουν ότι οι δυνατότητες των ψηφιακών τεχνολογιών και της Τεχνητής Νοημοσύνης που συμβάλλουν στην επίτευξη της ισότητας των φύλων μεγιστοποιούνται πλήρως και πρέπει να διασφαλίζουν ότι τα ανθρώπινα δικαιώματα και οι θεμελιώδεις ελευθερίες των κοριτσιών και των γυναικών, καθώς και η ασφάλεια και η ακεραιότητά τους δεν παραβιάζονται σε καμία περίπτωση σε οποιοδήποτε στάδιο του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης. Επιπλέον, η αξιολόγηση των ηθικών επιπτώσεων θα πρέπει να περιλαμβάνει μια εγκάρσια προοπτική του φύλου (UNESCO, 2021).

- **Πολιτισμός**

  Τα κράτη μέλη ενθαρρύνονται να ενσωματώνουν συστήματα Τεχνητής Νοημοσύνης, όπου χρειάζεται, στη διατήρηση, στον εμπλουτισμό, στην κατανόηση, στην προώθηση, στη διαχείριση και στην προσβασιμότητα της υλικής και άυλης πολιτιστικής κληρονομιάς, συμπεριλαμβανομένων των γλωσσών που απειλούνται με εξαφάνιση καθώς και των αυτόχθονων γλωσσών και γνώσεων, π.χ. εισαγωγή ή ενημέρωση εκπαιδευτικών προγραμμάτων που σχετίζονται με την εφαρμογή συστημάτων Τεχνητής Νοημοσύνης σε αυτούς τους τομείς και διασφαλίζοντας μια συμμετοχική προσέγγιση που απευθύνεται τόσο σε ιδρύματα όσο και στο κοινό (UNESCO, 2021).

- **Εκπαίδευση κι έρευνα**

  Τα κράτη μέλη θα πρέπει να συνεργαστούν με διεθνείς οργανισμούς, εκπαιδευτικά ιδρύματα και ιδιωτικές και μη κυβερνητικές οντότητες για την παροχή κατάλληλης εκπαίδευσης Τεχνητής Νοημοσύνης στο κοινό σε όλα τα επίπεδα σε όλες τις χώρες προκειμένου να ενδυναμωθούν οι άνθρωποι και να μειωθούν τα ψηφιακά χάσματα και οι ανισότητες ψηφιακής πρόσβασης που προκύπτουν από την ευρεία υιοθέτηση συστημάτων Τεχνητής Νοημοσύνης (UNESCO, 2021).

- **Επικοινωνία κι ενημέρωση**

  Τα κράτη μέλη θα πρέπει να χρησιμοποιούν συστήματα Τεχνητής Νοημοσύνης για να βελτιώσουν την πρόσβαση σε πληροφορίες και γνώση. Αυτό μπορεί να περιλαμβάνει υποστήριξη σε ερευνητές, ακαδημαϊκό κόσμο, δημοσιογράφους, στο ευρύ κοινό και στους προγραμματιστές, για ενίσχυση της ελευθερίας της έκφρασης, των ακαδημαϊκών και επιστημονικών ελευθεριών, της πρόσβασης σε πληροφορίες και της αυξημένης προληπτικής αποκάλυψης επίσημων δεδομένων και πληροφοριών (UNESCO, 2021).

- **Οικονομία κι εργασία**

  Τα κράτη μέλη θα πρέπει να αξιολογήσουν και να αντιμετωπίσουν τον αντίκτυπο των συστημάτων Τεχνητής Νοημοσύνης στις αγορές εργασίας και τις επιπτώσεις τους στις εκπαιδευτικές απαιτήσεις, σε όλες τις χώρες και με ιδιαίτερη έμφαση στις χώρες όπου η οικονομία είναι έντασης εργασίας. Αυτό μπορεί να περιλαμβάνει την εισαγωγή ενός ευρύτερου φάσματος διεπιστημονικών δεξιοτήτων σε όλα τα επίπεδα εκπαίδευσης για να παρέχει στους σημερινούς εργαζόμενους και στις νέες γενιές δίκαιη ευκαιρία να βρουν θέσεις εργασίας σε μια ταχέως μεταβαλλόμενη αγορά και να διασφαλίσει την επίγνωσή τους σχετικά με τις ηθικές πτυχές των συστημάτων Τεχνητής Νοημοσύνης. Δεξιότητες όπως το «μαθαίνοντας πώς να μαθαίνεις», η επικοινωνία, η κριτική σκέψη, η ομαδική εργασία, η ενσυναίσθηση και η ικανότητα μεταφοράς της γνώσης σε διάφορους τομείς, θα πρέπει να διδάσκονται παράλληλα με ειδικές, τεχνικές δεξιότητες, καθώς και εργασίες χαμηλής ειδίκευσης. Η διαφάνεια σχετικά με τις δεξιότητες που απαιτούνται και η ενημέρωση των προγραμμάτων σπουδών γύρω από αυτές είναι βασικές (UNESCO, 2021).

- **Υγεία και κοινωνική ευημερία**

  Τα κράτη μέλη θα πρέπει να προσπαθούν να χρησιμοποιούν αποτελεσματικά συστήματα Τεχνητής Νοημοσύνης για τη βελτίωση της ανθρώπινης υγείας και την προστασία του δικαιώματος στη ζωή, συμπεριλαμβανομένου του μετριασμού των εστιών ασθενειών, ενώ οικοδομούν και διατηρούν τη διεθνή αλληλεγγύη για την αντιμετώπιση των παγκόσμιων κινδύνων και αβεβαιοτήτων για την υγεία και να διασφαλίζουν ότι η ανάπτυξη συστημάτων Τεχνητής Νοημοσύνης στην υγειονομική περίθαλψη είναι συνεπείς με το διεθνές δίκαιο και τις υποχρεώσεις που απορρέουν από το δίκαιο των ανθρωπίνων δικαιωμάτων. Τα κράτη μέλη θα πρέπει να διασφαλίζουν ότι οι φορείς που εμπλέκονται στα συστήματα Τεχνητής Νοημοσύνης στον τομέα της υγείας λαμβάνουν υπόψη τη σημασία των σχέσεων του ασθενούς με την οικογένειά του και με το προσωπικό υγειονομικής περίθαλψης (UNESCO, 2021).

Παρακολούθηση και Αξιολόγηση

Τα κράτη μέλη θα πρέπει, να παρακολουθούν και να αξιολογούν με αξιοπιστία και διαφάνεια τις πολιτικές, τα προγράμματα και τους μηχανισμούς που σχετίζονται με τη δεοντολογία της Τεχνητής Νοημοσύνης, χρησιμοποιώντας έναν συνδυασμό ποσοτικών και ποιοτικών προσεγγίσεων. Για την υποστήριξη των κρατών μελών, η UNESCO μπορεί να συνεισφέρει με:

- ανάπτυξη μιας μεθοδολογίας της UNESCO για την αξιολόγηση ηθικών επιπτώσεων των τεχνολογιών Τεχνητής Νοημοσύνης, καθοδήγηση για την εφαρμογή της σε όλα τα στάδια του κύκλου ζωής του συστήματος Τεχνητής Νοημοσύνης και υλικό για ανάπτυξη ικανοτήτων για την υποστήριξη των προσπαθειών των κρατών μελών να εκπαιδεύσουν κυβερνητικούς αξιωματούχους, υπευθύνους χάραξης πολιτικής και άλλους συναφείς φορείς Τεχνητής Νοημοσύνης στην εν λόγω μεθοδολογία.
- ανάπτυξη μιας μεθοδολογίας αξιολόγησης ετοιμότητας της UNESCO για να βοηθήσει τα κράτη μέλη να προσδιορίσουν την κατάστασή τους για την εφαρμογή της παρούσας Σύστασης.
- ανάπτυξη μιας μεθοδολογίας της UNESCO για την αξιολόγηση εκ των προτέρων και εκ των υστέρων της αποτελεσματικότητας και της αποδοτικότητας των πολιτικών για την ηθική και τα κίνητρα της Τεχνητής Νοημοσύνης έναντι καθορισμένων στόχων.
- ενίσχυση της έρευνας και της ανάλυσης βάσει στοιχείων και της υποβολής εκθέσεων σχετικά με τις πολιτικές που αφορούν την ηθική της Τεχνητής Νοημοσύνης.
- συλλογή και διάδοση προόδου, καινοτομιών, ερευνητικών εκθέσεων, επιστημονικών δημοσιεύσεων, δεδομένων και στατιστικών σχετικά με τις πολιτικές για την ηθική της Τεχνητής Νοημοσύνης, μέσω υφιστάμενων πρωτοβουλιών, για την υποστήριξη της ανταλλαγής βέλτιστων πρακτικών και αμοιβαίας μάθησης για την προώθηση της εφαρμογής της παρούσας Σύστασης.

Οι διαδικασίες παρακολούθησης και αξιολόγησης θα πρέπει να διασφαλίζουν την ευρεία συμμετοχή όλων των ενδιαφερομένων, συμπεριλαμβανομένων, ενδεικτικά, των ευάλωτων ατόμων ή των ατόμων σε ευάλωτες καταστάσεις. Θα πρέπει να διασφαλιστεί η κοινωνική, πολιτιστική και η ποικιλομορφία των φύλων, με σκοπό τη βελτίωση των διαδικασιών μάθησης και την ενίσχυση των συνδέσεων μεταξύ των πορισμάτων, της λήψης αποφάσεων, της διαφάνειας και της λογοδοσίας για τα αποτελέσματα (UNESCO, 2021).

Συμπεράσματα

Οι πολλαπλές πτυχές της Τεχνητής Νοημοσύνης και η υιοθέτηση διαφορετικών προσεγγίσεων από εθνικά κράτη, Ευρωπαϊκή Ένωση, δημόσιο και ιδιωτικό τομέα, παγκόσμιους οργανισμούς, ακαδημαϊκή κοινότητα, έχει ως αποτέλεσμα την ύπαρξη πολλαπλών ορισμών της, οι οποίοι αντικατοπτρίζουν τη διαφορετική στρατηγική στοχοθέτηση των αντίστοιχων πολιτικών, οικονομικών, πολιτιστικών και κοινωνικών συστημάτων. Η ανυπαρξία, ωστόσο, ενός τυπικού ορισμού για το τι περιλαμβάνει στην πραγματικότητα η Τεχνητή Νοημοσύνη δυσκολεύει μια κοινή κατανόηση του τομέα, των δυνατοτήτων, του πεδίου και των επιπτώσεων εφαρμογής του και δυσχεραίνει την εφαρμογή ενιαίων πολιτικών παρεμβάσεων πρόληψης και αντιμετώπισης των ανησυχιών και κινδύνων που προκύπτουν. Αναγνωρίζοντας την ανάγκη ύπαρξης ενός κοινά αποδεκτού ορισμού που να περιλαμβάνει το εύρος των τεχνολογικών συστημάτων και να ενσωματώνει κατανόηση των εφαρμογών και της χρήσης τους, το Παρατηρητήριο για την Τεχνητή Νοημοσύνη δημιούργησε μια ταξινομία με στόχο τη χαρτογράφηση και τον χαρακτηρισμό του παγκόσμιου τοπίου της Τεχνητής Νοημοσύνης.

Η έλλειψη ενός κοινά κατανοητού ορισμού της Τεχνητής Νοημοσύνης έχει ως συνέπεια ο όρος να χρησιμοποιείται ως μια γενική έννοια που περιλαμβάνει οτιδήποτε το οποίο υποστηρίζεται ότι εμφανίζει χαρακτηριστικά που ορισμένοι περιγράφουν ως «έξυπνα». Σε μια προσπάθεια αποσαφήνισης της τεχνολογίας της Τεχνητής Νοημοσύνης προχωρήσαμε στην κατηγοριοποίηση των βασικών της τεχνολογιών, αποδίδοντας την εξελικτική της πορεία μέσω της ανάπτυξης διαφορετικών προσεγγίσεων, από τα Έμπειρα Συστήματα στη Μηχανική Μάθηση και από τη Μηχανική Μάθηση στην κατάκτηση της Γενικής Τεχνητής Νοημοσύνης που, ίσως, μελλοντικά οδηγήσει στην Υπερνοημοσύνη.

Η υιοθέτηση και χρήση της Τεχνητής Νοημοσύνης αποτελεί μια απαιτητική πρόκληση για το δημόσιο τομέα. Η ευρύτερη υιοθέτησή της στον δημόσιο τομέα συνήθως ακολουθεί μόλις τεθούν σε εφαρμογή οι κατάλληλες προϋποθέσεις, όπως επαρκής ψηφιακή υποδομή, επαρκείς ψηφιακές δεξιότητες, νομικά πλαίσια που επιτρέπουν την εφαρμογή της και ψηφιακές στρατηγικές.

Ορισμένες εθνικές στρατηγικές ευρωπαϊκών κρατών μελών για την Τεχνητή Νοημοσύνη φαίνεται να επικεντρώνονται περισσότερο στη συνεργασία δημόσιου και ιδιωτικού τομέα για την ανάπτυξη και την υιοθέτηση της Τεχνητής Νοημοσύνης στο δημόσιο τομέα, ενώ άλλες είναι περισσότερο προσανατολισμένες στην αντιμετώπιση των φραγμών που

σχετίζονται με τα δεδομένα. Επιπλέον, κάποιες άλλες στρατηγικές εστιάζουν στη βελτίωση της εσωτερικής ικανότητας των δημοσίων διοικήσεων και δημοσίων υπάλληλων, ως μια άλλη σημαντική παράμετρος που θα τονώσει την υιοθέτηση της Τεχνητής Νοημοσύνης στο δημόσιο τομέα.

Οι περιπτώσεις χρήσης Τεχνητής Νοημοσύνης που αναλύθηκαν, αναδεικνύουν τις ευκαιρίες της τεχνολογίας που ήδη αξιοποιούνται στο δημόσιο τομέα και οι οποίες μπορούν να εμπνεύσουν και να αξιοποιηθούν από άλλους δημόσιους οργανισμούς στην εφαρμογή της Τεχνητής Νοημοσύνης. Οι εν λόγω εφαρμογές εξυπηρετούν διάφορους τομείς του δημόσιου τομέα, κάνοντας χρήση διαφορετικών τεχνολογιών Τεχνητής Νοημοσύνης και στοχεύουν, κυρίως, στη βελτίωση της ποιότητας των δημόσιων υπηρεσιών ή στη βελτίωση της διοικητικής αποτελεσματικότητας ενώ άλλες στη βελτίωση των ικανοτήτων των κυβερνήσεων να γίνουν πιο ανοικτές, αυξάνοντας τη διαφάνεια ή τη συμμετοχή των πολιτών στη λήψη αποφάσεων.

Οι πιθανές ευκαιρίες και τα οφέλη των τεχνολογιών Τεχνητής Νοημοσύνης για το δημόσιο τομέα φαίνονται εντυπωσιακά, αλλά συνοδεύονται από ορισμένους ηθικούς κινδύνους, οι οποίοι αν δεν αντιμετωπιστούν επιτυχώς, βλάπτουν τη νομιμότητα των εισροών, της διεκπεραίωσης και των εκροών των δημόσιων οργανισμών με δυσμενείς επιπτώσεις για τους εμπλεκόμενους. Η εσφαλμένη χρήση της Τεχνητής Νοημοσύνης μπορεί να οδηγήσει σε άδικα αποτελέσματα, ενισχύοντας μεροληπτικές προκαταλήψεις. Επιπλέον, οι αποφάσεις που λαμβάνονται με ή από συστήματα Τεχνητής Νοημοσύνης μπορεί να γίνουν πιο αδιαφανείς και πιο δύσκολο να αιτιολογηθούν. Οι πρακτικές συλλογής και ανάλυσης δεδομένων θα μπορούσαν να απειλήσουν περαιτέρω την ιδιωτική ζωή των πολιτών ή να καταστήσουν την επιτήρησή τους από τις κυβερνήσεις, ως συνήθη διαδικασία, απειλώντας τις πολιτικές ελευθερίες. Αυτές οι αντιληπτές ηθικές ανησυχίες καθιστούν λιγότερο πιθανό τις δημόσιες διοικήσεις και οργανισμούς να ξεκινήσουν έργα που σχετίζονται με την Τεχνητή Νοημοσύνη ή, μετά από ένα επιτυχημένο πιλοτικό έργο, ανεπίλυτα ηθικά ζητήματα μπορεί να σταματήσουν την υλοποίηση. Ως εκ τούτου, η διασφάλιση της επίλυσης ηθικών διλημμάτων και ανησυχιών σχετικά με την Τεχνητή Νοημοσύνη θα καθιστούσε τη βιώσιμη χρήση των τεχνολογιών της πιο πιθανή στο δημόσιο τομέα. Επομένως, οι κυβερνήσεις θα πρέπει να γνωρίζουν αυτές τις διαφορετικές επιπτώσεις που μπορεί να έχει η Τεχνητή Νοημοσύνη στην κοινωνία και να λαμβάνουν τα κατάλληλα μέτρα για να τις αποτρέψουν, μετριάζοντας τους κινδύνους προκειμένου να διασφαλιστεί η ανθρωποκεντρική χρήση και η αξιοπιστία της τεχνολογίας.

Οι αλγόριθμοι έχουν φτάσει σε τέτοιο επίπεδο ώστε να επηρεάζουν όλο και περισσότερο τα άτομα, τους οργανισμούς και την κοινωνία. Η απειλή της αλγοκρατίας υφίσταται και σχετίζεται με την αδιαφάνεια ορισμένων αλγοριθμικών συστημάτων διακυβέρνησης, τα οποία λειτουργούν με τρόπους που είναι απροσπέλαστοι και αδιαφανείς για την ανθρώπινη λογική και κατανόηση, εγείροντας ζητήματα νομιμότητας σε σχέση με τις διαδικασίες λήψης αποφάσεων του δημόσιου τομέα.

Αν και έχουμε μια καλά αναπτυγμένη κατανόηση της ελλαττωματικής φύσης της ανθρώπινης λήψης αποφάσεων, όπως και ένα σύνολο νομικών και θεσμικών μηχανισμών που αποσκοπούν στην παροχή διασφαλίσεων έναντι των χειρότερων υπερβολών του, όσο ατελείς κι αν είναι, δεν έχουμε μια ισοδύναμη ολοκληρωμένη και συστηματική αναφορά και εμπειρία των πιθανών ελαττωμάτων και μειονεκτημάτων που σχετίζονται με τη λήψη αλγοριθμικών αποφάσεων. Οι ηθικές ανησυχίες που σχετίζονται τόσο με τη διαδικασία και τα αποτελέσματα που δημιουργούνται από την αλγοριθμική λήψη αποφάσεων όσο και από τη χρήση τέτοιων συστημάτων για την πρόβλεψη και εξατομίκευση υπηρεσιών μεταφράζονται σε κινδύνους για τα δικαιώματα των ατόμων και των επηρεαζόμενων ομάδων. Θα μπορούσαμε να θεωρήσουμε ότι αυτοί οι κίνδυνοι που συνδέονται με τα αλγοριθμικά συστήματα λήψης αποφάσεων είναι κίνδυνοι συστημικού χαρακτήρα, οι οποίοι, εάν αποτύχουμε να διασφαλίσουμε ότι εφαρμόζονται επαρκή μέτρα, μπορεί να διαβρώσουν τα ηθικά, πολιτικά και πολιτιστικά θεμέλια της κοινωνίας με αποτέλεσμα να υπονομεύσουν το δημοκρατικό πολιτικό σύστημα και μαζί του την ατομική μας ελευθερία, αυτονομία και ικανότητα αυτοδιάθεσης.

Οι «Κατευθυντήριες Γραμμές Δεοντολογίας για Αξιόπιστη Τεχνητή Νοημοσύνη» της Ευρωπαϊκής Ένωσης και η «Σύσταση για την Ηθική της Τεχνητής Νοημοσύνης» της UNESCO, αν και δεν είναι δεσμευτικές, προσπαθούν να αντιμετωπίσουν τις ανησυχίες που προκύπτουν από τη χρήση αλγοριθμικών συστημάτων, δίνοντας κατευθύνσεις που μπορούν να συμβάλλουν στην επίτευξη ηθικής και αξιόπιστης Τεχνητής Νοημοσύνης. Ωστόσο, οι πολιτιστικές διαφορές θα καταστήσουν πολύ δύσκολη την επίτευξη κοινής κατανόησης σε παγκόσμιο επίπεδο σχετικά με τα ηθικά πλαίσια και την αποτελεσματική εφαρμογή τους. Η ηθική δεν μπορεί να συναχθεί απλώς από τους κώδικες. Η ερμηνεία τους θα παίξει τον καθοριστικότερο ρόλο γιατί αυτή η ερμηνεία θα ποικίλλει πολύ ανάλογα με το πολιτισμικό πλαίσιο. Ορισμένοι από τους κώδικες που αναπτύχθηκαν, κυρίως, με βάση τις αρχές του δυτικού διαφωτισμού, αμφισβητούνται ήδη από ειδικούς από άλλα μέρη του κόσμου που ισχυρίζονται ότι έρχονται σε αντίθεση με τις ινδουιστικές ή ισλαμικές αξίες. Είναι,

επομένως, εξαιρετικά ενδιαφέρουσα η προσπάθεια της UNESCO να «παγκοσμιοποιήσει» τις ηθικές αρχές για την Τεχνητή Νοημοσύνη και η υιοθέτησή τους από 193 χώρες κρίνεται ιδιαίτερα σημαντική.

Βιβλιογραφία

1. Acquisti, A., Brandimarte, L., Lowenstein, G. (2015). *Privacy and Human Behavior in the Age of Information*. Science 347(6221): 509-14.
2. Bishop, M. & Trout, JD. (2002). 50 Years of Successful Predictive Modeling Should be Enough: Lessons for Philosophy of Science. *Philosophy of Science: PSA 2000 Symposium Papers*, 2002 69 (supplement): S197-S208.
3. Black, J. (1997). *Rules and Regulators.* Oxford University Press.
4. Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: How might the New Zealand government respond? *Policy Quarterly*, 13(4), 36–44.
5. Casey, A., Niblett, A. (2016). '*Self-Driving Laws'*. 66 University of Toronto Law Journal 429.
6. Citron, D. and Pasquale, F. (2014). *The Scored Society: Due Process for Automated Predictions*. Washington Law Review 86: 101.
7. Cranor, L., Frischmann, B., Harkins, R. (2013). '*Panel I: Disclosure and Notice Practices in Private Data Collection*'. 32 Cardozo Arts & Entertainment Law Journal 781.
8. Dourish, P. (2016). 'Algorithms and Their Others: Algorithmic Culture in Context' *Big Data & Society* 3: 2 doi.org/10.1177/2053951716665128
9. Endicott, T. (2005). 'The Value of Vagueness'. In: Bhatia, V., Engberg, J., Gotti, M., Heller, D. *Vagueness in Normative Texts*.
10. Endicott, T. (2015). *Administrative Law.* 3rd edn. Oxford University Press.
11. Ferguson, A. (2017). 'Policing Predictive Policing'. Washington University Law Review 1115.
12. Galligan, D. (1996). *Due Process and Fair Procedures*. Oxford University Press: Oxford.
13. Gardner, J. (2006). '*The Mark of Responsibility (With a Postscript on Accountability)*': In MW Dowdle (ed). Public Accountability. Cambridge University Press.
14. Gillespie, T. (2013). *'The Relevance of Algorithms.*' In Gillespie, T, Boczkowski, P & Foot, K. (eds.) Media Technologies: Essays on Communication, Materiality and Society, Cambridge. MIT Press: MA.
15. Hildebrandt, M. (2015). *Smart Technologies and the End(s) of Law*. Edward Elgar.
16. Hildebrandt, M. Gutwirth, S. (2010). *Profiling the European Citizen*. Springer.
17. Jasanoff, S. (2016). *The Ethics of Invention*. WW Norton & Co.
18. Kamiran, F., Zliobaite, I. (2012). 'Explainable and Non-explainable Discrimination in Classification' in B Custers, T Calders, B Schermer, and T Zarsky (eds). *Discrimination and Privacy in the Information Society*: Springer.
19. Kitchin, R. (2014a). 'Big Data, New Epistemologies and Paradigm Shifts' *Big Data & Society* 1-12.
20. Kosinski, M., Stillwell, D., Graepel, T. (2013). '*Private Traits and Attributes Are Predictable from Digital Records of Human Behaviours'*. Proceedings National Academy of Science 110 (15) 5802-5805.
21. Lee, M. (2017), 'The Legal Institutionalization of Public Participation in the EU Governance of Technology' in R Brownsword, E Scotford, and K Yeung (eds), *The Oxford Handbook of Law, Regulation and Technology*. Oxford University Press.
22. Matthias, A. (2004). 'The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata'. 6 Ethics and Information Technology.
23. Mayer-Schonberger, V. and Cukier, K. (2013) *Big Data*. London: John Murray.

24. Mayer-Schonberger, V. and Cukier, K. (2013). *Big Data: A Revolution that Will Transform How we Live Work and Think*. John Murray.
25. Mehr, H. (2017). *Artificial intelligence for citizen services and government*. Cambridge, MA: Harvard Kennedy School, Ash Center for Democratic Governance and Innovation.
26. Nissenbaum, H. (2011). '*A Contextual Approach to Privacy Online'*. 140(4) Daedalus, 32.
27. O'Neil, C. (2016). '*Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy'*. Allen Lane: New York.
28. Oswald, M. (2018). '*Algorithm-Assisted Decision-Making in the Public Sector: Framing the Issues Using Administrative Law Rules Governing Discretionary Power'*. Philosophical Transactions of the Royal Society. A doi: 10.1098.
29. Parens, E. (2010). '*The Ethics of Memory Blunting and the Narcissism of Small Differences'*. 3(2) Neuroethics, 99.
30. Pasquale, F. (2015). *The Black Box Society*. Harvard University Press: Boston.
31. Pentland, A. (2014). *Social Physics*. London: Penguin Press.
32. Richards, NM. (2013). '*The Dangers of Surveillance'*. 126 Harvard Law Review 1934.
33. Rifkin, J. (2014). *The Zero Marginal Cost Society*: The Internet of Things, The Collaborative Commons and the Eclipse of Capitalism. Palgrave MacMillan.
34. Roth, A. (2016). '*Trial by Machine'*. 104 Georgetown Law Journal 1245.
35. Seaver, N. (2013). *Knowing Algorithms*. In Media in Transition 8, Cambridge MA.
36. Skitka, LJ., Mosier, K., Burdick, MD. (2000). '*Accountability and Automation Bias'*. International Journal of Human-Computer Studies 701.
37. Thaler, R., Sunstein, C. (2008). *Nudge*. Penguin Books.
38. Townley, C., Morrison, E., Yeung, K. (2017). '*Big Data and Personalized Price Discrimination in EU Competition Law'*. 36(1) Yearbook of European Law 683.
39. Veale, M., Edwards, L. (2018). *'Clarity, Surprises, and Further Questions in the Article 29 Working Party Draft Guidance on automated Decision-Making and Profiling*'. 34 Computer Law & Security Review 398.
40. Wardle, C. Derakhshan, H. (2017). '*Information Disorder: Towards an Interdisciplinary Framework for Research and Policy Making'* Report DGI (2017) 09. Strasbourg. Council of Europe.
41. Weller, A. (2017). 'Challenges for Transparency' Paper presented at 2017 ICML Workshop on Human Interpretability in Machine Learning. Sydney, NSW, Australia.
42. Yeung, K. (2012). 'Nudge as Fudge' *Modern Law Review* 75(1): 122.
43. Yeung, K. (2017). "Hypernudge": Big Data as a Mode of Regulation by Design'. *Information, Communication & Society* 20(1): 118-136.
44. Yeung, K. (2018a). 'Algorithmic Regulation: A Critical Interrogation'. *4 Regulation & Governance*: 505.
45. Yeung, K. (2018b). 'Five Fears about Mass Predictive Personalization in an Age of Surveillance Capitalism'. *International Data Privacy Law* 8(3): 258–69.
46. Yeung, K. (2019). '*Why Worry about Decision-Making by Machine?*'. Oxford University Press.
47. Zarsky, T. (2011). Governmental Data-Mining and Its Alternatives. *Penn State Law Review* 116: 285
48. Zarsky, T. (2012). Automated Predictions: Perception, Law and Policy. *Communications of the ACM* 15(9): 33-35
49. Zuboff, S. (2019). '*The Age of Surveillance Capitalism'*. Profile Books.

Ηλεκτρονικές πηγές

1. AuroraAI (2019). *AuroraAI – Towards a Humancentric Society*. Available: https://vm.fi/documents/10623/1464506/AuroraAI+development+and+implementation+plan+2019%E2%80%932023.pdf (accessed 12/7/22).
2. Berryhill, J., Kok Heang, K., Clagher, R., McBride, K. (2019). *Hello, World: Artificial intelligence and its use in the public sector.* OECD Working Papers on Public Governance, No 36. OECD Publishing. Available: https://www.oecd-ilibrary.org/governance/hello-world_726fd39d-en (accessed 12/7/22).
3. Boden, M., Bryson, J., Caldwell, D. (2011). *'Principles of Robotics'* Engineering and Physical Sciences Research Council. Available: www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/ (accessed 6/5/22).
4. Brundage, M. Avin, S., Clark, J. (2018). '*The Malicious Use of AI: Forecasting, Prevention and Mitigation*'. Available: https://maliciousaireport.com/ (accessed 6/5/22).
5. Bumbulsky, J. (2013). *Chaotic Storage Lessons*. Medium. Available: https://medium.com/tech-talk/e3b7de266476 (accessed 10/5/22).
6. Buranyi, St. (2018). *Dehumanizing, impenetrable, frustrating: the grim reality of job hunting in the age of AI.* The Guardian. Available: https://www.theguardian.com/inequality/2018/mar/04/dehumanising-impenetrablefrustrating-the-grim-reality-of-job-hunting-in-the-age-of-ai (accessed 10/5/22).
7. Chen, S. (2019). *"Is Fraud-Busting AI Systems Being Turned Off for Being Too Efficient?"* South China Morning Post, February 4, 2019. Available: https://www.scmp.com/news/china/science/article/2184857/chinas-corruption-busting-ai-system-zero-trust-being-turned-being (accessed 12/07/22).
8. Danaher, J. (2014). *Rule by Algorithm? Big Data and the Threat of Algocracy*. Philosophical Disquisitions. Available: http://philosophicaldisquisitions.blogspot.ie/2014/01/rule-by-algorithm-big-data-and-threat.html (accessed 10/5/22).
9. Danaher, J. (2016). *The Logical Space of Algocracy (Redux).* Philosophical Disquisitions. Available: http://philosophicaldisquisitions.blogspot.com/2016/11/the-logical-space-of-algocracy-redux.html (accessed 10/5/22).
10. Danaher, J. (2016). *The Threat of Algocracy: Reality, Resistance and Accomodation*. Springer Link. Available: https://doi.org/10.1007/s13347-015-0211-1 (accessed 10/5/22).
11. Danaher, J. (2018). *Freedom in the age of Algocracy.* Academia. Available: https://www.academia.edu/41763719/Freedom_in_an_Age_of_Algocracy (accessed 10/5/22).
12. Davidow, B. (2014), *'Welcome to Algorithmic Prison-The Use of Big Data to Profile Citizens is Subtly, Silently Constraining Freedom'*. The Atlantic. Available: https://www.theatlantic.com/technology/archive/2014/02/welcome-to-algorithmic-prison/283985 (accessed 6/5/22).
13. European Commission (2020). AI Watch. Defining Artificial Intelligence. Available: https://publications.jrc.ec.europa.eu/repository/handle/JRC118163 (accessed 6/5/22)
14. European Commission (2021). *Humans and Societies in the age of Artificial Intelligence.* Available: https://op.europa.eu/en/publication-detail/-

/publication/a72ac1a9-98e2-11eb-b85c-01aa75ed71a1/language-en (accessed 6 May 2022).
15. European Parliamentary Research Service (2020). *Artificial intelligence: How does it work, why does it matter, and what can we do about it?*. Available: https://www.europarl.europa.eu/RegData/etudes/STUD/2020/641547/EPRS_STU%282020%29641547_EN.pdf (accessed 6 May 2022).
16. Greenfield, R. (2012). Inside the Method to Amazon's Beautiful Warehouse Madness. *The Wire*. Available: http://www.thewire.com/technology/2012/12/inside-method-amazons-beautiful-warehouse-madness/59563/ (accessed 10/5/22).
17. High-Level Expert Group on AI (2019). *Ethics Guidelines for Trustworthy Artificial Intelligence*. Available: https://op.europa.eu/en/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1 (accessed 6 May 2022).
18. High-Level Expert Group on AI (2020). *Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment*. Available: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment (accessed 6 May 2022).
19. Jack, I. (2018). 'I Feared the Grenfell Tributes Would Be Mawkish. I Was Wrong'.The Guardian. Available: https://www.theguardian.com/commentisfree/2018/may/26/grenfelltower-tributes-titanic-victims (accessed 6/5/22).
20. Kitchin, R. (2014b). *Thinking critically about researching algorithms*. The Programmable City Working Paper 5. Available: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2515786 (accessed 10/5/22).
21. Knightarchive, W. (2017). *The Dark Sectet at the Heart of AI*. MIT Technology Review. Available: https://www.technologyreview.com/2017/04/11/5113/the-dark-secret-at-the-heart-of-ai/ (accessed 10/5/22).
22. Misuraca, G., van Noordt, C. (2020). *AI Watch—Artificial Intelligence in public services* (JRC EUR 30255). Publications Office of the European Union. https://op.europa.eu/en/publication-detail/-/publication/4c72dd88-bcda-11ea-811c-01aa75ed71a1/language-en (accessed 12/7/22).
23. Molinari, F., van Noordt, C., Vaccari, L., Pignatelli, F., & Tangi, L. (2021). *AI Watch-Beyond pilots: Sustainable implementation of AI in public services*. EUR 30868 EN, Publications Office of the European Union, Luxembourg, 2021. https://publications.jrc.ec.europa.eu/repository/handle/JRC126665 (accessed 12/7/22).
24. Morozov, E. (2013). *The Real Privacy Problem*. MIT Technology Review. Available: http://www.technologyreview.com/featuredstory/520426/the-real-privacy-problem/ (accessed 10/5/22).
25. Peter, F. (2014). *Political Legitimacy*. In Edward N. Zalta (ed) The Stanford Encyclopedia of Philosophy Spring 2014 Edition. Available: http://plato.stanford.edu/archives/spr2014/entries/legitimacy/
26. Santiso, C. (2022). Artificial Intelligence in the public sector: *An engine for innovation in government…if we get it right*. OECD Observatory of Public Sector Innovation (OPSI). Available: https://oecd-opsi.org/blog/ai-an-engine-for-innovation/ (accessed 12/7/22).
27. Tangi, L., van Noordt, C., Combetto, M., Gattwinkel, D., Pignatelli, F. (2022). *AI Watch-European Landscape on the use of Artificial Intelligence by the Public Sector.* EUR 31088 EN. Publications Office of the European Union, Luxembourg, 2022. Available: https://publications.jrc.ec.europa.eu/repository/handle/JRC129301 (accessed 12/7/22).

28. Unesco (2021). *Recommendation on the Ethics of Artificial Intelligence*. Available: https://unesdoc.unesco.org/ark:/48223/pf0000380455 (accessed 6/5/22).
29. Williams, J. (2018), *'Technology is Driving Us to Distraction'*. The Observer. Available: https://www.theguardian.com/commentisfree/2018/may/27/world-distraction-demands-new-focus (accessed 6/5/22).
30. Wirtz, B. Weyerer, J. Geyer, C. (2018). *Artificial intelligence and Public Sector-Applications and Challenges*. International Journal of Public Administration. Available:file:///C:/Users/roula/Downloads/2018_Artificial_Intelligence_and_the_Public_Sector_Applications_and_Challenges_-_ohne_Angaben%20(1).pdf (accessed 12/7/22).
31. World Bank (2020). *Artificial Intelligence in the Public Sector: Maximizing Opportunities, Managing Risks*. Equitable Growth, Finance and Institutions Insight;. World Bank, Washington, DC. © World Bank. Available: https://openknowledge.worldbank.org/handle/10986/35317 (accessed 12/7/22).