



ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΟΙΚΟΝΟΜΙΑΣ & ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ

ΤΜΗΜΑ ΔΗΜΟΣΙΑΣ ΔΙΟΙΚΗΣΗΣ

Η εφαρμογή της τεχνητής νοημοσύνης στη δημόσια διοίκηση

ΔΙΔΑΚΤΟΡΙΚΗ ΔΙΑΤΡΙΒΗ

Νικόλαος Μπούκας

Αθήνα, 2023

Τριμελής Συμβουλευτική Επιτροπή

Επιβλέπουσα: Επίκουρη Καθηγήτρια Φερενίκη Παναγοπούλου

Μέλη:

Ομότιμη Καθηγήτρια Ισμήνη Κριάρη

Καθηγήτρια Πηνελόπη Φουντεδάκη

Copyright © Νικόλαος Μπούκας, 2023

All rights reserved. Με επιφύλαξη παντός δικαιώματος.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας διδακτορικής διατριβής εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της διδακτορικής διατριβής για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Η έγκριση της διδακτορικής διατριβής από το Πάντειον Πανεπιστήμιο Κοινωνικών και Πολιτικών Επιστημών δεν δηλώνει αποδοχή των γνώμων του συγγραφέα.

Ευχαριστίες

Ολόψυχα θέλω να εκφράσω το ευχαριστώ και την ευγνωμοσύνη μου, πρωτίστως, στην επιβλέπουσα της διδακτορικής μου διατριβής, Επίκουρη Καθηγήτρια του Τμήματος Δημόσιας Διοίκησης του Παντείου Πανεπιστημίου, κ. Φερενίκη Παναγοπούλου-Κουτνατζή, η οποία μου έκανε την απροσδόκητη τιμή να με επιλέξει ως υποψήφιο διδάκτορά της. Ο δρόμος της εκπόνησης της παρούσας μελέτης ήταν ανηφορικός, αλλά οι καίριες συμβουλές, η στοχευμένη καθοδήγηση, η ενθάρρυνση για κάθε επόμενο βήμα της συγγραφής, υπό το φως της άρτιας κατάρτισης, αστείρευτης γνώσης και καθησυχαστικής παρότρυνσης προς την κατάκτηση του ποθούμενου στόχου, κατέστησαν την πορεία αυτή σε μια ομαλή ανάβαση.

Επίσης, από καρδιάς θέλω να εκφράσω την ευγνωμοσύνη μου στην τ. Πρύτανη του Παντείου Πανεπιστημίου, Καθηγήτρια κ. Ισμήνη Κριάρη και στην Αντιπρύτανη του Παντείου Πανεπιστημίου, Καθηγήτρια κ. Πηνελόπη Φουντεδάκη, οι οποίες αποδέχθηκαν να συνθέσουν την Τριμελή μου Επιτροπή και χάρη του «ανεξίτηλου» Ακαδημαϊκού τους αποτυπώματός προσέδωσαν ιδιαίτερη αίγλη στην προσπάθειά μου.

Επιπλέον, θέλω να ευχαριστήσω τον εξάίρετο Αναπληρωτή Καθηγητή και Πρόεδρο του Τμήματος Δημόσιας Διοίκησης του Παντείου Πανεπιστημίου κ. Αντώνιο Χάνο, τον οποίο είχα την τύχη και χαρά να τον ζήσω ως Καθηγητή στο Μ.Π.Σ. Δίκαιο, Τεχνολογία και Οικονομία και ο οποίος αποδέχθηκε να είναι μέλος της Επταμελούς Επιτροπής κατά την υποστήριξη της παρούσας διατριβής.

κλείνοντας, θέλω να ευχαριστήσω και τα υπόλοιπα μέλη της Επταμελούς Επιτροπής, καθώς και την οικογένειά μου και τους εκλεκτούς μου φίλους, οι οποίοι με στήριξαν στην προσπάθειά μου κατά τη διάρκεια της εκπόνησης της παρούσας διατριβής.

Περιεχόμενα

Περίληψη	7
Abstract	8
Εισαγωγή	9
Αντικείμενο – Σκοπός.....	9
Η Επικέντρωση του Ερωτήματος	11
Ακολουθούμενη Μεθοδολογία	12
Σύντομη ανασκόπηση κεφαλαίων.....	14
Κεφάλαιο 1 Ορισμοί.....	16
1.1 Η Τεχνητή Νοημοσύνη	16
1.1.1 Η εξέλιξη της τεχνητής νοημοσύνης	16
1.1.2 Εννοιολογία.....	21
1.2 Ασφάλεια και «βρώμικα» δεδομένα	29
1.3 Μηχανική Μάθηση	31
1.4 «Έξυπνα-Εξημερωμένα» Αντικείμενα.....	33
1.5 Η Κατανόηση της Τεχνολογίας.....	34
1.6 Σύνοψη – Συμπεράσματα	36
Κεφάλαιο 2 Νομοθετικά Ζητήματα	38
2.1 Μορφές δικαιοϋκτών ρυθμίσεων της τεχνητής νοημοσύνης	38
2.2 Η Εφαρμογή των ‘Νορμών’ και τα Νομικά Ενδεχόμενα επί ζητημάτων Ηθικής	46
2.3 Σύνοψη – Συμπεράσματα	66
Κεφάλαιο 3 Ζητήματα Ηθικής.....	71
3.1 Η Ουδετερότητα στις Νέες Τεχνολογίες	71
3.2 Ο Διαχωρισμός του Ανθρώπινου και του Μη Ανθρώπινου Χαρακτήρα	74
3.3 Η Δημιουργία και Αξιοποίηση Απρόβλεπτων Μηχανών	76

3.4 Ο Απρόβλεπτος Χαρακτήρας των Μηχανών και η Απόδοση Ευθυνών	81
3.5 Ο Σχεδιασμός Ηθικού Πλαισίου για το Δίκαιο της Τεχνητής Νοημοσύνης	83
3.6 Ηθικά Ζητήματα Αυτονομίας των Μηχανών	89
3.7 Η Εφαρμογή της Τεχνητής Νοημοσύνης και τα Θεμελιώδη Δικαιώματα.....	96
3.8 Οι κοινωνικές προκλήσεις της τεχνητής νοημοσύνης.....	100
3.9 Εθνικές προσπάθειες τεχνολογικής προόδου	107
3.10 Η κατανόηση της τεχνολογίας	114
3.11 Σύνοψη – Συμπεράσματα	119
Κεφάλαιο 4 Δημόσια Διοίκηση	121
4.1 Τα Πεδία Εφαρμογής της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα.....	121
4.2 Οι Προκλήσεις του Δημοσίου Τομέα για μία Πολυδιάστατη Ενσωμάτωση της Τεχνητής Νοημοσύνης	130
4.2.1 Πρώτη Διάσταση: Το Κανονιστικό Πλαίσιο στην Ενσωμάτωση της Τεχνητής Νοημοσύνης στον Έλεγχο και τη Διακυβέρνηση.....	133
4.2.2 Δεύτερη Διάσταση: Κοινωνικά Ζητήματα της Πολύπλευρης Ενσωμάτωσης της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα	139
4.2.3 Τρίτη Διάσταση: Ηθικά Ζητήματα της Ενσωμάτωσης της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα	146
4.2.4 Τέταρτη Διάσταση: Τεχνολογικά Ζητήματα της Ενσωμάτωσης της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα	152
4.3 Η Τεχνητή Νοημοσύνη και οι Διατομεακές Συνεργασίες στον Δημόσιο Τομέα	157
4.3.1 Η Χαρτογράφηση των Διατομεακών Συνεργασιών στον Δημόσιο Τομέα μέσω της Αξιοποίησης της Τεχνητής Νοημοσύνης	157
4.3.2 Οι Προκλήσεις για Επιτυχείς Διατομεακές Συνεργασίες	162
4.3.3 Οι Προσδιοριστικοί Παράγοντες της Επιτυχίας και οι Ευκαιρίες των Επιτυχών Διατομεακών Συνεργασιών	168
4.4 Η Τεχνητή Νοημοσύνη ως Εργαλείο στη Διαχείριση του Ανθρωπίνου Δυναμικού.....	177
4.4.1 Εισαγωγικές Παρατηρήσεις	177
4.4.2 Ο Κύκλος Ζωής της Ενσωμάτωσης της Τεχνητής Νοημοσύνης στη Διαχείριση του Ανθρωπίνου Δυναμικού	184
4.4.3 Οι Προκλήσεις στην Επιτυχή Αξιοποίηση της Τεχνητής Νοημοσύνης στη Διαχείριση του Ανθρωπίνου Δυναμικού	190
4.4.4 Οι σύγχρονες παρεχόμενες υπηρεσίες προς τους πολίτες.....	213
4.4.5 Η συμβολή της τεχνητής νοημοσύνης στην παροχή των εν λόγω υπηρεσιών.....	217

4.4.6 Οι σύγχρονες ανάγκες των κυβερνήσεων για την τεχνολογική τους ανέλιξη	228
4.5 Σύνοψη - Συμπεράσματα.....	231
Κεφάλαιο 5 Συμπεράσματα και προτάσεις	237
Βιβλιογραφία	246

Πίνακες

Πίνακας 1. Από την UNESCO.....	63
Πίνακας 2. Από τη μελέτη του Chagal-Feferkorn	90
Πίνακας 3	185
Πίνακας 4. Περιπτώσεις δημόσιων υπηρεσιών που χρησιμοποιούν εφαρμογές τεχνητής νοημοσύνης για την εξυπηρέτηση των πολιτών	219

Διαγράμματα

Διάγραμμα 1. Κύκλος ζωής τεχνητής νοημοσύνης	184
--	-----

Περίληψη

Σκοπός της εν λόγω διατριβής είναι η διερεύνηση των διαφορετικών πτυχών της ενσωμάτωσης της τεχνητής νοημοσύνης στη δημόσια διοίκηση, καθώς επίσης και η διενέργεια μίας προσπάθειας διαλεύκανσης των ηθικών διλημάτων που συνεπάγεται η ενσωμάτωση αυτή. Η ενσωμάτωση και αξιοποίηση της τεχνητής νοημοσύνης στο σύνολο των πτυχών της δημόσιας διοίκησης δεν αρκεί να πραγματοποιηθεί μόνο επειδή αυτή αποτελεί ένα καινοτόμο τεχνολογικό σύστημα. Ακόμη και αν επέλθει πλήρης εξάλειψη των ηθικών κινδύνων που απορρέουν από την αξιοποίηση συστημάτων τεχνητής νοημοσύνης, δεν μπορεί να καταστεί εφικτή αν δεν εφαρμοστεί κάτω από ειδικά σχεδιασμένες στρατηγικές που θέτουν ως επίκεντρο την εξυπηρέτηση των πολιτών. Άρα, η ερώτηση που χρήζει απάντησης δεν πρέπει να αφορά στο με ποιον τρόπο μπορεί να χρησιμοποιηθεί η τεχνητή νοημοσύνη για την επίλυση των προβλημάτων των δημόσιων υπηρεσιών, αλλά στο ποια προβλήματα είναι αυτά που πρέπει να επιλυθούν μέσω της χρήσης της τεχνητής νοημοσύνης, το πώς μπορούν να επιλυθούν και για ποιον λόγο. Γενικότερα, θεωρούμε ότι την αποκλειστική ευθύνη για τα αποτελέσματα των ενεργειών και αποφάσεων των συστημάτων τεχνητής νοημοσύνης πρέπει να την έχει πάντα ο άνθρωπος, ενώ τα συστήματα αυτά δεν θα πρέπει να χρησιμοποιούνται όταν υπάρχει η πιθανότητα να καθοδηγήσουν, προς συγκεκριμένη κατεύθυνση, το ανθρώπινο υποσυνείδητο ή την ανθρώπινη συμπεριφορά, ώστε να καταλήξει να προκληθεί βλάβη προς τον ίδιο ή άλλον άνθρωπο. Σε κάθε περίπτωση, η από κοινού δημιουργία ενός κανονιστικού πλαισίου ηθικής και δεοντολογίας για την χρήση των συστημάτων τεχνητής νοημοσύνης, από τους κυβερνητικούς φορείς και από τους ίδιους τους πολίτες, κρίνεται αμοτεροβαρώς επικερδής, ωφέλιμη και αναγκαία

Λέξεις-Κλειδιά: Δημόσια Διοίκηση, Τεχνητή Νοημοσύνη, Ηθικά Ζητήματα, Συστήματα Τεχνητής Νοημοσύνης.

Abstract

The implementation of Artificial Intelligence in Public Administration

The purpose of this thesis is to investigate the different aspects of the integration of artificial intelligence in public administration, as well as to try to clarify the ethical dilemmas that this integration entails. The integration and utilization of artificial intelligence in all aspects of public administration is not enough to take place just because it is an innovative technological system. Even if the ethical risks arising from the use of artificial intelligence are completely eliminated, it cannot be achieved unless it is implemented under specially designed strategies that focus on serving citizens. So, the question that needs to be answered should not be how artificial intelligence can be used to solve public service problems, but what problems are to be solved through the use of artificial intelligence, how can be resolved and why. Generally, we believe that the responsibility for the results of the actions and decisions of artificial intelligence systems should always be held by humans, while these systems should not be used when there is a possibility that they will guide, in a specific direction, the human subconscious or human behavior, so that it ends up causing harm to the same or another person. In any case, the joint creation of a regulatory framework of ethics for the use of artificial intelligence systems, by government organizations and by the citizens, is considered mutually beneficial and necessary.

Keywords: Public Administration, Artificial Intelligence, Ethical Issues, Artificial Intelligence Systems.

Εισαγωγή

Αντικείμενο – Σκοπός

Βασικό αντικείμενο της παρούσας διατριβής αποτελεί η διερεύνηση των πολυδιάστατων πτυχών της ενσωμάτωσης και εφαρμογής της τεχνητής νοημοσύνης σε επιμέρους διαφορετικά πεδία της δημόσιας διοίκησης. Όπως θα φανεί διεξοδικά στην ανάλυση των κεφαλαίων της εν λόγω μελέτης, η τεχνητή νοημοσύνη αποτελεί ένα νεόκοπο πεδίο, το οποίο έχει ήδη αποτελέσει αντικείμενο πληθώρας μελετών και ερευνών επί της σύγχρονης βιβλιογραφίας, για την ενσωμάτωση και αξιοποίησή της τόσο στον δημόσιο όσο και στον ιδιωτικό τομέα. Μάλιστα, η εν λόγω επιτυχής ενσωμάτωση δεν θα ήταν υπερβολικό να χαρακτηρισθεί ως μία πολλά υποσχόμενη διαδικασία για τη βελτίωση των εσωτερικών λειτουργικών διεργασιών και τη λήψη αποφάσεων, τόσο στο πλαίσιο των επιχειρηματικών οντοτήτων όσο και στο αντίστοιχο των δημοσίων οργανισμών. Ο λόγος έγκειται στο ότι μέσω της ανωτέρω αναφερόμενης διαδικασίας επιτυχούς ενσωμάτωσης της τεχνητής νοημοσύνης αυτοματοποιούνται διαδικασίες, οι οποίες μέχρι πρότινος απαιτούσαν τη συμβολή του ανθρώπινου παράγοντα. Ως εκ τούτου, μέσω αυτής, δύνανται να αποφευχθούν υποκειμενικές και πιθανότατα λανθασμένες ανθρώπινες κρίσεις, οι οποίες ενδεχομένως οδηγούσαν σε υπονόμηση της εύρυθμης λειτουργίας και δράσης του εκάστοτε άμεσα εμπλεκόμενου, ιδιωτικού ή δημοσίου φορέα.

Όμως, η ενσωμάτωση και η αξιοποίηση της τεχνητής νοημοσύνης στους σύγχρονους οργανισμούς δεν είναι τόσο μονοδιάστατη, καθώς εσωκλείει και κινδύνους, οι οποίοι πρέπει να προβλεφθούν και να αντιμετωπιστούν μέσω της υιοθέτησης των κατάλληλων δικλίδων ασφαλείας και ελέγχου των κινήσεων και αποφάσεων, οι οποίες λαμβάνονται αυτοματοποιημένα από την πλευρά των καινοτόμων μηχανημάτων που υπάγονται στη σφαίρα του πεδίου της τεχνητής νοημοσύνης.

Επί της εν λόγω διατριβής δεν αποσκοπείται μόνο η απλή ανασκόπηση των πτυχών της δημόσιας διοίκησης, στις οποίες δύναται να εφαρμοστεί και να αξιοποιηθεί η τεχνητή νοημοσύνη. Το εν λόγω ζήτημα αποτελεί τον γενικό στόχο αυτής. Επί της ουσίας, ο βασικός σκοπός της παρούσας μελέτης είναι η ανάλυση των πτυχών εφαρμογής της τεχνητής νοημοσύνης επί του τομέα της δημόσιας διοίκησης. Εξ αυτού, όμως,

προκύπτουν επιμέρους ειδικοί στόχοι, οι οποίοι επιδιώκεται να καλυφθούν. Πιο συγκεκριμένα, επιδιώκεται η αποσαφήνιση των ωφελειών και των μεταρρυθμίσεων, που επέρχονται ως απόρροιας μίας επιτυχούς ενσωμάτωσης της τεχνητής νοημοσύνης σε εκάστοτε αναλυόμενη πτυχή του δημοσίου τομέα, η παράθεση και ανάλυση των προκλήσεων, τις οποίες καλείται να αντιμετωπίσει η κάθε άμεσα εμπλεκόμενη δημόσια υπηρεσία ή ο αρμόδιος φορέας, προκειμένου να επιτευχθεί η ενσωμάτωση και βέλτιστη αξιοποίηση των νέων τεχνολογιών τεχνητής νοημοσύνης, καθώς επίσης και η περιγραφή των ηθικών κυρίως κινδύνων που συνεπάγεται η εν λόγω αξιοποίηση, αν αυτή διενεργηθεί δίχως την προγενέστερη λήψη των απαιτούμενων μέτρων για την μη παραβίαση των συνταγματικών και ανθρωπίνων δικαιωμάτων.

Η Επικέντρωση του Ερωτήματος

Όπως ήδη αναφέρθηκε, οι μελέτες επί του ζητήματος της ενσωμάτωσης της τεχνητής νοημοσύνης στον ιδιωτικό και δημόσιο τομέα είναι αρκετές¹, καθώς η τεχνητή νοημοσύνη έχει προσεγγίσει το ερευνητικό ενδιαφέρον τόσο των Ελλήνων όσο και των αλλοδαπών επιστημόνων. Στην παρούσα μελέτη, τα ερωτήματα που τίθενται για την τεχνητή νοημοσύνη επιχειρείται να απαντηθούν με αξιοποίηση ερευνών και με επικέντρωση σε συγκεκριμένες πτυχές της δημόσιας διοίκησης, όπως για παράδειγμα, στη σύναψη διατομεακών συνεργασιών, στη διαχείριση του ανθρώπινου δυναμικού, καθώς επίσης και στην εξυπηρέτηση των πολιτών. Το ζήτημα της ενσωμάτωσης της τεχνητής νοημοσύνης προσεγγίζεται πολυδιάστατα σε εκφάνσεις της δημόσιας διοίκησης, ενώ συν τοις άλλοις, τίθενται υπόψη και σοβαρά ανακύπτοντα ζητήματα ηθικής και ενδεχόμενης παραβίασης συνταγματικών δικαιωμάτων, που ενδέχεται να προκληθούν από τη συγκεκριμένη διαδικασία ενσωμάτωσης.

Επομένως, κατόπιν της ολοκλήρωσης της διατριβής, οι δημόσιες αρχές και οι επίσημοι κυβερνητικοί φορείς πιθανότατα θα είναι σε θέση να σχηματίσουν πλήρη και ολοκληρωμένη άποψη για τις απόρροιες, τις συνθήκες και τα οφέλη, καθώς επίσης και ποιες προκλήσεις θα κληθούν να αντιμετωπίσουν, προκειμένου να επιτύχουν τη βέλτιστη ενσωμάτωση της τεχνητής νοημοσύνης στο πεδίο της καθημερινής τους δράσης και λειτουργίας. Τέλος, θεωρείται πως θα δύνανται να λάβουν γνώση όλων των κινδύνων που ελλοχεύει μία τέτοιου είδους ενσωμάτωση, κυρίως επί ζητημάτων ηθικής και παραβίασης των ανθρωπίνων δικαιωμάτων.

¹ Βλ. L. Sanzogni, G. Guzman, & P. Busch, Artificial intelligence and knowledge management: questioning the tacit dimension, *Prometheus*, Αρ. 35, Τεύχος 1, 2017, σελ. 37–56, Βλ. επίσης A. Bleicher, Demystifying the black box that is AI: Humans are increasingly entrusting our security, health and safety to Black Box intelligent machines, *Scientific American*, 2017, στο <https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/>, Βλ. επίσης W. Knight, The dark secret at the heart of AI, *MIT Technology Review*, Αρ. 120, Τεύχος 3, 2017, σελ. 54–61, Βλ. επίσης H. Lu, Y. Li, M. Chen, H. Kim, & S. Serikawa, Brain Intelligence: Go beyond Artificial Intelligence, *Mobile Networks and Applications*, Αρ. 23, Τεύχος 2, 2017, σελ. 368–375, Βλ. επίσης G.A. Hodge, & C. Greve, Public-private partnerships: an international performance review, *Public Administration Review*, Αρ. 67, 2007, σελ. 545–558.

Ακολουθούμενη Μεθοδολογία

Η διαδικασία συγγραφής της παρούσας διατριβής εκκινεί από τη συλλογή και μελέτη της σχετικής βιβλιογραφίας. Το ιδιαίτερο χαρακτηριστικό του υλικού, που αφορά στο πραγματευόμενο θέμα, είναι το πρόσφατο των σχετικών μελετών. Πράγματι, η σχετική βιβλιογραφία εντοπίζεται κυρίως στην τελευταία εικοσαετία.

Η μεθοδολογία που ακολουθήθηκε βασίστηκε σε τέσσερις επιμέρους άξονες, οι οποίοι είναι οι κάτωθι:

(1) Ο πρώτος άξονας έγκειται στην βαθύτερη κατά το δυνατόν μελέτη και αφομοίωση των υφιστάμενων κανόνων και αρχών που διέπουν ζητήματα άμεσα συνυφασμένα με την τεχνητή νοημοσύνη και την υιοθέτησή της σε διάφορες πτυχές του δημοσίου τομέα.

(2) Ο δεύτερος άξονας έγκειται στη συλλογή όλων των απαιτούμενων επιστημονικών μελετών, οι οποίες έχουν ολοκληρωθεί και είναι άμεσα συναρτημένες και επικεντρωμένες στο υπό ανάλυση ζήτημα της παρούσας μελέτης. Παράλληλα, στον ίδιο άξονα, διενεργήθηκε και η συλλογή και μελέτη σχετικών με το συγκεκριμένο ζήτημα, επιστημονικών συγγραμμάτων. Τόσο οι επιστημονικές μελέτες όσο και τα επιστημονικά συγγράμματα διαδραμάτισαν καθοριστικό ρόλο στην πληρέστερη ερμηνεία, κατανόηση και ανάλυση των υφιστάμενων κανόνων και αρχών, που διέπουν ζητήματα συνδεδεμένα με την ενσωμάτωση της τεχνητής νοημοσύνης στον δημόσιο τομέα.

(3) Ο τρίτος άξονας έγκειται στην κατηγοριοποίηση των αντλούμενων πληροφοριακών δεδομένων από το ανωτέρω αναφερόμενο συλλεγόμενο βιβλιογραφικό και επιστημονικό υλικό, προκειμένου μέσω αυτής, να διαμορφωθεί η δομή της εν λόγω διατριβής, υπό την έννοια του επιμερισμού της σε κεφάλαια και υποκεφάλαια. Η κατασκευή της δομής της εν λόγω διατριβής, έγινε με γνώμονα την διασφάλιση και προάσπιση μιας ολιστικής προσέγγισης του πραγματευόμενου ζητήματος και την απόδοση μίας αξιόπιστης εικόνας της προβληματικής, η οποία είναι άμεσα συναρτημένη με το ζήτημα της πολυδιάστατης ενσωμάτωσης της τεχνητής νοημοσύνης στον δημόσιο τομέα.

(4) Ο τέταρτος και τελευταίος άξονας έγκειται στη συγγραφή των επιμέρους κεφαλαίων και υποκεφαλαίων της συγκεκριμένης διατριβής και στην κριτική

ανασκόπηση των αποτυπωμένων πληροφοριακών στοιχείων, καθώς επίσης και στη διαμόρφωση και παράθεση των τελικώς εξαγόμενων συμπερασμάτων.

Σύντομη ανασκόπηση κεφαλαίων

Στο πρώτο κεφάλαιο της παρούσας διατριβής γίνεται αναφορά στους ορισμούς βασικών εννοιών, που συνάδουν με το προς ανάλυση ζήτημα αυτής. Αρχικά, προσεγγίζεται εννοιολογικά το πεδίο της τεχνητής νοημοσύνης, κατόπιν ανάλυσης της εξέλιξης αυτής. Στη συνέχεια, αποσαφηνίζεται εννοιολογικά το ζήτημα της ασφάλειας και της ύπαρξης «βρώμικων»² δεδομένων στον σχεδιασμό συστημάτων τεχνητής νοημοσύνης. Έπειτα, δίνεται ο ορισμός της μηχανικής μάθησης και των έξυπνων – εξημερωμένων αντικειμένων. Ακολουθεί και η εννοιολογική αποσαφήνιση της κατανόησης της τεχνολογίας.

Στο δεύτερο κεφάλαιο παρουσιάζονται βασικά νομοθετικά ζητήματα τα οποία αφορούν στην ενσωμάτωση και την χρήση μηχανών τεχνητής νοημοσύνης στη σύγχρονη κοινωνία και τις επιμέρους πτυχές της. Αρχικά, παρουσιάζονται οι μορφές των δικαιικών ρυθμίσεων της τεχνητής νοημοσύνης, ενώ στη συνέχεια περιγράφονται ζητήματα εφαρμογής των νομών και των νομικών ενδεχόμενων επί ηθικών ζητημάτων. Στο σημείο αυτό, διευκρινίζεται ότι λεπτομερέστερη ανάλυση επί της ηθικής διάστασης της χρήσης της τεχνητής νοημοσύνης αναπτύσσεται στο τρίτο κεφάλαιο της παρούσας διατριβής. Ωστόσο, ζητήματα που αφορούν την εν λόγω ηθική διάσταση και τα οποία παρουσιάζονται στο συγκεκριμένο κεφάλαιο, είναι αυτά που έχουν άμεση συσχέτιση με το υφιστάμενο κανονιστικό και ρυθμιστικό πλαίσιο, καθώς επίσης και με προτάσεις που προκύπτουν από την πλευρά της Ευρωπαϊκής Επιτροπής.

Στο τρίτο κεφάλαιο αναλύονται τα ζητήματα ηθικής που προκύπτουν από την αξιοποίηση των συστημάτων τεχνητής νοημοσύνης. Αρχικά, περιγράφεται η ουδετερότητα των νέων τεχνολογιών, καθώς επίσης και ο διαχωρισμός του ανθρώπινου και μη ανθρώπινου χαρακτήρα. Στη συνέχεια, δίνεται έμφαση στη δημιουργία και την αξιοποίηση απρόβλεπτων μηχανών τεχνητής νοημοσύνης και κατόπιν τούτου, γίνεται εμβάθυνση στην απόδοση ευθυνών, δεδομένου του απρόβλεπτου χαρακτήρα που προαναφέρθηκε. Ακολούθως, γίνεται μνεία σε ζητήματα σχεδιασμού ενός ηθικού

² Όταν ομιλούμε περί «βρώμικων» δεδομένων αναφερόμαστε σε δεδομένα τα οποία προέρχονται από μη αξιόπιστες και μη έγκυρες πηγές. Επομένως, πρόκειται για δεδομένα, των οποίων η γνησιότητα, αξιοπιστία και εγκυρότητα αμφισβητούνται. Περισσότερες λεπτομέρειες βρίσκονται στην υποενότητα 1.2 της παρούσας διατριβής.

πλαίσιου γύρω από το πεδίο του δικαίου της τεχνητής νοημοσύνης και της χρήσης αυτής. Συνεπώς, περιγράφονται πολυδιάστατα ηθικά ζητήματα που είναι συνυφασμένα με τον βαθμό στον οποίο οι μηχανές τεχνητής νοημοσύνης εμφανίζουν αυτονομία. Συνεχίζοντας, αναφέρονται ζητήματα εφαρμογής της τεχνητής νοημοσύνης σε αντιστοιχία με τα υφιστάμενα θεμελιώδη δικαιώματα. Συνακόλουθα, προκύπτει η ανάγκη της περαιτέρω ανάλυσης του ζητήματος της κοινωνικής πρόκλησης που μπορεί να αναδειχθεί από τη χρήση των εν λόγω τεχνολογικών συστημάτων. Ακολουθεί η περιγραφή των καταβαλλόμενων, σε εθνικό επίπεδο, προσπαθειών για την τεχνολογική πρόοδο και την κατανόηση της τεχνολογίας.

Στο τέταρτο κεφάλαιο αναπτύσσεται η ενσωμάτωση και χρήση της τεχνητής νοημοσύνης σε επίπεδο δημόσιας διοίκησης. Αρχικά, επισημαίνονται τα πεδία, στα οποία μπορεί να τύχει εφαρμογής η τεχνητή νοημοσύνη στον δημόσιο τομέα, καθώς επίσης και οι προκλήσεις που δημιουργούνται από τη συνθήκη αυτή. Σε αυτό το σημείο παρουσιάζονται οι τέσσερις διαστάσεις χρήσης της τεχνητής νοημοσύνης στη δημόσια διοίκηση. Κατόπιν, περιγράφονται ζητήματα αξιοποίησης μηχανών τεχνητής νοημοσύνης σε περιπτώσεις διατομεακών συνεργασιών στο πλαίσιο του δημόσιου τομέα. Στη συνέχεια, αναφέρεται το ζήτημα της χρήσης της τεχνητής νοημοσύνης στον τομέα του ανθρώπινου δυναμικού των δημοσίων υπηρεσιών. Ακολουθώντας, λαμβάνοντας υπόψη ότι απώτερος σκοπός των δημοσίων υπηρεσιών είναι η κοινωνική και πολιτική ευρυθμία και η κάλυψη των αναγκών των πολιτών, ένα μέρος του κεφαλαίου αυτού αποτελεί η περιγραφή της χρήσης συστημάτων τεχνητής νοημοσύνης στην εξυπηρέτηση των πολιτών της σύγχρονης κοινωνίας. Κλείνοντας, παρουσιάζονται οι ανάγκες που πρέπει να ικανοποιηθούν από την πλευρά των κυβερνήσεων προκειμένου να καταστεί δυνατή η παροχή των ανωτέρω αναφερόμενων υπηρεσιών, υπό την αξιοποίηση μηχανών τεχνητής νοημοσύνης, στην εξυπηρέτηση των πολιτών.

Στο πέμπτο και τελευταίο κεφάλαιο της παρούσας διατριβής παρουσιάζονται τα τελικά συμπεράσματα αυτής, καθώς επίσης και προτάσεις που προκύπτουν από την κριτική ανάλυση και αξιολόγηση των δεδομένων που περιγράφηκαν στις προηγούμενες ενότητες.

Κεφάλαιο 1 Ορισμοί

1.1 Η Τεχνητή Νοημοσύνη

1.1.1 Η εξέλιξη της τεχνητής νοημοσύνης

Η ιστορία της Τεχνητής Νοημοσύνης είναι αρκετά ενδιαφέρουσα και ξεκίνησε πριν από περίπου εκατό χρόνια. Το 1920 δημοσιεύτηκε το έργο επιστημονικής φαντασίας με τίτλο «Rossumovi Univerzální Roboti» από τον Τσέχο συγγραφέα Karel Čapek, στο οποίο γίνεται για πρώτη φορά αναφορά στον όρο ρομπότ. Τα ρομπότ παρουσιάζονται ως ζωντανά πλάσματα, τα οποία μοιάζουν περισσότερο με τον όρο των κλώνων³. Εμφανίστηκαν ως δημιουργήματα του ανθρώπου, τα οποία αρχικά εργάζονταν για να διευκολύνουν και να εξυπηρετούν τον άνθρωπο, αλλά στη συνέχεια ακολούθησε η εξέγερση των ρομπότ, η οποία οδήγησε στην εξαφάνιση του ανθρώπινου είδους⁴.

Τα δύο σημαντικά στοιχεία, στα οποία γίνεται λόγος στο συγκεκριμένο έργο, αφενός αποτυπώνονται στον πρωτοεμφανιζόμενο όρο των ρομπότ που, αν και δεν προσομοιάζει με την εικόνα και χρήση τους κατά τη σύγχρονη εποχή, όμως είναι παρεμφερής με την έννοια των κλώνων και αφετέρου ότι εμβαθύνει στην εξιστόρηση των συνθηκών, κάτω από τις οποίες δημιουργήθηκαν τα ρομπότ, τα οποία αποτελούν μία μορφή τεχνητής νοημοσύνης που αρχικά συμβάλλει στη βελτίωση της ποιότητας διαβίωσης των ανθρώπων, αλλά αργότερα ακολουθεί η απρόσμενη και μη αναμενόμενη εξέγερση των ρομπότ, που απειλεί την ανθρωπότητα⁵.

Κομβικό ρόλο, σε αυτήν την ιστορική αναδρομή, διαδραμάτισε και ο Alan Turing, ο οποίος γεννήθηκε στις 23 Ιουνίου 1912 στο Λονδίνο. Είναι ιδιαίτερα δημοφιλής από την αποκρυπτογράφηση του κώδικα του «αινίγματος» (enigma). Πρόκειται για τον κώδικα που η ναζιστική Γερμανία χρησιμοποίησε για την πραγματοποίηση μυστικών

³Βλ. K. Capek, R.U.R. (Rossum's Universal Robots), Penguin Books, USA 2004, σελ. 79-85

⁴Βλ. T. Higbie, Why Do Robots Rebel? The Labor History of a Cultural Icon, Labor Studies in Working-Class History of the Americas, Αρ. 10, Τεύχος 1, 2013, σελ. 99–121.

⁵Βλ. K. Capek, R.U.R. (Rossum's Universal Robots), Penguin Books, USA 2004, σελ. 79-85

επικοινωνιών. Ο Turing, μέσα από τη μελέτη του, κατέληξε στην ανάπτυξη της «θεωρίας του υπολογισμού». Η συγκεκριμένη θεωρία εμβαθύνει στον βαθμό αποτελεσματικότητας επίλυσης πολύπλοκων προβλημάτων και παρουσιάστηκε στο μοντέλο της μηχανής Turing, ενός διαχρονικά δημοφιλούς επιτεύγματος στην επιστήμη της πληροφορικής⁶.

Λίγα χρόνια μετά τον δεύτερο Παγκόσμιο Πόλεμο, παρουσιάστηκε η λεγόμενη «δοκιμή Turing». Πρόκειται για μία σημαντική προσπάθεια χαρακτηρισμού των μηχανών ως «έξυπνες». Πίσω από το συγκεκριμένο εγχείρημα, υπήρχε η λογική ότι μία μηχανή μπορεί να χαρακτηριστεί «έξυπνη», αν επικοινωνεί με ένα πρόσωπο χρησιμοποιώντας τη φυσική γλώσσα, με την προϋπόθεση ότι ένα δεύτερο πρόσωπο αδυνατεί να διαχωρίσει ποιο από τα δύο αρχικά αναφερόμενα μέρη είναι ο άνθρωπος και ποιο η μηχανή⁷.

Η διάσκεψη του Dartmouth, αποτελεί ένα ακόμα σημείο άξιο αναφοράς σε αυτήν την ιστορική αναδρομή. Πιο συγκεκριμένα, το 1956 εντοπίζεται ίσως το πρώτο εργαστήριο Τεχνητής Νοημοσύνης και με αυτό γεννήθηκε το πεδίο της έρευνας της τεχνητής νοημοσύνης. Η έρευνα αυτή χαρακτηριζόταν από υψηλά επίπεδα αισιοδοξίας. Κατέληγαν μάλιστα σε πεποιθήσεις, σύμφωνα με τις οποίες οι μηχανές θα είναι σε θέση, μέσα σε είκοσι χρόνια, να προβούν στην εκτέλεση οποιασδήποτε εργασίας, την οποία μπορεί να κάνει ένας άνθρωπος⁸.

Επίσης, σύμφωνα με δηλώσεις του Marvin Minsky (MIT), «μέσα σε μια γενιά, το πρόβλημα της δημιουργίας της τεχνητής νοημοσύνης θα έχει επιλυθεί ουσιαστικά»⁹. Πρόκειται για μια δήλωση, που διατυπώθηκε κατά τη δεκαετία του 1960, ωστόσο δεν επιβεβαιώθηκε στα επόμενα έτη, καθώς η πρόοδος της τεχνητής νοημοσύνης επιβραδύνθηκε σημαντικά. Αυτό που είχε παραληφθεί να γίνει κατά τη διεξαγωγή των ανωτέρω αναφερόμενων ερευνών, ήταν η αναγνώριση των δυσκολιών που συνόδευαν

⁶Βλ. C. Petzold, *The Annotated Turing: A Guided Tour Through Alan Turing's Historic Paper on Computability and the Turing Machine*. Wiley, UK 2008, σελ. 57-63.

⁷Βλ. C. Petzold, *όπ.παρ.*, σελ. 57-63.

⁸Βλ. R. Kline, *Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence*. IEEE Annals of the History of Computing, Ap. 33, Τεύχος 4, 2011, σελ. 5–16.

⁹ Βλ. R. Kline, *όπ.παρ.*, σελ. 5-16.

τα ερευνητικά αυτά καθήκοντα. Αυτό, οδηγούσε σε μια υπέρμετρη και αδικαιολόγητη αισιοδοξία και στην διατύπωση υποσχέσεων, οι οποίες φυσικά δεν κατάφεραν να τηρηθούν¹⁰.

Λόγω των υπερβολικά αισιόδοξων ρυθμίσεων και των μη πραγματοποιηθεισών ανακαλύψεων, η αμερικανική και η βρετανική κυβέρνηση, οι οποίες συμμετείχαν στην ανωτέρω αναφερόμενη έρευνα, προέβησαν στη διακοπή της. Τα επόμενα χρόνια ονομάστηκαν ως ο «πρώτος χειμώνας της τεχνητής νοημοσύνης» και συνοδεύτηκαν από απώλεια του ενθουσιασμού και από έλλειψη επαρκών χρηματοδοτικών κεφαλαίων σε έρευνες που επικεντρώνονταν στον τομέα της τεχνητής νοημοσύνης. Επρόκειτο για μια κατάσταση, η οποία επικρατούσε κοντά στο έτος 1974¹¹.

Μετά από τον πρώτο χειμώνα της τεχνητής νοημοσύνης, διαφαίνεται ένα κλίμα επανόδου, το οποίο είναι γνωστό ως «σύστημα εμπειρογνωμόνων»¹². Τα συστήματα εμπειρογνωμόνων είναι προγράμματα, που απαντούν σε στοχευμένες ερωτήσεις και επιλύουν προβλήματα, που προκύπτουν σε έναν ορισμένο τομέα. Μιμούνται, επί της ουσίας, έναν ειδικό (εμπειρογνώμονα) σε ένα συγκεκριμένο κλάδο και επιλύουν προβλήματα, ακολουθώντας καθορισμένους κανόνες. Στα συστήματα εμπειρογνωμόνων εφαρμόζονται δύο διαφορετικές λειτουργίες, η «μηχανή της γνώσης», η οποία εμπεριέχει κανόνες που είναι συναρτημένοι με ένα συγκεκριμένο θέμα και η «μηχανή συμπερασμάτων», η οποία εφαρμόζει τους κανόνες γεγονότα από τη μηχανή γνώσης σε νέα δεδομένα¹³.

Το 1981, ένα εξειδικευμένο σύστημα ονόματι SID (Synthesis of Integral Design) σχεδίασε το 93% των πυλών της λογικής της CPU VAX 9000. Το σύστημα SID απαρτιζόταν από συνολικά 1.000 χειρόγραφους κανόνες. Ο τελικός σχεδιασμός της CPU χρειάστηκε 3 ώρες, για να υπολογίσει και να ξεπεράσει ποικιλοτρόπως, τους

¹⁰Βλ. R. Kline, όπ.παρ., σελ. 5-16.

¹¹Βλ. K. Foote, A Brief History of Artificial Intelligence, 2016 στο <https://www.dataversity.net/brief-history-artificial-intelligence/>

¹² Βλ. G. Shafer, Probability Judgment in Artificial Intelligence and Expert Systems, Statistical Science, Ap. 2, Τεύχος 1, 1987, σελ. 3-16.

¹³Βλ. G. Shafer, όπ.παρ., σελ. 3-16.

ειδικούς-εμπειρογνώμονες, δηλαδή τον ανθρώπινο παράγοντα¹⁴. Αυτό επιτεύχθηκε μέσω της δημιουργίας ταχύτερου κώδικα 64-bit, συγκριτικά με τον κώδικα που είχε σχεδιαστεί χειροκίνητα. Επίσης, το ποσοστό σφάλματος ανά πύλη, στο οποίο εντοπιζόταν κατά προσέγγιση 1 σφάλμα ανά 200 πύλες από τον ανθρώπινο παράγοντα, ήταν πολύ χαμηλότερο στην περίπτωση του μηχανολογικού συστήματος, γιατί κυμαινόταν σε περίπου 1 σφάλμα ανά 20.000 πύλες¹⁵.

Σειρά είχε ο λεγόμενος «δεύτερος χειμώνας» ο οποίος εμφανίστηκε στο προσκήνιο στα τέλη της δεκαετίας του 1980 και στις αρχές της δεκαετίας του 1990 μετά από μια σειρά οικονομικών δυσκολιών. Η μείωση της εμφάνισης εξειδικευμένων τεχνολογικών συστημάτων και τα κατασκευαστικά επιτεύγματα της Apple και της IBM οδήγησαν και πάλι στη μείωση του ενδιαφέροντος για τον τομέα της τεχνητής νοημοσύνης και στην, για ακόμη μια φορά, μείωση της χρηματοδοτικής ενίσχυσης ερευνών, που επικεντρώνονταν στον τομέα αυτόν¹⁶.

Άξιο αναφοράς είναι και το λεγόμενο «DeepBlue» που μετά από αρκετές διακυμάνσεις στην απόδοσή του, έγινε ο πρώτος υπολογιστής σκακιού, ο οποίος κατάφερε να κερδίσει τον παγκόσμιο σκακιστικό πρωταθλητή, Garry Kasparov. Πιο συγκεκριμένα, στις 11 Μαΐου 1997 ο υπολογιστής σκακιού της IBM, νίκησε τον Garry Kasparov μετά από έξι αγώνες, με σκορ 3½-2½¹⁷.

Το Deep Blue, χρησιμοποίησε τη λεγόμενη κλαδική και δενδροειδή αναζήτηση για να υπολογίσει μέχρι και είκοσι πιθανές κινήσεις του αντιπάλου του. Αξιολόγησε μάλιστα, τις θέσεις των πιονιών, μέσα από μια συνάρτηση αξίας, που γράφτηκε κυρίως με το χέρι, η οποία αργότερα βελτιστοποιήθηκε με την ανάλυση χιλιάδων παιχνιδιών¹⁸. Το 1997 το Deep Blue ήταν ο 259ος πιο ισχυρός υπερυπολογιστής με 11.38 GFLOPS. Για

¹⁴Βλ. E. Herrera-Viedma, Z. Vale, P. Nielsen, A. Martin Del Rey και R. Casado Vara, Distributed Computing and Artificial Intelligence, 16th International Conference, Special Sessions, Springer, 2019, σελ. 48-54.

¹⁵Βλ. J. Blake, VAX 9000 Series, Digital Technical Journal, Ap. 2, Τεύχος 4, 1990, σελ. 1-144.

¹⁶Βλ. K. Foote, *όπ.παρ.* στο <https://www.dataversity.net/brief-history-artificial-intelligence/>

¹⁷Βλ. D. Hassabis, Artificial Intelligence: Chess match of the century, Nature, Ap. 544, Τεύχος 765, 2017, σελ. 413-414.

¹⁸Βλ. D. Hassabis, *όπ.παρ.*, σελ. 413-414.

να καταστεί αυτό περισσότερο αντιληπτό, ο πιο ισχυρός υπερυπολογιστής το 1997 είχε 1.068 GFLOPS και μέχρι και τον Δεκέμβριο του 2017, ο πιο ισχυρός υπερυπολογιστής είχε 93.015 GFLOPS¹⁹. Τα FLOPS, αντιπροσωπεύουν λειτουργίες κυμαινόμενου σημείου ανά δευτερόλεπτο και το 'G' στο GFLOPS, σημαίνει Giga. Έτσι το ισοδύναμο 1 GFLOPS είναι 10,9 FLOPS²⁰.

Στις τελευταίες δύο δεκαετίες, ο κλάδος της τεχνητής νοημοσύνης γνωρίζει ταχύρρυθμη ανάπτυξη. Στην πρόοδο αυτή, σημαντικό ρόλο έχει διαδραματίσει η εξελισσόμενη πορεία του τομέα των μεγάλων δεδομένων (bigdata)²¹, καθώς επίσης και η παραγωγή ταχύτερων υπολογιστών, αλλά και το σύνολο των εξελίξεων των τεχνικών της μηχανικής μάθησης²². Επιπροσθέτως, χρησιμοποιώντας τα «Neural Networks», είναι εφικτή η υλοποίηση καθηκόντων αυξημένης πολυπλοκότητας, όπως είναι παραδείγματος χάριν η ανάλυση κειμένων και η επεξεργασία βίντεο. Μάλιστα, είναι αναμενόμενη και η βελτίωση των ήδη υφιστάμενων, κατά τη σημερινή εποχή, λύσεων, στους τομείς αυτούς²³.

¹⁹Βλ. D. Hassabis, *όπ.παρ.*, σελ. 413-414.

²⁰Βλ. M. Campbell, A.J. Hoane & F. Hsu, Deep Blue, Artificial Intelligence, Ap. 134, Τεύχος 1-2, 2002, σελ. 57-83.

²¹Βλ. Z.S.Y. Wong, J. Zhou & Q. Zhang, Artificial Intelligence for infectious disease Big Data Analytics, Infection, Disease & Health, 2018, σελ. 35-48.

²²Βλ. J. Liu, X. Kong, F. Xia, X. Bai, L. Wang, Q. Qing & I. Lee, Artificial Intelligence in the 21st Century. IEEE Access, Έκδοση 6, 2018, σελ. 34403-34421.

²³Βλ. J.S. Almeida, Predictive non-linear modeling of complex data by artificial neural networks, Current Opinion in Biotechnology, Ap. 13, Τεύχος 1, 2002, σελ. 72-76.

1.1.2 Εννοιολογία

Παρά το γεγονός ότι προσφάτως η τεχνητή νοημοσύνη έχει καταστεί ένας πολυσυζητημένος όρος, ωστόσο, ο κλάδος της τεχνητής νοημοσύνης δεν είναι νέος. Ο όρος «Τεχνητή Νοημοσύνη» δημιουργήθηκε το 1956 στο κολλέγιο του Dartmouth, για να προσδιορίσει ένα αναδυόμενο ερευνητικό πεδίο, στο οποίο εμβαθύνουν οι ερευνητές της φυσιολογίας του εγκεφάλου, καθώς επίσης και διακεκριμένοι αναλυτές της εγκεφαλικής λογικής, αλλά και οι μηχανικοί ηλεκτρονικών υπολογιστών. Πρόκειται, επομένως, για έναν τομέα, που αποτελεί τη σφαίρα ενδιαφέροντος μιας πολυδιάστατης ομάδας επιστημόνων, μελετητών και ερευνητών, προερχόμενων από πολλούς και μεταξύ τους διαφοροποιημένους τομείς²⁴.

Σε αυτό το πλαίσιο, η τεχνητή νοημοσύνη αποσαφηνίζεται εννοιολογικά με βάση δύο ικανότητες, από τις οποίες χαρακτηρίζεται²⁵:

- 1) Ως η ικανότητα των μηχανών να εκτελούν καθήκοντα επιδεικνύοντας ευφυή ανθρώπινη συμπεριφορά και
- 2) Ως η ικανότητα των μηχανών να συμπεριφέρονται ορθολογικά, δηλαδή ως έξυπνες οντότητες, που διακατέχονται από την απαιτούμενη και στοιχειώδη αντίληψη του περιβάλλοντος και τη λήψη μέτρων για την επίτευξη ορισμένων στόχων.

Η διαδικασία της διερεύνησης της τεχνητής νοημοσύνης και του πεδίου που αυτή πρεσβεύει, χαρακτηρίζεται τόσο από θετικές, όσο και από αρνητικές περιόδους. Αναλυτικότερα, ένα πρώτο κύμα ενθουσιασμού για τον εν λόγω επιστημονικό τομέα, εμφανίστηκε στα τέλη της δεκαετίας του 1950 και συνδέθηκε με την ικανότητα της δημιουργίας προγραμμάτων, ικανών να αποδείξουν ορισμένα μαθηματικά θεωρήματα ή να βοηθήσουν στη διεξαγωγή απλών παιχνιδιών ψυχαγωγίας²⁶.

Ο αρχικός αυτός ενθουσιασμός, ακολουθήθηκε από μια περίοδο διευρυμένης απογοήτευσης, επί της οποίας οι δοθείσες υποσχέσεις ότι τα συστήματα τεχνητής

²⁴Βλ. S.G. Tzafestas, Artificial intelligence, Roboethics, Ap. 79, 2016, σελ. 25-33

²⁵Βλ. S. Russell & P. Norvig, Artificial intelligence: A modern approach (Global Edition), Pearson Higher Ed, Englewood Cliffs, NJ 2016, σελ. 112-125.

²⁶Βλ. S. Russell & P. Norvig, όπ.παρ., σελ. 112-125.

νοημοσύνης ήταν σε θέση να επιδείξουν παρόμοια ή και υψηλότερα επίπεδα νοημοσύνης από τα αντίστοιχα ανθρώπινα, συνοδεύτηκαν από παραδειγματικές αποτυχίες, όπως συνέβη για παράδειγμα στον μεταφραστικό τομέα²⁷.

Η κατάσταση αυτή οφειλόταν εν μέρει στην ανωριμότητα της τεχνολογίας της πληροφορικής και εν μέρει στην απουσία της βέλτιστης λύσης στο πρόβλημα του τρόπου χαρτογράφησης και ενσωμάτωσης των γνώσεων των συμφραζομένων. Εν συνεχεία, ακολούθησε μια περίοδος, που διήρκεσε μέχρι και το τέλος του 1990 και στο πλαίσιο της οποίας παρατηρήθηκε μειωμένη χρηματοδοτική ενίσχυση του κλάδου της τεχνητής νοημοσύνης, γεγονός που αποδείκνυε το μειωμένο ενδιαφέρον του επενδυτικού και ερευνητικού κοινού επί του τομέα αυτού. Στη σημερινή εποχή, πολλοί συγγραφείς αναφέρονται στην τεχνητή νοημοσύνη χαρακτηρίζοντάς την ως μία νέα επανάσταση, ως συνέπεια του τρόπου με τον οποίο οι μηχανές έχουν μεταμορφώσει την ανθρώπινη εργασία²⁸.

Παρόμοια με τη βιομηχανική επανάσταση και την μετέπειτα ψηφιακή, όπου οι μηχανές προσέφεραν ποικίλους τρόπους για να συμπληρώσουν, να ενισχύσουν και να υποκαταστήσουν τη χειρωνακτική εργασία των ανθρώπων, αντίστοιχα και η νέα επανάσταση, η οποία υποστηρίζεται από την ανάπτυξη της τεχνητής νοημοσύνης, εκλαμβάνεται ως ένας τρόπος συμπλήρωσης, ενίσχυσης και υποκατάστασης σχεδόν του συνόλου των καθηκόντων που εκτελούνται σήμερα από τον άνθρωπο²⁹.

Παρά τον εν λόγω αυξανόμενο ενθουσιασμό, οι περισσότεροι ερευνητές συμφωνούν ότι αυτό που προβάλλεται σήμερα, είναι η διάδοση των ασθενών τεχνολογιών της τεχνητής νοημοσύνης, σε αντίθεση με την ισχυρή τεχνητή νοημοσύνη. Για λόγους πληρέστερης κατανόησης, διευκρινίζουμε ότι η τεχνητή νοημοσύνη αναφέρεται σε συστήματα με ανθρώπινη ή υπεράνθρωπη νοημοσύνη, τα οποία προσομοιώνουν τη σύνθετη ανθρώπινη ικανότητα της σκέψης και της ταυτόχρονης εκτέλεσης καθηκόντων που απαιτούν την παράλληλη αξιοποίηση της λογικής και της κρίσης,

²⁷Βλ. S. Russell & P. Norvig, *όπ.παρ.*, σελ. 112-125.

²⁸Βλ. S. Makridakis, *The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms*, *Futures*, Ap. 90, 2017, σελ. 46–60.

²⁹Βλ. S. Makridakis, *όπ.παρ.*, σελ. 46-60.

όπως είναι επί παραδείγματι οι ηθικές κρίσεις, η συμβολική συλλογιστική, η διαχείριση κοινωνικών καταστάσεων και ο ιδεασμός³⁰.

Η αδύναμη τεχνητή νοημοσύνη αναφέρεται σε συστήματα που μπορούν να πραγματώνουν καθήκοντα που απαιτούν απλές ανθρώπινες ικανότητες, όπως είναι για παράδειγμα η οπτική αντίληψη, το πλαίσιο κατανόησης, η πιθανοτική συλλογιστική αντιμετώπιση καταστάσεων και η αντιμετώπιση της πολυπλοκότητας από την οποία ενδεχομένως χαρακτηρίζονται ορισμένες περιστάσεις³¹.

Ωστόσο, έχει διατυπωθεί η άποψη ότι μονάχα η αδύναμη μορφή της τεχνητής νοημοσύνης παρουσιάζει ενδιαφέρον για εφαρμογές σε πραγματικό κόσμο, καθώς τα ισχυρά συστήματα τεχνητής νοημοσύνης, όπως αυτά περιγράφηκαν ανωτέρω, εξακολουθούν να θεωρούνται ως μία αμιγής προσπάθεια κερδοσκοπίας και ως ένα σενάριο επιστημονικής φαντασίας³².

Παρ' όλο που η τεχνητή νοημοσύνη ως όρος έχει διερευνηθεί εδώ και δεκαετίες, δεν υπάρχει καθολικός αποδεκτός ορισμός. Αυτό οδηγεί στο θεμελιώδες πρόβλημα της κατανόησης της τεχνητής νοημοσύνης στο σύνολό της³³.

Για να αποκτηθεί μία βασική κατανόηση της τεχνητής νοημοσύνης, κρίνεται φρόνιμο να οριστεί πρώτα η έννοια της «ευφυΐας» ως ένας ρητός όρος, προτού εφαρμοσθεί η νοημοσύνη στις μηχανές και τεθεί στο προσκήνιο ο περισσότερος σύνθετος όρος «τεχνητή νοημοσύνη». Λαμβάνοντας υπόψη μία ποικιλία προηγούμενων ορισμών, αναφέρεται ότι οι Legg και Hutter³⁴ αναπτύσσουν έναν ολοκληρωμένο ορισμό της νοημοσύνης, περιγράφοντάς την ως την ικανότητα αλληλεπίδρασης, εκμάθησης, υιοθέτησης και προσφυγής σε πληροφορίες από εμπειρίες, καθώς και της

³⁰Βλ. E. Brynjolfsson & A. McAfee, *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*, W. W. Norton & Company, New York 2014, σελ. 112-127.

³¹Βλ. S. Russell & P. Norvig, *όπ.παρ.*, σελ. 112-125.

³²Βλ. S. Russell & P. Norvig, *όπ.παρ.*, σελ. 112-125.

³³ Βλ. S. Legg, & M. Hutter, *A formal measure of machine intelligence*, In *Proc. 15th Annual Machine Learning Conference of Belgium and The Netherlands*, Benelearn 2016, 2006, σελ. 73–80.

³⁴ Βλ. S. Legg, & M. Hutter, *όπ.παρ.*, σελ. 73–80.

αντιμετώπισης της αβεβαιότητας. Σε συνδυασμό με αυτό, ο όρος «τεχνητή» σημαίνει ένα στοιχείο που παράγεται από ανθρώπους³⁵.

Με βάση αυτήν τη θεμελιώδη κατανόηση, φαίνεται σκόπιμο να αναπτυχθεί ένας πιο εξελιγμένος ορισμός της τεχνητής νοημοσύνης στο πλαίσιο της δημόσιας διαχείρισης, προκειμένου να καταστούν κατανοητά τα όσα θα αναφερθούν παρακάτω επί του τρέχοντος κεφαλαίου. Προσβλέποντας στη δημιουργία μιας ολοκληρωμένης και σαφούς κατανόησης του όρου, εντοπίσαμε έξι ορισμούς της τεχνητής νοημοσύνης από σχετικά άρθρα. Οι ορισμοί αυτοί είναι οι εξής:

- Η μελέτη των McCarthy et al.³⁶ βασίζεται στην εικασία ότι κάθε πτυχή της μάθησης ή οποιοδήποτε άλλο χαρακτηριστικό της νοημοσύνης μπορεί κατ' αρχάς να περιγραφεί με ακρίβεια και εν συνεχεία μπορεί να κατασκευαστεί μία μηχανή για να την προσομοιώσει.
- Οι Rich et al.³⁷ υποστηρίζουν ότι η τεχνητή νοημοσύνη αντιπροσωπεύει τη μελέτη του πώς να καταστούν οι υπολογιστές ικανοί να διεκπεραιώνουν εργασίες που, προς το παρόν, οι άνθρωποι διεκπεραιώνουν καλύτερα.
- Οι Russel & Norvig³⁸ υποστηρίζουν ότι η τεχνητή νοημοσύνη μπορεί να οργανωθεί σε τέσσερις κατηγορίες, ήτοι τα συστήματα που σκέφτονται σαν τους ανθρώπους, τα συστήματα που δρουν σαν τους ανθρώπους, τα συστήματα που σκέφτονται ορθολογικά και τα συστήματα που λειτουργούν ορθολογικά.
- Οι Adams et al.³⁹ υποστηρίζουν ότι η τεχνητή νοημοσύνη είναι ένα σύστημα που θα μπορούσε να μάθει, να αναπαράγει, και πιθανώς να υπερβαίνει την

³⁵ Βλ. E.A. Patrick, & J.M. Fattu, Artificial intelligence with statistical pattern recognition, Englewood Cliffs, NJ, Prentice Hall, 1986, σελ. 47-61.

³⁶ Βλ. J. McCarthy, M.L. Minsky, N. Rochester, & C.E. Shannon, A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955, AI Magazine, Αρ. 27, Τεύχος 4, 2006., σελ. 12-14.

³⁷ Βλ. E. Rich, K. Knight, & S.B. Nair, Artificial intelligence. Third edition, New Dehli, Tata McGraw-Hill, 2009, σελ. 3.

³⁸ Βλ. S. Russel & R. Norvig, Artificial intelligence: A modern approach, Third edition, Englewood Cliffs, Prentice Hall, 2010, σελ. 2.

³⁹ Βλ. S. Adams, I. Arel, J. Bach, R. Coop, R. Furlan, B. Goertzel, J.S. Hall, A. Samsonovich, M. Scheutz, M. Schlesinger, S.C. Shapiro, & J. Sowa, Mapping the landscape of human-level artificial general intelligence, AI Magazine, Αρ. 33, Τεύχος 1, 2012, σελ. 25-42.

ανθρώπινη απόδοση σε όλο το εύρος των γνωστικών και πνευματικών ικανοτήτων της.

- Οι Rosa et al.⁴⁰ θεωρούν ότι η τεχνητή νοημοσύνη αντικατοπτρίζει τα προγράμματα που μπορούν να μάθουν, να προσαρμοστούν, να είναι δημιουργικά και να λύνουν προβλήματα.
- Οι Thierer et al.⁴¹ κάνουν μνεία στην έκθεση της νοημοσύνης από μία μηχανή. Αναφέρουν ότι ένα σύστημα τεχνητής νοημοσύνης είναι ικανό να εκτελεί λειτουργίες υψηλού επιπέδου. Η τεχνητή νοημοσύνη μπορεί να προσεγγίζει τις ικανότητες ενός ανθρώπου ή να επεκταθεί ακόμα και πέραν αυτών. Αυτή η ιδέα επιμερίζεται στην υπόσταση αδύναμης και ισχυρής τεχνητής νοημοσύνης.

Όλοι οι προαναφερθέντες ορισμοί αναφέρονται στα χαρακτηριστικά των συστημάτων που βασίζονται στη μηχανή και την ανθρώπινη ευφυή συμπεριφορά⁴². Η τεχνητή νοημοσύνη προσπαθεί να αναπαράγει ανθρώπινες πρακτικές επίλυσης προβλημάτων για την επίτευξη πιο αποτελεσματικών λύσεων⁴³. Ένα ιδιαίτερο χαρακτηριστικό της είναι η αναπαραγωγή της ανθρώπινης σκέψης και μάθησης, καθώς και οι δυνατότητες επίλυσης προβλημάτων που αποσκοπούν στην ενίσχυση της ανθρώπινης αποδοτικότητας⁴⁴. Με βάση αυτούς τους ανωτέρω αναφερόμενους ορισμούς, είναι δυνατό να αντληθεί ένας ολοκληρωμένος ορισμός της τεχνητής νοημοσύνης που θα ληφθεί υπόψη ως δεδομένος για την μετέπειτα ανάλυση.

Συνεπώς, η τεχνητή νοημοσύνη αναφέρεται στην ικανότητα ενός υπολογιστικού συστήματος να επιδεικνύει έξυπνη συμπεριφορά η οποία μοιάζει με ανθρώπινη και που χαρακτηρίζεται από ορισμένες βασικές ικανότητες, όπως είναι η αντίληψη, η

⁴⁰ Βλ. A. Rosa, J. Feyereisl, & T.G. Team, A framework for searching for general artificial intelligence, 2016, στο <https://arxiv.org/pdf/1611.00685.pdf>

⁴¹ Βλ. A. Thierer, A. O'Sullivan Castillo, & R. Russell, Artificial intelligence and public policy, Mercatus research, Mercatus Center at George Mason University, 2017, στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁴² Βλ. S. Adams, I. Arel, J. Bach, R. Coop, R. Furlan, B. Goertzel, J.S. Hall, A. Samsonovich, M. Scheutz, M. Schlesinger, S.C. Shapiro, & J. Sowa, *όπ.παρ.*, σελ. 25–42, Βλ. επίσης J. McCarthy, M.L. Minsky, N. Rochester, & C.E. Shannon, *όπ.παρ.*, σελ. 12–14, Βλ. επίσης A. Thierer, *όπ.παρ.*, στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁴³ Βλ. S. Russel & R. Norvig, *όπ.παρ.*, σελ. 2.

⁴⁴ Βλ. E. Rich, K. Knight, & S.B. Nair, *όπ.παρ.*, σελ. 3, Βλ. επίσης A. Rosa, J. Feyereisl, & T.G. Team, *όπ.παρ.*, στο <https://arxiv.org/pdf/1611.00685.pdf>

κατανόηση, η μάθηση, η προσαρμογή και η δράση⁴⁵. Σύμφωνα με το ανωτέρω, μία εφαρμογή-μηχανή τεχνητής νοημοσύνης αναφέρεται στην ενσωμάτωση της τεχνολογίας τεχνητής νοημοσύνης σε ένα πεδίο εφαρμογών υπολογιστών με αλληλεπίδραση ανθρώπου-υπολογιστή και αλληλεπίδραση δεδομένων⁴⁶.

Στο πλαίσιο των πρόσφατων εξελίξεων και περαιτέρω καινοτομιών, η τεχνητή νοημοσύνη δύναται να επιμεριστεί σε τρεις κύριες κατηγορίες, ήτοι τη στενή τεχνητή νοημοσύνη (Artificial Narrow Intelligence - ANI), τη γενική τεχνητή νοημοσύνη (Artificial General Intelligence - AGI) και την τεχνητή υπερ-νοημοσύνη (Artificial Super Intelligence - ASI)⁴⁷.

Η στενή – αδύναμη τεχνητή νοημοσύνη είναι συνήθως ικανή να λύσει ένα συγκεκριμένο πρόβλημα και πρέπει να προγραμματιστεί από τους ανθρώπους⁴⁸. Αντίθετα, η γενική – ισχυρή τεχνητή νοημοσύνη είναι σε θέση αυτοτελούς εκμάθησης και μετάδοσης των εμπειριών και δεξιοτήτων της σε άλλες εργασίες, χωρίς καμία ανθρώπινη βοήθεια. Στην ουσία πρόκειται για ένα φαινομενικά ανθρώπινο σύστημα που διαθέτει γνωστικές ικανότητες και ενσωματώνει εμπειρίες από το ανθρώπινο περιβάλλον και μπορεί να επεξεργάζεται τα δεδομένα αυτά σε πολύ ταχύτερο ρυθμό

⁴⁵ Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, Τεχνητή Νοημοσύνη: Ο δρόμος προς έναν ψηφιακό συνταγματισμό. Μια ηθικο-συνταγματική θεώρηση, 2023, σελ. 31-32, Βλ. επίσης S. Adams, I. Arel, J. Bach, R. Coop, R. Furlan, B. Goertzel, J.S. Hall, A. Samsonovich, M. Scheutz, M. Schlesinger, S.C. Shapiro, & J. Sowa, *όπ.παρ.*, σελ. 25–42, Βλ. επίσης J. McCarthy, M.L. Minsky, N. Rochester, & C.E. Shannon, *όπ.παρ.*, σελ. 12–14, Βλ. επίσης A. Thierer, *όπ.παρ.*, στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>, Βλ. επίσης S. Russel & R. Norvig, *όπ.παρ.*, σελ. 2.

⁴⁶ Βλ. S. Russel & R. Norvig, *όπ.παρ.*, σελ. 2, Βλ. επίσης E. Rich, K. Knight, & S.B. Nair, *όπ.παρ.*, σελ. 3, Βλ. επίσης A. Rosa, J. Feyereisl, & T.G. Team, *όπ.παρ.*, στο <https://arxiv.org/pdf/1611.00685.pdf>

⁴⁷ Βλ. S. Adams, I. Arel, J. Bach, R. Coop, R. Furlan, B. Goertzel, J.S. Hall, A. Samsonovich, M. Scheutz, M. Schlesinger, S.C. Shapiro, & J. Sowa, *όπ.παρ.*, σελ. 25–42, Βλ. επίσης N.J. Nilsson, *Artificial intelligence: A new synthesis*, San Francisco, Morgan Kaufmann Publishers, 2003, σελ. 32-39, Βλ. επίσης A. Thierer, *όπ.παρ.*, στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁴⁸ Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, *όπ.παρ.*, σελ. 34, Βλ. επίσης A. Rosa, J. Feyereisl, & T.G. Team, *όπ.παρ.*, στο <https://arxiv.org/pdf/1611.00685.pdf>

συγκριτικά με τον άνθρωπο⁴⁹. Ως εκ τούτου, είναι αναπόφευκτη η σύγκριση των ανθρωπινων ικανοτήτων με την τεχνητή νοημοσύνη.

Η τεχνητή υπερ-νοημοσύνη, με τη σειρά της, αντιπροσωπεύει την ανάπτυξη λογισμικού πολύ πιο προηγμένου από τον ανθρώπινο εγκέφαλο. Εντούτοις, μέχρι σήμερα δεν υφίσταται τέτοιου είδους λογισμικό, αν και οι επιστήμονες θεωρούν ότι η τεχνητή υπερ-ευφυΐα θα εξελιχθεί αυτόματα από τη γενική τεχνητή νοημοσύνη⁵⁰.

Η δυνητικά συνδεδεμένη απώλεια ανθρώπινου ελέγχου, καθώς και η μεταφορά ανθρωπινων δραστηριοτήτων και γνώσεων στα μηχανήματα δημιουργεί ανησυχίες που σχετίζονται με κυβερνητικές πολιτικές και ηθικά ζητήματα⁵¹.

Στην περίπτωση αυτών των αμφιβολιών και ανησυχιών, σε συνδυασμό με τη διαφαινόμενη εξέλιξη από τη στενή τεχνητή νοημοσύνη στη γενική τεχνητή νοημοσύνη, είναι προφανές ότι η κυβέρνηση κάθε κράτους πρέπει να διαδραματίζει τον ρόλο του επόπτη της τεχνολογίας, γενικά, και του αποδέκτη της τεχνολογίας για συγκεκριμένες εφαρμογές τεχνητής νοημοσύνης. Ωστόσο, πριν συζητηθούν οι προκλήσεις που σχετίζονται με την τεχνητή νοημοσύνη, οι οποίες ακολουθούν στο αμέσως επόμενο υποκεφάλαιο, κρίνεται σκόπιμο να εξεταστεί η δυναμική και η χρηστικότητα των εφαρμογών τεχνητής νοημοσύνης στο δημόσιο τομέα. Όπως αναφέρθηκε ήδη, η τεχνητή νοημοσύνη δύναται να μειώσει τον διοικητικό φόρτο και να ενθαρρύνει την κατανομή πόρων⁵².

⁴⁹ Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 34, Βλ. Επίσης S. Adams, I. Arel, J. Bach, R. Coop, R. Furlan, B. Goertzel, J.S. Hall, A. Samsonovich, M. Scheutz, M. Schlesinger, S.C. Shapiro, & J. Sowa, όπ.παρ., σελ. 25–42.

⁵⁰ Βλ. C. Kreinczes, Artificial Intelligence Innovation Report, 2016, στο <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Artificial-Intelligence-Innovation-Report-2018-Deloitte.pdf>

⁵¹ Βλ. D.G. Johnson, & M. Verdicchio, AI anxiety. Journal of the Association for Information Science and Technology, Αρ. 68, Τεύχος 9, 2017, σελ. 2267–2270, Βλ. επίσης M. Boyd, & N. Wilson, Rapid developments in artificial intelligence: How might the New Zealand government respond, Policy Quarterly, Αρ. 13, Τεύχος 4, 2017, σελ. 36–44

⁵² Βλ. W.D. Eggers, T. Fishman, & P. Kishnani, AI-augmented human services: Using cognitive technologies to transform program delivery, 2017, στο https://www2.deloitte.com/content/dam/insights/us/articles/4152_AI-human-services/4152_AI-human-services.pdf

Μολονότι του διαφαινόμενου οφέλους των μειωμένων διοικητικών επιβαρύνσεων, υπάρχει ακόμη περιορισμένη γνώση σχετικά με τους τύπους και την ευρύτερη δυναμική των εφαρμογών τεχνητής νοημοσύνης για τις κυβερνήσεις. Επιπροσθέτως, εν όψει των αναδυόμενων προκλήσεων, υφίσταται ένα χάσμα μεταξύ των προσδοκιών των πολιτών και των κυβερνητικών ικανοτήτων επί του συγκεκριμένου πεδίου⁵³.

Τέλος, επισημαίνεται ότι δεν θα πρέπει να συγχέεται το πεδίο της τεχνητής νοημοσύνης με τα μεγάλα δεδομένα (Big Data). Τα μεγάλα δεδομένα αναφέρονται σε μεγάλο όγκο «ετερόκλητων» δεδομένων, προερχόμενων από διαφορετικές πηγές και τα οποία επεξεργάζονται και αξιοποιούν τα συστήματα τεχνητής νοημοσύνης προκειμένου να λειτουργήσουν. Ενδεικτικά παραδείγματα μεγάλων δεδομένων, ανάλογα με το πεδίο δραστηριοποίησης του εκάστοτε αλγορίθμου, αποτελούν τα στατιστικά στοιχεία, οι βιβλιογραφικές αναφορές, αρχεία και λοιπές πληροφορίες⁵⁴.

⁵³ Βλ. H. Mehr, Artificial intelligence for citizen services and government. Cambridge, MA, Harvard Kennedy School, Ash Center for Democratic Governance And Innovation, 2017, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁵⁴Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 148-149, Βλ. Επίσης B. Wirtz, J. Weyerer, & C. Geyer, Artificial Intelligence and the Public Sector—Applications and Challenges, International Journal of Public Administration, 2018, σελ. 1–20.

1.2 Ασφάλεια και «βρώμικα» δεδομένα

Για λόγους αποφυγής παρερμηνειών, επισημαίνεται ότι ως ασφάλεια στις μηχανές τεχνητής νοημοσύνης νοείται η *«ικανότητα λειτουργίας τους χωρίς να θέτουν σε κίνδυνο ή να προκαλούν βλάβη στους ανθρώπους»*⁵⁵. Ουσιαστικά, τη στιγμή κατά την οποία μία μηχανή τεχνητής νοημοσύνης έχει σχεδιαστεί με τέτοιον τρόπο ώστε η χρήση της να είναι επωφελής για τους πολίτες της κοινωνίας, δίχως να προκαλεί ατυχήματα ή βλάβες, τότε υφίστανται αυξημένα επίπεδα ασφάλειας στη χρήση της.

Από την πλευρά των επιστημόνων, ωστόσο, η ασφάλεια που αφορά στη χρήση μηχανών τεχνητής νοημοσύνης ορίστηκε με τον χαρακτηρισμό *«ευεργετική τεχνητή νοημοσύνη»* ή *«ισχυρή τεχνητή νοημοσύνη»*, όπου υποστηρίζεται ότι οι άνθρωποι *«πρέπει να γίνουν πιο προνοητικοί παρά αντιδραστικοί»*⁵⁶.

Η ασφάλεια αποτελεί ένα ζήτημα που σχετίζεται άμεσα με τους κατασκευαστές και προγραμματιστές των μηχανών τεχνητής νοημοσύνης και των αρμοδίων προσώπων για την χάραξη πολιτικών, από τις οποίες ρυθμίζονται κατασκευαστικά και σχεδιαστικά ζητήματα των εν λόγω μηχανών⁵⁷.

Στο σημείο αυτό, τονίζεται ότι ένα βασικό ζήτημα που εγείρεται πάνω στον όρο της *«ασφάλειας της τεχνητής νοημοσύνης»* αφορά στην απειλή ή την καταχώριση *«βρώμικων δεδομένων»* (dirty data) στα εθνικά συστήματα βάσεων δεδομένων⁵⁸. Τα βρώμικα δεδομένα είναι ένας όρος που χρησιμοποιείται συνήθως στο πεδίο της εξόρυξης δεδομένων και αφορά *«δεδομένα που λείπουν, λανθασμένα δεδομένα και μη τυπικές αναπαραστάσεις των ιδίων δεδομένων»*⁵⁹.

⁵⁵Βλ. N. Malli, M. Jacobs & S. Villeneuve, *όπ.παρ.*, σελ. 11.

⁵⁶Βλ. M. Tegmark, *όπ.παρ.*, σελ. 94.

⁵⁷Βλ. M. Tegmark, *όπ.παρ.*, σελ. 95.

⁵⁸Βλ. W. Kim, B. Choi, E. Hong, S. Kim & L. Doheon, Data Mining And Knowledge Discovery, Τεύχος 7, Αρ. 1, 2003, σελ. 81-99.

⁵⁹Βλ. R. Richardson, J. Schultz & K. Crawford, Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems, And Justice. New York University Law Review Online, Forthcoming, Τεύχος 94, Αρ. 192, 2019, 195.

Στην πρόσφατη βιβλιογραφία, μάλιστα, παρατηρήθηκε η προσθήκη μίας εννοιολογικής προέκτασης του ανωτέρω αναφερόμενου ορισμού, τη στιγμή κατά την οποία τα βρώμικα δεδομένα επισημαίνεται ότι περιλαμβάνουν και «*δεδομένα που προέρχονται ή επηρεάζονται από διεφθαρμένες, μεροληπτικές και παράνομες πρακτικές, συμπεριλαμβανομένων δεδομένων που έχει υποστεί σκόπιμη χειραγώγηση ή «πλάκα», καθώς και δεδομένα που παραμορφώνονται από ατομικές και κοινωνικές προκαταλήψεις*»⁶⁰.

Διακρίνουμε ένα σύνολο ορισμών από τους οποίους καταδεικνύεται η παρουσία προβλημάτων στη διαδικασία συλλογής δεδομένων στα πλαίσια των κυβερνητικών δομών⁶¹. Πρόκειται για ένα σύνολο δεδομένων που είναι βασισμένα σε αποφάσεις που έχουν ληφθεί από ανθρώπους⁶² ή ακόμα και σε ερμηνείες που έχουν γίνει επί συγκεκριμένων γεγονότων και ανθρώπων και ως εκ τούτου, στερούνται αντικειμενικότητας και χαρακτηρίζονται από την ανθρώπινη μεροληψία⁶³. Εφόσον οι κυβερνήσεις στηρίζονται κατά κόρον σε δεδομένα με υψηλά επίπεδα αντικειμενικότητας, προκειμένου να προβούν στην εκτέλεση των κυβερνητικών τους καθηκόντων, αναπόφευκτα αυξάνεται ο κίνδυνος υπονόμησης της αρχής της ισότητας και της δικαιοσύνης⁶⁴.

⁶⁰Βλ. R. Richardson, J. Schultz & K. Crawford, *όπ.παρ.*, 2019, 195.

⁶¹Βλ. R. Richardson, J. Schultz & K. Crawford, *όπ.παρ.*, 2019, 226

⁶²Βλ. R. Richardson, J. Schultz & K. Crawford, *όπ.παρ.*, 2019, 226

⁶³Βλ. R. Richardson, J. Schultz & K. Crawford, *όπ.παρ.*, 2019, 226.

⁶⁴Βλ. R. Richardson, J. Schultz & K. Crawford, *όπ.παρ.*, 2019, 226.

1.3 Μηχανική Μάθηση

Πρέπει να αναφερθεί ότι με τον όρο «μηχανική μάθηση» νοείται ένα ευρύ σύνολο τεχνικών, από τις οποίες καθίσταται εφικτή η προσαρμογή και εκμάθηση από τα διαθέσιμα δεδομένα, έτσι ώστε να επιτυγχάνεται βέλτιστη απόδοση σε μία συγκεκριμένη εργασία. Με τον τρόπο αυτόν, μπορεί να επιτυγχάνεται και μέγιστη ακρίβεια προβλέψεων⁶⁵. Ένας αναλυτής δεδομένων προβαίνει στην επιλογή του αλγορίθμου μηχανικής μάθησης, καθορίζει τη βέλτιστη δυνατή μέθοδο μετρήσεων για την αξιολόγηση της ακρίβειας των δεδομένων που θα εξαχθούν από τη διαδικασία της μέτρησης και στη συνέχεια, «εκπαιδεύει» τον αλγόριθμο χρησιμοποιώντας το σχετικό δείγμα κατάρτισης⁶⁶. Ορισμένοι από τους πιο συχνά χρησιμοποιούμενους αλγόριθμους πρόβλεψης, όπως είναι επί παραδείγματι η υλικοτεχνική παλινδρόμηση και τα νευρωνικά δίκτυα, συνάγουν τη μεταβλητή του αποτελέσματος, που θεωρείται ενδιαφέρουσα και που προκύπτει από στατιστικές συσχετίσεις, μεταξύ των μεταβλητών που έχουν προσδιοριστεί προς περαιτέρω παρατήρηση και ανάλυση. Κατόπιν, το επίπεδο ακρίβειας των προκαταρκτικών μοντέλων αξιολογείται στα δεδομένα ανάπτυξης μέχρι τη στιγμή της σταθεροποίησής του σε ένα αποδεκτό επίπεδο⁶⁷.

Κατά τη διαδικασία των προσλήψεων, παραδείγματος χάριν, είναι πιθανό να παρατηρηθεί ποια συγκεκριμένα χαρακτηριστικά των αιτούντων έχουν εμφανίσει συσχέτιση με αυξημένα επίπεδα εργασιακής αποδοτικότητας. Στη συνέχεια, τα χαρακτηριστικά αυτά μπορούν να χρησιμοποιηθούν στο στάδιο της επιλογής υποψήφιων εργαζομένων για έναν οργανισμό⁶⁸.

Οι συγκεκριμένοι αλγόριθμοι εμφανίζουν διαφορές σε ορισμένες περιπτώσεις επί των χρησιμοποιούμενων τεχνικών, συγκριτικά με υφιστάμενες παραδοσιακές

⁶⁵Βλ. Y. Zhao, M. Hryniewicki, F. Cheng, B. Fu & X. Zhu, Employee Turnover Prediction with Machine Learning: A Reliable Approach. Intelligent Systems and Applications, 2018, σελ. 737–758.

⁶⁶Βλ. Y. Zhao, M. Hryniewicki, F. Cheng, B. Fu & X. Zhu, όπ.παρ., σελ. 737–758.

⁶⁷Βλ. Y. Zhao, M. Hryniewicki, F. Cheng, B. Fu & X. Zhu, όπ.παρ., σελ. 737–758.

⁶⁸Βλ. M.K. Lee, D. Kusbit, E. Metsky & L. Dabbish, Working with machines: The impact of algorithmic, data-driven management on human workers, New York, ACM Press, 2015, σελ. 1603-1612.

προσεγγίσεις, οι οποίες συνήθως αξιοποιούνται στη διαχείριση ανθρωπίνων πόρων⁶⁹. Η μηχανική μάθηση, αντίθετα, προβαίνει στη δημιουργία αλγορίθμων, από τους οποίους μπορεί να γίνει χρήση πολλών μεταβλητών, οι οποίες μπορεί να μην εντοπίζονται στο πλαίσιο των θεωρητικών προσεγγίσεων που έχουν ως αντικείμενο τους προσδιοριστικούς παράγοντες που συναντάμε στην εργασιακή αποδοτικότητα⁷⁰. Ένα από τα βασικά σημεία της μηχανικής μάθησης αφορά στη διερεύνηση μη παραδοσιακών προσδιοριστικών παραγόντων και μεταβλητών. Αυτό συμβαίνει λόγω του ότι απώτερος σκοπός είναι η οικοδόμηση ενός καλύτερου, ποιοτικότερου και πιο αξιόπιστου προληπτικού συστήματος⁷¹.

⁶⁹Βλ. M.K. Lee, D. Kusbit, E. Metsky& L. Dabbish, *όπ.παρ.*, σελ. 1603-1612.

⁷⁰Βλ. M.K. Lee, D. Kusbit, E. Metsky& L. Dabbish, *όπ.παρ.*, σελ. 1603-1612.

⁷¹Βλ. M.K. Lee, D. Kusbit, E. Metsky& L. Dabbish, *όπ.παρ.*, σελ. 1603-1612.

1.4 «Έξυπνα-Εξημερωμένα» Αντικείμενα

Στην έκθεση του Ευρωπαϊκού Κοινοβουλίου επισημαίνονται ως «έξυπνες» μηχανές, αυτές που η αυτονομία τους μπορεί να αποδειχθεί από τον τρόπο που διασυνδέονται με το περιβάλλον στο οποίο καλούνται να λειτουργήσουν και από το πόσο μπορούν να μεταβάλλουν τις δράσεις τους καθώς προσπαθούν να προσαρμοστούν στις νέες συνθήκες που επικρατούν στο συγκεκριμένο περιβάλλον⁷².

Χρήσιμο είναι να επισημάνουμε ότι υπάρχουν απλές περιπτώσεις τεχνικών αντικειμένων που αποτελούν ανθρώπινα κατασκευάσματα, όπως είναι για παράδειγμα ένα τραπέζι ή μία καρέκλα, που είναι «εξημερωμένα» για να χρησιμοποιούνται με ασφάλεια από τον άνθρωπο και να βελτιώνουν την ποιότητα και τις συνθήκες της καθημερινής του διαβίωσης. Διευκρινίζεται ότι ως «εξημερωμένα» ή «έξυπνα» νοούνται τα τεχνικά αντικείμενα, που κατά τη χρήση τους από τον άνθρωπο, η συμπεριφορά τους είναι προβλέψιμη αναφορικά με την επιρροή που θα ασκήσουν στον χρήστη τους⁷³. Λαμβάνοντας αυτό υπόψη, υποστηρίζεται ότι τα ρομπότ και οι «έξυπνοι» αλγόριθμοι δεν θεωρούνται οικιακές τεχνολογίες⁷⁴, δεδομένου ότι η εποχή της αλληλεπίδρασης με τους ανθρώπους, σε όλο το ιστορικό υπόβαθρο που διαθέτει αυτή η σχέση αλληλεπίδρασης και συνύπαρξης, δεν έχει επιτρέψει ακόμη την πρόβλεψη περισσότερων κινδύνων, προκειμένου αυτοί να ελεγχθούν ή να αναχαιτιστούν⁷⁵.

⁷²Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, & L. Floridi, Artificial Intelligence and the “Good Society”: the US, EU, and UK approach, *Science and Engineering Ethics*, Apr. 24, 2017, σελ. 505-528.

⁷³Βλ. R. Bogue, Domestic robots: Has their time finally come, *Industrial Robot: An International Journal*, Apr.44, Τεύχος 2, 2017, σελ. 129–136.

⁷⁴Βλ. D. Spoladore, S. Arlati & M. Sacco, Semantic and Virtual Reality-Enhanced Configuration of Domestic Environments: The Smart Home Simulator, *Mobile Information Systems*, 2017, σελ. 1–15.

⁷⁵Βλ. C. Wong, E. Yang, X.T. Yan & D. Gu, Adaptive and intelligent navigation of autonomous planetary rovers — A survey, 2017 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), 2017, σελ. 237-244.

1.5 Η Κατανόηση της Τεχνολογίας

Στο σημείο αυτό, η ανασκόπηση της βιβλιογραφίας που πραγματοποιείται αποσκοπεί στη διερεύνηση της σημασιολογικής αποσαφήνισης της «κατανόησης της τεχνολογίας» και των προϋποθέσεων που πρέπει να συντρέχουν για την πλήρη επίγνωση αυτής. Η κατανόηση της τεχνολογίας, αναλυτικότερα, υποδηλώνει την παρουσία ενός τεχνολογικού πλαισίου, από το οποίο μπορούν να προβλέπονται οι «κοινές υποθέσεις, γνώσεις και προσδοκίες σχετικά με το σκοπό, το πλαίσιο και τη σημασία της τεχνολογίας»⁷⁶.

Οι Bostrom, Dafoe και Flynn⁷⁷, οι οποίοι έχουν εμβαθύνει στον κλάδο της τεχνητής νοημοσύνης, θεωρούν ότι η κατανόηση της τεχνολογίας αντικατοπτρίζει τη δυνατότητα και τάση των κυβερνητικών παραγόντων να ενσωματώσουν τις απαραίτητες τεχνικές αρχές και γνώσεις στη διαδικασία λήψης πολιτικών αποφάσεων. Στην ίδια μελέτη, μάλιστα, επισημαίνεται η αναγκαιότητα της κατανόησης της τεχνολογίας μεταξύ των υπευθύνων χάραξης πολιτικών, εξαιτίας της απουσίας και της ταυτόχρονης μεγάλης σημασίας της πολιτικής υπηρεσίας στην πρόοδο που γνωρίζει ο κλάδος της τεχνητής νοημοσύνης⁷⁸. Συνεχίζοντας, ο κοινωνιολόγος Ulrich Hilpert⁷⁹ διευκρίνισε ότι η κατανόηση της τεχνολογίας, υπό το πρίσμα της τεχνητής νοημοσύνης, είναι απόλυτα εξαρτώμενη από τον βαθμό στον οποίο έχουν κατανοηθεί οι απαιτούμενες συνθήκες για τη διασφάλιση της αποτελεσματικής και καινοτόμου τεχνολογικής προόδου.

Από την πλευρά του Hilpert⁸⁰ επισημαίνεται ο στρατηγικός ρόλος που διαδραματίζει η ακαδημαϊκή έρευνα και η προθυμία των κυβερνήσεων να προχωρήσουν σε χρηματοδοτικές ενισχύσεις δαπανηρών τομεακών ερευνών. Εξετάζοντας την

⁷⁶Βλ. W. J. Orlikowski & C.D.Gash, Technological Frames: Making Sense Of Information Technology In Organizations, ACM Transactions On Information Systems, Τεύχος12, Αρ. 2, 1994, σελ. 174-207.

⁷⁷Βλ. N. Bostrom, A. Dafoe & C. Flynn, όπ.παρ., σελ. 24.

⁷⁸Βλ. N. Bostrom, A. Dafoe & C. Flynn, όπ.παρ., σελ. 19.

⁷⁹Βλ. U. Hilpert, State Policies And Techno-Industrial Innovation, London, Routledge, 1991, σελ. 25.

⁸⁰Βλ. U. Hilpert, όπ.παρ., σελ. 26.

κατανόηση της τεχνολογίας σε διεθνές επίπεδο, οι Kuhlman και Edler⁸¹ αναφέρονται στη χρήση διακρατικών πολιτικών, όπως συμβαίνει σε ευρωπαϊκό επίπεδο, μέσω της οποίας υποστηρίζεται η οικονομική και κοινωνική συνοχή των ευρωπαϊκών προγραμμάτων και πολιτικών στο ζήτημα της τεχνολογικής ενσωμάτωσης και δη της τεχνητής νοημοσύνης στην κοινωνία των πολιτών.

⁸¹Βλ. S. Kuhlmann&J. Edler, Scenarios Of Technology And Innovation Policies In Europe: Investigating Future Governance, Technological Forecasting And Social Change, Τεύχος70, Αρ. 7,2003, σελ. 619-637.

1.6 Σύνοψη – Συμπεράσματα

Στο παρόν κεφάλαιο αποσαφηνίστηκε η τεχνητή νοημοσύνη ως η ικανότητα των μηχανών να προβαίνουν σε εκτέλεση καθηκόντων επιδεικνύοντας ευφυή ανθρώπινη συμπεριφορά και ως η ικανότητα των μηχανών να συμπεριφέρονται ως έξυπνες οντότητες, χαρακτηριζόμενες από στοιχειώδη αντίληψη του περιβάλλοντός τους.

Αν και ιστορικά, η τεχνητή νοημοσύνη τέθηκε για πρώτη φορά στο προσκήνιο, τουλάχιστον ιδεολογικά και όχι πρακτικά, το 1920, ωστόσο περισσότερο έντονη ήταν η ουσιαστική της παρουσία κατά τη διάρκεια της βιομηχανικής επανάστασης, όπου από την πλευρά των μηχανών προσφέρονταν ποικίλοι τρόποι υποβοήθησης, ενίσχυσης και υποκατάστασης της ανθρώπινης χειρωνακτικής εργασίας. Με αυτόν τον τρόπο προσεγγίζεται η αξιοποίηση της τεχνητής νοημοσύνης και στα μετέπειτα χρόνια, ακόμη και κατόπιν της πρόσφατης ψηφιακής επανάστασης. Επί της ουσίας, απώτερος σκοπός είναι η αξιοποίηση της τεχνητής νοημοσύνης για λόγους εξυπηρέτησης και διευκόλυνσης της ανθρώπινης εργασίας.

Η τεχνητή νοημοσύνη επιμερίζεται σε αδύναμη και ισχυρή. Τα παρόντα συστήματα τεχνητής νοημοσύνης εντάσσονται στο πλαίσιο της αδύναμης τεχνητής νοημοσύνης, καθώς σε περίπτωση κατά την οποία τεθεί σε εφαρμογή και αξιοποίηση η ισχυρή τεχνητή νοημοσύνη, τότε θα ομιλούμε περί απόλυτης αντικατάστασης της ανθρώπινης εργασίας. Η απόλυτη αυτή αντικατάσταση συνεπάγεται και την πλήρη απουσία της κριτικής και αφαιρετικής ικανότητας των ανθρώπων.

Μέσω της χρήσης συστημάτων τεχνητής νοημοσύνης, ιδιαίτερη προσοχή πρέπει να δίνεται στα λεγόμενα βρώμικα δεδομένα, από τα οποία αναδύονται ζητήματα ασφάλειας. Σε αυτό το σημείο, κομβικό ρόλο διαδραματίζουν τα υποστηριζόμενα από την πρόταση της Ευρωπαϊκής Επιτροπής για τη χρήση συστημάτων τεχνητής νοημοσύνης και την επικινδυνότητα αυτής, όταν ενσωματώνεται σε σύγχρονες κοινωνικές, πολιτικές και οικονομικές διαστάσεις και κυρίως σε ζητήματα λήψης αποφάσεων επί αυτών των διαστάσεων.

Το συγκεκριμένο ζήτημα αναλύεται διεξοδικότερα στο επόμενο κεφάλαιο, ωστόσο, λαμβάνοντας υπόψη το ζήτημα των βρώμικων δεδομένων που τέθηκε στην παρούσα ενότητα, επισημαίνεται ότι η διασφάλιση της ύπαρξης επαρκούς ανθρώπινης εποπτείας στο στάδιο του σχεδιασμού και της κατασκευής των συστημάτων τεχνητής

νοημοσύνης, μπορεί να αποτρέψει την καταχώριση βρώμικων δεδομένων που δύνανται να οδηγήσουν στην καταπάτηση των ανθρωπίνων δικαιωμάτων και να εγείρουν ακόμα και ζητήματα κοινωνικής προστασίας και ασφάλειας των πολιτών.

Η ποιότητα των δεδομένων που καταχωρίζονται στο κατασκευαστικό και σχεδιαστικό στάδιο των μηχανών τεχνητής νοημοσύνης διαδραματίζει καθοριστικό ρόλο στην εξάλειψη των κινδύνων που συνοδεύονται από την ύπαρξη βρώμικων δεδομένων. Σε κάθε περίπτωση, η επιτυχής και ασφαλής ενσωμάτωση της χρήσης συστημάτων τεχνητής νοημοσύνης, ακόμα και σε διαδικασίες λήψης πολιτικών αποφάσεων, επέρχεται μόνο κατόπιν προηγούμενης απόλυτης κατανόησης της τεχνολογίας.

Κεφάλαιο 2 Νομοθετικά Ζητήματα

2.1 Μορφές δικαιοϋκόν ρυθμίσεων της τεχνητής νοημοσύνης

Η τεχνητή νοημοσύνη αποτελεί ένα φαινόμενο, το οποίο έχει χαρακτηριστικά που δύσκολα μπορούν να αντιμετωπιστούν κατά τρόπο ενιαίο από πλευράς δικαίου.

Στο επικρατές προσκίνητο, η Δύση, ήτοι Αμερική και Ευρώπη, θα προσπαθήσει να βρει κανόνες δεοντολογίας, υπό μία μορφή *soft law*. Ήδη, η Δύση έχει αρχίσει να επεξεργάζεται τρόπους με τους οποίους θα περιορίσει την εξάπλωση της τεχνητής νοημοσύνης, κάτι που κατά την άποψή μας, δεν θα φέρει αποτέλεσμα, διότι ο ρους της ιστορίας και της τεχνολογίας έχει αποδείξει ότι η εν λόγω τεχνολογική εξέλιξη δεν δύναται να αποτραπεί. Εν προκειμένω, η τεχνητή νοημοσύνη αντιμετωπίζεται τελείως διαφορετικά στο σύγχρονο κόσμο, λόγω της υπόστασης διαφορετικών κοινωνικών και διεθνών συστημάτων⁸².

Είναι δύσκολο να υπάρξει ένα παγκοσμιοποιημένο σύστημα ελέγχου της τεχνητής νοημοσύνης, λόγω του ότι η Κίνα, που αυτή τη στιγμή είναι η πρωτοπόρος στην τεχνητή νοημοσύνη, δεν θα θελήσει να θέσει σε ισχύ κανόνες δεοντολογίας. Άρα, καταλήγουμε στην ύπαρξη ενός περιβάλλοντος, με την Κίνα στην πρωτοκαθεδρία αυτής της καινοτόμου τεχνολογίας, το οποίο διαφοροποιείται τελείως από αυτό που μπορεί να συμβαίνει στη Δύση.

Η τεχνητή νοημοσύνη περιλαμβάνει εφαρμογές οι οποίες αφορούν πολλές βιομηχανίες, κυβερνητικές υπηρεσίες και ομάδες συμφερόντων, δημιουργώντας μία συντονισμένη ρυθμιστική δυσκολία αντιμετώπισης και νομοθετικής οριοθέτησης της εξέλιξής της. Επιπλέον, ο ρυθμός ανάπτυξης της τεχνητής νοημοσύνης υπερβαίνει κατά πολύ την ικανότητα του παραδοσιακού ρυθμιστικού συστήματος να συμβαδίζει

⁸²Βλ. D. Castro, M. McLaughlin & E. Chivot, Who Is Winning the AI Race: China, the EU or the United States, 2019, στο <https://www.datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/>

με αυτήν. Πρόκειται για μία πρόκληση που είναι διεθνώς γνωστή με τον όρο «pacing problem» και επηρεάζει πολλές αναδυόμενες τεχνολογίες⁸³.

Από την πλευρά των εθνικών κυβερνήσεων υπάρχει απροθυμία παρεμπόδισης των καινοτόμων αναδυόμενων τεχνολογιών, λόγω των αυξημένων επιπέδων ανταγωνισμού σε διεθνές βεληνεκές⁸⁴. Ωστόσο, μία προσπάθεια οριοθέτησης της χρήσης της τεχνητής νοημοσύνης πραγματοποιήθηκε από την θέσπιση του Γενικού Κανονισμού για την Προστασία Δεδομένων (ΓΚΠΔ). Μάλιστα, αξιοσημείωτη είναι η απάντηση που δόθηκε από την Jouroνα, Πρόεδρο της Ευρωπαϊκής Επιτροπής, επί του ζητήματος της τεχνητής νοημοσύνης, όπου χαρακτηριστικά αναφέρθηκε ότι «Οι εφαρμογές τεχνητής νοημοσύνης πρέπει να συμμορφώνονται με τα θεμελιώδη δικαιώματα, συμπεριλαμβανομένου του δικαιώματος στην προστασία των δεδομένων προσωπικού χαρακτήρα. Στην ανθρωποκεντρική της προσέγγιση, η Επιτροπή σκοπεύει ταυτόχρονα σε δύο στόχους: αφενός την αφομοίωση και ανάπτυξη της τεχνολογίας και αφετέρου την αποτελεσματική προστασία των πολιτών της ΕΕ. Οι στόχοι αυτοί είναι αλληλένδετοι: για την προώθηση της αφομοίωσης και της καινοτομίας, η Επιτροπή πρέπει να διασφαλίσει ότι οι χρήστες — οι πολίτες, οι καταναλωτές, αλλά και οι επιχειρήσεις και οι δημόσιες αρχές που χρησιμοποιούν την τεχνητή νοημοσύνη — έχουν εμπιστοσύνη στην τεχνολογία και ότι δικαιολογημένα έχουν αυτή την εμπιστοσύνη. Όταν οι αιτήσεις αφορούν την επεξεργασία δεδομένων προσωπικού χαρακτήρα, ισχύει ο Γενικός Κανονισμός για την Προστασία Δεδομένων (ΓΚΠΔ). Ο Γενικός Κανονισμός για την Προστασία Δεδομένων είναι τεχνολογικά ουδέτερος και, ως εκ τούτου, δεν αφορά ειδικά την τεχνητή νοημοσύνη. Ωστόσο, είναι συναφής και εφαρμόζεται σε πολλές εφαρμογές της τεχνητής νοημοσύνης. Οι κανόνες του Γενικού Κανονισμού για την Προστασία Δεδομένων, συμπεριλαμβανομένου του άρθρου 22, δεν παρεμποδίζουν την καινοτομία, δεδομένου ότι δεν απαγορεύουν τη γενική αυτοματοποιημένη λήψη αποφάσεων, αλλά επιτρέπουν την επεξεργασία αυτή υπό συγκεκριμένες προϋποθέσεις και διασφαλίσεις. Κάθε εφαρμογή τεχνητής νοημοσύνης που περιλαμβάνει δεδομένα προσωπικού χαρακτήρα πρέπει να

⁸³Βλ. G.E. Marchant, The Growing Gap between Emerging Technologies and the Law, Στο Gary E. Marchant et al.(Επιμ) The Growing Gap between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem, Springer, UK 2011, σελ 19-33.

⁸⁴Βλ. W. Wallach &G. Marchant, An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics, 2018, στο https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_77.pdf

είναι σύμφωνη με τις αρχές προστασίας των δεδομένων, συμπεριλαμβανομένης της νομιμότητας και της διαφάνειας της επεξεργασίας, του περιορισμού στον σκοπό και του περιορισμού της αποθήκευσης. Εναπόκειται στην επόμενη Επιτροπή να καθορίσει το περιεχόμενο της προσέγγισης της ΕΕ στην τεχνητή νοημοσύνη»⁸⁵.

Ο Γενικός Κανονισμός για την Προστασία Δεδομένων αποτελεί την πρώτη επίσημη απόδειξη ότι η τεχνητή νοημοσύνη χρήζει συμπερίληψης σε επίσημες νομοθετικές ρυθμίσεις, προκειμένου να οριοθετηθεί το πλαίσιο ανάπτυξης και εξέλιξής της. Πάντα όμως, θα εγείρονται ζητήματα του ποιες είναι οι πτυχές της τεχνητής νοημοσύνης που πρέπει να ρυθμιστούν -οριοθετηθούν- νομοθετικά, ποιος είναι ο τρόπος με τον οποίο πρέπει να νομοθετηθούν και από ποιον.

Δεδομένης της, μέχρι προ θέσπισης του Γενικού Κανονισμού για την Προστασία Δεδομένων, απουσίας επίσημων νομοθετικών και δικαϊκών ρυθμίσεων για το ζήτημα της τεχνητής νοημοσύνης, το «κενό διακυβέρνησης» καλυπτόταν κυρίως από τους λεγόμενους «soft laws», οι οποίοι εξακολουθούν να υφίστανται. Επί της ουσίας, πρόκειται για μηχανισμούς «μη δεσμευτικού δικαίου» που περιλαμβάνουν διάφορους τύπους μέσων, οι οποίοι καθιερώνουν βασικές προσδοκίες, που η άμεση κυβερνητική εκτέλεση των οποίων δεν είναι εφικτή. Επίσης, αντικατοπτρίζουν διαφορετικές προσεγγίσεις, όπως για παράδειγμα βέλτιστες πρακτικές, προτεινόμενες κατευθύνσεις, πρότυπα και κώδικες δεοντολογίας. Ορισμένες τέτοιες προσεγγίσεις μέσω softlaws έχουν ήδη προταθεί ή εφαρμόζονται για την τεχνητή νοημοσύνη. Ενώ οι softlaws διακατέχονται από ορισμένες σοβαρές ελλείψεις, όπως είναι η έλλειψη του υποχρεωτικού χαρακτήρα της εκτέλεσης και εφαρμογής τους, ωστόσο περιλαμβάνουν στρατηγικές, μέσω των οποίων μπορεί να βοηθηθεί η αποτελεσματική διαχείριση του τρόπου με τον οποίο εξελίσσεται και αναπτύσσεται ο τομέας της τεχνητής νοημοσύνης⁸⁶.

Ήδη, μέχρι σήμερα έχει διαφανεί η ταχεία έγχυση πρωτοβουλιών και προτάσεων τύπου soft law στο πεδίο της τεχνητής νοημοσύνης. Μία από τις αρχικές προσπάθειες soft law

⁸⁵Βλ. Ευρωπαϊκό Κοινοβούλιο, Κοινοβουλευτικές Ερωτήσεις, 2019, στο https://www.europarl.europa.eu/doceo/document/E-9-2019-002411-ASW_EL.html

⁸⁶Βλ. O.A. Stefan, M. Avbelj, M. Eliantonio, M. Hartlapp, E. Korkea-aho & N. Rubio, EU Soft Law in the EU Legal Order: A Literature Review, 2019, King's College London Law School Research Paper Forthcoming, at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3346629

για την τεχνητή νοημοσύνη ήταν ο «χάρτης δεοντολογίας των ρομπότ» (Robot Ethics Charter) τον οποίο είχε ορίσει η κυβέρνηση της Νότιας Κορέας το 2007, παρά το ότι εντέλει καμία ολοκληρωμένη έκδοση του εν λόγω χάρτη δεοντολογίας δεν δημοσιεύτηκε διαδικτυακά⁸⁷.

Ίσως η πιο ολοκληρωμένη πρωτοβουλία softlaw, η οποία αφορά το ζήτημα της τεχνητής νοημοσύνης, ξεκίνησε το έτος 2016 από το Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (Institute of Electric and Electronic Engineers - IEEE)⁸⁸. Σκοπός της συγκεκριμένης προσπάθειας είναι η προώθηση δημοσίων συζητήσεων για τον τρόπο δημιουργίας κοινωνικά ηθικών εφαρμογών, οι οποίες αναφέρονται σε περιπτώσεις «έξυπνων» τεχνολογιών, επιτυγχάνοντας ευθυγράμμιση της χρήσης τους με συγκεκριμένους κώδικες δεοντολογίας και αξίες, προσβλέποντας στην ανθρώπινη ευημερία⁸⁹.

Μία επιπρόσθετη πρωτοβουλία τύπου soft law εντοπίζεται στη δημιουργία μίας σειράς προτύπων IEEE, αναφορικά με τη διακυβέρνηση και τις ηθικές πτυχές της τεχνητής νοημοσύνης. Το IEEE έχει δώσει την επίσημη έγκριση για τη δημιουργία των ακόλουθων προτύπων, με τις επιτροπές καθορισμού προτύπων που έχουν δημιουργηθεί για την ανάπτυξη καθενός εξ αυτών. Τα πρότυπα αυτά είναι τα κάτωθι:

IEEE P7000TM - Πρότυπη διαδικασία για την αντιμετώπιση των δεοντολογικών ανησυχιών κατά το σχεδιασμό του συστήματος

IEEE P7001TM - Διαφάνεια των αυτόνομων συστημάτων

IEEE P7002TM - Διαδικασία προστασίας προσωπικών δεδομένων

IEEE P7003TM - Εξετάσεις αλγοριθμικών παραμέτρων

IEEE P7004TM - Πρότυπο για τη διαχείριση των μαθητών και των φοιτητών

⁸⁷Βλ. New Scientist, South Korea creates ethical code for righteous robots, 2007, στο <https://www.newscientist.com/article/dn11334-south-korea-creates-ethical-code-for-righteous-robots/#ixzz6FcZVDBtt>

⁸⁸Βλ. IEEE, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2002^a, στο https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

⁸⁹Βλ. IEEE, Ethically Aligned Design – Version II Request for Input, 2020c, στο https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf

IEEE P7005TM - Πρότυπο για τη διαφάνεια της διαχείρισης των δεδομένων των εργοδοτών

IEEE P7006TM - Πρότυπο για την Τεχνητή Νοημοσύνη Προσωπικών Δεδομένων

IEEE P7007TM - Δεοντολογικό Πρότυπο για Ρομποτική και Συστήματα Αυτοματισμού σε συνδυασμό με την Ηθική

IEEE P7008TM - Πρότυπο για τον ηθικό χειρισμό ρομποτικών και ευφύων μηχανών, και συστημάτων αυτοματισμού

IEEE P7009TM - Πρότυπο για ασφαλή σχεδιασμό αυτόνομων και ημιαυτόνομων συστημάτων

IEEE P7010TM - Πρότυπο μετρήσεων ευημερίας για την ηθική τεχνητή νοημοσύνη και τα αυτόνομα συστήματα

IEEE P7011TM - Πρότυπο για τη διαδικασία εντοπισμού και αξιολόγησης της αξιοπιστίας των ειδησεογραφικών πηγών

IEEE P7012TM - Πρότυπο για μη αυτόματους όρους προστασίας προσωπικών δεδομένων

IEEE P7013TM - Πρότυπα ενσωμάτωσης και εφαρμογής για την αυτοματοποιημένη τεχνολογία ανάλυσης προσώπου

Αυτά τα δεκατέσσερα πρότυπα που είναι συναρτημένα με την τεχνητή νοημοσύνη είχαν προγραμματιστεί να οριστικοποιηθούν μέχρι το τέλος του 2021 και θα παράσχουν ένα ευρύ σύνολο απαιτήσεων σχετικών με τον χειρισμό και την αντιμετώπιση των εξελίξεων που διαδραματίζονται στο πεδίο αυτό⁹⁰.

Το 2017, το Ινστιτούτο «Future of Life» συγκάλεσε συνάντηση πολλών κορυφαίων επαγγελματιών του πεδίου της τεχνητής νοημοσύνης και εμπειρογνομόνων στο συνεδριακό κέντρο Asilomar⁹¹. Στη διάσκεψη αυτή, οι συμμετέχοντες συμφώνησαν σε

⁹⁰Βλ. IEEE, IEEE Global Initiative for Ethical Considerations in Artificial Intelligence (AI) and Autonomous Systems (AS) Drives, together with IEEE Societies, New Standards Projects, Releases New Report on Prioritizing Human Well-Being, 2017, στο https://standards.ieee.org/news/2017/ieee_p7004.html

⁹¹Βλ. Future of Life Institute, Beneficial AI 2017, 2020, στο: <https://futureoflife.org/bai-2017/?cn-reloaded=1>

23 αρχές που διέπουν την έρευνα και τις εφαρμογές της τεχνητής νοημοσύνης. Αυτές οι αρχές, μεταξύ άλλων, περιλαμβάνουν τη λεγόμενη «Διαφάνεια των αποτυχιών» (Failure Transparency), όπου εάν ένα σύστημα τεχνητής νοημοσύνης προκαλέσει οποιασδήποτε μορφής βλάβη, πρέπει να είναι δυνατόν να διαπιστωθεί γιατί προκλήθηκε η βλάβη αυτή. Μία ακόμη αρχή είναι αυτή της «Ευθύνης» (Responsibility), κατά την οποία οι σχεδιαστές και οι κατασκευαστές προηγμένων συστημάτων τεχνητής νοημοσύνης είναι άμεσα εμπλεκόμενοι στις ηθικές συνέπειες της χρήσης, της κατάχρησης και των ενεργειών τους και έχουν την ευθύνη να διαμορφώσουν αυτές τις επιπτώσεις. Η τρίτη αρχή αφορά την «Προσδιορίσιμη Αξία» (Value Alignment), όπου τα πληροφοριακά συστήματα πρέπει να σχεδιάζονται με τέτοιο τρόπο, έτσι ώστε να διασφαλίζεται η ευθυγράμμιση της χρήσης τους με το σύνολο των ανθρώπινων αξιών⁹².

Από τον βιομηχανικό κλάδο, έχουν υιοθετηθεί εξειδικευμένοι soft laws για το ζήτημα της τεχνητής νοημοσύνης. Για παράδειγμα, από το Συμβούλιο Βιομηχανίας Πληροφορικής (Information Technology Industry Council - ITI) έχει αναπτυχθεί ένα σύνολο αρχών που αφορά αμιγώς στον συγκεκριμένο τομέα⁹³. Από τις κυβερνήσεις, έχει επίσης υποστηριχθεί η αξιοποίηση των soft laws, επί διαχειριστικών ζητημάτων στον κλάδο της τεχνητής νοημοσύνης.

Στις 25 Απριλίου του έτους 2018, ορίστηκαν συνολικά επτά αρχές από την Ευρωπαϊκή Επιτροπή και ακολούθησε η δημοσίευση σχετικού εγγράφου για τη στρατηγική διαχείριση της τεχνητής νοημοσύνης⁹⁴.

Μέχρι και το 2018, από την Ευρωπαϊκή Επιτροπή δεν είχαν προταθεί νέα διαχειριστικά μέτρα για τον κλάδο της τεχνητής νοημοσύνης⁹⁵. Αντίθετα, σχεδιάστηκαν σχετικές

⁹²Βλ. Future of Life, Asilomar AI Principles, 2020, <https://futureoflife.org/ai-principles/?cn-reloaded=1>

⁹³Βλ. ITI, ITI Unveils First Industry-Wide Artificial Intelligence Policy Principles, 2017, στο <https://www.itic.org/news-events/news-releases/iti-unveils-first-industry-wide-artificial-intelligence-policy-principles>

⁹⁴Βλ. European Commission, Communication Artificial Intelligence for Europe, 2018^a, στο <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe>

⁹⁵Βλ. European Parliament, European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), 2017, στο https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html

κατευθυντήριες γραμμές. Χαρακτηριστικό παράδειγμα είναι η δημοσίευση του «Σχεδίου συντονισμένης δράσης για την τεχνητή νοημοσύνη» (Coordinated Action Plan on Artificial Intelligence) τον Δεκέμβριο του 2018, στο οποίο εκτίθενται οι στόχοι και τα σχέδιά της για μια πανευρωπαϊκή στρατηγική συναρτημένη με την τεχνητή νοημοσύνη. Εντούτοις, η Ευρωπαϊκή Επιτροπή τόνισε ότι ενώ η αυτορρύθμιση μπορεί να παράσχει ένα πρώτο σύνολο σημείων αναφοράς, βάσει των οποίων μπορούν να αξιολογηθούν οι αναδυόμενες εφαρμογές και τα αποτελέσματα αυτών, οι δημόσιες αρχές πρέπει, από τη δική τους πλευρά, να διασφαλίσουν ότι τα κανονιστικά πλαίσια για την ανάπτυξη και τη χρήση τεχνολογιών τεχνητής νοημοσύνης ευθυγραμμίζονται με τις αξίες και τα θεμελιώδη δικαιώματα⁹⁶. Το 2021 δημοσιεύτηκε η πιο πρόσφατη πρόταση της Ευρωπαϊκής Επιτροπής. Χαρακτηριστικά αναφέρεται ότι «η παρούσα πρόταση υλοποιεί την πολιτική δέσμευση της προέδρου von der Leyen, η οποία ανακοίνωσε τις πολιτικές κατευθύνσεις της για την Επιτροπή 2019-2024...ότι η Επιτροπή θα προτείνει νομοθεσία για μια συντονισμένη ευρωπαϊκή προσέγγιση όσον αφορά τις επιπτώσεις της ΤΝ σε επίπεδο ανθρώπων και ηθικής»⁹⁷.

Νωρίτερα από την προαναφερθείσα έκθεση της Ευρωπαϊκής Επιτροπής, προκειμένου να καλυφθεί το ενδιάμεσο στάδιο μέχρι την υπόσταση επίσημης νομοθετικής ρύθμισης, τον Απρίλιο του 2018, στο Ηνωμένο Βασίλειο το UK House of Lords δημοσίευσε μια λεπτομερή έκθεση για την τεχνητή νοημοσύνη σε επίπεδο soft law⁹⁸. Η εν λόγω έκθεση συνιστά την εφαρμογή ενός διακλαδικού δεοντολογικού κώδικα ή ενός «κώδικα τεχνητής νοημοσύνης», κατάλληλου για εφαρμογή σε οργανισμούς του δημόσιου και ιδιωτικού τομέα, που αξιοποιούν πτυχές της τεχνητής νοημοσύνης. Ένας τέτοιος κώδικας θα πρέπει να περιλαμβάνει την ανάγκη να εξεταστεί η δημιουργία συμβουλευτικών επιτροπών δεοντολογίας σε εταιρείες ή οργανισμούς, που

⁹⁶Βλ. European Commission, Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Coordinated Plan on Artificial Intelligence COM/2018/795 final, 2018b, στο <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:795:FIN>

⁹⁷Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

⁹⁸Βλ. UK Parliament, Artificial Intelligence Committee - AI in the UK: ready, willing and able, 2017, στο <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm>

αναπτύσσουν ή χρησιμοποιούν την τεχνητή νοημοσύνη στην καθημερινή τους λειτουργία⁹⁹.

Η έκθεση αυτή, μάλιστα, έκανε μνεία σε ζητήματα σχετικά με τις πιθανές επιζήμιες συνέπειες της πρόωρης ρύθμισης, όπως ότι ο ρυθμός της τεχνολογικής εξέλιξης είναι δύσκολο να ακολουθηθεί από την κανονιστική ή ειδική νομοθεσία, σε σημείο που αν νομοθετηθεί κάτι επισήμως, μπορεί κατά τη στιγμή της έναρξης ισχύος του να είναι πλέον ξεπερασμένο. Αναφέρεται, επίσης, ότι τα διδάγματα από τη νομοθετική ρύθμιση ζητημάτων που αφορούσαν παλαιότερες τεχνολογικές εξελίξεις καθιστούν εναργές ότι μία αυστηρή και νομοθετικά λεπτομερής προσέγγιση των προκυπτουσών απαιτήσεων δεν είναι χρήσιμη. Κατόπιν των ανωτέρω, λοιπόν, το UK House of Lords κατέληξε στο συμπέρασμα ότι η ειδική νομοθετική ρύθμιση ζητημάτων συνυφασμένων με την τεχνητή νοημοσύνη, επί του παρόντος, είναι ακατάλληλη. Εντούτοις, εν καιρώ, ο εν λόγω κώδικας τεχνητής νοημοσύνης θα μπορούσε να αποτελέσει τη βάση για τη θέσπιση μίας επίσημης κανονιστικής ρύθμισης, εάν και όταν αυτό κριθεί απαραίτητο¹⁰⁰.

⁹⁹Βλ. UK Parliament, Artificial Intelligence Committee - AI in the UK: ready, willing and able, 2017, στο <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm>

¹⁰⁰Βλ. UK Parliament, Artificial Intelligence Committee - AI in the UK: ready, willing and able, 2017, στο <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm>

2.2 Η Εφαρμογή των ‘Νορμών’ και τα Νομικά Ενδεχόμενα επί ζητημάτων Ηθικής

Η ικανότητα συγκέντρωσης εμπειριών και διδαγμάτων από τη μαζική επεξεργασία δεδομένων, σε συνδυασμό με την ικανότητα που έχουν οι μηχανές τεχνητής νοημοσύνης στο να προβαίνουν σε ανεξάρτητη και αυτόνομη λήψη αποφάσεων, αποτελούν τις βασικές προϋποθέσεις για τη νομική ευθύνη που απορρέει από την αξιοποίηση και λειτουργία των μηχανών αυτών. Όμως, λόγω του ότι η τεχνητή νοημοσύνη δεν θεωρείται νομικό αντικείμενο, καταδεικνύεται ότι δεν μπορεί να οριστεί εξ ολοκλήρου υπεύθυνη για ζημιές και βλάβες που μπορεί να δημιουργήσει¹⁰¹.

Υπό αυτήν την έννοια, όπως ορίζεται στις διατάξεις του άρθρου 12 της Σύμβασης των Ηνωμένων Εθνών περί ζητήματος χρήσης ηλεκτρονικών επικοινωνιών σε διεθνείς συμβάσεις (*United Nations Convention on the Use of Electronic Communications in International Contracts*¹⁰²), η ευθύνη για τη βλάβη που ενδεχομένως προκληθεί από τη χρήση της τεχνητής νοημοσύνης καταλογίζεται στο ίδιο το φυσικό ή νομικό πρόσωπο για το οποίο κατασκευάστηκε η μηχανή αυτή¹⁰³. Η εν λόγω διάταξη ερμηνεύεται στη βάση του ότι ένα μηχάνημα δεν είναι δυνατόν να έχει δική του αυτοβουλία, αλλά τα αποτελέσματά του προκύπτουν από την πρόθεση και θέληση των ιδιοκτητών του.

Στο σημείο αυτό, επισημαίνεται ότι από το Ευρωπαϊκό Κοινοβούλιο, στις 16 Φεβρουαρίου του 2017 εκδόθηκε ψήφισμα που περιελάμβανε ένα σύνολο συστάσεων της Ευρωπαϊκής Επιτροπής συνυφασμένων με το κανονιστικό πλαίσιο αστικού δικαίου για τον κλάδο της ρομποτικής. Το έγγραφο που εκδόθηκε από το Ευρωπαϊκό Κοινοβούλιο, με τίτλο «*Συστάσεις της Ευρωπαϊκής Επιτροπής για τους κανόνες αστικού δικαίου στη ρομποτική 2015/2103 – Recommendations from the European Commission*

¹⁰¹Βλ. T.C. King, N. Aggarwal, M. Taddeo & L. Floridi, Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions, Science and Engineering Ethics, Ap. 26, 2019, σελ. 89-120.

¹⁰²Βλ. A. Raymond, Electronic Commerce and the New Uncitral Convention, International Law Practicum, Ap. 19, Τεύχος 1, 2006, σελ. 66-74.

¹⁰³Βλ. P. Čerka, J. Grigienė & G. Sirbikytė, Liability for damages caused by artificial intelligence, Computer Law & Security Review, Ap. 31, Τεύχος 3, 2015, σελ. 376-389.

on Civil Law Rules in Robotics 2015/2103- INL»¹⁰⁴ προέβη στην πρόταση της ίδρυσης ενός ευρωπαϊκού οργανισμού με αντικείμενο ζητήματα τεχνητής νοημοσύνης και ρομποτικής, ώστε να διασφαλιστεί¹⁰⁵ η τήρηση βασικών τεχνικών και ηθικών κανόνων¹⁰⁶. Επίσης, επειδή υπάρχει αναντιστοιχία μεταξύ τεχνολογίας και της ηθικής για τη χρήση αυτής, στο πλαίσιο της παραπάνω πρότασης επισημαίνεται ότι η αξιοπρέπεια πρέπει να τίθεται στη βάση της διαμόρφωσης μίας νέας ψηφιακής ηθικής¹⁰⁷.

Άξια αναφοράς είναι η πρόσφατη πρόταση της Ευρωπαϊκής Επιτροπής¹⁰⁸ περί θέσπισης εναρμονισμένων κανόνων για την τεχνητή νοημοσύνη, σκοπός της οποίας είναι: (1) η διασφάλιση ότι τα κατασκευασμένα συστήματα τεχνητής νοημοσύνης, που είναι διαθέσιμα στην αγορά, συνοδεύονται από τα απαιτούμενα επίπεδα ασφαλείας και

¹⁰⁴Βλ. European Parliament, Motion for a European Parliament Resolution, 2017, στο: https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html, Βλ. Επίσης Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 255-258.

¹⁰⁵ Ο τύπος ασφάλισης που πρέπει να εφαρμόζεται στην περίπτωση των έξυπνων ρομπότ, καθώς και το ποια εμπλεκόμενα μέρη και οργανισμοί πρέπει να φέρουν αυτό το βάρος της ευθύνης, εξακολουθεί να είναι ένα ανοιχτό ζήτημα. Η πρόσφατη έκθεση της Ευρωπαϊκής Ένωσης (2015/2103 (INL)) εξέδωσε συστάσεις σχετικά με το εν λόγω θέμα, προτείνοντας όχι μόνο την υποχρεωτική εγγραφή, αλλά και τη δημιουργία ασφαλίσεων και αποζημιώσεων. Σύμφωνα με το Ευρωπαϊκό Κοινοβούλιο, η ασφάλιση θα μπορούσε να ληφθεί τόσο από τον καταναλωτή όσο και από την εταιρεία σε παρόμοιο καθεστώς και περιβάλλον συνθηκών με αυτό που χρησιμοποιείται από την ασφάλιση αυτοκινήτων. Το ταμείο θα μπορούσε να είναι είτε γενικό (για όλα τα αυτόνομα ρομπότ) είτε μεμονωμένο (για κάθε κατηγορία ρομπότ), αποτελούμενο από τέλη που καταβάλλονται κατά τη στιγμή της διάθεσης του μηχανήματος στην αγορά ή των εισφορών που καταβάλλονται περιοδικά καθ' όλη τη διάρκεια ζωής των ρομπότ. Αξίζει να σημειωθεί ότι, σε αυτήν την περίπτωση, οι εταιρείες θα είναι υπεύθυνες για την αντιμετώπιση αυτού του βάρους και της ευθύνης που φέρουν. Παρά την πρόταση αυτή, ωστόσο, το ζήτημα αυτό εξακολουθεί να βρίσκεται υπό συζήτηση, με νέες εναλλακτικές λύσεις και πιο ενδιαφέροντα μοντέλα - όπως επί παραδείγματι ιδιωτικά κεφάλαια, συγκεκριμένα αρχεία, και λοιπές παρεμφερείς προτάσεις αυτού του βεληνεκούς - τα οποία όμως δεν θα αποτελέσουν αντικείμενο μεγαλύτερης ανάλυσης σε αυτή τη διατριβή.

¹⁰⁶Βλ. European Parliament, όπ.παρ., στο: https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html

¹⁰⁷Βλ. C. Nebeker, J. Torous & R.J. Bartlett Ellis, Building the case for actionable ethics in digital health research supported by artificial intelligence, BMC Medicine, Αρ. 17, Τεύχος 1, 2019, σελ. 1-7, Βλ. επίσης I.V. Ponkin, & A.I. Redkina Artificial Intelligence from the Point of View of Law, Rudn Journal of Law, Αρ. 22, Τεύχος 1, 2018, σελ. 91-109.

¹⁰⁸ Βλ. Ευρωπαϊκή Επιτροπή, Πρόταση για τη θέσπιση εναρμονισμένων κανόνων σχετικά με την τεχνητή νοημοσύνη (πράξη για την τεχνητή νοημοσύνη) και για την τροποποίηση ορισμένων νομοθετικών πράξεων της Ένωσης, 2021, στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

η χρήση τους συνεπάγεται την τήρηση του ισχύοντος νομοθετικού πλαισίου, (2) η διασφάλιση της «ασφάλειας του δικαίου» για λόγους διευκόλυνσης των επενδυτικών κινήσεων και προώθηση της καινοτομίας στο πλαίσιο της τεχνητής νοημοσύνης, (3) η ενίσχυση της διακυβέρνησης και της αποτελεσματικότητας της επιβολής και θέσπισης της επικρατούσας νομοθεσίας, για την διασφάλιση και προάσπιση των θεμελιωδών δικαιωμάτων και των αναγκών ασφάλειας, τα οποία εφαρμόζονται στην περίπτωση των κατασκευασμένων συστημάτων τεχνητής νοημοσύνης και (4) η προώθηση της ανάπτυξης μίας ενιαίας αγοράς, στην οποία θα επικρατούν ασφαλή, αξιόπιστα και νόμιμα συστήματα τεχνητής νοημοσύνης, ενώ παράλληλα θα είναι εφικτή η πρόληψη του κινδύνου να κατακερματιστεί η αγορά¹⁰⁹.

Στο πλαίσιο της συγκεκριμένης πρότασης¹¹⁰, τα συστήματα που κρίνονται ως υψηλού κινδύνου για την εφαρμογή συστημάτων τεχνητής νοημοσύνης είναι αυτά που συνδέονται με την «βιομετρική ταυτοποίηση και κατηγοριοποίηση των φυσικών προσώπων»¹¹¹, την «διαχείριση και λειτουργία υποδομών ζωτικής σημασίας»¹¹², την «εκπαίδευση και επαγγελματική κατάρτιση»¹¹³, την «απασχόληση και διαχείριση των εργαζομένων», καθώς επίσης και την πρόσβαση αυτών στο καθεστώς αυτοαπασχόλησης¹¹⁴. Επιπροσθέτως, σύμφωνα με την συγκεκριμένη πρόταση, ως συστήματα τεχνητής νοημοσύνης υψηλού κινδύνου είναι αυτά που υπάγονται στον

¹⁰⁹ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>, βλ. επίσης Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 119-120.

¹¹⁰ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

¹¹¹ Πρόκειται για συστήματα τεχνητής νοημοσύνης τα οποία προορίζονται προς χρήση για την εξ αποστάσεως βιομετρική ταυτοποίηση στα φυσικά πρόσωπα, η οποία πραγματοποιείται τόσο σε πραγματικό όσο και σε μεταγενέστερο χρόνο.

¹¹² Πρόκειται για συστήματα τεχνητής νοημοσύνης, τα οποία προορίζονται για χρήση ως «κατασκευαστικά στοιχεία ασφαλείας», τα οποία θα συντρέχουν αφενός στη λειτουργία και αφετέρου στη διαχείριση της οδικής κυκλοφορίας, καθώς επίσης και σε ζητήματα παροχής θέρμανσης, νερού, φυσικού αερίου, ηλεκτρικής ενέργειας κ.λπ., βλ. επίσης Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 121.

¹¹³ Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 121-122.

¹¹⁴ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>, βλ. επίσης Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 122.

τομέα της πρόσβασης των πολιτών σε βασικές υπηρεσίες και παροχές, οι οποίες προσφέρονται είτε από τον ιδιωτικό είτε από τον δημόσιο τομέα¹¹⁵, καθώς επίσης και τα συστήματα που είναι συνυφασμένα με την απονομή της Δικαιοσύνης και την τήρηση της Δημοκρατίας¹¹⁶. Συστήματα τεχνητής νοημοσύνης υψηλού κινδύνου, επιπλέον, είναι αυτά που αναφέρονται σε θέματα επιβολής των υφιστάμενων νομοθετικών πλαισίων¹¹⁷, αλλά και αυτά που προορίζονται για ζητήματα διαχείρισης μεταναστευτικών θεμάτων, διασυνοριακών ελεγκτικών διαδικασιών και ασύλου¹¹⁸.

Στην ίδια πρόταση της Ευρωπαϊκής Επιτροπής¹¹⁹, για λόγους διασφάλισης των χαμηλών επιπέδων επικινδυνότητας και για λόγους διασφάλισης της χρηστής και ηθικής αξιοποίησης των νέων παραγόμενων συστημάτων τεχνητής νοημοσύνης, προτείνεται η κατάρτιση πλήρους τεχνικού φακέλου, στον οποίο θα περιλαμβάνεται η γενική περιγραφή του κατασκευασμένου συστήματος τεχνητής νοημοσύνης¹²⁰, η λεπτομερής περιγραφή των επιμέρους στοιχείων από τα οποία αυτό απαρτίζεται¹²¹, η

¹¹⁵Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 122.

¹¹⁶ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>, βλ. επίσης Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 122.

¹¹⁷Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 122.

¹¹⁸ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>, βλ. επίσης Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 122.

¹¹⁹ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

¹²⁰ Στην γενική περιγραφή του κατασκευασμένου συστήματος τεχνητής νοημοσύνης πρέπει οπωσδήποτε να περιλαμβάνεται ο σκοπός που επιδιώκει μέσα από την λειτουργία του, τα υπεύθυνα άτομα για την ανάπτυξη του, ο τρόπος με τον οποίο είτε λειτουργεί είτε αλληλοεπιδρά με άλλα λογισμικά, τα οποία δεν είναι αναπόσπαστα μέρη του, η/οι επικαιροποιημένη/ες έκδοση/εις του, οι μορφές με τις οποίες είναι δυνατόν να διατεθεί στην αγορά, τα λογισμικά και υλικά στα/με τα οποία πρόκειται να λειτουργήσει, καθώς επίσης και φωτογραφικό υλικό που αποδεικνύει με λεπτομέρεια τα εξωτερικά χαρακτηριστικά του, αλλά και την εσωτερική του διαρρύθμιση.

¹²¹ Πρέπει υποχρεωτικά να περιλαμβάνονται: το σύνολο των σταδίων και των μεθόδων που ακολουθήθηκαν για την ανάπτυξη και κατασκευή του συστήματος τεχνητής νοημοσύνης, καθώς επίσης και οι σχεδιαστικές προδιαγραφές του, η περιγραφή της γενικής αρχιτεκτονικής του, τα απαιτούμενα μέτρα της ανθρώπινης εποπτείας κατά την κατασκευή και λειτουργία του, οι προκαθορισμένες αλλαγές για την βελτίωση των επιδόσεων της λειτουργίας του εν λόγω συστήματος τεχνητής νοημοσύνης, καθώς επίσης και οι διαδικασίες δοκιμών και επικυρώσεων που εκτελέστηκαν, με σχετική μνηεία στους χρησιμοποιούμενους αξιολογικούς δείκτες.

απόδοση λεπτομερών πληροφοριών αναφορικά με την λειτουργία, τον έλεγχο και την παρακολούθηση του κατασκευασμένου συστήματος τεχνητής νοημοσύνης, η λεπτομερής περιγραφή του συστήματος διαχείρισης κινδύνου που περιλαμβάνει, καθώς επίσης και οποιαδήποτε μεταβολή πρόκειται να δεχτεί το εν λόγω σύστημα τεχνητής νοημοσύνης, καθ' όλη τη διάρκεια του κύκλου ζωής του¹²².

Πέραν των ανωτέρω, ένα νέο κατασκευασμένο σύστημα τεχνητής νοημοσύνης, στον τεχνικό του φάκελο, πρέπει να περιλαμβάνει πλήρη κατάλογο με αναφορά στα εναρμονισμένα πρότυπα τα οποία έχουν εφαρμοστεί απόλυτα η μερικώς, και στοιχεία αναφοράς των οποίων έχουν δημοσιευθεί σε «*Επίσημη Εφημερίδα της Ευρωπαϊκής Ένωσης*»¹²³. Συν τοις άλλοις, ο τεχνικός φάκελος του κατασκευασμένου συστήματος τεχνητής νοημοσύνης πρέπει να περιλαμβάνει αντίγραφο της «*δήλωσης συμμόρφωσης της Ευρωπαϊκής Επιτροπής*», καθώς επίσης και περιγραφή του εφαρμοσμένου συστήματος για την αξιολογική διαδικασία των επιδόσεων της λειτουργίας του, κατόπιν της διάθεσής του στην αγορά¹²⁴.

Επισημαίνεται, ότι στην δήλωση συμμόρφωσης που προαναφέρθηκε¹²⁵ πρέπει υποχρεωτικά να γίνεται αναφορά αφενός την ονομασία και αφετέρου στον τύπο του κατασκευασμένου συστήματος τεχνητής νοημοσύνης, καθώς επίσης και στο σύνολο των επιμέρους χαρακτηριστικών στοιχείων που συμβάλλουν στην ταυτοποίηση και μοναδική του αναγνώριση. Επίσης, πρέπει να περιλαμβάνεται το ονοματεπώνυμο και η διεύθυνση του παρόχου του εν λόγω συστήματος τεχνητής νοημοσύνης, ενώ για την έκδοση της εν προκειμένω δήλωσης συμμόρφωσης, αποκλειστική ευθύνη φέρει ο πάροχος του συστήματος τεχνητής νοημοσύνης¹²⁶.

¹²² Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

¹²³ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

¹²⁴ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

¹²⁵ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

¹²⁶ Βλ. Ευρωπαϊκή Επιτροπή, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN>

Η απόδοση νομικής προσωπικότητας σε «έξυπνες» μηχανές είναι μία πιθανή λύση στο σύνολο των νομικής φύσεως προκλήσεων που ενδεχομένως αναδυθούν από την αυτονομία των εν λόγω μηχανών¹²⁷. Σύμφωνα με τη Γνωμοδότηση της Ευρωπαϊκής Οικονομικής και Κοινωνικής Επιτροπής (ΕΟΚΕ) με θέμα «*Η τεχνητή νοημοσύνη — Η επίδραση της τεχνητής νοημοσύνης στην (ψηφιακή) ενιαία αγορά, στην παραγωγή, στην κατανάλωση, στην απασχόληση και στην κοινωνία*», υπ' αριθμόν 2017/C288/01¹²⁸, χαρακτηριστικά επισημαίνεται ότι πρέπει οπωσδήποτε να ξεκινήσει ειδική αξιολόγηση της νομοθεσίας και των κανονιστικών ρυθμίσεων της Ευρωπαϊκής Ένωσης προκειμένου να επέλθει η αναθεώρηση ή η επικαιροποίηση των ζητημάτων που αναφέρονται στην απόδοση νομικής προσωπικότητας¹²⁹ σε μηχανές τεχνητής νοημοσύνης¹³⁰. Συγκεκριμένα, επισημαίνεται ότι από την πλευρά της ΕΟΚΕ υπάρχει διαφωνία στο να θεσπιστεί μορφή νομικής προσωπικότητας για τις μηχανές τεχνητής νοημοσύνης. Συνεχίζοντας, στην γνωμοδότηση αυτή χαρακτηριστικά αναφέρεται ότι δεδομένων των ανωτέρω «*Η προληπτική και οι διορθωτική επίδραση του δικαίου περί αστικής ευθύνης θα καθίστατο κενή περιεχομένου. Επιπλέον, κάτι τέτοιο θα δημιουργούσε δυνατότητες για καταχρήσεις, αλλά και ηθικούς κινδύνους, όσον αφορά την ανάπτυξη αλλά και την χρήση της τεχνητής νοημοσύνης*»¹³¹.

Ωστόσο, στην ίδια γνωμοδότηση επισημαίνεται ότι από την πλευρά του Ευρωπαϊκού Κοινοβουλίου έχουν διατυπωθεί συστάσεις αναφορικά με τους κανόνες του αστικού δικαίου ως προς το ζήτημα της ρομποτικής και ως προς το ενδεχόμενο δημιουργίας νομικής προσωπικότητας στις μηχανές τεχνητής νοημοσύνης, προκειμένου να αναλαμβάνονται από πλευράς τους ζητήματα αστικής ευθύνης για ζημιές που

¹²⁷Βλ. U. Pagallo, Robots of Just War: A Legal Perspective, Philosophy & Technology, Αρ. 24, Τεύχος 3, 2011, σελ. 307–323.

¹²⁸Βλ. ΕΟΚΕ, Γνωμοδότηση της Ευρωπαϊκής Οικονομικής και Κοινωνικής Επιτροπής (ΕΟΚΕ) με θέμα «*Η τεχνητή νοημοσύνη — Η επίδραση της τεχνητής νοημοσύνης στην (ψηφιακή) ενιαία αγορά, στην παραγωγή, στην κατανάλωση, στην απασχόληση και στην κοινωνία*», υπ' αριθμόν 2017/C288/01, 2017, στο <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52016IE5369&from=ET>

¹²⁹ Επισημαίνεται ότι η εν λόγω Πρόταση δεν έγινε τελικά αποδεκτή.

¹³⁰Βλ. ΕΟΚΕ, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52016IE5369&from=ET>

¹³¹Βλ. ΕΟΚΕ, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52016IE5369&from=ET>

ενδεχομένως προκληθούν από αυτές¹³². Με αυτή την τοποθέτηση διαφωνεί η ΕΟΚΕ, επιχειρηματολογώντας υπέρ του «απαράδεκτου ηθικού κινδύνου που ενέχει μία τέτοια προσέγγιση». Πιο συγκεκριμένα επισημαίνεται ότι σε περίπτωση κατά την οποία «υλοποιηθεί αυτή η πρόταση, απαλείφεται η προληπτική διορθωτική λειτουργία που απορρέει από το δίκαιο περί αστικής ευθύνης, εφόσον ο κατασκευαστής δεν θα αναλαμβάνει πλέον την ευθύνη, η οποία θα έχει μεταβιβαστεί στο ρομπότ»¹³³.

Μάλιστα, υπάρχει επιχειρηματολογία υπέρ του κινδύνου αθέμιτης χρήσης και κατάχρησης των συστημάτων τεχνητής νοημοσύνης, σε περίπτωση κατά την οποία τους αποδοθεί νομική προσωπικότητα. Σύμφωνα με την άποψη της ΕΟΚΕ, η τελική αστική ευθύνη πρέπει να αποδίδεται σε φυσικό πρόσωπο και όχι σε κάποια μηχανή, προκειμένου να αποφευχθεί ακόμα και ο κίνδυνος της κατάχρησης της δεύτερης¹³⁴. Ακολουθώντας, σκοπεύοντας στην εύρεση μίας τελικής λύσης στο συγκεκριμένο ζήτημα και να ακολουθηθεί μία συγκεκριμένη κατεύθυνση στο δίλημμα της απόδοσης νομικής προσωπικότητας στις μηχανές τεχνητής νοημοσύνης ή όχι, κρίνεται απαραίτητος ο ακριβής και λεπτομερής προσδιορισμός των ενεργειών που μπορεί να εκτελέσει ένα μηχανήμα τεχνητής νοημοσύνης, σε ποιες πτυχές της καθημερινότητας και της κοινωνίας μπορεί να χρησιμοποιηθεί και μέχρι ποιο επίπεδο, τόσο σε βραχυπρόθεσμο, όσο και σε μεσοπρόθεσμο, αλλά και σε μακροπρόθεσμο χρονικό ορίζοντα.

Από την πλευρά της Ευρωπαϊκής Επιτροπής επισημαίνεται ότι ενώ τα οφέλη από την χρήση μηχανών τεχνητής νοημοσύνης στην κοινωνία είναι πολλαπλά, όπως είναι για παράδειγμα η παροχή βελτιωμένων υπηρεσιών ιατρικής περίθαλψης μέσω της υποβοήθησης των υπηρεσιών που παρέχονται από τις εν λόγω μηχανές, ωστόσο δεν εκλείπουν οι περιπτώσεις που συνοδεύονται από υψηλά επίπεδα επικινδυνότητας¹³⁵. Αυτές οι περιπτώσεις πρέπει σε κάθε περίπτωση να αποφευχθούν, προκειμένου να

¹³²Βλ. ΕΟΚΕ, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52016IE5369&from=ET>

¹³³Βλ. ΕΟΚΕ, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52016IE5369&from=ET>

¹³⁴Βλ. ΕΟΚΕ, όπ.παρ., στο <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52016IE5369&from=ET>

μειωθεί η πιθανότητα εμφάνισης ανεπιθύμητων αποτελεσμάτων. Χαρακτηριστικά, από την πλευρά της Ευρωπαϊκής Επιτροπής δίνεται το εξής παράδειγμα: «Για παράδειγμα, η αδιαφάνεια πολλών αλγορίθμων μπορεί να δημιουργήσει αβεβαιότητα και να παρεμποδίσει την αποτελεσματική επιβολή της ισχύουσας νομοθεσίας για την ασφάλεια και τα θεμελιώδη δικαιώματα». Συνεπώς, η Ευρωπαϊκή Επιτροπή προτείνει την πραγματοποίηση νομοθετικών δράσεων για τη διασφάλιση της εύρυθμης λειτουργίας της εσωτερικής αγοράς, αναφορικά με την ενσωμάτωση συστημάτων τεχνητής νοημοσύνης. Συγκεκριμένα, προτείνονται «εφαρμογές όπως συστήματα βιομετρικής ταυτοποίησης ή αποφάσεις τεχνητής νοημοσύνης που άπτονται σημαντικών προσωπικών συμφερόντων, όπως στους τομείς των προσλήψεων, της εκπαίδευσης, της υγειονομικής περίθαλψης ή της επιβολής του νόμου»¹³⁶.

Σκοπός της πρότασης της Ευρωπαϊκής Επιτροπής για την εφαρμογή ενός επαρκούς κανονιστικού πλαισίου γύρω από το πεδίο της τεχνητής νοημοσύνης είναι η διασφάλιση της προστασίας της ασφάλειας και των δικαιωμάτων των πολιτών και των χειριστών αυτών των μηχανών¹³⁷. Επίσης, μέσω της πρότασης αυτής επιδιώκεται η αύξηση των επιπέδων εμπιστοσύνης των πολιτών στην αποδοχή της τεχνητής νοημοσύνης¹³⁸.

Εμμένοντας στην ανάλυση της θέσης της Ευρωπαϊκής Επιτροπής γύρω από το πεδίο της τεχνητής νοημοσύνης και ειδικότερα γύρω από το ζήτημα της απόδοσης νομικής προσωπικότητας στις εν λόγω μηχανές, διευκρινίζεται ότι αν η ενσωμάτωση των συστημάτων τεχνητής νοημοσύνης στη σύγχρονη κοινωνία δεν γίνει με δέουσα επιμέλεια και προσοχή, τότε ενδέχεται να δημιουργηθούν κίνδυνοι αναφορικά με την ασφάλεια των χειριστών των μηχανών αυτών και με την καταπάτηση των δικαιωμάτων των πολιτών¹³⁹. Ακολούθως, κλονίζεται το αίσθημα δικαίου, τόσο για τους πολίτες όσο και για τα νομικά πρόσωπα και αυτό μπορεί να οδηγήσει σε επιβράδυνση της

¹³⁶Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹³⁷Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹³⁸Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹³⁹Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

υιοθέτησης της χρήσης συστημάτων τεχνητής νοημοσύνης, τόσο από τις επιχειρήσεις όσο και από τα μέλη της κοινωνίας. Ο λόγος έγκειται στα χαμηλά επίπεδα εμπιστοσύνης που αναμένεται να επιδείξουν προς τις μηχανές αυτές.

Στο σημείο αυτό, διευκρινίζεται ότι για λόγους ανάληψης ηγετικού ρόλου από την πλευρά της Ευρωπαϊκής Επιτροπής στο πεδίο των συστημάτων της τεχνητής νοημοσύνης, αξιοσημείωτη είναι η έκθεση που συντάχθηκε από τον Stephane Sejourne (Renew Europe) κατά την οποία διευκρινίζεται ότι είναι απαραίτητη η παρουσία ενός αποτελεσματικού συστήματος δικαιωμάτων διανοητικής ιδιοκτησίας και δικλίδων ασφαλείας, προκειμένου να μην υπονομευτή η προστασία «καινοτόμων προγραμματιστών μέσω του συστήματος διπλωμάτων ευρεσιτεχνίας της ΕΕ»¹⁴⁰. Επιπροσθέτως, στην έκθεση αυτή διευκρινίζεται ότι αυτές οι απαιτήσεις δεν πρέπει να αντίκεινται στην προάσπιση και προστασία των συμφερόντων των ανθρώπων και κατασκευαστών των μηχανών αυτών, αλλά ούτε και να οδηγούν στην καταπάτηση των αρχών δεοντολογίας της ΕΕ¹⁴¹. Για τον λόγο αυτόν, η Ευρωπαϊκή Επιτροπή προτείνει την σαφήνεια της διάκρισης μεταξύ «των ανθρωπίνων δημιουργιών που υποβοηθούνται από την τεχνητή νοημοσύνη και των δημιουργιών που παράγονται αυτόνομα από την τεχνητή νοημοσύνη»¹⁴².

Ωστόσο, αν από την πλευρά των αρχών υπάρχει ανταπόκριση μέσω της θέσπισης του κατάλληλου ρυθμιστικού και κανονιστικού πλαισίου, τότε μπορεί να προληφθεί ο κίνδυνος του κατακερματισμού της εσωτερικής αγοράς¹⁴³. Η Ευρωπαϊκή Επιτροπή προτείνει, χαρακτηριστικά, την εφαρμογή του ανωτέρω προτεινόμενου νομικού και κανονιστικού πλαισίου, τόσο στην περίπτωση δημόσιων όσο και στην περίπτωση ιδιωτικών φορέων, εντός και εκτός της Ευρωπαϊκής Ένωσης, υπό την προϋπόθεση ότι το σύστημα τεχνητής νοημοσύνης είναι διαθέσιμο στην ευρωπαϊκή αγορά ή η χρήση

¹⁴⁰Βλ. Ευρωπαϊκό Κοινοβούλιο, Τεχνητή νοημοσύνη: το ΕΚ πρωτοστατεί για την πρώτη δέσμη κανόνων της ΕΕ, 2020, στο <https://www.europarl.europa.eu/news/el/press-room/20201016IPR89544/techniti-noimosuni-to-ek-protostatei-gia-tin-proti-desmi-kanonon-tis-ee>

¹⁴¹Ευρωπαϊκό Κοινοβούλιο, όπ.παρ. στο <https://www.europarl.europa.eu/news/el/press-room/20201016IPR89544/techniti-noimosuni-to-ek-protostatei-gia-tin-proti-desmi-kanonon-tis-ee>

¹⁴²Ευρωπαϊκό Κοινοβούλιο, όπ.παρ. στο <https://www.europarl.europa.eu/news/el/press-room/20201016IPR89544/techniti-noimosuni-to-ek-protostatei-gia-tin-proti-desmi-kanonon-tis-ee>

¹⁴³Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

αυτού ασκεί άμεση ή έμμεση επιρροή στα φυσικά και νομικά πρόσωπα της ευρωπαϊκής αγοράς¹⁴⁴. Επισημαίνεται ότι «μπορεί να αφορά τόσο τους παρόχους (π.χ. έναν υπεύθυνο ανάπτυξης ενός εργαλείου αξιολόγησης βιογραφικών σημειωμάτων) όσο και τους χρήστες συστημάτων τεχνητής νοημοσύνης υψηλού κινδύνου (π.χ. μια τράπεζα που αγοράζει το εν λόγω εργαλείο αξιολόγησης). Η πρόταση δεν έχει εφαρμογή σε ιδιωτικές, μη επαγγελματικές χρήσεις»¹⁴⁵. Στο σημείο αυτό, διευκρινίζεται ότι η άσκηση άμεσης ή έμμεσης επιρροής στα φυσικά και νομικά πρόσωπα της ευρωπαϊκής αγοράς από την πλευρά των μηχανημάτων τεχνητής νοημοσύνης, είναι άμεσα συνυφασμένη με τον βαθμό στον οποίο έχει αποδοθεί νομική προσωπικότητα σε αυτά.

Επιπροσθέτως, στην τοποθέτηση της Ευρωπαϊκής Επιτροπής επισημάνθηκαν ορισμένες κατηγορίες κινδύνου που πηγάζουν από τη μη χρηστή ενσωμάτωση των συστημάτων τεχνητής νοημοσύνης στις πτυχές της κοινωνίας και της δημόσιας διοίκησης. Με βάση αυτές τις κατηγορίες κινδύνου, μπορούν να ληφθούν σχετικές αποφάσεις με την απόδοση ή μη, νομικής προσωπικότητας στα συστήματα τεχνητής νοημοσύνης.

Αρχικά, επισημάνθηκε ο λεγόμενος μη αποδεκτός κίνδυνος¹⁴⁶. Στο πλαίσιο αυτό, προβλέπεται η απαγόρευση ενός περιορισμένου τμήματος επιβλαβών χρήσεων των συστημάτων τεχνητής νοημοσύνης. Πρόκειται για χρήσεις των μηχανών τεχνητής νοημοσύνης, μέσω των οποίων καταπατώνται οι αξίες της Ευρωπαϊκής Ένωσης, λόγω παραβίασης θεμελιωδών δικαιωμάτων. Χαρακτηριστικά παραδείγματα καταπατημένων δικαιωμάτων είναι τα εξής: «κοινωνική βαθμολόγηση από κυβερνήσεις, εκμετάλλευση ευάλωτων χαρακτηριστικών των παιδιών, χρήση τεχνικών που απευθύνονται στο υποσυνείδητο και —με περιορισμένες εξαιρέσεις— συστήματα εξ

¹⁴⁴Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁴⁵Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁴⁶Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683, Βλ. Επίσης Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 121.

αποστάσεως βιομετρικής ταυτοποίησης σε ζωντανή μετάδοση σε δημοσίως προσβάσιμους χώρους τα οποία χρησιμοποιούνται για σκοπούς επιβολής του νόμου»¹⁴⁷.

Στην κατηγορία του υψηλού κινδύνου, ένα μέρος των συστημάτων τεχνητής νοημοσύνης ενδέχεται να έχουν αρνητικό αντίκτυπο σε ζητήματα ανθρώπινης ασφάλειας και καταπάτησης των ανθρωπίνων δικαιωμάτων¹⁴⁸. Η Ευρωπαϊκή Επιτροπή επισυνάπτει αντίστοιχο κατάλογο των συστημάτων τεχνητής νοημοσύνης που υπάγονται στην κατηγορία υψηλού κινδύνου και στον οποίο περιλαμβάνονται ακόμα και ζητήματα ασφαλείας που εγείρονται στην περίπτωση του κατασκευαστικού σταδίου των μηχανών αυτών και η κάλυψη των οποίων γίνεται από την τομεακή νομοθεσία της Ευρωπαϊκής Ένωσης¹⁴⁹. Προκειμένου να αποφευχθεί η κατηγορία του υψηλού κινδύνου από την χρήση των εν λόγω μηχανών, θα πρέπει να διασφαλίζεται η ποιότητα του συνόλου των δεδομένων που αξιοποιούνται και επεξεργάζονται από αυτές, καθώς επίσης και τα επαρκή επίπεδα διαφάνειας και πληροφόρησης των χειριστών τους. Ωστόσο, σε αυτές τις περιπτώσεις η ανθρώπινη εποπτεία είναι απαραίτητη¹⁵⁰. Σε περίπτωση κατά την οποία υπάρξει παραβίαση επί θεμάτων ασφαλείας, καθώς επίσης και σε ανθρώπινα δικαιώματα, οι εθνικές αρχές θα είναι σε θέση να εισέρχονται στα απαιτούμενα πληροφοριακά δεδομένα προκειμένου να διερευνήσουν τον βαθμό στον οποίο η αξιοποίηση των συστημάτων τεχνητής νοημοσύνης είναι ευθυγραμμισμένη με το εκάστοτε ισχύον νομοθετικό πλαίσιο.

Χαρακτηριστικό παράδειγμα κατηγορίας υψηλού κινδύνου είναι η χρήση μηχανών τεχνητής νοημοσύνης, προοριζόμενων για εξ αποστάσεως βιομετρική ταυτοποίηση προσώπων¹⁵¹. Η Ευρωπαϊκή Επιτροπή προτείνει ότι τέτοιες περιπτώσεις θα εισέρχονται από προγενέστερη αξιολόγηση συμμόρφωσης από την πλευρά τρίτων

¹⁴⁷Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁴⁸Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁴⁹Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁵⁰Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁵¹Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 123.

μερών. Στο πλαίσιο της αξιολογικής αυτής διαδικασίας θα συμπεριλαμβάνονται και οι απαιτήσεις τεκμηρίωσης, καθώς επίσης και η υποχρεωτική ανθρώπινη εποπτεία, η οποία θα υφίσταται από το στάδιο του σχεδιασμού και της κατασκευής των μηχανών αυτών. Συν τοις άλλοις, ο κίνδυνος μειώνεται κατόπιν της εκ των προτέρων διασφάλισης της υψηλής ποιότητας των δεδομένων που εισάγονται στις μηχανές αυτές και από τα οποία αποκλείονται περιπτώσεις διακριτικής μεταχείρισης συγκεκριμένων πληθυσμιακών κατηγοριών. Είναι σαφές ότι από τη χρήση μηχανών τεχνητής νοημοσύνης που εμπίπτουν στην κατηγορία του υψηλού κινδύνου, υπάρχει ο κίνδυνος κλονισμού της ανθρώπινης αξιοπρέπειας, της επίδειξης ασέβειας απέναντι στην οικογενειακή και την ιδιωτική ζωή των πολιτών, ενώ παράλληλα ελλοχεύει ο κίνδυνος της προστασίας προσωπικών δεδομένων και κοινωνικών διακρίσεων. Κατά συνέπεια, η χρήση μηχανών τεχνητής νοημοσύνης που εμπίπτει στην κατηγορία του υψηλού κινδύνου, όπως το παράδειγμα που προαναφέρθηκε, επί του παρόντος απαγορεύεται από την πλευρά της Ευρωπαϊκής Επιτροπής.

Όσον αφορά την κατηγορία του περιορισμένου κινδύνου, καθίσταται απαραίτητη η επιβολή συγκεκριμένων απαιτήσεων διαφάνειας για ορισμένες περιπτώσεις χρήσης μηχανών τεχνητής νοημοσύνης¹⁵². Επί παραδείγματι, τέτοιες περιπτώσεις μπορεί να αφορούν περιστατικά που κατά την χρησιμοποίηση των συστημάτων τεχνητής νοημοσύνης υπάρχει αυξημένος κίνδυνος παραποίησης των πραγματικών δεδομένων. Ο κίνδυνος αυτός εμφανίζεται, για παράδειγμα, στην περίπτωση χρήσης των λεγομένων «διαλογικών» ρομπότ¹⁵³.

Προκειμένου να αποφευχθεί αυτός ο κίνδυνος κρίνεται απαραίτητη η καλή γνώση και εξοικείωση των χειριστών των μηχανημάτων αυτών, έτσι ώστε να διασφαλιστεί η αυξημένη ικανότητα αλληλεπίδρασης μαζί τους¹⁵⁴. Ωστόσο, από την πλευρά της Ευρωπαϊκής Επιτροπής υποστηρίζεται ότι η πλειοψηφία των συστημάτων τεχνητής νοημοσύνης, που προς το παρόν χρησιμοποιούνται στην Ευρωπαϊκή Ένωση, εμπίπτουν

¹⁵²Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁵³Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 123.

¹⁵⁴Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

στην κατηγορία του ελάχιστου κινδύνου¹⁵⁵. Στην κατηγορία αυτή, η χρήση των συστημάτων τεχνητής νοημοσύνης διενεργείται με την επιφύλαξη των ισχυουσών διατάξεων της νομοθεσίας, χωρίς την ύπαρξη επιπρόσθετων νομικών υποχρεώσεων¹⁵⁶. Εντούτοις, η Ευρωπαϊκή Επιτροπή επισημαίνει ότι τίθεται στη διακριτική ευχέρεια των παρόχων μηχανών τεχνητής νοημοσύνης, η επιλογή της εφαρμογής συγκεκριμένων απαιτήσεων για την κατασκευή και χρήση μηχανών τεχνητής νοημοσύνης που έχουν υψηλά επίπεδα αξιοπιστίας, καθώς επίσης και η καθιέρωση εθελοντικών κωδίκων δεοντολογίας¹⁵⁷. Πρόκειται για ένα πλαίσιο προτάσεων, μέσω των οποίων θα επέλθει σαφής διευκρίνιση και οριοθέτηση των χρήσεων των συστημάτων τεχνητής νοημοσύνης, γεγονός που θα οδηγήσει στην ταχύτερη απάντηση στο ερώτημα για την απόδοση ή μη, νομικής προσωπικότητας τα συστήματα τεχνητής νοημοσύνης.

Από την άλλη πλευρά, σε περίπτωση που η χρήση ενός τεχνικού αντικειμένου τεχνητής νοημοσύνης καταστεί επιβλαβής για έναν άνθρωπο ή μία ομάδα ανθρώπων, δημιουργείται μία ακόμη διάσταση ευθύνης, ένας άλλος τύπος ευθύνης που αναφέρεται στον επιμερισμό της ευθύνης σε διάφορα μέρη που είναι συνυπόλογα¹⁵⁸. Ομιλούμε περί επιμερισμού της ευθύνης, διότι υπόλογοι δεν είναι μόνο οι ιδιοκτήτες των μηχανών, αλλά και οι χειριστές τους.

Μία ακόμη διάσταση ευθύνης που προκύπτει για την πρόκληση βλάβης από τον χειρισμό μηχανών τεχνητής νοημοσύνης αναφέρεται στον βαθμό και στη δυνατότητα πρόβλεψης και εκ των προτέρων αποφυγής της προκαλούμενης ζημίας¹⁵⁹. Υποδεικνύεται, λοιπόν, ότι τόσο οι ιδιοκτήτες της μηχανής, όσο και οι χειριστές της, είναι υπεύθυνοι να προβλέψουν πιθανές ζημιές που ενδεχομένως προκληθούν σε

¹⁵⁵Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 123-124.

¹⁵⁶Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁵⁷Βλ. European Parliament, όπ.παρ., στο: https://ec.europa.eu/commission/presscorner/detail/el/QANDA_21_1683

¹⁵⁸Βλ. P. Čerka, J. Grigienė & G. Sirbikytė, όπ.παρ., σελ. 376–389.

¹⁵⁹Βλ. G. Hallevy, The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control, Akron Intellectual Property Journal, Αρ.4, Τεύχος 2, 2010, σελ. 171–201.

άλλους ανθρώπους από τη χρήση της μηχανής και βάσει της εν λόγω πρόβλεψης να αποτραπεί η πρόκληση των βλαβών.

Ωστόσο, στην περίπτωση κατά την οποία αποδοθεί νομική προσωπικότητα στα συστήματα τεχνητής νοημοσύνης, η ευθύνη μετακυλίεται από τα φυσικά πρόσωπα στο ίδιο το μηχάνημα. Εξ αυτού, ισχυροποιείται η άποψη και επιχειρηματολογία της ΕΟΚΕ περί αύξησης του κινδύνου κατάχρησης των συστημάτων τεχνητής νοημοσύνης από τα φυσικά πρόσωπα, την στιγμή κατά την οποία θα απαλλάσσονται της ηθικής και αστικής ευθύνης των συνεπειών τους.

Ο George S. Cole¹⁶⁰ αναφέρεται σε προκαθορισμένους τύπους αστικής ευθύνης. Πρόκειται για την ευθύνη από το ίδιο το μηχάνημα (ευθύνη προϊόντος), την ευθύνη που αφορά στην υπηρεσία που παρέχεται από την αξιοποίηση του μηχανήματος, την ευθύνη που προκύπτει από την μη χρηστή χρήση του συγκεκριμένου μηχανήματος και τέλος, την ευθύνη που πηγάζει από αμελή χρήση της μηχανής. Τα βασικά στοιχεία για τη θεμελίωση της ευθύνης προϊόντος είναι ότι η τεχνητή νοημοσύνη πρέπει να λαμβάνεται υπόψη ως ένα «προϊόν», ότι ο εναγόμενος πρέπει να είναι πωλητής της τεχνητής νοημοσύνης, ότι το προϊόν τεχνητής νοημοσύνης πρέπει να φτάσει στον παραλήπτη του χωρίς καμία αλλαγή, ότι η τεχνητή νοημοσύνη συνοδεύεται από ελαττωματικά στοιχεία και ότι τα εν λόγω ελαττωματικά στοιχεία οδηγούν στην πρόκληση βλάβης¹⁶¹. Συνεπώς, στην προκειμένη περίπτωση, η θέση συγκεκριμένων προτύπων για την χρήση μηχανών τεχνητής νοημοσύνης είναι απαραίτητη. Τα πρότυπα αυτά πρέπει να καθορίζονται από την εκάστοτε αρμόδια εταιρεία που προβαίνει στην αξιοποίηση και χρήση τέτοιων μηχανών.

Από την πλευρά των δικαστικών αρχών, σε παγκόσμιο βεληνεκές, επί του παρόντος, δεν έχει καθοριστεί ένας σαφής ορισμός για την ευθύνη που απορρέει από τη χρήση τεχνικών αντικειμένων τεχνητής νοημοσύνης από φυσικά ή νομικά πρόσωπα¹⁶².

¹⁶⁰Βλ. G.S. Cole, Tort Liability for Artificial Intelligence And Expert Systems, Computer/Law Journal, Αρ. 10, Τεύχος 2, 1990, σελ. 127–231.

¹⁶¹Βλ. G.S. Cole, όπ.παρ., σελ. 127–231.

¹⁶²Βλ. T. Kerikmäe, E. & Pärn-Lee, Legal dilemmas of Estonian artificial intelligence strategy: in between of e-society and global race, AI & Society, 2020, σελ. 1-12, Βλ. επίσης S. O'Sullivan, N. Nevejans, C. Allen, A. Blyth, S. Leonard, U. Pagallo, U., & H. Ashrafian, Legal, Regulatory and Ethical Frameworks or Standards for AI and Autonomous Robotic Surgery, The International Journal of Medical Robotics and Computer Assisted Surgery, Αρ. 15, 2018, σελ. 1-12.

Οπότε, επιβεβαιώνεται η επιτακτική ανάγκη καθορισμού αυστηρών προτύπων από τα ίδια τα νομικά πρόσωπα που κατέχουν και αξιοποιούν τέτοιου είδους μηχανές¹⁶³, καθώς επίσης και η ανάγκη διαμόρφωσης σαφών κατευθυντήριων γραμμών από νομικής και νομολογικής άποψης¹⁶⁴.

Κρίσιμο είναι να αναφερθεί ότι η διάκριση μεταξύ του κανόνα της αυστηρής ευθύνης και του κανόνα της αμέλειας είναι δυνατόν να επιφέρει ποικίλες συνέπειες στην ανθρώπινη μεταχείριση. Ειδικότερα επί ζητημάτων προφύλαξης των εν δυνάμει θυμάτων της χρήσης μηχανών τεχνητής νοημοσύνης¹⁶⁵. Πιο συγκεκριμένα, μέσω της αυστηρής ευθύνης, απότοκος είναι η ανάδυση σημαντικών κινήτρων των χρηστών ή των κατασκευαστών των μηχανών αυτών, ώστε να προβούν σε επιμελείς δράσεις για τη μείωση του κόστους της προκαλούμενης ζημίας από τις εν λόγω μηχανές¹⁶⁶. Ουσιαστικά, όμως, στην περίπτωση του συγκεκριμένου μοντέλου, είτε ο χρήστης της μηχανής τεχνητής νοημοσύνης είτε ο κατασκευαστής της αποκρίνονται ακόμη και στην περίπτωση που έχουν οριστεί υψηλά επίπεδα ασφαλείας και προφύλαξης¹⁶⁷. Όμως, αυτό δεν συνεπάγεται την απαγκίστρωση από την υιοθέτηση σωστής προσέγγισης και συμπεριφοράς απέναντι στην χρήση μηχανών τεχνητής νοημοσύνης, ώστε να μην προκαλούνται ανεπιθύμητες βλάβες και ζημιές σε άλλους ανθρώπους ή σε κοινωνικά σύνολα.

¹⁶³Βλ. G. Sartor, Cognitive automata and the law: electronic contracting and the intentionality of software agents, *Artificial Intelligence Law*, Ap. 17, Τεύχος 4, 2009, σελ. 253–290.

¹⁶⁴Βλ. P. Henman, Improving public services using artificial intelligence: possibilities, pitfalls, governance, *Asia Pacific Journal of Public Administration*, 2020, σελ. 1–13, Βλ. επίσης Y.H. Weng, Intersection of Tokku special zone, robots, and the law: a case study on legal impacts to humanoid robots, *International Journal of Social Robot*, Ap. 7, Τεύχος 5, 2015, σελ. 841–857.

¹⁶⁵Βλ. U. Pagallo, LegalAIze: Tackling the Normative Challenges of Artificial Intelligence and Robotics Through the Secondary Rules of Law, *Perspectives in Law, Business and Innovation*, 2017, σελ. 281–300, Βλ. επίσης M. Wagner, The dehumanization of international humanitarian law: legal, ethical, and political implications of autonomous weapons systems, *Vanderbilt Journal of Transnatl Law*, Ap. 47, 2014, σελ. 1371–1424.

¹⁶⁶Βλ. M. Coeckelbergh, Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability, *Science and Engineering Ethics*, Ap. 26, 2019, σελ. 2051–2068.

¹⁶⁷Βλ. I. Sagredo-Olivenza, P.P. Gomez-Martin, M.A. Gomez-Martin, &P.A Gonzalez-Calero, Trained Behavior Trees: Programming by Demonstration to Support AI Game Designers, *IEEE Transactions on Games*, 2017, σελ. 1–1.

Παρά ταύτα, επισημαίνεται η παρουσία συγκεκριμένων επιπέδων ασφαλείας, εντός των οποίων οι χρήστες, αλλά και οι κατασκευαστές των μηχανών τεχνητής νοημοσύνης μπορούν να μειώσουν σε σημαντικό βαθμό την πιθανότητα πρόκλησης βλάβης από αυτές, εντός του πλαισίου της αστικής ευθύνης από την οποία διέπονται¹⁶⁸. Όμως, ακόμη και στην περίπτωση όπου οι χρήστες εφαρμόζουν αστικά υπεύθυνη συμπεριφορά και επιμέλεια, αν συνεπεία αυτής προκληθεί βλάβη από τη χρήση μηχανήματος τεχνητής νοημοσύνης, τότε ο χρήστης ή ο ιδιοκτήτης αυτού δεν απαλλάσσεται από την αποκατάσταση της προκληθείσας βλάβης¹⁶⁹.

Μέσω του κανόνα αμέλειας, ωστόσο, διαμορφώνεται διαφορετική προοπτική από τις ανωτέρω αναφερόμενες. Στην προκειμένη περίπτωση ο χρήστης του τεχνικού αντικειμένου τεχνητής νοημοσύνης, υποχρεούται να αποκαταστήσει το θύμα από την προκαλούμενη βλάβη της μηχανής τεχνητής νοημοσύνης, μόνο όταν και αν αποδειχθεί η ενοχή του. Αυτό σημαίνει ότι αν ο χρήστης συμπεριφέρεται επιμελώς, τότε το θύμα θα επιφορτισθεί το βάρος της ευθύνης¹⁷⁰. Επομένως, το κίνητρο των θυμάτων να υιοθετηθούν προληπτικά επίπεδα ευθύνης είναι μεγαλύτερο, λόγω του ότι εάν υποστούν οποιαδήποτε απώλεια, θα επωφεληθούν σε μεγάλο βαθμό¹⁷¹.

Εάν από τη χρήση μίας μηχανής με τεχνητή νοημοσύνη προκληθούν ζημιές λόγω απάτης ή αμέλειας¹⁷², ή λόγω ελαττωματικής κατασκευής ή αστοχίας του σχεδιασμού της μηχανής, λόγω εσφαλμένων προγραμματιστικών πράξεων, επακόλουθο θα ήταν οι κανόνες ευθύνης να αναδεικνύανε τις κατασκευαστικές αμέλειες των

¹⁶⁸Βλ. A.M. Kaplan, & M. Haenlein, The fairyland of Second Life: About virtual social worlds and how to use them, *Business Horizons*, Αρ. 52, Τεύχος 6, 2009, σελ. 563-572, Βλ. επίσης S. Shavell, *Foundations of economic analysis of law*, Cambridge, Belknap Press of Harvard University Press, 2004, σελ. 23-28.

¹⁶⁹Βλ. D. Garcia, Lethal Artificial Intelligence and Change: The Future of International Peace and Security, *International Studies Review*, Αρ. 20, Τεύχος 2, 2018, σελ. 334-341

¹⁷⁰Βλ. A. Etzioni, & O. Etzioni, Incorporating ethics into artificial intelligence, *The Journal of Ethics*, Αρ. 21, Τεύχος 4, 2017, σελ. 403-418, Βλ. επίσης J. Keating, & I. Nourbakhsh, Teaching artificial intelligence and humanity, *Communications of the ACM*, Αρ. 61, Τεύχος 2, 2018, σελ. 29-32.

¹⁷¹Βλ. D. Garcia, *όπ.παρ.*, σελ. 334-341.

¹⁷²Βλ. A. Leveringhaus, What's so bad about killer robots, *Journal of Applied Philosophy*, Αρ. 35, Τεύχος, 2, 2018, σελ. 341-358.

προγραμματιστών επί των συγκεκριμένων μηχανών τεχνητής νοημοσύνης¹⁷³. Ωστόσο, τις περισσότερες φορές, οι χρήστες δεν μπορούν να γνωρίζουν τον τρόπο με τον οποίο οι προγραμματιστές έχουν κατασκευάσει τις μηχανές τεχνητής νοημοσύνης που χρησιμοποιούν¹⁷⁴. Αυτό καταδεικνύει ότι οι δυσάρεστες συνέπειες της χρήσης των μηχανών αυτών δεν μπορούν να προβλεφθούν ή να αποφευχθούν.

Συνακόλουθα, γίνεται εμφανής μία επικίνδυνη διάσταση της χρήσης τεχνικών αντικειμένων τεχνητής νοημοσύνης, βασισμένων στη μηχανική μάθηση¹⁷⁵, όπου το ίδιο το εφαρμόσιμο λογισμικό οδηγείται, εκ σχεδιασμού του, στην εμφάνιση μη προβλεπόμενων συμπεριφορών που ενδεχομένως καταλήξουν σε μη αναμενόμενες βλάβες¹⁷⁶. Οι βλάβες αυτές μπορούν να προβλεφθούν ή να αποφευχθούν μόνο στην περίπτωση που οι χρήστες έχουν στη διάθεσή τους όλα τα απαραίτητα δεδομένα που αφορούν ακόμα και στον προγραμματισμό των μηχανημάτων, αλλά και τον τρόπο με τον οποίο αυτά καταλήγουν να επεξεργάζονται τις πληροφορίες, οι οποίες έχουν τεθεί στη διάθεσή τους¹⁷⁷. Η ισχύουσα νομοθεσία δεν κρίνεται επαρκής για την εγγύηση της υπόστασης δίκαιων διατάξεων του πλαισίου από το οποίο διέπεται η τεχνητή νοημοσύνη και οι κίνδυνοι¹⁷⁸ που ελλοχεύουν από τη χρήση της¹⁷⁹.

¹⁷³Βλ. Z. Scherz, D. Goldberg & Z. Fund, Cognitive Implications of Learning Prolog—Mistakes and Misconceptions, *Journal of Educational Computing Research*, Αρ. 6, Τεύχος 1, 1990, σελ. 89–110.

¹⁷⁴Βλ. A.S. Rich, & T.M. Gureckis, Lessons for artificial intelligence from the study of natural stupidity, *Nature Machine Intelligence*, Αρ. 1, Τεύχος 4, 2019, σελ. 174–180, Βλ. επίσης A. Kaplan, & M. Haenlein, Rulers of the world, unite- The challenges and opportunities of artificial intelligence, *Business Horizons*, Αρ. 63, Τεύχος 1, 2019, σελ. 1-14.

¹⁷⁵Βλ. E. Neri, F. Coppola, V. Miele, C. Bibbolino, & R. Grassi, Artificial intelligence: Who is responsible for the diagnosis, *La Radiologia Medica*, Αρ. 125, 2020, σελ. 517-521.

¹⁷⁶Βλ. F. Rossi, Building trust in artificial intelligence, *Journal of International Affairs*, Αρ. 72, Τεύχος 1, 2019, σελ. 127-133, Βλ. επίσης T. VoPham, J.E. Hart, F. Laden, & Y.Y. Chiang, Emerging trends in geospatial artificial intelligence (geoAI): Potential applications for environmental epidemiology, *Environmental Health*, Αρ. 17, Τεύχος 40, 2018, σελ. 1-6.

¹⁷⁷Βλ. C. Mosser, Mechanical dreams of humanity: When machines long for humanity, *Interdisciplinary Humanities*, Αρ. 35, Τεύχος 1, 2018, σελ. 78-90, Βλ. επίσης T. Saarikko, U.H. Westergren, & T. Blomquist, The internet of things: Are you ready for what's coming, *Business Horizons*, Αρ. 60, Τεύχος 5, 2017, σελ. 667-676.

¹⁷⁸Βλ. P. Tambe, P. Cappelli, & V. Yacubovich, Artificial intelligence in human resources management: Challenges and a path forward. *California Management Review*, Αρ. 61, Τεύχος 4, 2019, σελ. 15-42.

¹⁷⁹Βλ. P. McCorduck, *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*, Natick, AK Peters Ltd, 2004, σελ. 56-69, Βλ. επίσης W. Philipps, *This is why we can't have*

Ο κάτωθι πίνακας περιλαμβάνει την ακριβή δομή που αναπτύχθηκε από την UNESCO¹⁸⁰, επί της οποίας εμπεριέχονται βασικές παράμετροι. Οι εν λόγω παράμετροι είναι βοηθητικές στην σκέψη και περίσκεψη των ανωτέρω αναφερόμενων ζητημάτων, κάνοντας μία παράλληλη προσπάθεια εντοπισμού των αρμοδίων και εμπλεκόμενων φορέων.

Πίνακας 1. Από την UNESCO¹⁸¹

Απόφαση που λαμβάνεται από ρομπότ	Ανθρώπινη συμμετοχή	Τεχνολογία	Ευθύνη	Κανονισμός λειτουργίας
Κατασκευασμένο από πεπερασμένο σύνολο επιλογών, σύμφωνα με προκαθορισμένα αυστηρά κριτήρια.	Κριτήρια που είναι εφαρμοσμένα επί νομικών πλαισίων	Αμιγής χρήση της μηχανής: αξιοποίηση ντετερμινιστικών αλγόριθμων - ρομπότ	Ο κατασκευαστής της μηχανής τεχνητής νοημοσύνης	Νομικά πρότυπα, περιλαμβανομένης της εθνικής ή διεθνούς νομοθεσίας

nice things: Mapping the relationship between online trolling and mainstream culture, Cambridge, The MIT Press, 2015, σελ. 71-75.

¹⁸⁰Βλ. UNESCO, Report of COMEST on robotics ethics, Paris, World Commission on the Ethics of Scientific Knowledge and Technology, 2020, στο: <https://unesdoc.unesco.org/ark:/48223/pf0000253952>

¹⁸¹Βλ. UNESCO, όπ.παρ., διαθέσιμο σε: <https://unesdoc.unesco.org/ark:/48223/pf0000253952>

Ανεπάρκεια επιλογών, με αυξημένα επίπεδα ευελιξίας, που συνάδουν με προκαθορισμένες πολιτικές	Η λήψη της απόφασης έχει ανατεθεί στη μηχανή τεχνητής νοημοσύνης (ρομπότ)	Η δράση έχει αναληφθεί αμιγώς από την πλευρά των ρομπότ (γνωστικά ρομπότ)	Ο κατασκευαστής της μηχανής τεχνητής νοημοσύνης, ο πωλητής, καθώς επίσης και οι τελικοί χρήστες της	Κώδικες πρακτικής και χρήσης της μηχανής τεχνητής νοημοσύνης (Αρχή της προφύλαξης)
Η λήψη αποφάσεων προκύπτει ως απόρροια της αλληλεπίδρασης μηχανής τεχνητής νοημοσύνης και ανθρώπου	Η λήψη των αποφάσεων της μηχανής ελέγχεται από τον ίδιο τον άνθρωπο	Ο άνθρωπος έχει την ικανότητα άσκησης ελέγχου επί της μηχανής τεχνητής νοημοσύνης, σε περιπτώσεις όπου οι δράσεις της μηχανής είναι πιθανό να οδηγηθούν σε πρόκληση βλάβης ή ζημίας	Την ευθύνη την έχει αμιγώς ο άνθρωπος	Κώδικες ηθικής και δεοντολογίας για τη χρήση μηχανών τεχνητής νοημοσύνης

Αν και η ανωτέρω αναφερόμενη δομή είναι απλοϊκή και προσφέρει σημαντική γνώση ως προς το ζήτημα της χρηστής χρήσης και της απόδοσης ευθυνών, αποτελεί ένα σύνθετο πλάνο με αυξημένη πολυπλοκότητα, που δυσχεραίνει το έργο των αρμοδίων διαμόρφωσης κώδικα ηθικής και δεοντολογίας για τη χρήση των μηχανών αυτών, και τελικά δεν κρίνεται αρκετή για την εφαρμογή μίας δίκαιης και επαρκούς δράσης¹⁸².

¹⁸²Βλ. C. Malhotra, V. Kotwaf, & S. Dalai, Ethical Framework for Machine Learning, 2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K), 2018, σελ. 1-8, Βλ. επίσης A. Gómez de Ágreda, Ethics of autonomous weapons systems and its applicability to any AI systems, Telecommunications Policy, Αρ. 44, Τεύχος 6, 2020, σελ. 1-15.

2.3 Σύνοψη – Συμπεράσματα

Συμπεραίνουμε, ότι η αξιοποίηση της τεχνητής νοημοσύνης ενδείκνυται μόνο σε περίπτωση κατά την οποία δεν είναι αλόγιστη και καταχρηστική και δεν καταπατά τα ανθρώπινα δικαιώματα και την ηθική. Πρόκειται για ένα ζήτημα που επιδιώκεται να καλυφθεί από τους λεγόμενους *soft laws*, που έχουν υιοθετηθεί σε Αμερική και Ευρώπη. Ιδανική θα ήταν η ανάπτυξη ενός παγκοσμιοποιημένου συστήματος διασφάλισης της ηθικής αξιοποίησης και ορθολογικής χρήσης της τεχνητής νοημοσύνης. Όμως, αυτό προς το παρόν φαίνεται δύσκολο, καθώς απαιτεί και την σύμφωνη γνώμη της Κίνας, η οποία επί του παρόντος πρωτοπορεί στις τεχνολογίες τεχνητής νοημοσύνης.

Την πιο πρόσφατη και, επί του παρόντος, σημαντικότερη νομοθετικά δεσμευτική προσπάθεια για την οριοθέτηση της τεχνητής νοημοσύνης, αποτελεί ο Γενικός Κανονισμός για την Προστασία Δεδομένων και ιδίως το άρθρο 22 αυτού. Εντούτοις, η ρύθμιση των ζητημάτων δεδομένων προσωπικού χαρακτήρα που απορρέουν από τη χρήση των συστημάτων τεχνητής νοημοσύνης, δεν δύνανται να απαντηθούν σε ικανοποιητικό βαθμό στις προβλέψεις του Κανονισμού. Το γεγονός αυτό οφείλεται στο ότι οι διατάξεις του Γενικού Κανονισμού για την Προστασία Δεδομένων είναι προσανατολισμένες στην προστασία πιο παραδοσιακών μορφών επεξεργασίας δεδομένων, που διαφοροποιούνται σημαντικά από τη λειτουργία των αλγορίθμων τεχνητής νοημοσύνης¹⁸³.

Επιπρόσθετο ζήτημα άξιο αναφοράς είναι η Πρόταση Κανονισμού της Ευρωπαϊκής Επιτροπής, μέσω της οποίας τονίζεται η σημασία της ενσωμάτωσης συστημάτων τεχνητής νοημοσύνης σε διάφορες πτυχές του σύγχρονου κοινωνικοοικονομικού περιβάλλοντος. Η Πρόταση αυτή θα καταλήξει στον λεγόμενο νόμο (Πράξη) Τεχνητής νοημοσύνης (*Artificial Intelligence Act*), σκοπός του οποίου θα είναι «η μετάβαση από τη συζήτηση περί ηθικής σε δεσμευτικούς νομικούς κανόνες που διασφαλίζουν θεμελιώδεις αρχές διαφάνειας, επεξηγησιμότητας των διεργασιών, ισότητας, λογοδοσίας»¹⁸⁴.

¹⁸³Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, *όπ.παρ.*, σελ. 126 επ..

¹⁸⁴Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, *όπ.παρ.*, σελ. 221.

αναφορικά με τις κατηγορίες κινδύνου που ενδέχεται να προκύψουν από την μη χρηστή ενσωμάτωση των συστημάτων τεχνητής νοημοσύνης την κοινωνία, η Ευρωπαϊκή Επιτροπή επισημαίνει ότι η επί του παρόντος αξιοποίηση συστημάτων τεχνητής νοημοσύνης εμπίπτει στην κατηγορία του ελάχιστου κινδύνου. Όμως, σε περίπτωση κατά την οποία επέλθει εντονότερη αξιοποίηση των συστημάτων τεχνητής νοημοσύνης, μέσω της ανάθεσης ακόμα πιο ενεργού ρόλου σε αυτά, ειδικότερα σε ζητήματα λήψης αποφάσεων, τα οποία έρχονται σε άμεση συσχέτιση με πληθυσμιακά χαρακτηριστικά, με ζητήματα δημόσιας και ατομικής προστασίας και με ζητήματα προάσπισης των ανθρωπίνων δικαιωμάτων, τότε, η αξιοποίηση συστημάτων τεχνητής νοημοσύνης αδιαμφισβήτητα συνοδεύεται από υψηλότερο επίπεδο επικινδυνότητας.

Σε αυτήν την περίπτωση, θεωρούμε ότι οι μηχανές τεχνητής νοημοσύνης δεν θα πρέπει να συμμετέχουν σε διαδικασίες λήψης αποφάσεων, με κίνδυνο την καταπάτηση των ανθρωπίνων δικαιωμάτων και των αξιών κοινωνικής δικαιοσύνης και ισότητας, χωρίς η χρήση τους να συνοδεύεται από την καθ' όλα αναγκαία και απαραίτητη ανθρώπινη εποπτεία. Πιο συγκεκριμένα, συλλογικά αξιολογώντας τα δεδομένα τα οποία παρατέθηκαν στο παρόν κεφάλαιο, προτείνουμε την αποφυγή της χρήσης των συστημάτων τεχνητής νοημοσύνης σε περίπτωση κατά την οποία υπάρχει -έστω και χαμηλή- πιθανότητα επιρροής του ανθρώπινου υποσυνείδητου εκ της χρήσης αυτής. Πρόκειται για χρήσεις συστημάτων τεχνητής νοημοσύνης που δύνανται να ασκήσουν επιρροή στην ανθρώπινη συμπεριφορά, κατά τέτοιον τρόπο, που να προκαλείται ή να πρόκειται να προκληθεί, είτε στο ίδιο είτε σε άλλο άτομο, αφενός ψυχολογική και αφετέρου σωματική βλάβη. Επίσης, προτείνουμε την αποφυγή της χρήσης των συστημάτων τεχνητής νοημοσύνης σε περίπτωση κατά την οποία αυτή συνεπάγεται την εκμετάλλευση των ευαίσθητων σημείων συγκεκριμένων κατηγοριών ανθρώπων, που θεωρούνται ευάλωτοι είτε λόγω κάποιας σωματικής και διανοητικής τους αναπηρίας, είτε λόγω ηλικίας, είτε για οποιοδήποτε επιπρόσθετο δημογραφικό, σωματικό και ψυχολογικό χαρακτηριστικό τους στοιχείο, το οποίο ευθύνεται για την υπαγωγή τους στις κοινωνικά τρωτές ομάδες.

Επίσης, προτείνουμε την αποφυγή της χρήσης συστημάτων τεχνητής νοημοσύνης από το σύνολο των δημόσιων αρχών. Όταν αυτή γίνεται για λόγους ζητημάτων αξιολόγησης και εν συνεχεία κατηγοριοποίησης των επιπέδων της αξιοπιστίας των ατόμων, είτε μόνιμα είτε για συγκεκριμένο χρονικό διάστημα, βάσει της συμπεριφοράς τους ή βάσει των διαφόρων εξατομικευμένων προσωπικών, σωματικών ή/και

ψυχολογικών χαρακτηριστικών τους στοιχείων. Με αυτό, εννοούμε ότι χρειάζεται να αποφεύγεται η χρήση των συστημάτων τεχνητής νοημοσύνης σε περίπτωση κατά την οποία επιδιώκεται μέσω αυτής απόδοση κοινωνικής βαθμολογίας, που αποτέλεσμα αυτής θα είναι η διαφορετική μεταχείριση συγκεκριμένων κατηγοριών προσώπων, η οποία αναπόφευκτα οδηγεί στην ένταση των κοινωνικών διακρίσεων και ανισοτήτων.

Επιζήμια θεωρούμε ότι πρόκειται να καταστεί η χρήση συστημάτων τεχνητής νοημοσύνης όταν αυτή προορίζεται για διαδικασίες ταυτοποίησης των ανθρώπων, από τις οποίες στη συνέχεια λαμβάνονται επιτόπιες αποφάσεις. Χαρακτηριστικό παράδειγμα μιας τέτοιας χρήσης είναι η ταυτοποίηση ενός ατόμου και η λήψη επιτόπου μέτρων για το αν το άτομο αυτό επιτρέπεται να εισέλθει σε έναν συγκεκριμένο χώρο (π.χ. δημόσια υπηρεσία) ή όχι. Σε αυτές τις περιπτώσεις, ωστόσο, τα συστήματα τεχνητής νοημοσύνης μπορεί να αποδειχθούν ιδιαίτερος χρήσιμα όταν καλούνται να ανιχνεύσουν πρόσωπα με συγκεκριμένα χαρακτηριστικά στοιχεία, τα οποία τους έχουν δοθεί από τις αρμόδιες δημόσιες αρχές, όπως είναι επί παραδείγματι εγκληματίες και καταζητούμενοι.

Η χρήση των συστημάτων τεχνητής νοημοσύνης σε αυτές τις περιπτώσεις ενδείκνυται, καθώς είναι ευκολότερη η αναγνώριση και η ταυτοποίηση συγκεκριμένων χαρακτηριστικών στοιχείων που φέρουν καταζητούμενα πρόσωπα, ώστε να εντοπίζονται άμεσα και σε πραγματικό χρόνο και να ειδοποιούνται έγκαιρα οι δημόσιες αρχές. Ωστόσο, για την αξιοποίηση των συστημάτων τεχνητής νοημοσύνης για τον εντοπισμό και την ανίχνευση εγκληματικών στοιχείων και καταζητούμενων προσώπων, προτείνουμε ότι πρέπει να είναι διαθέσιμη αντίστοιχη άδεια από την εκάστοτε αρμόδια δικαστική και αστυνομική αρχή, προκειμένου να μην υπονομεύεται ο σεβασμός επί των ανθρωπίνων δικαιωμάτων.

Μάλιστα, η ανθρώπινη εποπτεία θεωρείται απαραίτητη τόσο στο στάδιο της κατασκευής, του σχεδιασμού και του προγραμματισμού των εν λόγω μηχανών, προκειμένου να διασφαλιστεί η καταχώριση πληροφοριών και δεδομένων υψηλής ποιότητας σε αυτές, όσο και στο στάδιο της πρακτικής επεξεργασίας, εξαγωγής νέων δεδομένων και λήψης αποφάσεων, ήτοι κατά τη διάρκεια του ενεργού, πλέον, ρόλου των μηχανών στο κοινωνικό περιβάλλον. Συνεχίζοντας, λαμβάνοντας υπόψη ότι η τεχνητή νοημοσύνη αποτελεί έναν κλάδο που γνωρίζει εκθετική ανάπτυξη και εξέλιξη, αλληλεπιδρώντας σε ολοένα και μεγαλύτερο βαθμό με τον ανθρώπινο παράγοντα και

την ανθρώπινη διαβίωση, είναι απαραίτητη η θέσπιση νέων τύπων ευθύνης και λογοδοσίας, αναφορικά με το ζήτημα της χρήσης μηχανών τεχνητής νοημοσύνης και του βαθμού αυτονομίας τους στη λήψη αποφάσεων.

Η πρόταση του Ευρωπαϊκού Κοινοβουλίου περί απόδοσης «νομικής προσωπικότητας» στα τεχνικά αντικείμενα τεχνητής νοημοσύνης, πρόκειται για μια σύσταση που δεν μπορεί να παραβλεφθεί, τη στιγμή κατά την οποία αποτελεί τον μελλοντικά υποχρεωτικό εφαρμοζόμενο κανονισμό, αναφορικά με τα επίπεδα αυτονομίας που παρέχονται στις μηχανές τεχνητής νοημοσύνης. Εντούτοις, αξιολογώντας τις παρούσες τεχνολογικές συνθήκες, αν και η συγκεκριμένη πρόταση δείχνει συναισθηματικά και οικονομικά αρκετά δελεαστική, θα αποδεικνυόταν «ηθικά περιττή και οικονομικά προβληματική»¹⁸⁵. Αυτό, αιτιολογείται στο γεγονός ότι, επί του παρόντος, η απόδοση της ευθύνης από βλάβη που προκαλείται από μια αυτόνομη μηχανή, αποδίδεται «καλύτερα σε φυσικά πρόσωπα ή σε υφιστάμενες κατηγορίες νομικών προσώπων, και όπου αυτό δεν είναι εφικτό, επικαιροποιημένοι νόμοι που απευθύνονται σε φυσικά πρόσωπα αποτελούν καλύτερη απάντηση από τη δημιουργία μιας νέας¹⁸⁶ κατηγορίας προσώπων».

Κατά της συγκεκριμένης πρότασης του Ευρωπαϊκού Κοινοβουλίου ασκήθηκε έντονη κριτική, ιδίως όσον αφορά τη σύγκριση μεταξύ μίας συσκευής τεχνητής νοημοσύνης και ενός φυσικού προσώπου. Ισχυρότερη κριτική ασκήθηκε από την ΕΟΚΕ, μέσω επίσημης Γνωμοδότησής της, της οποίας η στάση αντίκειται στη θέση που υποστηρίζει το Ευρωπαϊκό Κοινοβούλιο, αναφορικά με το ζήτημα της απόδοσης ή μη, νομικής προσωπικότητας στις μηχανές. Αναλυτικότερα, ο αντίλογος επιχειρηματολογεί υπέρ του κινδύνου της κατάχρησης και της μη χρηστής χρήσης των μηχανών τεχνητής νοημοσύνης από τους ανθρώπους, σε περίπτωση κατά την οποία δεν αποδίδεται σε αυτούς αστική ευθύνη, αλλά σε αυτές καθαυτές τις μηχανές για ενδεχόμενες βλάβες που θα προκαλέσει η χρήση τους. Εξ όσων διαφάνηκε, ο αντίλογος υπερίσχυσε των αναφερόμενων επί της εν λόγω πρότασης, τη στιγμή κατά την οποία δεν έγινε αποδεκτή.

¹⁸⁵Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 241-242.

¹⁸⁶Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 231 επ..

Σε κάθε περίπτωση, υποστηρίζεται ότι η αστική ευθύνη θα πρέπει να βαρύνει ένα ή περισσότερα φυσικά πρόσωπα και όχι τεχνικά αντικείμενα και μηχανές. Εντούτοις, θεωρούμε ότι το δίλημμα της απόδοσης ή μη, νομικής προσωπικότητας στις μηχανές τεχνητής νοημοσύνης μπορεί να επιλυθεί σε περίπτωση κατά την οποία συγκεκριμενοποιηθεί σε μεγάλο βαθμό ο ρόλος που θα διαδραματίζουν τα τεχνικά αντικείμενα στην καθημερινότητα της κοινωνίας των πολιτών, καθώς επίσης και στην δημόσια διοίκηση. Σαφώς, η αυστηρή αυτή οριοθέτηση δεν θα πρέπει να γίνει μόνο σε βραχυπρόθεσμο, αλλά και σε μεσοπρόθεσμο, καθώς επίσης και σε μακροπρόθεσμο χρονικό ορίζοντα. Όσο πιο οριοθετημένος και περιορισμένος είναι ο ρόλος που θα διαδραματίζουν τα συστήματα τεχνητής νοημοσύνης κατά τη χρήση τους, τόσο συντομεύει η διαδρομή για την απάντηση στο προϋπάρχον δίλημμα περί απόδοσης ή μη αστικής ευθύνης και λογοδοσίας στις μηχανές.

Κεφάλαιο 3 Ζητήματα Ηθικής

3.1 Η Ουδετερότητα στις Νέες Τεχνολογίες

Ο Peter-Paul Verbeek στο έργο του «*Moralizing Technology: Understanding and Designing the Morality of Things - Κατανόηση και σχεδιασμός της ηθικής των πραγμάτων*» διερευνά ηθικά ζητήματα, μέσω της υιοθέτησης των οποίων επιτυγχάνεται η βέλτιστη δυνατή διασύνδεση των τεχνολογικών καινοτομιών και της ανθρωπότητας. Ερμηνεύοντας την προσέγγιση του Peter-Paul Verbeek, η τεχνολογία μπορεί να ληφθεί υπόψη ως «*ηθικός διαμεσολαβητής*», μέσω του οποίου διαμορφώνεται ο τρόπος αλληλεπίδρασής της με τον ευρύτερο κόσμο. Συνεπώς, καταλήγει να είναι καθοδηγητής πιθανών συμπεριφορών¹⁸⁷. Μάλιστα, αν θεωρηθεί δεδομένο ότι από τα τεχνολογικά επιτεύγματα μπορεί να επηρεαστεί ο τρόπος με τον οποίο οι άνθρωποι σκέφτονται και αλληλεπιδρούν, συμπεραίνεται ότι ουδέν τεχνολογικό επίτευγμα δύναται να θεωρηθεί ηθικά ουδέτερο.

Τα τεχνικά αντικείμενα, όπως εξηγείται από τον θεωρητικό Peter Kroes¹⁸⁸, είναι πράγματα - αντικείμενα- τα οποία έχουν κατασκευαστεί και παραχθεί από τον ίδιο τον άνθρωπο και διαθέτουν μία συγκεκριμένη λειτουργία και ένα καθορισμένο πλάνο χρήσης. Απαρτίζονται από προϊόντα τα οποία προέρχονται μέσω της συμβολής της τεχνολογίας και υποστηρίζουν την επίλυση πρακτικών ζητημάτων που προκύπτουν στην καθημερινότητα του ανθρώπου, συμπεριλαμβανομένων εκείνων που συνδέονται με τις επιθυμίες και τις ανάγκες του¹⁸⁹.

Τα τεχνικά αντικείμενα διακατέχονται από την ανάγκη τήρησης των κανόνων χρήσης τους από την πλευρά του εκάστοτε χρήστη, καθώς και από τη δημιουργία παραμέτρων,

¹⁸⁷ Βλ. P. Verbeek, *Moralizing Technology: Understanding and Designing the Morality of Things*, London, The University of Chicago Press, 2011, σελ. 47-52.

¹⁸⁸ Βλ. P. Vermaas, P. Kroes, I. van de Poel, M. Franssen & W. Houkes, *A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems. Synthesis Lectures on Engineers, Technology, and Society*, Ap. 6, Τεύχος 1, 2011, σελ. 1-134.

¹⁸⁹ Βλ. P. Vermaas, P. Kroes, I. van de Poel, M. Franssen & W. Houkes, *όπ.παρ.*, σελ. 1-134.

άμεσα συναρτημένων με τους ρόλους των ατόμων και των κοινωνικών ιδρυμάτων, σε σχέση με τα εν λόγω τεχνικά αντικείμενα και με τη χρήση τους¹⁹⁰.

Όπως προαναφέρθηκε, τα τεχνικά αντικείμενα διαθέτουν τα δικά τους σαφή λειτουργικά και χρηστικά χαρακτηριστικά. Επιπροσθέτως, όμως, αξιολογούνται ως προς τον βαθμό χρησιμότητας και χρηστικότητάς τους. Άρα, ο σκοπός που καλύπτει η χρήση ενός συγκεκριμένου τεχνικού αντικειμένου είναι ένα ιδιάζουσας σημασίας ζήτημα, προκειμένου να είναι εφικτή η τελική αξιολόγησή του¹⁹¹.

Στο σημείο αυτό, επισημαίνεται ότι η αμφισβήτηση του βαθμού στον οποίο οι ανθρώπινες δράσεις είναι ηθικές, επεκτείνεται στο πεδίο του βαθμού ηθικής από τον οποίο χαρακτηρίζεται η χρήση των τεχνικών αντικειμένων. Θεωρείται σχεδόν σίγουρο ότι η χρήση των τεχνικών αντικειμένων είναι δυνατόν να οδηγήσει σε τροποποίηση των συνθηκών της καθημερινής διαβίωσης των ανθρώπων. Άλλωστε, η επίτευξη των ατομικών και κοινωνικών στόχων, που τίθενται από τους ίδιους τους ανθρώπους, καθίσταται δυνατή υπό την υποβοήθηση των εν λόγω τεχνικών αντικειμένων, δεδομένου ότι οι στόχοι που τίθενται από τους ανθρώπους στο στάδιο της παραγωγής των τεχνικών αντικειμένων ευθυγραμμίζονται με τα επιμέρους χαρακτηριστικά στοιχεία των συγκεκριμένων αντικειμένων. Επομένως, είναι ευκόλως αντιληπτό ότι τα τεχνικά αντικείμενα διαθέτουν έναν εγγενώς ηθικό χαρακτήρα¹⁹².

Συνεπώς, παράλληλα με τα τεχνικά αντικείμενα, τα οποία μπορούν να αντιπροσωπεύουν τόσο απλούστερα αντικείμενα καταβάλλοντας μία μικρή προσπάθεια αλληλεπίδρασης και επιρροής, όσο και πιο περίπλοκες τεχνολογίες, τίθενται στο προσκήνιο και τα κοινωνικοτεχνικά συστήματα, στα οποία περιλαμβάνονται δίκτυα διασύνδεσης των τεχνικών αντικειμένων με τους ανθρώπους. Τα εν λόγω κοινωνικοτεχνικά συστήματα προσφέρουν τη δυνατότητα περισσότερων

¹⁹⁰ Βλ. I. Kalpouzos, Double elevation: Autonomous weapons and the search for an irreducible law of war, *Leiden Journal of International Law*, Αρ. 33, Τεύχος 2, 2020, σελ. 289-312.

¹⁹¹ Βλ. P. Vermaas, P. Kroes, I. van de Poel, M. Franssen & W. Houkes, *όπ.παρ.*, σελ. 1-134.

¹⁹² Βλ. D. Cawthorne & A. Robbins-van Wynsberghe, An Ethical Framework for the Design, Development, Implementation, and Assessment of Drones Used in Public Healthcare, *Science and Engineering Ethics*, Αρ. 26, Τεύχος 5, 2020, σελ. 2867-2891.

διαστάσεων αλληλεπίδρασης μεταξύ ανθρώπων και τεχνικών αντικειμένων, γεγονός που οδηγεί στην ανάδυση περισσότερων απρόβλεπτων γεγονότων και συνθηκών¹⁹³.

Πρόκειται για μία ιδέα ακόμα περισσότερο θεμελιώδη, αν εξεταστεί υπό το πρίσμα κανονιστικής ανάλυσης¹⁹⁴. Εξαιτίας των αυξημένων επιπέδων πολυπλοκότητάς της, που αφορά την αλληλεπίδραση πολλών διαφορετικών μερών -όπως επιβεβαιώνεται από τον Bruno Latour¹⁹⁵ στη θεωρία του «δικτύου ηθοποιών» (Actor Network Theory¹⁹⁶) που ο ίδιος ανέπτυξε- δημιουργούνται τεχνικά και κοινωνικά συστήματα με μικρότερη προβλεπτική ικανότητα. Επιπλέον, δημιουργείται μεγαλύτερη δυσκολία στην αποτροπή ανεπιθύμητων απορροιών¹⁹⁷ και στην ευθύνη των επιμέρους υφιστάμενων παραγόντων σε περίπτωση βλάβης¹⁹⁸, δεδομένου ότι η τεχνολογική δράση, που αντικατοπτρίζεται στο ανωτέρω αναφερόμενο κοινωνικοτεχνικό σύστημα, αποτελεί ένα σύνολο ενεργειών των χρηστών, οι οποίοι εμπλέκονται στο ανωτέρω δίκτυο σε μία εσωτερική σχέση-διάδραση¹⁹⁹.

¹⁹³ Βλ. D.A. Norman, A. Ortony & D.M. Russell, Affect and machine design: Lessons for the development of autonomous machines, IBM Systems Journal, Αρ.42, Τεύχος 1, 2003, σελ. 38-44.

¹⁹⁴ Βλ. P. Vermaas, P. Kroes, I. van de Poel, M. Franssen & W. Houkes, όπ.παρ., σελ. 1-134.

¹⁹⁵ Βλ. M. Misri, Building Entrepreneurship Culture in Madrasah Actor-Network Theory Bruno Latour Perspective, Basic and Applied Educational Research Journal, Αρ.1, Τεύχος 1, 2020, σελ. 24-36.

¹⁹⁶ Βλ. M. Müller & C. Schurr, Assemblage thinking and actor-network theory: conjunctions, disjunctions, cross-fertilisations, Transactions of the Institute of British Geographers, Αρ.41, Τεύχος 3, 2016, σελ. 217-229.

¹⁹⁷ Βλ. D.A. Norman, A. Ortony & D.M. Russell, όπ.παρ., σελ. 38-44.

¹⁹⁸ Βλ. D. Cawthorne & A. Robbins-van Wynsberghe, όπ.παρ., σελ. 2867-2891.

¹⁹⁹ Βλ. K. Barad, Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. Signs: Journal of Women in Culture and Society, Αρ.28, Τεύχος 3, 2003, σελ. 801-831.

3.2 Ο Διαχωρισμός του Ανθρώπινου και του Μη Ανθρώπινου Χαρακτήρα

Για να καταστεί εναργής η διαφορά μεταξύ των εννοιών του τεχνικού αντικειμένου, δηλαδή ενός προϊόντος της τεχνολογίας, και του κοινωνικοτεχνικού συστήματος, προτείνεται η παραδοχή ότι το τεχνικό αντικείμενο αντιπροσωπεύεται επί παραδείγματι από ένα αεροσκάφος και το κοινωνικοτεχνικό σύστημα αντιπροσωπεύεται από ένα περίπλοκο σύστημα αεροπορίας²⁰⁰.

Για παράδειγμα, στην περίπτωση ενός σοβαρού αεροπορικού ατυχήματος, είναι απαραίτητη η ανάλυση όλων των δεδομένων και στοιχείων που βρίσκονταν στο πεδίο ελέγχου εκάστου εμπλεκόμενου αιτιολογικού παράγοντα, καθώς επίσης και του συνόλου των επικρατούντων τεχνικών δεδομένων. Επομένως, είναι πιθανή η κατάληξη σε μία σχέση πολύπλοκης αλληλεπίδρασης μεταξύ διαφόρων συστατικών και αιτιολογικών στοιχείων, που η ύπαρξή τους επέφερε ως απότοκο την πρόκληση του συγκεκριμένου αεροπορικού ατυχήματος²⁰¹.

Επιπροσθέτως, το εν λόγω δυσμενές αποτέλεσμα είναι συχνά απρόβλεπτο, λόγω της αυτονομίας του συστήματος που βασίζεται σε ένα διασκορπισμένο και καταναμημένο παράγοντα, ο οποίος εμπλέκεται με όλα τα επιμέρους συστατικά στοιχεία και με όλες τις επιλεχθείσες και εφαρμοσθείσες ενέργειες και δράσεις²⁰².

Αυτά τα πολύπλοκα συστήματα οδηγούν σε μία αναπόφευκτη συζήτηση για την ευθύνη και την ηθική σχετικά με τις μηχανές της τεχνολογίας και όλα τα παράγωγα κατασκευάσματα αυτής και τον ρόλο τους στα κοινωνικοτεχνικά συστήματα. Ζητήματα, όπως η ευθύνη των προγραμματιστών και η ύπαρξη ηθικής σε μη

²⁰⁰ Βλ. P. Berger & T. Luckmann, *The social construction of reality. A treatise in the sociology of knowledge*, London, The Penguin Press, 1966, σελ. 68-73, Βλ. επίσης K.A. Pettersen, N. McDonald & O.A. Engen, *Rethinking the role of social theory in socio-technical analysis: a critical realist approach to aircraft maintenance*, *Cognition, Technology & Work*, Αρ. 12, Τεύχος 3, 2009, σελ. 181–191.

²⁰¹ Βλ. S. Nyholm, *Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci*, *Science and Engineering Ethics*, Αρ. 24, Τεύχος 4, 2018, σελ. 1201–1219.

²⁰² Βλ. S. Soltanzadeh, J. Galliot & N. Jevglevskaja, *Customizable Ethics Settings for Building Resilience and Narrowing the Responsibility Gap: Case Studies in the Socio-Ethical Engineering of Autonomous Systems*, *Science and Engineering Ethics*, 2020, σελ. 1-16.

ανθρώπινους παράγοντες, με έμφαση στις μηχανές της τεχνολογίας, χρειάζονται απαραίτητως μία απάντηση ή, τουλάχιστον, την εκδήλωση πολλαπλών προβληματισμών που θα οδηγήσουν σε μία ευρεία δημόσια συζήτηση²⁰³.

Η θεωρία που αναπτύχθηκε από τον Bruno Latour αποτελεί ορόσημο στην πρόοδο που έχει σημειωθεί στην αντιμετώπιση του ζητήματος της αλληλεπίδρασης μεταξύ τεχνικών αντικειμένων και ανθρώπων. Ένα μελανό σημείο της είναι η τοποθέτηση στο ίδιο επίπεδο του συνόλου των τεχνικών αντικειμένων, τα οποία, όμως, χαρακτηρίζονται από διαφορετικό βαθμό πολυπλοκότητας²⁰⁴.

Λαμβάνοντας υπόψη αυτές τις συνθήκες, από νομική και κανονιστική οπτική γωνία, η ανάθεση διαφορετικών καταστάσεων στις μηχανές της τεχνολογίας και στα κοινωνικοτεχνικά συστήματα, ανάλογα με την ικανότητά τους για ενεργό δράση και επιρροή, είναι δικαιολογημένη και θα πρέπει να διαθέτει, συν τοις άλλοις, διαφορετικές ηθικές καταστάσεις και διαφορετικά επίπεδα ευθυνών²⁰⁵.

Κρίνεται, λοιπόν, απαραίτητο να διακρίνουμε την επιρροή και τη σημασία που διαδραματίζει το κάθε εμπλεκόμενο μέρος στο συλλογικό σύστημα και, ιδίως, στην περίπτωση της δημόσιας σφαίρας.

²⁰³ Βλ. S. Soltanzadeh, J. Galliot & N. Jevglevskaja, όπ.παρ, σελ. 1-16.

²⁰⁴ Βλ. M. Misri, όπ.παρ., σελ. 24-36.

²⁰⁵ Βλ. D.A. Norman, A. Ortony & D.M. Russell, όπ.παρ., σελ. 38-44, Βλ. επίσης D. Cawthorne & A. Robbins-van Wynsberghe, όπ.παρ., σελ. 2867-2891.

3.3 Η Δημιουργία και Αξιοποίηση Απρόβλεπτων Μηχανών

Στο σημείο αυτό, θα δοθεί έμφαση σε συγκεκριμένα αντικείμενα και τεχνολογίες, στοχεύοντας σε προηγμένους αλγόριθμους μηχανικής μάθησης (machine learning) ή τεχνητής νοημοσύνης, θεωρώντας τα ως τεχνικά αντικείμενα που είναι συνδεδεμένα με συγκεκριμένα τεχνικά και κοινωνικά συστήματα και τα οποία διαθέτουν υψηλά επίπεδα αυτονομίας και χαμηλά επίπεδα προβλεπτικής ικανότητας²⁰⁶. Τα υψηλότερα επίπεδα αυτονομίας είναι βασισμένα στη διαδικασία επεξεργασίας μεγάλων δεδομένων (big data)²⁰⁷.

Οι Wallach και Allen²⁰⁸ υποστηρίζουν ότι όσο πιο έξυπνα είναι τα αντικείμενα, όπως είναι επί παραδείγματι τα ρομπότ τρίτης γενιάς, τόσο πιο αυτόνομα γίνονται και αναλαμβάνουν περισσότερες και μεγαλύτερες ευθύνες. Για τον λόγο αυτόν, είναι απαραίτητο να προγραμματίζονται με ηθικές δεξιότητες λήψης αποφάσεων, για την ασφάλεια του ανθρώπινου παράγοντα²⁰⁹.

Ο Peter-Paul Verbeek²¹⁰, διερευνώντας σε μεγαλύτερο βάθος το εν προκειμένω ζήτημα, ενώ επικεντρώθηκε στην ηθική των πραγμάτων, υποστήριξε ότι τη στιγμή κατά την οποία η λειτουργία των μηχανών γίνεται σε «ανοιχτά» κοινωνικά περιβάλλοντα, όπως σε συνδεδεμένα δημόσια δίκτυα, ο σχεδιασμός μίας μορφής λειτουργικής ηθικής είναι ολοένα και επιτακτικότερος. Η εν λόγω μορφή λειτουργικής ηθικής θα χαρακτηρίζεται από υψηλά επίπεδα ευαισθησίας επί ηθικής φύσεως ζητήματα και η εφαρμογή της θα είναι συχνή σε συνθήκες που μπορούν εύκολα να προβλεφθούν²¹¹.

²⁰⁶ Βλ. F. Alaiერი & A. Vellino, Ethical Decision Making in Robots: Autonomy, Trust and Responsibility, Lecture Notes in Computer Science, 2016, σελ. 159–168.

²⁰⁷ Βλ. L. Wang, C. Yang & R.Q. Hu, Autonomous Traffic Offloading in Heterogeneous Ultra-Dense Networks Using Machine Learning, IEEE Wireless Communications, 2019, σελ. 1–8.

²⁰⁸ Βλ. W. Wallach C. & Allen, Moral Machines: Teaching Robots Right from Wrong, Oxford, Oxford University Press, 2008, σελ. 64–78.

²⁰⁹ Βλ. F. Alaiერი & A. Vellino, όπ.παρ., σελ. 159–168.

²¹⁰ Βλ. P. Verbeek, όπ.παρ., σελ. 58–61.

²¹¹ Βλ. P. Bremner, L.A., Dennis, M. Fisher & A.F. Winfield, On Proactive, Transparent, and Verifiable Ethical Reasoning for Robots, Proceedings of the IEEE, Αρ. 107, Τεύχος 3, 2019, σελ. 541–561.

Ένα παράδειγμα αποτελεί το ρομπότ Tay που αναπτύχθηκε από την εταιρεία Microsoft. Πρόκειται για ένα ρομπότ που η χρήση του ήταν βοηθητική στην παρουσίαση των συνεπειών, που θα μπορούσαν να διαμορφωθούν από ένα τεχνικό αντικείμενο στο ευρύτερο κοινωνικό σύνολο. Η εταιρεία Microsoft ξεκίνησε το 2016 ένα πρόγραμμα που ήταν επικεντρωμένο στον τομέα της ανάπτυξης της τεχνητής νοημοσύνης και το οποίο ονομάστηκε Tay²¹². Το συγκεκριμένο ρομπότ χαρακτηριζόταν από μεγάλη ικανότητα μάθησης και βασισμένο σε συγκεκριμένες διαδικτυακές αλληλεπιδράσεις του με διάφορους ανθρώπους, κατέληγε στη διαμόρφωση της δικής του «κοσμοθεωρίας». Η εν λόγω κοσμοθεωρία ήταν βασισμένη στη διαμόρφωση αντίστοιχων εκφράσεων που βασίζονταν στις ανωτέρω αναφερόμενες αλληλεπιδράσεις. Τα αποτελέσματα του εγχειρήματος αυτού ήταν ιδιαιτέρως απογοητευτικά και ανησυχητικά και αυτός ήταν ο λόγος που η εταιρεία αναγκάστηκε να απενεργοποιήσει το ρομπότ άμεσα²¹³.

Στόχος του εγχειρήματος αυτού ήταν το ρομπότ να μπορεί να αλληλεπιδρά με άλλους χρήστες στα μέσα κοινωνικής δικτύωσης, εφαρμόζοντας ανθρώπινα επικοινωνιακά πρότυπα. Πρόκειται για ένα chatbot²¹⁴ που μέσα σε λίγες μόνο ώρες είχε καταλήξει στη δημοσίευση κοινωνικά και ηθικά απροσάρμοστων σχολίων²¹⁵.

Το «Google Photos» αποτελεί ένα παρεμφερές παράδειγμα. Επρόκειτο για ένα πρόγραμμα που δημιουργήθηκε το 2018 και μάθαινε από τους χρήστες τον τρόπο για να βάζει αυτόματα ετικέτες (tags) στις φωτογραφίες. Ωστόσο, τα αποτελέσματά του ήταν, επίσης, εντελώς απογοητευτικά, και παρατηρήθηκε, μεταξύ άλλων, ότι το bot χαρακτήριζε τους έγχρωμους ανθρώπους ως γορίλλες, καθώς δεν μπορούσε να διακρίνει τη διαφορά στα ιδιαίτερα και λεπτομερή χαρακτηριστικά τους στοιχεία²¹⁶.

²¹² Βλ. F. Alaieri & A. Vellino, *όπ.παρ.*, σελ. 159–168.

²¹³ Βλ. M.J. Wolf, K.W. Miller & F.S. Grodzinsky, Why We Should Have Seen That Coming, *The ORBIT Journal*, Ap. 1, Τεύχος 2, 2017, σελ. 1–12.

²¹⁴ Βλ. A.I. Canhoto & F. Clear, Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential, *Business Horizons*, 2019, σελ. 1-11.

²¹⁵ Βλ. M.J. Wolf, K.W. Miller & F.S. Grodzinsky, *όπ.παρ.*, σελ. 1–12.

²¹⁶ Βλ. A. Etzioni & O. Etzioni, The ethics of robotic caregivers. *Interaction Studies, Social Behaviour and Communication in Biological and Artificial Systems*, Ap. 18, Τεύχος 2, 2017, σελ. 174–190.

Η υλοποίηση προγραμμάτων, ικανών να μάθουν και να προσαρμόζονται επιτυχώς στο να εκτελούν λειτουργίες που σχετίζονται με ανθρώπους, είναι ένα ζήτημα το οποίο δημιουργεί νέες προκλήσεις κανονιστικής και ηθικής φύσεως²¹⁷, τη στιγμή κατά την οποία οδηγεί σε απρόσμενα και απρόβλεπτα αποτελέσματα²¹⁸.

Συγκεκριμένα, η αξιοποίηση μηχανών τεχνητής νοημοσύνης, σκοπός των οποίων είναι η αλληλεπίδρασή τους με ανθρώπους μέσω του διαδικτύου, δεν μπορεί να καταστεί εφικτή αν δεν συνοδεύεται από τη δέουσα ηθική περίσκεψη για την κάλυψη των απαιτήσεων και προϋποθέσεων που πρέπει να πληρούνται από τις μηχανές τέτοιου τύπου²¹⁹. Ο λόγος έγκειται στο ότι οι μηχανές αυτές, παρά την απουσία ανθρώπινης διάστασης, καλούνται να ενεργήσουν ως κοινωνικοί παράγοντες και να ασκήσουν επιρροή στο περιβάλλον με το οποίο αλληλεπιδρούν. Καθίσταται εναργές ότι τροφή για σκέψη δεν θα αποτελέσει μόνο η «κατασκευή» και η «χρήση» νέων τεχνολογιών, αλλά κυρίως ο σωστός ηθικός προσανατολισμός για την ανάπτυξή τους²²⁰.

Επανερχόμενοι στο παράδειγμα της εταιρείας Microsoft, επισημαίνεται ότι η ίδια υποστήριξε πως η αποτυχία του ρομπότ Tay προέκυψε ως συνέπεια των χρηστών, οι οποίοι σκοπίμως εκμεταλλεύτηκαν τη λογισμική και λειτουργική αδυναμία του εν λόγω τεχνικού αντικειμένου²²¹. Όμως, οι Miller et al.²²² τονίζουν ότι αυτή η αιτιολόγηση της εταιρείας δεν μπορεί να την απαλλάξει από την ευθύνη πρόκλησης ηθικής βλάβης στους χρήστες που αλληλεπίδρασαν με το συγκεκριμένο ρομπότ. Το ότι η λειτουργική αποτυχία του ρομπότ αυτού δεν ήταν αναμενόμενη ούτε από την ίδια την εταιρεία, αποδεικνύει τον έντονα απρόβλεπτο χαρακτήρα των τεχνικών αντικειμένων που εμπίπτουν στο πλαίσιο της τεχνητής νοημοσύνης.

²¹⁷ Βλ. P. Bremner, L.A., Dennis, M. Fisher & A.F. Winfield, *όπ.παρ.*, σελ. 541–561.

²¹⁸ Βλ. A. Etzioni & O. Etzioni, *όπ.παρ.*, σελ. 174–190.

²¹⁹ Βλ. P. Bremner, L.A., Dennis, M. Fisher & A.F. Winfield, *όπ.παρ.*, σελ. 541–561.

²²⁰ Βλ. K.W. Miller, M.J. Wolf & F.S. Grodzinsky, Why We Should Have Seen That Coming: Comments on Microsoft's Tay Experiment, and Wider Implications, *ORBIT Journal*, Αρ.1, Τεύχος 2, 2017, σελ. 54-64.

²²¹ Βλ. A.I. Canhoto & F. Clear, *όπ.παρ.*, σελ. 1-11.

²²² Βλ. K.W. Miller, M.J. Wolf & F.S. *όπ.παρ.*, σελ. 54-64.

Η προσπάθεια που έχει καταβληθεί για την ολοένα και μεγαλύτερη προσαρμοστικότητα των τεχνικών αντικειμένων τεχνητής νοημοσύνης στην αλληλεπίδρασή τους με τους ανθρώπους, έχει οδηγήσει στον περιορισμό του απρόβλεπτου χαρακτήρα τους. Αυτό σημαίνει πως μπορούν να ενεργήσουν ως αντικείμενα επιτέλεσης συγκεκριμένων δράσεων, οι οποίες έχουν εκ των προτέρων καθοριστεί από τους κατασκευαστές τους, αλλά και ως αντικείμενα διαμόρφωσης των πλέον καταλληλότερων πλάνων δράσης ανάλογα με την εκάστοτε περίσταση στην οποία καλούνται να λειτουργήσουν. Επομένως, καταλήγουν να ασκούν επιρροή στο ευρύτερο κοινωνικό σύνολο, με τρόπο που δεν μπορεί να ελεγχθεί απόλυτα από τους ανθρώπους. Αξιοσημείωτο κρίνεται το γεγονός ότι είναι εφικτή η προσαρμογή των αλγορίθμων με σκοπό την ανεύρεση νέων μεθόδων κάλυψης των καθηκόντων τους²²³, λειτουργία που καθιστά ακόμη πιο έντονο τον απρόβλεπτο χαρακτήρα τους, γεγονός το οποίο οδηγεί στην ανάδυση νέων κινδύνων²²⁴. Αυτό επισημαίνεται και στον Κώδικα Δεοντολογίας της Ένωσης Υπολογιστικών Μηχανημάτων (*Code of Ethics of the Association for Computing Machinery*)²²⁵.

Για να περιορίσουν τις πιθανότητες αρνητικών απορροιών, οι προγραμματιστές λογισμικού πρέπει να προβούν στην επιτυχή αναγνώριση των απρόβλεπτων και εν δυνάμει επικίνδυνων ενεργειών των κατασκευασμάτων τους, ώστε να μην επιτρέψουν την άμεση αλληλεπίδρασή τους με άλλους ανθρώπους, προκειμένου να αποφευχθούν οι ανεπιθύμητες καταστάσεις που είχαν δημιουργηθεί, λόγου χάρη, στην περίπτωση της Microsoft που προαναφέρθηκε²²⁶. Κατόπιν, το καταναλωτικό κοινό πρέπει να είναι ενημερωμένο για τις ευπάθειες ενός λογισμικού που χαρακτηρίζεται από έντονα απρόβλεπτο χαρακτήρα και ενδέχεται να αναπτύξει απρόσμενη συμπεριφορά²²⁷. Αν μη τι άλλο, οι μηχανές είναι «φορτωμένες» με την παροχή πολλών υπηρεσιών προς τους

²²³ Βλ. P. Domingos, *The Master Algorithm: how the quest for the ultimate learning machine will remake our world*, New York, Basic Books, 2015, σελ. 59-67, Βλ. επίσης D. Doneda & V.A.F. Almeida, *What Is Algorithm Governance?* IEEE Internet Computing, Αρ.20, Τεύχος 4, 2016, σελ. 60-63.

²²⁴ Βλ. D. Doneda & V.A.F. Almeida, *όπ.παρ.*, σελ. 60-63.

²²⁵ Βλ. R.E. Anderson, *Social Impacts of Computing: Codes of Professional Ethics*, Social Science Computer Review, Αρ.10, Τεύχος 4, 1992, σελ. 453-469.

²²⁶ Βλ. U. Pagallo, *What Robots Want: Autonomous Machines, Codes and New Frontiers of Legal Responsibility*, Human Law and Computer Law: Comparative Perspectives, 2013, σελ. 47-65.

²²⁷ Βλ. K.W. Miller, M.J. Wolf & F.S. *όπ.παρ.*, σελ. 54-64.

ανθρώπους, και δύνανται να επηρεάσουν άλλους, αλλά και να επηρεαστούν οι ίδιες σημαντικά από το κοινωνικοτεχνικό σύστημα, δημιουργώντας ακόμη πιο αυτόνομα και απρόβλεπτα δίκτυα²²⁸.

Παρά το γεγονός της επί του παρόντος απουσίας μίας μηχανής τεχνητής νοημοσύνης, η οποία χαρακτηρίζεται από πλήρη αυτονομία, ωστόσο, με την παρούσα ταχύρρυθμη ανάπτυξη της τεχνολογίας και της επιστήμης της πληροφορικής, είναι πιθανή η κατασκευή μηχανών που θα δύνανται να οδηγηθούν στην αυτόνομη λήψη αποφάσεων. Ακολούθως, όμως, εγείρονται απορίες και ερωτήματα επί του ποιος θα είναι ο αρμόδιος για τις συνέπειες που ενδεχομένως προκληθούν από τις αυτόνομες αποφάσεις των μηχανών αυτών.

²²⁸ Βλ. S. Beck, The problem of ascribing legal responsibility in the case of robotics, AI & SOCIETY, Αρ.31, Τεύχος 4, 2015, σελ. 473–481.

3.4 Ο Απρόβλεπτος Χαρακτήρας των Μηχανών και η Απόδοση Ευθυνών

Επαίοντες, προερχόμενοι από διαφορετικά επιστημονικά πεδία, εκφράζουν τις ανησυχίες τους για την εκχώρηση ικανοτήτων και δυνατοτήτων αυτονομίας στη «σκέψη» και «δράση» των μηχανών τεχνητής νοημοσύνης. Διακατέχονται από την πεποίθηση ότι η αυτονομία αυτή πρόκειται να τους προσφέρει την ικανότητα να πράττουν παρεκκλίνοντας από το ήδη δοθέν κανονιστικό πλαίσιο²²⁹. Στο σημείο αυτό εδράζεται και η σπουδαιότητα της διερεύνησης και εκτίμησης της άσκησης έντονης επιρροής από τους προγραμματιστές, που είναι αρμόδιοι για την κατασκευή των μηχανών τεχνητής νοημοσύνης, αλλά και από άλλους βασικούς παράγοντες, από τους οποίους διαδραματίζεται σημαντικός ρόλος στο στάδιο της λειτουργικής ανάπτυξης των μηχανών αυτών²³⁰.

Συχνά, στο στάδιο του σχεδιασμού των μηχανών τεχνητής νοημοσύνης, τα προκύπτοντα αποτελέσματα δεν μπορούν να προσδιοριστούν με ακρίβεια και βεβαιότητα. Ο λόγος έγκειται στη μερική εξάρτησή τους από άλλα πεδία, τα οποία δεν ανήκουν στη σφαίρα δράσης των προγραμματιστών των τεχνικών αντικειμένων τεχνητής νοημοσύνης. Μάλιστα, συλλογιζόμενοι την αυξημένη πολυπλοκότητα των διαδικασιών λήψης αποφάσεων, η απόδοση συγκεκριμένων εξηγήσεων από τον άνθρωπο, για την προκαλούμενη ζημία, είναι δυσχερής. Δύσκολη είναι και η απόδειξη για πιθανή βλάβη ή κατασκευαστικό σφάλμα της μηχανής²³¹.

Λαμβάνοντας υπόψη ότι η συμπεριφορά των μηχανών τεχνητής νοημοσύνης δεν δύναται να χαρακτηρίζεται από απόλυτα επίπεδα προβλεψιμότητας, λόγω της επίδρασης, μεταξύ άλλων, πολλών διαφορετικών παραγόντων κοινωνικοτεχνικής φύσεως και αυτομάθησης²³², ο προσδιορισμός συγκεκριμένης αιτιώδους σχέσης²³³ για

²²⁹ Βλ. U. Pagallo, *όπ.παρ.*, σελ. 47–65.

²³⁰ Βλ. D.C. Vladeck, *Machines without principals: liability rules and artificial intelligence*, *Washington Law Review*, Αρ. 89, Τεύχος 1, 2014, σελ. 117–150.

²³¹ Βλ. P. Čerka, J. Grigienė, & G. Sirbikytė, *όπ.παρ.*, σελ. 376–389.

²³² Βλ. S. Vallor, G.A. & Bekey, *Artificial Intelligence and the Ethics of Self-learning Robots*, In P. Lin, K. Abney, & R. Jenkins (Eds.), *Robot Ethics 2.0*, Oxford, Oxford University Press, 2017, σελ. 338–353.

²³³ Η αιτιώδης σχέση είναι η σχέση μεταξύ της συμπεριφοράς ενός μέρους και του αποτελέσματος που παράγεται από αυτό. Η εξέταση της αιτιώδους σχέσης που καθορίζει και χαρακτηρίζει τις συμπεριφορές, είτε αυτές εν τέλει είναι θετικές είτε αρνητικές, προκάλεσε το αποτέλεσμα που προβλέπει ο νόμος.

την προκληθείσα βλάβη και την ανθρώπινη συμπεριφορά είναι ανέφικτος²³⁴ ή εξαιρετικά δύσκολος²³⁵.

Βάσει της παρούσας ισχύουσας νομοθεσίας, η ανωτέρω περιγραφόμενη κατάσταση είναι δυνατόν να καταλήξει σε «κατανεμημένη ανευθυνότητα» μεταξύ των εμπλεκόμενων μερών²³⁶. Οπότε, προσδιορίζεται το πιθανό ανακύπτον αποτέλεσμα που πηγάζει εκ της ελλιπούς αναγνώρισης της αιτιώδους σχέσης²³⁷ που υφίσταται μεταξύ της χρηστικής συμπεριφοράς, των επιλογών του προγραμματιστή κατά το στάδιο της παραγωγής της μηχανής και της βλάβης που έχει προκληθεί²³⁸.

Επομένως, για να υποδηλωθεί ότι κάποιος έχει προκαλέσει ένα συγκεκριμένο γεγονός, είναι απαραίτητο να δημιουργηθεί μία σύνδεση (σχέση) μεταξύ της συμπεριφοράς του και του αποτελέσματος που παράγεται.

²³⁴ Πρέπει να αναφερθεί ότι, για να ενισχυθεί η διαφάνεια και η δυνατότητα λογοδοσίας σε αυτό το τεχνικά ρυθμιζόμενο πλαίσιο, υπάρχει σήμερα ένα αυξανόμενο κίνημα στην κοινωνία των πολιτών που απαιτεί την ανάπτυξη εξηγήσιμης τεχνητής νοημοσύνης. Επίσης, η συζήτηση σχετικά με το δικαίωμα εξήγησης για αλγοριθμικές και αυτόνομες αποφάσεις που υπάγονται επί του Γενικού Κανονισμού Προστασίας Δεδομένων (GDPR) είναι ένας επιπλέον τρόπος επίτευξης των στόχων της διαφάνειας και της λογοδοσίας στην τεχνητή νοημοσύνη, όπως αναφέρθηκε και ανωτέρω στο Κεφάλαιο 2 της παρούσης μελέτης, καθώς οι αλγόριθμοι λαμβάνουν πιο κρίσιμες αποφάσεις εκ μέρους μας και είναι όλο και πιο δύσκολο να εξηγήσει και να κατανοήσει κανείς τις διαδικασίες που ακολουθούνται από πλευράς τους.

²³⁵ Βλ. H.Y. Liu, & K. Zawieska, From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence, *Ethics and Information Technology*, Ap. 22, 2017, σελ. 321-333.

²³⁶ Βλ. M. Coeckelbergh, *όπ.παρ.*, σελ. 2051-2068.

²³⁷ Αυτό το νομικό φαινόμενο ονομάζεται επίσης από άλλους συγγραφείς ως «κενό ευθύνης».

²³⁸ Βλ. P. Andras, L. Esterle, M. Guckert, T.A. Han, P.R. Lewis, K. Milanovic, & S. Wells, Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems, *IEEE Technology and Society Magazine*, Ap. 37, Τεύχος 4, 2018, σελ. 76-83, Βλ. επίσης G. Dosi, & M.E. Virgillito, Whither the evolution of the contemporary social fabric, New technologies and old socio-economic trends, *International Labour Review*, Ap.158, Τεύχος 4, 2019, σελ. 593-625.

3.5 Ο Σχεδιασμός Ηθικού Πλαισίου για το Δίκαιο της Τεχνητής Νοημοσύνης

Λόγω του αυξημένου ερευνητικού ενδιαφέροντος γύρω από το φαινόμενο της τεχνητής νοημοσύνης, αναγκαία κρίνεται η προώθηση εκτενούς συζήτησης αναφορικά με την αποτύπωση ηθικών κατευθυντήριων γραμμών, από τις οποίες πρέπει απαραίτητως να συνοδεύεται το κατασκευαστικό και προγραμματιστικό πλαίσιο των μηχανών τεχνητής νοημοσύνης²³⁹.

Τον Απρίλιο του 2019, δημοσιεύθηκαν εκ της Ευρωπαϊκής Επιτροπής οι «Οδηγίες Δεοντολογίας για Αξιόπιστη Τεχνητή Νοημοσύνη» (Ethics Guidelines for Trustworthy Artificial Intelligence), οι οποίες περιέλαμβαναν ζητήματα συνυφασμένα με το ηθικό πλαίσιο που διέπει τη χρήση μηχανών τεχνητής νοημοσύνης²⁴⁰. Βάσει των εν λόγω οδηγιών, η τεχνητή νοημοσύνη, προκειμένου η χρήση της να καταστεί αξιόπιστη και ασφαλής, πρέπει να ευθυγραμμίζεται με το υφιστάμενο νομοθετικό πλαίσιο και συγχρόνως, να επιδεικνύεται ο απαιτούμενος σεβασμός προς το σύνολο των ισχυόντων ηθικών και δεοντολογικών κανόνων²⁴¹.

Από τις εν λόγω κατευθυντήριες γραμμές προτείνονται συνολικά επτά απαιτήσεις, που είναι απαραίτητο να πληρούνται από τις μηχανές τεχνητής νοημοσύνης, ώστε να χαρακτηριστούν ως αξιόπιστες κατά τη χρήση τους. Πρόκειται για τις εξής²⁴²:

● **Ανθρώπινη υπηρεσία και εποπτεία:** Τα τεχνικά αντικείμενα τεχνητής νοημοσύνης πρέπει να βοηθούν ουσιαστικά τον ανθρώπινο παράγοντα. Αυτό σημαίνει ότι πρέπει να οδηγούν τους ανθρώπους στη λήψη σωστών αποφάσεων. Παράλληλα, δεν πρέπει να αντίκεινται στα ανθρώπινα δικαιώματα, ενώ είναι απαραίτητο να διασφαλίζεται η

²³⁹ Βλ. A. Etzioni, & O. Etzioni, *όπ.παρ.*, σελ. 403-418, Βλ. επίσης J.R. Geis, A.P. Brady, C.C Wu, J. Spencer, E. Ranshaert, J.L. Jaremko, & M. Kohli, *Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement*, *Canadian Association of Radiologists Journal*, Ap. 70, 2019, σελ. 329-334.

²⁴⁰ Βλ. N.A. Smuha, *The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence*, *Computer Law Review International*, Ap. 20, Τεύχος 4, 2019, σελ. 97 - 106.

²⁴¹ Βλ. High-Level Expert Group on AI (HLEG AI), *Ethics guidelines for trustworthy AI*. Brussels: European Commission, 2019, στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

²⁴² Βλ. High-Level Expert Group on AI (HLEG AI), *όπ.παρ.*, στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

ευρυθμία των βέλτιστων εποπτικών μηχανισμών κατά το στάδιο της χρήσης των μηχανών τεχνητής νοημοσύνης. Αυτό μπορεί να καταστεί δυνατό μέσω των προσεγγίσεων: η ανθρώπινη παρέμβαση «human-in-the-loop» (HITL), η ανθρώπινη εποπτεία «human-on-the-loop» (HOTL) και ο ανθρώπινος έλεγχος «human-in-command» (HIC).²⁴³

- **Τεχνική αντοχή και ασφάλεια:** Τα συστήματα τεχνητής νοημοσύνης πρέπει να είναι ανθεκτικά και ασφαλή. Πρέπει να είναι ασφαλή, υπό την έννοια του ότι κατά την χρήση τους πρέπει να διασφαλίζεται η δυνατότητα και το περιθώριο εφαρμογής ενός εναλλακτικού πλάνου δράσης (Plan B) σε περίπτωση που κάτι δεν λειτουργήσει σωστά. Επίσης, θα πρέπει να είναι ακριβή, αξιόπιστα και αναπαραγωγίσιμα. Αυτός είναι ο μόνος τρόπος για να διασφαλιστεί ότι οποιαδήποτε ακούσια βλάβη μπορεί να ελαχιστοποιηθεί και να προληφθεί²⁴⁴.

- **Απόρρητο και διακυβέρνηση δεδομένων:** Πέραν της ανωτέρω αναφερόμενης διασφάλισης του σεβασμού στα προσωπικά και ευαίσθητα δεδομένα, είναι απαραίτητη η εξασφάλιση των βέλτιστων μηχανισμών διακυβέρνησης δεδομένων. Αυτό μπορεί να καταστεί εφικτό μέσω της διασφάλισης των υψηλών επιπέδων ποιότητας των δεδομένων και της νομιμότητας στην πρόσβαση σε αυτά²⁴⁵.

- **Διαφάνεια:** Οι μηχανισμοί ιχνηλασιμότητας, που έχουν αναπτυχθεί στο πλαίσιο της κατασκευαστικής διαδικασίας των μηχανημάτων τεχνητής νοημοσύνης,

²⁴³ Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 219-220, Βλ. επίσης High-Level Expert Group on AI (HLEG AI), όπ.παρ., στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Βλ. επίσης L. Holmberg, Human in Command Machine Learning, 2020, στο <https://www.diva-portal.org/smash/get/diva2:1426167/FULLTEXT01.pdf>, Βλ. επίσης S.S. Tavallaey & C. Ganz, Automation to Autonomy. 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 2019, σελ. 31-34.

²⁴⁴ Βλ. High-Level Expert Group on AI (HLEG AI), όπ.παρ., στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Βλ. επίσης C.H. Hoffmann & B. Hahn, Decentered ethics in the machine era and guidance for AI regulation, AI & Society, Ap. 35, 2019, σελ. 635-644.

²⁴⁵ Βλ. High-Level Expert Group on AI (HLEG AI), όπ.παρ., στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Βλ. επίσης C.H. Hoffmann & B. Hahn, όπ.παρ., σελ. 635-644.

διαδραματίζουν καθοριστικό ρόλο στην εξασφάλιση των υψηλών επιπέδων διαφάνειας, από τα οποία διέπεται η χρήση των μηχανών τεχνητής νοημοσύνης. Επίσης, οι αποφάσεις που λαμβάνονται κατά την αξιοποίηση των μηχανών τεχνητής νοημοσύνης, πρέπει να εξηγούνται με την απόδοση στοιχείων που είναι προσαρμοσμένα στο εκάστοτε ενδιαφερόμενο και εμπλεκόμενο μέρος. Η γνώση των περιορισμών των μηχανών τεχνητής νοημοσύνης από την πλευρά των χρηστών τους, είναι ένα σημείο εξίσου σημαντικό για τη διασφάλιση της εν λόγω διαφάνειας²⁴⁶.

- **Ποικιλομορφία, μη διάκριση και δικαιοσύνη**²⁴⁷: Οι αλγοριθμικές προκαταλήψεις πρέπει να αποφεύγονται γιατί μπορούν να καταλήξουν σε ανεπιθύμητα αποτελέσματα, όπως είναι η περιθωριοποίηση συγκεκριμένων πληθυσμιακών ομάδων. Οι μηχανές και τα συστήματα τεχνητής νοημοσύνης δεν πρέπει να είναι προσβάσιμα σε συγκεκριμένες κοινωνικές ομάδες, διότι με τον τρόπο αυτόν εντατικοποιούνται οι κοινωνικής φύσεως διακρίσεις²⁴⁸.

- **Κοινωνική και περιβαλλοντική ευημερία**: Όλοι οι άνθρωποι πρέπει να επωφελούνται από την αξιοποίηση των μηχανών τεχνητής νοημοσύνης. Για τον λόγο αυτόν είναι απαραίτητη η διασφάλιση της βιωσιμότητάς τους και η γνώση των κοινωνικών τους συνεπειών²⁴⁹.

²⁴⁶ Βλ. High-Level Expert Group on AI (HLEG AI), όπ.παρ., στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Βλ. επίσης C.H. Hoffmann & B. Hahn, όπ.παρ., σελ. 635-644.

²⁴⁷ Για τους σκοπούς της παρούσας μελέτης, παρ' όλο που η «δικαιοσύνη» μπορεί να καταστεί κατανοητή ως ένας ευρύτερος όρος, στο συγκεκριμένο σημείο αναφέρεται λιγότερο στο θέμα της τεχνητής νοημοσύνης και είναι περισσότερο εστιασμένη στην αλγοριθμική δικαιοσύνη. Δεν εμπίπτει στο πεδίο της εν λόγω διατριβής να επεκταθεί σε ζητήματα εξειδικευμένα επί της αλγοριθμικής δικαιοσύνης. Μία βαθύτερη διερεύνηση της εν λόγω έννοιας, ωστόσο, απαιτείται από την πλευρά μίας μελέτης που θα είναι εξειδικευμένη σε αυτήν και δεν θα επικεντρώνεται σε άλλα πεδία ενδιαφέροντος, αλλά θα είναι αφοσιωμένη στη συζήτηση και ανάλυση της κάθε κατευθυντήριας αρχής και οδηγίας, από αυτές που προαναφέρθηκαν.

²⁴⁸ Βλ. High-Level Expert Group on AI (HLEG AI), όπ.παρ., στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Βλ. επίσης C.H. Hoffmann & B. Hahn, όπ.παρ., σελ. 635-644.

²⁴⁹ Βλ. High-Level Expert Group on AI (HLEG AI), όπ.παρ., στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Βλ. επίσης C.H. Hoffmann & B. Hahn, όπ.παρ., σελ. 635-644.

● **Ευθύνη:** Πρόκειται για την ανάγκη ανάπτυξης μηχανισμών διασφάλισης υψηλών επιπέδων ευθύνης και υπευθυνότητας αναφορικά με τη χρήση των μηχανών τεχνητής νοημοσύνης. Με αυτό είναι άμεσα συναρτημένα τα ζητήματα έννομης προστασίας που συνάδουν με τη χρήση των μηχανών τεχνητής νοημοσύνης ²⁵⁰.

Σε παρεμφερές πλαίσιο με την ανωτέρω αναφερόμενη πρωτοβουλία, επισημαίνεται ότι αρκετά κράτη, νομικές οντότητες, κυβερνητικοί και μη κυβερνητικοί φορείς έχουν προβεί στη δημοσίευση κατευθύνσεων γύρω από το ζήτημα της αξιοποίησης της τεχνητής νοημοσύνης. Οι κατευθύνσεις αυτές περιλαμβάνουν αντίστοιχες αρχές με τις προαναφερθείσες, αποσκοπώντας στη διασφάλιση των θετικών πτυχών της χρήσης της τεχνητής νοημοσύνης και της μείωσης των επιπέδων κινδύνου της περαιτέρω ανάπτυξής της. Πρόσφατες τέτοιου είδους πρωτοβουλίες είναι αυτές των:

- «Future of Life Institute – Asilomar AI» ²⁵¹
- «USC center for AI in society» ²⁵²
- «Berkman Klein Center» ²⁵³
- «AI Austin» ²⁵⁴
- «Future of Humanity Institute» ²⁵⁵
- «Institute Electrical and Electronic Engineers IEEE» ²⁵⁶
- «Center for human-compatible AI» ²⁵⁷

²⁵⁰ Βλ. High-Level Expert Group on AI (HLEG AI), όπ.παρ., στο <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>, Βλ. επίσης C.H. Hoffmann & B. Hahn, όπ.παρ., σελ. 635-644.

²⁵¹ Βλ. Future of Life, όπ.παρ., στο <https://futureoflife.org/ai-principles/?cn-reloaded=1>

²⁵² Βλ. USC Center of AI in Society, 2020, στο <https://cais.usc.edu>

²⁵³ Βλ. J. Fjeld, & A. Nagy, Principled Artificial Intelligence, 2015, στο <https://cyber.harvard.edu/publication/2020/principled-ai>

²⁵⁴ Βλ. H. Austin, Austin Tech Company Proposes Ethics Principles for Design of Digital Humans, 2019, στο <https://honestAustin.medium.com/austin-tech-company-proposes-ethics-principles-for-design-of-digital-humans-bab506260186>

²⁵⁵ Βλ. Future of Humanity Institute, Centre for the Governance of AI, 2020, στο <https://www.fhi.ox.ac.uk/govai/>

²⁵⁶ Βλ. IEEE, The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2020, στο <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html>

²⁵⁷ Βλ. Center for Human-Compatible AI, 2020, στο <https://humancompatible.ai>

- «Centre for the study on existential risks»²⁵⁸
- «K&L gates endowment for ethics»²⁵⁹
- «World Economic Forum’s Council on the Future of AI and Robotics»²⁶⁰
- «Machine Intelligence Research Institute»²⁶¹
- «AI-100»²⁶²
- «Leverhulme center for future of intelligence»²⁶³
- «Partnership on AI»²⁶⁴
- «Open AI»²⁶⁵
- «Data & Society (New York, US)»²⁶⁶
- «Foundation for Responsible Robotics»²⁶⁷
- «AI Now Initiative»²⁶⁸

Πέραν των εξελίξεων ως προς το ζήτημα των ηθικής φύσεως κατευθύνσεων που διαμορφώθηκαν από τις ανωτέρω πρωτοβουλίες, μία συζήτηση που είναι ιδιαιτέρως

²⁵⁸ Βλ. Centre for the Study of Existential Risk, A report on recent activities, 2020, στο https://www.cser.ac.uk/media/uploads/files/CSER_report_to_donors_2020_general_for_print.pdf

²⁵⁹ Βλ. K & L Gates Hub, Artificial Intelligence and Machine Learning, 2020, στο <https://www.klgates.com/artificial-intelligence-machine-learning#LangCode=en-US>

²⁶⁰ Βλ. World Economic Forum, Global Future Council on Artificial Intelligence for Humanity, 2020, στο <https://www.weforum.org/communities/gfc-on-artificial-intelligence-for-humanity>

²⁶¹ Βλ. Machine Intelligence Research Institute, A Guide to MIRI’s Research, 2020, στο <https://intelligence.org/research-guide/>

²⁶² Βλ. Stanford University, One Hundred Year Study on Artificial Intelligence (AI100), 2020, στο <https://ai100.stanford.edu/people>

²⁶³ Βλ. Leverhulme Centre for the Future of Intelligence. 2020, στο <http://lcfi.ac.uk>

²⁶⁴ Βλ. J. Yang, Bridging AI Principles to Practice with ABOUT ML, 2020, στο <https://www.partnershiponai.org/bridging-ai-principles-to-practice-with-about-ml/>

²⁶⁵ Βλ. Open AI, 2020, στο <https://openai.com/about/>

²⁶⁶ Βλ. M. Latonero, Governing Artificial Intelligence: Upholding Human Rights & Dignity, 2020, στο https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf

²⁶⁷ Βλ. Foundation for Responsible Robotics, 2020, στο <https://responsiblerobotics.org>

²⁶⁸ Βλ. AI Now Institute, 2020, στο <https://ainowinstitute.org>

πολύπλοκη, αφορά στον βαθμό αυτονομίας των μηχανών τεχνητής νοημοσύνης²⁶⁹. Η μελέτη των διαφορετικών επιπέδων αυτονομίας τους είναι αναγκαία προκειμένου να οριοθετηθεί, όπου αυτό είναι απαραίτητο, και να προαχθεί η ανθρώπινη παρέμβαση ώστε να αποφευχθούν ανεπιθύμητες συνέπειες. Τα διαφορετικά επίπεδα αυτονομίας που μπορεί να συνοδεύουν τις μηχανές τεχνητής νοημοσύνης, πρέπει να ασκούν άμεση επιρροή στους νομικούς και ηθικούς συλλογισμούς που αφορούν στη χρήση τους²⁷⁰.

²⁶⁹ Βλ. L. Kunze, N. Hawes, T. Duckett, M. Hanheide, & T. Krajnik, Artificial Intelligence for Long-Term Robot Autonomy: A Survey, IEEE Robotics and Automation Letters, 2018, Ap. 3, Τεύχος 4, σελ. 4023–4030.

²⁷⁰ Βλ. A. Osório & A. Pinto, Information, uncertainty and the manipulability of artificial intelligence autonomous vehicles systems, International Journal of Human-Computer Studies, Ap.130,2019, σελ. 40-46.

3.6 Ηθικά Ζητήματα Αυτονομίας των Μηχανών

Βασιζόμενος στον ορισμό του Ευρωπαϊκού Κοινοβουλίου περί «έξυπνων μηχανών»²⁷¹, που αναφέρθηκε σε προηγούμενο κεφάλαιο, ο Chagal-Feferkorn διεξήγαγε έρευνα που επικεντρώθηκε στην ανάλυση των επιπέδων αυτονομίας των έξυπνων μηχανών. Σκοπός της έρευνάς του ήταν να παράσχει βοήθεια στο ευρύτερο κοινό, επί του ζητήματος της διαφοροποίησης της ευθύνης που υποκρύπτεται σε κάθε περίπτωση²⁷². Ο ίδιος επεσήμανε ότι για να βρεθεί λύση στο θέμα της απόδοσης ευθυνών, είναι σημαντική η κατανόηση των πολυδιάστατων επιπέδων αυτονομίας των έξυπνων μηχανών.

Δύο μετρήσεις που αφορούν στον βαθμό αυτονομίας των έξυπνων μηχανών είναι η ικανότητα της μηχανής να προχωρήσει σε πλήρη αντικατάσταση της ανθρώπινης παρέμβασης και η ελευθερία κάτω από την οποία η μηχανή μπορεί να δράσει συγκριτικά με τον άνθρωπο. Πρόκειται για πολύπλοκες μετρήσεις με πληθώρα υπο-αναλύσεων. Μάλιστα, ο Chagal-Feferkorn ανέφερε ότι σε αυτές τις μετρήσεις θα πρέπει να λαμβάνεται υπόψη και ο βαθμός συμμετοχής της μηχανής τεχνητής νοημοσύνης στη λήψη αποφάσεων²⁷³.

Ο κάτωθι πίνακας παρουσιάζεται στη μελέτη του Chagal-Feferkorn και αντικατοπτρίζει μία μορφή μέτρησης, στην οποία εκδηλώνεται ο βαθμός που οι μηχανές τεχνητής νοημοσύνης μπορούν να αντικαταστήσουν την ανθρώπινη παρέμβαση στη διεξαγωγή εργασιών αυξημένων επιπέδων πολυπλοκότητας και στη λήψη αποφάσεων²⁷⁴.

²⁷¹Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, & L. Floridi, Artificial Intelligence and the “Good Society”: the US, EU, and UK approach, Science and Engineering Ethics, Ap. 24, 2017, σελ. 505-528.

²⁷²Βλ. K.A. Chagal-Feferkorn, Am I an Algorithm or a Product-When Products Liability Should Apply to Algorithmic Decision-Makers, Stanford Law & Policy Review, Ap. 30, 2019, σελ. 61-114.

²⁷³ Βλ. K.A. Chagal-Feferkorn, όπ.παρ., σελ. 61-114.

²⁷⁴ Βλ. K.A. Chagal-Feferkorn, όπ.παρ., σελ. 61-114.

Πίνακας 2. Από τη μελέτη του Chagal-Feferkorn²⁷⁵

	Ρομπότ Roomba	Αυτόματος πιλότος	Αυτόνομο όχημα	Ρομπότ- γιατρός
Τα ποσοστά επιτυχίας δεν είναι μετρήσιμα;				
Υπεύθυνος για περισσότερα από δύο στάδια κύκλου OODA;			+	+
Επιλέγει ανεξάρτητα τον τύπο συλλεγόμενων πληροφοριών;			?	+
Επιλέγει ανεξάρτητα τις πηγές των συλλεγόμενων πληροφοριών;				+

²⁷⁵Βλ. Κ.Α. Chagal-Feferkorn, όπ.παρ., σελ. 61–114.

Υπόσταση δυναμικής φύσης των πηγών των συλλεγόμενων πληροφοριών;				+
Αντικαθιστά επαγγελματίες σε περίπλοκα πεδία;		?	?	+
Φύση ζωής και θανάτου των αποφάσεων;		+	+	+
Απαιτούνται αποφάσεις σε πραγματικό χρόνο;		+	+	?

Ένα κριτήριο που ο Chagal-Feferkorn χρησιμοποιεί στη μελέτη του είναι ο λεγόμενος «κύκλος OODA (Observe-Orient-Decide-Act)»²⁷⁶. Ο εν λόγω κύκλος περιλαμβάνει τις πτυχές της παρατήρησης, του προσανατολισμού, της λήψης απόφασης και της δράσης. Λαμβάνοντας υπόψη την αυξημένη πολυπλοκότητα του ζητήματος της αυτονομίας των μηχανών τεχνητής νοημοσύνης, ο Chagal-Feferkorn²⁷⁷ τονίζει την ιδιάζουσα σημασία της λεπτομερούς παρατήρησης των στοιχείων από τα οποία απαρτίζονται τα επιμέρους

²⁷⁶ Βλ. K.A. Chagal-Feferkorn, *όπ.παρ.*, σελ. 61–114, Βλ. επίσης R.J. Curts & D.E. Campbell, *Avoiding Information Overload Through the Understanding of OODA Loops, A Cognitive Hierarchy and Object-Oriented Analysis and Design*, 2001, στο http://dodccrp.org/events/6th_ICCRTS/Tracks/Papers/Track4/018_tr4.pdf

²⁷⁷ Βλ. K.A. Chagal-Feferkorn, *όπ.παρ.*, σελ. 61–114.

συστήματα κάτω από τα οποία γίνεται η λήψη αποφάσεων. Η εκδήλωση των συστημάτων αυτών διεξάγεται σε τέσσερα στάδια, κατ' ακολουθήση των οριζόμενων από τον κύκλο OODA²⁷⁸. Κάθε στάδιο ασκεί διαφορετική επιρροή στην αιτιολόγηση της απόδοσης ευθυνών αναφορικά με τη χρήση έξυπνων μηχανών²⁷⁹.

Τα ανωτέρω αναφερόμενα στάδια, αναλύονται ως κάτωθι:

(Α) **Παρατήρηση:** Το στάδιο αυτό αφορά τη συλλογή στοιχείων και πληροφοριών που πηγάζουν από το σύνολο των διαθέσιμων βάσεων δεδομένων²⁸⁰.

(Β) **Προσανατολισμός:** Πρόκειται για ένα στάδιο που είναι συνδεδεμένο με τη διαδικασία της επεξεργασίας και ανάλυσης των συλλεγόμενων στοιχείων²⁸¹.

(Γ) **Απόφαση:** Το συγκεκριμένο στάδιο αφορά τη διαδικασία λήψης αποφάσεων προκειμένου να αναληφθούν και να εφαρμοστούν συγκεκριμένες δράσεις²⁸².

(Δ) **Πράξη:** Στο τελευταίο στάδιο εφαρμόζονται στην πράξη οι αποφάσεις που λήφθηκαν στο αμέσως προηγούμενο στάδιο²⁸³.

Δεδομένων των επιμέρους σταδίων από τα οποία απαρτίζεται ο κύκλος OODA²⁸⁴, εξάγεται το συμπέρασμα ότι όσο μεγαλύτερο είναι το πλήθος των χαρακτηριστικών του συστήματος που βρίσκονται σε αντιστοιχία με «μη έξυπνες» μηχανές, τόσο πιο

²⁷⁸ Ο OODA είναι ένας κύκλος (loop) προσανατολισμού και απαρτίζεται από τα χαρακτηριστικά «παρατήρηση-προσανατολισμός-απόφαση-πράξη». Πρόκειται για μία στρατηγική που αναπτύχθηκε πρωτίστως από τον στρατιωτικό Τζον Μπόιντ για να εξηγήσει τον τρόπο με τον οποίο τα άτομα και οι οργανισμοί μπορούν να αποκομίσουν οφέλη και κέρδη σε αβέβαιες και χαοτικές καταστάσεις.

²⁷⁹ Βλ. R.J. Curts & D.E. Campbell, *όπ.παρ.*, στο http://dodccrp.org/events/6th_ICCRTS/Tracks/Papers/Track4/018_tr4.pdf, Βλ. επίσης J. Wade, J. Buenfil, & P. Collopy, *A Systems Engineering Approach for Artificial Intelligence: Inspired by the VLSI Revolution of Mead & Conway*, INSIGHT, Αρ. 23, Τεύχος 1, 2020, σελ. 41–47.

²⁸⁰ Βλ. S. McIntosh, *The Wingman-Philosopher of MiG Alley: John Boyd and the OODA Loop*, Air Power History, Αρ. 58, Τεύχος 4, σελ. 24-33.

²⁸¹ Βλ. S. McIntosh, *όπ.παρ.*, σελ. 24-33.

²⁸² Βλ. S. McIntosh, *όπ.παρ.*, σελ. 24-33.

²⁸³ Βλ. S. McIntosh, *όπ.παρ.*, σελ. 24-33.

²⁸⁴ Βλ. K.A. Chagal-Feferkorn, *όπ.παρ.*, σελ. 61–114.

πιθανή είναι η ευθυγράμμιση με το καταναλωτικό δίκαιο²⁸⁵. Όμως, οι έξυπνες μηχανές, επειδή διαθέτουν εξειδικευμένα χαρακτηριστικά, ενδεχομένως να ακολουθήσουν διαφορετική ταξινόμηση από τις μηχανές που δεν χαρακτηρίζονται ως «έξυπνες». Αυτό καταδεικνύει την αναγκαιότητα σχεδιασμού ενός διαφοροποιημένου πλαισίου απόδοσης ευθυνών και αντιμετώπισης της χρήσης τους.

Παρά το γεγονός ότι η τεχνολογική πρόοδος δεν έχει υποστηρίξει, επί του παρόντος, μηχανές που να διαθέτουν επαρκή επίπεδα αυτονομίας, προκειμένου να επέλθει η πλήρης αντικατάσταση του ανθρώπου σε περίπλοκες και χρονοβόρες διαδικασίες, όπως είναι η περίπτωση των ρομπότ-γιατρών του κύκλου OODA που παρατέθηκε ανωτέρω, ωστόσο, εάν αυτό το επίπεδο τείνει να πραγματωθεί, θα πρέπει να είναι διαθέσιμοι θεωρητικοί μηχανισμοί, προκειμένου να εφαρμοστεί ένας τύπος ανάθεσης ευθύνης χωρίς να προκληθούν ανεπιθύμητα αποτελέσματα που υποκρύπτονται πίσω από την προαγωγή και αξιοποίηση των τεχνολογικών καινοτομιών²⁸⁶.

Λόγω του ότι η απόλυτη ρομποτική αυτονομία δεν προβλέπεται να υλοποιηθεί στο άμεσο βραχυπρόθεσμο μέλλον, κάτι που ενδεχομένως δεν ισχύει για την περίπτωση της «γενικής τεχνητής νοημοσύνης», ωστόσο, παρατηρείται ένα σημαντικό «ρεύμα» ενάντια στην θέσπιση νομοθεσίας που εξειδικεύεται στην κάλυψη του πεδίου αυτού. Πρόσφατα, περισσότεροι από εκατόν πενήντα εμπειρογνώμονες στους τομείς της τεχνητής νοημοσύνης, της ρομποτικής, του εμπορίου, της νομοθεσίας και της δεοντολογίας, προερχόμενοι από δεκατέσσερα διαφορετικά κράτη, προέβησαν στην υπογραφή «ανοιχτής» επιστολής, με την οποία καταγγέλλεται η Ευρωπαϊκή πρόταση χορήγησης νομικής προσωπικότητας στα τεχνικά αντικείμενα τεχνητής νοημοσύνης²⁸⁷. Μέσω της συγκεκριμένης κίνησης, εκδηλώνεται η πεποίθηση ότι οι έξυπνες μηχανές

²⁸⁵ Βλ. H.W. Micklitz, Do Consumers and Businesses Need a New Architecture of Consumer Law, A Thought Provoking Impulse, Yearbook of European Law, Ap. 32, Τεύχος 1, 2013, σελ. 266–367, Βλ. επίσης K.A. Chagal-Feferkorn, όπ.παρ., σελ. 61–114, Βλ. επίσης G.W. Ng, & W.C. Leung, Strong Artificial Intelligence and Consciousness, Journal of Artificial Intelligence and Consciousness, Ap. 07, Τεύχος 01, 2020, σελ. 63–72.

²⁸⁶ Βλ. K.A. Chagal-Feferkorn, όπ.παρ., σελ. 61–114.

²⁸⁷ Βλ. H. Tavani, Can Social Robots Qualify for Moral Consideration, Reframing the Question about Robot Rights. Information, Ap.9, Τεύχος 4, 2018, σελ. 73, Βλ. επίσης D.J. Gunkel, The other question: can and should robots have rights, Ethics and Information Technology, Ap. 20, Τεύχος 2, 2017, σελ. 87–99, Βλ. επίσης U. Pagallo, Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots. Information, Ap. 9, Τεύχος 9, 2018, σελ.230.

τεχνητής νοημοσύνης δεν μπορούν να αποκτήσουν ηθική υπόσταση. Επίσης, υποστηρίζεται ότι δεν πρέπει να αποκτήσουν δικαιώματα, κάτι που θα πραγματοποιηθεί αν εφαρμοστεί η πρόταση της Ευρωπαϊκής Επιτροπής.

Πρέπει να επισημανθεί ότι ένα εκ των βασικότερων στοιχείων που χρήζουν προσοχής είναι η ταχύτητα κάτω από την οποία ολοκληρώνεται η διαδικασία της μάθησης των «έξυπνων» μηχανών²⁸⁸. Το σημείο αυτό μπορεί να είναι αντιπροσωπευτικό των υφιστάμενων περιορισμών της διαδικασίας εκπαίδευσης, ασκώντας περιοριστικές πιέσεις στη νομική και ηθική ευθύνη των μηχανών²⁸⁹. Όμως, απόρροια αυτών είναι η δημιουργία των εξής ερωτημάτων: Με ποιον τρόπο θα μπορούσε κάποιος να επιβάλλει οποιαδήποτε τιμωρία σε μία μηχανή; Και τι είδους τιμωρία θα μπορούσε να είναι αυτή; Θα είχε νόημα η εν λόγω διαδικασία;

Σε αυτήν την περίπτωση, επικρατούν δύο επιλογές, αφενός η αποκατάσταση και αφετέρου η αποζημίωση. Η αποκατάσταση αφορά τον επαναπρογραμματισμό του υπεύθυνου τεχνικού αντικειμένου²⁹⁰. Η αποζημίωση αφορά την αποκατάσταση της βλάβης που προκλήθηκε από την αξιοποίηση του τεχνικού αντικειμένου²⁹¹. Σε αμφότερες των περιπτώσεων, το ανωτέρω αναφερόμενο ευρωπαϊκό ψήφισμα, που αφορά την απόδοση ηθικής ευθύνης απέναντι στις έξυπνες μηχανές, μπορεί να αποκτήσει ουσία²⁹².

Εντούτοις, όταν διερευνάται η πρόκληση βλαβών, που προκύπτουν σε σύνθετα τεχνικά και κοινωνικά συστήματα, τότε υπονομεύεται η ασφαλής υιοθέτηση αιτιώδους σχέσης

²⁸⁸ Βλ. C.C. Beltran-Hernandez, D. Petit, I.G. Ramirez-Alpizar, T. Nishi, S. Kikuchi, T. Matsubara, & K. Harada, Learning Force Control for Contact-rich Manipulation Tasks with Rigid Position-controlled Robots, IEEE Robotics and Automation Letters, 2020, σελ. 1.

²⁸⁹ Βλ. V. Nekrasov, Criminal responsibility of robots: actual problem and areas for further development, Gosudarstvo i pravo, Αρ. 5, 2019, σελ. 134-140, Βλ. επίσης G. Magyar, M. Vircikova, & P. Sincak, Interactive Q-learning for social robots that learn from the Wizard: A pilot study, 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2016, σελ. 9-12.

²⁹⁰ Βλ. J. Wassermann, V. Vonesek, & A. Vick, Distributed Industrial Robot Control Using Environment Perception and Parallel Path Planning Cloud Services, 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 2018, σελ. 1263-1265.

²⁹¹ Βλ. C. Reed, How should we regulate artificial intelligence, Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, Αρ.376, Τεύχος 2128, 2018, σελ. 1-12.

²⁹² Βλ. E.M. Kaufman, Reprogramming consent: implications of sexual relationships with artificially intelligent partners, Psychology & Sexuality, 2020, σελ. 1-12.

και απόδοσης ευθύνης προς κάποιο συγκεκριμένο μέρος²⁹³. Ακολούθως, σε τέτοιου είδους περίπλοκες καταστάσεις, όταν προκαλείται βλάβη, δόκιμη κρίνεται η έμφαση στην οικονομική αποζημίωση και όχι σε αυτό καθ' αυτό το βλαπτικό συμβάν ²⁹⁴.

Συνεπεία αυτών, εγείρεται ο κίνδυνος της εκούσιας και κακόβουλης αποφυγής της αναλογούσας ευθύνης, από την πλευρά των νομικών οντοτήτων και των φυσικών προσώπων που κατέχουν στην κυριότητά τους τη μηχανή τεχνητής νοημοσύνης που προκάλεσε τη βλάβη. Πρόκειται για τη λεγόμενη δυσανάλογη απόδοση ευθύνης²⁹⁵. Προς αποφυγή εμφάνισης ανάλογων προβλημάτων, προτείνεται ο σχεδιασμός και προγραμματισμός των τεχνικών αντικειμένων τεχνητής νοημοσύνης με πλήρη προσανατολισμό στην προάσπιση αρχών, όπως είναι η δικαιοσύνη, η αξιοπιστία, η ασφάλεια, η προστασία της ιδιωτικής ζωής και των δεδομένων, η περιεκτικότητα, η διαφάνεια και η λογοδοσία²⁹⁶.

²⁹³ Βλ. C. Cauffman, Robo-liability: The European Union in search of the best way to deal with liability for damage caused by artificial intelligence, *Maastricht Journal of European and Comparative Law*, Αρ. 25, Τεύχος 5, 2018, σελ. 527–532, Βλ. επίσης C. Reed, *όπ.παρ.*, σελ. 1-12.

²⁹⁴ Βλ. S. Sholla, R.N. Mir, & M.A. Chishti, Eventuality of an Apartheid State of Things, *International Journal of Technoethics*, Αρ.9, Τεύχος 2, 2018, σελ. 62–76, Βλ. επίσης V. Özdemir, & N. Hekim, Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, *The Internet of Things and Next-Generation Technology Policy*, *OMICS: A Journal of Integrative Biology*, Αρ. 22, Τεύχος 1, 2018, σελ. 65–76.

²⁹⁵ Βλ. S. Sholla, R.N. Mir, & M.A. Chishti, *όπ.παρ.*, σελ. 62–76.

²⁹⁶ Βλ. J.R. Geis, A.P. Brady, C.C. Wu, J. Spencer, E. Ranshaert, J.L. Jaremko & M. Kohli, *όπ.παρ.*, σελ. 329-334, Βλ. επίσης S. O'Sullivan, N. Nevejans, C. Allen, A. Blyth, S. Leonard, U. Pagallo, & H. Ashrafian, H. *όπ.παρ.*, σελ. 1-12.

3.7 Η Εφαρμογή της Τεχνητής Νοημοσύνης και τα Θεμελιώδη Δικαιώματα

Ένα σημείο που αξίζει προσοχής είναι ότι τα ελαττώματα είναι αναμενόμενα και ότι μπορούν να θεωρηθούν ακόμη και επιθυμητά για την ταχύτερη βελτίωση ενός μηχανήματος τεχνητής νοημοσύνης. Επομένως, ένα κανονιστικό πλαίσιο που θα εξαλείψει όλα τα υφιστάμενα και τυχόντα μελλοντικά αναδυόμενα ελαττώματα ή ζημιές, θα ήταν μη αποδεκτό. Τα ρομπότ που είναι εμπνευσμένα από την τεχνολογία τεχνητής νοημοσύνης είναι προϊόντα με εγγενώς απρόβλεπτους κινδύνους²⁹⁷.

Η ιδέα που υποκρύπτεται πίσω από την έρευνα του πεδίου της μηχανικής μάθησης αναφέρεται στο ότι τα ρομπότ πρέπει να ανακαλύψουν και να αποκτήσουν νέους τρόπους αλληλεπίδρασης, χωρίς τη ρητή οδηγία του σχεδιαστή και προγραμματιστή τους²⁹⁸. Η γενική τεχνητή νοημοσύνη, η οποία είναι βέβαιο ότι έχει μία βλέψη με ορίζοντα που είναι στραμμένος στο μέλλον, έγκειται στο ότι η ανωτέρω αναφερόμενη αυτονομία των μηχανών τεχνητής νοημοσύνης δύναται να υιοθετηθεί, ακόμη και χωρίς την υπόσταση κάποιας σιωπηρής οδηγίας ή συμφωνίας²⁹⁹. Συνεπώς, θα ήταν δυνατό να υποστηριχθεί ότι αυτές οι τεχνολογίες είναι απρόβλεπτες εκ του σχεδιασμού τους και εκ της πρώτης θέσης τους σε λειτουργία.

Από νομικής άποψης, θεμελιώδης είναι η απόδοση διάχυτης ευθύνης, που είναι δυνητικά επιμερισμένη στην ευθύνη πολλών και διαφορετικών παραγόντων, οι οποίοι συνάδουν με τη δημόσια διοίκηση. Υπό αυτήν την έννοια, πρέπει να συλλογιστούμε το πεδίο, στο οποίο γίνονται παραδοχές για την απόδοση ευθύνης. Το ερώτημα που γεννάται, λοιπόν, δεν είναι μόνο το πώς πρέπει να καταστούν υπεύθυνα τα

²⁹⁷ Βλ. M.J. Rosenstrauch, & J. Kruger, Safe human-robot-collaboration-introduction and experiment using ISO/TS 15066, 2017 3rd International Conference on Control, Automation and Robotics (ICCAR), 2017, σελ. 740-744.

²⁹⁸ Βλ. P. Jamshidi, J. Camara, B. Schmerl, C. Kaestner, & D. Garlan, Machine Learning Meets Quantitative Planning: Enabling Self-Adaptation in Autonomous Robots, 2019 IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), 2019, σελ. 39-50.

²⁹⁹ Βλ. C. Yi Tan, Artificial Intelligence, Artificial Persons, and the Law, Becoming Human: Artificial Intelligence Magazine, 2018, στο <https://becominghuman.ai/artificial-intelligence-artificial-persons-and-the-law-2cce322743b6>

εμπλεκόμενα μέρη, αλλά και το πώς να κατανεμηθεί εύλογα και δίκαια η ευθύνη αυτή³⁰⁰.

Η ιδέα της από κοινού ευθύνης μεταξύ των διαφόρων παραγόντων που εμπλέκονται σε ένα κοινωνικοτεχνικό δίκτυο φαίνεται ως μία λογική προοπτική, η οποία προϋποθέτει, προκειμένου να αποδοθεί δίκαιη ευθύνη σε καθένα από τα εμπλεκόμενα μέρη, την ανάλυση των σφαιρών ελέγχου, καθώς επίσης και του βαθμού στον οποίο επηρεάζουν τις διαδραματιζόμενες καταστάσεις. Όμως, απαραίτητη είναι η ανάλυση επιπλέον παραγόντων, τεχνολογικών και ανθρώπινων, εστιάζοντας στον βαθμό που υπάρχει μεταξύ τους ενδοσύνδεση³⁰¹.

Ωστόσο, υπάρχει ακόμα σημαντική απόκλιση από την επίτευξη μίας λογικής συναίνεσης³⁰², ως προς το ζήτημα της θέσπισης των βέλτιστων νομικής φύσεως παραμέτρων για το ρυθμιστικό πλαίσιο έξυπνων μηχανών και πραγμάτων, αν και ήδη έχουν τεθεί στο προσκήνιο πολλές εξελίξεις σχετικά με τις ηθικές κατευθυντήριες γραμμές³⁰³ που πρέπει να ακολουθηθούν κατά τη χρήση μηχανών τεχνητής νοημοσύνης³⁰⁴.

Οι μηχανές τεχνητής νοημοσύνης μπορούν να επηρεάσουν τις σχέσεις μεταξύ των ανθρώπων, τη διαμόρφωση συμπεριφορών και τις παγκόσμιες απόψεις, ειδικά και πιο επιδραστικά, όταν μία λειτουργική τους πτυχή χαρακτηρίζεται από διαφορετικά επίπεδα αυτονομίας και από αυξημένα επίπεδα τεχνολογικής πολυπλοκότητας. Αυτό παρατηρείται συχνά στην περίπτωση των μηχανών τεχνητής νοημοσύνης, οι οποίες

³⁰⁰ Βλ. M. Janssen, P. Brous, E. Estevez, L.S. Barbosa, & T. Janowski, Data governance: Organizing data for trustworthy Artificial Intelligence, *Government Information Quarterly*, Αρ. 37, Τεύχος 3, 2020, σελ. 1-8.

³⁰¹ Βλ. K. Barad, Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter, *Signs: Journal of Women in Culture and Society*, Αρ. 28, Τεύχος 3, 2003, σελ. 801-831.

³⁰² Βλ. J. Habermas, S. Lennox, & F. Lennox, The Public Sphere: An Encyclopedia Article (1964), *New German Critique*, Αρ.3, 1974, σελ. 49-55.

³⁰³ Κατά την δική μας άποψη, υποστηρίζεται ότι η εν λόγω συναίνεση πρέπει να οικοδομηθεί σύμφωνα με την πρόταση του Jurgen Habermas, δηλαδή μέσω διαλεκτικών συγκρούσεων στη δημόσια σφαίρα.

³⁰⁴ Βλ. E.F. Villaronga, P. Kieseberg, & T. Li, Humans forget, machines remember: Artificial intelligence and the Right to Be Forgotten, *Computer Law & Security Review*, Αρ.34, Τεύχος 2, 2018, σελ. 304–313.

διαθέτουν ικανότητα μάθησης και πραγματοποίησης συλλογισμών, βασισμένων στην εφαρμογή τεχνικών βαθιάς μάθησης³⁰⁵.

Λόγω των αυξημένων επιπέδων επικινδυνότητας από τα οποία συνοδεύεται η εξέλιξη της τεχνολογίας, ενισχυόμενη από την εξίσου ταχύρρυθμη πρόοδο της τεχνητής νοημοσύνης και του διαδικτύου των πραγμάτων, το κράτος δικαίου θα πρέπει να έχει ως βασική προϋπόθεση την συμπεριφορά κατά την ανάπτυξη των τεχνολογιών³⁰⁶. Αυτό σημαίνει ότι η συμπεριφορά πρέπει να καθορίζει την τεχνολογική ανάπτυξη και όχι το αντίστροφο.

Προκειμένου να διασφαλιστεί η ευρυθμία της εφαρμογής των υφιστάμενων νομοθετικών και κανονιστικών πλαισίων, η πρόοδος της τεχνολογίας πρέπει να αποτελεί μία συνθήκη που θα ενσωματώνει ηθικούς κανόνες³⁰⁷. Για να καταστεί αυτό δυνατό, απαραίτητη είναι η πλήρης κατανόηση του βαθμού στον οποίο οι μηχανές τεχνητής νοημοσύνης και το σύνολο των τεχνικών αντικειμένων μπορούν να επηρεάσουν την ανθρώπινη διαβίωση, αποσκοπώντας στη βελτίωση αυτής. Στην προκειμένη περίπτωση, πρέπει να δοθεί ακόμα μεγαλύτερη προσοχή στις μηχανές υψηλών επιπέδων αυτονομίας³⁰⁸.

Διευκρινίζεται, επίσης, η σημασία της καθοδήγησης της ανάπτυξης και κατασκευής τεχνολογικών αντικειμένων από το ίδιο το νομοθετικό πλαίσιο. Συνεπώς, θα είναι εφικτή η προαγωγή της ευαισθησίας για τη συναρμογή με αξίες, όπως είναι η προάσπιση της ιδιωτικής ζωής και της ηθικής³⁰⁹.

³⁰⁵ Βλ. D. Ihde, *The Experience of Technology: Human-Machine Relations, Cultural Hermeneutics*, Ap. 2, Τεύχος 3, 1975, σελ. 267–279.

³⁰⁶ Βλ. M. Queudot, & M.J. Meurs, *Artificial Intelligence and Predictive Justice: Limitations and Perspectives*, *Recent Trends and Future Technology in Applied Intelligence*, 2018, σελ. 889–897.

³⁰⁷ Βλ. E. Magrani, *Threats of the internet of things in a techno-regulated society*, *ACM SIGCAS Computers and Society*, Ap. 47, Τεύχος 3, 2017, σελ. 124–138.

³⁰⁸ Βλ. S. Livingston, & M. Risse, *The Future Impact of Artificial Intelligence on Humans and Human Rights, Ethics & International Affairs*, Ap.33, Τεύχος 2, 2019, σελ. 141–158.

³⁰⁹ Βλ. J.R. Geis, A.P. Brady, C.C. Wu, J. Spencer, E. Ranshaert, J.L. Jaremko, & M. Kohli, *όπ.παρ.*, σελ. 329–334, Βλ. επίσης E. Magrani, *όπ.παρ.*, σελ. 124–138.

Ο επανασχεδιασμός του νομοθετικού πλαισίου είναι αναγκαίος, έτσι ώστε η τεχνολογική πρόοδος και εξέλιξη να μην αποτελεί ένα απλό εργαλείο αλλά έναν σοβαρό ρυθμιστή της λήψης αποφάσεων, με ευαισθησία στις συνταγματικές αξίες. Ομιλούμε, επομένως, περί ενός σχεδιασμού ευαίσθητου στην αξία³¹⁰. Η εν λόγω ενέργεια είναι απαραίτητη, τη στιγμή που η σύγχρονη κοινωνία είναι «τεχνο-ρυθμιζόμενη».

³¹⁰ Βλ. C. Cath, Governing artificial intelligence: ethical, legal and technical opportunities and challenges, *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, Αρ.376, Τεύχος 2133, 2018, σελ. 1-8.

3.8 Οι κοινωνικές προκλήσεις της τεχνητής νοημοσύνης

Όπως επισημάνθηκε από τον Brey, οι επιπτώσεις της αξιοποίησης της τεχνητής νοημοσύνης στην κοινωνία είναι άμεσα συναρτημένες με την εκάστοτε επικρατούσα θεωρητική προοπτική³¹¹. Επί παραδείγματι, από τον Philip Meurs³¹² υποστηρίχθηκε η άποψη ότι οι συνέπειες της χρήσης της τεχνολογίας στο κοινωνικό σύνολο είναι απόλυτα εξαρτημένες αφενός από τον ορισμό που αποδίδεται στην τεχνολογία και αφετέρου από τις επιμέρους προοπτικές και διαστάσεις κάτω από τις οποίες αυτή χρησιμοποιείται. Αναλυτικότερα, επισημαίνεται ότι στην περίπτωση των περισσότερο «στενών» ορισμών, συνήθως, η κοινωνική προοπτική είναι διαχωρισμένη από τα δομικά μέρη της τεχνολογίας³¹³. Αντιθέτως, από την άποψη των κοινωνικών θεωριών, συνυπολογίζονται ποικίλοι παράγοντες.

Παραδείγματος χάριν, από την πλευρά της κοινωνικής κονστρουκτιβιστικής θεωρίας γίνεται επίκληση προς τους δομικούς παράγοντες στο να συμβάλλουν στη διαμόρφωση της ανάπτυξης της τεχνολογίας στην ευρύτερη κοινωνία³¹⁴. Η βάση της συγκεκριμένης θεωρίας εδράζεται στο ότι η επίδραση που έχουν οι σχετικές επιρροές κοινωνικής φύσεως σε συνδυασμό με τους δομικούς παράγοντες, από τους οποίους διαμορφώνονται οι κοινωνικές δομές, έχει ιδιαίτερη σημασία στην περίπτωση κατά την οποία γίνεται ακριβής προσδιορισμός των κοινωνικών δομών που μπορούν να επηρεάσουν την τεχνολογική ανάπτυξη³¹⁵.

Πρόκειται για μία θεωρία, η οποία είναι βασισμένη στην κοινωνική κονστρουκτιβιστική έμφαση, η οποία αποδίδεται στον βαθμό στον οποίο είναι μεταξύ τους συναφείς οι παράγοντες και τα αποτελέσματα των δράσεων των κοινωνικών

³¹¹ Βλ. P. Brey, *The Strategic Role of Technology In A Good Society, Technology In Society*, Τεύχος 52, 2018, σελ. 39-45.

³¹² Βλ. P. Meurs, *Technology, Society And Democracy: The Social Impact Of, And Democratic Control Over Technology, With Special Reference To Information Technology And Data Protection*, Ph.D., London School of Economics and Political Science, 1990, σελ. 25.

³¹³ Βλ. P. Meurs, *όπ.παρ.*, σελ. 25.

³¹⁴ Βλ. H.K. Klein & D.L. Kleinman, *The Social Construction Of Technology: Structural Considerations, Science, Technology, & Human Values*, Τεύχος 27, Αρ., 1, 2002, σελ. 28-52.

³¹⁵ Βλ. H.K. Klein & D.L. Kleinman, *όπ.παρ.*, σελ. 28.

φορέων, στις εξελίξεις και την εν γένει πορεία και ανάπτυξη του κλάδου της τεχνολογίας. Στο σημείο αυτό, επισημαίνεται ότι από την πλευρά των Klein & Kleinman³¹⁶ οι υφιστάμενες δομές λαμβάνονται υπόψη ως «κανόνες παιχνιδιού», οι οποίοι εφαρμόζονται υπό συνθήκες, στις οποίες οι «παίκτες» μπορούν είτε να βρίσκονται σε περιορισμό κινήσεων και δράσεων, είτε να κινούνται και να δρουν κάτω από απόλυτη ελευθερία, λαμβάνοντας υπόψη και την υποκείμενη διαμόρφωση ισχύος. Ωστόσο, η έμφαση που αποδίδεται αφορά στις συνέπειες που καταλήγουν να παρουσιάζονται σε διαφορετικά κοινωνικά επίπεδα, τα οποία υπήρχαν ακόμα και πριν την φυσική διαδικασία σχεδιασμού των ανωτέρω αναφερόμενων κανόνων, οι οποίοι μπορούν να σκίσουν επιρροή στην διαδικασία της λήψης αποφάσεων επί ζητημάτων συνυφασμένο με την τεχνολογική ανάπτυξη³¹⁷.

Συν τοις άλλοις, από την συγκεκριμένη θεωρία υποδεικνύεται ο συναφείς χαρακτήρας των γνωστικών δομών από τις οποίες συγκροτείται ένας οργανισμός, η δυναμική που χαρακτηρίζει τις εσωτερικές ομάδες που τον απαρτίζουν, οι πηγές προέλευσης και η ποικιλομορφία των αποδοθεισών εξουσιών, καθώς επίσης και ο βαθμός στον οποίο οι κοινωνικές αξίες είναι θεσμοποιημένες³¹⁸. Υπό το πρίσμα της κοινωνικής κονστρουκτιβιστικής προοπτικής, συνεπώς, προσφέρεται ποικιλομορφία επιμέρους κατηγοριών, από τις οποίες είναι εφικτή η δυνητική ανίχνευση της κοινωνικής επιρροής που υπάρχει στην ευρύτερη διαδικασία διαμόρφωσης της τεχνολογίας και του βαθμού και τρόπου με τον οποίο αυτή αξιοποιείται, συνδυαστικά με το τελικό αποτέλεσμα που αυτή καταλήγει να έχει στο γενικό κοινωνικό σύνολο. Οπότε, στο πλαίσιο της συγκεκριμένης θεωρίας υποστηρίζεται ότι οι δομές της κοινωνίας πρέπει να είναι αναγνωρισμένες εκ των προτέρων, δηλαδή πριν την τελική ανάπτυξη της τεχνολογίας της τεχνητής νοημοσύνης³¹⁹. Ακολούθως, η τεχνολογία τεχνητής νοημοσύνης δεν μπορεί παρά να είναι συμμορφωμένη και ευθυγραμμισμένη με έναν ευεργετικό τρόπο αξιοποίησης της για το κοινό όφελος της κοινωνίας, των πολιτών και των κυβερνήσεων.

³¹⁶ Βλ. H.K. Klein & D.L. Kleinman, *όπ.παρ.*, σελ. 35.

³¹⁷ Βλ. H.K. Klein & D.L. Kleinman, *όπ.παρ.*, σελ. 37.

³¹⁸ Βλ. H.K. Klein & D.L. Kleinman, *όπ.παρ.*, σελ. 28-52.

³¹⁹ Βλ. H.K. Klein & D.L. Kleinman, *όπ.παρ.*, σελ. 28-52.

Αναφορικά με την περίπτωση των κοινωνικών προκλήσεων που προκύπτουν από την αξιοποίηση της τεχνολογίας της τεχνητής νοημοσύνης, η μεροληψία και ο βαθμός λογοδοσίας αντικατοπτρίζουν τα βασικότερα ζητήματα φιλοσοφικών συζητήσεων γύρω από τον συγκεκριμένο προβληματισμό³²⁰. Τη στιγμή κατά την οποία η λειτουργία της τεχνητής νοημοσύνης χαρακτηρίζεται από υψηλά επίπεδα αυτονομίας, το βασικότερο ερώτημα που τίθεται και το οποίο αφορά στην πτυχή της λογοδοσίας, έγκειται στο ποιος είναι τελικά ο νομικά και ηθικά υπεύθυνος σε περίπτωση που προκληθεί ζημιά από την χρήση της τεχνητής νοημοσύνης³²¹. Πρόκειται για ένα ερώτημα που τέθηκε και στο δεύτερο κεφάλαιο της διατριβής αυτής, όπου όμως προτείνουμε τη μη χρήση των συστημάτων τεχνητής νοημοσύνης σε γενικευμένες καταστάσεις³²², αλλά σε εξειδικευμένες καταστάσεις και κατόπιν της έκδοσης σχετικής άδειας από την αρμόδια αστυνομική και δικαστική αρχή³²³. Όσον αφορά το δεύτερο ζήτημα που τίθεται και είναι σχετικό με την αξιοποίηση της τεχνητής νοημοσύνης, δηλαδή το ζήτημα της μεροληψίας, στην βιβλιογραφία αναφέρεται ότι οι προκαταλήψεις αποτελούν την συνέπεια της αξιοποίησης συστημάτων τεχνητής νοημοσύνης και μηχανικής μάθησης³²⁴.

Σε επίπεδο κοινωνίας, η προκατάληψη αποτελεί μία ήδη γνωστή έννοια, όμως σε επίπεδο αξιοποίησης της τεχνολογίας, η προκατάληψη είναι μία έννοια άμεσα συνυφασμένη με την αξιοποίηση συστημάτων τεχνητής νοημοσύνης³²⁵. Στο σημείο αυτό, πρέπει να τονιστεί ότι οι προκαταλήψεις συνοδεύονται από κοινωνικούς

³²⁰ Βλ. FLI, AI Policy Challenges And Recommendations. AI Policy. Future of Life Institute, 2019^a, στο <https://futureoflife.org/ai-policy-challenges-and-recommendations/>

³²¹ Βλ. FLI, όπ.παρ., στο <https://futureoflife.org/ai-policy-challenges-and-recommendations/>

³²² Όπως παραδείγματος χάριν στην ταυτοποίηση των ανθρώπων για τη λήψη επιτόπιων αποφάσεων ή σε περιπτώσεις που χρησιμοποιούνται τρωτά και ευαίσθητα σημεία των κοινωνικά ευπαθών ομάδων ή επηρεάζεται το ανθρώπινο υποσυνείδητο και ωθείται στη διαστρέβλωση της ανθρώπινης συμπεριφοράς υπέρ ή κατά συγκεκριμένων ανθρώπων

³²³ Όπως παραδείγματος χάριν σε περιπτώσεις ταυτοποίησης και ανίχνευσης των χαρακτηριστικών στοιχείων των ανθρώπων για τον εύκολο και άμεσο εντοπισμό εγκληματικών στοιχείων και καταζητούμενων.

³²⁴ Βλ. M. Tegmark, Life 3.0: Being Human in The Age Of Artificial Intelligence, London, Penguin Books, 2017, σελ. 321.

³²⁵ Βλ. K. Crawford, The Trouble With Bias – NIPS, Video, 2017, στο <https://www.youtube.com/watch?v=6Uao14eIyGc>

ορισμούς οι οποίοι είναι μεταξύ τους αντικρουόμενοι. Ωστόσο, όταν ομιλούμε για το πεδίο της τεχνητής νοημοσύνης, οι προκαταλήψεις και μεροληψίες είναι συνδεδεμένες με ζητήματα ορθολογικών διακρίσεων που αφορούν την διαδικασία της παραγωγής και του προγραμματισμού των χρησιμοποιούμενων συστημάτων τεχνητής νοημοσύνης³²⁶.

Επιπροσθέτως, επισημαίνεται ότι το ζήτημα της μεροληψίας, όταν αναφέρεται σε δεδομένα εκπαίδευσης, δύναται να οδηγήσει στην ύπαρξη μίας «λογικής διάκρισης». Ο λόγος έγκειται στο ότι η λειτουργία των αυτόνομων τεχνολογικών συστημάτων τεχνητής νοημοσύνης δεν χρειάζεται την παρέμβαση του ανθρώπινου παράγοντα. Αυτό καταδεικνύει ότι τέτοιου τύπου τεχνολογικά συστήματα μπορούν να προχωρήσουν στην λήψη αποφάσεων και μέτρων, σύμφωνα με τον προγραμματισμό και τον αλγόριθμο που διαθέτουν, ακόμη και αν χαρακτηρίζονται από μηδενικά επίπεδα προκατάληψης³²⁷. Προκειμένου να καταστεί αυτή η διατύπωση πιο κατανοητή, δίνεται το παράδειγμα των αλγοριθμικών προκαταλήψεων που υπάρχουν με βάση το φύλο ή την αναγνώριση προσώπου των πολιτών μίας κοινωνίας³²⁸.

Μία από τις βασικότερες αιτίες της αλγοριθμικής μεροληψίας εντοπίζεται στα δεδομένα, με βάση τα οποία γίνεται η εκπαίδευση των αυτόνομων τεχνολογικών συστημάτων τεχνητής νοημοσύνης. Για παράδειγμα, η αλγοριθμική μεροληψία συναρτάται με τον βαθμό στον οποίο εντοπίζονται ή καταχωρίζονται ελλιπή δεδομένα, καθώς επίσης και με τον βαθμό στον οποίο δεν επικρατεί διαφάνεια και κοινωνική αμεροληψία στη διαδικασία της καταχώρισης των δεδομένων στα εν λόγω τεχνολογικά συστήματα³²⁹. Συνακόλουθα, ελλοχεύει ο κίνδυνος της καταχώρισης κοινωνικά μεροληπτικών δεδομένων. Όπως επισημάνθηκε από την Crawford³³⁰, η μεροληψία, όταν συνδέεται με το πεδίο της τεχνητής νοημοσύνης δεν είναι απόλυτα ξεκαθαρισμένη, λόγω του ότι τα εν λόγω τεχνολογικά συστήματα επηρεάζουν σε ολόένα και μεγαλύτερο βαθμό την καθημερινότητα των πολιτών, ενώ παράλληλα

³²⁶ Βλ. K. Crawford, όπ.παρ., στο <https://www.youtube.com/watch?v=6Uao14eIyGc>, Βλ. επίσης D. Danks, & A.J. London, Algorithmic Bias In Autonomous Systems. Proceedings Of The Twenty-Sixth International Joint Conference On Artificial Intelligence (IJCAI-17), 2017, σελ. 4691-4698.

³²⁷ Βλ. D. Danks, & A.J. London, όπ.παρ., σελ. 4691-4698.

³²⁸ Βλ. K. Crawford, όπ.παρ., στο <https://www.youtube.com/watch?v=6Uao14eIyGc>

³²⁹ Βλ. K. Crawford, όπ.παρ., στο <https://www.youtube.com/watch?v=6Uao14eIyGc>

³³⁰ Βλ. K. Crawford, όπ.παρ., στο <https://www.youtube.com/watch?v=6Uao14eIyGc>

ασκούν επιρροή ακόμα και σε ζητήματα κοινωνικού κόστους, σε περίπτωση που σημειωθεί κάποιο σφάλμα ή κάποια παράλειψη που επηρεάζει τον βαθμό μεροληψίας των καταχωρισμένων δεδομένων.

Πρόκειται για ένα σύνολο ηθικών ζητημάτων που συνάδουν με την κοινωνική αξιοποίηση της τεχνητής νοημοσύνης και τα οποία προκύπτουν ακόμα και από την αξιοποίηση της πλειοψηφίας των σύγχρονων και προηγμένων τεχνολογικών συστημάτων, ακόμη και αν δεν εμπίπτουν στο πεδίο της τεχνητής νοημοσύνης. Ωστόσο, πρέπει να αναφερθεί ένα βασικό σημείο στο οποίο διαφοροποιούνται τα προκύπτοντα ηθικά ζητήματα από την πλευρά των συστημάτων τεχνητής νοημοσύνης και από την πλευρά των λοιπών προηγμένων τεχνολογιών. Αναλυτικότερα, όπως επισημάνθηκε από τους Campolo et al.³³¹, η μεγαλύτερη διαφορά έγκειται στην ικανότητα που έχουν οι τεχνολογίες τεχνητής νοημοσύνης να λαμβάνουν αυτόνομες αποφάσεις που σχετίζονται με την καθημερινή ανθρώπινη διαβίωση και δραστηριότητα³³².

Ακολούθως, επανέρχεται στο προσκήνιο ένα παρεμφερές ηθικής φύσεως ερώτημα, το οποίο αναλύθηκε στο δεύτερο κεφάλαιο της εν λόγω διατριβής. Το ερώτημα αυτό αφορά τον βαθμό στον οποίο θα πρέπει να καθίσταται επιτρεπτό στα μηχανήματα τεχνητής νοημοσύνης να λαμβάνουν αποφάσεις αυτόνομα³³³ και τον βαθμό στον οποίο θα πρέπει να καταστεί απαραίτητη και αναγκαία η ανθρώπινη εμπλοκή³³⁴.

Οι West, Whittaker και Crawford³³⁵, επεσήμαναν ότι οι πρόσφατες συζητήσεις που διεξάγονται γύρω από το ζήτημα της ηθικής διάστασης της χρήσης μηχανών τεχνητής νοημοσύνης σε κοινωνικά ζητήματα, λαμβάνουν υπόψη τους παράγοντες της ευθυγράμμισης των κοινωνικών αξιών με τη λειτουργία των μηχανών αυτών και της απουσίας κοινωνικών διακρίσεων στις αποφάσεις που λαμβάνονται από τα εν λόγω

³³¹ Βλ. A. Campolo, M. Sanfilippo, M. Whittaker & K. Crawford, AI Now 2017 Report, New York, AI Now Institute, 2017, σελ. 30.

³³² Βλ. N. Malli, M. Jacobs & S. Villeneuve, Intro To AI For Policymakers: Understanding The Shift. March 2018 AI + Public Policy Conference, Ontario, Brookfield Institute, 2018, σελ. 9.

³³³ Βλ. M. Tegmark, *όπ.παρ.*, σελ. 9.

³³⁴ Βλ. A. Campolo, M. Sanfilippo, M. Whittaker & K. Crawford, *όπ.παρ.*, σελ. 30

³³⁵ Βλ. S. West, M. Whittaker, & K. Crawford, Discriminating Systems: Gender, Race, And Power In AI, New York, AI Now Institute, 2019, σελ. 6.

τεχνολογικά συστήματα. Καταδεικνύεται, λοιπόν, ότι αυτά καθ' αυτά τα τεχνολογικά περιβάλλοντα τεχνητής νοημοσύνης δεν πρωταγωνιστούν σε τέτοιου είδους δεοντολογικούς προβληματισμούς. Όσον αφορά την παρουσία ποικιλομορφίας, οι αδυναμίες και τα μελανά σημεία της χρήσης συστημάτων τεχνητής νοημοσύνης αφορούν τη δυναμική τους. Πρόκειται για μία δυναμική που απορρέει αφενός από θέματα κοινωνικών διακρίσεων που προκύπτουν από τη χρήση συστημάτων τεχνητής νοημοσύνης και αφετέρου από την αναγκαιότητα του σωστού προγραμματισμού των εν λόγω συστημάτων³³⁶.

Στην ίδια μελέτη³³⁷ δίνεται έμφαση στην περίπτωση των ομοιογενών κοινωνικών περιβαλλόντων, εντός των οποίων γίνεται η ανάπτυξη των συστημάτων τεχνητής νοημοσύνης. Πιο συγκεκριμένα, η μελέτη αναφέρει ότι στην πλειοψηφία των περιπτώσεων, τα συστήματα τεχνητής νοημοσύνης αναπτύσσονται από προγραμματιστές ανδρικού φύλου και λευκού χρώματος δέρματος, οι οποίοι αποτελούν μέλη του ανθρώπινου δυναμικού κολοσσιαίων εταιρειών τεχνολογίας, γεγονός, το οποίο είναι πολύ πιθανό να εκδηλώσει ζήτημα κοινωνικών διακρίσεων κατά τον προγραμματισμό και την εκπαίδευση των τεχνολογικών συστημάτων³³⁸.

Η ομοιογένεια των κοινωνικών και δημογραφικών χαρακτηριστικών των μηχανικών και προγραμματιστών των συστημάτων τεχνητής νοημοσύνης επιφέρει ως άμεση απότοκο την παρουσία μη πλήρων ή παραποιημένων δεδομένων αναφορικά με το ποιος μπορεί να θεωρείται από το μηχάνημα τεχνητής νοημοσύνης ως «κανονικός άνθρωπος»³³⁹. Ο λόγος της δημιουργίας αυτού του προβλήματος έγκειται στο ότι οι κατασκευαστές των εν λόγω μηχανών και οι πεποιθήσεις από τις οποίες διακατέχονται επί κοινωνικών ζητημάτων μπορεί να μην ευθυγραμμίζονται με ζητήματα ποικιλομορφίας μεταξύ των τελικών χρηστών των μηχανών της τεχνητής νοημοσύνης³⁴⁰.

³³⁶ Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 6.

³³⁷ Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 8.

³³⁸ Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 7.

³³⁹ Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 7.

³⁴⁰ Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 8.

Οι ανωτέρω αναφερόμενες κοινωνικές διακρίσεις είναι επίσης ορατές στη φύση από την οποία διέπονται τα συστήματα τεχνολογίας τεχνητής νοημοσύνης. Πιο συγκεκριμένα, οι μηχανές τεχνητής νοημοσύνης «μαθαίνουν» να προβαίνουν σε διακρίσεις στους τύπους δεδομένων που διαχειρίζονται³⁴¹. Η μελέτη των West, Whittaker και Crawford³⁴² προτρέπει τη διαδικασία επανεκτίμησης *«της χρήσης συστημάτων τεχνητής νοημοσύνης για την ταξινόμηση, τον εντοπισμό και την πρόβλεψη της φυλής και του φύλου»*³⁴³. Πρόκειται για μία πρόταση, η οποία θεωρούμε ότι αν τύχει εφαρμογής, μπορεί να οδηγήσει στην καταπολέμηση ενός προβλήματος, στο οποίο ο επιστημονικός και ακαδημαϊκός κόσμος, αλλά και οι αρμόδιοι για την χάραξη πολιτικής επί της χρήσης των μηχανών τεχνητής νοημοσύνης στην κοινωνία, δεν έχουν δώσει τόσο μεγάλη έμφαση.

Όσον αφορά την κοινωνική αξιοποίηση της τεχνητής νοημοσύνης, το ζήτημα της ασφάλειας πρέπει να λαμβάνεται υπόψη από το σύνολο των εμπλεκόμενων μερών, ως η μεγαλύτερη πρόκληση του σταδίου του σχεδιασμού, του προγραμματισμού και της ρύθμισης των μηχανών τεχνητής νοημοσύνης³⁴⁴.

Η μελέτη των Richardson, Schultz και Crawford³⁴⁵ απέδειξε τον αρνητικό αντίκτυπο της τάσης των κυβερνήσεων να επιδιώκουν την εφαρμογή διαδικασιών χάραξης πολιτικής και λήψης αποφάσεων έχοντας ως κεντρικό άξονα τα δεδομένα και την αξιοποίηση μηχανημάτων τεχνητής νοημοσύνης. Η εν λόγω μελέτη επισημαίνει την απουσία αποτελεσματικής λύσης για να αντιμετωπιστεί το πρόβλημα των *«βρώμικων δεδομένων»*³⁴⁶. Ομοίως, αναφέρει την απουσία κάποιας τεχνολογίας, από την οποία θα είναι εφικτή η διαδικασία ανίχνευσης και αντιμετώπισης των ελλιπών ή λανθασμένων δεδομένων³⁴⁷.

³⁴¹ Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 6.

³⁴² Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 6

³⁴³ Βλ. S. West, M. Whittaker, & K. Crawford, *όπ.παρ.*, σελ. 3.

³⁴⁴ Βλ. N. Malli, M. Jacobs & S. Villeneuve, *όπ.παρ.*, σελ. 11.

³⁴⁵ Βλ. R. Richardson, J. Schultz & K. Crawford, *όπ.παρ.*, 2019, 225.

³⁴⁶ Βλ. W. Kim, B. Choi, E. Hong, S. Kim & L. Doheon, *όπ.παρ.*, σελ. 81-99.

³⁴⁷ Βλ. W. Kim, B. Choi, E. Hong, S. Kim & L. Doheon, *όπ.παρ.*, σελ. 81-99.

3.9 Εθνικές προσπάθειες τεχνολογικής προόδου

Το εν λόγω υποκεφάλαιο αποσκοπεί στη συζήτηση επί του ζητήματος των καταβαλλόμενων εθνικών προσπαθειών για την αξιοποίηση τόσο μηχανών τεχνητής νοημοσύνης όσο και Τεχνολογιών Γενικού Σκοπού (GPTs). Συγκεκριμένα, παρουσιάζεται μία εκτενής επισκόπηση μίας σειράς πρόσφατων εθνικών εξελίξεων, οι οποίες είναι συνυφασμένες με το ζήτημα της ανάπτυξης της τεχνολογίας της τεχνητής νοημοσύνης, καθώς επίσης και με θεωρητικές προσεγγίσεις που έχουν πραγματοποιηθεί γύρω από το πεδίο της Επιστήμης και της Τεχνολογίας και με την τεχνική κατανόηση των ζητημάτων αυτών σε κυβερνητικό επίπεδο.

Αναλυτικότερα, από το έτος 2016 και μετά, οι εθνικές κυβερνήσεις προσπαθούν να εξελίσσουν τον τρόπο λειτουργίας του δημοσίου τομέα, προσβλέποντας στην ευθυγράμμιση του με τον τρόπο λειτουργίας του ιδιωτικού και μία εξ αυτών των προσπαθειών αντικατοπτρίζεται στη χρήση συστημάτων τεχνητής νοημοσύνης σε κυβερνητικό επίπεδο³⁴⁸. Τα εν λόγω συστήματα χρησιμεύουν επί ζητημάτων δημοσίευσης αναφορών επικοινωνιών και αξιολογήσεων που συνάδουν με την μελλοντική ανάπτυξη και πορεία των τεχνολογιών τεχνητής νοημοσύνης. Σε αυτήν την προσπάθεια πρωτοστατεί σε παγκόσμιο επίπεδο η Κίνα, η οποία μάλιστα προέβη στη δημοσίευση της εθνικής της στρατηγικής γύρω από το ζήτημα της τεχνητής νοημοσύνης, η οποία φέρει τίτλο «*Made in China 2025*»³⁴⁹.

Κατά τη διάρκεια του 2016, η κινεζική κυβέρνηση προχώρησε στη δημοσίευση της λεγόμενης «*Τριετούς Καθοδήγησης για το Σχέδιο Τεχνητής Νοημοσύνης Internet Plus (2016-2018)*», η οποία συνοδεύτηκε από το «*Τριετές σχέδιο δράσης για την προώθηση της ανάπτυξης μιας βιομηχανίας τεχνητής νοημοσύνης νέας γενιάς (2018– 2020)*»³⁵⁰. Αμφότερες οι δημοσιεύσεις θα μπορούσαν να χαρακτηριστούν ως ξεκάθαρες προθέσεις της κινεζικής κυβέρνησης για την ταχεία παροχή και γνωστοποίηση

³⁴⁸ Βλ. G.C. Allen, *Understanding China's AI Strategy: Clues To Chinese Strategic Thinking On Artificial Intelligence And National Security*, Washington, CNASDC, 2019, σελ. 4

³⁴⁹ Βλ. G.C. Allen, *όπ.παρ.*, σελ. 4.

³⁵⁰ Βλ. T. Dutton, *An Overview Of National AI Strategies*. Medium, Politics + AI, 2018, στο <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

κατευθυντήριων γραμμών και οδηγιών³⁵¹, αναφορικά με ζητήματα έναρξης ισχύος μίας «από πάνω προς τα κάτω» προσέγγισης στον τομέα της ανάπτυξης της τεχνολογίας της τεχνητής νοημοσύνης στην Κίνα³⁵².

Όσον αφορά στην περίπτωση της Αμερικής, κατά το έτος 2016, υπό την προεδρεία Ομπάμα, δημοσιεύτηκαν συνολικά τρεις εκθέσεις, οι οποίες εσώκλειαν τον ορισμό, την περιγραφή της επικρατούσας κατάστασης και την παρουσία ευκαιριών έρευνας και ανάπτυξης γύρω από το πεδίο της τεχνητής νοημοσύνης. Πρόκειται για τις δημοσιεύσεις που έφεραν τους εξής τίτλους: «*Προετοιμασία για το μέλλον της τεχνητής νοημοσύνης*», «*Το Εθνικό Στρατηγικό Σχέδιο Έρευνας και Ανάπτυξης Τεχνητής Νοημοσύνης*» και «*Τεχνητή Νοημοσύνη, Αυτοματισμός και Οικονομία*»³⁵³.

Τα ανωτέρω συγκεκριμένα δημοσιεύματα εξ Αμερικής, αναδύουν μία προσέγγιση που χαρακτηρίζεται από υψηλά επίπεδα φιλελευθερισμού και η οποία είναι προσανατολισμένη στην αγορά. Αυτό συμβαίνει ειδικότερα κατά τον καθορισμό του κυβερνητικού ρόλου επί ζητημάτων ανάπτυξης μηχανών τεχνητής νοημοσύνης³⁵⁴.

Επιπρόσθετες περιπτώσεις κρατών, όπως είναι παραδείγματος χάριν ο Καναδάς, η Σιγκαπούρη, η Ιαπωνία και η Φινλανδία, κατά το έτος 2017, εντάσσονται στις πρώτες περιπτώσεις από τις οποίες έγιναν κυβερνητικές δημοσιεύσεις των αντίστοιχων εθνικών στρατηγικών επί της αξιοποίησης και ανάπτυξης της τεχνητής νοημοσύνης. Αυτό διενεργήθηκε στο πλαίσιο μίας συντονισμένης προσπάθειας προώθησης και καθοδήγησης της ανάπτυξης αξιόπιστων συστημάτων τεχνητής νοημοσύνης³⁵⁵.

³⁵¹ Βλ. J. Ding, *Deciphering China's AI Dream. Governance Of AI Program*, Oxford, University of Oxford, 2018, σελ. 8.

³⁵² Βλ. FLI, *AI Policy - China - Future Of Life Institute*. Future Of Life Institute, 2019, στο <https://futureoflife.org/ai-policy/>

³⁵³ Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo & L. Floridi, *όπ.παρ.*, σελ. 505–528, Βλ. επίσης A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, *Artificial General Intelligence: Coordination & Great Powers*. 2018 Foresight Institute Strategy Meeting On AGI, Palo Alto, Foresight Institute, 2018, σελ. 16

³⁵⁴ Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo & L. Floridi, *όπ.παρ.*, σελ. 508.

³⁵⁵ Βλ. T. Dutton, *όπ.παρ.*, στο <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

Επιπλέον κράτη, κατά το έτος 2018, προσχώρησαν στην ανωτέρω αναφερόμενη τάση. Σε αυτές τις περιπτώσεις συμπεριλαμβάνονται και διεθνείς οργανισμοί, όπως είναι επί παραδείγματι ο ΟΟΣΑ και ο ΟΗΕ³⁵⁶. Σε συνδυασμό με τις ανωτέρω αναφερόμενες κυβερνητικές στρατηγικές, η ακαδημαϊκή κοινότητα επί του θέματος της τεχνητής νοημοσύνης, αλλά και λοιποί μη κυβερνητικοί οργανισμοί, προέβησαν στη δημοσίευση μίας σειράς αρχών, αποσκοπώντας στην ανάπτυξη του κλάδου της τεχνητής νοημοσύνης, με τέτοιον τρόπο που να είναι επωφελής για την ανθρωπότητα και να μην υπάρχει απόκλιση από τις ανθρώπινες και κοινωνικές αρχές και αξίες³⁵⁷.

Κατά το έτος 2019, αναπτύχθηκαν από τον ΟΟΣΑ οι πρώτες διεθνείς αρχές, οι οποίες προέβλεπαν σαφείς κατευθύνσεις και οδηγίες που έπρεπε να ακολουθηθούν για λόγους αξιοπιστίας ανάπτυξης της τεχνητής νοημοσύνης σε διεθνές βεληνεκές³⁵⁸. Ο ταχύς ρυθμός των πρόσφατων δημοσιευμένων εθνικών στρατηγικών, σχετικών με το πεδίο της τεχνητής νοημοσύνης, αποδεικνύει την έντονη προθυμία των κρατών να συμμετάσχουν σε έναν διεθνή αγώνα επί του θέματος της ηγεσίας της τεχνητής νοημοσύνης και του ολοένα και εντεινόμενου επιστημονικού ανταγωνισμού επί πολιτικών που συνάδουν με τη χρήση των μηχανών τεχνητής νοημοσύνης³⁵⁹.

Η Ελλάδα, προσανατολισμένη στη διαχείριση και εδραίωση του ψηφιακού μετασχηματισμού που απορρέει από τη χρήση των συστημάτων τεχνητής νοημοσύνης, προχώρησε το 2022 στη θέσπιση του ν.4961. Ο συγκεκριμένος νόμος θεσπίζει υποχρεώσεις τόσο για φορείς του δημοσίου όσο και για ιδιωτικούς, οι οποίοι δραστηριοποιούνται από το στάδιο της παραγωγής, της διανομής έως και της χρήσης των νέων τεχνολογιών. Οι ρυθμίσεις του νόμου στο Μέρος Α και Β, αναφέρονται στη ψηφιακή αναβάθμιση της δημόσιας διοίκησης και στην αξιοποίηση προηγμένων τεχνολογιών, αντίστοιχα. Είναι σημαντικό να αναφερθεί ότι οι ρυθμίσεις του νόμου

³⁵⁶Βλ. T. Dutton, όπ.παρ., στο <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>

³⁵⁷Βλ. FLI, όπ.παρ., στο: <https://futureoflife.org/ai-policy-challenges-and-recommendations/>. Βλ. επίσης FLI, όπ.παρ., στο <https://futureoflife.org/ai-policy/>

³⁵⁸ Βλ. OECD, Artificial Intelligence In Society, Paris, OECD Publishing, 2019, στο <https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm>

³⁵⁹ Βλ. OECD, όπ.παρ., στο <https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm>

δεν θίγουν τα δικαιώματα και τις υποχρεώσεις που προβλέπονται στον Γενικό Κανονισμό Προστασίας Δεδομένων³⁶⁰.

Προ της διερεύνησης των θεωριών που είναι συνυφασμένες με τον τρόπο με τον οποίο η τεχνολογία καθίσταται κατανοητή μεταξύ των κυβερνήσεων, κρίνεται σημαντική η ανασκόπηση της υφιστάμενης βιβλιογραφίας, από την οποία αντλούνται εξηγήσεις και ερμηνείες για τις πρόσφατες κυβερνητικές προσπάθειες πάνω στο θέμα της διαχείρισης της ανάπτυξης της τεχνολογίας της τεχνητής νοημοσύνης, η οποία σχετίζεται και με κοινωνικά ζητήματα. Η μελέτη των Hagendijk και Irwin³⁶¹ προέβη στην εξέταση της δημόσιας συζήτησης και διακυβέρνησης επί ζητημάτων επιστήμης και ανάπτυξης της τεχνολογίας. Από τους εν λόγω μελετητές εντοπίστηκαν τάσεις, από τις οποίες χαρακτηρίζεται η ευρωπαϊκή διακυβέρνηση σε ζητήματα επιστήμης και τεχνολογικής ανάπτυξης³⁶².

Οι παρατηρήσεις που απορρέουν από τη μελέτη αυτήν, υποστηρίζουν ότι επικρατούν ολοένα και αυξανόμενες κυβερνητικές τάσεις, στις οποίες η λειτουργία δημοσίων πλατφορμών ανατίθεται σε εμπειρογνώμονες επί ζητημάτων τεχνολογίας και επιστήμης. Συν τοις άλλοις, επικρατεί η τάση έναρξης δημοσίων διαβουλεύσεων γύρω από ζητήματα πολιτικών διαδικασιών. Ωστόσο, η δημόσια εμπλοκή, στο σημείο αυτό, φαίνεται να είναι οριοθετημένη σε συζητήσεις ηθικής φύσεως και σε πρωταρχικές φάσεις χάραξης πολιτικής³⁶³.

Εντούτοις, τα ανωτέρω ευρήματα δεν ευθυγραμμίζονται με την περίπτωση της Αμερικής και τις τεχνολογικές πολιτικές της. Η τοποθέτηση της αμερικανικής κυβέρνησης για τον περιορισμένο ρόλο της στην ανάπτυξη της τεχνητής νοημοσύνης διαφοροποιείται και πρεσβεύει την άποψη ότι οι υπεύθυνοι χάραξης πολιτικής θα είναι σε θέση να αντιμετωπίσουν τις τεχνολογικές προκλήσεις, μέσω συγκεκριμένων

³⁶⁰ Βλ. Φερενίκη Παναγοπούλου-Κουτνατζή, όπ.παρ., σελ. 144-148.

³⁶¹ Βλ. R. Hagendijk & A. Irwin, Public Deliberation And Governance: Engaging With Science And Technology In Contemporary Europe, Minerva, Τεύχος 44, Αρ. 2, 2006, σελ. 167-184.

³⁶² Βλ. R. Hagendijk & A. Irwin, όπ.παρ., σελ. 167-184.

³⁶³ Βλ. R. Hagendijk & A. Irwin, όπ.παρ., σελ. 167-184.

δημοσιευμένων κανονισμών και ενεργειών στις οποίες θα προβεί η αμερικανική κυβέρνηση³⁶⁴.

Ταυτόχρονα με το ζήτημα της ολοένα και αυξανόμενης συμμετοχής εμπειρογνομόνων για την υιοθέτηση και αξιοποίηση της προηγμένης τεχνολογίας στην καθημερινή λειτουργία των δημόσιων και κοινωνικών φορέων και οργανισμών, η μελέτη που διεξήχθη από τους Hagendijk και Irwin³⁶⁵ επεσήμανε ένα ακόμα σημαντικό ζήτημα. Πρόκειται για την απόδοση ιδιαίτερης έμφασης στον διεθνή ανταγωνισμό που υφίσταται επί του θέματος της διακυβέρνησης της τεχνολογίας, ειδικότερα τη στιγμή κατά την οποία οι δημόσιες διαβουλεύσεις λαμβάνονται υπόψη ως εμπόδια στη χάραξη πολιτικής.

Από τα ανωτέρω δύναται να εξαχθεί το συμπέρασμα ότι οι καταβαλλόμενες προσπάθειες των ευρωπαϊκών κυβερνήσεων για την επίτευξη δημόσιας δέσμευσης είναι προσοδοφόρες σε περιπτώσεις κατά τις οποίες οι κυβερνήσεις προβαίνουν στη δήλωση σαφών και λεπτομερών δεσμεύσεων για την υιοθέτηση μέτρων και την αξιοποίηση των αποτελεσμάτων των δημοσίων διαβουλεύσεων. Επίσης, είναι προσοδοφόρες όταν οι κυβερνήσεις δεν προβαίνουν σε διαχωρισμό των θεμάτων επιστημονικού και δημοσίου ενδιαφέροντος στις διεξαγόμενες δημόσιες διαβουλεύσεις. Τέλος, είναι προσοδοφόρες όταν οι κυβερνήσεις περιλαμβάνουν το σύνολο των σχετικών διεθνών δεσμεύσεών τους και τα αντίστοιχα πλαίσιά τους³⁶⁶.

Στο σημείο αυτό, επισημαίνεται ότι η μελέτη των Hagendijk και Irwin³⁶⁷ καταλήγει στο ότι από την πλευρά των ευρωπαϊκών κυβερνήσεων επικρατεί η τάση διαχωρισμού των επιστημονικών ζητημάτων από τα ηθικά ζητήματα κατά τη διάρκεια των επίσημων ακολουθούμενων διεργασιών χάραξης πολιτικής³⁶⁸. Μάλιστα, θεωρούμε ότι ο εν λόγω διαχωρισμός θεμάτων δημοσίου και επιστημονικού ενδιαφέροντος καταλήγει να είναι περιοριστικός στην απόλυτη επιρροή των μη κυβερνητικών οργανισμών και του

³⁶⁴ Βλ. A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, *όπ.παρ.*, σελ. 16.

³⁶⁵ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 175.

³⁶⁶ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 182.

³⁶⁷ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 182.

³⁶⁸ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 182.

πολιτικού λόγου στη χάραξη πολιτικής επί ζητημάτων χρηστής και ηθικής διαχείρισης και αξιοποίησης της τεχνολογίας και των μηχανών τεχνητής νοημοσύνης.

Σε γενικές γραμμές, σε ευρωπαϊκό βεληνεκές, είναι σχετικά μικρός ο αριθμός των ειδημόνων που λαμβάνει μέρος σε δημόσιες διαβουλεύσεις επί ζητημάτων διαχείρισης και αξιοποίησης της τεχνολογίας σε επίπεδο κοινωνίας και κυβέρνησης³⁶⁹. Ωστόσο, τόσο οι ευρωπαϊκές όσο και οι βρετανικές πολιτικές και στρατηγικές τίθενται υπέρ της εφαρμογής του συστήματος των διαβουλευτικών δημοκρατικών διαδικασιών, επί του πλαισίου των ακολουθούμενων πολιτικών σε ζητήματα έρευνας και τεχνολογίας³⁷⁰.

Χαρακτηριστικό παράδειγμα είναι το ότι η κυβέρνηση του Ηνωμένου Βασιλείου ανέκαθεν διακατεχόταν από μία κουλτούρα δημόσιας δέσμευσης. Όμως, αυτή η κουλτούρα αφορούσε αμιγώς την περίπτωση θεμάτων που ήταν συνυφασμένα με την θέσπιση κανονισμών και την απόδοση οδηγιών για την αξιοποίηση της τεχνολογίας, παρά το ότι η πλειοψηφία των πολιτικών της ήταν με τέτοιο τρόπο σχεδιασμένες, ώστε να εξυπηρετούσαν τον ιδιωτικό τομέα και τον κλάδο της βιομηχανίας³⁷¹.

Η κρίση του ιού των «τρελών αγελάδων»³⁷² ανάγκασε τη βρετανική κυβέρνηση να εκτιμήσει τη σπουδαιότητα της δημόσιας συμμετοχής στις πολιτικές που χαράσσονται επί του ζητήματος της διαχείρισης των τεχνολογικών πόρων και της τεχνητής νοημοσύνης επί ζητημάτων δημοσίου ενδιαφέροντος, λόγω της σημαντικής επιρροής που άσκησε ο έντονος φόβος στην κοινωνία των πολιτών³⁷³.

Επιπροσθέτως, επισημαίνεται ότι από την πλευρά των υπευθύνων χάραξης πολιτικής της Ευρωπαϊκής Επιτροπής υιοθετήθηκαν παρεμφερείς απόψεις και εκτιμήθηκε η δημόσια συμμετοχή επί ζητημάτων διαχείρισης της τεχνολογίας στη δημόσια διοίκηση. Αυτό, πιο συγκεκριμένα, συνέβη κατά τη διάρκεια της κρίσης των γενετικών μεταλλάξεων³⁷⁴ και οι πολίτες ένιωσαν αποκλεισμένοι από τις πολιτικές εξελίξεις που

³⁶⁹ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 175.

³⁷⁰ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 175.

³⁷¹ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 179.

³⁷² Στη βιβλιογραφία αναφέρεται με τα ακρωνύμια BSE (Bovine Spongiform Encephalopathy)

³⁷³ Βλ. R. Hagendijk & A. Irwin, *όπ.παρ.*, σελ. 190.

³⁷⁴ Στη βιβλιογραφία αναφέρονται ως Genetically Modified Organism (GMO).

απασχολούσαν τους ιδίους και τη διαβίωσή τους³⁷⁵. Στο σημείο αυτό, επισημαίνεται ότι το επιχείρημα της Ευρωπαϊκής Επιτροπής υπέρ της εντατικοποίησης της συμμετοχής του κοινού σε αποφάσεις και πολιτικές επί ζητημάτων χρήσης της τεχνολογίας και της τεχνητής νοημοσύνης ήταν ότι η τεχνολογία έχει πλέον καταστεί αναπόσπαστο σκέλος της κοινωνίας των πολιτών. Αυτό, μάλιστα, πρέπει να λαμβάνεται υπόψη στο σύνολο των ακολουθούμενων διαδικασιών κατά τη διάρκεια της χάραξης στρατηγικών και πολιτικών³⁷⁶.

³⁷⁵ Βλ. S. Arnaldi, G. Quaglio, M. Ladikas, H. O'Kane, T. Karapiperis, K. Srinivas & Y. Zhao, Responsible Governance In Science And Technology Policy: Reflections From Europe, China And India, Technology In Society, Τεύχος 42, 2015, σελ. 81-92.

³⁷⁶ Βλ. S. Arnaldi, G. Quaglio, M. Ladikas, H. O'Kane, T. Karapiperis, K. Srinivas & Y. Zhao, όπ.παρ., σελ. 85.

3.10 Η κατανόηση της τεχνολογίας

Στην υποενότητα αυτή διενεργείται μία ανασκόπηση των συζητήσεων που εκτυλίσσονται στον ακαδημαϊκό κόσμο επί του ζητήματος της κατανόησης της τεχνολογίας στο πλαίσιο της διακυβέρνησης και των πολιτικών που χαράζονται επί της αξιοποίησης της τεχνητής νοημοσύνης. Θα αναπτυχθούν οι επικρατούσες θεωρητικές έννοιες και προοπτικές, ενώ θα εξεταστεί σε θεωρητικό πλαίσιο το επίπεδο και η πτυχή της κατανόησης της τεχνολογίας μεταξύ των διαφορετικών εθνών.

Κατά τη διάρκεια μίας εποχής, όπου ένας μικρός αριθμός πολυεθνικών οργανισμών ασκεί τον βασικό έλεγχο του συνόλου των μελλοντικών -βραχυπρόθεσμων και μακροπρόθεσμων- εξελίξεων της τεχνολογικής ανάπτυξης, είναι επιτακτική ανάγκη τα λαμβανόμενα μέτρα από την πλευρά των εθνικών κυβερνήσεων, επί ζητημάτων διαχείρισης των τεχνολογικών πόρων, να είναι προσανατολισμένα στη χρηστή χρήση της τεχνολογίας προς όφελος των πολιτών και της εν γένει κοινωνίας. Όμως, αυτό εξαρτάται από την ερευνητική προοπτική του τρόπου με τον οποίο μπορεί κάποιος να προβεί στη μελέτη των διαφορετικών πολιτικών προοπτικών για τον κλάδο της τεχνολογίας και την εξέλιξη που αυτός γνωρίζει³⁷⁷.

Για λόγους πληρέστερης κατανόησης της ανωτέρω διατύπωσης, επισημαίνεται ότι η οργανωσιακή πολιτική κουλτούρα προς την υιοθέτηση συγκεκριμένων πτυχών της τεχνολογίας, όπως είναι επί παραδείγματι η τεχνητή νοημοσύνη, δύναται να ασκήσει είτε θετική είτε αρνητική επιρροή στις πολιτικές διαδικασίες, αλλά και σε ζητήματα αντιπροσώπευσης των πολιτικών παραγόντων, ενώ παράλληλα, πρέπει να αντιμετωπιστούν ζητήματα πολιτικής που αφορούν την ενσωμάτωση και αξιοποίηση της τεχνολογίας. Από κοινωνικής και κονστρουκτιβιστικής πτυχής, οι επικρατούσες αντιλήψεις των φορέων, επί του ζητήματος της διαχείρισης της τεχνολογίας, διαδραματίζουν κομβικό ρόλο στο στάδιο της ανάλυσης και του σχεδιασμού των

³⁷⁷ Βλ. P. Dawson, C. Clausen & K. Nielsen, Political Processes In Management, Organization And The Social Shaping Of Technology, Technology Analysis & Strategic Management, Τεύχος 12, Αρ.1, 2000, σελ. 5-15.

διαδικασιών και των πολιτικών για την υιοθέτηση της τεχνολογίας και δη της τεχνητής νοημοσύνης³⁷⁸.

Εξ αυτής της οπτικής, η μελέτη που διεξήχθη από τους Dawson, Clausen και Nielsen³⁷⁹ επισημαίνει ότι είναι εφικτή η διερεύνηση της κατανόησης της τεχνολογίας και του πώς αυτή λειτουργεί σε περιβάλλοντα της κοινωνίας των πολιτών, ειδικότερα όταν ακολουθεί ανάλυση αντιλήψεων, κοινωνικής διαμόρφωσης της τεχνολογίας και λήψης αποφάσεων επί πολιτικών διαδικασιών που θα εφαρμοστούν από τις εκάστοτε αρμόδιες κυβερνήσεις. Η μελέτη των Duettmann, Afanasjeva, Armstrong, Braley, Cussins, Ding, Eckersley, Guan, Vance και Yampolskiy³⁸⁰ είναι αντιπροσωπευτικό παράδειγμα της ανωτέρω αναφερόμενης θεωρίας. Πιο αναλυτικά, η μελέτη αυτή υπογραμμίζει τη μείωση του πλήθους των πολιτικών της αμερικανικής κυβέρνησης, οι οποίες αφορούν τη διαχείριση της προηγμένης τεχνολογίας. Αυτό έχει ως αποτέλεσμα τον περιορισμό της ικανότητας της αμερικανικής κυβέρνησης να προβεί στον σχεδιασμό και την αποτελεσματική εφαρμογή πολιτικών επί του ζητήματος της ενσωμάτωσης της τεχνολογίας και της τεχνητής νοημοσύνης στην κοινωνία των πολιτών. Η εν λόγω επιβράδυνση δεν οφείλεται στην υφιστάμενη διχόνοια και διχογνωμία των αντίπαλων κομμάτων της ομοσπονδιακής κυβέρνησης της Αμερικής, αλλά στις προσωπικές απόψεις από τις οποίες διακατέχονται κορυφαίοι πολιτικοί παράγοντες³⁸¹.

Τονίζεται ότι σε περίπτωση που ένα κράτος επιθυμεί να συμμετέχει στον διεθνή ανταγωνισμό και να μην χάσει το ανταγωνιστικό του πλεονέκτημα από άλλα κράτη, δεν νοείται να μην διαθέτει στρατηγικά καθοδηγούμενη έρευνα επί του ζητήματος της κατανόησης της τεχνολογίας³⁸². Οι Kuhlmann και Edler³⁸³, οι οποίοι έδωσαν ιδιαίτερη έμφαση στην προαγωγή της καινοτομίας, εξήγησαν τη σημασία της αναγνώρισης της

³⁷⁸ Βλ. P. Dawson, C. Clausen & K. Nielsen, *όπ.παρ.*, σελ. 8.

³⁷⁹ Βλ. P. Dawson, C. Clausen & K. Nielsen, *όπ.παρ.*, σελ. 8.

³⁸⁰ Βλ. A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, *όπ.παρ.*, σελ. 16

³⁸¹ Βλ. A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, *όπ.παρ.*, σελ. 16.

³⁸² Βλ. U. Hilpert, *όπ.παρ.*, σελ. 26.

³⁸³ Βλ. S. Kuhlmann & J. Edler, *όπ.παρ.*, σελ. 619.

υπόστασης και θέσπισης περιφερειακών και διακρατικών στρατηγικών, πολιτικών και προγραμμάτων για λόγους ενίσχυσης των εθνικών καταβαλλόμενων προσπαθειών στο θέμα της υιοθέτησης προηγμένων τεχνολογιών και δη της τεχνητής νοημοσύνης στην κοινωνία των πολιτών.

Παρατηρείται ότι κάποια κράτη διαθέτουν ειδικές γνώσεις επί τεχνολογικών θεμάτων, οι οποίες εκτείνονται ακόμα και σε πολιτικό και στρατηγικό επίπεδο. Μέσω των γνώσεων αυτών θα ήταν εφικτή η μεταξύ τους αλληλοσυμπλήρωση. Επί παραδείγματι, σχετικά με τη Γαλλία, η ένταξη στην Ευρωπαϊκή Ένωση προσφέρει ένα ευρύ φάσμα κεφαλαίων υπέρ της τεχνολογικής καινοτομίας, της ανάπτυξης προγραμμάτων και δικτύων συνεργασίας και της παροχής επιπλέον ευκαιριών για την ενίσχυση της ανταγωνιστικότητας των κρατών-μελών³⁸⁴.

Στο σημείο αυτό, διευκρινίζεται ότι ο ιδιωτικός τομέας **συγκλίνει** σε μεγαλύτερο βαθμό προς τον στρατιωτικό και τον πολιτικό τομέα, κάτι που δεν ευθυγραμμίζεται με τον αυστηρότερο διαχωρισμό μεταξύ ιδιωτικού και του δημόσιου τομέα, όπου η απόκλιση είναι μεγαλύτερη, ειδικότερα στο παράδειγμα της Αμερικής³⁸⁵. Η διαφορά στην κατανόηση της τεχνολογίας στην Κίνα και στην Αμερική πηγάζει, συν τοις άλλοις, από τη σχέση που έχει εκάστη εθνική κυβέρνηση με τον ιδιωτικό τομέα που σχεδιάζει και κατασκευάζει μηχανές τεχνητής νοημοσύνης³⁸⁶. Κρίνοντας από την αποτελεσματικότητα της κινεζικής κυβέρνησης στο ζήτημα της κατανόησης της τεχνολογίας, οι συμπράξεις μεταξύ του ιδιωτικού και του δημόσιου τομέα ή η συμμετοχή ιδιωτών στις εθνικές προσπάθειες, με σκοπό την ανάπτυξη πολιτικών επί του θέματος της αξιοποίησης της τεχνολογίας στην κοινωνία των πολιτών, θα μπορούσαν να ενισχύσουν περαιτέρω την κατανόηση της τεχνολογίας που επικρατεί σε ένα κράτος.

³⁸⁴ Βλ. S. Kuhlmann & J. Edler, *όπ.παρ.*, σελ. 620, Βλ. επίσης A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, *όπ.παρ.*, σελ. 17.

³⁸⁵ Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo & L. Floridi, *όπ.παρ.*, σελ. 508, Βλ.επίσης A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, *όπ.παρ.*, σελ. 17, Βλ. επίσης B. Hall, *The Assessment: Technology Policy*, Oxford Review Of Economic Policy, Τεύχος 18, Αρ. 1, 2002, σελ. 1-9.

³⁸⁶ Βλ. A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, *όπ.παρ.*, σελ. 18.

Ορισμένες μελέτες που έχουν επικεντρωθεί στο θέμα της τεχνητής νοημοσύνης υποστηρίζουν ότι οι κυβερνήσεις που δεν διακατέχονται από αυξημένα επίπεδα κατανόησης της τεχνολογίας θα οδηγηθούν στην αναπόφευκτη δημιουργία επικοινωνιακού χάσματος³⁸⁷. Σύμφωνα με εκθέσεις που έχουν καταρτιστεί από επιστήμονες του κλάδου της τεχνητής νοημοσύνης, τονίζεται ότι είναι αναγκαία και άκρως απαραίτητη η συμμετοχή του ιδιωτικού στον δημόσιο τομέα, διαμέσου εφαρμογής προγραμμάτων ανταλλαγών των πόρων των δύο τομέων και μέσω έργων διατομεακής συνεργασίας, με σκοπό τη γεφύρωση του ανωτέρω αναφερόμενου χάσματος επικοινωνίας³⁸⁸. Σε περίπτωση κατά την οποία οι εθνικές κυβερνήσεις προβούν σε αποκλεισμό των ιδιωτικών τεχνολογικών επιχειρήσεων από τις διαδικασίες χάραξης πολιτικής, τότε το επικοινωνιακό» χάσμα θα είναι βέβαιο³⁸⁹.

Επισημαίνεται ότι η Αμερική και η Κίνα διαθέτουν ισχυρούς εμπορικούς τομείς τεχνητής νοημοσύνης, που αποτελούν τους πρωτοστάτες στο μεγαλύτερο μέρος της τεχνολογικής ανάπτυξης του πεδίου της τεχνητής νοημοσύνης³⁹⁰. Στη μελέτη που πραγματοποιήθηκε από τον Hilpert³⁹¹, δόθηκε έμφαση στη σπουδαιότητα της παρουσίας χειριστών δικτύου τεχνολογίας και επικοινωνίας, εντός μίας εθνικής τεχνολογικής δομής³⁹². Η παρουσία του εν λόγω δικτύου θα ήταν δυνατό να οδηγήσει σε αύξηση του επιπέδου τεχνολογικής ανταγωνιστικότητας, καθώς επίσης και σε αύξηση των επιπέδων διαθεσιμότητας των εμπειρογνομόνων επί ζητημάτων τεχνολογίας³⁹³. Μάλιστα, οι εν λόγω εμπειρογνώμονες θα μπορούσαν ενδεχομένως να συνεργαστούν με τις εθνικές κυβερνήσεις. Επιστρέφοντας ξανά στο παράδειγμα της Αμερικής, αναφέρεται ότι η αμερικανική κυβέρνηση ξεκίνησε συνεργασίες

³⁸⁷ Βλ. FLI. όπ.παρ., στο: <https://futureoflife.org/ai-policy-challenges-and-recommendations/>

³⁸⁸ Βλ. FLI. όπ.παρ., στο: <https://futureoflife.org/ai-policy-challenges-and-recommendations/>

³⁸⁹ Βλ. A. Duettmann, O. Afanasjeva, S. Armstrong, R. Braley, J. Cussins, J. Ding, P. Eckersley, M. Guan, A. Vance & R. Yampolskiy, όπ.παρ., σελ. 18.

³⁹⁰ Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo & L. Floridi, όπ.παρ., σελ. 513, Βλ. επίσης J Ding, όπ.παρ., σελ. 27, Βλ. επίσης FLI, όπ.παρ., στο: <https://futureoflife.org/ai-policy/>

³⁹¹ Βλ. U. Hilpert, όπ.παρ., σελ. 16.

³⁹² Βλ. U. Hilpert, όπ.παρ., σελ. 16.

³⁹³ Βλ. U. Hilpert, όπ.παρ., σελ. 16.

αυτορρύθμισης, οι οποίες δυνητικά θα μπορούσαν να γεφυρώσουν ένα υφιστάμενο επικοινωνιακό χάσμα³⁹⁴.

³⁹⁴ Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo & L. Floridi, *όπ.παρ.*, σελ. 513.

3.11 Σύνοψη – Συμπεράσματα

Είναι προφανές ότι, λαμβάνοντας υπόψη την έκταση της μελέτης και διερεύνησής τους που διαφάνηκε στο παρόν κεφάλαιο, τα έξυπνα αντικείμενα, συμπεριλαμβανομένων των μηχανών τεχνητής νοημοσύνης, έχουν τη δυνατότητα να ασκήσουν σημαντική επιρροή στον τρόπο που σκέφτεται και οργανώνεται ο άνθρωπος στην κοινωνία. Συνακόλουθα, η επιστημονική και νομική πρόοδος δεν δύναται να αποστασιοποιηθεί από τα ζητήματα νομικής και ηθικής φύσεως που εμπλέκονται σε αυτήν την πραγματικότητα.

Υπό αυτό το πρίσμα, η ανάδυση νέων επιστημονικών και οντολογικών πτυχών προκύπτει ως επιτακτική ανάγκη. Είναι απαραίτητο, σε κάθε περίπτωση, τα τεχνικά αντικείμενα τεχνητής νοημοσύνης να ληφθούν υπόψη, όχι ως απλοϊκά εργαλεία, αλλά ως μηχανές που έχουν άμεση εμπλοκή με ηθικά ζητήματα, τη στιγμή που καταλήγουν σε έντονες αλληλεπιδράσεις με τα μέλη μίας κοινωνίας.

Για να σχεδιαστούν επαρκή όρια, εν όψει της σύγχρονης εποχής ανάδυσης της τεχνητής νοημοσύνης, πρέπει να αναγνωριστούν τα αντικείμενα και οι μηχανές ως αυτοτελή εμπλεκόμενα μέρη, τα οποία διαθέτουν μετα-ανθρωπιστική προοπτική. Η προσέγγιση αυτή όμως δεν πρέπει να γίνει με τρόπο που παραγκωνίζει και δεν σέβεται τα ανθρώπινα δικαιώματα.

Η συζήτηση γύρω από το θέμα της ηθικής της τεχνητής νοημοσύνης εξακολουθεί να κυριαρχεί στο μελετητικό και ερευνητικό γίγνεσθαι. Οι όποιες δυσκολίες που προκύπτουν από τεχνολογικούς μετασχηματισμούς υψηλής πολυπλοκότητας δεν είναι δυνατόν να οδηγήσουν σε αποτροπή της θέσπισης νέου κανονιστικού πλαισίου, το οποίο θα δύναται να οδηγήσει σε μείωση των κινδύνων που συνοδεύουν τις νέες δραστηριότητες και, κατά συνέπεια, την αποκατάσταση ζημιών.

Η ακριβής κατεύθυνση που πρέπει να ακολουθηθεί εξακολουθεί να παραμένει αβέβαιη. Παρ' όλα αυτά, είναι ήδη δυνατό να φανταστούμε τις επιμέρους δυνατότητες που υφίστανται και που μπορούν να χρησιμεύσουν ως σημαντικές παράμετροι. Η νέα πρόκληση, άλλωστε, δεν είναι η τεχνολογική καινοτομία και η εξέλιξη και ανάπτυξή της, αλλά η διακυβέρνηση του ψηφιακού κόσμου και ο τρόπος με τον οποίο η τεχνολογική καινοτομία θα καταστεί αξιοποιήσιμη και διαχειρίσιμη από τον

ανθρώπινο παράγοντα ή αν τελικά θα καταλήξει να τον επισκιάσει, με ό,τι συνεπάγεται αυτό για την ανθρώπινη υπόσταση.

Επίσης, επισημαίνεται ότι η ενσωμάτωση της τεχνητής νοημοσύνης στους δημόσιους και κυβερνητικούς οργανισμούς έχει καταστεί εναργής στην κάλυψη διαδικαστικών ζητημάτων εξυπηρέτησης των πολιτών στις καθημερινές τους απαιτήσεις. Ο δημόσιος τομέας, προκειμένου να ενσωματώσει σε ακόμα μεγαλύτερο βαθμό την τεχνητή νοημοσύνη στις διάφορες πτυχές του, προϋποθέτει στοχευμένη οργάνωση και προγραμματισμό, ώστε να διασφαλίζονται αφενός ζητήματα προστασίας και προάσπισης των προσωπικών και ευαίσθητων δεδομένων των πολιτών και αφετέρου ζητήματα κάλυψης του ηθικού κινδύνου που προκύπτει από την αλληλεπίδραση της τεχνητής νοημοσύνης με τον ανθρώπινο παράγοντα.

Κεφάλαιο 4 Δημόσια Διοίκηση

4.1 Τα Πεδία Εφαρμογής της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα

Στη συνέχεια παρουσιάζονται ορισμένες δυναμικές εφαρμογές της τεχνητής νοημοσύνης, με αντίστοιχες λεπτομερείς περιγραφές και παραδείγματα, οι οποίες χαρακτηρίζονται από αυξημένη πιθανότητα αξιοποίησης στο πλαίσιο του δημοσίου τομέα. Για λόγους παρουσίασης των σύγχρονων χρησιμοποιούμενων εφαρμογών επί του τομέα της τεχνητής νοημοσύνης, επιλέχθηκαν συγκεκριμένα παραδείγματα που αναφέρονται στην τελευταία οκταετία. Στην προκειμένη περίπτωση, εντοπίστηκαν συνολικά δέκα διαφορετικοί τομείς του δημόσιου τομέα, στους οποίους εφαρμόζεται η τεχνητή νοημοσύνη:

(1) Λογισμικό διαχείρισης γνώσης βασισμένο στην τεχνητή νοημοσύνη (Knowledge Management - KM): Η δημιουργία αξίας της τεχνητής νοημοσύνης έγκειται στην παραγωγή και συστηματοποίηση γνώσεων, ήτοι τη συλλογή, την καταγραφή, την ταξινόμηση, τη μεταμόρφωση και την κοινή χρήση γνώσεων³⁹⁵. Επίσης, ειδικά συστήματα δύνανται να υποστηρίζουν την κωδικοποίηση της διαχείρισης γνώσης, ενώ η χρήση νευρωνικών δικτύων επιτρέπει την ανάλυση, διανομή και κοινή χρήση γνώσεων με άλλους³⁹⁶. Η αξιοποίηση της τεχνολογίας αυτής στον δημόσιο τομέα θα μπορούσε να πραγματοποιηθεί σε περιπτώσεις κλινικής τεκμηρίωσης που βασίζονται στην τεχνητή νοημοσύνη³⁹⁷.

(2) Συστήματα αυτοματοποίησης διεργασιών τεχνητής νοημοσύνης (Process Automation Systems): Το ουσιώδες στοιχείο της τεχνητής νοημοσύνης εντοπίζεται στην αυτοματοποίηση των τυπικών εργασιών και στην εκτέλεση επίσημων λογικών διεργασιών με απρόβλεπτες συνθήκες, σε σταθερή ποιότητα³⁹⁸. Με αυτό

³⁹⁵ Βλ. H. Lu, Y. Li, M. Chen, H. Kim, & S. Serikawa, *όπ.παρ.*, σελ. 368–375.

³⁹⁶ Βλ. L. Sanzogni, G. Guzman, & P. Busch, *όπ.παρ.*, σελ. 37–56.

³⁹⁷ Βλ. S.Y. Lin, T.D. Shanafelt, & S.M. Asch, Reimagining clinical documentation with artificial intelligence, *Mayo Clinic Proceedings*, Ap. 93, Τεύχος 5, 2018, σελ. 563–565.

³⁹⁸ Βλ. H. Lu, Y. Li, M. Chen, H. Kim, & S. Serikawa, *όπ.παρ.*, σελ. 368–375.

εξασφαλίζεται η διατήρηση της ήδη υφιστάμενης ποιότητας και η διασφάλιση των σταθερών επιπέδων ποιότητας της προσφερόμενης επιπρόσθετης αξίας από τη χρήση της τεχνητής νοημοσύνης. Επίσης, οι σύνθετες διαδικασίες ανθρώπινης δράσης δύνανται να ανατεθούν σε συστήματα αυτοματισμού, τα οποία μπορούν να υποστηρίξουν τους ανθρώπους στην εκτέλεση των εργασιών τους³⁹⁹. Μπορεί, συν τοις άλλοις, να περιλαμβάνεται η αξιολόγηση βάσει κανόνων, η επεξεργασία ροής εργασίας, διάφορες προτάσεις, η εξόρυξη δεδομένων, η συλλογιστική βάσει περιπτώσεων, καθώς και η έξυπνη τεχνολογία αισθητήρων⁴⁰⁰. Ο αυτοματισμός της ρομποτικής διαδικασίας έχει αναδειχθεί ως μία υπο-περιοχή των περαιτέρω αναδυομένων τεχνολογικών καινοτομιών⁴⁰¹. Αυτό, αξιοποιεί την ικανότητα των ρομπότ ή των μηχανών που βασίζονται στην τεχνητή νοημοσύνη να μιμούνται την ανθρώπινη αλληλεπίδραση με τις διεπαφές χρήστη στα λογισμικά συστήματα⁴⁰². Στη συγκεκριμένη περίπτωση, η τεχνολογία αυτή θα μπορούσε να χρησιμοποιηθεί για αυτοματοποιημένες διαγνώσεις εικόνας⁴⁰³, αλλά και για την αλληλεπίδραση ανθρώπου-υπολογιστή σε επαναλαμβανόμενες εργασίες, όπως είναι επί παραδείγματι η εισαγωγή δεδομένων και λοιπών χρήσιμων πληροφοριών⁴⁰⁴.

(3) Εικονικοί πράκτορες (Virtual Agents): Το σημείο αναφοράς εδράζεται στην υπόσταση ενός συστήματος που είναι βασισμένο σε υπολογιστές και το οποίο αλληλοεπιδρά με τον χρήστη μέσω ανάλυσης ομιλίας, όρασης υπολογιστή και γραπτής

³⁹⁹ Βλ. D. Ali, & S. Frimpong, Artificial intelligence, machine learning and process automation: existing knowledge frontier and way forward for mining sector, *Artificial Intelligence Review*, Ap. 53, 2020, σελ. 6025-6042.

⁴⁰⁰ Βλ. B. Qela, & H.T. Mouftah, Observe, Learn, and Adapt (OLA)—An Algorithm for Energy Management in Smart Homes Using Wireless Sensors and Artificial Intelligence, *IEEE Transactions on Smart Grid*, Ap. 3, Τεύχος 4, 2012, σελ. 2262–2272.

⁴⁰¹ Βλ. W. Wang, & K. Siau, Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity, *Journal of Database Management*, Ap. 30, Τεύχος 1, 2019, σελ. 61–79.

⁴⁰² Βλ. A. Chun, & H. Wai, Using AI for E-government automatic assessment of immigration application forms, *Proceedings of the National Conference on Artificial Intelligence, AAAI 2007, Vancouver, 2007*, σελ. 1684–1691.

⁴⁰³ Βλ. M. Collier, R. Fu, & L. Yin, Artificial intelligence: Healthcare's new nervous system, 2017, στο https://www.accenture.com/t20170418T023052Z__w__/_au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

⁴⁰⁴ Βλ. D. Jefferies, The automated city: Do we still need humans to run public services, 2016, στο <https://www.theguardian.com/cities/2016/sep/20/automated-city-robots-run-public-services-councils>

εισαγωγής δεδομένων. Επίσης, μπορεί να περιλαμβάνει καθολική μετάφραση σε πραγματικό χρόνο και συστήματα επεξεργασίας της φυσικής γλώσσας, καθώς και συναισθηματικό υπολογισμό⁴⁰⁵. Μία επιπρόσθετη αξία αναφέρεται στην υπόσταση λογισμικού που είναι σε θέση να εκτελεί εργασίες για λογαριασμό των ανθρώπων. Οι υπο-περιοχές είναι chatbots και είδωλα (avatars)⁴⁰⁶. Η αξιοποίηση της εν λόγω τεχνολογίας στην περίπτωση του δημοσίου τομέα θα ήταν δυνατό να επιτευχθεί για την κατανομή εργασιών, με βάση τον αντίστοιχο τομέα ευθύνης μίας συγκεκριμένης υπηρεσίας, όπως είναι για παράδειγμα οι έξυπνες υπηρεσίες ανθρωπίνων πόρων⁴⁰⁷. Επιπροσθέτως, οι εικονικοί πράκτορες θα μπορούσαν να χρησιμεύσουν στον δημόσιο τομέα ως εικονικοί βοηθοί στον κλάδο της νοσηλευτικής⁴⁰⁸ ή θα μπορούσαν να χρησιμοποιηθούν υπό την μορφή ενός chatbot⁴⁰⁹ για να βοηθούν τους πρόσφυγες που ζητούν άσυλο, στη συμπλήρωση και αναζήτηση εγγράφων⁴¹⁰.

(4) Προγνωστική ανάλυση (Predictive Analytics) και οπτικοποίηση δεδομένων (Data Visualization): Στην προκειμένη περίπτωση, η δημιουργία αξίας της τεχνητής νοημοσύνης αφορά το ότι αυτές οι αναλύσεις βασίζονται σε ποσοτική και στατιστική ανάλυση των δεδομένων⁴¹¹. Πιο συγκεκριμένα, επιδίδονται στην επεξεργασία μεγάλων

⁴⁰⁵ Βλ. K. Weitz, D. Schiller, R. Schlagowski, T. Huber, & E. André, Do you trust me, Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents - IVA '19, 2019, σελ. 7-9.

⁴⁰⁶ Βλ. L. Ciechanowski, A. Przegalinska, M. Magnuski, & P. Gloor, In the shades of the uncanny valley: An experimental study of human–chatbot interaction, Future Generation Computer Systems, Ap. 92, 2018, σελ. 539-548.

⁴⁰⁷ Βλ. Y. Zheng, Y. Han, L. Cui, C. Miao, C. Leung, & Q. Yang, SmartHS: An AI platform for improving government service provision, The Thirtieth AAAI Conference on Innovative Applications of Artificial Intelligence (IAAI-18), 2018, σελ. 7704–7711, στο <https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16041/16369>

⁴⁰⁸ Βλ. M. Collier, R. Fu, & L. Yin, όπ.παρ., στο https://www.accenture.com/t20170418T023052Z__w_/au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

⁴⁰⁹ Βλ. L. Ciechanowski, A. Przegalinska, M. Magnuski, & P. Gloor, όπ.παρ., σελ. 539-548.

⁴¹⁰ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁴¹¹ Βλ. P. Balthazar, P. Harri, A. Prater, & N.M. Safdar, Protecting Your Patients' Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics, Journal of the American College of Radiology, Ap. 15, Τεύχος 3, 2018, σελ. 580–586, Βλ. επίσης X. Qin, Y. Luo, N. Tang, & G. Li, DeepEye: An automatic big data visualization framework, Big Data Mining and Analytics, Ap. 1, Τεύχος 1, 2018, σελ. 75–82.

δεδομένων για αναφορά, προδιαγραφική και προγνωστική ανάλυση. Η μηχανική μάθηση εμφανίζεται ως μία τεχνική υπο-περιοχή που είναι βασισμένη σε αλγόριθμους, οι οποίοι έχουν τη δυνατότητα να μάθουν από τα υφιστάμενα και συλλεγόμενα δεδομένα⁴¹². Η αξιοποίηση της εν λόγω τεχνολογίας στην περίπτωση του δημοσίου τομέα θα ήταν δυνατό να επιτευχθεί για τον έλεγχο και την παρακολούθηση των επιδόσεων σε δημόσιους χώρους, από τις αστυνομικές υπηρεσίες, για τον προσδιορισμό τρομοκρατικών απειλών και σημείων εγκληματικότητας για προληπτική δράση⁴¹³. Η προγνωστική ανάλυση θα μπορούσε να αξιοποιηθεί και στον προσδιορισμό καταστάσεων υψηλού κινδύνου εγκληματικότητας για την ασφάλεια των δημόσιων μεταφορών⁴¹⁴, αλλά και ως ένα πρότυπο ελέγχου προκειμένου να προβλέπονται τα επίπεδα ύδατος⁴¹⁵.

(5) Αναλύσεις ταυτότητας (Identity Analytics): Η δημιουργία αξίας, στη συγκεκριμένη περίπτωση είναι συνδεδεμένη με την υπόσταση ενός λογισμικού που συνδυάζει μεγάλα δεδομένα, προηγμένα αναλυτικά στοιχεία, τη δυνατότητα διαχείρισης πρόσβασης ταυτότητας⁴¹⁶ για τον έλεγχο της εισόδου σε συστήματα πληροφορικής και την αυτοματοποίηση ελέγχων ταυτότητας βάσει κινδύνου⁴¹⁷. Μπορεί να περιλαμβάνει βαθιά και μηχανική μάθηση, συναισθηματικό υπολογισμό και το λεγόμενο τεχνητό ανοσοποιητικό σύστημα (artificial immune system)⁴¹⁸. Η

⁴¹² Βλ. P. Balthazar, P. Harri, A. Prater, & N.M. Safdar, *όπ.παρ.*, σελ. 580–586.

⁴¹³ Βλ. D.J. Power, Big Brother can watch us, *Journal of Decision Systems*, Αρ. 25, Τεύχος 1, 2016, σελ. 578–588.

⁴¹⁴ Βλ. G.N. Kouziokas, The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment, *Transportation Research Procedia*, Αρ. 24, 2017, σελ. 467–473.

⁴¹⁵ Βλ. G.N. Kouziokas, *όπ.παρ.*, σελ. 467–473.

⁴¹⁶ Βλ. M. Swan, Blockchain for Business: Next-Generation Enterprise Artificial Intelligence Systems, *Blockchain Technology: Platforms, Tools and Use Cases*, Αρ. 111, 2018, σελ. 121–162.

⁴¹⁷ Βλ. M.C. Michel, & M.C. King, Categorization of Discoverable Cyber Attributes for Identity Protection, Privacy, and Analytics, *SoutheastCon*, 2018, σελ. 1-5.

⁴¹⁸ Βλ. O. Igbe, I. Darwish, & T. Saadawi, Distributed Network Intrusion Detection Systems: An Artificial Immune System Approach, 2016 IEEE First International Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 2016, σελ. 101-106, Βλ. επίσης SP. aurabh, & B. Verma, An efficient proactive artificial immune system based anomaly detection and prevention system, *Expert Systems with Applications*, Αρ. 60, 2016, σελ. 311–320.

συγκεκριμένη τεχνολογία θα μπορούσε να αξιοποιηθεί στον δημόσιο τομέα ως ένα λογισμικό αναγνώρισης προσώπου, για την επαλήθευση ή τον εντοπισμό εγκληματιών σε δημόσιους χώρους⁴¹⁹, ή ακόμα και για την ανίχνευση απάτης τεχνητής νοημοσύνης για τη διασφάλιση κυβερνητικών δεδομένων⁴²⁰.

(6) Γνωστική Ρομποτική (Cognitive Robotics) & Αυτόνομα Συστήματα (Autonomous Systems): Η παρουσία συστημάτων που διαθέτουν γνωστικές λειτουργίες υψηλότερου επιπέδου είναι το σημείο αναφοράς⁴²¹. Οι λειτουργίες αυτές περιλαμβάνουν την αναπαράσταση γνώσης και είναι σε θέση να μαθαίνουν και να ανταποκρίνονται⁴²². Μερικές φορές η εν λόγω εκμάθηση και ανταπόκριση σχετίζεται με τον συναισθηματικό υπολογισμό για τον προσδιορισμό και την προσαρμογή της ανθρώπινης συμπεριφοράς, καθώς και με την απόκριση στα αντίστοιχα συναισθήματα⁴²³. Η αξιοποίηση αυτής της τεχνολογίας στην περίπτωση του δημοσίου τομέα θα ήταν δυνατό να επιτευχθεί μέσω της δημιουργίας ηλεκτρικών μηχανοκίνητα οχημάτων για τις δημόσιες συγκοινωνίες⁴²⁴ ή ακόμη και στο πεδίο της χειρουργικής, η οποία θα είναι υποβοηθούμενη από ρομπότ⁴²⁵.

⁴¹⁹ Βλ. D.J. Power, *όπ.παρ.*, σελ. 578–588.

⁴²⁰ Βλ. T. Hemken, & C. GraySmart move: Technologies make their mark on public service, 2016, στο https://www.accenture.com/t20170224T040950Z__w__/us-en/_acnmedia/PDF-36/Accenture-Public-Service-Emerging-Technologies-Slideshare.pdf

⁴²¹ Βλ. A. Lieto, M. Bhatt, A. Oltramari, & D. Vernon, The role of cognitive architectures in general artificial intelligence, *Cognitive Systems Research*, Αρ. 48, 2018, σελ. 1–3, Βλ. επίσης A. Aly, S. Griffiths, & F. Stramandinoli, Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics, *Cognitive Systems Research*, Αρ. 43, 2017, σελ. 313–323.

⁴²² Βλ. A. Lieto, M. Bhatt, A. Oltramari, & D. Vernon, *όπ.παρ.*, σελ. 1–3.

⁴²³ Βλ. A. Aly, S. Griffiths, & F. Stramandinoli, *όπ.παρ.*, σελ. 313–323.

⁴²⁴ Βλ. Christchurch International Airport, New Zealand's first trial of autonomous vehicle announced. 2016, στο <https://www.scoop.co.nz/stories/SC1610/S00008/new-zealands-first-trial-of-autonomous-vehicle-announced.htm>, Βλ. επίσης D. Jefferies, *όπ.παρ.*, στο <https://www.theguardian.com/cities/2016/sep/20/automated-city-robots-run-public-services-councils>

⁴²⁵ Βλ. M. Collier, R. Fu, & L. Yin, *όπ.παρ.*, στο https://www.accenture.com/t20170418T023052Z__w__/au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

(7) Συστήματα σύστασης (Recommendation Systems): Η αξία εντοπίζεται στην ύπαρξη ενός συστήματος φιλτραρίσματος πληροφοριών⁴²⁶. Πρόκειται για συστήματα που βασίζονται σε λογισμικά και τα οποία προβάλλουν εξατομικευμένες πληροφορίες για να προβλέψουν τις προτιμήσεις των ανθρώπων⁴²⁷. Η αξιοποίηση της εν λόγω τεχνολογίας στην περίπτωση του δημοσίου τομέα θα ήταν δυνατό να επιτευχθεί μέσω της υπόστασης μίας ηλεκτρονικής υπηρεσίας για κυβερνητικά γραφεία, με την οποία θα παρέχονται εξατομικευμένες πληροφορίες προς τους υπαλλήλους⁴²⁸.

(8) Ευφυείς ψηφιακοί βοηθοί (Intelligent Digital Assistants - IDA): Η αξία που δημιουργείται από την τεχνητή νοημοσύνη αναφέρεται στην υπόσταση λογισμικών βασισμένων σε ανάλυση ομιλίας (speech analytics)⁴²⁹. Τα λογισμικά αυτά βασίζονται στην παροχή μίας διαισθητικής διεπαφής μεταξύ ενός χρήστη και ενός συστήματος ή μίας συσκευής, για λόγους αναζήτησης πληροφοριών ή για την ολοκλήρωση απλών εργασιών⁴³⁰. Η αξιοποίηση της εν λόγω τεχνολογίας στην περίπτωση του δημοσίου τομέα θα ήταν δυνατό να επιτευχθεί για τη διασύνδεση ομοσπονδιακών προγραμμάτων προς δημοσίευση διαθέσιμων πληροφοριών για τους πολίτες⁴³¹. Επίσης, οι ευφυείς ψηφιακοί βοηθοί θα ήταν χρήσιμοι στην παροχή βοήθειας προς τους κατοίκους ώστε

⁴²⁶ Βλ. J. Castro, R. Yera, & L. Martínez, An empirical study of natural noise management in group recommendation systems, *Decision Support Systems*, Ap. 94, 2017, σελ. 1–11.

⁴²⁷ Βλ. N. Nikzad–Khasmakhi, M.A. Balafar, & M. Reza Feizi–Derakhshi, The state-of-the-art in expert recommendation systems, *Engineering Applications of Artificial Intelligence*, Ap. 82, 2019, σελ. 126–147.

⁴²⁸ Βλ. E. Cortés-Cediel, I. Cantador, & O. Gil, Recommender systems for e-governance in smart cities. In Y. Jie, S. Zhu, B. Alessandro; Z. Jie & L. Martha (Eds.) *Proceedings of the international workshop on citizens for recommender systems - CitRec '17*, Como, Italy, August 27, 2017, New York, ACM press, 2017, στο https://www.researchgate.net/publication/319043437_Recommender_systems_for_e-governance_in_smart_cities_State_of_the_art_and_research_opportunities

⁴²⁹ Βλ. G. Gauglitz, Artificial vs. human intelligence in analytics, *Analytical and Bioanalytical Chemistry*, 2019, σελ. 1-2.

⁴³⁰ Βλ. N.J. Fast, & J. Schroeder, Power and Decision Making: New Directions for Research in the Age of Artificial Intelligence, *Current Opinion in Psychology*, Ap. 33, 2019, σελ. 172-176.

⁴³¹ Βλ. J. Herman, Federal pilot to integrate public services into intelligent personal assistants. 2017, στο <https://digital.gov/2017/04/24/federal-pilot-to-integrate-public-services-into-intelligent-personal-assistants/>, Βλ. επίσης J. Herman, Opening public services to artificial intelligence assistants, 2017 στο <https://www.gsa.gov/blog/2017/06/06/Opening-Public-Services-to-Artificial-Intelligence-Assistants>

να εντοπίσουν πληροφορίες και να συμπληρώσουν φόρμες εφαρμογών χρησιμοποιώντας αναλυτικά στοιχεία ομιλίας και συναισθηματικό υπολογισμό⁴³².

(9) Ανάλυση ομιλίας (Speech Analytics): Η παρουσία λογισμικού για έξυπνη αναγνώριση και επεξεργασία γλώσσας είναι το σημείο δημιουργίας αξίας της εν λόγω περίπτωσης⁴³³. Η λειτουργία του αναφέρεται στην κατανόηση και απάντηση στη φυσική γλώσσα⁴³⁴. Μία ακόμα προστιθέμενη αξία αφορά τη δυνατότητα μετάφρασης από προφορική σε γραπτή γλώσσα ή από μία φυσική γλώσσα σε μία άλλη. Μπορεί, μάλιστα, να περιλαμβάνει συστήματα καθολικής μετάφρασης και επεξεργασίας της φυσικής γλώσσας σε πραγματικό χρόνο. Στον δημόσιο τομέα, η συγκεκριμένη τεχνολογία θα μπορούσε να αξιοποιηθεί για την παροχή υπηρεσιών, μέσω της πραγματοποίησης μεταφράσεων σε πραγματικό χρόνο⁴³⁵ για λόγους διευκόλυνσης των δια ζώσης επικοινωνιών. Επίσης, η ανάλυση ομιλίας, στην περίπτωση του δημόσιου τομέα, είναι χρήσιμη και για την παροχή βοήθειας διοικητικής ροής εργασιών με μετατροπή της φωνής σε κείμενο⁴³⁶.

(10) Γνωστική ανάλυση ασφαλείας (Cognitive Security Analytics) & Ευφυΐα κινδύνου (Threat Intelligence): Το σημείο αναφοράς εντοπίζεται στην πρόσθετη εφαρμογή γνωστικών τεχνολογιών για την ανάλυση πληροφοριών ασφαλείας, μέσω της φυσικής επεξεργασίας γλωσσών και της μηχανικής μάθησης⁴³⁷. Η ερμηνεία και οργάνωση πληροφοριών και η παροχή συλλογιστικής είναι ορισμένες επιπρόσθετες

⁴³² Βλ. D. Jefferies, *όπ.παρ.*, στο <https://www.theguardian.com/cities/2016/sep/20/automated-city-robots-run-public-services-councils>

⁴³³ Βλ. T.H. Davenport, From analytics to artificial intelligence, *Journal of Business Analytics*, 2018, σελ. 1-8.

⁴³⁴ Βλ. G. Gauglitz, *όπ.παρ.*, σελ. 1-2.

⁴³⁵ Βλ. Microsoft, Machine translation, 2018, στο <https://www.microsoft.com/en-us/translator/mt.aspx>

⁴³⁶ Βλ. M. Collier, R. Fu, & L. Yin, *όπ.παρ.*, στο https://www.accenture.com/t20170418T023052Z__w__/au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf

⁴³⁷ Βλ. R.O. Andrade, & S.G. Yoo, Cognitive security: A comprehensive study of cognitive science in cybersecurity, *Journal of Information Security and Applications*, Apr. 48, 2019, σελ. 1-13.

αξίες που προσφέρονται από τα εν λόγω συστήματα⁴³⁸. Η αξιοποίηση της εν λόγω τεχνολογίας στην περίπτωση του δημοσίου τομέα θα ήταν δυνατό να επιτευχθεί για εφαρμογές, όπως είναι παραδείγματος χάριν το Watson, για την ασφάλεια στον κυβερνοχώρο και την υποστήριξη της ανάλυσης της ανθρώπινης ασφάλειας στο δημόσιο τομέα⁴³⁹.

Μία σημαντική διαπίστωση που προκύπτει από την παραπάνω ανάλυση, αφορά την προσπάθεια πραγματοποίησης δοκιμών αξιοποίησης της τεχνητής νοημοσύνης από διαφορετικούς τύπους δημοσίων οργανισμών σε παγκόσμιο βεληνεκές. Ακόμη και σε αυτήν την περίπτωση, όμως, η ταχύτητα αξιοποίησης της τεχνητής νοημοσύνης από την πλευρά των ιδιωτικών επιχειρήσεων είναι μεγαλύτερη συγκριτικά με την αντίστοιχη των δημόσιων οργανισμών⁴⁴⁰. Παρά ταύτα, δεν μπορεί να παραβλεφθεί η πιλοτική αξιοποίησή της από την πλευρά των δημόσιων οργανισμών, με σκοπό τη βελτίωση της ποιότητας εξυπηρέτησης των πολιτών⁴⁴¹.

Σε αυτό το υπόβαθρο, οι αναφερόμενες περιοχές εφαρμογών τεχνητής νοημοσύνης «Αυτοματοποίηση διεργασίας τεχνητής νοημοσύνης», «Αναγνωριστικό πρόβλεψης», «Ανάλυση ταυτότητας», «Εικονικοί πράκτορες» και «Γνωστική ρομποτική» έχουν ιδιαίτερη σημασία και ουσιαστικά οφέλη για τον δημόσιο τομέα. Οι εν λόγω εφαρμογές είναι δυνατόν να αυξήσουν την αποδοτικότητα και να οδηγήσουν σε εξοικονόμηση κόστους μέσω της αυτοματοποίησης των ακολουθούμενων διαδικασιών, βοηθώντας, με τον τρόπο αυτόν, την κατανομή πόρων και μειώνοντας, παράλληλα, τον χρόνο αναμονής και τις διοικητικές επιβαρύνσεις⁴⁴².

⁴³⁸ Βλ. V. Mavroeidis, & S. Bromander, *Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence*, European Intelligence and Security Informatics Conference (EISIC), 2017, σελ. 91-98.

⁴³⁹ Βλ. R.O. Andrade, & S.G. Yoo, *όπ.παρ.*, σελ. 1-13, Βλ. επίσης V. Dheap, *IBM QRadar advisor with Watson: Revolutionizing the way security analysts work*, 2017, στο <https://securityintelligence.com/ibm-qradar-advisor-with-watson-revolutionizing-the-way-security-analysts-work/>

⁴⁴⁰ Βλ. H. Mehr, *όπ.παρ.*, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁴⁴¹ Βλ. J. Herman, *όπ.παρ.*, στο <https://digital.gov/2017/04/24/federal-pilot-to-integrate-public-services-into-intelligent-personal-assistants/>, Βλ. επίσης J. Herman, *όπ.παρ.*, στο <https://www.gsa.gov/blog/2017/06/06/Opening-Public-Services-to-Artificial-Intelligence-Assistants>

⁴⁴² Βλ. H. Mehr, *όπ.παρ.*, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf, Βλ. επίσης W.D.

Κατά συνέπεια, δημιουργούνται ευκαιρίες για τους δημόσιους υπαλλήλους, καθώς υπό τις συνθήκες αυτές δύνανται να επικεντρωθούν σε συγκεκριμένες εργασίες και δραστηριότητες, ενώ οι μηχανές θα είναι σε θέση να φέρνουν εις πέρας διάφορες διαδικασίες ρουτίνας. Ταυτόχρονα, ωστόσο, παραμένει η απειλή της ανθρώπινης αντικατάστασης, δημιουργώντας, επομένως, ηθικούς προβληματισμούς σχετικά με τη χρήση της τεχνολογίας της τεχνητής νοημοσύνης σε δημόσιο πλαίσιο. Για παράδειγμα, το ζήτημα της ευθύνης σε περίπτωση ελαττωματικών ή λανθασμένων αποφάσεων, οι οποίες λαμβάνονται από τις μηχανές και την τεχνολογία τεχνητής νοημοσύνης δεν έχει ακόμη διευκρινιστεί⁴⁴³.

Συνεπώς, η τεχνητή νοημοσύνη και η υπόσταση και αξιοποίησή της οδηγεί σε ένα εξαιρετικά φλέγον δίλημμα: Από τη μία πλευρά, οι κυβερνήσεις μπορούν να χρησιμοποιήσουν τη δυναμική των εφαρμογών της τεχνητής νοημοσύνης προσβλέποντας στη βελτίωση διάφορων δημοσίων υποθέσεων και στην αύξηση της αποτελεσματικότητας των εσωτερικών τους διαδικασιών, ενώ από την άλλη, οι υφιστάμενες και επερχόμενες απειλές καταδεικνύουν την απαίτηση η τεχνητή νοημοσύνη να συνοδεύεται από πολιτικές και ρυθμίσεις που βασίζονται σε αρχές και βασικές κοινωνικές αξίες προκειμένου να αποφέρει συλλογικά οφέλη⁴⁴⁴.

Eggers, T. Fishman, & P. Kishnani, όπ.παρ., στο https://www2.deloitte.com/content/dam/insights/us/articles/4152_AI-human-services/4152_AI-human-services.pdf

⁴⁴³ Βλ. M. Boyd, & N. Wilson, όπ.παρ., σελ. 36–44.

⁴⁴⁴ Βλ. M. Boyd, & N. Wilson, όπ.παρ., σελ. 36–44.

4.2 Οι Προκλήσεις του Δημοσίου Τομέα για μία Πολυδιάστατη Ενσωμάτωση της Τεχνητής Νοημοσύνης

Όπως φάνηκε εκ της ανωτέρω ανάλυσης, υπάρχουν ποικίλες ευκαιρίες για την εφαρμογή και αξιοποίηση της τεχνολογίας τεχνητής νοημοσύνης στο δημόσιο τομέα, με εναργείς δυνατότητες αύξησης της αποτελεσματικότητάς του⁴⁴⁵. Ωστόσο, οι σχετικές προκλήσεις δεν πρέπει να παραμερίζονται, καθώς ενδέχεται να εμποδίσουν την εφαρμογή και τη χρήση καινοτόμων εφαρμογών⁴⁴⁶. Στο πλαίσιο της υφιστάμενης και διαθέσιμης βιβλιογραφίας είναι δυνατός ο εντοπισμός αρκετών μελετών, οι οποίες επικεντρώνονται στις προκλήσεις από τις οποίες συνοδεύεται η αξιοποίηση και ενσωμάτωση της τεχνητής νοημοσύνης στον δημόσιο τομέα ⁴⁴⁷.

Λόγω των πιθανών κοινωνικών και ηθικών επιπτώσεων της εφαρμογής της τεχνητής νοημοσύνης, έχει πυροδοτηθεί μία έντονη συζήτηση στην επιστήμη και την κοινωνία. Κάποιες έρευνες που επικεντρώνονται στα ζητήματα αυτά, μεταξύ άλλων, είναι των Bartlett⁴⁴⁸, των Johnson & Verdicchio⁴⁴⁹, του Mehr⁴⁵⁰, του Müller⁴⁵¹ και των Quraishi et al.⁴⁵². Κάποιες άλλες έρευνες έχουν επικεντρωθεί στην εφαρμογή της τεχνολογίας της τεχνητής νοημοσύνης στην επιχειρηματική συμπεριφορά, στην περίπτωση των δημοσίων υπηρεσιών. Συγκεκριμένα, πρόκειται για τις έρευνες που

⁴⁴⁵ Βλ. W.G. Sousa, E.R.P. de, Melo, P.H.D.S. Bermejo, R.A.S. Farias, & A.O. Gomes, How and where is artificial intelligence in the public sector going, A literature review and research agenda, Government Information Quarterly, 2019, σελ. 1-14.

⁴⁴⁶ Βλ. B.W. Wirtz, J.C. Weyerer & C. Geyer, Artificial Intelligence and the Public Sector—Applications and Challenges, International Journal of Public Administration, 2018, σελ. 1–20.

⁴⁴⁷ Βλ. W.G. Sousa, E.R.P. de, Melo, P.H.D.S. Bermejo, R.A.S. Farias, & A.O. όπ.παρ., σελ. 1-14, Βλ. επίσης B.W. Wirtz, J.C. Weyerer & C. Geyer, όπ.παρ., σελ. 1–20.

⁴⁴⁸ Βλ. S.J. Bartlett, The case for government by artificial intelligence, SSRN Journal, 2017, σελ. 1–15.

⁴⁴⁹ Βλ. D.G. Johnson, & M. Verdicchio, όπ.παρ., σελ. 2267–2270.

⁴⁵⁰ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁴⁵¹ Βλ. V. Müller, Risks of general artificial intelligence, Journal of Experimental & Theoretical Artificial Intelligence, Αρ. 26, Τεύχος 3, 2014, σελ. 297– 301.

⁴⁵² Βλ. F.F. Quraishi, S.A. Wajid, & P. Dhiman, Social and ethical impact of artificial intelligence on public: A case study of university students, International Journal of Scientific Research in Science, Engineering and Technology, Αρ. 3, Τεύχος 8, 2017, σελ. 463– 467.

πραγματοποιήθηκαν από τους Amodei et al.⁴⁵³, του Scherer⁴⁵⁴ και τους Thierer et al.⁴⁵⁵. Από την πλευρά τους, οι Boyd & Wilson⁴⁵⁶, Power⁴⁵⁷ και Scherer⁴⁵⁸ επικεντρώθηκαν στις νομικής φύσεως προϋποθέσεις για την αξιοποίηση της τεχνητής νοημοσύνης στο δημόσιο τομέα.

Με βάση τις ανωτέρω αναφερόμενες και πολυδιάστατες τάσεις συζήτησης, προτείνουμε τέσσερις μεγάλες διαστάσεις προκλήσεων, από τις οποίες συνοδεύεται το φαινόμενο της τεχνητής νοημοσύνης⁴⁵⁹. Αυτές είναι το υφιστάμενο νομοθετικό και κανονιστικό πλαίσιο που είναι συνυφασμένο με την τεχνητή νοημοσύνη, η κοινωνία της τεχνητής νοημοσύνης, η ηθική της τεχνητής νοημοσύνης και η εφαρμογή της τεχνολογίας της τεχνητής νοημοσύνης⁴⁶⁰. Αυτές οι τέσσερις διαστάσεις χρησιμεύουν ως βάση για τη μοντελοποίηση των προκλήσεων που σχετίζονται με την εισαγωγή των εφαρμογών της τεχνητής νοημοσύνης στον δημόσιο τομέα⁴⁶¹.

Αν και οι προαναφερθείσες μελέτες εξετάζουν τις προκλήσεις της τεχνητής νοημοσύνης ξεχωριστά και λαμβάνοντας υπόψη ότι τα θέματα αυτά δεν είναι νεοεισαχθέντα στο επιστημονικό γίγνεσθαι, θεωρούμε ότι δεν παρέχεται από την

⁴⁵³ Βλ. D. Amodei, C., Olah, J. Steinhardt, P. Christiano, J. Schulman, & D. Mané, Concrete problems in AI safety, In CoRR abs/1606.06565, 2016, στο <https://arxiv.org/pdf/1606.06565.pdf>

⁴⁵⁴ Βλ. M.U. Scherer, Regulating artificial intelligence systems: Risk, challenges, competencies, and strategies, Harvard Journal of Law & Technology, Ap. 29, Τεύχος 2, 2016, σελ. 353–400.

⁴⁵⁵ Βλ. A. Thierer, A. O’Sullivan Castillo, & R. Russell, όπ.παρ., στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁴⁵⁶ Βλ. M. Boyd, & N. Wilson, όπ.παρ., σελ. 36–44.

⁴⁵⁷ Βλ. D.J. Power, όπ.παρ., σελ. 578–588.

⁴⁵⁸ Βλ. M.U. Scherer, όπ.παρ., σελ. 353–400.

⁴⁵⁹ Βλ. F.F. Quraishi, S.A. Wajid, & P. Dhiman, όπ.παρ., σελ. 463–467, Βλ. επίσης D. Amodei, C., Olah, J. Steinhardt, P. Christiano, J. Schulman, & D. Mané, όπ.παρ., στο <https://arxiv.org/pdf/1606.06565.pdf>, Βλ. επίσης S.J. Bartlett, όπ.παρ., σελ. 1–15.

⁴⁶⁰ Βλ. A. Thierer, A. O’Sullivan Castillo, & R. Russell, όπ.παρ., στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>, Βλ. επίσης B.W. Wirtz, J.C. Weyerer & C. Geyer, όπ.παρ., σελ. 1–20.

⁴⁶¹ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf, Βλ. επίσης M.U. Scherer, όπ.παρ., σελ. 353–400, Βλ. επίσης W.G. Sousa, E.R.P. de Melo, P.H.D.S. Bermejo, R.A.S. Farias, & A.O. όπ.παρ., σελ. 1–14.

υφιστάμενη βιβλιογραφία μία ολοκληρωμένη επισκόπηση των προκλήσεων της τεχνητής νοημοσύνης. Παρατηρώντας και μελετώντας το εν λόγω ζήτημα πιο προσεκτικά, εντοπίσαμε, συν των ανωτέρω, την υπόσταση υπο-προκλήσεων που αποδίδονται στις τέσσερις κύριες διαστάσεις που περιγράφονται σε καθένα από τα ακόλουθα υποκεφάλαια.

4.2.1 Πρώτη Διάσταση: Το Κανονιστικό Πλαίσιο στην Ενσωμάτωση της Τεχνητής Νοημοσύνης στον Έλεγχο και τη Διακυβέρνηση

Ο νόμος και το ευρύτερο κανονιστικό πλαίσιο για την τεχνητή νοημοσύνη αναφέρονται στη γενική διακυβέρνησή της και επομένως αφορούν στη συνολική ικανότητα διαχείρισης και ελέγχου του εξελισσόμενου αυτού φαινομένου, καθώς και τον κοινωνικό και οικονομικό του αντίκτυπο. Λόγω του ευρέος πεδίου εφαρμογής, η διακυβέρνηση της τεχνητής νοημοσύνης σχετίζεται με μία ποικιλία και πολυδιάσταση νομικών και κανονιστικών ζητημάτων που αναφέρονται σε δεδομένα, αλγόριθμους, υποδομές, καθώς επίσης και στον ανθρώπινο παράγοντα⁴⁶².

Η διαχείριση της τεχνητής νοημοσύνης με υπεύθυνο και ευεργετικό τρόπο αντιπροσωπεύει, ως εκ τούτου, ένα πολύπλοκο έργο και μία μεγάλη πρόκληση, η οποία μάλλον έχει παραμεληθεί, ιδιαίτερα όταν πρόκειται για μακροπρόθεσμα ζητήματα⁴⁶³. Οι γνώσεις από κορυφαία ερευνητικά ιδρύματα, όπως η Future Society στο Harvard Kennedy School ή το Future of Humanity Institute στο Πανεπιστήμιο της Οξφόρδης υποδεικνύουν τον κρίσιμο ρόλο και την αναγκαιότητα της χάραξης πολιτικής που σχετίζεται με την τεχνητή νοημοσύνη⁴⁶⁴, τονίζοντας την ανάγκη για κάποιας μορφής παγκόσμιου διοικητικού συμβουλίου⁴⁶⁵. Η αποσαφήνιση του νομικού καθεστώτος της τεχνητής νοημοσύνης είναι απαραίτητη για την εξάλειψη της αβεβαιότητας και τον προσδιορισμό της νομικής ευθύνης και υπευθυνότητας, όταν οι εφαρμογές τεχνητής

⁴⁶² Βλ. U. Gasser, AI and the law: Setting the stage, Berkman Klein Center for Internet & Society at Harvard University, 2017, στο <https://medium.com/berkman-klein-center/ai-and-the-law-setting-the-stage-48516fda1b11>

⁴⁶³ Βλ. N. Bostrom, A. Dafoe, & C. Flynn, Policy desiderata in the development of machine superintelligence, Future of Humanity Institute, University of Oxford, 2016, στο <https://www.fhi.ox.ac.uk/wp-content/uploads/Policy-Desiderata-in-the-Development-of-Machine-Superintelligence.pdf>

⁴⁶⁴ Βλ. N. Bostrom, A. Dafoe, & C. Flynn, *όπ.παρ.*, στο <https://www.fhi.ox.ac.uk/wp-content/uploads/Policy-Desiderata-in-the-Development-of-Machine-Superintelligence.pdf>, Βλ. επίσης N. Mialhe, The policy challenges of automation, Field Actions Science Reports, The Journal of Field Actions, Ap. 17, 2017, σελ. 66–71.

⁴⁶⁵ Βλ. M. Boyd, & N. Wilson, *όπ.παρ.*, σελ. 40.

νοημοσύνης προκαλούν βλάβες και ηθικές ζημιές⁴⁶⁶. Συνακόλουθα, αποκαλύπτονται τέσσερις σημαντικές πτυχές σε σχέση με τον νόμο και το κανονιστικό πλαίσιο της τεχνητής νοημοσύνης, συμπεριλαμβανομένης της διακυβέρνησης των αυτόνομων πληροφοριακών συστημάτων, της ευθύνης, της λογοδοσίας και της ιδιωτικής ζωής και ασφάλειας⁴⁶⁷.

Η διακυβέρνηση των αυτόνομων συστημάτων νοημοσύνης αναφέρεται στην πρόκληση της κατανόησης και του ελέγχου των αποφάσεων και των ενεργειών των συστημάτων τεχνητής νοημοσύνης και των αλγορίθμων που συχνά αναφέρονται ως «μαύρα κουτιά»⁴⁶⁸. Σε σχέση με αυτό, ο Castelvechi⁴⁶⁹ υπογραμμίζει την αδιαφάνεια των συστημάτων τεχνητής νοημοσύνης και σημειώνει ότι αντί να αποθηκεύουν ό,τι έχουν μάθει σε ένα τακτοποιημένο μπλοκ ψηφιακής μνήμης, διαδίδουν τις πληροφορίες με τρόπο που είναι εξαιρετικά δύσκολο να αποκρυπτογραφηθούν, και αυτό με τη σειρά του καθιστά επίσης δύσκολη τη διεξαγωγή επαρκούς ελέγχου στην εν λόγω ακολουθούμενη διαδικασία.

Η διακυβέρνηση έχει διακλαδικό χαρακτήρα, αναφερόμενη τόσο σε μία κυβερνητική δράση όσο και στην υπόσταση κανόνων ή μέτρων, συμπεριλαμβανομένων των λοιπών σχετικών ενδιαφερομένων μερών, όπως είναι οι εταιρείες τεχνολογίας τεχνητής νοημοσύνης ή οι μη κερδοσκοπικές οργανώσεις⁴⁷⁰. Κατά συνέπεια, οι κυβερνήσεις και όλοι οι άλλοι σχετικοί παράγοντες πρέπει να διασφαλίσουν την υπόσταση ορισμένων βασικών απαιτήσεων, όσον αφορά την επεξήγηση, τη διαφάνεια, τη δικαιοσύνη και την υπευθυνότητα επί των συστημάτων τεχνητής νοημοσύνης και τους αλγόριθμους που διέπουν τη ζωή των ανθρώπων, αναλαμβάνοντας το δύσκολο καθήκον της

⁴⁶⁶ Βλ. IEEE, Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems, Edited by The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, 2017, στο http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf

⁴⁶⁷ Βλ. M. Boyd, & N. Wilson, όπ.παρ., σελ. 40, Βλ. επίσης IEEE, όπ.παρ., στο http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf

⁴⁶⁸ Βλ. A. Bleicher, όπ.παρ., στο <https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/>, Βλ. επίσης W. Knight, όπ.πρ., σελ. 54–61

⁴⁶⁹ Βλ. D. Castelvechi, Can we open the black box of AI, Nature, Apr. 538, Τεύχος 7623, 2016, σελ. 20–23.

⁴⁷⁰ Βλ. N. Bostrom, A. Dafoe, & C. Flynn, όπ.παρ., στο <https://www.fhi.ox.ac.uk/wp-content/uploads/Policy-Desiderata-in-the-Development-of-Machine-Superintelligence.pdf>

εισαγωγής μηχανισμών διακυβέρνησης που ελαχιστοποιούν τους κινδύνους και τις πιθανές παγίδες⁴⁷¹.

Επομένως, οι κυβερνήσεις, σε παγκόσμιο βεληνεκές, κρίνεται απαραίτητο να συμφωνήσουν σε ενιαίες αποδεκτές αρχές και κανονισμούς για τα συστήματα τεχνητής νοημοσύνης που ενσωματώνουν, αξιοποιήσιμα, πάντα, υπό των επιταγών της δημοκρατίας και του σεβασμού των ανθρωπίνων δικαιωμάτων⁴⁷². Ο σχεδιασμός και η καθιέρωση ενός παγκόσμιου και ευέλικτου συστήματος διακυβέρνησης τεχνητής νοημοσύνης, το οποίο όχι μόνο θα ανταποκρίνεται στις διάφορες πτυχές της τεχνητής νοημοσύνης, αλλά θα ενσωματώνει πολιτιστικές ποικιλίες και διαφορετικά εθνικά νομικά συστήματα είναι περίπλοκο και, συνεπώς, αντιπροσωπεύει μία σημαντική πρόκληση στο πλαίσιο της ενσωμάτωσης και αξιοποίησής της στον δημόσιο τομέα⁴⁷³.

Η ευθύνη και η υπευθυνότητα συνδέονται αρρήκτως με την πτυχή της διακυβέρνησης και αναφέρονται στον καθορισμό του νομικού πλαισίου σχετικά με το ποιος είναι υπεύθυνος και αρμόδιος για αποφάσεις που λαμβάνονται από την τεχνητή νοημοσύνη. Παραδείγματος χάριν, το ερώτημα του ποιος είναι υπεύθυνος προκύπτει όταν ένα αυτόνομο όχημα για τις δημόσιες συγκοινωνίες βλάπτει έναν πεζό. Είναι ο σχεδιαστής του υλικού ή του λογισμικού, ο προμηθευτής ή ο χειριστής, οι αρχές ή ακόμη και η ίδια η ενσωμάτωση της τεχνητής νοημοσύνης υπεύθυνη για τις συνέπειες τυχουσών αποφάσεων που λαμβάνονται κατά τη λειτουργία της;⁴⁷⁴

Δεδομένου ότι τα συστήματα τεχνητής νοημοσύνης μαθαίνουν ενώ λειτουργούν και ενεργούν αυτόνομα, οι προγραμματιστές ή οι χειριστές τους ενδέχεται να μην είναι σε θέση να ελέγχουν ή να προβλέπουν την επακόλουθη συμπεριφορά τους⁴⁷⁵. Επομένως, τα συστήματα τεχνητής νοημοσύνης είναι πιθανόν να αψηφούν τον άμεσο ανθρώπινο έλεγχο, γεγονός που οδηγεί στο λεγόμενο κενό της ευθύνης, σύμφωνα με το οποίο οι

⁴⁷¹ Βλ. U. Gasser, & V.A.F. Almeida, A layered model for AI governance, IEEE Internet Computing, Αρ. 21, Τεύχος 6, 2017, σελ. 58–62, Βλ. επίσης I. Rahwan, Society-in-the-loop: Programming the algorithmic social contract, Ethics and Information Technology, Αρ. 20, Τεύχος 1, 2018, σελ. 5–14.

⁴⁷² Βλ. M. Boyd, & N. Wilson, όπ.παρ., σελ. 36–44.

⁴⁷³ Βλ. U. Gasser, & V.A.F. Almeida, όπ.παρ., σελ. 58–62.

⁴⁷⁴ Βλ. IEEE, όπ.παρ., στο http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf

⁴⁷⁵ Βλ. D.G. Johnson, Technology with no human responsibility, Journal of Business Ethics, Αρ. 127, Τεύχος 4, 2015, σελ. 707–715.

άνθρωποι δεν μπορούν να θεωρηθούν υπεύθυνοι για τη συμπεριφορά των συστημάτων αυτών, λόγω της έλλειψης ελέγχου και της εν τέλει μηδαμινής επιρροής τους⁴⁷⁶.

Στην αντίθετη πλευρά, ο De George⁴⁷⁷ υποστηρίζει ότι οι άνθρωποι είναι πάντα υπεύθυνοι για τις συνέπειες που σχετίζονται με την τεχνολογία. Παρ' όλο που υπήρξαν διάφορες προσεγγίσεις που αντιμετωπίζουν την πρόκληση της ευθύνης και της λογοδοσίας, στο πλαίσιο της τεχνητής νοημοσύνης, και επιδιώκουν να ξεπεράσουν το κενό ευθύνης, μέχρι στιγμής δεν υπάρχει συναίνεση για τον τρόπο αντιμετώπισης της εν λόγω μείζονος υφιστάμενης πρόκλησης⁴⁷⁸. Ο Johnson⁴⁷⁹ υποδεικνύει ότι η υπέρβαση αυτής της πρόκλησης είναι ζήτημα ανθρώπινης απόφασης, καθώς και οικοδόμησης πολιτικής και κοινωνικής συναίνεσης, επισημαίνοντας ότι σε περίπτωση κατά την οποία θα υπάρξει κενό ευθύνης, αυτό εξαρτάται απόλυτα από τις ανθρώπινες επιλογές και όχι από την τεχνολογική πολυπλοκότητα.

Το απόρρητο και η ασφάλεια αναφέρονται στην πρόκληση της διατήρησης της προστασίας της ιδιωτικής σφαίρας των ατόμων και των δεδομένων τους και των πόρων του δικτύου, επί των υφιστάμενων ή επερχόμενων απειλών ασφάλειας που πηγάζουν από την αξιοποίηση της τεχνητής νοημοσύνης. Καταδεικνύεται ότι τα δεδομένα που συλλέγονται από τα άτομα και τα οποία υποβάλλονται σε επεξεργασία, θα πρέπει να συνοδεύονται από τη συγκατάθεσή τους και δε θα πρέπει η εν λόγω διαδικασία να παρεκκλίνει από τους αντίστοιχους ισχύοντες νόμους και κανονισμούς⁴⁸⁰.

⁴⁷⁶ Βλ. A. Matthias, The responsibility gap: Ascribing responsibility for the actions of learning automata, *Ethics and Information Technology*, Αρ.6, Τεύχος 3, 2004, σελ. 175–183.

⁴⁷⁷ Βλ. R.T. De George, *The ethics of information technology and business*, Malden, Blackwell Publishing, 2003, σελ. 46-59.

⁴⁷⁸ Βλ. S.L. Anderson, Machine metaethics, In M. Anderson & S. L. Anderson (Eds.), *Machine ethics* New York, Cambridge University Press, 2011, σελ. 21-27, Βλ. επίσης P.M. Asaro, A body to kick, but still no soul to damn: Legal perspectives on robotics, In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics*, Cambridge, MIT Press, 2012, σελ. 169-186, Βλ. επίσης M. Nagenborg, R. Capurro, J. Weber, & C. Pingel, Ethical regulations on robotics in Europe, *AI & Society*, Αρ. 22, Τεύχος 3, 2008, σελ. 349–366, Βλ. επίσης M. Santoro, D. Marino, & G. Tamburrini, Learning robots interacting with humans: From epistemic risk to responsibility, *AI & Society*, Αρ. 22, Τεύχος 3, 2008, σελ. 301–314

⁴⁷⁹ Βλ. D.G. Johnson, *όπ.παρ.*, σελ. 714.

⁴⁸⁰ Βλ. D.G. Johnson, *όπ.παρ.*, σελ. 707-715.

Σύμφωνα με τον Rössler⁴⁸¹, υπάρχουν τρεις τρόποι με τους οποίους είναι δυνατόν με τα συστήματα τεχνητής νοημοσύνης να προκύψουν παραβιάσεις της ιδιωτικής σφαίρας, ήτοι ως παράνομη παρέμβαση στις ενέργειες κάποιου, ως παράνομη επιτήρηση και ως παράνομες εισβολές σε δωμάτια ή κατοικίες⁴⁸². Για παράδειγμα, τα συστήματα τεχνητής νοημοσύνης, όπως είναι οι εφαρμογές ρομπότ, είναι ευάλωτα σε επιθέσεις στον κυβερνοχώρο, οι οποίες είναι ιδιαίτερα επικίνδυνες λόγω του ότι οι επιτιθέμενοι αποκτούν πρόσβαση στο φυσικό περιβάλλον διαβίωσης των ανθρώπων και, ιδίως, στον πιο ευαίσθητο τομέα της ζωής τους εντός του πλαισίου της ιδιωτικής σφαίρας⁴⁸³. Επιπλέον, ενδέχεται να προκύψουν ανησυχίες για το απόρρητο, επί παραδείγματι, σε σχέση με την κυβέρνηση που βασίζεται στην αξιοποίηση συστημάτων τεχνητής νοημοσύνης⁴⁸⁴.

Επί του συγκεκριμένου ζητήματος, μία πρόσφατη έρευνα αποτυπώνει ότι η συντριπτική πλειοψηφία των πολιτών ανησυχεί για την απειλή που δημιουργεί η τεχνητή νοημοσύνη στην ιδιωτική τους ζωή⁴⁸⁵. Καθίσταται εναργές ότι η τεχνητή νοημοσύνη δύναται να έχει σημαντικό αντίκτυπο στην ιδιωτική ζωή των ανθρώπων, καθιστώντας την προστασία της σημαντική πρόκληση στο πλαίσιο της επιτυχούς της ενσωμάτωσης στον δημόσιο τομέα⁴⁸⁶.

Από τη μία πλευρά, τα συστήματα τεχνητής νοημοσύνης πρέπει να περιέχουν προηγμένα λογισμικά ασφαλείας στον κυβερνοχώρο, προκειμένου να εξασφαλίσουν την ασφάλεια των δεδομένων και την προστασία της ιδιωτικής σφαίρας των

⁴⁸¹ Βλ. B. Rössler, *The value of privacy*, Cambridge, Polity Press, 2005, σελ. 9.

⁴⁸² Βλ. R. Calo, *Peeping HALs: Making sense of artificial intelligence and privacy*, *European Journal of Legal Studies*, Αρ. 2, Τεύχος 3, 2010, σελ. 168–192.

⁴⁸³ Βλ. R.M. Calo, *Robots and privacy*, In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics*, Cambridge, MIT Press, 2012, σελ. 187-201.

⁴⁸⁴ Βλ. U. Gasser, & V.A.F. Almeida, *όπ.παρ.*, σελ. 58–62.

⁴⁸⁵ Βλ. Z. Soo, *The increasing use of artificial intelligence is stoking privacy concerns in China*, 2018, στο <http://www.scmp.com/business/companies/article/2135713/increasing-use-artificial-intelligence-stoking-privacy-concerns>

⁴⁸⁶ Βλ. Y.A. De Montjoye, A. Farzanehfar, J. Hendrickx, & L. Rocher, *Solving artificial intelligence's privacy problem*, *Field Actions Science Reports, The Journal of Field Actions Special Issue*, Αρ. 17, 2017, σελ. 80–83.

πολιτών⁴⁸⁷. Από την άλλη πλευρά, όμως, η νομοθεσία και η χάραξη πολιτικής πρέπει να προσαρμοστούν στις νέες εξελίξεις και τις μεταβαλλόμενες συνθήκες που προκύπτουν από την ενσωμάτωση της τεχνητής νοημοσύνης⁴⁸⁸. Αυτό περιλαμβάνει, μεταξύ άλλων, τον χειρισμό ασυμφωνιών που υφίστανται μεταξύ δικαιοδοσιών όσον αφορά τα πρότυπα απορρήτου, καθώς επίσης και λοιπά παρεμφερή ζητήματα, όπως επί παραδείγματι ζητήματα ηθικής⁴⁸⁹.

⁴⁸⁷ Βλ. J. Holdren, & M. Smith, Preparing for the future of artificial intelligence, Edited by Executive Office of the President National Science and Technology Council Committee on Technology, 2016, στο https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

⁴⁸⁸ Βλ. A. Krausová, Intersections between law and artificial intelligence, International Journal of Computer, Apr. 27, Τεύχος 1, 2017, σελ. 55–68.

⁴⁸⁹ Βλ. U. Gasser, *όπ.παρ.*, στο <https://medium.com/berkman-klein-center/ai-and-the-law-setting-the-stage-48516fda1b11>

4.2.2 Δεύτερη Διάσταση: Κοινωνικά Ζητήματα της Πολύπλευρης Ενσωμάτωσης της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα

Η κοινωνία της τεχνητής νοημοσύνης αναφέρεται στην εξέλιξη της κοινωνικής ζωής και της ανθρώπινης αλληλεπίδρασης που οδηγείται από την τεχνητή νοημοσύνη και τις κοινωνικές προκλήσεις οι οποίες σχετίζονται με αυτές τις αλλαγές. Η τεχνητή νοημοσύνη έχει καταστεί ένα τμήμα του καθημερινού κόσμου, ασκώντας επιρροή στις καθημερινές συνήθειες των πολιτών, καθώς επίσης και σε πολλούς βασικούς τομείς της κοινωνίας, όπως είναι η υγειονομική περίθαλψη, οι μεταφορές και η οικονομία⁴⁹⁰. Ταυτόχρονα, υφίστανται αυξανόμενες ανησυχίες σχετικά με τη μελλοντική ανάπτυξη της τεχνητής νοημοσύνης και τις διάχυτες αμφιβολίες που εμφανίζονται στην κοινωνία, λόγω των δυνητικά δυσμενών επιπτώσεων της στην ανθρωπότητα και το κοινωνικό σύνολο⁴⁹¹.

Επί παραδείγματι, ερευνητές της τεχνητής νοημοσύνης έλαβαν μία πρωτοβουλία, με τη μορφή ανοιχτών επιστολών που δημοσιεύθηκαν από το Future of Life Institute, η οποία έχει υποστηριχθεί ευρέως από διάσημους ηγέτες και εμπειρογνώμονες του κλάδου, όπως τον Stephen Hawking ή τον Elon Musk. Στις εν λόγω επιστολές, δίνονται απαντήσεις για την απειλή που μπορεί να δημιουργήσει η τεχνητή νοημοσύνη, εκφράζοντας τις ανησυχίες τους σχετικά με την αξιοποίησή της και τις αρνητικές επιπτώσεις που μπορεί να επιφέρει στην ανθρωπότητα και το κοινωνικό σύνολο, καθώς και υποστηρίζοντας μία ανάπτυξη τεχνητής νοημοσύνης η οποία είναι επωφελής για την κοινωνία⁴⁹².

Οι φόβοι και οι αντιρρήσεις για την ενσωμάτωση της τεχνητής νοημοσύνης ανιχνεύονται, επίσης, στους ίδιους τους πολίτες. Η έρευνα της Ευρωπαϊκής Επιτροπής

⁴⁹⁰ Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, & L. Floridi, Artificial intelligence and the 'good society': The US, EU, and UK approach, Science and Engineering Ethics, Ap. 24, Τεύχος 2, 2018, σελ. 505–528.

⁴⁹¹ Βλ. D.G. Johnson, & M. Verdicchio, όπ.παρ., σελ. 2267–2270.

⁴⁹² Βλ. Future of Life Institute, Autonomous weapons: An open letter from AI & robotics researchers, 2015, στο <https://futureoflife.org/open-letter-autonomous-weapons/>, Βλ. επίσης Future of Life Institute, Research priorities for robust and beneficial artificial intelligence, 2015, στο <https://futureoflife.org/ai-open-letter>

για τα αυτόνομα συστήματα το 2015, καταδεικνύει ότι μεγάλο ποσοστό πολιτών εκφράζει τις επιφυλάξεις του σχετικά με την τεχνητή νοημοσύνη σε διάφορους τομείς της εφαρμογής της. Πιο συγκεκριμένα, η πλειονότητα των ερωτηθέντων αισθάνεται αμήχανα με συστήματα τεχνητής νοημοσύνης που παρέχουν υπηρεσίες σε ηλικιωμένους ή που χρησιμοποιούνται σε ιατρικές χειρουργικές επεμβάσεις⁴⁹³.

Αξίζει να αναφερθεί ότι στη βιβλιογραφία εντοπίζονται τρεις πτυχές αναφορικά με την πρόκληση της τεχνητής νοημοσύνης για την κοινωνία. Αυτές περιλαμβάνουν την αντικατάσταση και τον μετασχηματισμό του εργατικού δυναμικού, την κοινωνική αποδοχή και εμπιστοσύνη στην τεχνητή νοημοσύνη και τον μετασχηματισμό της αλληλεπίδρασης μεταξύ ανθρώπου-μηχανής (H2M) και μηχανής σε μηχανή (M2M)⁴⁹⁴.

Η αντικατάσταση και ο μετασχηματισμός του εργατικού δυναμικού αναφέρεται στον αντίκτυπο της τεχνητής νοημοσύνης στην αγορά εργασίας και αντιπροσωπεύει μία από τις πιο διαδεδομένες προκλήσεις για την κοινωνία⁴⁹⁵. Υπάρχει αυξανόμενη ανησυχία ότι η περαιτέρω πρόοδος και η ενσωμάτωση της τεχνητής νοημοσύνης θα οδηγήσει σε ανεργία, καθώς οι εργασιακές δραστηριότητες και οι θέσεις εργασίας υπόκεινται ολοένα και περισσότερο σε ένα καθεστώς αυτοματοποίησης⁴⁹⁶.

Διάφορα ευρήματα μελετών υποστηρίζουν τις ανωτέρω αναφερόμενες ανησυχίες, αν και οι εκτιμήσεις των εμπειρογνομόνων στο εν προκειμένω ζήτημα ποικίλλουν έντονα. Μία μελέτη της PwC αποτυπώνει ότι περίπου το ένα τρίτο των θέσεων εργασίας απειλείται σε μεγάλες βιομηχανικές χώρες⁴⁹⁷, τη στιγμή που και οι μελέτες του

⁴⁹³ Βλ. European Commission, Special Eurobarometer 427: Autonomous systems. Report, 2015, στο http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_427_en.pdf

⁴⁹⁴ Βλ. M. Boyd, & N. Wilson, *όπ.παρ.*, σελ. 36-44.

⁴⁹⁵ Βλ. C. Bataller, & J. Harris, Turning artificial intelligence into business value-Today, 2016, στο <https://www.accenture.com/nz-en/insight-artificial-intelligence-business-value>, Βλ. επίσης M. Boyd, & N. Wilson, *όπ.παρ.*, σελ. 36-44.

⁴⁹⁶ Βλ. H. Mehr, *όπ.παρ.*, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf, Βλ. επίσης A. Thierer, A. O'Sullivan Castillo, & R. Russell, *όπ.παρ.*, στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁴⁹⁷ Βλ. J. Brown, T. Gosling, B. Sethi, B. Sheppard, C. Stubbings, J. Sviokla, W. Jon, Z. Daria, F. Liz, Workforce of the future: The competing forces shaping 2030, 2018, στο <https://www.pwc.com/gx/en/services/people-organisation/workforce-of-the-future/workforce-of-the-future-the-competing-forces-shaping-2030-pwc.pdf>, Βλ. επίσης J. Hawksorth, B. Kupelian, R.

Πανεπιστημίου της Οξφόρδης⁴⁹⁸ και του Παγκοσμίου Ινστιτούτου McKinsey⁴⁹⁹ καταλήγουν σε παρεμφερή συμπεράσματα, δηλώνοντας ότι περίπου τα μισά από τα επαγγέλματα ή τις δραστηριότητες του εργατικού δυναμικού θα μπορούσαν να αυτοματοποιηθούν και συνακόλουθα αυξάνεται ο εν λόγω κίνδυνος. Κάποιες επιπρόσθετες εκτιμήσεις αντλούν μία ακόμη πιο απαισιόδοξη εικόνα της κατάργησης της εργασίας που προκαλείται από την αυτοματοποίηση, η οποία προβλέπεται να αγγίζει ένα ποσοστό της τάξεως του 80-90% μέσα στα επόμενα 15 χρόνια⁵⁰⁰.

Αν και οι προσεγγίσεις που οδηγούν σε αυτά τα στοιχεία έχουν δεχθεί έντονη κριτική⁵⁰¹, καθίσταται προφανές ότι η τεχνητή νοημοσύνη προωθεί σε μεγάλο βαθμό την αντικατάσταση του ανθρώπινου δυναμικού και η κοινωνία θα αντιμετωπίσει την τεχνολογική ανεργία σε ευρεία κλίμακα. Εκτός από τη συγκεκριμένη πρόκληση της υποκατάστασης του εργατικού δυναμικού, η κάλυψη πολλών θέσεων εργασίας από την τεχνητή νοημοσύνη αναμένεται να μεταβάλλει τα χαρακτηριστικά των απαιτήσεών τους, με αποτέλεσμα οι εργαζόμενοι να βιώσουν μία μετατόπιση των ρόλων τους προς μία περισσότερο εποπτική λειτουργία η οποία θα απαιτεί και ειδικές δεξιότητες τεχνητής νοημοσύνης⁵⁰².

Η ανάπτυξη της τεχνητής νοημοσύνης οδηγεί, επίσης, στην εμφάνιση εντελώς νέων προφίλ εργασίας, όπως είναι οι επιστήμονες δεδομένων ή οι μηχανικοί μηχανικής μάθησης, οι οποίοι, επί του παρόντος, είναι λίγοι. Ο ιδιωτικός και ο δημόσιος τομέας

Berriman, & D. Mckellar, UK economic outlook, 2017, στο <https://www.pwc.co.uk/economic-services/ukeo/pwc-uk-economic-outlook-full-report-march-2017-v2.pdf>

⁴⁹⁸ Βλ. C.B. Frey, & M.A. Osborne, The future of employment. How Susceptible are Jobs to Computerisation. Technological Forecasting and Social Change, Ap. 114, 2017, σελ. 254– 280.

⁴⁹⁹ Βλ. J. Manyika, M. Chui, M. Miremadi, J. Bughin, K. George, P. Willmott, & M. Dewhurst, A future that works: Automation, employment, and productivity, 2017, στο https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works_Full-report.ashx

⁵⁰⁰ Βλ. R. Lever, Tech world debate on robots and jobs heats up, 2017, στο <https://phys.org/news/2017-03-tech-world-debate-robots-jobs.html>

⁵⁰¹ Βλ. R.D. Atkinson, In defense of robots. National Review, LXIX(7), 2017, στο <https://www.nationalreview.com/2017/04/robots-jobs-industrial-future/>

⁵⁰² Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf, Βλ. επίσης C. Bataller, & J. Harris, Turning artificial intelligence into business value-Today, 2016, στο <https://www.accenture.com/nz-en/insight-artificial-intelligence-business-value>

πρέπει να προετοιμαστούν και να αντιμετωπίσουν τις εν λόγω μεγάλες προκλήσεις, προκειμένου να προστατεύσουν τα οικονομικά συστήματα και την κοινωνία από τις δυσμενείς επιπτώσεις που δύνανται να προκύψουν από αυτήν τη μετάβαση⁵⁰³.

Η κοινωνική αποδοχή και εμπιστοσύνη στην τεχνητή νοημοσύνη συνδέεται στενά και εξαρτάται απόλυτα από την επίλυση λοιπών προκλήσεων, όπως είναι η ασφάλεια που πρέπει να συνοδεύει την ενσωμάτωσή της, η προστασία της ιδιωτικής ζωής, οι διακρίσεις και η αντικατάσταση του εργατικού δυναμικού που δύνανται να επηρεάσουν αρνητικά τους πολίτες με άμεσο τρόπο. Μία αναντιστοιχία μεταξύ των προσδοκιών των πολιτών ή των χρηστών και της πραγματικότητας των συστημάτων τεχνητής νοημοσύνης είναι πιθανό να ασκήσει αρνητική επιρροή στην αποδοχή της νέας τεχνολογίας⁵⁰⁴. Πέραν αυτού, ο συνδυασμός της ανεργίας και της αύξησης του κέρδους μέσω της τεχνητής νοημοσύνης δύναται να ενισχύσει την κοινωνική ανισότητα και να διευρύνει το χάσμα μεταξύ των πλουσίων και των φτωχών, το οποίο με τη σειρά του οδηγεί σε κοινωνική δυσαρέσκεια αντί κοινωνική αποδοχή⁵⁰⁵.

Επιπλέον, προκύπτει η πρόκληση να μειωθούν οι κοινωνικές ανησυχίες, να προετοιμαστεί η επόμενη γενιά για μία ψηφιακή εποχή ξεκινώντας και προωθώντας έναν κοινωνικό διάλογο προσβλέποντας στην οικοδόμηση εμπιστοσύνης προς τα συστήματα τεχνητής νοημοσύνης. Οι Hameed et al.⁵⁰⁶ τονίζουν τη σημασία της κατανόησης και της συνεκτίμησης των παραγόντων που οδηγούν στην αποδοχή των συστημάτων τεχνητής νοημοσύνης στην κοινωνία. Ωστόσο, πρέπει, σε γενικές

⁵⁰³ Βλ. EY, The growing impact of AI on business. Cambridge,; Massachusetts Institute of Technology, 2018, στο <https://www.technologyreview.com/s/611013/the-growing-impact-of-ai-on-business/>

⁵⁰⁴ Βλ. I.A. Hameed, Z.H. Ten, N.B. Thomsen, & D. Xiaodong, User acceptance of social robots, The Ninth International Conference on Advances in Computer-Human Interactions, 2016, σελ. 274–279.

⁵⁰⁵ Βλ. M. Boyd, & N. Wilson, *όπ.παρ.*, σελ. 36-44, Βλ.επίσης G. Wisskirchen, B.T. Biacabe, U. Bormann, A. Muntz, G. Niehaus, G. Jiménez Soler, & B.V. Brauchitsch, Artificial intelligence and robotics and their impact on the workplace, 2017, στο <https://www.ibanet.org/Document/Default.aspx?DocumentUId=c06aa1a3-d355-4866-beda-9a3a8779ba6e>

⁵⁰⁶ Βλ. I.A. Hameed, Z.H. Ten, N.B. Thomsen, & D. Xiaodong, *όπ.παρ.*, σελ. 274–279.

γραμμές, να έχουμε κατά νου ότι η αποδοχή της τεχνολογίας καθ' αυτή εξαρτάται από το εκάστοτε υφιστάμενο και επικρατές πλαίσιο⁵⁰⁷.

Όσον αφορά τον μετασχηματισμό της αλληλεπίδρασης H2M / M2M, αυτός συνδέεται με την κατανόηση της επικοινωνίας μεταξύ ανθρώπων και μηχανών ή μεταξύ των μηχανών τεχνητής νοημοσύνης και αναφέρεται στις προκλήσεις που σχετίζονται με αυτές τις αλληλεπιδράσεις. Λόγω των εξελίξεων στην ομιλία, την χειρονομία και την αναγνώριση προτύπων, η τεχνητή νοημοσύνη έχει ουσιαστικά τη δυνατότητα να μεταβάλλει τις αλληλεπιδράσεις μας⁵⁰⁸ και η μελλοντική αλληλεπίδραση H2M θεωρείται ότι χαρακτηρίζεται από συστήματα που μπορούν να προβλέψουν τις κινήσεις των ανθρώπων και να καθοδηγούνται από την φωνή, την επικοινωνία μέσω της γλώσσας του σώματος και από το περιβάλλον⁵⁰⁹. Εντούτοις, δεδομένου ότι τα περισσότερα συστήματα τεχνητής νοημοσύνης έχουν σχεδιαστεί και αναπτυχθεί από τεχνικούς εμπειρογνώμονες που δεν είναι εξοικειωμένοι με γνωστικές πτυχές, η αλληλεπίδραση H2M δεν θεωρείται ιδανική και θα πρέπει οι χρήστες να προσαρμοστούν στις εν λόγω μηχανές, με αποτέλεσμα την συχνά εσφαλμένη επικοινωνία και την πρόκληση αδικαιολόγητου άγχους⁵¹⁰.

Προκειμένου να υπερκεραστεί η συγκεκριμένη πρόκληση, θα πρέπει η αλληλεπίδραση H2M και M2M να καταστεί πιο διαισθητική, γεγονός που απαιτεί ισχυρότερη εξέταση των επιμέρους υφιστάμενων γνωστικών συστατικών⁵¹¹. Συνεπώς, ο χρήστης πρέπει,

⁵⁰⁷ Βλ. N. Petit, Artificial intelligence and automated law enforcement: A review paper, SSRN Journal, 2018, σελ. 1-12.

⁵⁰⁸ Βλ. C. Cath, S. Wachter, B. Mittelstadt, M. Taddeo, & L. Floridi, *όπ.παρ.*, σελ. 505–528.

⁵⁰⁹ Βλ. B. Ice, Gesture technology, invisible user interface and the evolution of human-to-machine interaction, 2015, στο <https://www.marketingmag.com.au/hubs-c/gesture-technology-invisible-user-interface-evolution-human-machine-interaction/>

⁵¹⁰ Βλ. S.G. Lee, & P. Sathikh, A framework for effective human-to-machine communication for artificial interactive systems, In U. Lindemann, S. Venkataraman, Y. S. Kim, S. W. Lee, P. Badke-Schaub, & K. Sato (Eds.), Proceedings of the 19th international conference on engineering design (ICED13), Design for harmonies, Τεύχος 7, Human behaviour in design, Korea, Design Society, 2013, σελ. 75-77.

⁵¹¹ Βλ. K. Ducatel, M. Bogdanowicz, F. Scapolo, J. Leitjen, & J.C. Burgelman, That's what friends are for - Ambient intelligence (AmI) and the information society in 2010, In A. Zerdick, A. Picot, K. Schrape, J.C. Burgelman, R. Silverstone, V. Feldmann, C. Wernick & C. Wolff (Eds.), E-merging media: Communication and the media economy of the future, Berlin, Springer, 2005, σελ. 181-200.

μεταξύ άλλων, να κατανοήσει τον τρόπο με τον οποίο η τεχνητή νοημοσύνη επηρεάζει πραγματικά την αλληλεπίδραση⁵¹².

Επιπροσθέτως, ο σχεδιασμός ενός συστήματος τεχνητής νοημοσύνης, εναρμονισμένος με τον χρήστη, απαιτεί πολλούς περισσότερους τρόπους επικοινωνίας του χρήστη με το σύστημα⁵¹³. Εκτός από τη λεκτική επικοινωνία, η μη λεκτική επικοινωνία είναι ένα σημαντικό χαρακτηριστικό της ανθρώπινης συμπεριφοράς και αλληλεπίδρασης. Η μη λεκτική επικοινωνία αντιπροσωπεύει πτυχές της επικοινωνίας που αφορούν τις εκφράσεις του προσώπου, τη στάση του σώματος και χειρονομίες⁵¹⁴. Η ικανότητα κατανόησης των συναισθηματικών εκφράσεων είναι ζωτικής σημασίας και αποτελεί σοβαρή πρόκληση για τα λογικά συστήματα τεχνητής νοημοσύνης⁵¹⁵, με αποτέλεσμα μία διαφορετική συμπεριφορά αλληλεπίδρασης τύπου H2M.

Αυτή η κατάσταση μπορεί, επίσης, να συμβάλει στο πρόβλημα της ομοιογένειας που αντιμετωπίζει σήμερα η αλληλεπίδραση τύπου H2M, σύμφωνα με το οποίο τα συστήματα τεχνητής νοημοσύνης, όπως είναι τα ρομπότ, είναι σε θέση να αλληλοεπιδρούν με τους ανθρώπους μόνο με ομοιόμορφο τρόπο και δεν έχουν τη δυνατότητα να τους προσεγγίσουν με μεταβλητό τρόπο σε κάθε λεπτομέρεια για να επιτύχουν μία πληθώρα αμοιόδρομων επικοινωνιών⁵¹⁶. Προκειμένου να αντικαταστήσει την αλληλεπίδραση μεταξύ ανθρώπων, η τεχνητή νοημοσύνη θα χρειαστεί μία

⁵¹² Βλ. L.E. Holmquist, Intelligence on tap: Artificial intelligence as a new design material, Interactions, Αρ. 24, Τεύχος 4, 2017, σελ. 28–33.

⁵¹³ Βλ. L.E. Holmquist, *όπ.παρ.*, σελ. 28–33, Βλ. επίσης P. Mankins, Can we design trust between humans and artificial intelligence, 2015, στο <https://www.fastcompany.com/3047500/can-we-design-trust-between-humans-and-artificial-intelligence>

⁵¹⁴ Βλ. T.J. Dodds, B.J. Mohler, & H.H. Bühlhoff, Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments, PLOS ONE, Αρ. 6, Τεύχος 10, 2011, σελ. 1-12.

⁵¹⁵ Βλ. S. Banerjee, S. Pradeep Kumar, & J. Bajpai, A comparative study on decision-making capability between human and artificial intelligence, In B. K. Panigrahi, M. N. Hoda, V. Sharma, & S. Goel (Eds.), Nature inspired computing: Proceedings of CSI 2015, Singapore, Springer, 2018, σελ. 203-210.

⁵¹⁶ Βλ. T. Tanaka, & K. Kobayashi, Developing a dividual model using a modular neural network for human-robot interaction, Journal of Robotics, Networking and Artificial Life, Αρ. 2, Τεύχος 1, 2015, σελ. 34–39.

πραγματική κατανόηση του τρόπου με τον οποίο λειτουργεί ο κόσμος και των προθέσεων του εκάστοτε μέρους αλληλεπίδρασης⁵¹⁷.

Επιπλέον, μία άλλη μεγάλη πρόκληση αφορά τις ταχέως αναπτυσσόμενες και αυξανόμενες αλληλεπιδράσεις τύπου M2M, οι οποίες απαιτούν κατάλληλο ανθρώπινο έλεγχο. Δεδομένου ότι οι αυτο-αλγόριθμοι καθίστανται όλο και πιο περίπλοκοι και οι άνθρωποι δεν διαδραματίζουν ενεργό ρόλο σε αυτές τις αλληλεπιδράσεις, αλλά μάλλον αναλαμβάνουν έναν μετριοπαθή ρόλο, ελλοχεύει ο κίνδυνος της απώλειας της πρόσβασης στις εν λόγω αλληλεπιδράσεις, οι οποίες, με τον τρόπο αυτόν, ενδέχεται να αψηφούν τον ανθρώπινο έλεγχο και την κατανόηση. Όπως και με τη διακυβέρνηση τεχνητής νοημοσύνης, σε γενικές γραμμές η πρόκληση έγκειται στη διασφάλιση του ότι οι αλληλεπιδράσεις τύπου M2M, στην προκειμένη περίπτωση, είναι διαφανείς και υπόλογες, αλλά και ελεγχόμενες από τον άνθρωπο⁵¹⁸.

⁵¹⁷ Βλ. L.E. Holmquist, *όπ.παρ.*, σελ. 28–33.

⁵¹⁸ Βλ. Z. Minchev, Challenges to human factor for advance persistent threats proactive identification in modern social networks, *Information & Security: An International Journal*, Ap. 34, 2016, σελ. 57–70.

4.2.3 Τρίτη Διάσταση: Ηθικά Ζητήματα της Ενσωμάτωσης της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα

Μία έντονα συζητημένη πρόκληση σε σχέση με την ενσωμάτωση της τεχνητής νοημοσύνης στον δημόσιο τομέα αναφέρεται στην ηθική της τεχνητής νοημοσύνης. Οι κύριες ηθικές πτυχές της τεχνητής νοημοσύνης αφορούν τον τομέα της ηθικής των ρομπότ ή της ηθικής των μηχανών. Από τη μία πλευρά, αυτό περιλαμβάνει εκτιμήσεις του κατά πόσον η ανάπτυξη και η χρήση ορισμένων εφαρμογών τεχνητής νοημοσύνης και οι συνέπειές τους είναι ηθικές και ηθικά δικαιολογημένες⁵¹⁹. Από την άλλη πλευρά, αφορούν το ζήτημα του πώς πρέπει να ενσωματωθούν οι ηθικές αρχές στα συστήματα αυτά προκειμένου να διασφαλιστεί ότι ενεργούν ηθικά⁵²⁰.

Η ηθική της τεχνητής νοημοσύνης δεν ακολουθεί μόνο τον κωδικοποιημένο νόμο, αλλά, μεταξύ άλλων, αναγνωρίζει κοινωνικούς κανόνες και πρότυπα που αναφέρονται σε εύλογες υποχρεώσεις, καθώς επίσης και εκείνους που περιλαμβάνουν τις αρετές της πίστης και της τιμιότητας. Αυτό απαιτεί, επιπροσθέτως, μία συνεχή μελέτη των ηθικών πεποιθήσεων και συμπεριφορών για την εξασφάλιση λογικών και βάσιμων προτύπων⁵²¹. Καθίσταται εναργές ότι η πρόκληση της ηθικής της τεχνητής νοημοσύνης καλύπτει ένα ευρύ φάσμα πτυχών, οι οποίες κυμαίνονται από τη διαμόρφωση κανόνων για την ανθρώπινη συμπεριφορά, τη συμβατότητα της κρίσης της μηχανής έναντι της ανθρώπινης αξίας, έως τα ηθικά διλήμματα και τις διακρίσεις της⁵²².

Η διαμόρφωση κανόνων τεχνητής νοημοσύνης για την ανθρώπινη συμπεριφορά αναφέρεται στις απόρροιες, για τον πληθυσμό, οι οποίες προκύπτουν από τη λήψη αποφάσεων με βάση την τεχνητή νοημοσύνη και αντιπροσωπεύει μία από τις μεγαλύτερες ηθικές προκλήσεις που σχετίζονται με αυτήν. Τα συστήματα τεχνητής νοημοσύνης έχουν σχεδιαστεί, κατά το πλείστον, για την αναπαραγωγή ή τη μίμηση

⁵¹⁹ Βλ. L. Patrick., K. Abney, & G.A. Bekey, Robot ethics: The ethical and social implications of robotics, Cambridge, MIT Press, 2012, σελ. 36-38, Βλ. επίσης S. Russell, Take a stand on AI weapons, Nature, Αρ. 521, Τεύχος 7553, 2015, σελ. 415-416.

⁵²⁰ Βλ. S.L. Anderson, όπ.παρ., σελ. 21-27.

⁵²¹ Βλ. M. Velasquez, C. Andre, T. Shanks, S.J. Meyer, & M.J. Meyer, What is ethics, 2010, στο <https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/>

⁵²² Βλ. L. Patrick., K. Abney, & G.A. όπ.παρ., σελ. 36-38.

της ανθρώπινης συμπεριφοράς και τη λήψη αποφάσεων αντί του ανθρώπου, με στόχο τη μεγιστοποίηση της αποτελεσματικότητας και της αποδοτικότητας, καθώς και την ελαχιστοποίηση των σφαλμάτων, μέσω της λήψης της σωστής ή καλύτερης απόφασης. Συνεπώς, τα συστήματα τεχνητής νοημοσύνης σκοπεύουν να σκέφτονται και να ενεργούν ορθολογικά και να αναπαράγουν τη φυσική διαδικασία λήψης αποφάσεων σε σύγκριση με τους ανθρώπους⁵²³.

Ωστόσο, καθώς τα συστήματα τεχνητής νοημοσύνης δεν είναι άψογα και ενδέχεται να προκύψουν λανθασμένες κρίσεις που να αποκλίνουν από την ανθρώπινη προοπτική, η λήψη αποφάσεων μέσω της τεχνολογίας, δύναται να αποτελέσει απειλή για τα ανθρώπινα όντα⁵²⁴. Για την αποφυγή εσφαλμένης κρίσης, τα συστήματα τεχνητής νοημοσύνης βασίζονται σε αμερόληπτα δεδομένα υψηλής ποιότητας, των οποίων η παροχή και αξιοπιστία αποτελεί από μόνη της μία ήδη σημαντική πρόκληση⁵²⁵.

Μία ακόμη σημαντική πρόκληση προκύπτει από το ότι οι άνθρωποι, σε αντίθεση με τα συστήματα τεχνητής νοημοσύνης, δεν ενεργούν απλώς οργανικά ορθολογικά, αλλά χαρακτηρίζονται και από επιπρόσθετα στοιχεία, τα οποία ασκούν επιρροή και είναι δυνατό να καθορίζουν, σε σημαντικό βαθμό, τις αποφάσεις και τις συμπεριφορές τους, όπως είναι η συνείδηση ή το συναίσθημα⁵²⁶. Αυτά τα χαρακτηριστικά ή οι έννοιες αποτελούν σημαντικές απαιτήσεις για την επίτευξη αυτού, που ο Solum⁵²⁷ αποκαλεί Συνταγματική Προσωπικότητα και η οποία εκφράζει την ισότητα των συστημάτων τεχνητής νοημοσύνης με τους ανθρώπους⁵²⁸. Δεδομένου ότι τα συστήματα τεχνητής νοημοσύνης δεν διαθέτουν τα εν λόγω ανθρώπινα χαρακτηριστικά, ενδέχεται να μην

⁵²³ Βλ. S. Banerjee, S. Pradeep Kumar, & J. Bajpai, *όπ.παρ.*, σελ. 203-210.

⁵²⁴ Βλ. A. Krausová, *όπ.παρ.*, 55–68.

⁵²⁵ Βλ. ΕΥ. *όπ.παρ.*, στο <https://www.technologyreview.com/s/611013/the-growing-impact-of-ai-on-business/>

⁵²⁶ Βλ. S. Banerjee, S. Pradeep Kumar, & J. Bajpai, *όπ.παρ.*, σελ. 203-210.

⁵²⁷ Βλ. L.B. Solum, Legal personhood for artificial intelligences, *North Carolina Law Review*, Ap. 70, Τεύχος 4, 1992, σελ. 1231–1288.

⁵²⁸ Βλ. A. Krausová, *όπ.παρ.*, 55–68.

έχουν δικαίωμα εμπλοκής και παρέμβασης στα δικαιώματα που είναι συνυφασμένα με τη Συνταγματική Προσωπικότητα⁵²⁹.

Σε συνέχεια των ανωτέρω, γεννάται το ερώτημα και η πρόκληση του εάν είναι ορθό και ηθικά δικαιολογημένο τα συστήματα τεχνητής νοημοσύνης να θεσπίζουν κανόνες για την ανθρώπινη συμπεριφορά. Επιπλέον, είναι ανοιχτό προς συζήτηση το εάν είναι επιθυμητό να τεθούν σε εφαρμογή συστήματα τεχνητής νοημοσύνης, τα οποία αποφασίζουν να αποσπαστούν από ανθρώπινα χαρακτηριστικά, όπως το συναίσθημα ή η συνείδηση, ειδικά σε καταστάσεις όπου η ανθρώπινη διάσταση και παρέμβαση πρέπει να διαδραματίζει σημαντικό ρόλο και να υπάρχει αντίστοιχο νομικό περιθώριο ή τουλάχιστον το πεδίο της διακριτικής ευχέρειας. Παραδείγματος χάριν, είναι προβληματικό ότι μία απόφαση είναι δυνατόν να αλλάξει απλώς λόγω του ότι μία μηχανή το αποφασίζει και όχι ένας άνθρωπος⁵³⁰. Στο σημείο αυτό, τα συστήματα τεχνητής νοημοσύνης αυτομάθησης μπορούν να μην αποφασίζουν πλέον σε κανονιστική ηθική βάση και ως εκ τούτου να καταστούν απρόβλεπτα για τα ανθρώπινα όντα. Επί παραδείγματι, το προαναφερθέν σύστημα αυτοματοποίησης της διαδικασίας τεχνητής νοημοσύνης για φόρμες αίτησης μετανάστευσης ενδέχεται να απορρίψει ένα ερώτημα λόγω προκατειλημμένων δεδομένων, ενώ ένας εργαζόμενος, επί τη βάσει των ίδιων δεδομένων, θα μπορούσε να αποδεχτεί το αίτημα⁵³¹.

Η συμβατότητα της κρίσης της μηχανής, έναντι της ανθρώπινης αξίας, βασίζεται στην υπόθεση ότι η κρίση της ανθρώπινης αξίας διαφέρει από την πιθανή κρίση της αξίας της μηχανής και συνεπώς, συνδέεται στενά με τις προαναφερθείσες συζητήσεις σχετικά με τη διαμόρφωση κανόνων τεχνητής νοημοσύνης για την ανθρώπινη συμπεριφορά⁵³². Λαμβάνοντας υπόψη ότι η λήψη αποφάσεων από τον άνθρωπο, σε αντίθεση με εκείνη των συστημάτων τεχνητής νοημοσύνης, δεν επηρεάζεται μόνο από ορθολογικές αρχές, αλλά και από παράγοντες όπως το συναίσθημα και η συνείδηση, συνακόλουθο είναι ότι οι εκτιμήσεις της αξίας τους μπορεί να διαφέρουν σε ορισμένες καταστάσεις⁵³³.

⁵²⁹ Βλ. L.B. Solum, *όπ.παρ.*, σελ. 1231–1288.

⁵³⁰ Βλ. S. Vallor, G.A. & Bekey, *όπ.παρ.*, σελ. 338-353.

⁵³¹ Βλ. H.Y. Liu, & K. Zawieska, *όπ.παρ.*, σελ. 321-333.

⁵³² Βλ. S. Banerjee, S. Pradeep Kumar, & J. Bajpai, *όπ.παρ.*, σελ. 203-210.

⁵³³ Βλ. S. Banerjee, S. Pradeep Kumar, & J. Bajpai, *όπ.παρ.*, σελ. 203-210.

Στο εν λόγω πεδίο, οι έρευνες και μελέτες που έχουν ήδη διεξαχθεί έχουν επιστήσει, επίσης, την προσοχή τους στην πρόκληση της δημιουργίας ηθικής συνοχής μεταξύ των συστημάτων τεχνητής νοημοσύνης και των ανθρώπων. Πρόκειται για μία συνοχή η οποία είναι δυνατόν να προκύψει σε σχέση με τη λήψη αποφάσεων με βάση την τεχνητή νοημοσύνη⁵³⁴. Ο Turilli, για παράδειγμα, επισημαίνει τη σημασία της ευθυγράμμισης των συστημάτων τεχνητής νοημοσύνης με τις ίδιες ηθικές αρχές των ανθρώπων, προκειμένου να επιτευχθεί ηθική συνέπεια για τη συνολική συμπεριφορά ή απόδοση⁵³⁵.

Εντούτοις, παρά τις φιλοδοξίες του πεδίου της ηθικής της μηχανής να ενσωματώσει ηθικές αρχές σε συστήματα τεχνητής νοημοσύνης για να διασφαλίσει ότι ενεργούν ηθικά⁵³⁶, οι ηθικές αρχές, όπως χρησιμοποιούνται από τους υπεύθυνους λήψης αποφάσεων, μπορεί να αποδειχθεί δύσκολο να καθοριστούν και να υπολογιστούν⁵³⁷. Επομένως, ο αλγόριθμος αυτομάθησης και η προκύπτουσα αυτονομία μηχανής διαδραματίζουν καθοριστικό ρόλο. Τα συστήματα τεχνητής νοημοσύνης μπορεί να αναπτύξουν το δικό τους σύστημα αξιών και να δημιουργήσουν το δικό τους πλαίσιο αναφοράς, το οποίο δύναται να καθίσταται ασυμβίβαστο με το ανθρώπινο σύστημα αξιών. Στη χειρότερη περίπτωση, η μηχανική κρίση θα αντιταχθεί στην κρίση της ανθρώπινης αξίας και αυτό είναι δυνατόν να οδηγήσει σε μία ανεξάρτητη τεχνολογία τεχνητής νοημοσύνης η οποία θα τείνει να προκαλεί βλάβες και να στρέφεται εναντίον των ανθρώπων⁵³⁸.

Τα ηθικά διλήμματα αναφέρονται σε καταστάσεις, στις οποίες τα συστήματα τεχνητής νοημοσύνης πρέπει να επιλέξουν μεταξύ αντικρουόμενων εναλλακτικών λύσεων,

⁵³⁴ Βλ. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, & L. Floridi, The ethics of algorithms: Mapping the debate, *Big Data & Society*, Αρ.3, Τεύχος 2, 2016, σελ. 1–21.

⁵³⁵ Βλ. M. Turilli, Ethical protocols design, *Ethics and Information Technology*, Αρ.9, Τεύχος 1, 2007, 49–62.

⁵³⁶ Βλ. M. Anderson, & S.L. Anderson, *Machine ethics*, New York, Cambridge University Press, 2011, σελ. 26–39.

⁵³⁷ Βλ. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, & L. Floridi, *όπ.παρ.*, σελ. 11

⁵³⁸ Βλ. H.Y. Liu, & K. Zawieska, *όπ.παρ.*, σελ. 321–333.

δεδομένου ότι καμία ηθική επιλογή δεν έχει ανεπιθύμητες ηθικές συνέπειες⁵³⁹. Τέτοια διλήμματα τίθενται, επί παραδείγματι, όταν πρόκειται για εφαρμογές τεχνητής νοημοσύνης, όπως είναι τα ρομπότ υγειονομικής περίθαλψης ή τα οχήματα αυτόνομης οδήγησης, οι οποίες μπορεί να βοηθήσουν ή να βλάψουν τα ανθρώπινα όντα και όπου τα συστήματα πρέπει να λάβουν ηθική απόφαση μεταξύ δύο αρνητικών εναλλακτικών λύσεων⁵⁴⁰.

Για παράδειγμα, σε περίπτωση κατά την οποία ένα αυτόνομο όχημα πρόκειται να εμπλακεί σε ατύχημα, το σύστημα τεχνητής νοημοσύνης μπορεί να αντιμετωπίσει μία απόφαση που έχει μόνο δύο κακές επιλογές: ή θα πρέπει να προστατέψει τους επιβάτες του με κάθε κόστος για το άλλο εμπλεκόμενο όχημα, ή το αντίστροφο. Αυτό το παράδειγμα καθιστά σαφές ότι ορισμένες εφαρμογές, ιδίως εκείνες που περιλαμβάνουν αποφάσεις ζωής ή θανάτου, πρέπει να διαθέτουν μία ηθική ικανότητα συλλογιστικής, πριν εφαρμοστούν πλήρως αυτόνομα στον δημόσιο τομέα⁵⁴¹.

Όπως αναφέρθηκε προηγουμένως, το πεδίο της ηθικής της μηχανής αντιμετωπίζει την πρόκληση της ενσωμάτωσης ηθικών αρχών σε συστήματα τεχνητής νοημοσύνης και της διδασκαλίας τους για ηθική συμπεριφορά⁵⁴². Ωστόσο, η μεγάλη πρόκληση παραμένει να καταστεί κατανοητό τι είναι σημαντικό να αιτιολογηθεί επιτυχώς σε ηθικές καταστάσεις που είναι συνδεδεμένες με την τεχνητή νοημοσύνη⁵⁴³.

Επιπροσθέτως, ένα ακόμα φλέγον ζήτημα είναι αυτό που αφορά την πρόκληση της πρόληψης της ανισότητας και της αδικίας που προκαλείται από εφαρμογές τεχνητής νοημοσύνης. Προηγούμενες έρευνες καταδεικνύουν ότι τα συστήματα τεχνητής νοημοσύνης και το προφίλ αυτών, είναι δυνατόν να οδηγήσουν σε διακρίσεις και ως

⁵³⁹ Βλ. P.H. Ditto, & B. Liu, Deontological dissonance and the consequentialist crutch, In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil*, Washington, American Psychological Association, 2012, σελ. 51-70.

⁵⁴⁰ Βλ. B. Deng, Machine ethics: The robot's dilemma, *Nature*, Ap. 523, Τεύχος 7558, 2015, σελ. 24–26.

⁵⁴¹ Βλ. V. Conitzer, W. Sinnott-Armstrong, J.S. Borg, Y. Deng, & M. Kramer, Moral decision making frameworks for artificial intelligence, *Proceedings of the 31st AAAI Conference on Artificial Intelligence*, AAAI 2017, San Francisco, USA, 2017, σελ. 4831–4835.

⁵⁴² Βλ. M. Anderson, & S.L. Anderson, *όπ.παρ.*, σελ. 46-51.

⁵⁴³ Βλ. B. Deng, *όπ.παρ.*, σελ.2 5.

εκ τούτου να παραβιάζουν ηθικές αρχές, όπως η ισότητα και η δικαιοσύνη⁵⁴⁴. Δεδομένου ότι οι άνθρωποι προγραμματίζουν τους αλγόριθμους των συστημάτων τεχνητής νοημοσύνης ή χρησιμεύουν ως πηγή εισόδου δεδομένων, τα συστήματα αυτά δύνανται, επίσης, να λάβουν τα δεδομένα και τις προκαταλήψεις τους από τους ανθρώπους⁵⁴⁵. Για παράδειγμα, τα συστήματα τεχνητής νοημοσύνης μπορεί να απορροφήσουν το φύλο ή τις φυλετικές προκαταλήψεις που κρύβονται στα πρότυπα της ανθρώπινης γλώσσας και ακολούθως να ενισχύουν τα ανθρώπινα στερεότυπα⁵⁴⁶.

Ως απάντηση στα εν λόγω αναφερόμενα μειονεκτήματα και ενδεχομένως ακόμη και στην υπόσταση παράνομων διακρίσεων, ποικίλες προσεγγίσεις εμφανίζονται ολοένα και περισσότερο στο πλαίσιο της εξόρυξης δεδομένων και της μηχανικής μάθησης για τον εντοπισμό και την πρόληψη των διακρίσεων⁵⁴⁷. Εντούτοις, παρά την υπόσταση αυτών των εμβρυικών προσπαθειών για την εξουδετέρωση των διακρίσεων σε σχέση με την τεχνητή νοημοσύνη, η πρόκληση της εξάλειψης των διακρίσεων από τα συστήματα αυτά παραμένει και ως εκ τούτου μένει ζωντανή και η αναγκαιότητα, καθώς και η απαίτηση για περαιτέρω μηχανισμούς ανίχνευσης και πρόληψης αντίστοιχων καταστάσεων⁵⁴⁸.

⁵⁴⁴ Βλ. A. Thierer, A. O'Sullivan Castillo, & R. Russell, *όπ.παρ.*, στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁵⁴⁵ Βλ. D.K. Citron, & F.A. Pasquale, The scored society: Due process for automated predictions. *Washington Law Review*, Ap.89, Τεύχος 1, 2014, σελ. 1–33.

⁵⁴⁶ Βλ. D. Bass, & E. Huet, Researchers combat gender and racial bias in artificial intelligence, 2017, στο <https://www.bloomberg.com/news/articles/2017-12-04/researchers-combat-gender-and-racial-bias-in-artificial-intelligence>, Βλ. επίσης H. Devlin, AI programs exhibit racial and gender biases, research reveals, 2017, στο <https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals>

⁵⁴⁷ Βλ. S. Barocas, Data mining and the discourse on discrimination, *Proceedings of the Data Ethics Workshop, Conference on Knowledge Discovery and Data Mining (KDD)*, New York, 2014, στο <https://www.cs.yale.edu/homes/jf/Barocas-Taxonomy.pdf>, Βλ. επίσης T. Bolukbasi, K.W. Chang, J.Y. Zou, V. Saligrama, & A.T. Kalai, Man is to computer programmer as woman is to homemaker? Debiasing word embeddings, In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems 29*, *Neural information processing systems 2016*, Barcelona, Curran Associates, 2016, σελ. 4349-4357, Βλ. επίσης S. Hajian, & J. Domingo-Ferrer, A methodology for direct and indirect discrimination prevention in data mining, *IEEE Transactions on Knowledge and Data Engineering*, Ap.25, Τεύχος 7, 2013, σελ. 1445–1459.

⁵⁴⁸ Βλ. B.D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, & L. Floridi, *όπ.παρ.*, σελ. 1-21.

4.2.4 Τέταρτη Διάσταση: Τεχνολογικά Ζητήματα της Ενσωμάτωσης της Τεχνητής Νοημοσύνης στον Δημόσιο Τομέα

Σε γενικές γραμμές, η εφαρμογή της τεχνολογίας είναι συνυφασμένη με τις πτυχές από τις οποίες συνίσταται μία πρωτοβουλία ή δράση και με τις προϋποθέσεις και τους όρους που συνοδεύουν τη δράση αυτή⁵⁴⁹. Επομένως, η ενσωμάτωση της τεχνητής νοημοσύνης στο δημόσιο τομέα απαιτεί μία στοχαστική και στρατηγική πορεία δράσης, για να υπάρξει άμεσο όφελος από τις πολλά υποσχόμενες δυνατότητές της και τελικά να δημιουργηθεί αξία από αυτήν⁵⁵⁰.

Όμως, παρά το γεγονός ότι κυβερνητικοί οργανισμοί, παγκοσμίως, έχουν ξεκινήσει πρωτοβουλίες για την εφαρμογή και την υιοθέτηση της τεχνητής νοημοσύνης στον δημόσιο τομέα, η αξιοποίησή της αποτελεί μία απαιτητική πρόκληση για τον τομέα αυτόν⁵⁵¹. Σύμφωνα με τους Thierer et al.⁵⁵², το μεγαλύτερο μέρος των εφαρμοσμένων λύσεων τεχνητής νοημοσύνης θα παραμείνει, πιθανώς, αδύναμο και εξαιρετικά εξειδικευμένο μέσα στις επόμενες δεκαετίες. Αυτός ο φαινομενικός αγώνας εφαρμογής της τεχνητής νοημοσύνης στον δημόσιο τομέα συνδέεται με ορισμένα ζητήματα που αναστέλλουν την πρόοδό της. Βάσει κάποιων διεξαχθεισών ερευνών, είναι δυνατός ο εντοπισμός τεσσάρων ζητημάτων που σχετίζονται με τη μεγάλη πρόκληση της εφαρμογής της τεχνολογίας της τεχνητής νοημοσύνης, όπως είναι για παράδειγμα η ασφάλεια της τεχνητής νοημοσύνης, η ποιότητα και ενσωμάτωση συστημάτων και δεδομένων, η οικονομική σκοπιμότητα, καθώς και η εξειδίκευση⁵⁵³.

⁵⁴⁹ Βλ. J.A. Durlak, & E.P. DuPre, Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation, *American Journal of Community Psychology*, Ap. 41, Τεύχος 3– 4, 2008, σελ. 327–350.

⁵⁵⁰ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁵⁵¹ Βλ. C. Cath, όπ.παρ., σελ. 1-8.

⁵⁵² Βλ. A. Thierer, A. O’Sullivan Castillo, & R. Russell, όπ.παρ., στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁵⁵³ Βλ. A. Thierer, A. O’Sullivan Castillo, & R. Russell, όπ.παρ., στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

Η ασφάλεια της τεχνητής νοημοσύνης έχει αναφερθεί σε προηγούμενη έρευνα ως σημαντικός παράγοντας κινδύνου ή ως μία ακόμα πρόκληση της τεχνητής νοημοσύνης και αναφέρεται στην εξασφάλιση της ασφαλούς απόδοσης και του αντίκτυπου της υιοθέτησης και αξιοποίησης αυτής⁵⁵⁴. Πρόκειται για μία τοποθέτηση που δεν περιλαμβάνει απλώς θέματα ασφάλειας πληροφοριών, αλλά ευρύτερα θέματα ασφάλειας. Σύμφωνα με αυτήν την άποψη, περίπλοκες και κρίσιμες για την ασφάλεια καταστάσεις, οι οποίες προκύπτουν από περιστάσεις, όπως είναι παραδείγματος χάριν ότι οι μηχανές τεχνητής νοημοσύνης μπορούν να αφομοιώσουν μία αρνητική συμπεριφορά ή να παρεξηγήσουν και παρερμηνεύσουν κάποιες υφιστάμενες συνθήκες του περιβάλλοντος στο οποίο δρουν, περιλαμβάνονται επίσης σε αυτήν την τοποθέτηση⁵⁵⁵. Στο εν λόγω αναφερόμενο πεδίο, λοιπόν, οι Bostrom & Yudkowsky υπογραμμίζουν τη σημασία και την ανάγκη, η τεχνολογία τεχνητής νοημοσύνης να είναι ανθεκτική έναντι των δυσμενών χειρισμών από τους ανθρώπους⁵⁵⁶.

Η Google, μία από τις κορυφαίες εταιρείες στην έρευνα τεχνητής νοημοσύνης, έχει εντοπίσει πολλαπλά ζητήματα ασφάλειας που έχουν ήδη συμβεί στην πράξη. Στην περίπτωση των εφαρμογών που βασίζονται στην ενισχυτική μάθηση, πρέπει να διασφαλιστεί ότι το σύστημα της τεχνητής νοημοσύνης μαθαίνει χωρίς να εκτελεί καταστροφικές ενέργειες. Επιπλέον, είναι απαραίτητο να αποφευχθούν αρνητικές παρενέργειες, όπως είναι η διατάραξη του εργασιακού περιβάλλοντος, κατά την εκτέλεση των εργασιών για τις οποίες η μηχανή είναι πραγματικά κατασκευασμένη⁵⁵⁷. Επί παραδείγματι, ένα ρομπότ που βοηθά σε μία χειρουργική επέμβαση θα πρέπει να δύναται να μάθει, χωρίς να θέτει σε κίνδυνο τον ασθενή, δοκιμάζοντας διάφορες μεθόδους χειρουργικής επέμβασης. Κατά συνέπεια, η εφαρμογή και η πρόοδος της τεχνητής νοημοσύνης συνδέεται άμεσα με την πρόληψη ατυχημάτων και την

⁵⁵⁴ Βλ. M. Boyd, & N. Wilson, *όπ.παρ.*, σελ. 36-44.

⁵⁵⁵ Βλ. A. Conn, *Artificial intelligence: The challenge to keep it safe*, San Francisco, CA: Future of Life Institute, 2017, στο <https://futureoflife.org/2017/09/21/safety-principle/>

⁵⁵⁶ Βλ. N. Bostrom, A. Dafoe, & C. Flynn, *όπ.παρ.*, στο <https://www.fhi.ox.ac.uk/wp-content/uploads/Policy-Desiderata-in-the-Development-of-Machine-Superintelligence.pdf>

⁵⁵⁷ Βλ. D. Amodei, C., Olah, J. Steinhardt, P. Christiano, J. Schulman, & D. Mané, *όπ.παρ.*, στο <https://arxiv.org/pdf/1606.06565.pdf>

εξασφάλιση της ασφαλούς λειτουργίας των εφαρμογών της για την προστασία της ανθρωπότητας⁵⁵⁸.

Η ποιότητα και η ενσωμάτωση του συστήματος και των δεδομένων έχουν βαρύνουσα σημασία, λόγω του ότι το σύστημα τεχνητής νοημοσύνης οφείλει την «ευφυΐα» του στα παρεχόμενα δεδομένα από τα οποία μαθαίνει και ως εκ τούτου, τα δεδομένα θεωρούνται ως η «καρδιά» των υφιστάμενων συστημάτων τεχνητής νοημοσύνης⁵⁵⁹. Συγκεκριμένα, δεδομένα χαμηλής ποιότητας ή μη αξιόπιστα δεδομένα αντιπροσωπεύουν μία σημαντική πρόκληση για τους οργανισμούς⁵⁶⁰ και ακολούθως, η συλλογή, η συγκέντρωση, η αποθήκευση και η χρήση αμερόληπτων και σχετικών δεδομένων είναι απαραίτητη για την επιτυχή ενσωμάτωση της τεχνητής νοημοσύνης στον δημόσιο τομέα, καθώς τα ανακριβή ή αναξιόπιστα δεδομένα είναι δυνατόν να οδηγήσουν σε αποτυχίες⁵⁶¹.

Υπό αυτές τις συνθήκες, η δημιουργία ενός εξελιγμένου και υψηλής ποιότητας συστήματος τεχνητής νοημοσύνης που είναι ικανό να ενσωματώνει δεδομένα και να διαχειρίζεται τις αλληλεξαρτήσεις μεταξύ των δεδομένων, των τεχνολογιών και των διαδικασιών είναι απαραίτητη, αλλά ταυτόχρονα αποτελεί ιδιαίτερη πρόκληση στην εφαρμογή λύσεων τεχνητής νοημοσύνης⁵⁶².

Η οικονομική σκοπιμότητα διαδραματίζει, συν τοις άλλοις, σημαντικό ρόλο στην εφαρμογή της τεχνολογίας τεχνητής νοημοσύνης και ο ανεπαρκής προϋπολογισμός αποτελεί μία εκ των μεγαλύτερων προκλήσεων που αντιμετωπίζουν οι οργανισμοί κατά την έναρξη προγραμμάτων τεχνητής νοημοσύνης⁵⁶³. Κατά συνέπεια, πριν από την

⁵⁵⁸ Βλ. M. Boyd, & N. Wilson, *όπ.παρ.*, σελ. 36-44.

⁵⁵⁹ Βλ. A. Thierer, A. O'Sullivan Castillo, & R. Russell, *όπ.παρ.*, στο <https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf>

⁵⁶⁰ Βλ. EY, *όπ.παρ.*, στο <https://www.technologyreview.com/s/611013/the-growing-impact-of-ai-on-business/>

⁵⁶¹ Βλ. H. Mehr, *όπ.παρ.*, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁵⁶² Βλ. P. Gerbert, M. Hecker, S. Steinhäuser, & P. Ruwolt, Putting artificial intelligence to work. Edited by BCG Henderson Institute, Munich, Germany, The Boston Consulting Group, 2017, στο <https://www.bcg.com/publications/2017/technology-digital-strategy-putting-artificial-intelligence-work>

⁵⁶³ Βλ. EY, *όπ.παρ.*, στο <https://www.technologyreview.com/s/611013/the-growing-impact-of-ai-on-business/>, Βλ. επίσης PwC, Sherlock in health: How artificial intelligence may improve quality and

ανάπτυξη και την έναρξη μίας εφαρμογής τεχνητής νοημοσύνης, θα πρέπει να εξετάζεται το συνολικό κόστος και τα αναμενόμενα έσοδα, προκειμένου να εκτιμηθεί εάν μία λύση τεχνητής νοημοσύνης είναι βιώσιμη. Οπότε, υπάρχουν δύο βασικοί παράγοντες κόστους που καθιστούν τη χρηματοοικονομική σκοπιμότητα σημαντική πρόκληση, κατά την ενσωμάτωση της τεχνητής νοημοσύνης στον δημόσιο τομέα. Αφενός η επένδυση για τη δημιουργία μίας εξελιγμένης τεχνολογικής υποδομής για τη συλλογή και την αποθήκευση δεδομένων είναι ιδιαίτερος υψηλού κόστους⁵⁶⁴ και αφετέρου υπάρχει μεγάλη ζήτηση για περιορισμένο αριθμό εμπειρογνομόνων τεχνητής νοημοσύνης, η οποία σχετίζεται με την αύξηση του κόστους εκπαίδευσης και αύξηση των μισθών⁵⁶⁵.

Οι εμπειρογνώμονες θα διαθέτουν κατάλληλες δεξιότητες ώστε να είναι σε θέση να υποστηρίξουν και να προωθήσουν τις επερχόμενες καινοτομίες⁵⁶⁶. Ήδη, τα τελευταία χρόνια, η παγκόσμια ζήτηση για εμπειρογνώμονες τεχνητής νοημοσύνης έχει αυξηθεί εκθετικά⁵⁶⁷, ωστόσο, όπως έχει ήδη αναφερθεί, υπάρχει έλλειψη ειδικών και εξειδικευμένων επαγγελματιών σε θέματα τεχνητής νοημοσύνης, γεγονός που δημιουργεί έναν ανασταλτικό παράγοντα στο πλαίσιο της ανάπτυξης και εφαρμογής της στον δημόσιο τομέα⁵⁶⁸.

Στο συγκεκριμένο ζήτημα, η κυβέρνηση του κάθε κράτους διαδραματίζει πρωταγωνιστικό ρόλο και πρέπει να δώσει ιδιαίτερη έμφαση στην ανάπτυξη και την

efficiency, whilst reducing health-care costs in Europe, 2017, στο <https://www.pwc.de/de/gesundheitswesen-und-pharma/studie-sherlock-in-health.pdf>

⁵⁶⁴ Βλ. A. Roberts, Five big challenges to AI adoption and success, 2017, στο <https://www.clickz.com/five-big-challenges-to-ai-adoption-and-success/112795/>

⁵⁶⁵ Βλ. J. Bughin, E. Hazan, S. Ramaswamy, M. Chui, T., Allas, P. Dahlström, H. Nicolaus., T. Monica, Artificial intelligence: The next digital frontier, 2017, στο <https://www.mckinsey.com/~media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx>

⁵⁶⁶ Βλ. J. Holdren, & M. Smith, όπ.παρ., στο https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

⁵⁶⁷ Βλ. J.F. Gagné, The global artificial intelligence talent report, 2018, στο <https://jfgagne.ai/talent/>

⁵⁶⁸ Βλ. ΕΥ. όπ.παρ., στο <https://www.technologyreview.com/s/611013/the-growing-impact-of-ai-on-business/>

προώθηση ενός καλά μορφωμένου και ποικίλου εργατικού δυναμικού, προκειμένου να είναι ικανό να οικοδομήσει και να δημιουργήσει μία βιώσιμη, αξιόπιστη και έγκυρη βάση γνώσεων σχετικά με την τεχνητή νοημοσύνη⁵⁶⁹.

⁵⁶⁹ Βλ. J. Holdren, & M. Smith, *όπ.παρ.*, στο https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

4.3 Η Τεχνητή Νοημοσύνη και οι Διατομεακές Συνεργασίες στον Δημόσιο Τομέα

4.3.1 Η Χαρτογράφηση των Διατομεακών Συνεργασιών στον Δημόσιο Τομέα μέσω της Αξιοποίησης της Τεχνητής Νοημοσύνης

Η παροχή υπηρεσιών από τους δημόσιους οργανισμούς, συχνά, πραγματοποιείται μέσω της υιοθέτησης διαφορετικών οργανωτικών δομών. Μέσω των εν λόγω δομών, αρκετές φορές, ο δημόσιος τομέας καταλήγει να έρχεται σε ευθυγράμμιση με τον τρόπο λειτουργίας και οργάνωσης των νομικών οντοτήτων του ιδιωτικού τομέα⁵⁷⁰. Σε αυτό το πλαίσιο εμπίπτουν και τα συνεργατικά σχήματα που αναπτύσσονται μεταξύ δημοσίων και ιδιωτικών οργανισμών, τα οποία αντικατοπτρίζονται στο πρίσμα τυποποιημένων δομών. Χαρακτηριστικά παραδείγματα τέτοιου είδους δομών είναι οι συμπράξεις δημόσιου και ιδιωτικού τομέα.

Η διατομεακή συνεργασία έχει αποκτήσει παγκόσμια διάσταση σημασία, ιδίως στις ευρωπαϊκές χώρες⁵⁷¹. Στο Ηνωμένο Βασίλειο έχουν αναπτυχθεί περισσότερες από επτακόσιες συμπράξεις δημόσιου και ιδιωτικού τομέα, η αξία των οποίων καταλήγει να ξεπερνά το ποσό των 50 δισεκατομμυρίων λιρών για τη δημιουργία νοσοκομειακών και σχολικών μονάδων, φυλακών, δρόμων, γεφυρών και λοιπών συγκοινωνιακών και οδικών υποδομών, καθώς επίσης και στρατιωτικού εξοπλισμού⁵⁷². Παρά τα υψηλά επίπεδα δημοτικότητας που φέρουν, οι εν λόγω συνεργασίες συνεπάγονται πολλές διαχειριστικές πολυπλοκότητες⁵⁷³ και, ως απόρροια τούτων, ένα μεγάλο ποσοστό τους,

⁵⁷⁰ Βλ. J.D. Donahue, & R.J. Zeckhauser RJ., Collaborative governance: private roles for public goals in turbulent times, Princeton, Princeton University Press, 2011, σελ. 37-46.

⁵⁷¹ Βλ. EPEC, A Guide to Preparing and Procuring a PPP Project Public-Private Partnerships in the Western Balkans, 2018, στο <https://www.wbif.eu/storage/app/media/Library/8.%20Public%20Private%20Partnership/3.%203-PPP-Preparation-and-Procurement-Guide-FINAL-310818.pdf>

⁵⁷² Βλ. HM Treasury, Private Finance Initiative Projects: 2013 Summary Data, Report number PU 1587, 2013, στο <https://www.gov.uk/government/publications/private-finance-initiative-projects-2013-summary-data>

⁵⁷³ Βλ. C. Huxham, Theorizing collaboration practice, Public Management Review, Αρ. 3, 2003, σελ. 401-423, Βλ. επίσης T. Bovaird, Public-private partnerships: from contested concepts to prevalent practice, International Review of Administration Science, Αρ. 70, 2004, σελ. 199-215.

συχνά, δεν καταγράφει ικανοποιητικά αποτελέσματα⁵⁷⁴. Βέβαια, αυτό δεν συνεπάγεται την πλήρη απουσία περιπτώσεων που μπορούν να χαρακτηριστούν ως επιτυχημένες⁵⁷⁵.

Άξια αναφοράς είναι η υπόσταση τριμερών συνεργατικών ρυθμίσεων επί του κλάδου της τεχνητής νοημοσύνης. Για παράδειγμα, το Πανεπιστήμιο του Essex προχώρησε σε σύμπραξη με το «Essex County Council», ιδρύοντας το «Ινστιτούτο Ανάλυσης και Επιστήμης των Δεδομένων (Institute for Analytics and Data Science - IADS)». Το Ινστιτούτο Ανάλυσης και Επιστήμης Δεδομένων αποτελεί ένα κέντρο αριστείας του Πανεπιστημίου του Essex, στο πλαίσιο του οποίου «συνδέονται» μελετητές, επιχειρήσεις, ιδρύματα και κυβερνήσεις, για τη διεκπεραίωση έργων άμεσα συναρτημένων με την τεχνητή νοημοσύνη. Ο στόχος της σύμπραξης αυτής έγκειται στην αξιοποίηση των διαθέσιμων πόρων και δεδομένων του δημόσιου τομέα, με την κατοχή τεχνογνωσίας επί ζητημάτων τεχνητής νοημοσύνης, που είναι διαθέσιμη αφενός στο Πανεπιστήμιο και αφετέρου στις επιχειρήσεις, με απώτερο σκοπό την παροχή υπηρεσιών προς όφελος της κοινότητας του Essex⁵⁷⁶.

Οι κυβερνήσεις ανά τον κόσμο προβαίνουν σε συνεργασίες με τα πανεπιστήμια και διάφορους λοιπούς φορείς, μέσω πλατφορμών εργαστηριακής πολιτικής, προσβλέποντας στον συνδυασμό και στη συλλογική αξιοποίηση διαφορετικών δυνατοτήτων επίλυσης προκύπτοντων προβλημάτων. Τα εν λόγω διατομεακά εργαστήρια βοηθούν στη συνέργεια της γνώσης, έτσι ώστε η κυβέρνηση να καθίσταται ικανή να εργαστεί για την ανεύρεση λύσεων επί ζητημάτων που είναι συνυφασμένα με το πεδίο της επιστήμης της ανάλυσης δεδομένων και της τεχνητής νοημοσύνης, σε περιπτώσεις πολυδιάστατων τομέων που περιλαμβάνουν από ιατρικής φύσεως ζητήματα έως και θέματα που εμπίπτουν στο πλαίσιο των λεγόμενων «έξυπνων» πόλεων. Επί παραδείγματι, το GovLab των ΗΠΑ, προβαίνει στη δημιουργία εφαρμοσμένων ερευνητικών πλαισίων, τα οποία λειτουργούν υποστηρικτικά προς τις κυβερνητικές αρχές στην προσέγγιση των υφιστάμενων και μελλοντικών πιθανών προβλημάτων. Επίσης, επισημαίνεται ότι η «Μεθοδολογία της Καινοτομίας», που αναπτύχθηκε από το People's Lab, χρησιμοποιείται από ομάδες ειδημόνων,

⁵⁷⁴ Βλ. W.L. Waugh, & G. Streib, Collaboration and leadership for effective emergency management. *Public Administration Review*, Ap. 66, 2006, 131–140.

⁵⁷⁵ Βλ. G.A. Hodge, & C. Greve, *όπ.παρ.*, σελ. 545–558.

⁵⁷⁶ Βλ. G.A. Hodge, & C. Greve, *όπ.παρ.*, σελ. 545–558.

αποσκοπώντας στην προσέλκυση κυρίων προβλημάτων του δημοσίου τομέα, μέσω τεσσάρων επιμέρους φάσεων, από τις οποίες απελευθερώνεται η τεχνογνωσία άλλων μερών για τη δημιουργία βέλτιστων λύσεων. Αυτά τα εργαστήρια τεχνητής νοημοσύνης και επιστήμης των δεδομένων παρέχουν ένα δίκτυο κατάρτισης και ανάπτυξης δεξιοτήτων προς τους δημόσιους υπαλλήλους και προτείνουν επικοινωνιακές στρατηγικές και τεχνικές συμβουλές για την επίτευξη «έξυπνης» διαχείρισης⁵⁷⁷.

Τα απορρέοντα οφέλη από την αξιοποίηση μηχανών τεχνητής νοημοσύνης συνθέτουν ένα σύνολο γραμμών κατευθύνσεων προς τη σύναψη συνεργασιών μεταξύ δημοσίου και ιδιωτικού τομέα. Πρόκειται για οφέλη, άμεσα συνδεδεμένα με ελεγκτικά και προβλεπτικά ζητήματα που αφορούν τα επίπεδα της ζήτησης για συγκεκριμένες υπηρεσίες, καθώς επίσης και ζητήματα αυτόματης απόκρισης της ζήτησης στα υφιστάμενα επίπεδα προσφοράς. Επίσης, συνάδουν με τη διευκόλυνση της διαδικασίας προσδιορισμού ομάδων υψηλού κινδύνου, αλλά και με την ανάπτυξη συγκεκριμένων παρεμβατικών δράσεων που αποσκοπούν στην αύξηση της παραγωγικότητας, στον περιορισμό του κόστους και στην επίτευξη βέλτιστης απόδοσης, για την καλύτερη δυνατή κάλυψη των αναγκών των πολιτών - πελατών⁵⁷⁸.

Τα κοινωνικά οφέλη που προκύπτουν από τέτοιου είδους συνεργασίες είναι επικεντρωμένα στη βελτίωση της ποιότητας της παροχής υπηρεσιών από δημόσιους οργανισμούς, καθώς επίσης και στη συρρίκνωση των διοικητικών επιβαρύνσεων. Για παράδειγμα, από το Essex προβλέφθηκε ο κίνδυνος της υπόστασης ανεκπαιδευτων, άνεργων και μη καταρτισμένων ανθρώπων ηλικιακού εύρους μεταξύ 14-18 ετών και λόγω της πρόβλεψης αυτής της δυσμενούς κατάστασης, προσχώρησε στη σύναψη συνεργασίας με τις σχολικές μονάδες, σκοπεύοντας στην ανάπτυξη παρεμβάσεων παροχής επιπρόσθετης υποστήριξης και της συνακόλουθης ενθάρρυνσης της συμμετοχής στην εκπαιδευτική διαδικασία, των ανθρώπων που διατρέχουν υψηλό

⁵⁷⁷ Βλ. G.A. Hodge, & C. Greve, *όπ.παρ.*, σελ. 545-558.

⁵⁷⁸ Βλ. McKinsey Global Institute, *Artificial Intelligence: The next digital frontier*, Report, 2017, στο <https://www.mckinsey.com/~media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx>

κίνδυνο να υπαχθούν στην εν λόγω πληθυσμιακή κατηγορία των ανεκπαίδευστων, ανέργων και ακατάρτιστων⁵⁷⁹.

Ένα ακόμα παράδειγμα πηγάζει από το Ηνωμένο Βασίλειο και αφορά την περίπτωση του προγράμματος «Εκπαίδευση της κυριαρχίας» του οργανισμού Defense Advanced Research Projects, όπου χρησιμοποιείται η τεχνητή νοημοσύνη με σκοπό να μειωθεί ο χρόνος που απαιτείται για τους υπηρετούντες στο Ναυτικό, ώστε να καταστούν τεχνικοί εμπειρογνώμονες. Η μείωση του απαιτούμενου χρονικού διαστήματος είναι αξιοπρόσεκτη, καθώς ενώ παλαιότερα χρειαζόταν η έλευση ετών, πλέον απαιτούνται μόλις μερικοί μήνες. Το αποτέλεσμα αυτό επιτεύχθηκε μέσω της δημιουργίας ενός ψηφιακού «δασκάλου» που εφαρμόζει τη μηχανική μάθηση σε αρχάριους επί του εν λόγω τομέα, χρησιμοποιώντας εξειδικευμένη αλληλεπίδραση⁵⁸⁰.

Οι διατομεακές συνεργατικές προσπάθειες γύρω από την τεχνητή νοημοσύνη οργανώνονται θεσμικά, συνήθως, γύρω από γραφεία ανάλυσης δεδομένων (Offices for Data Analytics - ODAs). Τα γραφεία αυτά χρησιμοποιούνται, κυρίως, ως συνεργατικοί οργανωτικοί φορείς για την προαγωγή της κοινής χρήσης δεδομένων, προσβλέποντας στη βελτίωση των παρεχόμενων δημοσίων υπηρεσιών. Αυτό που διαφοροποιεί τα γραφεία ανάλυσης δεδομένων από τα εργαστήρια πολιτικής (policy laboratories) είναι ότι τα εργαστήρια πολιτικής λειτουργούν ως δικτυακές πλατφόρμες, που συνήθως προέρχονται από τα πανεπιστήμια, ενώ τα γραφεία ανάλυσης δεδομένων είναι «φυσικά» γραφεία, τα οποία συνδέονται με τους δημόσιους και δημοτικούς κυβερνητικούς φορείς. Ένα χαρακτηριστικό παράδειγμα είναι το γραφείο ανάλυσης δεδομένων της Νέας Υόρκης (MODA NYC). Πρόκειται για ένα φορέα που διαδραματίζει τον ρόλο του «ενοποιητή» της διαδικασίας συλλογής δεδομένων εκ του συνόλου των παρεχόμενων υπηρεσιών σε μία συγκεκριμένη πόλη. Η αξιοποίησή του

⁵⁷⁹ Βλ. Society TR, Machine learning: the power and promise of computers that learn by example, 2017, στο <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf>

⁵⁸⁰ Βλ. OFA Intelligence, Executive Office of the President National Science and Technology Council 2016 Preparing for the future of Artificial Intelligence, 2016, στο https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

έχει ως στόχο την προώθηση και διασφάλιση της δημόσιας τάξης και της επακόλουθης αποτελεσματικής αντιμετώπισης εγκληματικών πράξεων⁵⁸¹.

Ενώ η δράση των γραφείων ανάλυσης δεδομένων εμφανίζεται κυρίως σε ποικίλες και διαφορετικές κυβερνητικές υπηρεσίες, οι μη κυβερνητικοί οργανισμοί συμβάλλουν ομοίως στην ανταλλαγή γνώσεων και τη συλλογική και συνδυαστική ανάλυση αυτών. Για παράδειγμα, στην κατεύθυνση ανάπτυξης «*προτύπων και πρωτόκολλων για την κοινή χρήση δεδομένων, το γραφείο ανάλυσης δεδομένων της Νέας Υόρκης παρουσιάζει ουσιαστική συνεργασία με την τεχνογνωσία που διατίθεται από την πλευρά του NYU Center for Urban Science and Progress, καθώς και με το Columbia University Institute for Data Science and Engineering*»⁵⁸².

Μία τέτοιου είδους διατομεακή συνεργασία καθίσταται εξαιρετικά ελπιδοφόρα, αλλά ταυτόχρονα αποτελεί και μία μεγάλη πρόκληση. Αναλυτικότερα, ενώ τα οφέλη της αλληλεπίδρασης μεταξύ των επιμέρους τομέων για την εφαρμογή στρατηγικών τεχνητής νοημοσύνης είναι πολλαπλά, η διατομεακή εργασία και συνεργασία έχει αποδειχθεί εξαιρετικά περίπλοκη⁵⁸³. Με τις επικείμενες προκλήσεις και ευκαιρίες που προκύπτουν ασχολούνται τα δύο επόμενα υποκεφάλαια.

⁵⁸¹ Βλ. R. Yasin, How analytics is making NYC's streets and buildings safer, GCN Magazine, 2013, στο <https://gcn.com/articles/2013/10/04/gcn-award-nyc-databridge.aspx>

⁵⁸² Βλ. R. Yasin, *όπ.παρ.*, στο <https://gcn.com/articles/2013/10/04/gcn-award-nyc-databridge.aspx>

⁵⁸³ Βλ. W. Eggers, & J. Bellman, The journey to government's digital transformation, Deloitte, 2015, στο https://www2.deloitte.com/content/dam/insights/us/articles/digital-transformation-in-government/DUP_1081_Journey-to-govt-digital-future_MASTER.pdf

4.3.2 Οι Προκλήσεις για Επιτυχείς Διατομεακές Συνεργασίες

Ενώ οι προκλήσεις των συνεργασιών μεταξύ νομικών οντοτήτων του ιδιωτικού τομέα έχουν διερευνηθεί σε ικανοποιητικό βαθμό⁵⁸⁴, ωστόσο, δεν έχει διεξαχθεί αρκετή έρευνα γύρω από το πεδίο των μελανών σημείων των συνεργατικών σχημάτων που δραστηριοποιούνται στον δημόσιο τομέα. Αρκετοί ερευνητές, όπως είναι για παράδειγμα οι Stoker et al.⁵⁸⁵ εμβάθυναν στο ότι η ανάπτυξη του κοινωνικού κεφαλαίου, αποτελεί έναν αποτελεσματικό τρόπο, μέσω του οποίου μπορούν να αντιμετωπιστούν οι προκλήσεις που αναδύονται από τις συναπτόμενες συνεργασίες μεταξύ δύο ή περισσότερων δημοσίων οργανισμών. Οι Innes et al.⁵⁸⁶ στη μελέτη που διεξήγαγαν έδωσαν έμφαση σε θέματα άσκησης περιβαλλοντικής πολιτικής, τονίζοντας τα οφέλη που προκύπτουν από την ανάπτυξη ευέλικτων δικτυακών δομών, για λόγους υπερκέρρασης των εμποδίων που τίθενται από την πλευρά των γραφειοκρατικών και πεπαλαιωμένων θεσμικών περιορισμών.

Σε μία πιο πρόσφατη ερευνητική προσέγγιση, οι Andrews et al.⁵⁸⁷ προέβησαν στη μελέτη εμπειρικών έργων, σχετιζόμενων με τις διαφορές που προκύπτουν μεταξύ των ιδιωτικών και δημόσιων νομικών οντοτήτων. Οι ερευνητές επιχειρηματολόγησαν για τον τρόπο με τον οποίο οι διαφορές μεταξύ των οργανισμών που δραστηριοποιούνται σε αυτούς τους διαφορετικούς τομείς οδηγούν σε αμφισβήτηση των δια-οργανωτικών συνεργατικών σχημάτων. Η εν λόγω αμφισβήτηση, μάλιστα, στηρίζεται στην ανάμιξη του δημοσίου και ιδιωτικού περιβάλλοντος, καθώς πρόκειται για δύο τομείς που έχουν

⁵⁸⁴ Βλ. N. Riccucci, *Public administration: traditions of inquiry and philosophies of knowledge*, Washington, Georgetown University Press, 2010, σελ. 47-52.

⁵⁸⁵ Βλ. G. Stoker, G. Smith, & W. Maloney, *Building social capital in city politics: scope and limitations at interorganisational level*, *Political Studies*, Ap. 52, 2004, σελ. 508-530.

⁵⁸⁶ Βλ. J. Innes, L. Kaplan, S. Connick, & D. Booher, *Collaborative governance in the CALFED program: adaptive policy making for California water*, *Ecological Societies*, Ap. 15, 2006, σελ. 35, Βλ. επίσης J. Innes, S. Connick, & D. Booher, *Informality as a planning strategy: collaborative water management in the CALFED bay-delta program*, *Journal of American Planning Association*, Ap. 73, 2007, σελ. 195-210.

⁵⁸⁷ Βλ. R. Andrews, M. Esteve, & T. Ysa, *Public-private joint-ventures: mixing oil and water*, *Public Money Management*, Ap. 35, 2015, σελ. 265-272.

διαφορετικές δομές και αξίες, καταλήγοντας σε αναπόφευκτα διαφορετικούς στόχους⁵⁸⁸.

Οι αποκλίνουσες προσεγγίσεις επί ζητημάτων διαχείρισης κινδύνων αποτελούν μία ακόμη πτυχή προκλήσεων αναφορικά με τον συνδυασμό διαφορετικών περιβαλλόντων σε περιπτώσεις συνεργατικών συμπράξεων μεταξύ ιδιωτικού και δημοσίου τομέα. Επί του συγκεκριμένου ζητήματος, μάλιστα, η μελέτη που διεξήχθη από τους Klijn & Teisman⁵⁸⁹, καταλήγει να υποστηρίζει ότι οι πολιτικοί κίνδυνοι που αναλαμβάνουν οι κυβερνητικοί φορείς δεν μπορούν να ευθυγραμμιστούν επιτυχώς με τους τύπους κινδύνου που καλούνται να αντιμετωπίσουν οι δραστηριοποιούμενες στον ιδιωτικό τομέα νομικές οντότητες, όπως είναι για παράδειγμα ο κίνδυνος αγοράς.

Η ανταγωνιστική θεσμική λογική θεωρείται ως μία επιπρόσθετη θεμελιώδης πρόκληση, κατά την ανάπτυξη συνεργατικών εγχειρημάτων⁵⁹⁰. Πρακτικά, αυτό ασκεί αρνητική επιρροή στους τιθέμενους στόχους της συναπτόμενης συμφωνίας. Επί παραδείγματι, οι διαχειριστές των συνεργατικών επιχειρήσεων ενδέχεται να δυσκολεύονται να προσφέρουν δημόσια αξία θυσιάζοντας την επίτευξη υψηλότερων επιπέδων κερδοφορίας. Αντιστρόφως, ενδεχομένως να υφίσταται δυσκολία στη μεγιστοποίηση του κέρδους για λόγους ικανοποίησης των συμφερόντων του μετοχικού κοινού⁵⁹¹.

Ο συνδυασμός των διαφορετικών οργανωτικών δομών μπορεί, επίσης, να οδηγήσει σε ανεπιτυχή συνεργασία. Υπό αυτήν την έννοια, οι οργανισμοί του δημόσιου τομέα αναγνωρίζονται ως αμιγώς προσανατολισμένοι στην τήρηση των υφιστάμενων νόμων και κανονισμών, λόγω της ανάγκης ικανοποίησης των εκάστοτε υφιστάμενων νομικών

⁵⁸⁸ Βλ. P. Nutt, Comparing public and private sector decision-making practices, *Journal of Public Administration Research and Theory*, Ap. 16, 2006, σελ. 289–318, Βλ. επίσης G.A. Hodge, & C. Greve, *όπ.παρ.*, σελ. 545–558, Βλ. επίσης R. Andrews, M. Esteve, & T. Ysa, *όπ.παρ.*, σελ. 265–272.

⁵⁸⁹ Βλ. E. Klijn, & G. Teisman, Institutional and strategic barriers to public–private partnerships—an analysis of Dutch cases, *Public Money Management*, Ap. 23, 2003, σελ. 137–146.

⁵⁹⁰ Βλ. A. Saz-Carranza, & F. Longo, Managing competing institutional logics in public–private joint ventures, *Public Management Review*, Ap. 14, 2012, σελ. 331–357.

⁵⁹¹ Βλ. J. Shaoul, A. Shepherd, A. Stafford, & P. Stapleton, Losing control in joint ventures: the case of building schools for the future, Report, Institute of Chartered Accountants of Scotland, 2013, στο https://www.research.manchester.ac.uk/portal/files/32803045/FULL_TEXT.PDF

απαιτήσεων. Αντίθετα, ο Rainey⁵⁹² περιγράφει τον τρόπο με τον οποίο οι ιδιωτικές νομικές οντότητες δεν γίνονται αποδέκτες ομοίων πιέσεων σε ζητήματα πολιτικής λογοδοσίας και ως εκ τούτου, δεν δέχονται ασφυκτικές πιέσεις από γραφειοκρατικές διεργασίες.

Η επικρατούσα απόκλιση μεταξύ των οργανωτικών αξιών και του δομικού συστήματος των ιδιωτικών και των δημοσίων οργανισμών αποτελεί, συνήθως, τον κύριο λόγο αποτυχίας των συναφθεισών συνεργασιών⁵⁹³. Η σημασία αυτής της διαφοροποίησης είναι ιδιαίτερος σημαντική σε περιπτώσεις διατομεακών συνεργασιών, τη στιγμή κατά την οποία έκαστος τομέας περιλαμβάνει ένα σύνολο διαφορετικών προσδιοριστικών παραγόντων και μεταβλητών από τις οποίες εξαρτάται η αποτελεσματικότητα και αποδοτικότητα του. Πιο συγκεκριμένα, Το ανθρώπινο δυναμικό που στελεχώνει τους δημόσιους οργανισμούς έχει μεγαλύτερα κίνητρα υπέρ της εξυπηρέτησης των πολιτών και της κάλυψης των αναγκών τους, κάτι που δεν συμβαίνει στην περίπτωση των εργαζομένων στον ιδιωτικό τομέα, οι οποίοι έχουν κίνητρα να διασφαλίσουν το συμφέρον της επιχειρηματικής μονάδας για την οποία εργάζονται⁵⁹⁴. Βάσει αυτού του συλλογισμού, η μεγαλύτερη πρόκληση που προκύπτει από τη δημιουργία συνεργατικών σχημάτων μεταξύ ιδιωτικού και δημοσίου τομέα είναι η απόδοση των κατάλληλων κινήτρων, τόσο στους δημοσίους όσο και στους ιδιωτικούς υπαλλήλους, έτσι ώστε αμφότεροι να καταλήξουν να αλλάξουν τις αρχές και τις αξίες τους, κατευθυνόμενοι από την ατομικιστική προς μία συλλογική προσέγγιση.

Ο τομέας της τεχνητής νοημοσύνης παρέχει αξιοπρόσεκτες επενδύσεις σε δεξιότητες και στον ανθρώπινο παράγοντα, γεγονός που έχει άμεση θετική επίπτωση στην ευρύτερη οικονομία ενός κράτους. Επίσης, υφίσταται μείωση του χάσματος που υπάρχει μεταξύ των κυβερνητικών πληροφοριών και των πληροφοριών που κατευθύνονται προς αξιοποίηση από τον δημόσιο τομέα ενός κράτους και οι οποίες

⁵⁹² Βλ. H.G. Rainey, Public management—recent research on the political context and managerial roles, structures and behaviors, *Journal of Management*, Ap. 15, 1989, σελ. 229–250.

⁵⁹³ Βλ. M. Huang, K. Kao, T. Lu, & W Pu, Control mechanisms as enhancers of international joint venture performance, Paper presented at the Academy of Management Annual Meeting, Chicago, 2009, σελ. 14.

⁵⁹⁴ Βλ. F.K. Sonnenberg, Partnering: entering the age of co-operation, *Journal of Business Strategy*, Ap. 13, 1992, σελ. 49–52, Βλ. επίσης J.L. Perry, & L.R. Wise, The motivational bases of public service, *Public Administration Reviews*, Ap. 50, 1990, σελ. 367–373.

πηγάζουν από την τεχνητή νοημοσύνη. Επιπλέον, οι Chen et al.⁵⁹⁵ επισήμαναν την έλλειψη τεχνικής ορολογίας, που απαιτείται για την εξατομίκευση των διαφορετικών δημόσιων στοιχείων και πληροφοριών που προκύπτουν από διαφορετικούς δημόσιους οργανισμούς, ώστε αυτοί να αξιοποιηθούν συλλογικά. Οι μελετητές τόνισαν τη σπουδαιότητα της ικανότητας του εκάστοτε χρήστη των συστημάτων τεχνητής νοημοσύνης, στον χειρισμό νέων συστημάτων. Ομοίως σημαντική προκύπτει και η ενθάρρυνση της διαδικασίας ανταλλαγής δεδομένων, όχι μόνο σε ομαδικό, αλλά και σε εξατομικευμένο πλαίσιο ⁵⁹⁶.

Στη μελέτη που διεξήχθη από τους Weber et al.⁵⁹⁷ αποδείχτηκε ότι η μεταφορά, η λήψη και η επεξεργασία των πληροφοριών μεταξύ των εμπλεκόμενων μερών αποτελεί μία διαδικασία που συνθέτει μία αέναη πρόκληση για το σύνολο των προβλημάτων που αφορούν τον δημόσιο τομέα και που εμπίπτουν στο πλαίσιο των δια-οργανωτικών ρυθμίσεων. Ως εκ τούτου, τα διοικητικά και διευθυντικά στελέχη πρέπει να επιδιώκουν τη στρατηγική ενθάρρυνση των υφισταμένων τους, ώστε να καταλήγουν στον διαμοιρασμό νέων δεδομένων και στην περαιτέρω βελτίωση των δεξιοτήτων τους. Η αναβάθμιση των δεξιοτήτων, συμπεριλαμβανομένης της επικοινωνίας με τους συναδέλφους τους, αποτελεί κομβικό σημείο στην επίτευξη συλλογικής βελτίωσης της κατάκτησης γνώσης⁵⁹⁸.

Το πεδίο της τεχνητής νοημοσύνης μπορεί να αντιμετωπίσει το ανωτέρω αναφερόμενο ζήτημα δημιουργώντας τα λεγόμενα «Data Trusts». Η χρησιμοποίησή τους εξασφαλίζει την παροχή σαφών πλαισίων για την ασφάλεια της χρήσης και διαχείρισης πληροφοριών⁵⁹⁹. Πρόκειται για μία μορφή κεντρικού συστήματος δεδομένων, που προσφέρει τεχνική ικανότητα για την ανταλλαγή και επεξεργασία πληροφοριών μεταξύ

⁵⁹⁵ Βλ. Y. Chen, & J. Lee, Collaborative data networks for public service: governance, management, and performance, *Public Management Reviews*, Ap. 20, 2018, σελ. 672–690.

⁵⁹⁶ Βλ. Y. Chen, & J. Lee, *όπ.παρ.*, σελ. 672–690.

⁵⁹⁷ Βλ. E. Weber, & A. Khademian, Wicked problems, knowledge challenges, and collaborative capacity builders in network settings, *Public Administration Reviews*, Ap. 68, 2008, σελ. 334–349.

⁵⁹⁸ Βλ. A. Cabrera, & E. Cabrera, Knowledge-sharing dilemmas, *Organizational Studies*, Ap. 23, 2002, σελ. 687–710.

⁵⁹⁹ Βλ. HM Government, 2017 Industrial strategy: building a Britain fit for the future, *Industrial Strategy Artificial Intelligence Sector Deal, White Paper*, 2017, σελ. 30.

δια-οργανωτικών νορμών, υπό την παρουσία και τήρηση συγκεκριμένων προτύπων για τα διαρρέοντα δεδομένα⁶⁰⁰. Μάλιστα, πέραν της υιοθέτησης και τήρησης συγκεκριμένων προτύπων κοινής χρήσης πληροφοριών, έχει προταθεί και η δημιουργία προτύπων για τις διαδικασίες της συγκέντρωσης της πληροφορίας και της εγγύησης της υψηλής της ποιότητας. Με τον τρόπο αυτόν, υπό τη χρήση των εν λόγω προτύπων, προάγεται η τεχνική ικανότητα της σύμπραξης και συνεργασίας⁶⁰¹.

Πέραν από τις ανησυχίες της ηθικής χρήσης της τεχνητής νοημοσύνης και της δικαιοσύνης στην ανταλλαγή και την εφαρμογή της, οι ανεπιθύμητες απόρροιες είναι, επιπροσθέτως, άμεσα συνυφασμένες με τις επακόλουθες αποφάσεις που λαμβάνονται σχετικά με τον ανθρώπινο παράγοντα και η λήψη των οποίων γίνεται από τους ίδιους τους ανθρώπους. Πρόκειται για ένα σύνολο αποφάσεων, οι οποίες αναμένεται να αντικατασταθούν από την τεχνητή νοημοσύνη⁶⁰². Συγκεκριμένα, εγείρονται αρκετές ανησυχίες αναφορικά με τον τρόπο με τον οποίο μπορεί να προασπιστεί η δικαιοσύνη, η λογοδοσία και η ισοτιμία στη διαδικασία λήψης αποφάσεων και στο πώς το πεδίο της μηχανικής μάθησης μπορεί να προκαλέσει αντιδράσεις στην πολυπλοκότητα που χαρακτηρίζει το ανθρώπινο περιβάλλον. Ειδικά όσον αφορά την ποινική δικαιοσύνη, πρέπει να ασκούνται διαρκώς πιέσεις από οργανισμούς όπως το «Κέντρο Δεοντολογίας Δεδομένων και Καινοτομίας», επί ζητημάτων ενσωμάτωσης αμερόληπτων και ποιοτικών δεδομένων. Σε διαφορετική περίπτωση, υπάρχει ο κίνδυνος της αύξησης των προβλημάτων που δημιουργούνται από προκαταλήψεις επί των τεχνολογικών διεπαφών και, μεταξύ άλλων, ελλοχεύει ο κίνδυνος της παρουσίας έντονων διακρίσεων. Εντούτοις, η ανάλυση δεδομένων, μέσω της τεχνητής νοημοσύνης, είναι δυνατόν να χρησιμοποιηθεί και για λόγους πρόβλεψης και ανίχνευσης κρουσμάτων προκατάληψης, με σκοπό την πρόληψη των εν λόγω αναφερόμενων διακρίσεων⁶⁰³.

⁶⁰⁰ Βλ. Y. Chen, & J. Lee, *όπ.παρ.*, σελ. 672-690.

⁶⁰¹ Βλ. Y. Chen, & J. Lee, *όπ.παρ.*, σελ. 672-690.

⁶⁰² Βλ. J. Holdren, & M. Smith, *όπ.παρ.*, στο https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

⁶⁰³ Βλ. C. Munoz, & M. Smith, Big risks, big opportunities: the intersection of big data and civil rights, Obama White House Archives, 2016, στο <https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights>

Επίσης, επισημαίνεται ότι η ίδρυση του λεγόμενου «Κέντρου Δεοντολογίας Δεδομένων και Καινοτομίας» οδηγεί ένα βήμα πιο κοντά στην ηθική αξιοποίηση των τεχνικών αντικειμένων τεχνητής νοημοσύνης. Παράλληλα, ο «Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR)» και το αντίστοιχο νομοσχέδιο προστασίας δεδομένων που ισχύει στην Αγγλία, κατοχυρώνουν την απαιτούμενη ασφάλεια δικαίου ως προς την αξιοποίηση των πληροφοριών και την κοινή χρήση τους⁶⁰⁴.

⁶⁰⁴ Βλ. HM Government, 2017 Industrial strategy: building a Britain fit for the future, Industrial Strategy Artificial Intelligence Sector Deal, White Paper, 2017, σελ. 30.

4.3.3 Οι Προσδιοριστικοί Παράγοντες της Επιτυχίας και οι Ευκαιρίες των Επιτυχών Διατομεακών Συνεργασιών

Επί του ζητήματος της απόδοσης των διατομεακών συνεργασιών έχουν καταγραφεί εκτεταμένες έρευνες⁶⁰⁵. Στο πλαίσιο της υφιστάμενης βιβλιογραφίας είναι δυνατή η διάκριση επτά βασικών στρατηγικών ηγεσίας, ήτοι η διευκολυντική ηγεσία, η ικανοποίηση των κοινών στόχων μέσω της συλλογής και ανταλλαγής γνώσεων και τέλος, αυτή της επικοινωνίας και κοινωνικοποίησης. Οι εν λόγω στρατηγικές αναλύονται ως κάτωθι:

(1) Διευκολυντική ηγεσία: Εν αντιθέσει με την κλασική ιδέα της υπόστασης ενός ιεραρχικά ανώτερου ηγέτη, ο οποίος έχει την τάση να επιβάλλει τις απόψεις του στους οπαδούς-υφισταμένους του, στηριζόμενος στη θέση εξουσίας που κατέχει εντός του οργανισμού, η συγκεκριμένη ηγετική στρατηγική υποστηρίζει τον σεβασμό και την ανάπτυξη αποδοτικών σχέσεων μεταξύ των μελών μίας ομάδας, καθώς επίσης και την εποικοδομητική διαχείριση συγκρούσεων, αλλά και την ειλικρινή έκφραση στάσεων και πεποιθήσεων⁶⁰⁶. Η μετα-ανάλυση των Ansell & Gash⁶⁰⁷, για τη διαχείριση της συνεργατικής διακυβέρνησης, καταλήγει στο συμπέρασμα ότι οι ηγέτες των συνεργασιών θα πρέπει να προωθούν την ευρεία και ενεργό συμμετοχή, καθώς επίσης και να εξασφαλίζουν την άσκηση ευρείας επιρροής και τη διενέργεια ευρύτερων ελέγχων.

Επομένως, η διευκολυντική ηγεσία κρίνεται ουσιαστικά αποδεκτή στην εφαρμογή της για λόγους προώθησης της βέλτιστης δυνατής συνεργασίας μεταξύ των μελών των

⁶⁰⁵ Βλ. M. McGuire, Collaborative public management: assessing what we know and how we know it, *Public Administration Reviews*, Αρ. 66, 2006, σελ. 33–43, Βλ. επίσης P. Kenis, & K. Provan, Towards an exogenous theory of public network performance, *Public Administration*, Αρ. 87, 2009, σελ. 440–456, Βλ. επίσης M.P. Mandell, Getting results through collaboration: networks and network structures for public policy and management, Westport, Quorum Books, 2001, σελ. 26–31, Βλ. επίσης R. Agranoff, & M. McGuire, Multinetwork management: collaboration and the hollow state in local economic policy, *Journal of Public Administration Research and Theory*, Αρ. 8, 1998, 67–91.

⁶⁰⁶ Βλ. D. Day, J. Fleenor, L. Atwater, R. Sturm, & R. McKee, Advances in leader and leadership development: a review of 25 years of research and theory, *Leadership Quart*, Αρ. 25, 2014, 63–82.

⁶⁰⁷ Βλ. C. Ansell, & A. Gash, Collaborative governance in theory and practice, *Journal of Public Administration Research and Theory*, Αρ. 18, 2007, σελ. 543–571.

ομάδων και για την απόδοση μέγιστων κινήτρων ενεργούς συμμετοχής τους στην επιτέλεση των υποχρεώσεων και την κάλυψη των συλλογικών και ατομικών αρμοδιοτήτων⁶⁰⁸. Ωστόσο, σημειώνεται ότι το διευκολυντικό στυλ ηγεσίας δεν αποτελεί το μοναδικό βέλτιστο υιοθετούμενο στυλ ηγεσίας. Αυτό αιτιολογείται και από την ανάλυση των Hovik et al.⁶⁰⁹, οι οποίοι υποστήριξαν ότι το στυλ ηγεσίας που εφαρμόζεται ανά περιπτώσεις, είναι δυνατόν να είναι απόλυτα εξαρτημένο από τα χαρακτηριστικά στοιχεία που διέπουν μία σχέση επαγγελματικής συνεργασίας. Μάλιστα, τα ιδιαίτερα γνωρίσματα της προσωπικότητας και του χαρακτήρα ενός διευθυντικού στελέχους, αποτελούν βασικούς προσδιοριστικούς παράγοντες της ικανότητάς του να κατανοήσει πώς μπορούν να αξιοποιηθούν οι διαφορετικές δεξιότητες από τις οποίες χαρακτηρίζονται οι υφιστάμενοί του. Οι ίδιοι, επιπροσθέτως, προέβησαν στον εντοπισμό τεσσάρων βασικών ρόλων ηγετών που αποσκοπούν στην προαγωγή συνεργατικού κλίματος στην εργασία. Πιο συγκεκριμένα, πρόκειται για αυτόν του συντονιστή που διασφαλίζει τη ροή πληροφοριών, την αποσαφήνιση ρόλων και τη συμμόρφωση, αυτόν του καταλύτη που δημιουργεί κίνητρα, ευαισθητοποιεί και διασφαλίζει την ιδιοκτησία, αυτόν του διαμεσολαβητή που διαδραματίζει τον ρόλο του μεσίτη και διευκολύνει τη συζήτηση και τον υγιή διάλογο και τέλος, αυτόν του συνδετικού κρίκου, μέσα από τον οποίον επέρχεται γεφύρωση των χασμάτων των συνεργαζόμενων οργανισμών.

(2) Κοινοί στόχοι: Ο ορισμός των στόχων του οργανισμού συσχετίστηκε θετικά με την οργανωσιακή απόδοση των οργανισμών που λειτουργούν στο πλαίσιο του δημόσιου τομέα⁶¹⁰. Ωστόσο, δεδομένου ότι οι συνεργασίες συνεπάγονται τη δημιουργία κοινών αξιών και την απαλλοτρίωση των παλαιών ισχυουσών, η υπόσταση μικρών επιμέρους διαφοροποιημένων κινήτρων, μεταξύ των συνεργαζόμενων οργανισμών, είναι δυνατόν να καταλήξει στη δημιουργία εντάσεων, οι οποίες με τη σειρά τους μπορούν να κλονίσουν τους κοινούς στόχους και να τους μετατρέψουν σε ατομικούς⁶¹¹. Η

⁶⁰⁸ Βλ. C. Ansell, & A. Gash, *όπ.παρ.*, σελ. 543-571.

⁶⁰⁹ Βλ. S. Hovik, & G. Hanssen, The impact of network management and complexity on multi-level coordination, *Public Administration*, Αρ. 93, 2015, σελ. 506-523.

⁶¹⁰ Βλ. G.A. Boyne, & A. Chen, Performance targets and public service improvement, *Journal of Public Administration Research and Theory*, Αρ. 17, 2007, σελ. 455-477.

⁶¹¹ Βλ. A. Madhok, Opportunism and trust in joint venture relationships: an exploratory study and a model, *Scandinavian Journal of Management*, Αρ. 11, 1995, σελ. 57-74.

παράλληλη επιδίωξη διαφορετικών στόχων οδηγεί τα διοικητικά στελέχη στη συνέχιση της επιτέλεσης των εργασιών τους με τον συνήθη τρόπο, λόγω του ότι δεν γνωρίζουν σε ποιους στόχους πρέπει ακριβώς να επικεντρωθούν⁶¹². Αυτό έρχεται σε ευθυγράμμιση με τον ισχυρισμό του Jensen⁶¹³, ο οποίος αναφέρει ότι το να ζητείται από τα διοικητικά στελέχη των συνεργασιών να επιδιώκουν πολλαπλούς στόχους δημιουργεί προβλήματα. Ο ίδιος, δεν αναφέρεται στην πρόκληση σύγχυσης και αποπροσανατολισμού από τον βασικό στόχο, αλλά αναφέρεται περισσότερο στη δημιουργία της διατήρησης μίας προκαταρκτική στάσης από την πλευρά των διοικητικών στελεχών⁶¹⁴.

Ακόμη και αν το σύνολο των εμπλεκόμενων σε μία συνεργασία μερών, έρχεται σε αρχική ευθυγράμμιση με τον βασικό τιθέμενο στόχο, ωστόσο, δεν αποκλείεται η παρουσία σημείων στα οποία εντοπίζονται διαφοροποιήσεις στους στόχους που έχει θέσει το κάθε μέρος χωριστά. Η σημασία των στόχων που τίθενται προς κάλυψη από τα συνεργατικά μέρη έγκειται στο ότι διαδραματίζουν το ρόλο του κινητήριου μοχλού της διαδικασίας λήψης αποφάσεων. Επίσης, ταυτόχρονα, διαδραματίζουν τον ρόλο των προτύπων αναφοράς, βάσει των οποίων θα αξιολογηθεί αν και κατά πόσο επετεύχθη ο τελικός στόχος⁶¹⁵.

(3) Συγκέντρωση και ανταλλαγή γνώσεων: Για να υπερκεραστούν τα διαδικαστικά ζητήματα, καθώς επίσης και τα ζητήματα που απορρέουν από τη δυναμική της συνεργατικής διακυβέρνησης, προτάθηκε από τους Chen & Lee⁶¹⁶ ότι οι διοικητικές και διαχειριστικές φύσεως δραστηριότητες των μερών που εμπλέκονται σε μία συνεργασία, είναι απαραίτητο να εμβαθύνουν στην ανάπτυξη θεσμικών ικανοτήτων που να προάγουν την κοινή και συλλογική δράση, όπως είναι λόγου χάρη η ανάπτυξη κοινών προτύπων για τη συλλογή και επεξεργασία των δεδομένων. Σε τεχνικό επίπεδο,

⁶¹² Βλ. S.K. Ethiraj, & D. Levinthal, Hoping for A to Z while rewarding only A: complex organizations and multiple goals, *Organizational Science*, Ap. 20, 2009, σελ. 4–21.

⁶¹³ Βλ. M.C. Jensen, Value maximization, stakeholder theory, and the corporate objective function, *Journal of Application of Corporate Finance*, Ap. 14, 2001, σελ. 8–21.

⁶¹⁴ Βλ. S.K. Ethiraj, & D. Levinthal, όπ.πρ. σελ. 4-21.

⁶¹⁵ Βλ. D. Farnham, & S. Horton, *Managing the new public services*. London, Macmillan, 1993, σελ. 34-42.

⁶¹⁶ Βλ. Y. Chen, & J. Lee, όπ.παρ., σελ. 672-690.

η εμπειριστατωμένη μελέτη περίπτωσης, που διεξήχθη από την πλευρά των εν λόγω μελετητών, οδηγεί στη διαπίστωση ότι οι ομοσπονδιακοί οργανισμοί καταλήγουν να αντιμετωπίζουν προκλήσεις από τη διαχείριση των συνεργατικών δικτύων δεδομένων, τα οποία καθίστανται απολύτως αναγκαία για την προαγωγή και προώθηση της διαδικασίας της ανταλλαγής δεδομένων σε όλες τις δικαιοδοσίες, κάτι που απαιτείται για την παρουσία ενός ολοκληρωμένου και κοινού σχεδιασμού μεταφοράς δεδομένων⁶¹⁷.

Διευκρινίζεται ότι η ανάπτυξη κοινών προτύπων συγκέντρωσης και επεξεργασίας δεδομένων επιφέρει πιο θετικά αποτελέσματα μέσω της παρουσίας σταθερών και βασικών συνεργατών, έτσι ώστε οι ομάδες εργασίας να μπορούν να είναι ευθυγραμμισμένες με τις επιμέρους λειτουργικές ευθύνες που διέπουν το σύνολο του συνεργατικού πλαισίου⁶¹⁸.

Παραδείγματος χάριν, οι εργαζόμενοι της πρώτης γραμμής θα μπορούσαν να αποκαλούνται και ως «τράπεζες γνώσης», μέσω των οποίων είναι δυνατή υποστήριξη και η πρόταση σχεδιασμών διαδικαστικής τυποποίησης. Οι συγγραφείς⁶¹⁹ κατέληξαν στη διαπίστωση ότι τα μέλη του GIS διατηρούν τακτική επαφή και επικοινωνία με περιφερειακές δραστηριότητες σχεδιασμού μεταφορών των δεδομένων και μπορούν να προσφέρουν εξειδικευμένες πληροφορίες για τη λειτουργία των δεδομένων και της τεχνολογίας που βρίσκεται πίσω από αυτά.

(4) Επικοινωνία: Στη μελέτη των Ansell & Gash⁶²⁰ περιγράφηκαν τα αποτελέσματα από περιπτώσεις όπου ακολουθήθηκε μία συγκεκριμένη στρατηγική επικοινωνίας, στο πλαίσιο μίας συνεργασίας. Οι μελετητές αναφέρουν ότι αν τα συνεργαζόμενα μέρη έχουν την προθυμία ανεύρεσης και αξιοποίησης των απαιτούμενων πόρων και διάθεσης του αναγκαίου χρόνου, τότε ομιλούμε περί επιτυχημένης συνεργασίας που

⁶¹⁷ Βλ. Y. Chen, & J. Lee, όπ.παρ., σελ. 672-690.

⁶¹⁸ Βλ. Y. Chen, & J. Lee, όπ.παρ., σελ. 672-690.

⁶¹⁹ Βλ. Y. Chen, & J. Lee, όπ.παρ., σελ. 672-690.

⁶²⁰ Βλ. C. Ansell, & A. Gash, Collaborative platforms as a governance strategy, Journal of Public Administration Research and Theory, Apr. 28, 2017, σελ. 16–32.

καταλήγει σε απτά αποτελέσματα. Όμως, αυτό δεν δύναται να επιτευχθεί αν δεν υπάρχουν κοινές αξίες που χτίζονται στο πλαίσιο μίας κοινής δράσης⁶²¹.

Επίσης, θετικές αλληλεπιδράσεις μεταξύ των συνεργαζόμενων μερών πηγάζουν από τη διαδικασία απόκτησης γνώσης μέσω της μάθησης, η οποία επιτυγχάνεται όταν η συνεργασία συνάπτεται και εφαρμόζεται σε χώρους και πλαίσια που χαρακτηρίζονται από πολιτική ουδετερότητα⁶²². Επιπλέον, οι συνεργασίες μπορεί να στεφτούν από επιτυχία σε περιπτώσεις κατά τις οποίες είναι δυνατή η άσκηση αρχιτεκτονικής μόχλευσης. Αυτή δύναται να επιτευχθεί μέσω της ανάπτυξης κοινών στοιχείων, σχεδίων και προτύπων, τα οποία με τη σειρά τους μπορούν να αναδιαμορφωθούν, παίρνοντας ως άμεση απόρροια πολλαπλασιαστικά αποτελέσματα που βασίζονται σε αυτές τις προϋπάρχουσες προσπάθειες⁶²³.

(5) Κοινωνικοποίηση: Σε περίπτωση κατά την οποία οι διαχειριστές-διευθυντές των συνεργατικών σχημάτων καθιστούν διαφανή και δελεαστικό τον αντίκτυπο των προσπαθειών της συνεργασίας στα βασικά εμπλεκόμενα μέρη του συνεργατικού σχήματος, ακολούθως, ασκείται θετική επιρροή στη συνεργασία⁶²⁴. Αυτό έρχεται σε ευθυγράμμιση με την άποψη ότι, όταν οι εργαζόμενοι αντιλαμβάνονται τα θετικά αποτελέσματα των ενεργειών τους σε ένα άτομο, επόμενο είναι να οδηγούνται σε αύξηση της εργασιακής τους απόδοσης και των καταβαλλόμενων προσπαθειών τους. Τα διαφανή αποτελέσματα, τα οποία προκύπτουν μέσα από ειδικούς δείκτες μέτρησης που προσδιορίζονται στο πλαίσιο του εκάστοτε συνεργατικού σχήματος, μπορούν να διευκολύνουν την ανάπτυξη περισσότερων ιδεών στο σύνολο των πλαισίων συνεργασίας, όπου μπορεί να είναι δυσχερής η εφαρμογή μίας συγκεκριμένης ιδέας, που φέρει κατεύθυνση από «πάνω προς τα κάτω». Αυτό συμβαίνει συνήθως στις αποκεντρωμένες οργανωτικές δομές. Επί παραδείγματι, ο τομέας υγείας των ΗΠΑ βρίσκεται σε πειραματικό στάδιο με την εν λόγω ολιστική προσέγγιση, που αφορά τη

⁶²¹ Βλ. C. Ansell, & A. Gash, *όπ.παρ.*, σελ.16-32.

⁶²² Βλ. A. Saz-Carranza, & S. Ospina, The behavioural dimension of governing inter-organizational goal-directed networks: managing the unity–diversity tension, *Journal of Public Administration Research and Theory*, Αρ. 21, 2011, σελ. 327–365.

⁶²³ Βλ. A. Saz-Carranza, & S. Ospina, *όπ.παρ.*, σελ. 327-365.

⁶²⁴ Βλ. S. Cuganesan, A. Hart, & C. Steele, Managing information sharing and stewardship for public-sector collaboration: a management control approach, *Public Management Review*, Αρ. 19, 2017, σελ. 862–879.

σφαιρική κατανόηση του ευρύτερου συστήματος, από το οποίο περιβάλλονται τα προβληματικά σημεία που συνοδεύουν τα συστήματα αποζημιώσεων από τη χρήση πληροφοριών που προκύπτουν από μεγάλα συνεργατικά σχήματα. Αυτό, είναι υλοποιήσιμο αν προηγηθεί συνδυαστική ανάλυση και επεξεργασία δεδομένων υγείας, από πολλά διαφορετικά μέρη, όπως είναι οι επαγγελματίες υγείας, οι ίδιοι οι ασθενείς, καθώς επίσης και η κοινότητα των ερευνητών⁶²⁵.

Αντιθέτως, ένα διασυνδεδεμένο «σύστημα μάθησης» οδηγεί σε ταχύτερη μετάδοση πληροφοριών, γεγονός που καταλήγει σε γρηγορότερη μεταφορά γνώσης σε πολλά και διαφορετικά επίπεδα. Η Page⁶²⁶ ανέφερε ότι διαδικασίες που είναι απαλλαγμένες αποκλεισμών και αξιοποιούνται για την ανάπτυξη δεικτών και αξιολογήσεων, διαδραματίζουν υποστηρικτικό ρόλο στις αμοτεροβαρείς υποψίες των μερών που εμπλέκονται σε ένα συνεργατικό σχήμα. Πρόκειται για υποψίες που όλα τα μέρη μπορεί να έχουν κατά την έναρξη της διαδικασίας της μεταρρύθμισης και είναι ένα ζήτημα που χρήζει περαιτέρω επεξήγησης, επειδή οι στόχοι, η αποστολή, το όραμα και οι καθημερινές δράσεις της συνεργασίας, δύνανται να ξεπεράσουν τόσο τα εμπλεκόμενα μέρη όσο και τους ίδιους τους συνεργαζόμενους οργανισμούς⁶²⁷.

Στη μελέτη των Thomson & Perry⁶²⁸ διατυπώνονται οι ανωτέρω αναφερόμενες υποψίες, τονίζοντας το ρόλο που διαδραματίζεται από την αμοιβαία ευθύνη. Καθίσταται εναργές ότι η γενική συναίνεση που είναι απαραίτητη για τη συνεργατική διαχείριση δεν είναι εφικτό να επιτευχθεί άνευ ισορροπίας.

(6) Εμπειρογνωμοσύνη: Γενικότερα, η λογική από την οποία διακατέχονται τα άτομα που καλούνται να διαχειριστούν την συναπτόμενη συνεργασία είναι η βάση για την πραγματοποίηση επιλογών αναφορικά με το «πώς» και «πότε» θα αξιοποιηθούν οι επιμέρους στρατηγικές λήψης, μεταβίβασης και επεξεργασίας δεδομένων που

⁶²⁵ Βλ. J. Podesta, J. Holdren, J. Zients, E. Moniz, & P. Pritzker, Big Data: seizing opportunities, preserving values, Report, Obama White House Archives, 2014, σελ. 22.

⁶²⁶ Βλ. S. Page, Entrepreneurial strategies for managing interagency collaboration, Journal of Public Administration Research and Theory, Ap. 13, 2003, σελ. 311–339.

⁶²⁷ Βλ. S. Page, όπ.παρ., σελ. 311-339.

⁶²⁸ Βλ. A. Thomson, & J. Perry, Collaboration processes: inside the black box, Public Administration Reviews, Ap.66, 2006, σελ. 20–32.

προέρχονται από το διαδίκτυο⁶²⁹. Τα επίπεδα εμπιστοσύνης των μερών που συμμετέχουν σε μία συνεργασία δύνανται να αυξηθούν μέσω της πρόσληψης τεχνικών για τη διαχείριση των δικτύων και της διακοπής της ενασχόλησης εμπειρογνομόνων από το εν λόγω πεδίο. Συνακόλουθα, βελτιώνεται η ποιότητα της συγκεκριμένης παρεχόμενης υπηρεσίας⁶³⁰.

Παρατηρώντας καλές πρακτικές επί συνεργασιών οργανισμών που δραστηριοποιούνται στον δημόσιο τομέα, από την «NYU Govlab» επισημάνθηκε ότι οι αρμόδιοι επί διαχειριστικών ζητημάτων στις εν λόγω συνεργασίες πρέπει απαραίτητα να δύνανται να προβούν στην αποτελεσματική αντιμετώπιση πιθανών θεσμικών ανεπαρειών. Η αντιμετώπιση αυτή είναι εφικτή εντείνοντας τον βαθμό συμμετοχής των ατόμων που διαχειρίζονταν παλαιές πληροφορίες και τα οποία απασχολούνται σε κυβερνητικούς φορείς, όπως είναι ο οργανισμός GIS⁶³¹.

Στη μελέτη του Copeland⁶³² επισημαίνεται ότι η ηγεσία που έχει κατεύθυνση από «πάνω προς τα κάτω», όταν εφαρμόζεται σε δημόσιους οργανισμούς, είναι επικίνδυνη για την απομάκρυνση ταλέντων στην ανάλυση δεδομένων. Οπότε, οι ίδιοι προτείνουν ότι η απασχόληση εμπειρογνομόνων σε τέτοιες θέσεις εργασίας, ενδείκνυται για την αποφυγή του κινδύνου αυτού στις συνεργασίες μεταξύ δημοσίων οργανισμών.

Παραδείγματος χάριν, στη συνεργασία που είχε συνάψει το τμήμα περιπολιών του δήμου του Ρότερνταμ “Traffic”, με το Πανεπιστήμιο του Σικάγο⁶³³, ξεκαθάρισε το τοπίο για το «πώς» μπορεί να λειτουργήσει αποδοτικά το σχήμα συνεργασίας, στο οποίο συμμετείχαν οι επιθεωρητές των περιπολιών που ήταν υπεύθυνοι για τους αυτοκινητόδρομους και οι επιθεωρητές που ήταν υπεύθυνοι για τη ρύθμιση της

⁶²⁹ Βλ. E. Weber, & A. Khademian, όπ.παρ., σελ. 334-349.

⁶³⁰ Βλ. S. Giest, Network capacity-building in high-tech sectors: opening the black box of cluster facilitation policy, Public Administration, Ap. 93, 2015, σελ. 471-489.

⁶³¹ Βλ. B. Blauer, Barriers to building a data practice in government, 2017, στο <https://sppd.thegovlab.org/lectures/barriers-to-building-a-data-practice-in-government.html>

⁶³² Βλ. E. Copeland, Lessons from piloting the London Office of Data Analytics, 2017, στο <https://www.oreilly.com/radar/lessons-from-piloting-the-london-office-of-data-analytics/>

⁶³³ Βλ. H. Chua, V. Vicente, & S. Mintzer, Looking through the lens of Rijkswaterstaat traffic inspectors and the municipality of Rotterdam, 2017, στο <https://www.dssgfellowship.org/2017/08/09/looking-through-the-lens-of-rijkswaterstaat-traffic-inspectors-and-the-municipality-of-rotterdam/>

κυκλοφορίας. Η τήρηση της νομιμότητας ανέκαθεν είναι η βάση της επιτυχίας του συγκεκριμένου συνεργατικού εγχειρήματος. Οι αρμόδιοι για τη διαχείριση της συνεργασίας αυτής απέκτησαν σημαντικό όφελος από το γεγονός της εξασφάλισης ότι η «δύναμη» των εμπειρογνομώνων είχε αξιοποιηθεί με χρηστό και αποδοτικό τρόπο. Συνεπώς, υπήρχε αξιοπιστία για την ποιότητα των δεδομένων που αφορούσαν τη λήψη, τη μεταφορά και την επεξεργασία πληροφοριών, σε ένα κοινό δίκτυο πληροφόρησης που χρησιμοποιούσαν τα μέρη που συμμετείχαν στη συνεργασία⁶³⁴.

(7) Sense-making: Στη μελέτη της Heen⁶³⁵ εξετάστηκε το πόσο εφικτό είναι να στεφτεί με επιτυχία μία συνεργασία στο πλαίσιο των παρεχόμενων υπηρεσιών πρωτοβάθμιας φροντίδας υγείας, η οποία απαρτιζόταν από μέρη με διαφορετικούς ρόλους. Στη συγκεκριμένη μελέτη επρόκειτο για σύναψη συνεργασίας μεταξύ των δήμων και των επαγγελματιών υγείας, με σκοπό την προαγωγή του συμφέροντος, της εξυπηρέτησης και της ασφάλειας των ασθενών. Το εγχείρημα αποδείχθηκε ασύμμετρο και δεν επιτεύχθηκε ποτέ η σύμπραξη των εμπλεκόμενων φορέων, καθώς τα κίνητρα των ιατρών προσβλέπανε στην άσκηση επιρροής στις δημοτικές αποφάσεις υπέρ του φορέα υγείας που αντιπροσώπευαν. Αντιθέτως, οι δημοτικές αρχές είχαν όφελος για τη σύναψη της συνεργασίας αυτής με τους ιατρούς, τη στιγμή κατά την οποία θα διασφάλιζαν υψηλότερα επίπεδα εξυπηρέτησης και κάλυψης των αναγκών των πολιτών. Στην προκειμένη περίπτωση, υπάρχει αναντιστοιχία μεταξύ προσφοράς και ζήτησης ανάγκης συνεργασίας, εφόσον οι ιατροί δεν επωφελούνται από τη συνεργασία αυτή, εν αντιθέσει με τις δημοτικές αρχές.

Τα ευρήματα που αφορούν τις μελέτες περίπτωσης αποτελούν απτά παραδείγματα για το ότι το πλαίσιο μίας μη ισορροπημένης αμοιβαιότητας προάγει καταστάσεις έμμεσης διαχείρισης του συνεργατικού σχήματος, καθώς επίσης και προσπάθειες δημιουργίας στρατηγικών για την οικοδόμηση εμπιστοσύνης και πειθούς⁶³⁶.

Τα διευθυντικά και διοικητικά στελέχη των συνεργατικών σχημάτων, λοιπόν, είναι απαραίτητο να κατανοήσουν την αναγκαιότητα της εκάστοτε κατάστασης. Επισημαίνεται ότι, η εμπιστοσύνη στον ρόλο της διπλωματίας βασίζεται συχνά στην

⁶³⁴ Βλ. S. Page, *όπ.παρ.*, σελ. 311-339.

⁶³⁵ Βλ. H. Heen, One size does not fit all, *Public Management Reviews*, Ap. 11, 2009, σελ. 235–253.

⁶³⁶ Βλ. H. Heen, *όπ.παρ.*, σελ. 235-253.

εξουσία και συνδέεται εξίσου συχνά με την αντίληψη ότι ένα εμπλεκόμενο μέρος έχει επιρροή σε ένα άλλο. Επιπροσθέτως, υποστηρίζεται ότι, καθώς η σύγκρουση είναι αναπόφευκτη, μπορεί να προκύψει αντίπαλος ρόλος της διαχείρισης του εκάστοτε συνεργατικού σχήματος, λόγω του υποχρεωτικού συμβατικού χαρακτήρα του συνεργατικού δικτύου. Τέλος, ο ρόλος του διαχειριστή ενός δικτύου συνεργατών υποδηλώνει ότι ένα δίκτυο έχει θεσμοθετηθεί και οι φορείς που εμπλέκονται σε αυτό έχουν πλήρως και νομίμως ενσωματωθεί, αλλά εξακολουθούν να απαιτούν την περαιτέρω ανάπτυξή του. Με παρόμοιο τρόπο, ορισμένες επιπρόσθετες μελέτες παρείχαν πληροφορίες για τις διαχειριστικές παρεμβάσεις που μπορούν να χρησιμοποιηθούν σε κάθε στάδιο της διαδικασίας της συνεργασίας, επισημαίνοντας τη σημασία της λογικής κατά την επιλογή του διαχειριστικού ρόλου που απαιτείται σε κάθε στάδιο του εκάστοτε υφιστάμενου συνεργατικού σχήματος⁶³⁷.

Εν ολίγοις, οι κύριες συνέπειες για αυτές τις στρατηγικές περιλαμβάνουν την αυξημένη επίλυση συγκρούσεων, τη διαμόρφωση ατζέντας χωρίς αποκλεισμούς, τη θεσμική ικανότητα, τον κοινό χαρακτήρα του υφιστάμενου σκοπού και την ισορροπία στην άσκηση εξουσίας και διοίκησης στο εκάστοτε συνεργατικό σχήμα.

⁶³⁷ Βλ. E. Weber, & A. Khademian, *όπ.παρ.*, σελ. 334-349.

4.4 Η Τεχνητή Νοημοσύνη ως Εργαλείο στη Διαχείριση του Ανθρώπινου Δυναμικού

4.4.1 Εισαγωγικές Παρατηρήσεις

Το ζήτημα του ψηφιακού μετασχηματισμού στη διαχείριση ανθρώπινου δυναμικού και η μετακύληση αυτού από τα μεγάλα δεδομένα (big data) στη μηχανική μάθηση (machine learning) και την τεχνητή νοημοσύνη, έχει κεντρίσει ιδιαίτερα το ενδιαφέρον ακαδημαϊκών και επιστημονικών συζητήσεων⁶³⁸. Ωστόσο, ο βαθμός στον οποίο οι συζητήσεις και το θεωρητικό πλαίσιο που αναπτύσσονται γύρω από τους συγκεκριμένους άξονες ευθυγραμμίζονται με την πραγματικότητα, αποτελεί ένα διαφορετικό ζήτημα, το οποίο μπορεί να θεωρηθεί ακόμα και πρόκληση⁶³⁹.

Είναι γεγονός ότι κατά τη διάρκεια των τελευταίων ετών πραγματοποιήθηκε σημαντική πρόοδος στους τομείς της αναγνώρισης προτύπων και της επεξεργασίας της φυσικής γλώσσας, καθώς επίσης και στον τομέα της εκμάθησης μέσω της αξιοποίησης νευρωνικών δικτύων⁶⁴⁰. Μάλιστα, η αξιοποίηση των νευρωνικών δικτύων έχει καταστεί ολοένα και πιο συνήθης σε περιβάλλοντα τα οποία χαρακτηρίζονται από ποικιλομορφία δεδομένων. Η ανωτέρω αναφερόμενη πρόοδος έχει οδηγήσει τον ακαδημαϊκό και επιστημονικό κόσμο ένα βήμα πλησιέστερα στο πεδίο και την ενσωμάτωση της τεχνητής νοημοσύνης, η οποία, όπως έχει ήδη αναφερθεί σε προηγούμενα κεφάλαια, αντιπροσωπεύει την ικανότητα των μηχανών να μιμούνται την προσαρμοστική λήψη αποφάσεων του ανθρώπου⁶⁴¹. Εντούτοις, ως προς το ζήτημα της διαχείρισης ανθρώπινου δυναμικού, έχει ήδη επικεντρωθεί η προσοχή στην ανάπτυξη του σταδίου των μεγάλων δεδομένων.

⁶³⁸ Βλ. M.C. Elish & D. Boyd, Situating methods in the magic of Big Data and AI, *Communication Monographs*, Τεύχος 85, Αρ. 1, 2017, σελ. 57–80.

⁶³⁹ Βλ. M.C. Elish & D. Boyd, *όπ.παρ.*, σελ. 57–80.

⁶⁴⁰ Βλ. L. Zhou, S. Pan, J. Wang & V. Vasilakos, Machine learning on big data: Opportunities and challenges, *Neurocomputing*, Αρ. 237, 2017, σελ. 350–361.

⁶⁴¹ Βλ. S. Barocas & A. Selbst, Big data's disparate impact. *SSRN Electronic Journal*, Αρ. 104, 2016, σελ. 671–732

Η ανάλυση δεδομένων σε τομείς όπως είναι για παράδειγμα αυτός του μάρκετινγκ, μολονότι εξακολουθούν να υφίστανται πολλαπλά ερωτήματα που χρήζουν απάντησης, εντούτοις, η κάλυψή τους τείνει να είναι περισσότερο σαφής, συγκριτικά με την αντίστοιχη που προέρχεται από τη διαχείριση ανθρώπινου δυναμικού, η οποία χαρακτηρίζεται από ιδιαίτερη πολυπλοκότητα⁶⁴². Χαρακτηριστικά παραδείγματα τέτοιου είδους ερωτημάτων, είναι το ποιοι είναι οι προβλεπτικοί παράγοντες για το αν το πελατειακό κοινό θα αγοράσει ένα συγκεκριμένο προϊόν ή υπηρεσία, ή ακόμα και με ποιον τρόπο οι αλλαγές στην εμφάνιση ενός προϊόντος μπορούν να ασκήσουν επιρροή στις πωλήσεις του. Πρόκειται για ερωτήματα, οι απαντήσεις των οποίων είναι εύκολα μετρήσιμες, ενώ απαιτούν δεδομένα τα οποία ήδη συλλέγονται ηλεκτρονικά μέσω της διαδικασίας των πωλήσεων⁶⁴³.

Επιπροσθέτως, το πλήθος των υφιστάμενων παρατηρήσεων, δηλαδή των πωλήσεων συγκεκριμένων προϊόντων και υπηρεσιών σε εθνικό επίπεδο, μεταξύ των ετών, διευκολύνει την απόδοση απάντησης στα ανωτέρω αναφερόμενα ερωτήματα, τα οποία δεν αφορούν στη διαχείριση ανθρώπινου δυναμικού αλλά άλλον επιστημονικό κλάδο. Οπότε, αιτιολογείται το γεγονός της ευκολότερης εφαρμογής και ενσωμάτωσης τεχνικών μεγάλων δεδομένων σε άλλα επιστημονικά πεδία, πλην αυτού της διαχείρισης ανθρωπίνων πόρων⁶⁴⁴. Επί παραδείγματι, στην περίπτωση του επιστημονικού πεδίου του μάρκετινγκ, αν και δεν είναι απόλυτα απαλλαγμένο από ηθικά διλήμματα, η ιδέα ότι οι οργανισμοί καλούνται να καταβάλουν προσπάθειες για την αύξηση των πωλήσεων των προϊόντων και υπηρεσιών τους είναι αποδεκτή. Το ίδιο συμβαίνει και με την περίπτωση της ιδέας ότι οι οργανισμοί καταβάλουν προσπάθειες για να ασκήσουν επιρροή στο εν δυνάμει ή υφιστάμενο πελατειακό κοινό για την αύξηση των αγορών των προϊόντων και υπηρεσιών που προσφέρει η εκάστοτε συγκεκριμένη εταιρία⁶⁴⁵.

⁶⁴² Βλ. S. Huang, & J. Xiang, Big Data Light up Wisdom of Human Resources Management System, Science & Technology for Chinese Mass Media, Ap. 12, 2013, σελ. 76-78.

⁶⁴³ Βλ. S. Huang, & J. Xiang, όπ.παρ., σελ. 76-78.

⁶⁴⁴ Βλ. H. Tian, The Concept, Thinking of Human Resources Management Based on the Era of Big Data, Human Resources, Ap. 20, 2014, σελ. 162-163.

⁶⁴⁵ Βλ. G.H. Zhou, Thinking on Human Resources Management Based on the Value of Big Data, Science Mosaic, Ap. 12, 2013, σελ. 104-107.

Η αποτελεσματική εφαρμογή της τεχνητής νοημοσύνης σε ζητήματα που είναι συνδεδεμένα με τη διαχείριση ανθρωπίνων πόρων παρουσιάζει διαφορετικές προκλήσεις, συγκριτικά με τα υπόλοιπα επιστημονικά πεδία⁶⁴⁶, όπως στο ανωτέρω παράδειγμα του μάρκετινγκ. Πιο συγκεκριμένα, παρά το γεγονός ότι σχεδόν το σύνολο των διεξαγόμενων εργασιών που αποσκοπούν στη βελτίωση της διαχείρισης των ανθρωπίνων πόρων δεν απαλλάσσονται από την αξιοποίηση τεχνικών αναλύσεων δεδομένων, λόγω χάρη αλγορίθμων που βοηθούν τη διαδικασία της πρόβλεψης των τελικών αποτελεσμάτων, όπως είναι το ποιος εργαζόμενος έχει καλύτερες προοπτικές εξέλιξης και προόδου στην εταιρία που εργάζεται, ωστόσο, εντοπίζονται ορισμένα προβληματικά σημεία⁶⁴⁷, τα οποία συνοψίζονται ως κάτωθι:

Το πρώτο εντοπισμένο προβληματικό σημείο αφορά την αυξημένη πολυπλοκότητα των αποτελεσμάτων και δεδομένων που εξάγονται από το τμήμα διαχείρισης ανθρωπίνων πόρων⁶⁴⁸. Επί παραδείγματι, δεν είναι απόλυτα ξεκάθαρη η απάντηση στο ερώτημα του ποιος μπορεί να θεωρείται «καλός» υπάλληλος. Πρόκειται για ένα ερώτημα, που επιδέχεται πολυδιάστατες απαντήσεις, η ακριβής μέτρηση και αποτύπωση των οποίων, για τις περισσότερες περιπτώσεις, θεωρείται αρκετά δυσχερές. Δεν είναι άλλωστε τυχαίο το γεγονός ότι οι τεχνικές αξιολόγησης της εργασιακής αποδοτικότητας έχουν επικριθεί σθεναρά για την πρόκληση ζητημάτων αξιοπιστίας και εγκυρότητας, καθώς επίσης και για το γεγονός ότι δεν μπορούν να απαλλαχτούν εύκολα από προκαταλήψεις, ενώ ταυτόχρονα, χαρακτηρίζονται από υψηλά επίπεδα υποκειμενικότητας. Όλα τα ανωτέρω, αντικατοπτρίζουν τους λόγους για τους οποίους η πλειοψηφία των εργοδοτών αποφεύγει την εφαρμογή πρακτικών αξιολόγησης της εργασιακής αποδοτικότητας του ανθρώπινου δυναμικού τους⁶⁴⁹.

Άλλωστε, υποστηρίζεται ότι οποιαδήποτε εργασία μπορεί να θεωρηθεί ως λογικά πολύπλοκη, εμπεριέχει τον υψηλό βαθμό εξάρτησης και επιρροής της από άλλες εργασίες και ως εκ τούτου, η ατομική εργασιακή αποδοτικότητα ενός υπαλλήλου δεν

⁶⁴⁶ Βλ. H. Tian, όπ.παρ., σελ. 162-163.

⁶⁴⁷ Βλ. G.H. Zhou, όπ.παρ., σελ. 104-107.

⁶⁴⁸ Βλ. X.F. Meng & X. Ci, Big Data Management: Concepts, Techniques and Challenges, Journal of Computer Research and Development, Ap. 50, 2013, σελ. 146-169.

⁶⁴⁹ Βλ. P. Cappelli & A. Tavis, The Performance Management Revolution, Harvard Business Review, 2017, στο <https://hbr.org/2016/10/the-performance-management-revolution>

είναι εύκολα διαχωρίσιμη από την ομαδική, τη στιγμή κατά την οποία για την διεκπεραίωση μιας συγκεκριμένης εργασίας δεν μπορεί να ευθύνεται αποκλειστικά και μόνο ένας εργαζόμενος⁶⁵⁰.

Ένα δεύτερο προβληματικό σημείο που εντοπίστηκε έγκειται στο ότι τα σύνολα δεδομένων που αφορούν στο ανθρώπινο δυναμικό ενός οργανισμού τείνουν να είναι αρκετά περιορισμένα, σε σύγκριση με τα πρότυπα που ορίζονται από την επιστήμη της ανάλυσης δεδομένων⁶⁵¹. Για παράδειγμα, ο αριθμός των απασχολούμενων ανθρώπων σε έναν οργανισμό, ακόμα και αν αυτός χαρακτηρίζεται από μεγάλο μέγεθος, είναι ασήμαντος συγκριτικά με το πλήθος των αγορών στις οποίες προβαίνουν οι πελάτες του οργανισμού αυτού ή της ευρύτερης αγοράς στον κλάδο που δραστηριοποιείται αυτός ο οργανισμός. Συν τοις άλλοις, σπάνια παρατηρείται αξιοπρόσεκτος και σημαντικός όγκος ενδιαφερόντων αποτελεσμάτων που προέρχονται από τον κλάδο της διαχείρισης ανθρωπίνων πόρων, όπως για παράδειγμα το ποιος ή ποιοι εργαζόμενοι κατέληξαν να απολυθούν λόγω κακής εργασιακής απόδοσης. Οι τεχνικές που ορίζονται στο πλαίσιο της επιστήμης των δεδομένων δεν χαρακτηρίζονται από αυξημένες αποδόσεις όταν καλούνται να προβαίνουν σε προβλέψεις σχετικά σπάνιων αποτελεσμάτων⁶⁵².

Ένα τρίτο εντοπισμένο πρόβλημα αναφέρεται στο ότι τα αποτελέσματα που πηγάζουν από τις αποφάσεις που λαμβάνονται στο πλαίσιο της διαχείρισης ανθρώπινου δυναμικού, όπως είναι το ποιοι εργαζόμενοι απολύονται και ποιοι προσλαμβάνονται, εσωκλείουν ιδιάζουσας σημασίας συνέπειες τόσο για τους ίδιους τους εργαζόμενους όσο και για το ευρύτερο κοινωνικό σύνολο⁶⁵³, γεγονός το οποίο καθιστά απαραίτητη την παρουσία κατάλληλα διαμορφωμένων νομικών πλαισίων, εντός των οποίων οι εργοδότες θα λαμβάνουν αποφάσεις χωρίς να καταπατούν θεμελιώδεις αρχές και αξίες.

⁶⁵⁰ Βλ. J. Pfeffer & R. Sutton, *Hard Facts, Dangerous Half-Truths and Total Nonsense: Profiting from Evidence-Based Management, Strategy and Leadership*, Τεύχος 34, Αρ. 2, 2006, σελ. 35-42.

⁶⁵¹ Βλ. X.F. Meng & X. Ci, *όπ.παρ.*, σελ. 146-169.

⁶⁵² Βλ. X.F. Meng & X. Ci, *όπ.παρ.*, σελ. 146-169.

⁶⁵³ Βλ. W. Niehueser & G. Boak, *Introducing artificial intelligence into a human resources function, Industrial and Commercial Training*, Τεύχος 52, Αρ. 2, 2020, σελ. 121-130.

Για το ευρύτερο κοινωνικό σύνολο, τα αποτελέσματα τέτοιου είδους αποφάσεων θεωρούνται ανησυχητικά, τη στιγμή κατά την οποία είναι άμεσα συνυφασμένα με ζητήματα απόδοσης δικαιοσύνης⁶⁵⁴. Επίσης, αποφάσεις που αφορούν το καθεστώς απασχόλησης των εργαζομένων σε έναν οργανισμό, υπόκεινται σε ένα σύνολο πολυδιάστατων και πολύπλοκων κοινωνικών και ψυχολογικών προβληματισμών. Χαρακτηριστικά παραδείγματα τέτοιου είδους προβληματισμών είναι οι προσωπικές αξίες, οι συμβατικές προσδοκίες, η εκτίμηση της απόδοσης δικαιοσύνης κατά τη λήψη αποφάσεων που είναι συναρτημένες με ζητήματα διαχείρισης ανθρώπινου δυναμικού και οι οποίες ασκούν επιρροή στα οργανωτικά αποτελέσματα, αλλά και διακριτικά στον ίδιο τον άνθρωπο⁶⁵⁵. Κατά συνέπεια, το να είναι ένας οργανισμός σε θέση να εξηγήσει και να αιτιολογήσει τις πρακτικές που χρησιμοποιεί στο πλαίσιο της διαχείρισης ανθρωπίνων πόρων καθίσταται πολύ πιο σημαντικό συγκριτικά με την περίπτωση άλλων επιστημονικών πεδίων. Τέλος, επακόλουθες είναι οι αντιδράσεις του ανθρώπινου δυναμικού, καθώς αποτελούν αναπόσπαστο μέρος του εσωτερικού περιβάλλοντος του οργανισμού, συμμετέχουν ενεργά στις δραστηριότητες αυτού και συνεπώς, έχουν την δυνατότητα είτε να αποδεχτούν είτε να απορρίψουν αλγοριθμικές αποφάσεις⁶⁵⁶.

Προκειμένου να καταστούν περισσότερο κατανοητές οι εν λόγω αναφερόμενες ανησυχίες, αρκεί να αναλογιστεί κανείς τη χρήση ενός αλγορίθμου για την πρόβλεψη του ποιοι εργαζόμενοι θα προσληφθούν σε μία εταιρεία, με κριτήριο την πλήρωση των χαρακτηριστικών που διέπουν τους «καλούς» υπαλλήλους. Ακόμα και αν ήταν εφικτή η απόδειξη της ύπαρξης μίας αιτιώδους σχέσης μεταξύ του φύλου του εργαζόμενου και της εργασιακής του αποδοτικότητας, ένας αλγόριθμος που προτείνει την πρόσληψη, επί παραδείγματι, περισσότερων ανθρώπων με ανοιχτό χρώμα δέρματος, μπορεί να μην είναι αξιόπιστος. Ο λόγος είναι ότι το ίδιο το μέγεθος της εργασιακής αποδοτικότητας

⁶⁵⁴ Βλ. P. Hogg, Artificial intelligence: HR friend or foe, Strategic HR Review, Τεύχος 12, Αρ. 2, 2019, σελ. 47-51.

⁶⁵⁵ Βλ. P. Hogg, όπ.παρ., σελ. 47-51.

⁶⁵⁶ Βλ. A. Abubakar, E. Behraves, H. Rezapouraghdam & S. Yildiz, Applying artificial intelligence technique to predict knowledge hiding behavior, International Journal of Information Management, Αρ. 49, 2019, σελ. 45-57.

μπορεί να αποτελεί έναν δείκτη που συνοδεύεται από υψηλά επίπεδα προκαταλήψεων⁶⁵⁷.

Επιπροσθέτως, είναι πιθανό τα χαρακτηριστικά του τρέχοντος απασχολούμενου ανθρώπινου δυναμικού σε έναν οργανισμό να αλλοιωθούν και να μετασχηματιστούν με το πέρασ των ετών και ακολούθως, οι ίδιοι άνθρωποι να καταλήξουν να έχουν διαφορετικά χαρακτηριστικά, όσον αφορά την εργασιακή τους αποδοτικότητα, συγκριτικά με την στιγμή που προσλήφθηκαν. Μία πιθανή λύση σε τέτοιου είδους ζητήματα θα μπορούσε να είναι η δημιουργία ενός αλγορίθμου, ο οποίος θα λειτουργεί με πιο αντικειμενικά κριτήρια, όπως είναι για παράδειγμα ο προσδιορισμός των ανθρώπων που απολύονται λόγω κακής εργασιακής αποδοτικότητας⁶⁵⁸.

Όμως, η κατασκευή ενός τέτοιου αλγορίθμου δεν αποτελεί ένα σενάριο εύκολα υλοποιήσιμο για την πλειοψηφία των οργανισμών, ειδικότερα για αυτούς που χαρακτηρίζονται από μικρομεσαίο μέγεθος. Επιπλέον, σε περίπτωση κατά την οποία αιτούντες εργασία ή οι ήδη απασχολούμενοι σε έναν οργανισμό ανακαλύψουν το περιεχόμενο και τον τρόπο λειτουργίας του αλγορίθμου πρόσληψής τους, όπως επί παραδείγματι το ότι χρησιμοποιεί αποδεικτικά στοιχεία που αντλούνται κατά τη διάρκεια των συνεντεύξεων αναφορικά με τα προβλήματα που είχαν δημιουργηθεί στους προηγούμενους χώρους απασχόλησης των εργαζομένων, τότε είναι πιθανό οι εργαζόμενοι να απαντήσουν διαφορετικά στις συνεντεύξεις, με σκοπό την παροχή λανθασμένων δεδομένων στον αλγόριθμο, γεγονός το οποίο θα τον καταστήσει ανούσιο και μη έγκυρο⁶⁵⁹.

Στο παρόν κεφάλαιο, καθεμία από τις ανωτέρω αναφερόμενες προκλήσεις αντιμετωπίζονται και αναλύονται ξεχωριστά, σε κάθε στάδιο του λεγόμενου κύκλου ζωής της τεχνητής νοημοσύνης. Ο κύκλος ζωής της τεχνητής νοημοσύνης απαρτίζεται συγκεκριμένα από τις Λειτουργίες, τη Δημιουργία Δεδομένων, τη Μηχανική Μάθηση και τη Λήψη Αποφάσεων, όπως θα αναλυθεί στην επόμενη ενότητα. Επίσης, για την

⁶⁵⁷ Βλ. E. Barends & D. Rousseau, Evidence-Based Management: How to Use Evidence to Make Better Organizational Decisions, Kogan Page, 2018, σελ. 36-42.

⁶⁵⁸ Βλ. E. Barends & D. Rousseau, *όπ.παρ.*, σελ. 45-48.

⁶⁵⁹ Βλ. J. Pfeffer & R. Sutton, *όπ.παρ.*, σελ. 35-42.

ανάλυση που πρόκειται να ακολουθήσει έχει δοθεί έμφαση σε βασικές ιδέες που απορρέουν από το πεδίο της διαχείρισης ανθρωπίνων πόρων⁶⁶⁰.

Κατόπιν, συνοψίζοντας τα στοιχεία που αντλήθηκαν από την ανάλυση που θα διεξαχθεί, δίνονται προτάσεις για τον τρόπο με τον οποίο θα ήταν εφικτή η πραγματοποίηση προόδου στην εφαρμογή εργαλείων μηχανικής μάθησης στον τομέα της διαχείρισης ανθρωπίνων πόρων, δεδομένης της παρουσίας των ανωτέρω αναφερόμενων προβληματικών και περιοριστικών σημείων. Συγκεκριμένα, δίνεται έμφαση στον ρόλο που διαδραματίζεται από την πλευρά των αιτιολογικών μοντέλων στον τομέα της μηχανικής μάθησης⁶⁶¹.

Επιπροσθέτως, προτείνεται ότι η τυχαιοποίηση μπορεί να καταστεί χρήσιμη μέθοδος, στο πλαίσιο της διαδικασίας λήψης αποφάσεων, δεδομένης της διατήρησης της αξίας της δικαιοσύνης και των δυσκολιών που ενδεχομένως αντιμετωπίζουν οι αναλυτές στη λήψη δίκαιων και έγκυρων αποφάσεων⁶⁶².

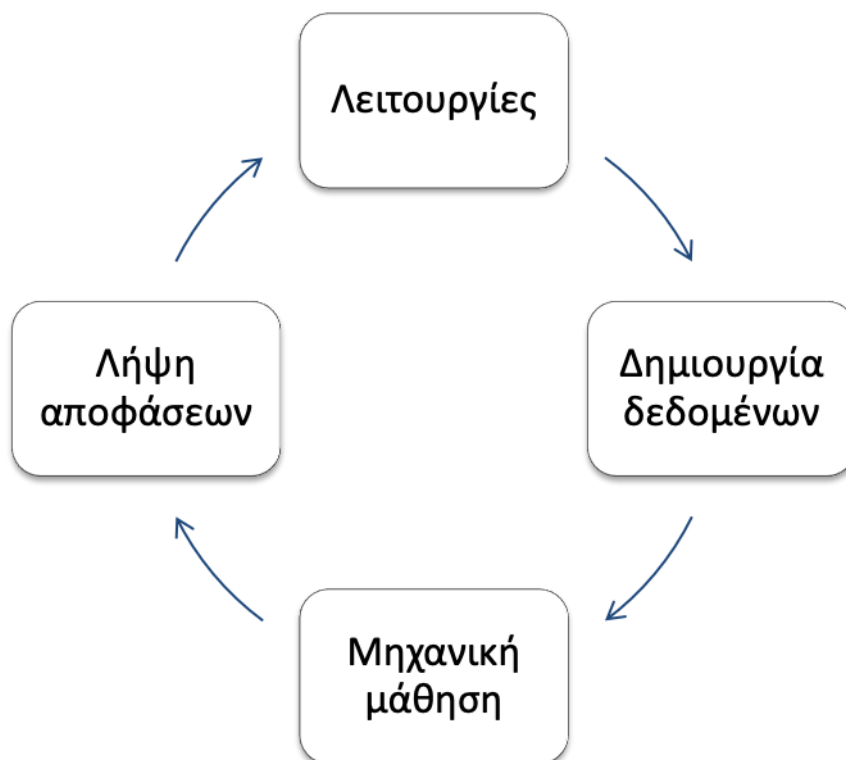
⁶⁶⁰ Βλ. D. Rousseau, *The Oxford Handbook of Evidence-Based Management*, Oxford University Press, 2014, σελ. 13-19.

⁶⁶¹ Βλ. J. Pearl, *The Book of Why: The New Science of Cause and Effect*, Basic Books, 2018, σελ. 23-29.

⁶⁶² Βλ. J. Denrell, C. Fang & C. Liu, *Change Explanations in Management Science*. *Organization Science*, Τεύχος 26, Αρ. 3, 2015, σελ. 923-940.

4.4.2 Ο Κύκλος Ζωής της Ενσωμάτωσης της Τεχνητής Νοημοσύνης στη Διαχείριση του Ανθρώπινου Δυναμικού

Αρχικά, παρουσιάζεται ο κύκλος ζωής της τεχνητής νοημοσύνης (Διάγραμμα 1), ο οποίος όπως προαναφέρθηκε απαρτίζεται από τα εξής στάδια: Λειτουργίες, δημιουργία δεδομένων, μηχανική μάθηση και λήψη αποφάσεων⁶⁶³.



Διάγραμμα 1. Κύκλος ζωής τεχνητής νοημοσύνης

Οι λειτουργίες αποτελούν και το κυρίαρχο σημείο ενδιαφέροντος του κύκλου ζωής της τεχνητής νοημοσύνης, καθώς αντικατοπτρίζουν ζητήματα όπως την αποσαφήνιση του τρόπου με τον οποίο μία νομική οντότητα προβαίνει στην πρόσληψη ανθρώπινου δυναμικού. Ένας από τους λόγους προσέλκυσης ενδιαφέροντος για το ζήτημα της εφαρμογής εργαλείων που απορρέουν από την επιστήμη των δεδομένων στο πλαίσιο της διαχείρισης ανθρώπινων πόρων έγκειται στο ότι η διαχείριση ανθρώπινου

⁶⁶³Βλ. D. Rousseau, όπ.παρ., σελ. 25-28.

δυναμικού περιλαμβάνει πολυδιάστατες και πολυπληθείς λειτουργίες, ενώ παράλληλα συγκεντρώνει μεγάλο μέρος των οικονομικών πόρων ενός οργανισμού. Αυτό επιβεβαιώνεται από το γεγονός ότι η πλειοψηφία του ποσοστού των οργανωσιακών δαπανών προορίζεται προς το ανθρώπινο δυναμικό που απασχολείται σε έναν οργανισμό⁶⁶⁴. Στον πίνακα που ακολουθεί, παρουσιάζονται οι πλέον συνηθέστερες λειτουργίες που εμπίπτουν στο πλαίσιο της διαχείρισης ανθρωπίνων πόρων, σε αντιστοίχιση με προβλέψεις που αφορούν στην ανάλυση του απασχολούμενου ανθρώπινου δυναμικού:

Πίνακας 3⁶⁶⁵

Λειτουργία διαχείρισης ανθρωπίνων πόρων	Προβλέψεις
Στρατολόγηση: Πρόκειται για το στάδιο του εντοπισμού των πιθανών υποψηφίων εργαζομένων σε έναν οργανισμό.	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Εξασφαλίζουμε «καλούς» εργαζόμενους για τον οργανισμό μας, που πληρούν τις προϋποθέσεις και τα κριτήρια που έχουμε θέσει;
Επιλογή: Η συγκεκριμένη λειτουργία αφορά στην επιλογή των εργαζομένων που θα κληθούν να εργαστούν σε έναν συγκεκριμένο οργανισμό.	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Έχουμε προχωρήσει στην πρόσληψη αυτών που θα αποδειχθούν ότι είναι οι καταλληλότεροι υπάλληλοι για τις συγκεκριμένες θέσεις εργασίας που στελεχώνουν;

⁶⁶⁴ Βλ. D. Rousseau, όπ.παρ., σελ. 46-51.

⁶⁶⁵ Βλ. D. Rousseau, όπ.παρ., σελ. 63-65.

Είσοδος: Πρόκειται για το στάδιο της εισόδου του εργαζόμενου στον οργανισμό. Το συγκεκριμένο στάδιο αναφέρεται στην πρώτη επαφή του εργαζόμενου με το εσωτερικό περιβάλλον του οργανισμού.	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Ποιες πρακτικές μπορούν να ωθήσουν τα νεοπροσληφθέντα άτομα για να προσφέρουν αξία στον οργανισμό το ταχύτερο δυνατόν;
Εκπαίδευση: Πρόκειται για το στάδιο, στο οποίο τίθεται σε πρακτική εφαρμογή η απάντηση που αποδόθηκε στο ερώτημα του ανθρώπινου δυναμικού στο αμέσως προηγούμενο στάδιο (πρακτικές που μπορούν να εφαρμοστούν για την απόδοση αξίας στον οργανισμό από τους νεοπροσληφθέντες).	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Ποιες εκπαιδευτικές παρεμβάσεις είναι χρήσιμες, σε ποια άτομα πρέπει να εφαρμοστούν και πώς, ώστε να επέλθει βελτίωση της εργασιακής αποδοτικότητάς τους;
Διαχείριση απόδοσης: Το συγκεκριμένο στάδιο σχετίζεται με τον προσδιορισμό και την αξιολόγηση της εργασιακής αποδοτικότητας του ανθρώπινου δυναμικού.	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Οι πρακτικές του οργανισμού αυτού, καταλήγουν να βελτιώνουν την αποδοτικότητα των υπαλλήλων;
Προαγωγή-Εξέλιξη: Το συγκεκριμένο στάδιο αφορά τον προσδιορισμό των εργαζομένων που προάγονται σε θέσεις υψηλότερης ευθύνης ή/και υψηλότερης ιεραρχίας.	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Μπορούμε να προβλέψουμε ποιοι είναι οι εργαζόμενοι που θα εμφανίσουν καλύτερη εργασιακή αποδοτικότητα όταν τους ανατεθούν νέοι εργασιακοί ρόλοι και περισσότερες ευθύνες;

Διαχείριση κύκλου εργασιών ανθρώπινου δυναμικού: Το συγκεκριμένο στάδιο αφορά ζητήματα παραμονής των ήδη απασχολούμενων εργαζομένων στον οργανισμό στον οποίο απασχολούνται. Εναλλακτικά, το συγκεκριμένο στάδιο μπορεί να εμπεριέχει και τον βαθμό πιστότητας των απασχολούμενων στον οργανισμό απασχόλησής τους.	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Μπορούμε να προβλέψουμε ποιοι εργαζόμενοι είναι πιθανό να παραιτηθούν από την εργασία τους ή να απολυθούν;
Οφέλη ανθρώπινου δυναμικού: Το συγκεκριμένο στάδιο εμπεριέχει τα κίνητρα που αποδίδονται από την πλευρά του οργανισμού προς τους εργαζόμενους, προκειμένου αυτοί να είναι σε θέση να αυξήσουν την εργασιακή τους αποδοτικότητα, τα επίπεδα της οργανωσιακής πιστότητας, αλλά και να είναι περισσότερο ικανοποιημένοι από το εργασιακό τους περιβάλλον και τις συνθήκες άσκησης των επαγγελματικών τους καθηκόντων.	Το ερώτημα που καλείται να απαντήσει το τμήμα διαχείρισης ανθρωπίνων πόρων στην προκειμένη περίπτωση είναι το εξής: Μπορούμε να προσδιορίσουμε ποια είναι τα σημαντικότερα οφέλη για το απασχολούμενο ανθρώπινο δυναμικό;

Κάθε μία από τις παραπάνω λειτουργίες που εμπίπτουν στο πλαίσιο της διοίκησης ανθρωπίνων πόρων εσωκλείει διοικητικά καθήκοντα. Κάθε ένα εξ αυτών ασκεί επιρροή στην απόδοση του ίδιου του οργανισμού ενώ παράλληλα, αφορά συγκεκριμένες θέσεις εργασίας και ρόλους, επαγγελματικές αρμοδιότητες και ευθύνες, την απόδοση γραπτών οδηγιών και κατευθύνσεων, καθώς επίσης και τις πραγματικές αλληλεπιδράσεις μεταξύ όλων των εμπλεκόμενων μερών που εντάσσονται στο εσωτερικό περιβάλλον του εκάστοτε οργανισμού. Από την πλευρά των ανωτέρω λειτουργιών προάγεται ένας συγκεκριμένος όγκος δεδομένων, με τη μορφή κειμένων,

ηλεκτρονικών ή ηχογραφημένων αρχείων και δεδομένων που είναι διαθέσιμα σε οποιαδήποτε άλλη μορφή⁶⁶⁶.

Καθώς οι παραπάνω λειτουργίες μετακυλίνουν στο πεδίο της εικονικής πραγματικότητας και τεχνητής νοημοσύνης, πολλές από τις εξόδους δεδομένων θα έχουν τη μορφή «ψηφιακής εξάτμισης», η οποία αφορά στην ανίχνευση δεδομένων από ψηφιακές δραστηριότητες, η χρήση των οποίων θα είναι χρήσιμη για τη δημιουργία αλγορίθμων πρόσληψης ανθρώπινου δυναμικού⁶⁶⁷. Χαρακτηριστικά παραδείγματα τέτοιου είδους ανιχνεύσιμων δεδομένων είναι οι διαδικτυακές εφαρμογές εργασίας, καθώς επίσης και η ενσωμάτωση συστημάτων αυτόματης αξιολόγησης των εργασιακών δεξιοτήτων των απασχολούμενων.

Τα πληροφοριακά συστήματα που εμπίπτουν στο πλαίσιο της διαχείρισης ανθρώπινου δυναμικού, καθώς επίσης και τα συστήματα παρακολούθησης των αιτούντων εργασία, η ανωτέρω αναφερόμενη «ψηφιακή εξάτμιση» και λοιποί δείκτες αποτελούν κρίσιμα στοιχεία στο στάδιο της παραγωγής δεδομένων. Η είσοδος δεδομένων πρέπει να διεξάγεται από πολλές και διαφορετικές βάσεις δεδομένων, ενώ εν συνεχεία, τα δεδομένα τα οποία συλλέγονται είναι απαραίτητο να μετατρέπονται σε μία κοινή μορφή και να μπορούν να ενοποιηθούν και να συνδυαστούν πριν την έναρξη της διαδικασίας ανάλυσής τους⁶⁶⁸.

Τέλος, το στάδιο της λήψης αποφάσεων είναι και το τελευταίο του κύκλου ζωής της τεχνητής νοημοσύνης. Πρόκειται για ένα στάδιο που είναι συνυφασμένο με τον τρόπο με τον οποίο χρησιμοποιούνται πληροφοριακά δεδομένα από τα μοντέλα μηχανικής μάθησης, στο πλαίσιο των καθημερινών λειτουργιών των οργανισμών. Άλλωστε, οι εργοδότες διαθέτουν τη διακριτική ευχέρεια για τον τρόπο με τον οποίο χρησιμοποιούν τις ιδέες που πηγάζουν από τα μοντέλα μηχανικής μάθησης. Αυτό καταδεικνύει ότι στη σύγχρονη εποχή επικρατούν υψηλά επίπεδα ευελιξίας, σε αντίθεση με παλαιότερες

⁶⁶⁶ Βλ. C. Prentice, S. Dominique Lopes & X. Wang, Emotional intelligence or artificial intelligence– an employee perspective, *Journal of Hospitality Marketing & Management*, 2019, σελ. 1–27.

⁶⁶⁷ Βλ. C. Prentice, S. Dominique Lopes & X. Wang, *όπ.παρ.*, σελ. 1–27.

⁶⁶⁸ Βλ. C. Prentice, S. Dominique Lopes & X. Wang, *όπ.παρ.*, σελ. 1–27.

εποχές, όπου πολλές διαδικασίες της διαχείρισης ανθρώπινου δυναμικού ήταν περισσότερο τυποποιημένες⁶⁶⁹.

⁶⁶⁹ Βλ. M.K. Lee, D. Kusbit, E. Metsky & L. Dabbish, *όπ.παρ.*, σελ. 1603-1612.

4.4.3 Οι Προκλήσεις στην Επιτυχή Αξιοποίηση της Τεχνητής Νοημοσύνης στη Διαχείριση του Ανθρώπινου Δυναμικού

Στο παρόν υποκεφάλαιο πραγματοποιείται λεπτομερής διερεύνηση των τεσσάρων ευρύτερων προκλήσεων που σχετίζονται με την εφαρμογή της τεχνητής νοημοσύνης στο πεδίο της διοίκησης ανθρώπινων πόρων και οι οποίες αναφέρθηκαν στην ενότητα των εισαγωγικών παρατηρήσεων του παρόντος κεφαλαίου. Συγκεκριμένα, πρόκειται για το ζήτημα της πολυπλοκότητας που συνοδεύει τη διοίκηση ανθρώπινου δυναμικού, την ύπαρξη λίγων δεδομένων, τους νομικούς και ηθικούς περιορισμούς και τις ενδεχόμενες αντιδράσεις του ανθρώπινου δυναμικού σχετικά με τη διαχείριση της τεχνητής νοημοσύνης⁶⁷⁰. Προκειμένου να καταστούν οι εν λόγω προκλήσεις κατανοητές, αναλύονται υπό το πρίσμα των ιδιαίτερων σταδίων του κύκλου ζωής της τεχνητής νοημοσύνης, ο οποίος περιγράφηκε στην προηγούμενη ενότητα.

⁶⁷⁰ Βλ. Y. Zhao, M. Hryniewicki, F. Cheng, B. Fu & X. Zhu, *όπ.παρ.*, σελ. 737–758, Βλ. επίσης C. Prentice, S. Dominique Lopes & X. Wang, *όπ.παρ.*, σελ. 1–27.

4.4.3.1 Στάδιο δημιουργίας δεδομένων

Η πολυπλοκότητα που ενυπάρχει στις επιμέρους και πολυπληθείς διαστάσεις της διοίκησης ανθρωπίνων πόρων εκδηλώνεται κυρίως στο στάδιο της δημιουργίας των δεδομένων. Η πιο σημαντική μορφή πολυπλοκότητας έγκειται στη δυσκολία μέτρησης του ποιος υπάλληλος μπορεί να θεωρηθεί «καλός» ή «κακός»⁶⁷¹. Αυτό καθίσταται ακόμη πιο έντονο αν ληφθεί υπόψη το γεγονός ότι οι εργασιακές απαιτήσεις από την πλευρά των οργανισμών είναι επίσης πολυδιάστατες και πολυπληθείς, ενώ ταυτόχρονα, η παρακολούθηση της εργασιακής αποτελεσματικότητας και αποδοτικότητας των υπαλλήλων είναι ανεπαρκής και οι προκαταλήψεις αναφορικά με ζητήματα αξιολόγησής τους είναι δεδομένες, αυξάνοντας την αξιολογική υποκειμενικότητα⁶⁷². Συν τοις άλλοις, οι εργασίες που χαρακτηρίζονται από αυξημένα επίπεδα πολυπλοκότητας, κατά το πλείστον, είναι μεταξύ τους αλληλοεξαρτώμενες και ακολούθως, η απόδοση του ανθρώπινου δυναμικού είναι συχνά αναπόσπαστη από την απόδοση των εργασιακών ομάδων⁶⁷³.

Υπό την απουσία ενός σαφούς ορισμού για το τι σημαίνει να είναι κάποιος «καλός» υπάλληλος, τα τμήματα ανθρώπινου δυναμικού των οργανισμών αντιμετωπίζουν δυσχέρειες στην αξιολόγηση και μέτρηση της εργασιακής αποδοτικότητας.

Όσον αφορά το ζήτημα της διαχείρισης των δεδομένων, είναι γεγονός πως δεν προκύπτουν «ψηφιακά ίχνη» από το σύνολο των λειτουργικών στοιχείων ενός οργανισμού. Αυτό καταδεικνύει ότι για τα εναπομείναντα διαθέσιμα δεδομένα, αφενός καθίσταται ιδιαιτέρως δυσχερής η εξαγωγή και μετατροπή τους σε μία εύχρηστη μορφή και αφετέρου, συγχρόνως, να συνοδεύεται από ένα εύλογο και ανεκτό κόστος για τους οργανισμούς⁶⁷⁴. Επίσης, οι περισσότεροι οργανισμοί δεν προβαίνουν στη διατήρηση δεδομένων σχετικά με τους αιτούντες εργασία, τους οποίους ελέγχουν και

⁶⁷¹ Βλ. J. Pfeffer & R. Sutton, *όπ.παρ.*, σελ. 35-42.

⁶⁷² Βλ. J. Pfeffer & R. Sutton, *όπ.παρ.*, σελ. 35-42.

⁶⁷³ Βλ. J. Pfeffer & R. Sutton, *όπ.παρ.*, σελ. 35-42.

⁶⁷⁴ Βλ. S. Naghibi & H. Pourghasemi, A Comparative Assessment Between Three Machine Learning Models and Their Performance Comparison by Bivariate and Multivariate Statistical Methods in Groundwater Potential Mapping, *Water Resources Management*, Τεύχος 29, Αρ. 14, 2015, σελ. 5217–5236.

με τους οποίους συνομιλούν σε συνεντεύξεις⁶⁷⁵. Αυτές οι επιλογές περιορίζουν τους τύπους αναλύσεων που μπορούν να πραγματοποιηθούν, καθώς επίσης και τα πιθανά εξαγόμενα συμπεράσματα.

Επιπροσθέτως, απουσιάζει μία ευρύτατα αποδεκτή «λίστα», στην οποία θα περιλαμβάνονται «τυπικές» μεταβλητές που οι οργανισμοί θα μπορούν να επιλέγουν ώστε να συγκεντρώνουν για αυτές στοιχεία, τα οποία θα ενημερώνονται διαρκώς μέσω των αποτελεσμάτων των εργασιών του απασχολούμενου ανθρώπινου δυναμικού κατά τη διάρκεια της καθημερινής δραστηριοποίησής τους⁶⁷⁶. Η απουσία αυτή περιορίζει σε ακόμη μεγαλύτερο βαθμό το επίπεδο στο οποίο οι βέλτιστες πρακτικές στην ανάλυση δεδομένων έχουν τη δυνατότητα να υιοθετηθούν στους οργανισμούς και δη στην περίπτωση διαχείρισης των ανθρωπίνων πόρων. Τα συμπεριφορικά μέτρα, τα οποία έχουν συγκεντρωθεί από έρευνες που είναι επικεντρωμένες στις στάσεις των εργαζομένων, παρουσιάζουν σημαντική διαφοροποίηση στον τρόπο με τον οποίο χρησιμοποιούνται και αναλύονται μεταξύ των διάφορων οργανισμών⁶⁷⁷. Ακόμα και οι εντοπισμένες διαφορές στη λογιστική κόστους, μεταξύ των οργανισμών, υποδεικνύουν ότι τα δεδομένα που λαμβάνουν υπόψη οι οργανισμοί σχετικά με το κόστος των διαφόρων λειτουργιών διαφέρουν πολύ μεταξύ τους⁶⁷⁸. Για παράδειγμα, κάποιοι οργανισμοί μπορεί να μη λαμβάνουν υπόψη το κόστος εκπαίδευσης του ανθρώπινου δυναμικού τους, ενώ αυτοί που το λαμβάνουν υπόψη μπορεί να διαφοροποιούνται στον τρόπο με τον οποίο κατηγοριοποιούν τα δεδομένα της συγκεκριμένης δαπάνης⁶⁷⁹.

Η πολυπλοκότητα των διαδικασιών της διαχείρισης ανθρωπίνων πόρων στο στάδιο δημιουργίας δεδομένων δεν αποτελεί ένα ζήτημα που επήλθε πρόσφατα στο ερευνητικό και μελετητικό προσκήνιο. Στο πλαίσιο της προσπάθειας αντιμετώπισης των ανωτέρω αναφερόμενων προκλήσεων, οι οργανισμοί δύνανται να αντλήσουν σημαντικά οφέλη από όλα όσα μπορούν να διδαχθούν μέσα από τη διαχείριση της

⁶⁷⁵ Βλ. S. Naghibi & H. Pourghasemi, όπ.παρ., σελ. 5217–5236.

⁶⁷⁶ Βλ. H. Tian, όπ.παρ., σελ. 162-163.

⁶⁷⁷ Βλ. S. Naghibi & H. Pourghasemi, όπ.παρ., σελ. 5217–5236.

⁶⁷⁸ Βλ. H. Tian, όπ.παρ., σελ. 162-163.

⁶⁷⁹ Βλ. H. Tian, όπ.παρ., σελ. 162-163.

εργασιακής απόδοσης του ανθρώπινου δυναμικού τους⁶⁸⁰. Ακολουθεί συνοπτική αναφορά σε σημεία διδαχών που μπορούν να εισπράξουν οι οργανισμοί:

- Δεν είναι απαραίτητη η αναζήτηση τέλειων και εξιδανικευμένων μέτρων εργασιακής απόδοσης. Άλλωστε, τέτοιου είδους μέτρα δεν υφίστανται⁶⁸¹. Είναι προτιμότερο για τους ίδιους τους οργανισμούς να προβαίνουν στην επιλογή προσιτών δίκαιων μέτρων για την αξιολόγηση της εργασιακής απόδοσης του ανθρώπινου δυναμικού τους, τα οποία ωστόσο θα μπορούν να τα τηρούν σε συνεχή βάση.
- Οι οργανισμοί είναι απαραίτητο να προβούν στη συλλογή πληροφοριών που προέρχονται από πολλαπλές πηγές και να τις ενημερώνουν διαρκώς με την πάροδο του χρόνου. Τη συγκεκριμένη λειτουργία μπορεί να υποστηρίξει η ενσωμάτωση ψηφιακών εργαλείων ανθρώπινου δυναμικού, που καθιστούν επιτρεπτή τη διαδικασία ταχύτατων αξιολογήσεων σε πραγματικό χρόνο, μεταξύ συναδέλφων, οι οποίοι μπορεί να χρησιμοποιούν κινητές συσκευές⁶⁸².
- Η αξιοποίηση αντικειμενικών μέτρων στην αξιολόγηση της εργασιακής αποδοτικότητας του ανθρώπινου δυναμικού είναι απαραίτητη και όποτε καθίσταται εφικτό, θα πρέπει να είναι βασισμένα σε εκ των προτέρων καθορισμένους στόχους⁶⁸³. Επίσης, μπορούν να συμπληρώνονται ακόμα και από περισσότερο υποκειμενικές αξιολογήσεις, όπως είναι λόγου χάρη το αν και κατά πόσο ένας εργαζόμενος ταιριάζει και ταυτίζεται με την κουλτούρα του οργανισμού στον οποίο απασχολείται.
- Σημαντική είναι και η ενσωμάτωση δεδομένων ανθρώπινου δυναμικού σύμφωνα με τα επιχειρηματικά και οικονομικά δεδομένα του εκάστοτε οργανισμού, για την ανάλυση επιπτώσεων στην βιωσιμότητά του⁶⁸⁴.

Η πολυπλοκότητα των διαστάσεων του ανθρώπινου δυναμικού δημιουργεί ένα επιπρόσθετο προβληματικό σημείο, το οποίο έγκειται στην παρουσία πολλών

⁶⁸⁰ Βλ. A. De Mauro, M. Greco, M. Grimaldi & P. Ritala, Human resources for Big Data professions: A systematic classification of job roles and required skill sets, Information Processing & Management, Τεύχος 54, Αρ. 5, 2018, σελ. 807–817.

⁶⁸¹ Βλ. A. De Mauro, M. Greco, M. Grimaldi & P. Ritala, *όπ.παρ.*, σελ. 807–817.

⁶⁸² Βλ. A. De Mauro, M. Greco, M. Grimaldi & P. Ritala, *όπ.παρ.*, σελ. 807–817.

⁶⁸³ Βλ. A. De Mauro, M. Greco, M. Grimaldi & P. Ritala, *όπ.παρ.*, σελ. 807–817.

⁶⁸⁴ Βλ. A. De Mauro, M. Greco, M. Grimaldi & P. Ritala, *όπ.παρ.*, σελ. 807–817.

διαφορετικών προμηθευτών λογισμικών διαχείρισης δεδομένων, καθένας από τους οποίους εξειδικεύεται σε ένα συγκεκριμένο έργο ή τομέα. Στους σύγχρονους οργανισμούς παρατηρείται το φαινόμενο να διαθέτουν ένα σύστημα ψηφιακής παρακολούθησης των επιπέδων εργασιακής απόδοσης των εργαζομένων, ένα διαφορετικό λογισμικό παρακολούθησης των δεδομένων που αφορούν τους αιτούντες εργασία, καθώς επίσης και ένα τρίτο διαφορετικό λογισμικό παρακολούθησης δεδομένων μισθοδοσίας, τα οποία προέρχονται από διαφορετικούς προμηθευτές⁶⁸⁵.

Επίσης, επισημαίνεται ότι πριν την έναρξη ψηφιοποίησης των τμημάτων ανθρώπινου δυναμικού, οι οργανισμοί πρέπει να προβούν στον καθορισμό για το ποια δεδομένα είναι απαραίτητα και στη συνέχεια να προχωρήσουν στον έλεγχο του τι είναι διαθέσιμο και υπάρχει δυνατότητα εξαγωγής και μεταφοράς του σε εύχρηστη μορφή με λογικό κόστος⁶⁸⁶. Για παράδειγμα, σε περίπτωση όπου ένας οργανισμός επιθυμεί να προβεί στην αξιοποίηση ενός αλγορίθμου μηχανικής μάθησης κατά την πρόσληψη νέων εργαζομένων, κρίνεται απαραίτητο να προχωρήσει στη συλλογή ιστορικών δεδομένων για τους υποψηφίους, οι οποίοι τελικά δεν προσλήφθηκαν⁶⁸⁷.

Για να αποδείξει τη δέσμευσή της στον ψηφιακό μετασχηματισμό, καθώς και να επωφεληθεί από αυτόν, η ανώτατη διοίκηση των οργανισμών πρέπει να δώσει προτεραιότητα στη διαδικασία της ανταλλαγής δεδομένων, σε βραχυπρόθεσμο χρονικό ορίζοντα, και στη συνέχεια, μακροπρόθεσμα, να επενδύσει στη διαδικασία της τυποποίησης των δεδομένων και της ενσωμάτωσής τους σε μία ενιαία ψηφιακή πλατφόρμα⁶⁸⁸.

Μία πιο συστηματική προσέγγιση περιλαμβάνει την εξέταση της ερευνητικής βιβλιογραφίας, προκειμένου να διαπιστωθεί τι γνωστικό υπόβαθρο έχει μέχρι στιγμής σχηματιστεί για τα διάφορα ερωτήματα γύρω από το εν λόγω ζήτημα, όπως

⁶⁸⁵ Βλ. E. Yadegaridehkordi, M. Hourmand, M. Nilashi, L. Shuib, A. Ahani & O. Ibrahim, Influence of big data adoption on manufacturing companies' performance: An integrated DEMATEL-ANFIS approach, Technological Forecasting and Social Change, 2018, σελ. 1-12.

⁶⁸⁶ Βλ. E. Yadegaridehkordi, M. Hourmand, M. Nilashi, L. Shuib, A. Ahani & O. Ibrahim, όπ.παρ., σελ. 1-12.

⁶⁸⁷ Βλ. E. Barends & D. Rousseau, όπ.παρ., σελ. 64-69.

⁶⁸⁸ Βλ. E. Barends & D. Rousseau, όπ.παρ., σελ. 64-69.

υποστηρίζεται από τους Barends & Rousseau⁶⁸⁹. Το γεγονός ότι αυτή η προσέγγιση φαίνεται να μην χρησιμοποιείται σε μεγάλη συχνότητα, αντικατοπτρίζει την αποσύνδεση του κλάδου της επιστήμης της διαχείρισης δεδομένων, η οποία φαίνεται να κατανοεί την ανάλυση δεδομένων αλλά όχι τους ανθρώπινους πόρους, από το πεδίο των ανθρώπινων πόρων, το οποίο δείχνει να κατανοεί τους ανθρώπινους πόρους αλλά όχι την ανάλυση δεδομένων⁶⁹⁰. Μακροπρόθεσμα, η τεχνητή νοημοσύνη μπορεί να καταστεί ικανή να αναλύσει την ίδια την ερευνητική βιβλιογραφία για να προσδιορίσει τα συγκεκριμένα ερευνητικά ερωτήματα που δύνανται να τεθούν και τα μοντέλα που μπορούν να δοκιμαστούν με τα διαθέσιμα δεδομένα⁶⁹¹.

Όσον αφορά την περίπτωση των «μικρών δεδομένων» (Small Data), αποτελούν μία θεμελιώδη ανησυχία για την ανάλυση των ανθρώπινων πόρων. Οι περισσότεροι οργανισμοί δεν προβαίνουν σε προσλήψεις πολλών εργαζομένων και δεν προχωρούν σε επαρκείς εκτιμήσεις εργασιακής απόδοσης αυτών, ενώ παράλληλα δεν συλλέγουν αρκετά δεδομένα από πολλές και διαφορετικές πηγές για το τρέχον ανθρώπινο δυναμικό τους μέσω χρήσης τεχνικών που εμπίπτουν στο πεδίο της μηχανικής μάθησης. Ο λόγος έγκειται στο ότι οι περισσότεροι οργανισμοί δεν απασχολούν τόσο μεγάλο αριθμό εργαζομένων. Η βιβλιογραφία που υπάγεται στο πεδίο της μηχανικής μάθησης έχει αποδείξει ότι η πρόσβαση σε μεγαλύτερα δεδομένα έχει σημαντικά οφέλη όσον αφορά την ακρίβεια των προβλέψεων στο τμήμα ανθρώπινων πόρων⁶⁹².

Ταυτόχρονα, τα μικρά δεδομένα είναι συχνά επαρκή για τον εντοπισμό αιτιωδών σχέσεων, τα οποία πρέπει να καταστούν κατανοητά από την πλευρά των διευθυντικών και διοικητικών στελεχών προκειμένου να λειτουργήσουν υπό το πρίσμα γνώσεων που απορρέουν από αναλύσεις επικεντρωμένες στους ανθρώπινους πόρους⁶⁹³. Η βιβλιογραφία διαχείρισης ανθρώπινων πόρων συνοδεύεται από ένα σημαντικό πλεονέκτημα έναντι της επιστήμης της ανάλυσης δεδομένων, το οποίο έγκειται στην

⁶⁸⁹ Βλ. E. Barends & D. Rousseau, *όπ.παρ.*, σελ. 64-69.

⁶⁹⁰ Βλ. D. Wang, Evidence-Based Human Resource Management in the Era of Big Data, *Chinese and Foreign Entrepreneurs*, Αρ. 8, 2013, σελ. 159.

⁶⁹¹ Βλ. D. Wang, *όπ.παρ.*, σελ. 159.

⁶⁹² Βλ. E. Junqué de Fortuny, D. Martens & F. Provost, Predictive modeling with big data: is bigger really better, *Big Data*, Τεύχος 1, Αρ. 4, 2013, σελ. 215-226.

⁶⁹³ Βλ. J. Pearl, *όπ.παρ.*, σελ. 23-29.

άρθρωση αιτιώδους σχέσης, σε αντίθεση με την πρόβλεψη που γίνεται μέσω συσχετίσεων μεταξύ των παρατηρούμενων μεταβλητών. Μόλις πρόσφατα, ορισμένοι ισχυρισμοί που πηγάζουν από την κοινότητα της επιστήμης της πληροφορικής, διατύπωσαν το πρόβλημα της αιτιότητας ως κρίσιμο για το μέλλον της τεχνητής νοημοσύνης στις ανθρώπινες υποθέσεις⁶⁹⁴.

Όσο λιγότερα δεδομένα διατίθενται, τόσο περισσότερη θεωρία είναι απαραίτητη προκειμένου να εντοπιστούν αιτιολογικοί προγνωστικοί παράγοντες του ενδιαφέροντος αποτελέσματος. Η θεωρία, μπορεί συν τοις άλλοις, να προέλθει από ειδικές γνώσεις και διαχειριστική εμπειρία⁶⁹⁵. Η διαχείριση της τεχνητής νοημοσύνης απαιτεί από τους εκάστοτε αρμόδιους να γνωστοποιήσουν τις υποθέσεις τους, και να ασκήσουν πίεση στους υπολοίπους ενδιαφερόμενους φορείς για την ακρίβειά τους⁶⁹⁶. Η διατύπωση τέτοιου είδους υποθέσεων, συχνά, μετατρέπεται σε ένα ανταγωνιστικό περιβάλλον μεταξύ των ενδιαφερομένων μερών. Πρόκειται για ένα μέρος όπου απαιτείται η επισημοποίηση της διαδικασίας που προϋποθέτει τη συμβολή των ενδιαφερομένων μερών⁶⁹⁷.

Σε περιπτώσεις κατά τις οποίες μία τυπική διαδικασία αναδύει πολυπληθείς διαφωνίες, μία εναλλακτική λύση δύναται να περιλαμβάνει τη συλλογή πρόσθετων δεδομένων από τυχαιοποιημένες δοκιμές προκειμένου να δοκιμαστούν οι αντίστοιχες αιτιολογικές υποθέσεις. Η Google αποτελεί ένα χαρακτηριστικό παράδειγμα για την πραγματοποίηση δοκιμών για όλες τις πτυχές του ανθρώπινου δυναμικού της, οι οποίες εμπεριείχαν ακόμα και ζητήματα πραγματοποίησης του βέλτιστου αριθμού συνεντεύξεων ανά υποψήφιο⁶⁹⁸. Εάν οι συζητήσεις, οι δοκιμές και η πειθώ των διοικητικών στελεχών των οργανισμών δεν οδηγούν σε μία λογική συναίνεση σχετικά με το αιτιολογικό μοντέλο που δημιουργεί το αποτέλεσμα του ενδιαφέροντος, οι αναλύσεις της τεχνητής νοημοσύνης είναι πιθανό να καταστούν αντιπαραγωγικές. Σε

⁶⁹⁴ Βλ. J. Pearl, *όπ.παρ.*, σελ. 23-29.

⁶⁹⁵ Βλ. J. Pearl, *όπ.παρ.*, σελ. 34-38.

⁶⁹⁶ Βλ. J. Pearl, *όπ.παρ.*, σελ. 34-38.

⁶⁹⁷ Βλ. J. Pearl, *όπ.παρ.*, σελ. 34-38.

⁶⁹⁸ Βλ. L. Bock, *Work Rules! Insights from Inside Google That Will Transform How You Live and Lead*, Hachette Book Group, 2015, σελ. 37-41.

αυτήν την περίπτωση θα πρέπει να αποφεύγονται, έως ότου ακολουθήσει η συλλογή περισσότερων ή ποιοτικότερων δεδομένων⁶⁹⁹.

Αξίζει να επισημανθεί ότι ένα θετικό στοιχείο της αξιοποίησης και συνεργασίας των οργανισμών με πολλούς διαφορετικούς προμηθευτές, το οποίο αναφέρθηκε ανωτέρω ως προβληματικό λόγω της μη συμβατότητας των λογισμικών ανάλυσης δεδομένων, είναι η ικανότητά τους να συνδυάζουν δεδομένα από πολλούς διαφορετικούς οργανισμούς για να δημιουργήσουν τους αλγορίθμους τους⁷⁰⁰. Ανάλογες προσεγγίσεις έχουν ήδη χρησιμοποιηθεί με τυπικές δοκιμές κατά το στάδιο της διαδικασίας προσλήψεων ή για την ανάθεση ρόλων προώθησης πωλήσεων. Χαρακτηριστικό παράδειγμα είναι η εταιρεία ADP, η οποία διαχειρίζεται ως εξωτερικός συνεργάτης ζητήματα μισθοδοσίας πολλών διαφορετικών εταιρειών. Η ADP κατάφερε να αξιοποιήσει την εν λόγω κλίμακα δυνατοτήτων για να δημιουργήσει προγνωστικά μοντέλα αποζημιώσεων και απολύσεων. Οι εταιρείες –πελάτες, μάλιστα, προθυμοποιούνται να διαθέσουν τα δεδομένα τους για την ολοκλήρωση του εγχειρήματος αυτού, με αντάλλαγμα την πρόσβαση που θα αποκτήσουν σε ζητήματα συγκριτικής αξιολόγησης του ανθρώπινου δυναμικού τους⁷⁰¹.

Σε γενικές γραμμές και σύμφωνα με τα ανωτέρω αναφερόμενα, τονίζεται ότι η επιπλοκή των μεμονωμένων οργανισμών, ωστόσο, έγκειται στο ότι πρέπει να γνωρίζουν σε ποιο βαθμό το πλαίσιό τους είναι επαρκώς διακριτό, ώστε ένας αλγόριθμος που βασίζεται σε δεδομένα άλλων οργανισμών, να είναι δυνατόν να εξάγει εύστοχες προβλέψεις και για τους ίδιους. Όπως θα συζητηθεί εκτενέστερα στις επόμενες ενότητες, τέτοιου είδους σημεία είναι απαραίτητα για την αντιμετώπιση νομικών προβλημάτων.

⁶⁹⁹ Βλ. L. Bock, *όπ.παρ.*, 37-41.

⁷⁰⁰ Βλ. L. Bock, *όπ.παρ.*, 46-58.

⁷⁰¹ Βλ. L. Bock, *όπ.παρ.*, 46-58.

4.4.3.1.1 Αντιδράσεις των εργαζομένων στις προσπάθειες συλλογής δεδομένων

Το γεγονός ότι οι εργαζόμενοι μπορούν να προσαρμόσουν τις απαντήσεις τους και τα δεδομένα που θα παράσχουν στον οργανισμό απασχόλησής τους ανάλογα με το πώς θεωρούν ότι αυτά πρόκειται να χρησιμοποιηθούν, αποτελεί σημαντικό μέλημα για όλα τα εγχειρήματα ανάλυσης δεδομένων ανθρώπινου δυναμικού. Λόγω της ευάλωτης αυτής περίπτωσης επικρατεί ένα κλίμα ανησυχίας, ως προς το ζήτημα της ανεύρεσης εναλλακτικών πηγών δεδομένων, τα οποία θα ήταν δυνατό να θεωρηθούν ως περισσότερο αυθεντικά⁷⁰². Πολλοί οργανισμοί, άλλωστε, προβαίνουν στην συλλογή πληροφοριών από τα μέσα κοινωνικής δικτύωσης, τις οποίες αξιοποιούν σε ζητήματα συνυφασμένα με τις διαδικασίες προσλήψεων. Επί παραδείγματι, αναζητούν αποδεικτικά στοιχεία κακής ή ανάρμοστης συμπεριφοράς των υποψήφιων εργαζομένων ή αποδεικτικά στοιχεία καταλληλότητάς τους για την πρόσληψή τους στον συγκεκριμένο οργανισμό και τη στελέχωση συγκεκριμένης θέσης εργασίας.

Προς την ίδια κατεύθυνση, τα τραπεζικά ιδρύματα διαθέτουν ακόμη αυστηρότερους κανονισμούς αναφορικά με την επίβλεψη του απασχολούμενου ανθρώπινου δυναμικού τους και έχουν προβεί, εδώ και καιρό, στην ανάλυση δεδομένων ηλεκτρονικού ταχυδρομείου για στοιχεία σχετικά με υπεξαίρέσεις, καθώς και για τον εντοπισμό εν δυνάμει ύποπτων όρων, όπως για παράδειγμα, η «παρενόχληση» στην επισκεψιμότητα μέσω ηλεκτρονικού ταχυδρομείου⁷⁰³ μπορεί να αποτελέσει αφορμή διεξαγωγής εσωτερικής έρευνας για τον εντοπισμό προβλημάτων στο χώρο εργασίας. Σε περίπτωση κατά την οποία αυτό καταστεί αντιληπτό από το ανθρώπινο δυναμικό, τότε οι εργαζόμενοι θα δίνουν μεγαλύτερη προσοχή στις λέξεις και τους όρους που χρησιμοποιούν στα μηνύματα ηλεκτρονικού ταχυδρομείου, είτε για να προάγουν τη διεξαγωγή περαιτέρω ερευνών είτε για να την αποτρέψουν⁷⁰⁴.

⁷⁰² Βλ. H. Tian, *όπ.παρ.*, σελ. 162-163, Βλ. επίσης W. Niehueser & G. Boak, *όπ.παρ.*, σελ. 121-130.

⁷⁰³ Βλ. X. Qiao, Design of Abnormal Behavior Detection System in the State Grid Business Office, In Sun X., Zhang X., Xia Z., Bertino E. (Eds.) *Advances in Artificial Intelligence and Security*, Springer, 2021, σελ. 510-520, Βλ. επίσης M. Fethi & F. Pasiouras, Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey, *European Journal of Operational Research*, Τεύχος 204, Αρ. 2, 2010, σελ. 189–198.

⁷⁰⁴ Βλ. M. Fethi & F. Pasiouras, *όπ.παρ.*, σελ. 189–198.

Η εταιρεία Vibe⁷⁰⁵ αποτελεί ένα χαρακτηριστικό παράδειγμα προμηθευτή, η οποία χρησιμοποιεί εργαλεία επεξεργασίας φυσικής γλώσσας για τη μέτρηση της έντασης και δυναμικότητας των σχολίων που δημοσιεύουν οι εργαζόμενοι στους εσωτερικούς πίνακες συνομιλίας τους, βοηθώντας με τον τρόπο αυτόν στην πρόβλεψη του κινδύνου μαζικής αποχώρησης ανθρώπινου δυναμικού. Τέτοιου είδους εφαρμογές, ωστόσο, έρχονται αντιμέτωπες με ορισμένες βασικές προκλήσεις, σε περίπτωση που εφαρμοστούν στο χώρο εργασίας. Πιο συγκεκριμένα, όταν οι υπάλληλοι συνειδητοποιήσουν ότι οι θέσεις τους χρησιμοποιούνται για τη λήψη μέτρων μαζικών απολύσεων, αυτό μπορεί να ασκήσει επιρροή στο τι και πώς επιλέγουν να δημοσιεύουν στις εσωτερικές τους συνομιλίες. Κατόπιν, εντοπίζονται ακόμα και ζητήματα που μπορεί να προκύψουν αναφορικά με το αν οι απασχολούμενοι θεωρούν ότι η χρήση αυτών των δεδομένων παραβιάζει το απόρρητό τους⁷⁰⁶.

Μάλιστα, αρκετοί οργανισμοί αναφέρουν ότι οι καλύτεροι προγνωστικοί παράγοντες σε ζητήματα ανθρώπινου δυναμικού δεν προέρχονται από παραδοσιακά ευρήματα βασισμένα στην ψυχολογία, αλλά από πηγές δεδομένων, όπως είναι τα μέσα κοινωνικής δικτύωσης. Τόσο αυτές όσο και παρεμφερείς εκτιμήσεις εμπίπτουν στο πεδίο προστασίας της ιδιωτικής σφαίρας, η οποία αποκτά νέες διαστάσεις στην ψηφιακή εποχή και είναι συναρτημένη με ζητήματα όπως η επαναπροσαρμογή δεδομένων και η διασπορά τους⁷⁰⁷.

Τα δεδομένα μπορούν να διατηρηθούν πολύ περισσότερο από τον αρχικό χρόνο που δημιουργήθηκαν, ακόμα και μετά την αξιοποίησή τους από έναν οργανισμό και οι οργανισμοί δύνανται να τα επεξεργαστούν και να τα συμβουλευτούν ακόμα και για σκοπούς που δεν έχουν κατ' αρχήν προβλεφθεί από τον δημιουργό τους. Για παράδειγμα, οι λέξεις που χρησιμοποιούνται κατά την ανταλλαγή μηνυμάτων ηλεκτρονικού ταχυδρομείου μεταξύ δύο ή περισσότερων συναδέλφων, είναι δυνατόν να αξιοποιηθούν για την πρόβλεψη του κινδύνου απόλυσης⁷⁰⁸. Επίσης, δύναται να

⁷⁰⁵ Βλ. X. Qiao, *όπ.παρ.*, σελ. 510-520.

⁷⁰⁶ Βλ. X. Qiao, *όπ.παρ.*, σελ. 510-520.

⁷⁰⁷ Βλ. C. Tucker, *Privacy, Algorithms, and Artificial Intelligence*, In Ajay A., Joshua G., and Avi G. (Eds.) *The Economics of Artificial Intelligence: An Agenda*, University of Chicago Press, 2017, σελ. 19-35.

⁷⁰⁸ Βλ. C. Tucker, *όπ.παρ.*, σελ. 19-35.

ασκήσει ακούσια επιρροή ακόμα και σε άλλα άτομα, όπως επί παραδείγματι στους φίλους των δημιουργών των δεδομένων που έχουν επισημανθεί (tag) σε αναρτήσεις και φωτογραφίες τους στα μέσα κοινωνικής δικτύωσης. Πρέπει να αναφερθεί, ότι οι οργανισμοί είναι απαραίτητο να μην παραβλέπουν το υφιστάμενο νομοθετικό πλαίσιο επί ζητημάτων προστασίας της ιδιωτικής σφαίρας, όπως τον Γενικό Κανονισμό για την Προστασία Δεδομένων Προσωπικού Χαρακτήρα (GDPR)⁷⁰⁹.

Όσον αφορά τις τεχνολογικές λύσεις στο ζήτημα της ιδιωτικότητας των δεδομένων, οι ειδήμονες επιδιώκουν την κατασκευή αλγορίθμων που να αντικατοπτρίζουν μεθόδους ανάλυσης δεδομένων που βασίζονται στην έννοια της διαφοροποιημένης ιδιωτικότητας⁷¹⁰. Στο πλαίσιο αυτό, τα δεδομένα «τυχαιοποιούνται» κατά τη διαδικασία συλλογής τους με αποτέλεσμα να μην αντλείται κανένα δεδομένο για ένα συγκεκριμένο άτομο, αλλά να αντλούνται μαζικά δεδομένα για μία πληθυσμιακή ομάδα⁷¹¹. Ακολούθως, οι αναλυτές δεδομένων δεν γνωρίζουν ποιου δεδομένα χρησιμοποιούνται στη διαδικασία της ανάλυσης.

⁷⁰⁹ Βλ. C. Tucker, *όπ.παρ.*, σελ. 19-35.

⁷¹⁰ Βλ. C. Tucker, *όπ.παρ.*, σελ. 19-35.

⁷¹¹ Βλ. C. Tucker, *όπ.παρ.*, σελ. 19-35.

4.4.3.2 Στάδιο μηχανικής μάθησης

Δεν προκαλεί έκπληξη το γεγονός ότι ένας αλγόριθμος μηχανικής μάθησης αποδίδει πολύ πιο ικανοποιητικά αποτελέσματα, συγκριτικά με οτιδήποτε εναλλακτικό έχει αξιοποιηθεί στο παρελθόν, αναφορικά με την πρόβλεψη του ποιοι υποψήφιοι εργαζόμενοι μπορούν να προσληφθούν και να στελεχώσουν το ανθρώπινο δυναμικό σε έναν οργανισμό. Πράγματι, η προηγούμενη έρευνα που έχει διεξαχθεί γύρω από το πεδίο της διαχείρισης ανθρωπίνων πόρων δεν έχει προσφέρει σημαντική βοήθεια στους οργανισμούς, από τη στιγμή που οι περισσότεροι προγνωστικοί παράγοντες, όπως είναι τα αποτελέσματα των τεστ προσωπικότητας και νοημοσύνης, προβλέπουν εξαιρετικά περιορισμένες εργασιακές επιδόσεις για τους εργαζόμενους⁷¹². Αυτή η έλλειψη δημιουργεί μία μεγάλη ευκαιρία για την ανάδυση της ανάλυσης δεδομένων και της υιοθέτησης της τεχνητής νοημοσύνης στο πεδίο του ανθρώπινου δυναμικού.

Ωστόσο, η εύρεση ποιοτικών δεδομένων για τη δημιουργία αλγορίθμων χαρακτηρίζεται από ιδιαίτερες συνθήκες. Μία κοινή προσέγγιση εδράζεται στη δημιουργία ενός αλγορίθμου βασιζόμενου στα χαρακτηριστικά των «καλύτερων» εργασιακών επιδόσεων και στη συνέχεια ακολουθεί η αξιολόγηση των αιτούντων εργασία. Για παράδειγμα, ο προμηθευτής HireVue βοηθά τους οργανισμούς-πελάτες του να πραγματοποιούν συνεντεύξεις υποψηφίων εργαζομένων μέσω βίντεο. Μέρος της προσφοράς του περιλαμβάνει αλγόριθμους που αναλύουν στοιχεία της μη λεκτικής επικοινωνίας ενός υποψηφίου που καταγράφονται στο βίντεο, όπως είναι ένα χαμόγελο ή μία συγκεκριμένη έκφραση προσώπου, για να προβλέψουν τη μελλοντική απόδοση του υποψηφίου εργαζομένου στον οργανισμό⁷¹³. Οι εν λόγω αλγόριθμοι «εκπαιδεύονται» σε δεδομένα που πηγάζουν από άλλους κορυφαίους και επιτυχημένους οργανισμούς, ομοειδείς του οργανισμού που τους χρησιμοποιεί για την αξιολόγηση των υποψηφίων εργαζομένων του.

⁷¹² Βλ. C. Fan, P. Fan, T. Chan & S. Chang, Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals, *Expert Systems with Applications*, Τεύχος 39, Αρ. 10, 2012, σελ. 8844–8851.

⁷¹³ Βλ. A. Páez, Negligent Algorithmic Discrimination, *Law and Contemporary Problems*, Τεύχος 84, Αρ. 2, 2021, σελ. 1-18.

Επιπλέον, η αυτοεπιλογή μπορεί να προκαλέσει αναληθείς συσχετίσεις μεταξύ των χαρακτηριστικών των εργαζομένων. Πρόκειται για το λεγόμενο φαινόμενο της σύγκρουσης στην επιστήμη των δεδομένων. Πολλοί οργανισμοί προκαλούν ακούσια τέτοιου είδους αποτελέσματα, όταν δημιουργούν προφίλ «καλύτερων εργασιακών επιδόσεων» από αναλύσεις επιδόσεων των μισθωτών τους και εν συνεχεία εφαρμόζουν το προφίλ αυτό σε υποψηφίους εργαζομένους τους στο στάδιο των προσλήψεων⁷¹⁴. Η ανάλυση, συνήθως, δεν λαμβάνει υπόψη το ότι οι εν λόγω εργαζόμενοι έχουν πιθανώς επιλεγεί από το σύνολο των υποψηφίων με βάση τους ίδιους και άλλους καθοριστικούς παράγοντες που εμφανίζουν θετική συσχέτιση με την εργασιακή αποδοτικότητα. Επιπροσθέτως, η ικανότητα του μοντέλου να «συνεχίζει να μαθαίνει» και να προσαρμόζεται σε νέες πληροφορίες, εκλείπει σε περιπτώσεις που η ροή νέων προσλήψεων περιορίζεται από τις προβλέψεις του τρέχοντος αλγορίθμου⁷¹⁵.

Η αυτοεπιλογή των εργαζομένων στο σύνολο των αιτούντων εργασία ενός οργανισμού δημιουργεί, επίσης, μία προκατειλημμένη εικόνα της προσφοράς εργασίας, ειδικά εάν ο οργανισμός διαθέτει μία ξεχωριστή μάρκα (brand) ή ταυτότητα, ανεξάρτητα από το αν αυτή είναι θετική ή αρνητική. Ως απόρροια αυτού, ο οργανισμός είναι δυνατόν να καταλήξει, λανθασμένα, στο συμπέρασμα ότι υπάρχει έλλειψη εργατικού δυναμικού και αφθονία προσφοράς εργασίας σε άλλες θέσεις. Αυτό μπορεί να προκαλέσει μία άσκοπη χαλάρωση ή, αντίστοιχα, ενίσχυση των κριτηρίων πρόσληψης⁷¹⁶.

Διάφορες πτυχές της διαδικασίας της μοντελοποίησης είναι επίσης δυνατόν να αποτελέσουν σημαντική πρόκληση για τους οργανισμούς και τη διαχείριση ανθρώπινου δυναμικού μέσω αξιοποίησης της τεχνητής νοημοσύνης. Για παράδειγμα, μπορεί να υπάρχουν περισσότερα από ένα μέτρο «προσαρμογής» με τα διαθέσιμα δεδομένα. Σε περίπτωση που εξετάζεται η διαφορά μεταξύ των πληθυσμών της πλειοψηφίας, όπως είναι επί παραδείγματι οι λευκοί υπάλληλοι, και των πληθυσμών της μειοψηφίας, όπως είναι επί παραδείγματι οι αфроαμερικανοί υπάλληλοι, οι αλγόριθμοι που μεγιστοποιούν την προγνωστική επιτυχία για τον πληθυσμό στο σύνολό του, είναι πιθανό να προχωρήσουν σε διακρίσεις έναντι της προγνωστικής

⁷¹⁴ Βλ. J. Pearl, *όπ.παρ.*, σελ. 34-38.

⁷¹⁵ Βλ. J. Pearl, *όπ.παρ.*, σελ. 36-37.

⁷¹⁶ Βλ. J. Pearl, *όπ.παρ.*, σελ. 42-46.

επιτυχίας για τον μειονοτικό πληθυσμό. Η δημιουργία ξεχωριστών αλγορίθμων για τον κάθε πληθυσμό, η οποία θα ήταν δυνατό να οδηγήσει σε πιο δίκαια αποτελέσματα, τίθεται σε σύγκρουση με τα νομικά και ηθικά πρότυπα της διαφορετικής μεταχείρισης. Ως εκ τούτου, υπάρχουν θεμελιώδεις αντισταθμίσεις σε αυτόν τον τομέα μεταξύ ακρίβειας και δικαιοσύνης που πρέπει να αντιμετωπιστούν σε οποιαδήποτε εφαρμογή μηχανικής μάθησης ανθρώπινου δυναμικού⁷¹⁷.

Παρά ταύτα, υπάρχουν πολλοί τρόποι καθορισμού δίκαιων προτύπων στη διαδικασία της λήψης αποφάσεων. Παραδείγματος χάριν, ένα πρότυπο θα μπορούσε να περιλαμβάνει τον αποκλεισμό της μεταβλητής της φυλής από τον αλγόριθμο, ενώ ένα άλλο θα μπορούσε να επικεντρώνεται μόνο στη μεγιστοποίηση της πιθανότητας εντοπισμού εργαζομένων, ανεξαρτήτως φυλής⁷¹⁸. Καταδεικνύεται, λοιπόν, ότι ακόμη και κατά την ανάπτυξη ενός αλγορίθμου μπορεί να είναι δύσκολη η σύγκλιση σε βασικά ζητήματα που έχουν θεμελιώδεις επιπτώσεις στο σχεδιασμό του.

Η κατανόηση τέτοιου είδους ζητημάτων έχει μέγιστη σημασία και δεν πρέπει να αγνοούνται ή να παραβλέπονται. Η άγνοια επί των εν λόγω θεμάτων δύναται να μην οδηγήσει σε σοβαρές συνέπειες σε τομείς όπως είναι το μάρκετινγκ, όπου η εσφαλμένη στόχευση διαφημίσεων θα έχει ενδεχομένως οικονομικές μα όχι ανθρώπινες επιπτώσεις⁷¹⁹, αλλά μπορεί να καταστεί πολύ σοβαρή και δαπανηρή σε ζητήματα που συνάδουν με τους ανθρώπινους πόρους, στο πλαίσιο των οποίων τυπικές αποφάσεις, όπως η πρόσληψη, η αμοιβή ή η προαγωγή, έχουν σημαντικές οργανωσιακές και νομικές συνέπειες⁷²⁰. Η παρουσία και καταχώριση του κατάλληλου όγκου και δειγμάτων δεδομένων στο στάδιο της Δημιουργίας Δεδομένων αποτελεί μόνο ένα βήμα προς την αποφυγή της προκατειλημμένης μηχανικής μάθησης⁷²¹.

⁷¹⁷ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, Can Data Science Change Human Resources, In Canals J., Heukamp F. (Eds.) The Future of Management in an AI World, Palgrave Macmillan, IESE Business Collection, 2020, σελ. 93-115.

⁷¹⁸ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

⁷¹⁹ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

⁷²⁰ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

⁷²¹ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

Η δεύτερη και πιο ολοκληρωμένη λύση στην πρόκληση αυτή είναι η λεγόμενη τυχαία ανακάλυψη. Αναλυτικότερα, οι αλγόριθμοι υπολογιστών τυχαίας ανακάλυψης αναπτύσσονται ενεργά και η ερμηνεία τους δεν απαιτεί προηγμένη εκπαίδευση⁷²². Με τη σειρά τους, οι αλγόριθμοι μηχανικής μάθησης με αιτιώδη γνώση επιτρέπουν την ελαχιστοποίηση ή την εξάλειψη της αιτιώδους εξάρτησης από παράγοντες που δεν ελέγχονται από ένα άτομο, όπως η καταγωγή τους ή η εθνικότητά τους⁷²³.

⁷²² Βλ. D. Malinsky & D. Danks, Causal discovery algorithms: A practical guide, *Philosophy Compass*, Τεύχος 13, Αρ. 1, 2017, σελ. 1-11.

⁷²³ Βλ. D. Malinsky & D. Danks, *όπ.παρ.*, σελ. 1-11.

4.4.3.3 Στάδιο λήψης αποφάσεων

Υπάρχουν τρεις κύριες προκλήσεις σε περιπτώσεις όπου οι υπεύθυνοι λήψης αποφάσεων προσπαθούν να εφαρμόσουν τις προβλέψεις που δημιουργούνται από τη μηχανική μάθηση. Η πρώτη αφορά τη δικαιοσύνη και τα νομικά ζητήματα, η δεύτερη αφορά την έλλειψη επεξηγήσεων του αλγορίθμου και η τρίτη αναφέρεται στο πώς θα αντιδράσουν οι εργαζόμενοι στις αλγοριθμικές αποφάσεις. Στην παρούσα ενότητα αναλύεται ξεχωριστά καθεμία από τις προκλήσεις αυτές.

Δικαιοσύνη

Στο πλαίσιο του ανθρώπινου δυναμικού επικρατούν αρκετά ζητήματα συναρτημένα με την αρχή της δικαιοσύνης. Ένα εξ αυτών είναι η αναγνώριση ότι κάθε αλγόριθμος είναι πιθανό να λαμβάνει υπόψη του δεδομένα, τα οποία έχει προγραμματιστεί να αξιολογεί κατά τη στιγμή της δημιουργίας του. Αναλυτικότερα, η αξιοποίηση υφιστάμενων δεδομένων, που εμπεριέχουν διακρίσεις, για το σχεδιασμό ενός αλγορίθμου πρόσληψης είναι πιθανό να έχει ως αποτέλεσμα τη δημιουργία ενός μοντέλου, το οποίο είναι δυνατόν να οδηγηθεί σε δυσανάλογες επιλογές λευκών και ανδρών εις βάρος των «έγχρωμων» και γυναικών υπαλλήλων. Η χρήση αυτών των αλγορίθμων οδηγεί στον κίνδυνο αναπαραγωγής δημογραφικής ποικιλομορφίας ή ακόμα και της απουσίας αυτής, η οποία παρατηρείται στα ιστορικά δεδομένα⁷²⁴.

Στον τομέα του ανθρώπινου δυναμικού παρατηρείται μία ευρέως διαδεδομένη πεποίθηση ότι οι αξιολογήσεις των υποψηφίων και των εργαζομένων διαμορφώνονται σε μεγάλο βαθμό από τις προκαταλήψεις του αξιολογητή, οι οποίες είναι συνηθέστερες αναφορικά με τα δημογραφικά στοιχεία των ανθρώπων. Οι αλγόριθμοι έχουν τη δυνατότητα να περιορίσουν την προκατάληψη αυτήν, μέσω της τυποποίησης της εφαρμογής κριτηρίων στα αποτελέσματα που θα εξάγουν και μέσω της αφαίρεσης πληροφοριών που δεν θεωρούνται σχετικές με την εργασιακή αποδοτικότητα και οι οποίες υπάρχει πιθανότητα να ασκήσουν επιρροή στις αποφάσεις πρόσληψης εργαζομένων, όπως είναι η εθνικότητα και το φύλο των υποψηφίων. Παράγοντες που μπορεί να είναι σημαντικοί για την προγνωστική δύναμη του αλγορίθμου, μπορεί παρ' όλα αυτά να θεωρηθούν ακατάλληλοι, όπως είναι για παράδειγμα η κοινωνική κατάσταση ενός ατόμου⁷²⁵.

Η πιθανότητα νομικών προκλήσεων που σχετίζονται ιδιαίτερα με τις δυσμενείς επιπτώσεις των αποφάσεων για την απασχόληση εγείρει, επίσης, ποικίλες ανησυχίες. Η λήψη ελεύθερων αποφάσεων από τα διοικητικά και διευθυντικά στελέχη των οργανισμών για ζητήματα που συνάδουν με την απασχόληση του ανθρώπινου δυναμικού, χωρίς σχετική καθοδήγηση, θα οδηγήσει πιθανότατα στην υιοθέτηση και

⁷²⁴ Βλ. C. Liu & J. Denrell, Performance persistence through the lens of chance models, *Academy of Management Proceedings*, Ap. 1, σελ. 1-4.

⁷²⁵ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

παρουσία εντονότερων προκαταλήψεων από αυτές που θα υπήρχαν και θα δημιουργούνταν από τους αλγόριθμους⁷²⁶. Όμως, αυτού του είδους οι προκαταλήψεις είναι πολύ πιο δύσκολο να κατηγορηθούν για λανθασμένες επιλογές ανθρώπινου δυναμικού, λόγω του ότι είναι ασυστηματικές και περιορίζονται σε συγκεκριμένες ομάδες εργασίας⁷²⁷. Στον αντίποδα, οι αλγόριθμοι που χρησιμοποιούνται σε έναν οργανισμό για αντίστοιχα ζητήματα, δύνανται να χαρακτηρίζονται από μικρότερα επίπεδα προκαταλήψεων, με τη διαδικασία ανίχνευσής τους να καθίσταται πιο εύκολη, οδηγώντας σε πιθανές προτεινόμενες ενέργειες για δράση. Ένα ερώτημα που προκύπτει από την υιοθέτηση και αξιοποίηση συγκεκριμένων αλγορίθμων είναι το αν οι φορείς λήψης αποφάσεων αξίζει να αναλάβουν αυτόν τον νομικό κίνδυνο προκειμένου να μειώσουν τη συνολική προκατάληψη⁷²⁸.

Ακόμη και σε περίπτωση που ένας αλγόριθμος αντιμετωπίζει τις ανησυχίες που παρατηρούνται σχετικά με παρουσία διακρίσεων σε προστατευόμενες ομάδες ανθρώπινου δυναμικού, δεν εκλείπουν οι ανησυχίες σχετικά με την απονομή δικαιοσύνης⁷²⁹. Υποθεθίσθω ότι με δεδομένες τις απαιτήσεις που συνοδεύονται από μια συγκεκριμένη θέση εργασίας, ένας αλγόριθμος καθόρισε τους δύο καλύτερους υποψηφίους για την κάλυψη της εν λόγω θέσης με 80% και 90% αντίστοιχα. Ένα ερώτημα που γεννάται αφορά το αν η προκύπτουσα διαφορά του 10% είναι μεγάλη ή μικρή, ως προς την ακρίβεια του αλγορίθμου. Ο διευθυντής προσλήψεων έχει τη διακριτική ευχέρεια να αναθεωρήσει όλες τις διαθέσιμες πληροφορίες και να κρίνει εμπειρικά επ' αυτού⁷³⁰.

Ένας τρόπος για τον μετριασμό τέτοιων ζητημάτων είναι η εισαγωγή τυχαίων παραλλαγών, οι οποίες αποτελούν έναν μη αναγνωρισμένο, αλλά σημαντικό μηχανισμό διαχείρισης⁷³¹. Η έρευνα δείχνει ότι οι εργαζόμενοι αντιλαμβάνονται την τυχαία ακολουθούμενη διαδικασία ως δίκαιη για τον καθορισμό σύνθετων και,

⁷²⁶ Βλ. C. Liu & J. Denrell, *όπ.παρ.*, σελ. 1-4.

⁷²⁷ Βλ. C. Liu & J. Denrell, *όπ.παρ.*, σελ. 1-4.

⁷²⁸ Βλ. C. Liu & J. Denrell, *όπ.παρ.*, σελ. 1-4.

⁷²⁹ Βλ. J. Denrell, C. Fang & C. Liu, *όπ.παρ.*, σελ. 923-940.

⁷³⁰ Βλ. J. Denrell, C. Fang & C. Liu, *όπ.παρ.*, σελ. 923-940.

⁷³¹ Βλ. J. Denrell, C. Fang & C. Liu, *όπ.παρ.*, σελ. 923-940.

συνεπώς, αβέβαιων αποτελεσμάτων⁷³². Οπότε, σε περίπτωση κατά την οποία ο αλγόριθμος είναι αξιόπιστος και οι δύο υποψήφιοι είναι «καλοί» για τη στελέχωση της συγκεκριμένης θέσης εργασίας, τότε είναι πιο θεμιτό να αποδεχτούμε την πολυπλοκότητα των αποφάσεων πρόσληψης και να προβούμε σε μία τυχαία επιλογή μεταξύ των δύο υποψηφίων, ενδεχομένως, με τις πιθανότητες επιλογής που είναι ανάλογες των αντίστοιχων ποσοστιαίων βαθμολογιών. Επομένως, η τυχαιοποίηση θα πρέπει να αποτελεί ένα εργαλείο διαχείρισης της τεχνητής νοημοσύνης στο στάδιο της λήψης αποφάσεων, αλλά και σε προηγούμενα στάδια του κύκλου ζωής της⁷³³.

Ωστόσο, η χρήση αλγορίθμων για την επίπληξη των απασχολούμενων σε έναν οργανισμό εγείρει ακόμα περισσότερα ερωτήματα⁷³⁴, όπου ένα εξ αυτών, λόγου χάρη, μπορεί να είναι το αν ένας αλγόριθμος που προβλέπει μελλοντικές συνεισφορές των απασχολούμενων προορίζεται να αξιοποιηθεί κάποια στιγμή για να λαμβάνει αποφάσεις απολύσεων, ή για προβλέψεις διενέργειας υποκλοπών εις βάρος του οργανισμού από έναν ήδη απασχολούμενό του.

Στο σημείο αυτό, εντοπίζεται ένα δίλημμα που αφορά στο εξής: Η ωφελμιστική έννοια της προώθησης του συλλογικού αγαθού θα ήταν δυνατόν να υποστηρίξει την αξιοποίηση προγνωστικών αλγορίθμων για την εξάλειψη προβλημάτων που συνδέονται με τη διαχείριση ανθρώπινου δυναμικού ή για τον περιορισμό των δαπανών ενός οργανισμού για τους ανθρώπινους πόρους του⁷³⁵; Η καντιανή δεοντολογική θέση, απεναντίας, υποστηρίζει ότι οι άνθρωποι πρέπει να κρίνονται με βάση τις δικές τους ενέργειες και ότι υπό το πρίσμα αυτό, η ανωτέρω αναφερόμενη τοποθέτηση δεν είναι αποδεκτή⁷³⁶.

Όσον αφορά την έννοια της δικαιοσύνης, αυτή είναι συναρτημένη ακόμα και με την αποδοχή που επικρατεί στη συμμετοχή στη λήψη αποφάσεων με τους περισσότερους εργαζόμενους. Η έρευνα αποδεικνύει σε ολοένα και εντονότερο βαθμό ότι οι

⁷³² Βλ. E. Lind & K. Van den Bos, When Fairness Works: Toward a General Theory of Uncertainty Management, *Research in Organizational Behavior*, Ap. 24, 2002, σελ. 181-223.

⁷³³ Βλ. E. Lind & K. Van den Bos, *όπ.παρ.*, σελ. 181-223.

⁷³⁴ Βλ. E. Lind & K. Van den Bos, *όπ.παρ.*, σελ. 181-223.

⁷³⁵ Βλ. E. Lind & K. Van den Bos, *όπ.παρ.*, σελ. 181-223.

⁷³⁶ Βλ. E. Lind & K. Van den Bos, *όπ.παρ.*, σελ. 181-223.

αλγόριθμοι είναι πιο αποδοτικοί συγκριτικά με την ανθρώπινη κρίση, σε περιπτώσεις που αξιοποιούνται για την πρόβλεψη επαναλαμβανόμενων αποτελεσμάτων, καθώς επίσης και για την πρόβλεψη αποτελεσμάτων εργασιακής αποδοτικότητας ή εντοπισμού των βέλτιστων περιπτώσεων υποψηφίων εργαζομένων⁷³⁷. Όμως, ένα ερώτημα που εγείρεται στο σημείο αυτό αφορά το σε περίπτωση που οι αλγόριθμοι αναλάβουν τη διαδικασία των προσλήψεων και οι αρμόδιοι επί τέτοιου είδους ζητημάτων δεν διαδραματίσουν κάποιον ρόλο σε αυτήν, αν οι δεύτεροι θα είναι εξίσου αφοσιωμένοι στις νέες προσλήψεις, σαν να είχαν πάρει οι ίδιοι την απόφαση πρόσληψης, συγκριτικά με την περίοδο που όντως αναλάμβαναν την ευθύνη αυτή.

⁷³⁷ Βλ. B. Cowgill, Bias and Productivity in Humans and Algorithms-Theory and Evidence from Résumé Screening, 2017, στο https://conference.iza.org/conference_files/MacroEcon_2017/cowgill_b8981.pdf

Έλλειψη εξηγήσεων

Με τα μοντέλα μηχανικής μάθησης, η ικανότητα να εξηγηθούν ποιοι είναι οι παράγοντες που χρησιμοποιούνται από έναν αλγόριθμο, προκειμένου αυτός να καταλήξει στη διενέργεια προβλέψεων, καθίσταται γενικά δυσκολότερη συγκριτικά με την περίπτωση των παραδοσιακών στατιστικών μοντέλων⁷³⁸. Ένας απλός κανόνας λήψης αποφάσεων βάσει της παλαιότητας των εργαζομένων σε έναν οργανισμό είναι εύκολα κατανοητός και διαφαίνεται αντικειμενικός ακόμα και αν τα αποτελέσματα που ενδεχομένως επιφέρει στο άμεσο ή μακροπρόθεσμο μέλλον δεν είναι αρεστά⁷³⁹. Ένας αλγόριθμος μηχανικής μάθησης που βασίζεται σε σταθμισμένο συνδυασμό συγκεκριμένων προσδιοριστικών παραγόντων και μεταβλητών που εμφανίζουν συσχέτιση με την εργασιακή αποδοτικότητα, μπορεί να καταστεί εξίσου αντικειμενικός με ένα παραδοσιακό στατιστικό μοντέλο. Ωστόσο, η πρώτη αναφερόμενη περίπτωση είναι δυσκολότερα κατανοητή, ειδικά όταν οι εργαζόμενοι προβαίνουν σε αναπόφευκτες συγκρίσεις μεταξύ τους. Οι αλγόριθμοι χαρακτηρίζονται από υψηλότερα επίπεδα ακρίβειας όσο υψηλότερα είναι τα επίπεδα πολυπλοκότητάς τους⁷⁴⁰, όμως, παράλληλα, αυτό τους καθιστά περισσότερο δυσνόητους και εξηγήσιμους.

⁷³⁸ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

⁷³⁹ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

⁷⁴⁰ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

Αντιδράσεις εργαζομένων στις αλγοριθμικές αποφάσεις

Οι αλλαγές στη διαδικασία της λήψης αποφάσεων σε έναν οργανισμό ασκούν αναπόφευκτη επιρροή στη συμπεριφορά του απασχολούμενου ανθρώπινου δυναμικού και ως εκ τούτου, μπορεί να δικαιολογηθεί η καταβολή αλλεπάλληλων προσπαθειών ανάπτυξης βέλτιστων κανόνων λήψης αποφάσεων⁷⁴¹. Τονίζεται, ωστόσο, ότι από την πλευρά των απασχολούμενων της πρώτης γραμμής και των προϊσταμένων τους, ενδεχομένως να προκληθούν αντιδράσεις αν λαμβάνονται αποφάσεις στις οποίες δεν συμμετέχει ο ανθρώπινος παράγοντας, δηλαδή τα στελέχη που απασχολούνται σε ιεραρχικά ανώτερες θέσεις από αυτούς⁷⁴².

Επί παραδείγματι, είναι ευρέως αποδεκτό ότι η σχέση ενός απασχολούμενου με τον προϊστάμενο ή διευθυντή του είναι ζωτικής σημασίας για την εργασιακή του αποδοτικότητα και ότι η ποιότητα αυτής της σχέσης εξαρτάται από την κοινωνική ανταλλαγή. Η κοινωνική ανταλλαγή, στην προκειμένη περίπτωση αφορά στο ότι ο προϊστάμενος ή διευθυντής «φροντίζει» τον υφιστάμενό του και ο υφιστάμενος ως «αντάλλαγμα» επιτελεί αποδοτικά και αποτελεσματικά τα εργασιακά του καθήκοντα προς ευχαρίστηση του προϊσταμένου ή διευθυντή του⁷⁴³. Ακόμη και όταν οι εργαζόμενοι έχουν χαμηλά επίπεδα οργανωσιακής πιστότητας και δέσμευσης, μπορεί να αισθάνονται δέσμευση απέναντι στον προϊστάμενο ή τον διευθυντή τους, ανεξάρτητα από το αν η δέσμευση αυτή κατευθύνεται και προς τον ευρύτερο οργανισμό⁷⁴⁴. Το ερώτημα που εγείρεται στην περίπτωση αυτή έγκειται στο πώς επηρεάζεται αυτή η κοινωνική ανταλλαγή, όταν οι αποφάσεις που είχαν ληφθεί από τον προϊστάμενο ή τον διευθυντή, πλέον λαμβάνονται από έναν αλγόριθμο.

Όταν το εργασιακό πρόγραμμα των απασχολούμενων δημιουργείται και καθορίζεται από ένα λογισμικό, ενδεχομένως θα επηρεαστεί η καλή θέληση και πίστη των εργαζομένων απέναντι είτε στον ευρύτερο οργανισμό είτε άμεσα στον προϊστάμενο ή

⁷⁴¹ Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

⁷⁴² Βλ. P. Cappelli, P. Tambe, & V. Yakubovich, *όπ.παρ.*, σελ. 93-115.

⁷⁴³ Βλ. B. Dietvorst, J. Simmons & C. Massey, Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them, *Management Science*, Τεύχος 64, Αρ. 3, 2016, σελ. 1155-1170.

⁷⁴⁴ Βλ. B. Dietvorst, J. Simmons & C. Massey, *όπ.παρ.*, σελ. 1155-1170.

διευθυντή τους⁷⁴⁵. Άλλωστε, όταν οι αποφάσεις επί ζητημάτων διαχείρισης και διαμόρφωσης του εργασιακού προγράμματος λαμβάνονται από έναν λογισμικό, αυτομάτως, οι εργαζόμενοι χάνουν το δικαίωμα παραπόνων αν το πρόγραμμα αυτό δεν τους εξυπηρετεί ή σε περίπτωση που χρειάζονται μία ευελιξία ή προσαρμογή. Σε γενικές γραμμές, έχει αποδειχτεί ότι οι άνθρωποι ανταποκρίνονται διαφορετικά σε αποφάσεις που λαμβάνονται από αλγόριθμους, εν συγκρίσει με αποφάσεις που λαμβάνονται από ανθρώπους⁷⁴⁶. Ειδικότερα δε, στην περίπτωση της διοίκησης ανθρώπινου δυναμικού, αν ληφθεί η απόφαση για μισθολογική αύξηση, αυτομάτως χτίζεται μία καλή σχέση μεταξύ υφιστάμενου και προϊσταμένου ή διευθυντή, σε περίπτωση που ο δεύτερος έχει μεσολαβήσει στη λήψη αυτής της απόφασης⁷⁴⁷, κατάσταση που δεν θα ισχύει στην περίπτωση όπου η εν λόγω απόφαση είναι απόρροια ενός αλγόριθμου.

Ωστόσο, μπορεί να υπάρχουν περιπτώσεις όπου η διαδικασία της λήψης αποφάσεων διευκολύνεται και οι αποφάσεις γίνονται άμεσα αποδεκτές όταν λαμβάνονται από έναν αλγόριθμο, παρά όταν λαμβάνονται από έναν άνθρωπο⁷⁴⁸. Αυτό ισχύει ειδικότερα σε περιπτώσεις που αυτές οι αποφάσεις έχουν αρνητικές συνέπειες για τους ανθρώπους. Πειραματικά δεδομένα υποδηλώνουν ότι η προθυμία αποδοχής και χρήσης αλγόριθμων εξαρτάται εν μέρει από τον τρόπο με τον οποίο οι αλγόριθμοι ενημερώνονται για την αντιμετώπιση των ενδεχόμενων λαθών και παραλήψεων⁷⁴⁹.

⁷⁴⁵ Βλ. B. Dietvorst, J. Simmons & C. Massey, *όπ.παρ.*, σελ. 1155-1170.

⁷⁴⁶ Βλ. B. Dietvorst, J. Simmons & C. Massey, *όπ.παρ.*, σελ. 1155-1170.

⁷⁴⁷ Βλ. B. Dietvorst, J. Simmons & C. Massey, *όπ.παρ.*, σελ. 1155-1170.

⁷⁴⁸ Βλ. B. Dietvorst, J. Simmons & C. Massey, *όπ.παρ.*, σελ. 1155-1170.

⁷⁴⁹ Βλ. B. Dietvorst, J. Simmons & C. Massey, *όπ.παρ.*, σελ. 1155-1170.

4.4.4 Οι σύγχρονες παρεχόμενες υπηρεσίες προς τους πολίτες

Είτε πρόκειται για εξυπηρέτηση των πολιτών επί ζητημάτων παροχής υπηρεσιών υγείας και περίθαλψης είτε πρόκειται για οποιαδήποτε άλλη ανάγκη εξυπηρέτησής τους από τις δημόσιες υπηρεσίες, οι πολίτες καλούνται συχνά να έρθουν σε επαφή με το ανθρώπινο δυναμικό που στελεχώνει τους δημόσιους οργανισμούς, προκειμένου να εξυπηρετηθούν⁷⁵⁰. Στον 20ό αιώνα υπήρξε έντονη αλληλεπίδραση μεταξύ πολιτών και δημοσίων υπηρεσιών ή κυβερνητικών φορέων, ενώ είναι γεγονός ότι η εξέλιξη που έχει γνωρίσει η τεχνολογία και ο κλάδος των πληροφοριακών συστημάτων κατά τη διάρκεια του 21^{ου} αιώνα μπορεί να εξυπηρετήσει σε ακόμα μεγαλύτερο βαθμό την κάλυψη των αναγκών των πολιτών από τους δημόσιους οργανισμούς⁷⁵¹.

Ωστόσο, η έλλειψη νεοτερισμού, όταν συνδυάζεται με τις γενικότερες τάσεις πολιτικής δυσπιστίας είναι δυνατό να επιφέρει μειωμένα επίπεδα ικανοποίησης των πολιτών από την εξυπηρέτησή τους μέσω των δημόσιων οργανισμών⁷⁵². Πιο συγκεκριμένα, το ευρύτερο κοινό, στην πλειοψηφία των περιπτώσεων, καταλήγει να είναι λιγότερο ικανοποιημένο από τις δημόσιες υπηρεσίες, σε αντίθεση με τις υπηρεσίες που προσφέρονται από νομικές οντότητες που δραστηριοποιούνται στον ιδιωτικό τομέα⁷⁵³. Συνεπώς, εύλογα, οι δημόσιες υπηρεσίες καταλαμβάνουν τις τελευταίες θέσεις ικανοποίησης των πολιτών, διότι καλύπτουν ανεπαρκώς τις ανάγκες τους, τόσο από θέμα ποιότητας όσο και από θέμα αμεσότητας εξυπηρέτησής τους. Βέβαια, η χαμηλή εμπιστοσύνη των πολιτών προς τις παρεχόμενες δημόσιες υπηρεσίες, είναι άμεσα

⁷⁵⁰ Βλ. H. Mehr, *όπ.παρ.*, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁷⁵¹ Βλ. H. Mehr, *όπ.παρ.*, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁷⁵² Βλ. M. Bovens, & A Wille, Deciphering the Dutch drop: ten explanations for decreasing political trust in The Netherlands, *International Review of Administrative Sciences*, Τεύχος 74, Αρ.2, 2008, σελ. 283–305.

⁷⁵³ Βλ. M. Wisniewski, Using SERVQUAL to assess customer satisfaction with public sector services, *Managing Service Quality: An International Journal*, Τεύχος 11, Αρ.6, 2001, σελ. 380–388.

συνυφασμένη με τα χαμηλά επίπεδα εμπιστοσύνης τους προς τις εκάστοτε αρμόδιες κυβερνήσεις⁷⁵⁴.

Ο τομέας της τεχνητής νοημοσύνης μπορεί να προσφέρει σημαντικές ευκαιρίες στις δημόσιες υπηρεσίες και τους κυβερνητικούς φορείς, αν αξιοποιηθεί με ορθολογικό και χρηστό τρόπο⁷⁵⁵. Μέσω της αξιοποίησης της τεχνητής νοημοσύνης, οι κυβερνητικοί φορείς και οι δημόσιοι οργανισμοί έχουν τη δυνατότητα να προχωρήσουν σε προβλέψεις των τάσεων της εποχής, που δημιουργούν νέες ανάγκες στους πολίτες. Επιπλέον, η τεχνητή νοημοσύνη μπορεί να βοηθήσει τους δημόσιους και κυβερνητικούς οργανισμούς να προβούν σε επανασχεδιασμό και αναβάθμιση των παλαιών τεχνολογικών συστημάτων τους, τα οποία πλέον θεωρούνται παρωχημένα⁷⁵⁶.

Ενώ μία πληθώρα συστημικών παραγόντων έχει προκαλέσει τα χαμηλά επίπεδα ικανοποίησης των πολιτών από τους δημόσιους οργανισμούς, όπως προαναφέρθηκε, δεν μπορεί να παραβλεφθεί η απαρχαιωμένη στελέχωση των υπηρεσιών του δημοσίου τομέα ως προς τον διαθέσιμο υλικοτεχνικό εξοπλισμό⁷⁵⁷. Για τον λόγο αυτόν, η εν λόγω ανεπάρκεια υλικοτεχνικών πόρων θεωρείται ένας επιπρόσθετος επιβαρυντικός παράγοντας για τα ήδη χαμηλά επίπεδα εμπιστοσύνης που επιδεικνύει το κοινό στους δημόσιους οργανισμούς.

Δεν είναι άλλωστε τυχαίο ότι στην περίπτωση των ΗΠΑ, από τα 90 δισεκατομμύρια δολάρια που δαπανώνται από την πλευρά της ομοσπονδιακής κυβέρνησης υπέρ του τομέα της τεχνολογίας, η πλειοψηφία του 75% προορίζεται στην συντήρηση και επισκευή των ξεπερασμένων υλικοτεχνολογικών συστημάτων⁷⁵⁸. Τα εν λόγω απαρχαιωμένα τεχνολογικά συστήματα είναι απόλυτα συνδεδεμένα με την πρόκληση καθυστερήσεων και διοικητικής φύσεως επιβαρύνσεων. Αυτό επιβεβαιώνεται και από το ότι, στην περίπτωση των ΗΠΑ, ένα ποσοστό της τάξεως του 53% των δημοσίων

⁷⁵⁴ Βλ. G. Duncan, Trust, distrust, and the end of politics-as-we-knew-it: the mood of the nation prior to election 2017, Kōtuitui, New Zealand Journal of Social Sciences Online, 2017, σελ. 1–18.

⁷⁵⁵ Βλ. K. Desouza, G. Dawson, & D. Chenok, D., Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector, Business Horizons, σελ. 1-9.

⁷⁵⁶ Βλ. K. Desouza, G. Dawson, & D. Chenok, D., όπ.παρ., σελ. 1-9.

⁷⁵⁷ Βλ. K. Desouza, G. Dawson, & D. Chenok, D., όπ.παρ., σελ. 1-9.

⁷⁵⁸ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

υπαλλήλων που συμμετείχαν σε συγκεκριμένη έρευνα⁷⁵⁹, δήλωσαν ότι έρχονταν συχνά αντιμέτωποι με έντονη γραφειοκρατία, η οποία ασκούσε αρνητική επιρροή στη δυνατότητα και τη θέλησή τους να ολοκληρώσουν τις εργασιακές τους υποχρεώσεις προς εξυπηρέτηση των πολιτών.

Πέραν των ανωτέρω, η πληθυσμιακή γήρανση είναι ένας επιπλέον επιβαρυντικός παράγοντας αφενός για την ανάγκη παροχής ποιοτικών υπηρεσιών από την πλευρά των δημόσιων οργανισμών⁷⁶⁰, ιδίως στον τομέα υγειονομικής περίθαλψης, και αφετέρου στην επαρκή στελέχωση των δημοσίων υπηρεσιών με παραγωγικό ανθρώπινο δυναμικό. Το πρόβλημα είναι ήδη τεταμένο, αν αναλογιστεί κανείς ότι σε σχετική έρευνα που διεξήχθη στην Αμερική⁷⁶¹, σχεδόν το ένα τρίτο του συνόλου των δημοσίων υπαλλήλων που απασχολείται σε υπηρεσίες Κοινωνικής Ασφάλισης αναμένεται να συνταξιοδοτηθεί εντός του 2022.

Κλείνοντας, γενικότερη παραδοχή είναι το γεγονός ότι το ευρύτερο κοινό έχει ανάγκη την ψηφιοποίηση των δημοσίων υπηρεσιών και την εξυπηρέτηση από τεχνολογικά εκσυγχρονισμένους δημόσιους οργανισμούς, έτσι ώστε η κάλυψη των αναγκών τους από τον δημόσιο τομέα να είναι ποιοτικά αντίστοιχη με αυτήν που απολαμβάνουν από τον ιδιωτικό⁷⁶². Σύμφωνα με σχετική έρευνα που διεξήχθη από τον οργανισμό «Accenture»⁷⁶³, διαπιστώθηκε ότι μόλις το 27% των πολιτών της Αμερικής εμφάνισε αυξημένα επίπεδα ικανοποίησης από την υπάρχουσα προσφορά ψηφιοποιημένων υπηρεσιών από τις δημόσιες αρχές. Η συντριπτική πλειοψηφία του 92% των

⁷⁵⁹ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁷⁶⁰ Βλ. J. Oude Mulders, Attitudes About Working Beyond Normal Retirement Age: The Role of Mandatory Retirement, *Journal of Aging & Social Policy*, 2019, σελ. 1–17.

⁷⁶¹ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁷⁶² Βλ. K. Desouza, G. Dawson, & D. Chenok, D., όπ.παρ., σελ. 1-9.

⁷⁶³ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

Αμερικανών πολιτών επεσήμανε ότι «οι βελτιωμένες ψηφιακές υπηρεσίες θα επηρέαζαν θετικά την άποψή τους για τον δημόσιο τομέα»⁷⁶⁴.

Λαμβάνοντας υπόψη τα ανωτέρω, ο τομέας της τεχνητής νοημοσύνης έχει τη δυνατότητα να προσφέρει την ευκαιρία στους δημόσιους οργανισμούς και τις κυβερνητικές υπηρεσίες να προβούν σε ενεργές προβλέψεις των αναγκών των πολιτών και να προχωρήσουν σε αναβάθμιση των παλαιών και τεχνολογικά ξεπερασμένων υλικοτεχνικών συστημάτων τους.

⁷⁶⁴ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

4.4.5 Η συμβολή της τεχνητής νοημοσύνης στην παροχή των εν λόγω υπηρεσιών

Ενώ ο κλάδος της τεχνητής νοημοσύνης έχει κατά καιρούς εμφανιστεί μέσω διαφόρων τεχνολογικών εφαρμογών, ειδικότερα σε πιλοτική μορφή κατά τη διάρκεια των τελευταίων δεκαετιών, ωστόσο, η ουσιαστική ενσωμάτωσή του τόσο σε εικονικά όσο και σε φυσικά περιβάλλοντα είναι εντονότερη στην πιο πρόσφατη εποχή⁷⁶⁵. Πιο συγκεκριμένα, η πρόσφατη εξέλιξη που έχει γνωρίσει ο τομέας της τεχνητής νοημοσύνης τον έχει καταστήσει ως τον βασικό κινητήριο μοχλό, κατευθύνοντας το μέλλον της παγκόσμιας οικονομίας στην εποχή που υποκινείται από τα «μεγάλα δεδομένα»⁷⁶⁶. Δεν είναι άλλωστε τυχαίο ότι στην έρευνα που διεξήχθη από τον οργανισμό «Accenture», εκτιμήθηκε ότι ο τομέας της τεχνητής νοημοσύνης, μέχρι το έτος 2035⁷⁶⁷ θα έχει οδηγήσει σε διπλασιασμό τον ρυθμό της οικονομικής ανάπτυξης των κρατών σε παγκόσμια βάση.

Διευκρινίζεται, επιπροσθέτως, ότι οι εκάστοτε κυβερνήσεις ανά τον κόσμο προβαίνουν σε δοκιμαστική εφαρμογή συστημάτων τεχνητής νοημοσύνης⁷⁶⁸. Ένα χαρακτηριστικό παράδειγμα συστηματικής αξιοποίησης συστημάτων τεχνητής νοημοσύνης στις δημόσιες υπηρεσίες, προσβλέποντας στην εξυπηρέτηση των πολιτών, αφορά ζητήματα κάλυψης αποριών και ερωτημάτων τους. Πριν την αξιοποίηση συστημάτων τεχνητής νοημοσύνης, όταν οι πολίτες ήθελαν να υποβάλλουν ένα ερώτημα σε κάποια δημόσια υπηρεσία, συνήθως έπρεπε να περιμένουν αρκετή ώρα στην τηλεφωνική τους γραμμή, ή να προβούν σε επίσκεψη στα φυσικά καταστήματα των δημοσίων υπηρεσιών ή να αναζητήσουν στο διαδίκτυο την απάντηση στο ερώτημά τους. Μέσω της χρήσης συστημάτων τεχνητής νοημοσύνης, επέρχεται σημαντική βελτίωση στον βαθμό που οι

⁷⁶⁵ Βλ. T.Q. Sun, & R. Medaglia, Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare, *Government Information Quarterly*, 2018, σελ. 1-16.

⁷⁶⁶ Βλ. I. Pencheva, M. Esteve, & S. Mikhaylov, Big Data and AI – A transformational shift for government: So, what next for research, *Public Policy and Administration*, 2018, σελ. 1-21.

⁷⁶⁷ Βλ. H. Mehr, *όπ.παρ.*, στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁷⁶⁸ Βλ. M. Kuziemski, & G. Misuraca, AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings, *Telecommunications Policy*, 2020, σελ. 1-13.

πολίτες μπορούν να έχουν άμεση πρόσβαση στην κάλυψη των αποριών τους, σε πραγματικό χρόνο, από τους ίδιους τους δημόσιους υπαλλήλους⁷⁶⁹.

Επιπροσθέτως, η τεχνητή νοημοσύνη χρησιμοποιείται ακόμη και σε ζητήματα διαμόρφωσης και συμπλήρωσης δημοσίων εγγράφων, ειδικότερα σε περιπτώσεις διαδικασιών ρουτίνας⁷⁷⁰, με αποτέλεσμα την ταχύτερη και βέλτιστη εξυπηρέτηση των πολιτών εξαλείφοντας τους τεράστιους χρόνους αναμονής και την φυσική τους προσέλευση στους χώρους των δημοσίων υπηρεσιών.

Μολονότι οι εφαρμογές τεχνητής νοημοσύνης που χρησιμοποιούνται από τους δημόσιους οργανισμούς χαρακτηρίζονται από αισθητά χαμηλότερους ρυθμούς υιοθέτησης, συγκριτικά με τους αντίστοιχους ρυθμούς που παρατηρούνται στην περίπτωση των επιχειρήσεων που δραστηριοποιούνται στον ιδιωτικό τομέα, ωστόσο, οι πιθανές περιπτώσεις αξιοποίησης συστημάτων τεχνητής νοημοσύνης στον δημόσιο τομέα προσομοιάζουν με τις αντίστοιχες περιπτώσεις εφαρμογών που χρησιμοποιούνται ήδη στον ιδιωτικό τομέα. Συνεπώς, η τεχνητή νοημοσύνη καταλήγει να είναι λειτουργική μόνο υπό την προϋπόθεση της συνεχούς μάθησης των μηχανών⁷⁷¹.

Πιο συγκεκριμένα, σε περίπτωση κατά την οποία ο σκοπός των εφαρμογών της τεχνητής νοημοσύνης είναι η απλή αντικατάσταση των παρόντων διαδραστικών συστημάτων που χρησιμοποιούνται στον δημόσιο τομέα για την εξυπηρέτηση των πολιτών, μέσω αυτοματοποιημένων φωνητικών αποκρίσεων ή της αυτοματοποίησης υπολογιστικών εργασιών ρουτίνας, τότε δεν θα καταλήξουν να είναι τόσο αποτελεσματικές όσο οι εφαρμογές που βασίζονται στη διαρκή μάθηση⁷⁷². Ο ουσιαστικός θετικός αντίκτυπος της χρήσης των εφαρμογών τεχνητής νοημοσύνης στους δημόσιους οργανισμούς θα επέλθει μόνο σε περίπτωση κατά την οποία μειωθεί αισθητά ο διοικητικός φόρτος των δημοσίων υπηρεσιών και ταυτόχρονα, οδηγήσει σε βελτίωση της ανθρώπινης εμπειρίας, αλλά όχι σε πλήρη αντικατάσταση του ανθρώπινου παράγοντα.

⁷⁶⁹ Βλ. M. Kuziemski, & G. Misuraca, *όπ.παρ.*, σελ. 1-13.

⁷⁷⁰ Βλ. M. Kuziemski, & G. Misuraca, *όπ.παρ.*, σελ. 1-13.

⁷⁷¹ Βλ. A. Kankanhalli, Y. Charalabidis, Y. & S. Mellouli, *IoT and AI for Smart Government: A Research Agenda*, *Government Information Quarterly*, Τεύχος 36, Αρ. 2, 2019, σελ. 304–309.

⁷⁷² Βλ. A. Kankanhalli, Y. Charalabidis, Y. & S. Mellouli, *όπ.παρ.*, σελ. 304–309.

Η σταδιακή και στρατηγική πλήρης εφαρμογή τέτοιου είδους εφαρμογών τεχνητής νοημοσύνης θα είναι δυνατόν να επιφέρει εναργή προσφορά ποιοτικών και αποτελεσματικών υπηρεσιών προς τους πολίτες, τη στιγμή κατά την οποία, παράλληλα, θα οδηγούν σε δυνητική μείωση του κόστους και σε αύξηση της ικανοποίησης των πολιτών λόγω των αναβαθμισμένων παρεχόμενων δημοσίων υπηρεσιών⁷⁷³. Στο ίδιο πνεύμα κινείται και η διατύπωση της H.R. Gilman, σύμφωνα με την οποία «...οι εφαρμογές της τεχνητής νοημοσύνης θα μπορούσαν να επιτρέψουν πιο άμεση αμφίδρομη αλληλεπίδραση μεταξύ πολιτών και κράτους»⁷⁷⁴.

Παρακάτω, παρουσιάζονται ορισμένες εναλλακτικές επιλογές παροχής δημόσιων υπηρεσιών προς τους πολίτες, στις οποίες η πιλοτική σταδιακή εφαρμογή της τεχνητής νοημοσύνης έχει ήδη ξεκινήσει και αναμένεται να ενταθεί στο άμεσο μέλλον, με σκοπό τις πιο ολοκληρωμένες και ποιοτικά αποδοτικότερες υπηρεσίες. Πρόκειται για τις εξής περιπτώσεις δημοσίων υπηρεσιών:

Πίνακας 4. Περιπτώσεις δημοσίων υπηρεσιών που χρησιμοποιούν εφαρμογές τεχνητής νοημοσύνης για την εξυπηρέτηση των πολιτών

Υπηρεσίες	Παραδείγματα
Απαντήσεις σε ερωτήματα των πολιτών ⁷⁷⁵	Στην περίπτωση μίας δημόσιας υπηρεσίας που εδρεύει στη βόρεια Καρολίνα, χρησιμοποιούνται τα λεγόμενα «chatbots». Πρόκειται για κειμενικά ή ακουστικά ηλεκτρονικά συστήματα συνομιλίας, η βάση των οποίων εντοπίζεται στον τομέα της τεχνητής νοημοσύνης. Τα ηλεκτρονικά αυτά συστήματα εξυπηρετούν την

⁷⁷³ Βλ. K. Desouza, G. Dawson, & D. Chenok, D., όπ.παρ., σελ. 1-9.

⁷⁷⁴ Βλ. H. Mehr, όπ.παρ., στο https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf

⁷⁷⁵ Βλ. S. Goldsmith, Artificial Intelligence Will Help Create a More Responsive Government, 2017, στο <https://www.govtech.com/opinion/artificial-intelligence-will-help-create-a-more-responsive-government.html>, Βλ. επίσης D. Pereira, Watson helps cities help citizens, 2017, στο <https://medium.com/@darylp/watson-assists-cities-with-311-3d7d6898d132>

	ανθρώπινη εργασία, καθώς απελευθερώνουν τον χρόνο των δημοσίων υπαλλήλων σε ζητήματα χειρισμού και κατανομής των τηλεφωνικών γραμμών των πολιτών μέσω του κέντρου τηλεφωνικής βοήθειας. Ειδικότερα σε περιπτώσεις που τα θέματα που αντιμετωπίζουν οι πολίτες είναι καθημερινής ρουτίνας, όπως άλλωστε έχει αποδειχτεί με το 90% των χρηστών που καλούν το τηλεφωνικό κέντρο της εν λόγω υπηρεσίας να ζητούν απλή υποστήριξη στην επαναφορά του κωδικού πρόσβασής τους, η εξυπηρέτηση γίνεται άμεσα και αυτόματα από το chatbot, χωρίς να χρειαστεί η ανθρώπινη παρέμβαση.
	Η χρήση chatbot παρατηρείται και από τη συνεργασία που συνάφθηκε μεταξύ κυβερνητικού φορέα της Σιγκαπούρης με την εταιρεία Microsoft. Σκοπός της συνεργασίας ήταν η προμήθεια της κυβερνητικής αρχής με ειδικά σχεδιασμένο chatbot για την εξυπηρέτηση των πολιτών σε συγκεκριμένες υπηρεσίες. Μάλιστα, το chatbot λειτουργούσε ως ψηφιακός εκπρόσωπος των πολιτών προς τον εν λόγω κυβερνητικό φορέα.
	Η συνεργασία των δημοτικών αρχών της Νέας Υόρκης με την εταιρεία IBM αποσκοπεί στην αξιοποίηση εφαρμογών τεχνητής νοημοσύνης για την ανάπτυξη ενός προηγμένου συστήματος διαχείρισης των αναγκών και απαιτήσεων των πολιτών. Οι εφαρμογές επιτυγχάνουν ταχύτερη εξυπηρέτηση σε καθημερινά τους ερωτήματα σχετικά με τις παρεχόμενες

	υπηρεσίες από την πόλη, εξοικονομώντας σημαντικό χρόνο. Πρόκειται για παρόμοια συνεργασία με αυτήν που η εταιρεία IBM είχε συνάψει με την πόλη Surrey της Κολομβίας, όπου η IBM ανέπτυξε την εφαρμογή «MySurrey». Μέσω της εφαρμογής αυτής οι πολίτες είχαν τη δυνατότητα να λαμβάνουν απαντήσεις αυτόματα από το chatbot, στα αντίστοιχα ηλεκτρονικά τους ερωτήματα. Η επιτυχία της συγκεκριμένης εφαρμογής ήταν η κάλυψη του 65% των ερωτημάτων των πολιτών χωρίς να χρειαστεί η ανθρώπινη παρέμβαση.
Αναζήτηση και αυτόματα συμπλήρωση δημοσίων εγγράφων ⁷⁷⁶	Η ανάπτυξη μίας εφαρμογής τεχνητής νοημοσύνης, η οποία μπορεί να χρησιμοποιηθεί δωρεάν από τους αντίστοιχους επαγγελματίες, σκοπεύει στην παροχή βοήθειας προς τους πρόσφυγες που ζητούν άσυλο στην Αμερική. Οι επαγγελματίες απαντούν σε ερωτήματα των προσφύγων αναφορικά με τα διαδικαστικά θέματα που πρέπει να ακολουθήσουν, όπως είναι για παράδειγμα η συμπλήρωση συγκεκριμένων εγγράφων και αιτήσεων ή αν πληρούνται οι απαιτούμενες προϋποθέσεις για να τους χορηγηθεί άσυλο. Στη συνέχεια, το chatbot συμβάλλει στην αυτόματη συμπλήρωση του απαραίτητου εγγράφου και δίνει όσες επιπλέον πληροφορίες χρειάζεται να γνωρίζει ο αιτών

⁷⁷⁶Βλ. J. Ebsworth, S. Johns, & M. Dodson, Surveillance capitalism: the hidden costs of the digital revolution, 2021, στο: <https://evangelicalfocus.com/jubilee-centre/13158/surveillance-capitalism-the-hidden-costs-of-the-digital-revolution>, Βλ. επίσης P. Clayton, & J. Clopton, Business curriculum redesign: Integrating data analytics, Journal of Education for Business, 2018, σελ. 1–7.

	για τη συνέχεια. Μάλιστα, το chatbot είναι με τέτοιο τρόπο σχεδιασμένο ώστε να χρησιμοποιεί απλή γλώσσα, που να είναι κατανοητοί σε ανθρώπους που δεν είναι φυσικοί ομιλητές της αγγλικής.
	Μέσω εφαρμογών τεχνητής νοημοσύνης εξυπηρετείται ταχύτατα η αναζήτηση και κατηγοριοποίηση μεγάλου όγκου εικόνων και εγγράφων. Ένας αμερικανικός κυβερνητικός φορέας αξιοποιεί εφαρμογές τεχνητής νοημοσύνης για να παράσχει βοήθεια στους πολίτες, μέσω της αναζήτησης εγγράφων σε περισσότερες από ένα εκατομμύριο σελίδες κειμένου. Η χρήση της εν λόγω εφαρμογής καταλήγει στην εξοικονόμηση διοικητικών οικονομικών επιβαρύνσεων, τη στιγμή κατά την οποία ο εν λόγω φορέας θα έπρεπε να έχει δαπανήσει εκατομμύρια δολάρια για την πραγματοποίηση επαρκών αναβαθμίσεων.
	Στο δικηγορικό επάγγελμα, οι εφαρμογές τεχνητής νοημοσύνης είναι χρήσιμες για την ανεύρεση της σχετικής νομολογίας, καθώς επίσης και για ζητήματα σάρωσης νομικών εγγράφων.
Κατεύθυνση και εξυπηρέτηση των πολιτών στις αρμόδιες δημόσιες υπηρεσίες⁷⁷⁷	Η πιλοτική εφαρμογή και αξιοποίηση προγραμμάτων τεχνητής νοημοσύνης είχε ξεκινήσει από τη μεξικανική κυβέρνηση, με σκοπό την ταξινόμηση των αιτήσεων που αποστέλλονταν από τους πολίτες, ώστε να δρομολογηθούν προς τη σωστή και πλέον

⁷⁷⁷Βλ. P. Clayton, & J. Clopton, όπ.παρ., σελ. 1–7.

	αρμόδια υπηρεσία ή στο σωστό τμήμα εντός μίας συγκεκριμένης δημόσιας υπηρεσίας.
	Η δημόσια υπηρεσία «Allstate Insurance» έχει συνάψει συνεργασία με την εταιρεία Amazon, με σκοπό την αξιοποίηση του chatbot «Alexa». Το chatbot παρέχει απαντήσεις σε ερωτήματα των πολιτών σχετικά με το ασφαλιστικό τους πρόγραμμα και την εξόφληση των ασφαλιστηρίων συμβολαίων τους. Μάλιστα, το chatbot κατευθύνει τους πολίτες, ανάλογα με το αίτημά τους, στο αρμόδιο τμήμα της ασφαλιστικής υπηρεσίας, ώστε να εξυπηρετηθούν από τον εκάστοτε αρμόδιο εκπρόσωπο.
Υπηρεσίες μεταφράσεων δημοσίων εγγράφων⁷⁷⁸	Μία πληθώρα δημοσίων υπηρεσιών, πλέον, χρησιμοποιώντας συστήματα τεχνητής νοημοσύνης παρέχει ποιοτικές υπηρεσίες μετάφρασης δημοσίων εγγράφων, τα οποία περιλαμβάνουν βασικές κρατικές πληροφορίες. Χαρακτηριστικό παράδειγμα είναι οι υπηρεσίες διεξαγωγής των Χειμερινών Ολυμπιακών Αγώνων της PyeongChang, από τις οποίες γινόταν χρήση μεταφραστικών υπηρεσιών που παρέχονταν προς τους ενδιαφερόμενους σε πραγματικό χρόνο. Οι υπηρεσίες αυτές υποστηρίχθηκαν από εφαρμογές που είναι βασισμένες στην τεχνητή νοημοσύνη.

⁷⁷⁸Βλ. D. Castro, & J. New, The Promise of Artificial Intelligence, 2016, στο: <https://datainnovation.org/2016/10/the-promise-of-artificial-intelligence/>

	Η μετάφραση των δραστηριοτήτων του οργανισμού Unbabel διενεργείται μέσω της αξιοποίησης της μηχανικής μάθησης σε συνδυασμό με το crowdsourcing. Η μετάφραση γίνεται σε συνολικά δεκατέσσερις διαφορετικές γλώσσες, μέσω χρήσης ειδικά διαμορφωμένου αλγορίθμου για τη μετάφραση μηνυμάτων ηλεκτρονικού ταχυδρομείου και ιστοσελίδων που αποσκοπούν στην εξυπηρέτηση του κοινού. Μάλιστα, ο εν λόγω οργανισμός έχει στελεχώσει εξειδικευμένο ανθρώπινο δυναμικό, για λόγους ελέγχου ποιότητας και ακρίβειας των μεταφρασμένων κειμένων, προκειμένου να αποφευχθούν λάθη και παρερμηνείες. Ωστόσο, οι υπηρεσίες εξακρίβωσης της πιστότητας των μεταφρασμένων κειμένων παρέχονται κατόπιν σχετικού αιτήματος και κατόπιν καταβολής 0,02 δολαρίων ανά συνομιλία.
Υπηρεσίες σύνταξης ολοκληρωμένων γραπτών μηνυμάτων και δημοσίων εγγράφων προς τους πολίτες ⁷⁷⁹	Από το Υπουργείο Οικονομίας, Εμπορίου και Βιομηχανίας της Ιαπωνίας έχει ξεκινήσει η πιλοτική εφαρμογή ενός συστήματος τεχνητής νοημοσύνης, με σκοπό την παροχή βοήθειας προς τους πολίτες, σε ερωτήματα που θέτουν στα γραφεία των μελών του κοινοβουλίου. Η

⁷⁷⁹Βλ. ICF, Colorado Child Welfare County Workload Study, 2014, στο: https://leg.colorado.gov/sites/default/files/1354s_highlights.pdf, Βλ. επίσης A. Baig, A. Dua, & V. Riefberg, Putting citizens first, McKinsey Center for Government, 2014, στο: <https://www.mckinsey.com/~media/mckinsey/industries/public%20and%20social%20sector/our%20insights/how%20us%20state%20governments%20can%20improve%20customer%20service/putting%20citizens%20first%20how%20to%20improve%20citizens%20experience%20and%20satisfaction%20with%20government%20services.pdf>

	σύνταξη απαντήσεων στα ερωτήματα των πολιτών γίνεται αυτόματα μέσω chatbot.
	Η σύνταξη δημοσίων εγγράφων αποτελεί μία ακόμη δυνατότητα που δίνεται μέσω της αξιοποίησης εφαρμογών τεχνητής νοημοσύνης. Η εφαρμογή τεχνητής νοημοσύνης «Natural Language Generation», ήδη αξιοποιείται σε πληθώρα ειδησεογραφικών σελίδων, όπως είναι για παράδειγμα το Associated Press και το Bloomberg. Η εφαρμογή αυτή χρησιμοποιείται με σκοπό την άντληση δεδομένων για την επιτυχή σύνταξη κειμένων, σε ταχύτατους ρυθμούς που μπορούν να φτάσουν μέχρι και τα 2.000 άρθρα ανά δευτερόλεπτο.

Ωστόσο, πέραν από τις ανωτέρω αναφερόμενες υπηρεσίες, οι οποίες ήδη εφαρμόζονται σε εμβρυϊκό στάδιο σε διάφορες δημόσιες αρχές σε διεθνές πλαίσιο, το μέλλον της αξιοποίησης της τεχνητής νοημοσύνης στη δημόσια διοίκηση είναι μακρύ και πολυδιάστατο⁷⁸⁰. Η αξιοποίηση των εφαρμογών τεχνητής νοημοσύνης από τις δημόσιες υπηρεσίες δεν μπορεί να μείνει προσκολλημένη σε υπηρεσίες μεταφράσεων δημοσίων εγγράφων, συμπλήρωσης δημοσίων αιτήσεων και αυτοματοποιημένων απαντήσεων στα ερωτήματα των πολιτών ή στην καθοδήγησή τους στο αρμόδιο τμήμα της υπηρεσίας που τους ενδιαφέρει για να εξυπηρετηθούν από τον κατάλληλο εκπρόσωπο.

Οι εφαρμογές τεχνητής νοημοσύνης μπορούν να αποδειχθούν χρήσιμες ακόμα και σε περιπτώσεις εξυπηρέτησης των πολιτών σε καταστάσεις έκτακτης ανάγκης, σε

⁷⁸⁰ Βλ. C. Wang, T.S. Teo, & M. Janssen, Public and private value creation using artificial intelligence: An empirical study of AI voice robot users in Chinese public sector, International Journal of Information Management, Ap. 61, 2021, σελ. 1-15

ζητήματα παροχής εξειδικευμένης και χαμηλού κόστους εκπαίδευσης, σε θέματα εντοπισμού κρουσμάτων δολιοφθοράς και απάτης, σε ζητήματα αναφοράς εγκληματικών ενεργειών, διενέργειας προβλέψεων για λόγους πρόληψης των απαιτούμενων παρεμβατικών κινήσεων, καθώς επίσης και για ζητήματα ενημέρωσης των πολιτών για κατασκευαστικές και επισκευαστικές εργασίες στις δημόσιες υποδομές⁷⁸¹. Επίσης, μπορούν να αξιοποιηθούν από τις δημόσιες υπηρεσίες για την αύξηση των επιπέδων ασφαλείας στο πλαίσιο της πρόληψης και αντιμετώπισης του ηλεκτρονικού εγκλήματος, μέσω της επιτυχούς πρόβλεψης για ενδεχόμενες κυβερνοεπιθέσεις ή για περιπτώσεις υποκλοπών προσωπικών και ευαίσθητων δεδομένων.

Μία ιδιαιτέρως σημαντική πρόκληση για τις κυβερνήσεις ανά τον κόσμο αποτελεί η πλήρης ενσωμάτωση στην κοινωνική και οικονομική ζωή των ατόμων με αναπηρίες, έτσι ώστε να αισθάνονται ότι έχουν την δυνατότητα να διεκδικούν ισότιμα τις ευκαιρίες που θα παρουσιάζονται στη ζωή τους. Η ανάπτυξη εφαρμογών μηχανικής μάθησης και τεχνητής νοημοσύνης, προσανατολισμένες και στην εξυπηρέτηση συνανθρώπων μας με αναπηρίες, είναι βέβαιο ότι θα επιλύσουν σε μέγιστο βαθμό πολλά από τα καθημερινά προβλήματα των ανθρώπων αυτών.

Για παράδειγμα, οι συνάνθρωποί μας, οι οποίοι βιώνουν απώλεια της όρασης, το σημαντικότερο ζήτημα που αντιμετωπίζουν στην καθημερινότητά τους και το οποίο τους περιορίζει ή και αποκλείει από την κοινωνική και οικονομική ζωή, αποτελεί η μετακίνησή τους. Η ανάπτυξη και υλοποίηση εφαρμογών με ικανότητα ανάλυσης βίντεο με ταυτόχρονη φωνητική ενημέρωση σε πραγματικό χρόνο του ευρύτερου χώρου μετακίνησης, κρίνονται ζωτικής σημασίας. Οι συγκεκριμένες εφαρμογές θα μπορούσαν να εδράζονται στην τεχνολογία που χρησιμοποιείται για την ανάπτυξη των αυτόνομων οχημάτων, αξιοποιώντας κάμερες, αισθητήρες τρισδιάστατης σάρωσης του περιβάλλοντος και υπολογισμού απόστασης τυχόν κινητών ή στατικών εμποδίων, καθώς και συστήματα γεωεντοπισμού, έτσι ώστε να καθίσταται, όσο το δυνατόν, ασφαλής η μετακίνηση και μέσω της τεχνολογίας τεχνητής νοημοσύνης να ενισχυθεί η αυτοπεποίθηση, η αυτονομία, η αξιοπρέπεια και το αίσθημα της ισότητας στους

⁷⁸¹ Βλ. C. Wang, T.S. Teo, & M. Janssen, όπ.παρ., σελ. 1-15.

συνανθρώπους μας, οι οποίοι, καθημερινά, έρχονται αντιμέτωποι με απρόβλεπτους παράγοντες, προκειμένου να πραγματοποιήσουν μία μετακίνησή τους.

Ήδη, καταγράφονται κάποιες προσπάθειες ανάπτυξης εφαρμογών, βασισμένες στην ανάλυση εικόνας, οι οποίες δύνανται να επεξεργαστούν στατικές εικόνες και να ενημερώσουν τον χρήστη τους ως προς το περιεχόμενό τους. Ωστόσο, οι εν λόγω εφαρμογές χρήζουν σημαντικών βελτιώσεων για να καταστούν έγκυρες και αξιόπιστες.

Μία ακόμη προσπάθεια των δημοσίων αρχών για τη διεύρυνση του πεδίου εφαρμογής των συστημάτων τεχνητής νοημοσύνης εκδηλώνεται με τη δημιουργία των «έξυπνων» πόλεων (Smart Cities)⁷⁸². Αποτελεί ένα εγχείρημα ιδιαίτερας ελπιδοφόρο, το οποίο, όμως, θα πρέπει να σχεδιαστεί λαμβάνοντας υπόψη πολύπλοκες και πολύπλευρες παραμέτρους, όπως λόγου χάρη, την βελτίωση της καθημερινότητας των ατόμων με αναπηρίες και την επακόλουθη συμμετοχή τους στη κοινωνική ζωή χωρίς περιορισμούς.

Επειδή η ανάπτυξη και υλοποίηση των εφαρμογών μηχανικής μάθησης και τεχνητής νοημοσύνης πραγματοποιείται από το αρχικό μέχρι το τελικό στάδιό τους, κυρίως, από εταιρείες που δραστηριοποιούνται στον ιδιωτικό τομέα, οι κυβερνήσεις, στο πλαίσιο ενός σύγχρονου δημοκρατικού και κοινωνικού κράτους, το οποίο σέβεται την ισότητα και προστατεύει τις ανθρώπινες αρχές και αξίες, θα πρέπει να λάβουν τα κατάλληλα μέτρα, προκειμένου οι εφαρμογές υψηλού κόστους, οι οποίες θα προορίζονται για την βελτίωση της καθημερινότητας των ατόμων με αναπηρίες, να παρέχονται στο πλαίσιο στοχευμένων προνοιακών οικονομικών παροχών.

⁷⁸²Βλ. E. Cortés-Cediel, I. Cantador, & O. Gil, όπ.παρ., στο: https://www.researchgate.net/publication/319043437_Recommender_systems_for_e-governance_in_smart_cities_State_of_the_art_and_research_opportunities

4.4.6 Οι σύγχρονες ανάγκες των κυβερνήσεων για την τεχνολογική τους ανέλιξη

Συστημικοί προσδιοριστικοί παράγοντες επαληθεύουν την ύπαρξη ποικίλων και σημαντικών περιθωρίων στους δημόσιους και κυβερνητικούς φορείς για την περαιτέρω τεχνολογική τους ανέλιξη, αξιοποιώντας σε μεγαλύτερο βαθμό εφαρμογές που βασίζονται στην τεχνητή νοημοσύνη⁷⁸³, καθιστώντας την πραγματοποίηση τεχνολογικών επενδύσεων αναγκαία και αναπόφευκτη⁷⁸⁴. Συνεπώς, οι δημόσιες υπηρεσίες θα επιτύχουν να έρθουν σε ευθυγράμμιση με τις σύγχρονες τάσεις των πολιτών και με τον τρόπο με τον οποίο οι ίδιοι προτιμούν να έρχονται σε επαφή με τους δημοσίους υπαλλήλους και να εξυπηρετούνται από τους δημόσιους οργανισμούς.

Οι κυβερνητικοί φορείς μπορούν να εντείνουν την αξιοποίηση εφαρμογών τεχνητής νοημοσύνης, παίρνοντας ως παράδειγμα τον τρόπο που αυτή πραγματοποιήθηκε από την πλευρά των νομικών οντοτήτων που δραστηριοποιούνται στον ιδιωτικό τομέα. Για την υλοποίηση του συγκεκριμένου εγχειρήματος, προτείνονται συνολικά έξι στρατηγικές, μέσω της αξιοποίησης των οποίων οι κυβερνητικοί φορείς μπορούν να ωφεληθούν για την περαιτέρω ενσωμάτωση συστημάτων τεχνητής νοημοσύνης στην καθημερινή τους λειτουργία⁷⁸⁵.

Αρχικά, οι κυβερνητικοί φορείς πρέπει να υιοθετήσουν πελατοκεντρικό χαρακτήρα και στον κεντρικό άξονα της λειτουργίας τους να τεθεί η βελτιωμένη εξυπηρέτηση των πολιτών και η ικανοποίηση των αναγκών τους⁷⁸⁶. Παράλληλα, θα πρέπει να επιδιώκουν και να παροτρύνουν την ενεργό συμμετοχή των πολιτών στις διάφορες δράσεις τους. Επίσης, προσπαθώντας να αναβαθμίσουν τον υλικοτεχνικό εξοπλισμό τους, οι κυβερνητικοί φορείς προτείνεται να συνεχίζουν, σε ένα βαθμό, να βασίζονται στους πόρους που ήδη διαθέτουν, συμπεριλαμβανομένου του ήδη υφιστάμενου ανθρώπινου

⁷⁸³ Βλ. D. Susar, & V. Aquaro, Artificial Intelligence. Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance, 2019, σελ. 418-426

⁷⁸⁴ Βλ. D. Susar, & V. Aquaro, όπ.παρ., σελ. 418-426.

⁷⁸⁵ Βλ. A. Alshahrani, D. Dennehy, & M. Mäntymäki, An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia, Government Information Quarterly, 2021, σελ. 1-14.

⁷⁸⁶ Βλ. N. Aoki, An experimental study of public trust in AI chatbots in the public sector, Government Information Quarterly, Τεύχος 37, Αρ. 4, 2020, σελ. 1-10.

δυναμικού τους. Άλλωστε, όπως ήδη έχει επισημανθεί, σκοπός της τεχνητής νοημοσύνης δεν είναι να αντικαταστήσει τον ανθρώπινο παράγοντα αλλά να διευκολύνει την εργασία του. Σε αυτή την λογική βασίζεται η τέταρτη στρατηγική, η οποία αφορά την αύξηση του απασχολούμενου ανθρώπινου δυναμικού στις δημόσιες υπηρεσίες, ώστε να είναι εφικτή η βέλτιστη εξυπηρέτηση των πολιτών σε όσο το δυνατόν μικρότερο χρονικό διάστημα, λαμβάνοντας συνδυαστικά τη βοήθεια των συστημάτων τεχνητής νοημοσύνης⁷⁸⁷.

Η πέμπτη στρατηγική αφορά στην επαρκή προετοιμασία των δημοσίων υπηρεσιών επί ζητημάτων ασφαλείας και προστασίας των προσωπικών δεδομένων των πολιτών. Αν μη τι άλλο, οι δημόσιες υπηρεσίες είναι υποχρεωμένες να ακολουθούν αυστηρά τα πρωτόκολλα περί απορρήτου. Τέλος, καλούνται να περιορίσουν τον ηθικό κίνδυνο που απορρέει από τη λήψη αποφάσεων εξ ολοκλήρου από μηχανές τεχνητής νοημοσύνης⁷⁸⁸, ζήτημα το οποίο αναλύθηκε διεξοδικά στα προηγούμενα κεφάλαια της παρούσας μελέτης.

Στον αντίποδα, κρίνεται αναγκαίο να επισημανθούν κάποιες αδυναμίες που περιβάλλουν τους κυβερνητικούς φορείς, οι οποίες δύνανται να επηρεάσουν αρνητικά την επιτυχή αξιοποίηση των συστημάτων τεχνητής νοημοσύνης.

Ίσως, από τα πιο κρίσιμα σημεία που δυσχεραίνουν την επιτυχή αξιοποίηση των συστημάτων τεχνητής νοημοσύνης είναι η ανεπαρκής κατανομή των διαθέσιμων πόρων των δημοσίων υπηρεσιών. Η ανεπάρκεια στην κατανομή των δημοσίων πόρων ως προς το ανθρώπινο δυναμικό, για παράδειγμα, έχει ως αποτέλεσμα την μεσολάβηση μεγάλων χρόνων αναμονής των πολιτών όταν επιθυμούν να θέσουν ερωτήματα στις δημόσιες υπηρεσίες. Προκειμένου να επιλυθεί το συγκεκριμένο προβληματικό σημείο, απαιτούνται αυξημένα επίπεδα διοικητικής υποστήριξης.

Ένα δεύτερο σημαντικό προβληματικό σημείο των κυβερνητικών και δημοσίων φορέων εδράζεται στην παρουσία μεγάλου πλήθους δεδομένων και πληροφοριών. Ο μεγάλος όγκος δεδομένων αποτελεί ανασταλτικό παράγοντα για την αποτελεσματική

⁷⁸⁷ Βλ. A. Alshahrani, D. Dennehy, & M. Mäntymäki, *όπ.παρ.*, σελ. 1-14.

⁷⁸⁸ Βλ. N. Aoki, *όπ.παρ.*, σελ. 1-10.

συνεργασία των απασχολουμένων δημοσίων υπαλλήλων⁷⁸⁹, οι οποίοι, όπως προαναφέρθηκε, δεν επαρκούν για την ταχεία κάλυψη των αναγκών των πολιτών. Προτείνεται ο συνδυασμός των εσωτερικών πληροφοριών με τις εξωτερικές, έτσι ώστε να επέλθει ως αποτέλεσμα η ουσιαστική εξόρυξη αξιόπιστων και έγκυρων δεδομένων.

Η ποικιλομορφία και διαφορετικότητα του ήδη μεγάλου όγκου δεδομένων απαιτεί την εξειδικευμένη επεξεργασία, ανάλυση και αποκωδικοποίηση οπτικών, χωροταξικών, γλωσσικών και ακουστικών πληροφοριών, προκειμένου να εξαχθούν αξιόπιστα δεδομένα⁷⁹⁰. Μάλιστα, η ανάλυση τόσο των ποσοτικών όσο και των ποιοτικών δεδομένων πρέπει να γίνεται σε τακτική βάση από εξειδικευμένο ανθρώπινο δυναμικό, το οποίο, όμως, απουσιάζει από την δυναμική των δημοσίων υπηρεσιών και συνθέτει ένα ακόμη «οδόφραγμα» στην πορεία για την περαιτέρω αξιοποίηση συστημάτων τεχνητής νοημοσύνης από τους δημόσιους φορείς⁷⁹¹.

⁷⁸⁹ Βλ. D. Susar, & V. Aquaro, *όπ.παρ.*,σελ. 418-426.

⁷⁹⁰ Βλ. M. Kuziemski, & G. Misuraca, *όπ.παρ.*, σελ. 1-13.

⁷⁹¹ Βλ. B. Wirtz, J. Weyerer, & C. Geyer, *όπ.παρ.*, σελ. 1–20.

4.5 Σύνοψη - Συμπεράσματα

Το σημείο εκκίνησης του παρόντος κεφαλαίου έγκειται στην ολοένα και αυξανόμενη συνάφεια της τεχνητής νοημοσύνης στην επιστήμη και την εν γένει καθημερινότητα, καθώς και στην πρωτοποριακή δυναμική της για τον δημόσιο τομέα σε παγκόσμιο επίπεδο, τόσο σε θετικό όσο και σε αρνητικό πλαίσιο. Αν και ορισμένες χώρες, ιδίως οι Ηνωμένες Πολιτείες και η Κίνα, έχουν αναγνωρίσει τη μεγάλη αξία της για τη δημόσια χρήση και έχουν ξεκινήσει διάφορες πρωτοβουλίες τεχνητής νοημοσύνης υψηλής έντασης κόστους, αποκαλύπτοντας το ευρύ φάσμα πιθανών τομέων εφαρμογής της, ωστόσο, καμία κυβέρνηση δεν έχει αντιμετωπίσει μέχρι στιγμής συνολικά ολόκληρο το πεδίο εφαρμογών της. Οι σχετικές προκλήσεις που ενδέχεται να παρεμποδίσουν την επιτυχή ενσωμάτωση της τεχνητής νοημοσύνης στον δημόσιο τομέα, παραμελούνται περιστασιακά από τις αρμόδιες κυβερνήσεις. Γίνεται φανερό ότι παρά τα μεγάλα οφέλη και τις πρόσφατες προσπάθειες επί του πεδίου της τεχνητής νοημοσύνης, η ενσωμάτωση αυτής στον δημόσιο τομέα αποτελεί ένα ιδιαίτερο εγχείρημα.

Δεδομένου ότι ο δημόσιος τομέας αντιπροσωπεύει ένα σχετικά νέο πεδίο εφαρμογής της συγκεκριμένης τεχνολογίας και ότι πολλές εφαρμογές έχουν χρησιμοποιηθεί ως καινοτόμα πιλοτικά έργα, οι κυβερνήσεις και οι δημόσιες αρχές ενδέχεται να μην γνωρίζουν το πλήρες φάσμα των εφαρμογών της, τις ευκαιρίες που προκύπτουν, αλλά ούτε τις συναφείς προκλήσεις. Αυτό επιδεινώνεται από το γεγονός ότι η αντίστοιχη έρευνα τεχνητής νοημοσύνης για τον δημόσιο τομέα βρίσκεται ακόμη σε εμβρυικό στάδιο και μέχρι στιγμής δεν παρέχει μία ολοκληρωμένη εικόνα των εφαρμογών και των προκλήσεων στον τομέα αυτόν.

Οι κατηγορίες που προσδιορίστηκαν στο παρόν κεφάλαιο αναδεικνύουν ήδη τις πρώτες ενδείξεις για τους πιθανούς τομείς εφαρμογής και τις προκλήσεις της τεχνητής νοημοσύνης, από τις οποίες πολλές σημαντικές πτυχές έχουν αναγνωριστεί με βεβαιότητα από προηγούμενες έρευνες, αλλά με μεμονωμένο και αποσπασματικό τρόπο.

Η ολιστική εξέταση των εφαρμογών και των προκλήσεων της τεχνητής νοημοσύνης στον δημόσιο τομέα αποκαλύπτει την αλληλεξάρτησή τους, καθώς οι προκλήσεις ενδέχεται να προκύψουν κατά την εισαγωγή των συγκεκριμένων εφαρμογών στο πεδίο

αυτό και, ως εκ τούτου, θα πρέπει να εξεταστούν από κοινού. Υπάρχουν, συν τοις άλλοις, αλληλεξαρτήσεις μεταξύ των προκλήσεων, γεγονός που επιτρέπει να εξαχθούν πιθανές συνέργειες όταν πρόκειται για την αντιμετώπιση και την υπέρβαση αυτών των προκλήσεων.

Για παράδειγμα, η πρόκληση της επίτευξης κοινωνικής αποδοχής και εμπιστοσύνης στην τεχνητή νοημοσύνη εξαρτάται από το εάν άλλες προκλήσεις πρόκειται να επηρεάσουν αρνητικά τους πολίτες, όπως είναι η ασφάλεια της τεχνητής νοημοσύνης, η προστασία της ιδιωτικής ζωής, οι διακρίσεις και η αντικατάσταση του εργατικού δυναμικού. Επιπροσθέτως, η υπόσταση κανονισμών για τον έλεγχο και τη διαχείριση των εφαρμογών τεχνητής νοημοσύνης απαιτεί την εκ των προτέρων αντιμετώπιση και συζήτηση ηθικών ζητημάτων ή θεμάτων που σχετίζονται με την ενσωμάτωσή της στον δημόσιο τομέα.

Από την περιγραφή που παρατέθηκε στο παρόν κεφάλαιο, αποδεικνύεται το ευρύ και πολύπλευρο φάσμα εφαρμογής της τεχνητής νοημοσύνης. Μέσω του εντοπισμού διαφόρων περιπτώσεων αξιοποίησης και προσθήκης αξίας σε συγκεκριμένες εφαρμογές του δημοσίου τομέα, από την τεχνητή νοημοσύνη, οδηγηθήκαμε στον σχηματισμό δέκα ειδικών κατηγοριών εφαρμογών τεχνητής νοημοσύνης.

Επίσης, στο παρόν κεφάλαιο πραγματοποιήθηκε μία ολοκληρωμένη προσέγγιση, η οποία εντόπισε συγκεκριμένες προκλήσεις της τεχνητής νοημοσύνης στον δημόσιο τομέα και διαμόρφωσε τέσσερις κύριες διαστάσεις, στις οποίες ανατέθηκαν οι εν λόγω προκλήσεις. Η συνάφεια και η σοβαρότητα αυτών των προκλήσεων είναι γεγονός. Παραδείγματος χάριν, η κοινωνική αποδοχή μπορεί να διαφέρει από χώρα σε χώρα και η πρόκληση της αντικατάστασης του εργατικού δυναμικού είναι δυνατόν να έχει μεγαλύτερο αντίκτυπο στην κοινωνία όσον αφορά την ανεργία και την ανισότητα σε χώρες με μεγαλύτερο δυναμικό αυτοματισμού και ψηφιοποίησης σε σύγκριση με άλλες χώρες που η εν λόγω δυναμική τους δεν είναι τόσο έντονη. Ωστόσο, παρατηρήσαμε ότι οι διαστάσεις των προκλήσεων της ενσωμάτωσης της τεχνητής νοημοσύνης στον δημόσιο τομέα είναι πολλές και σχετίζονται με την κοινωνία, την ηθική, την τεχνολογία και το ισχύον νομοθετικό και κανονιστικό πλαίσιο.

Πρέπει να γίνει ευρέως αποδεκτό ότι λόγω των ραγδαίων αλλαγών στην τεχνολογία της τεχνητής νοημοσύνης, καθίσταται εξαιρετικά δυσχερές για τις κυβερνήσεις να συμβαδίσουν με τον συγκεκριμένο ταχύ ρυθμό ανάπτυξης και να δημιουργήσουν

ολοκληρωμένους κανονισμούς και επαρκές νομοθετικό πλαίσιο, ενώ ταυτόχρονα αντιμετωπίζουν την ποικιλία των προκλήσεων που συνοδεύουν την ενσωμάτωσή της στον δημόσιο τομέα. Η ικανοποίηση ηθικών ανησυχιών, η οικοδόμηση μίας ευρείας κατανόησης και αποδοχής της στην κοινωνία, καθώς και η δημιουργία της αίσθησης ασφάλειας και ούτω καθεξής, από την πλευρά των κυβερνήσεων προς την ευρύτερη κοινωνία είναι ζητήματα που χρήζουν άμεσης κάλυψης, αλλά η εν λόγω κάλυψη αποτελεί μία διαδικασία πολύπλοκη, πολυδιάστατη και αναγκαστικά χρονοβόρα, μα ταυτοχρόνως απαραίτητη και αναγκαία.

Συν τοις άλλοις, στο παρόν κεφάλαιο διαφάνηκε ότι η τεχνητή νοημοσύνη εσωκλείει σημαντικές δυνατότητες, αλλά προκειμένου να αξιοποιηθεί η κοινωνικοοικονομικής φύσεως δυναμική της, είναι απαραίτητη η ανάληψη επιχειρηματικού ρόλου από την πλευρά των κυβερνητικών αρχών, οι οποίες καλούνται να ασκήσουν καινοτόμες πολιτικές, που αντικατοπτρίζονται μέσω συνεργατικών σχημάτων μεταξύ δημοσίων οργανισμών και νομικών οντοτήτων που δραστηριοποιούνται στον ιδιωτικό τομέα.

Η σύναψη τέτοιου είδους συνεργασιών με πυρήνα την τεχνητή νοημοσύνη, μπορεί να είναι ένα ζήτημα που δεν έχει αναλωθεί σε ακαδημαϊκές και ερευνητικές συζητήσεις, αλλά ήδη καταγράφει αξιοσημείωτη εμπειρία πρακτικής εφαρμογής του σε αρκετά εγχειρήματα που συναντώνται σε παγκόσμια βάση. Τα πιθανά οφέλη που μπορούν να προκύψουν μέσω συνεργασιών ιδιωτικού και δημοσίου τομέα, ιδίως με σκοπό την αναβάθμιση των παρεχόμενων δημοσίων υπηρεσιών προς τους πολίτες, είναι ευρύτατα και αναλύθηκαν διεξοδικά ανωτέρω. Ωστόσο, πρόκειται για μία σειρά ωφελειών οι οποίες δεν αναδύθηκαν από μόνες τους.

Για λόγους διασφάλισης της μέγιστης δυνατής δυναμικής της συνεργασίας μεταξύ δημοσίου και ιδιωτικού τομέα, επί του πεδίου της τεχνητής νοημοσύνης, είναι απαραίτητη η ανάλυση σε διεθνές επίπεδο, των προσδιοριστικών παραγόντων που πλαισίωσαν επιτυχή παρεμφερή εγχειρήματα.

Προσδοκώντας σε αυξημένες πιθανότητες επιτυχίας διατομεακών συνεργασιών στο πεδίο της τεχνητής νοημοσύνης, προτάσσεται η ανάγκη άσκησης της λεγόμενης διευκολυντικής ηγεσίας. Επιπροσθέτως, η θέση κοινών στόχων προς επίτευξη, από την πλευρά του συνόλου των εμπλεκόμενων στη συνεργασία μερών, είναι ένα σημαντικό ζήτημα που αναγνωρίζεται ευρέως ως προσδιοριστικός παράγοντας επίτευξης μίας καρποφόρας συνεργασίας μεταξύ ιδιωτικού και δημοσίου τομέα.

Η κοινωνικοποίηση αποτελεί έναν ακόμα σημαντικό προσδιοριστικό παράγοντα επιτυχούς συνεργασίας μεταξύ οργανισμών που δραστηριοποιούνται στον ιδιωτικό και τον δημόσιο τομέα. Συνεπώς, πίσω από το σύνολο των υφιστάμενων τεχνικών πολυδιαστάσεων για την ανάπτυξη πρωτοβουλιών στον τομέα της τεχνητής νοημοσύνης, τα άτομα που κρίνονται υπεύθυνα για την θέση πολιτικών κατευθύνσεων και στρατηγικών είναι υποχρεωμένα για τη μεταβίβαση δημόσιας αξίας. Επιπλέον, η αξιοποίηση της ειδικής γνώσης διασφαλίζει ότι οι εναλλακτικές και διαφοροποιημένες στάσεις στην επίλυση προβλημάτων καταλήγουν να ενσωματώνονται και να ευθυγραμμίζονται, προκειμένου να επέλθει η διασφάλιση της ποιότητας.

Ενώ η τεχνητή νοημοσύνη και η ευρύτερη χρήση της εξακολουθεί να βρίσκεται μακριά από την ενσωμάτωση σε τομείς που είναι άμεσα συνυφασμένοι με τη σωματική δραστηριότητα, ωστόσο, η ταχύτητα της προόδου προς εξειδικευμένα συστήματα τεχνητής νοημοσύνης στην υγειονομική περίθαλψη, τη βιομηχανία, τα μέσα κοινωνικής δικτύωσης, τη διαφήμιση και το μάρκετινγκ είναι αξιοσημείωτη. Πολύ λιγότερη πρόοδος εντοπίζεται σε ζητήματα που είναι συναρτημένα με τη διαχείριση των ανθρωπίνων πόρων και οφείλεται σε τέσσερεις λόγους, ήτοι στην πολυπλοκότητα των ζητημάτων ανθρώπινου δυναμικού, στην παρουσία περιορισμένων δεδομένων από δραστηριότητες αναφορικά με τη διαχείριση ανθρώπινου δυναμικού, στη δικαιοσύνη και στους νομικούς περιορισμούς, καθώς επίσης και στις αντιδράσεις των εργαζομένων στη διαχείριση της τεχνητής νοημοσύνης επί ζητημάτων που τους αφορούν.

Η αιτιολογική προσέγγιση μπορεί να οδηγήσει στην αντιμετώπιση αυτών των προκλήσεων στα επιμέρους στάδια του κύκλου ζωής της τεχνητής νοημοσύνης, διότι καθιστά δυνατές τις πρακτικές λήψεις αποφάσεων από μικρά δεδομένα και συν τοις άλλοις, κάνει τη λήψη αποφάσεων που υποστηρίζονται από την αξιοποίηση αλγορίθμων πιο δίκαιη και εξηγήσιμη. Τα εν λόγω οφέλη δεν είναι απαλλαγμένα από αντίστοιχες επιπτώσεις, καθώς οι οργανισμοί πρέπει να αποδεχτούν τη χαμηλότερη προγνωστική ισχύ των αλγορίθμων και να αναζητήσουν συναινετικά στοιχεία σχετικά με τις αιτιώδεις υποθέσεις. Ο πρώτος είναι ο κύριος λόγος για τον οποίο η κοινότητα της επιστήμης της ανάλυσης των δεδομένων είναι αρκετά επιφυλακτική για την αιτιολόγηση της χρήσης συστημάτων τεχνητής νοημοσύνης στη διοίκηση ανθρωπίνων πόρων. Η τυχαιοποίηση είναι μία δεύτερη αρχή που μπορεί να βοηθήσει στον προσδιορισμό της αιτιότητας και μπορεί επίσης να αποτελέσει έναν κανόνα λήψης αποφάσεων όταν υπάρχει χαμηλή προγνωστική ισχύ από την πλευρά των αλγορίθμων,

ενισχύοντας την αντίληψη της δικαιοσύνης μεταξύ των εργαζομένων. Τέλος, η διαδικασία επισημοποίησης είναι απαραίτητη για να διασφαλιστεί ότι οι παραδοχές που ενσωματώνονται στους αλγόριθμους είναι εύλογες και ότι τα αποτελέσματα αυτών είναι κατανοητά. Στην πορεία, η επισημοποίηση μπορεί να είναι περισσότερο δυναμική παρά αναγκαστική⁷⁹².

Ένα ερώτημα που δημιουργείται στο σημείο αυτό είναι το σε ποιο βαθμό οι αλλαγές που προτείνονται στο παρόν κεφάλαιο απαιτούν αναδιάρθρωση της λειτουργίας του ανθρώπινου δυναμικού των οργανισμών. Οι διευθύνσεις ανθρώπινου δυναμικού πρέπει να κατανοήσουν και να διευκολύνουν τα στάδια Δημιουργίας Δεδομένων και Μηχανικής Μάθησης του κύκλου ζωής της τεχνητής νοημοσύνης. Η ενσωμάτωση δεδομένων ανθρώπινου δυναμικού με επιχειρηματικά και οικονομικά δεδομένα θα καταστήσει εφικτή, στο Τμήμα Ανθρώπινου Δυναμικού, την ποσοτικοποίηση της συμβολής του στην κατώτατη γραμμή λειτουργίας του οργανισμού.

Επιπροσθέτως, τα διευθυντικά στελέχη θα πρέπει να ανανεώσουν τις δεξιότητές τους. Η τεχνητή νοημοσύνη πρέπει να συνεπάγεται την ενημέρωση και εκπαίδευσή τους επί της αξιοποίησης και ανάλυσης δεδομένων για ζητήματα που σχετίζονται με το ανθρώπινο δυναμικό και τη λήψη αποφάσεων επ' αυτού. Η βιβλιογραφία που εντοπίστηκε σχετικά με τη διαχείριση που είναι βασισμένη σε αποδεικτικά στοιχεία, προτείνει μία προσέγγιση τύπου Bayesian προσέγγιση για τη συστηματική ενημέρωση των διαχειριστικών πεποιθήσεων με νέες πληροφορίες⁷⁹³. Πρόκειται για ένα βοηθητικό σημείο για τη διαχείριση τεχνητής νοημοσύνης.

Η ένταση που επικρατεί μεταξύ της λογικής της εργασιακής αποδοτικότητας και της καταλληλότητας των απασχολούμενων σε έναν οργανισμό, ασκεί επιρροή στις περισσότερες οργανωτικές δράσεις⁷⁹⁴. Στην περίπτωση του ανθρώπινου δυναμικού, μάλιστα, η επιθυμία και προσπάθεια για εργασιακή αποτελεσματικότητα και οι ανησυχίες για απόδοση δικαιοσύνης δεν ευθυγραμμίζονται σε όλες τις περιπτώσεις.

⁷⁹²Βλ. P. Adler & B. Boris, Two Types of Bureaucracy: Enabling and Coercive, *Administrative Science Quarterly*, Ap. 41, 1996, σελ. 61-89.

⁷⁹³ Βλ. E. Barends & D. Rousseau, *όπ.παρ.*, σελ. 36-41.

⁷⁹⁴Βλ. J. March & H. Simon, *Organizations*, Oxford, Blackwell, 1993, σελ. 19-23.

Κλείνοντας, επισημαίνεται ότι ο ακρογωνιαίος λίθος της αξιοποίησης της τεχνητής νοημοσύνης στον δημόσιο τομέα δεν πρέπει να είναι άλλος, πέραν του πελατοκεντρικού χαρακτήρα της δημόσιας διοίκησης. Επομένως, η ενσωμάτωση μηχανών τεχνητής νοημοσύνης στην καθημερινή επαφή των πολιτών με τις δημόσιες αρχές θα επικεντρώνεται στην ποιοτικότερη και ταχύτερη εξυπηρέτησή τους, στο πλαίσιο της προσπάθειας καλλιέργειας μίας σχέσης εμπιστοσύνης μεταξύ του δημοσίου τομέα και του ευρύτερου κοινού.

Κεφάλαιο 5 Συμπεράσματα και προτάσεις

Κύριο αντικείμενο της διατριβής αυτής ήταν η διερεύνηση των διαφορετικών πτυχών της ενσωμάτωσης της τεχνητής νοημοσύνης στη δημόσια διοίκηση, καθώς επίσης και η διενέργεια μίας προσπάθειας διαλεύκανσης των ηθικών διλημάτων που συνεπάγεται η ενσωμάτωση αυτή. Όπως διαφάνηκε από τη συλλογική ανάλυση της διατριβής αυτής, το γεγονός ότι το πεδίο της τεχνητής νοημοσύνης έχει αναδύσει πρωτόγνωρα ζητήματα και ηθικά διλήμματα στις σύγχρονες κοινωνίες, κατέληξε να αποτελέσει τον πόλο έλξης πολλών μελετών και ερευνών, οι οποίες επικεντρώνονται σε θέματα ενσωμάτωσής της στον δημόσιο και ιδιωτικό τομέα.

Σαφώς, αν τα συστήματα τεχνητής νοημοσύνης δεν είχαν τη δυνατότητα και τις προδιαγραφές να προσφέρουν ουσιαστική υποβοήθηση στη λειτουργική ευρυθμία των δημοσίων και ιδιωτικών οργανισμών και στην ποιότητα διαβίωσης των σύγχρονων πολιτών, τότε ουδεμία συζήτηση θα γινόταν για το ζήτημα αυτό. Βασικός λόγος που το εν λόγω αντικείμενο μελέτης προσέλκυσε το ενδιαφέρον και την προσοχή μας είναι ότι η αξιοποίηση της τεχνητής νοημοσύνης τόσο σε οργανωσιακό όσο και σε κοινωνικό επίπεδο είναι αφενός πολυδιάστατη και αφετέρου πολύπλοκη. Ο βασικός λόγος για τον οποίο συμβαίνει αυτό είναι ότι εσωκλείονται κίνδυνοι, η πρόβλεψη και αντιμετώπιση των οποίων επιβάλλεται, προκειμένου να μην θιγούν τα ανθρώπινα δικαιώματα, αλλά και για να μην καταστεί η χρήση της τεχνητής νοημοσύνης ανήθικη και επιζήμια για τα μέλη των σύγχρονων κοινωνιών.

Η ενσωμάτωση και αξιοποίηση της τεχνητής νοημοσύνης σε όλες τις πτυχές της δημόσιας διοίκησης δεν αρκεί να πραγματοποιηθεί απλά και μόνο επειδή αφορά ένα προηγμένο και καινοτόμο τεχνολογικό σύστημα. Ακόμη και αν εξαλειφθούν όλοι οι ηθικοί κίνδυνοι που απορρέουν από αυτήν, δεν μπορεί να καταστεί εφικτή αν δεν εφαρμοστεί στο πλαίσιο ειδικά σχεδιασμένων στρατηγικών που θέτουν ως επίκεντρο την εξυπηρέτηση των πολιτών.

Οι δημόσιοι και κυβερνητικοί υπάλληλοι θα πρέπει, προς εκπλήρωση των εργασιακών τους καθηκόντων και της κάλυψης των επαγγελματικών τους υποχρεώσεων, να διακατέχονται από τις απαιτούμενες γνώσεις και δεξιότητες που θα απαντούν στην εκάστοτε ανακύπτουσα περίπτωση προς επίλυση κάποιου προβλήματος. Εντούτοις, η ενσωμάτωση της τεχνητής νοημοσύνης δεν αποσκοπεί στην πλήρη αντικατάσταση του

ανθρώπινου παράγοντα. Υπό αυτό το πρίσμα, οι μηχανές τεχνητής νοημοσύνης προσφέρονται ως βοηθητικά επιπρόσθετα εργαλεία που επιταχύνουν την επίλυση των προκύπτοντων προβλημάτων των δημοσίων υπαλλήλων, έτσι ώστε να επιτυγχάνεται ο σκοπός των δημοσίων υπηρεσιών, που δεν είναι άλλος παρά η εξυπηρέτηση των πολιτών.

Στην προκειμένη περίπτωση, επομένως, η ερώτηση που πρέπει να απαντηθεί δεν πρέπει να αφορά στο με ποιον τρόπο μπορεί να χρησιμοποιηθούν τα συστήματα τεχνητής νοημοσύνης για την επίλυση των προκύπτοντων προβλημάτων στις δημόσιες υπηρεσίες, αλλά να αφορά στο ποια προβλήματα είναι αυτά που πρέπει να επιλυθούν μέσω της χρήσης μηχανών τεχνητής νοημοσύνης, πώς μπορούν να επιλυθούν και για ποιον λόγο. Σε περίπτωση κατά την οποία η χρήση των μηχανών τεχνητής νοημοσύνης αποτελεί το βέλτιστο δυνατό μέσο επίτευξης του συγκεκριμένου σκοπού, τότε δεν ενδείκνυται η αξιοποίησή της από τη δημόσια διοίκηση. Σε διαφορετική περίπτωση, η αξιοποίηση της τεχνητής νοημοσύνης στις πτυχές της δημόσιας διοίκησης δεν είναι απαραίτητο να γίνει εξαναγκαστικά.

Σε γενικευμένη βάση, επομένως, θεωρούμε ότι την αποκλειστική ευθύνη για τα αποτελέσματα των ενεργειών και αποφάσεων των συστημάτων τεχνητής νοημοσύνης πρέπει να την έχει πάντα ο άνθρωπος. Μπορεί τα συστήματα τεχνητής νοημοσύνης να έχουν τη δυνατότητα της αυτομάθησης και αυτοεκπαίδευσης, αλλά ο άνθρωπος είναι αυτός που τα δημιούργησε. Αυτό σημαίνει ότι ήδη από το στάδιο του αρχικού σχεδιασμού του λογισμικού των συστημάτων τεχνητής νοημοσύνης, οι κατασκευαστές τους δύνανται να θέσουν τα όρια και τους περιορισμούς των επιπέδων αυτονομίας των μηχανών.

Αν τελικά οι μηχανές τεχνητής νοημοσύνης είναι τα πλέον πρόσφορα εργαλεία που δύνανται να εξυπηρετήσουν τον σκοπό των διαφορετικών πτυχών του δημόσιου τομέα, ακόμη και τότε, δεν είναι δυνατόν να αποτελούν το μοναδικό σημείο επαφής των πολιτών με τις δημόσιες υπηρεσίες. Μολονότι από την ανάλυση που προηγήθηκε στην παρούσα διδακτορική διατριβή εξάγεται το συμπέρασμα ότι οι οργανισμοί που προβαίνουν στην εξυπηρέτηση των πελατών τους χρησιμοποιώντας συστήματα τεχνητής νοημοσύνης από την αρχική μέχρι την τελική επαφή μαζί τους, καταλήγουν να επιτυγχάνουν υψηλότερα επίπεδα πελατειακής ικανοποίησης και πιστότητας, ωστόσο, στο πεδίο της δημόσιας διοίκησης η κατάσταση είναι ελαφρώς

διαφοροποιημένη. Άλλωστε, οι οργανισμοί που υιοθετούν εξ ολοκλήρου συστήματα τεχνητής νοημοσύνης κατά τη διάρκεια της επαφής τους με τους πολίτες, δραστηριοποιούνται στον ιδιωτικό τομέα.

Αυτό είναι και το κομβικό σημείο διαφοροποίησης των ιδιωτικών από τους δημόσιους οργανισμούς. Ο δημόσιος τομέας μπορεί να παραδειγματιστεί από τον ιδιωτικό, αλλά σε καμία περίπτωση δεν μπορεί να ευθυγραμμιστεί απόλυτα με αυτόν και με τις συνθήκες λειτουργίας του. Το βασικό σημείο διαφοροποίησης μεταξύ των δύο τομέων αφορά τη δομική τους φύση. Συνεπώς, οι δημόσιες υπηρεσίες, παραδειγματιζόμενες από καλές πρακτικές που εφαρμόζονται ήδη στον ιδιωτικό τομέα, αναφορικά με το ζήτημα της αξιοποίησης της τεχνητής νοημοσύνης, δύνανται να εκκινήσουν τις διαδικασίες υιοθέτησης μιας περισσότερο πολιτοκεντρικής προσέγγισης, δίνοντας προτεραιότητα στην κάλυψη των αναγκών των πολιτών και την καλύτερη δυνατή εξυπηρέτησή τους. Όμως, ειδικά όσον αφορά το ζήτημα της αξιοποίησης της τεχνητής νοημοσύνης από την πλευρά των δημοσίων υπηρεσιών, πρέπει απαραίτητως να μεσολαβήσει εξέταση για τον ακριβή προσδιορισμό του πότε και σε ποια σημεία μπορεί η τεχνητή νοημοσύνη να αποτελέσει το σημείο επαφής των πολιτών με την εκάστοτε δημόσια υπηρεσία.

Αναλυτικότερα, θεωρούμε ότι η τεχνητή νοημοσύνη δεν θα πρέπει να χρησιμοποιείται όταν από τη χρήση των εν λόγω μηχανών δύναται να καθοδηγηθεί, προς συγκεκριμένη κατεύθυνση, το ανθρώπινο υποσυνείδητο ή η ανθρώπινη συμπεριφορά, ώστε να καταλήξει σε πρόκληση βλάβης προς τον ίδιο ή άλλον άνθρωπο. Επίσης, προτείνεται η απαγόρευση της χρήσης των μηχανών τεχνητής νοημοσύνης όταν αυτή συνεπάγεται την εκμετάλλευση των ευαίσθητων σημείων συγκεκριμένων κοινωνικών ομάδων, που θεωρούνται τρωτές, αλλά και όταν γίνεται αποσκοπώντας στην ταυτοποίηση ανθρώπων για τη λήψη επιτόπιων αποφάσεων που τους αφορούν, είτε οι αποφάσεις αυτές είναι δικαστικής φύσεως είτε αφορούν στο αν θα τους επιτραπεί ή απαγορευτεί η είσοδος σε κάποιον χώρο ή υπηρεσία. Όμως, μπορούν να χρησιμοποιηθούν για τον εντοπισμό εγκληματικών στοιχείων και προσώπων που καταζητούνται από τις δημόσιες αρχές, εφόσον υπάρχει η έκδοση σχετικής άδειας είτε από την αντίστοιχη δικαστική αρχή είτε από την αστυνομία.

Αξίζει να επισημανθεί ότι η συμβολή των πολιτών στην υποστήριξη της ενσωμάτωσης της τεχνητής νοημοσύνης από τις δημόσιες υπηρεσίες είναι απαραίτητη, προκειμένου

αυτό το εγχείρημα να στεφθεί με επιτυχία και να μην απολέσει την πολιτοκεντρική του προσέγγιση. Καταδεικνύεται ότι η πλευρά των δημοσίων και κυβερνητικών φορέων πρέπει να προσβλέπει στη συμμετοχική προσέγγιση του ευρύτερου κοινού, κυρίως για λόγους απομυθοποίησης των μηχανών τεχνητής νοημοσύνης και μεγαλύτερης εξοικείωσης των πολιτών με αυτές. Συνακόλουθα, θα είναι εφικτή η εξάλειψη οποιουδήποτε φόβου ή δισταγμού υπάρχει από την πλευρά των πολιτών προς την τεχνητή νοημοσύνη. Οι ανοικτές συζητήσεις και επιμορφώσεις των πολιτών γύρω από το πεδίο της τεχνητής νοημοσύνης, με απώτερο σκοπό την εις βάθος κατανόηση του τρόπου λειτουργίας της και των ωφελειών που αναδύονται από αυτήν, κρίνονται επιτακτικής ανάγκης. Μάλιστα, η εξάλειψη των ανωτέρω αναφερόμενων ανησυχιών των πολιτών για την αλληλεπίδρασή τους με μηχανές τεχνητής νοημοσύνης, μπορεί να επιτευχθεί αν οι ίδιοι οι δημόσιοι και κυβερνητικοί οργανισμοί προβούν σε λεπτομερή ενημέρωση του κοινού για τα ακριβή σημεία στα οποία θα χρησιμοποιηθούν μηχανές τεχνητής νοημοσύνης στο πλαίσιο της καθημερινής τους εξυπηρέτησης, για τα οφέλη που θα προκύψουν από αυτήν την πρωτοβουλία, αλλά και για την οριοθέτηση αυτής. Η από κοινού δημιουργία ενός κανονιστικού πλαισίου ηθικής και δεοντολογίας για την χρήση της τεχνητής νοημοσύνης, από τους κυβερνητικούς φορείς και από τους ίδιους τους πολίτες, κρίνεται αμοτεροβαρώς επικερδής και ωφέλιμη.

Επιπροσθέτως, επειδή ο ολοσχερής υλικοτεχνικός ανασχεδιασμός των δημοσίων υπηρεσιών πρόκειται να αποβεί εξαιρετικά δαπανηρός, επισημαίνεται ότι δεν είναι απαραίτητος για την επιτυχή ενσωμάτωση της τεχνητής νοημοσύνης. Αντ' αυτού, η τεχνητή νοημοσύνη μπορεί να αξιοποιηθεί ακόμα και με τη χρήση των ήδη υφιστάμενων πόρων των δημοσίων υπηρεσιών ή ακόμη και να συναφθούν διατομεακές συνεργασίες μεταξύ οργανισμών δημοσίου και ιδιωτικού τομέα, προκειμένου να επωφεληθούν οι πρώτοι από τις υποδομές των δεύτερων. Άλλωστε, όπως προαναφέρθηκε, η αξιοποίηση της τεχνητής νοημοσύνης στους οργανισμούς που δραστηριοποιούνται στον ιδιωτικό τομέα βρίσκεται σε αρκετά μεταγενέστερο στάδιο από τα αντίστοιχα επίπεδα στα οποία κυμαίνεται στην περίπτωση των δημοσίων υπηρεσιών.

Επισημαίνεται ότι για την επιτυχή ενσωμάτωση της τεχνητής νοημοσύνης στις πτυχές του δημόσιου τομέα, οι δημόσιες υπηρεσίες πρέπει να έχουν λάβει όλα τα μέτρα και τις δικλίδες ασφαλείας περί προστασίας προσωπικών και ευαίσθητων δεδομένων των πολιτών, για λόγους προάσπισης και διατήρησης του απορρήτου. Αυτό αποτελεί

βασική προϋπόθεση προκειμένου οι κυβερνητικοί και δημόσιοι φορείς να προσχωρήσουν στην αξιοποίηση συστημάτων τεχνητής νοημοσύνης στο πλαίσιο της καθημερινής τους λειτουργίας. Επίσης, προς επίτευξη του συγκεκριμένου σκοπού, οι δημόσιες υπηρεσίες θα πρέπει να αποδώσουν ιδιαίτερη βαρύτητα στην διαδικασία ποιοτικής συγκέντρωσης και επεξεργασίας δεδομένων.

Η απόφαση σχετικά με το ποια δεδομένα είναι χρήσιμα και απαραίτητα, προκειμένου να εξυπηρετήσουν τη μελλοντική χρήση της τεχνητής νοημοσύνης, έχει βαρύνουσα σημασία. Πρόκειται για την λήψη απόφασης για την συγκέντρωση και διαχείριση του σωστού τύπου δεδομένων. Από την άλλη πλευρά, η διασφάλιση του απορρήτου των ευαίσθητων και προσωπικών δεδομένων των πολιτών είναι απαραίτητη, λόγω του ότι ουδεμία προσπάθεια ενσωμάτωσης της τεχνητής νοημοσύνης πρόκειται να στεφθεί με επιτυχία, αν προηγουμένως δεν κερδίσει την εμπιστοσύνη των πολιτών.

Ειδικότερα, το ζήτημα της διασφάλισης και προστασίας των προσωπικών και ευαίσθητων δεδομένων των πολιτών, μπορεί να επιλυθεί όταν τα μοναδικά δεδομένα που θα χρησιμοποιούνται από τις δημόσιες υπηρεσίες, θα είναι δεδομένα που ήδη βρίσκονται στην κυβερνητική βάση δεδομένων και έχουν χορηγηθεί σε κυβερνητικούς φορείς από τους ίδιους τους πολίτες. Όμως, οι ανησυχίες γύρω από το ζήτημα της προστασίας προσωπικών δεδομένων εγείρονται σε περιπτώσεις που δεν έχει παρασχεθεί η αντίστοιχη συναίνεση των πολιτών προς τις εκάστοτε αρμόδιες δημόσιες υπηρεσίες ή όταν οι εξωτερικές πληροφορίες αναμιγνύονται με εσωτερικές που βρίσκονται ήδη στην κατοχή κυβερνητικών φορέων. Η ανακρίβεια των δεδομένων αποτελεί ένα επιπρόσθετο ανησυχητικό ζήτημα που δημιουργεί κινδύνους στο πλαίσιο της προστασίας και ασφάλειας των προσωπικών δεδομένων. Πρόκειται για προκλήσεις, οι οποίες, απαραιτήτως, πρέπει να αντιμετωπίζονται από την πλευρά των δημοσίων και κυβερνητικών αρχών.

Ένα επιπρόσθετο ζήτημα, το οποίο αναλύθηκε διεξοδικά στην παρούσα διατριβή, αναφέρεται στην πρόκληση ηθικών κινδύνων και την ολοσχερή συμμετοχή των μηχανών τεχνητής νοημοσύνης στην διαδικασία λήψης αποφάσεων. Συστήματα τεχνητής νοημοσύνης, μπορεί να είναι ιδιαίτερος επιρρεπής στη λήψη μεροληπτικών αποφάσεων, ανάλογα με το είδος της εκπαίδευσης και τον προγραμματισμό που σχεδίασαν οι κατασκευαστές τους. Ο εν λόγω κίνδυνος υφίσταται ακόμα και σε

περιπτώσεις που η είσοδος πληροφοριών στο μηχάνημα τεχνητής νοημοσύνης πάσχει από κάποια δυσλειτουργία.

Προσβλέποντας στη μείωση του κινδύνου προκατάληψης στη λήψη αποφάσεων που λαμβάνονται από μηχανές τεχνητής νοημοσύνης και ακολούθως να διασφαλιστεί και επαληθευτεί ότι η εκάστοτε ληφθείσα απόφαση είναι και η πλέον ορθή, απαιτείται η συμμετοχή διεπιστημονικών ομάδων στην διαδικασία λήψης αποφάσεων. Με τον τρόπο αυτόν, δηλαδή με την παράλληλη συμβολή και συμμετοχή του ανθρώπινου παράγοντα στη διαδικασία λήψης αποφάσεων, δεν εξαλείφεται μόνο ο κίνδυνος της μεροληψίας αλλά και ο ηθικός κίνδυνος.

Σε κάθε περίπτωση, πρέπει να επισημανθεί ότι λαμβάνοντας υπόψη τα πολλαπλά και πολυδιάστατα ηθικά ζητήματα που εγείρονται και είναι συνυφασμένα με την ενσωμάτωση της τεχνητής νοημοσύνης στις πτυχές του δημόσιου τομέα, συμπεριλαμβανομένης της διαδικασίας λήψης αποφάσεων, αλλά συνυπολογίζοντας και τη διαρκή πρόοδο που γνωρίζει η επιστήμη της μηχανικής μάθησης, κρίνεται δόκιμη και φρόνιμη η αποφυγή λήψης κρίσιμων κυβερνητικών αποφάσεων που αφορούν στους πολίτες, από μηχανές τεχνητής νοημοσύνης. Παραδείγματος χάριν, η λήψη δικαστικών αποφάσεων αναφορικά με την επιβαλλόμενη ποινή εις βάρος των κατηγορουμένων, από την πλευρά μηχανών τεχνητής νοημοσύνης, υποκρύπτει έντονους κινδύνους ηθικής και μεροληψίας, ενώ τα αποτελέσματα μπορούν να αποδειχθούν καταστροφικά και εξαιρετικά δυσμενή για τον άνθρωπο.

Η συμμετοχή των μηχανών τεχνητής νοημοσύνης στη λήψη δικαστικών αποφάσεων, αποτελεί ένα χαρακτηριστικό παράδειγμα προς αποφυγή, τουλάχιστον μέχρι να προκύψει τεκμηριωμένη βελτίωση της μηχανικής μάθησης, ώστε να μπορούν επιβεβαιωμένα να εξαλειφθούν οι κίνδυνοι ηθικής και μεροληψίας. Μέχρι τότε, η χρήση της τεχνητής νοημοσύνης στον δημόσιο τομέα ενδείκνυται μόνο σε περιπτώσεις αναλύσεων και βελτιώσεων των επιμέρους ακολουθούμενων διαδικασιών, για λόγους καθημερινής εξυπηρέτησης των πολιτών και ταχύτερης κάλυψης των αναγκών τους, καθώς επίσης και για λόγους βελτίωσης της ποιότητας των δημόσιων υπηρεσιών. Επομένως, υποστηρίζεται ο αποκλεισμός της χρήσης μηχανών τεχνητής νοημοσύνης σε διαδικασίες λήψης στρατηγικής σημασίας αποφάσεων, τουλάχιστον όχι χωρίς την ανθρώπινη παρέμβαση.

Επί του διλήμματος περί απόδοσης ή μη νομικής προσωπικότητας στα τεχνικά αντικείμενα τεχνητής νοημοσύνης, τασσόμαστε υπέρ της θέσης της ΕΟΚΕ, αλλά και της Πρότασης της Ευρωπαϊκής Επιτροπής που εκδόθηκε το έτος 2021. Πιο συγκεκριμένα, αμφότερες οι τοποθετήσεις πρεσβεύουν ότι πρέπει να εξετάζεται το επίπεδο αυτοβουλίας της κάθε μηχανής και εν συνεχεία να καταλογίζεται η αντίστοιχη ευθύνη. Παραδείγματος χάριν, στην περίπτωση των αυτόνομων οχημάτων – οχήματα που δεν θα διαθέτουν χειριστήρια για να μπορεί ο επιβάτης να παρέμβει σε κάποια δύσκολη κατάσταση - η ευθύνη θα πρέπει να οδηγείται στον κατασκευαστή και στον προγραμματιστή τους. Ο λόγος έγκειται στο ότι αυτός είναι ο αρμόδιος -και υπεύθυνος- για την καταχώριση του συνόλου των απαραίτητων δεδομένων και των αναγκαίων παραμέτρων, έτσι ώστε η συγκεκριμένη μηχανή να είναι σε θέση να λειτουργεί άνευ υποψία πρόκλησης οποιουδήποτε είδους ατυχήματος, είτε για τους επιβάτες, είτε για τους πεζούς, είτε για τα άλλα διερχόμενα οχήματα. Βέβαια, σε αυτές του είδους τις αυτόνομες μηχανές, θα πρέπει οπωσδήποτε να ενσωματώνεται η τεχνολογία των «μαύρων κουτιών», προκειμένου να καταγράφεται οτιδήποτε συμβαίνει κατά τη χρήση τους, συμπεριλαμβανομένου και του περιβάλλοντος στο οποίο χρησιμοποιούνται.

Στις μηχανές τεχνητής νοημοσύνης, οι οποίες θα διαθέτουν την ευχέρεια να εκπαιδεύονται και να μαθαίνουν κατά την αλληλεπίδρασή τους με τους τελικούς χρήστες τους, στην περίπτωση πρόκλησης βλάβης, είτε στον χρήστη τους είτε σε κάποιον τρίτο, θα πρέπει να εξετάζεται αν η προκληθείσα βλαπτική πράξη ήταν αποτέλεσμα των δεδομένων που είχαν καταχωριστεί από το κατασκευαστικό τους στάδιο ή αν ήταν απόρροια της αλληλεπίδρασης και της υιοθέτησης καταχρηστικών – κακόβουλων συμπεριφορών του τελικού τους χρήστη. Συνακόλουθα, πρέπει να καταλογίζεται η ευθύνη είτε στον κατασκευαστή είτε στον τελικό χρήστη των μηχανών αυτών. Σε κάθε περίπτωση, η ευθύνη δεν θα πρέπει να αποδίδεται στη μηχανή τεχνητής νοημοσύνης, αλλά στο φυσικό ή στο νομικό πρόσωπο, το οποίο την έχει στην κυριότητά του.

Όσον αφορά την περίπτωση μηχανών τεχνητής νοημοσύνης, των οποίων η χρήση εγείρει ζητήματα που εμπίπτουν στην κατηγορία του «περιορισμένου» και του «ελάχιστου κινδύνου», αυτά θεωρούμε ότι είναι αναγκαίο να ενταχθούν σε ένα πλαίσιο ασφαλιστικών εταιρειών, το οποίο θα είναι παρεμφερές με τα ασφαλιστικά πλαίσια που επικρατούν στην ασφάλιση οχημάτων. Ο λόγος της αναγκαιότητας αυτής είναι ότι με τον τρόπο αυτόν θεωρούμε ότι θα επιταχύνονται οι διαδικασίες αποζημίωσης των

θυμάτων από τις επικείμενες βλάβες που ενδέχεται να προκληθούν από τη χρήση μηχανών τεχνητής νοημοσύνης. Μάλιστα, τα ασφάλιστρα θα πρέπει να καλύπτονται από τις εκάστοτε εταιρείες, οι οποίες θα έχουν σχεδιάσει και κατασκευάσει τις μηχανές τεχνητής νοημοσύνης.

Επιπροσθέτως, οι συγκεκριμένες ασφαλιστικές εταιρείες θεωρούμε ότι, πέραν από την κάλυψη των ανωτέρω αναφερόμενων ασφαλιστρών, είναι αναγκαίο να διαδραματίζουν έναν προληπτικό εποπτικό ρόλο επί των συγκεκριμένων μηχανών τεχνητής νοημοσύνης, διότι, προκειμένου να μειώσουν το ύψος των αποζημιώσεων που θα καλούνται να καταβάλλουν σε περιπτώσεις πρόκλησης βλάβης από τη χρήση τους, πριν την ενσωμάτωσή τους στο χαρτοφυλάκιό τους, θα εξετάζουν λεπτομερώς τόσο το πλαίσιο δραστηριοτήτων που θα δύνανται να εκτελούν, όσο και τα δεδομένα που θα καταχωρίζονται στο κατασκευαστικό στάδιο. Εξ αυτού, θα είναι σε θέση να γνωρίζουν εις βάθος το εύρος των δυνατοτήτων της εκάστοτε μηχανής τεχνητής νοημοσύνης, το οποίο θα γνωστοποιείται στους τελικούς χρήστες – πολίτες και ακολούθως, θα ενισχύεται η εμπιστοσύνη των τελευταίων προς τις εν λόγω μηχανές.

Είναι προφανές ότι η τεχνητή νοημοσύνη δύναται να ασκήσει μεγάλη επιρροή στον τρόπο αλληλεπίδρασης των πολιτών με τις δημόσιες υπηρεσίες και τις κυβερνητικές αρχές. Αυτό καταδεικνύει ότι τα συστήματα τεχνητής νοημοσύνης δεν αποτελούν την λύση των κυβερνητικών προβλημάτων, ούτε μπορούν να καλύψουν εξ ολοκλήρου τις αδυναμίες του δημόσιου τομέα, οι οποίες όπως προαναφέρθηκε είναι κυρίως δομικές. Ωστόσο, δύνανται να αποτελέσουν έναν τρόπο, στον οποίο οι δημόσιες υπηρεσίες μπορούν να βασιστούν για να αυξήσουν την αποτελεσματικότητά τους και να ενθαρρύνουν πρακτικά τον πολιτοκεντρικό τους χαρακτήρα. Όμως, το ότι τα αντικείμενα τεχνητής νοημοσύνης δεν δύνανται να διακατέχονται από συναισθήματα και αισθήματα, ενισχύει την θέση μας για το ότι δεν μπορεί να τους αποδοθεί νομική προσωπικότητα, ακόμη και αν χαρακτηρίζονται από υψηλά επίπεδα νοημοσύνης.

Εν κατακλείδι, η ενσωμάτωση των εφαρμογών τεχνητής νοημοσύνης στη δημόσια διοίκηση θα πρέπει να πραγματοποιηθεί, πρωτίστως, προς την κατεύθυνση της ενίσχυσης και συμπλήρωσης των εργασιακών καθηκόντων των δημοσίων υπαλλήλων. Η μετακύληση διάφορων εργασιών από τον ανθρώπινο παράγοντα σε μηχανές τεχνητής νοημοσύνης, θα πρέπει να προϋποθέτει την προγενέστερη σε βάθος εξέταση όλων των νομικών και ηθικών παραμέτρων, οι οποίοι θα συντρέχουν για τη συγκεκριμένη

εργασία που πρόκειται να ανατεθεί στις μηχανές, έτσι ώστε να αποφεύγονται επικείμενες απολύσεις, γεγονός το οποίο θα επηρεάσει αρνητικά την εμπιστοσύνη των πολιτών προς τις νέες τεχνολογίες και δη της τεχνητής νοημοσύνης. Μάλιστα, σημαντικό ρόλο για την επιτυχή ενσωμάτωση εφαρμογών τεχνητής νοημοσύνης σε πεδία της δημόσιας διοίκησης θα διαδραματίζει και ο εκάστοτε προϊστάμενος, ο οποίος θα πρέπει να διαθέτει την ικανότητα να αντιλαμβάνεται τα ιδιαίτερα γνωρίσματα των υφισταμένων του, στους οποίους θα ανατίθεται η αρμοδιότητα της αλληλεπίδρασης και, όπου ενδείκνυται, της τροφοδότησης των μηχανών τεχνητής νοημοσύνης με επικαιροποιημένα και νέα δεδομένα, προσβλέποντας σε αυξημένη αποδοτικότητα των παρεχόμενων υπηρεσιών και την επακόλουθη ικανοποίηση των πολιτών από τις βελτιωμένες και άμεσες δημόσιες υπηρεσίες.

Βιβλιογραφία

- ΕΟΚΕ. (2017). Γνωμοδότηση της Ευρωπαϊκής Οικονομικής και Κοινωνικής Επιτροπής (ΕΟΚΕ) με θέμα «Η τεχνητή νοημοσύνη — Η επίδραση της τεχνητής νοημοσύνης στην (ψηφιακή) ενιαία αγορά, στην παραγωγή, στην κατανάλωση, στην απασχόληση και στην κοινωνία», υπ' αριθμόν 2017/C288/01. Διαθέσιμο στο <https://eur-lex.europa.eu/legal-content/EL/TXT/PDF/?uri=CELEX:52016IE5369&from=ET> (4/7/2022)
- Ευρωπαϊκή Επιτροπή. (2021). Πρόταση για τη θέσπιση εναρμονισμένων κανόνων σχετικά με την τεχνητή νοημοσύνη (πράξη για την τεχνητή νοημοσύνη) και για την τροποποίηση ορισμένων νομοθετικών πράξεων της Ένωσης. Διαθέσιμο στο <https://eur-lex.europa.eu/legal-content/EL/TXT/HTML/?uri=CELEX:52021PC0206&from=EN> (2/1/2022)
- Ευρωπαϊκό Κοινοβούλιο. (2019) Κοινοβουλευτικές Ερωτήσεις. Διαθέσιμο στο: https://www.europarl.europa.eu/doceo/document/E-9-2019-002411-ASW_EL.html (19/5/2020)
- Ευρωπαϊκό Κοινοβούλιο. (2020). Τεχνητή νοημοσύνη: το ΕΚ πρωτοστατεί για την πρώτη δέσμη κανόνων της ΕΕ. Διαθέσιμο στο: <https://www.europarl.europa.eu/news/el/press-room/20201016IPR89544/techniti-noimosuni-to-ek-protostatei-gia-tin-protidesmi-kanonon-tis-ee> (5/7/2022)
- Φερενίκη Παναγοπούλου-Κουντατζή, Τεχνητή Νοημοσύνη: Ο δρόμος προς έναν ψηφιακό συνταγματισμό. Μια ηθικο-συνταγματική θεώρηση, (Εκδ. Παπαζήση 2023)
- Abubakar, A. M., Behraves, E., Rezapouraghdam, H., &Yildiz, S. B. (2019). Applying artificial intelligence technique to predict knowledge hiding behavior. International Journal of Information Management, 49, 45–57.
- Adams, S., Arel, I., Bach, J., Coop, R., Furlan, R., Goertzel, B., Hall, J. S., Samsonovich, A., Scheutz, M., Schlesinger, M., Shapiro, S. C., Sowa, J. (2012).

- Mapping the landscape of human-level artificial general intelligence. *AI Magazine*, 33 (1), 25–42.
- Adler, P., & Boris, B. (1996). Two Types of Bureaucracy: Enabling and Coercive. *Administrative Science Quarterly*, 41, 61-89.
- Agranoff, R., & McGuire, M. (1998). Multinetwork management: collaboration and the hollow state in local economic policy. *Journal of Public Administration Research and Theory*, 8, 67–91.
- AI Now Institute. (2020). Available at: <https://ainowinstitute.org> (4/10/2020)
- Alaieri, F., & Vellino, A. (2016). Ethical Decision Making in Robots: Autonomy, Trust and Responsibility. *Lecture Notes in Computer Science*, 159–168.
- Ali, D., & Frimpong, S. (2020). Artificial intelligence, machine learning and process automation: existing knowledge frontier and way forward for mining sector. *Artificial Intelligence Review*, 53, 6025-6042.
- Allen, G.C. (2019). Understanding China's AI Strategy: Clues To Chinese Strategic Thinking On Artificial Intelligence And National Security. Washington: CNASDC.
- Almeida, J. S. (2002). Predictive non-linear modeling of complex data by artificial neural networks. *Current Opinion in Biotechnology*, 13(1), 72–76.
- Alshahrani, A., Dennehy, D., & Mäntymäki, M. (2021). An attention-based view of AI assimilation in public sector organizations: The case of Saudi Arabia. *Government Information Quarterly*, 1-14.
- Aoki, N. (2020). An experimental study of public trust in AI chatbots in the public sector. *Government Information Quarterly*, 37(4), 1-10.
- Aly, A., Griffiths, S., & Stramandinoli, F. (2017). Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics. *Cognitive Systems Research*, 43, 313–323.
- Amodei, D., Olah, C., Steinhardt, J., Christiano, P., Schulman, J., & Mané, D. (2016). Concrete problems in AI safety. In *CoRR abs/1606.06565*, 1–29. Available at: <https://arxiv.org/pdf/1606.06565.pdf> (5/20/2020)

- Anderson, R. E. (1992). Social Impacts of Computing: Codes of Professional Ethics. *Social Science Computer Review*, 10(4), 453–469.
- Anderson, S. L. (2011). Machine metaethics. In M. Anderson & S. L. Anderson (Eds.), *Machine ethics* (pp. 21–27). New York, NY: Cambridge University Press.
- Anderson, M., & Anderson, S. L. (2011). *Machine ethics*. New York: Cambridge University Press.
- Andrade, R. O., & Yoo, S. G. (2019). Cognitive security: A comprehensive study of cognitive science in cybersecurity. *Journal of Information Security and Applications*, 48, 1-13.
- Andras, P., Esterle, L., Guckert, M., Han, T. A., Lewis, P. R., Milanovic, K., & Wells, S. (2018). Trusting Intelligent Machines: Deepening Trust Within Socio-Technical Systems. *IEEE Technology and Society Magazine*, 37(4), 76–83.
- Andrews, R., Esteve, M., & Ysa, T. (2015). Public–private joint-ventures: mixing oil and water? *Public Money Management*, 35, 265–272.
- Angwin, J., Larson, J., Mattu, S. & Kirchner, L. (2016). Machine Bias - There's software used across the country to predict future criminals. And it's biased against blacks. Available at: <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing> (12/2/2020)
- Ansell, C., & Gash, A. (2007). Collaborative governance in theory and practice. *Journal of Public Administration Research and Theory*, 18, 543–571.
- Arnaldi, S., Quaglio, G., Ladikas, M., O'Kane, H., Karapiperis, T., Srinivas, K. & Zhao, Y. (2015). Responsible Governance In Science And Technology Policy: Reflections From Europe, China And India. *Technology In Society*, 42, 81-92.
- Asaro, P. M. (2012). A body to kick, but still no soul to damn: Legal perspectives on robotics. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 169–186). Cambridge: MIT Press.
- Atkinson, R. D. (2017). In defense of robots. *National Review*, LXIX(7). Available at: <https://www.nationalreview.com/2017/04/robots-jobs-industrial-future/> (1/10/2020)

- Austin, H. (2019). Austin Tech Company Proposes Ethics Principles for Design of Digital Humans. Available at: <https://honestaustin.medium.com/austin-tech-company-proposes-ethics-principles-for-design-of-digital-humans-bab506260186> (4/10/2020)
- Baig, A., Dua, A., & Riefberg, V. (2014). Putting citizens first. McKinsey Center for Government. Available at: <https://www.mckinsey.com/~media/mckinsey/industries/public%20and%20social%20sector/our%20insights/how%20us%20state%20governments%20can%20improve%20customer%20service/putting%20citizens%20first%20how%20to%20improve%20citizens%20experience%20and%20satisfaction%20with%20government%20services.pdf> (2/11/2021)
- Balkin, J.M. (2015). The Path of Robotics Law. Yale Law School, 536.
- Balthazar, P., Harri, P., Prater, A., & Safdar, N. M. (2018). Protecting Your Patients' Interests in the Era of Big Data, Artificial Intelligence, and Predictive Analytics. *Journal of the American College of Radiology*, 15(3), 580–586.
- Bamberger, M. (2010). Technologies of Compliance: Risk and Regulation in a Digital Age. *Texas Law Review*, 88, 669-740.
- Banerjee, S., Pradeep Kumar, S., & Bajpai, J. (2018). A comparative study on decision-making capability between human and artificial intelligence. In B. K. Panigrahi, M. N. Hoda, V. Sharma, & S. Goel (Eds.), *Nature inspired computing: Proceedings of CSI 2015* (pp. 203–210.). Singapore: Springer.
- Barad, K. (2003). Posthumanist Performativity: Toward an Understanding of How Matter Comes to Matter. *Signs: Journal of Women in Culture and Society*, 28(3), 801-831.
- Barends, E., & Rousseau, D. (2018). Evidence-Based Management: How to Use Evidence to Make Better Organizational Decisions. UK: Kogan Page.
- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *SSRN Electronic Journal*, 104, 671–732.
- Barocas, S. (2014). Data mining and the discourse on discrimination. *Proceedings of the Data Ethics Workshop, Conference on Knowledge Discovery and Data*

- Mining (KDD), New York. Available at: <https://www.cs.yale.edu/homes/jf/Barocas-Taxonomy.pdf> (7/10/2020)
- Bartlett, S. J. (2017). The case for government by artificial intelligence. SSRN Journal, 1–15.
- Bass, D., & Huet, E. (2017). Researchers combat gender and racial bias in artificial intelligence. Available at: <https://www.bloomberg.com/news/articles/2017-12-04/researchers-combat-gender-and-racial-bias-in-artificial-intelligence> (7/10/2020)
- Bataller, C., & Harris, J. (2016). Turning artificial intelligence into business value- Today. Available at: <https://www.accenture.com/nz-en/insight-artificial-intelligence-business-value> (26/10/2020)
- Beck, S. (2015). The problem of ascribing legal responsibility in the case of robotics. *AI & SOCIETY*, 31(4), 473–481.
- Beltran-Hernandez, C. C., Petit, D., Ramirez-Alpizar, I. G., Nishi, T., Kikuchi, S., Matsubara, T., & Harada, K. (2020). Learning Force Control for Contact-rich Manipulation Tasks with Rigid Position-controlled Robots. *IEEE Robotics and Automation Letters*, 1–1.
- Berger, P., & Luckmann, T. (1966). The social construction of reality. A treatise in the sociology of knowledge. London: The Penguin Press.
- Berk, R. (2017). An impact assessment of machine learning risk forecasts on parole board decisions and recidivism. *Journal of Experimental Criminology*, 13(2), 193–216.
- Bigman, Y. E., Waytz, A., Alterovitz, R., & Gray, K. (2019). Holding Robots Responsible: The Elements of Machine Morality. *Trends in Cognitive Sciences*, 23(5), 365-368.
- Blake, J. (1990). VAX 9000 Series. *Digital Technical Journal*, 2(4), 1-144.
- Blauer, B. (2017). Barriers to building a data practice in government. Available at: <https://sppd.thegovlab.org/lectures/barriers-to-building-a-data-practice-in-government.html> (13/2/2021)

- Bleicher, A. (2017). Demystifying the black box that is AI: Humans are increasingly entrusting our security, health and safety to Black Box intelligent machines. *Scientific American*. Available at: <https://www.scientificamerican.com/article/demystifying-the-black-box-that-is-ai/> (3/10/2020)
- Bock, L. (2015). *Work Rules! Insights from Inside Google That Will Transform How You Live and Lead*. Hachette Book Group.
- Bogue, R. (2017). Domestic robots: Has their time finally come? *Industrial Robot: An International Journal*, 44(2), 129–136.
- Bolukbasi, T., Chang, K.W., Zou, J. Y., Saligrama, V., & Kalai, A. T. (2016). Man is to computer programmer as woman is to homemaker? Debiasing word embeddings. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems 29. Neural information processing systems 2016* (pp. 4349–4357). Barcelona: Curran Associates.
- Bostrom, N., Dafoe, A., & Flynn, C. (2016). Policy desiderata in the development of machine superintelligence. Future of Humanity Institute, University of Oxford. Available at: <https://www.fhi.ox.ac.uk/wp-content/uploads/Policy-Desiderata-in-the-Development-of-Machine-Superintelligence.pdf> (4/10/2020)
- Bovaird, T. (2004). Public–private partnerships: from contested concepts to prevalent practice. *International Review of Administration Science*, 70, 199–215.
- Bovens, M., & Wille, A. (2008). Deciphering the Dutch drop: ten explanations for decreasing political trust in The Netherlands. *International Review of Administrative Sciences*, 74(2), 283–305.
- Boyd, M., & Wilson, N. (2017). Rapid developments in artificial intelligence: How might the New Zealand government respond? *Policy Quarterly*, 13(4), 36–44.
- Boyne, G.A., & Chen, A. (2007). Performance targets and public service improvement. *Journal of Public Administration Research and Theory*, 17, 455–477.
- Bray, S. (2016). The System of Equitable Remedies. *Ucla Law review*, 530 (63), 531–593.

- Bremner, P., Dennis, L. A., Fisher, M., & Winfield, A. F. (2019). On Proactive, Transparent, and Verifiable Ethical Reasoning for Robots. *Proceedings of the IEEE*, 107(3), 541–561.
- Brey, P. (2018). The Strategic Role Of Technology In A Good Society. *Technology In Society*, 52, 39-45.
- Brown, J., Gosling, T., Sethi, B., Sheppard, B., Stubbings, C., Sviokla, J., Jon, W., Daria, Z., Liz, F. (2018). Workforce of the future: The competing forces shaping 2030. Available at: <https://www.pwc.com/gx/en/services/people-organisation/workforce-of-the-future/workforce-of-the-future-the-competing->
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. New York, NY: W. W. Norton & Company.
- Bughin, J., Hazan, E., Ramaswamy, S., Chui, M., Allas, T., Dahlström, P., Nicolaus., H., Monica, T. (2017). Artificial intelligence: The next digital frontier. Available at: <https://www.mckinsey.com/~/media/McKinsey/Industries/Advanced%20Electronics/Our%20Insights/How%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/MGI-Artificial-Intelligence-Discussion-paper.ashx> (4/10/2020)
- Cabrera, A., & Cabrera, E. (2002). Knowledge-sharing dilemmas. *Organizational Studies*, 23, 687–710.
- Calo, R. (2010). Peeping HALs: Making sense of artificial intelligence and privacy. *European Journal of Legal Studies*, 2(3), 168–192.
- Calo, R. M. (2012). Robots and privacy. In P. Lin, K. Abney, & G. A. Bekey (Eds.), *Robot ethics: The ethical and social implications of robotics* (pp. 187–201). Cambridge: MIT Press.
- Campolo, A., Sanfilippo, M., Whittaker, M. & Crawford, K. (2017). *AI Now 2017 Report*. New York: AI Now Institute.

- Canhoto, A. I., & Clear, F. (2019). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, 1-11.
- Capek, K. (2004). *R.U.R. (Rossum's Universal Robots)*. USA: Penguin Books.
- Cappelli, P. & Tavis, A. (2017). The Performance Management Revolution. *Harvard Business Review*. Available at: <https://hbr.org/2016/10/the-performance-management-revolution>
- Cappelli P., Tambe P., Yakubovich V. (2020) Can Data Science Change Human Resources?. In Canals J., Heukamp F. (Eds.) *The Future of Management in an AI World* (pp.. 93-115). Palgrave Macmillan: IESE Business Collection.
- Castelvecchi, D. (2016). Can we open the black box of AI? *Nature*, 538(7623), 20–23.
- Castro, D., & New, J. (2016). The Promise of Artificial Intelligence. Available at: <https://datainnovation.org/2016/10/the-promise-of-artificial-intelligence/> (3/11/2021)
- Castro, D., McLaughlin, M. & Chivot, E. (2019). Who Is Winning the AI Race: China, the EU or the United States. Available at: <https://www.datainnovation.org/2019/08/who-is-winning-the-ai-race-china-the-eu-or-the-united-states/> (2/3/2020)
- Castro, J., Yera, R., & Martínez, L. (2017). An empirical study of natural noise management in group recommendation systems. *Decision Support Systems*, 94, 1–11.
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2017). Artificial Intelligence and the “Good Society”: the US, EU, and UK approach. *Science and Engineering Ethics*, 24, 505-528.
- Cath, C. (2018). Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2133), 1-8.
- Cath, C., Wachter, S., Mittelstadt, B., Taddeo, M., & Floridi, L. (2018). Artificial intelligence and the ‘good society’: The US, EU, and UK approach. *Science and Engineering Ethics*, 24(2), 505–528.

- Cauffman, C. (2018). Robo-liability: The European Union in search of the best way to deal with liability for damage caused by artificial intelligence. *Maastricht Journal of European and Comparative Law*, 25(5), 527–532.
- Cawthorne, D., & Robbins-van Wynsberghe, A. (2020). An Ethical Framework for the Design, Development, Implementation, and Assessment of Drones Used in Public Healthcare. *Science and Engineering Ethics*, 26(5), 2867-2891.
- Center for Human-Compatible AI. (2020). Available at: <https://humancompatible.ai> (4/10/2020)
- Centre for the Study of Existential Risk. (2020). A report on recent activities. Available at: https://www.cser.ac.uk/media/uploads/files/CSER_report_to_donors_2020_general_for_print.pdf (4/20/2020)
- Čerka, P., Grigienė, J., & Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. *Computer Law & Security Review*, 31(3), 376–389.
- Chagal-Feferkorn, K. A. (2019). Am I an Algorithm or a Product? When Products Liability Should Apply to Algorithmic Decision-Makers. *Stanford Law & Policy Review*, 30, 61–114.
- Chen, D.L. & Eagle, J. (2016). Can Machine Learning Help Predict the Outcome of Asylum Adjudications. *Proceedings of the Association for Computing Machinery Conference on Artificial Intelligence and the Law*, Forthcoming.
- Chen, Y., & Lee, J. (2018). Collaborative data networks for public service: governance, management, and performance. *Public Management Reviews*, 20, 672–690.
- Christchurch International Airport (2016). New Zealand's first trial of autonomous vehicle announced. Available at: <https://www.scoop.co.nz/stories/SC1610/S00008/new-zealands-first-trial-of-autonomous-vehicle-announced.htm> (26/10/2020)
- Chua, H., Vicente, V., & Mintzer, S. (2017). Looking through the lens of Rijkswaterstaat traffic inspectors and the municipality of Rotterdam. Available at: <https://www.dssgfellowship.org/2017/08/09/looking-through-the-lens-of->

rijkswaterstaat-traffic-inspectors-and-the-municipality-of-rotterdam/
(12/2/2021)

- Chun, A., & Wai, H. (2007). Using AI for E-government automatic assessment of immigration application forms. *Proceedings of the National Conference on Artificial Intelligence, AAAI 2007, Vancouver*, 1684–1691.
- Ciechanowski, L., Przegalinska, A., Magnuski, M., & Gloor, P. (2018). In the shades of the uncanny valley: An experimental study of human–chatbot interaction. *Future Generation Computer Systems*, 92, 539–548.
- Citron, D. K., & Pasquale, F. A. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89(1), 1–33.
- Citron, D.K. (2008). Technological Due Process. *Washington University Law Review*, 85(6), 1249–1313.
- Clayton, P. R., & Clopton, J. (2018). Business curriculum redesign: Integrating data analytics. *Journal of Education for Business*, 1–7.
- Coeckelbergh, M. (2019). Artificial Intelligence, Responsibility Attribution, and a Relational Justification of Explainability. *Science and Engineering Ethics*, 26, 2051–2068.
- Cole, G. S. (1990). Tort Liability for Artificial Intelligence And Expert Systems. *Computer/Law Journal*, 10(2), 127–231.
- Collier, M., Fu, R. & Yin, L. (2017). Artificial intelligence: Healthcare’s new nervous system. Available at: https://www.accenture.com/t20170418T023052Z__w__/au-en/_acnmedia/PDF-49/Accenture-Health-Artificial-Intelligence.pdf (28/9/2020)
- Conitzer, V., Sinnott-Armstrong, W., Borg, J. S., Deng, Y., & Kramer, M. (2017). Moral decision making frameworks for artificial intelligence. *Proceedings of the 31st AAAI Conference on Artificial Intelligence, AAAI 2017, San Francisco, USA*, 4831–4835.
- Conn, A. (2017). Artificial intelligence: The challenge to keep it safe. San Francisco, CA: Future of Life Institute. Available at: <https://futureoflife.org/2017/09/21/safety-principle/> (2/10/2020)

- Copeland, E. (2017). Lessons from piloting the London Office of Data Analytics. Available at: <https://www.oreilly.com/radar/lessons-from-piloting-the-london-office-of-data-analytics/> (13/2/2021)
- Corbett-Davies, S., Pierson, E., Feller, A. & Goel, S. (2016). A computer program used for bail and sentencing decisions was labeled biased against blacks. It's actually not that clear. Available at: <https://www.washingtonpost.com/news/monkey-cage/wp/2016/10/17/can-an-algorithm-be-racist-our-analysis-is-more-cautious-than-propublicas/> (20/1/2020)
- Cortés-Cediel, M.E.; Cantador, I. & Gil, O. (2017). Recommender systems for e-governance in smart cities. In Yang, Jie; Sun, Zhu; Bozzon, Alessandro; Zhang, Jie & Larson, Martha (Eds.), Proceedings of the international workshop on citizens for recommender systems - CitRec '17, Como, Italy, August 27, 2017, New York, NY: ACM press. Available at: https://www.researchgate.net/publication/319043437_Recommender_systems_for_e-governance_in_smart_cities_State_of_the_art_and_research_opportunities (3/10/2020)
- Cowgill, B. (2017). Bias and Productivity in Humans and Algorithms. Theory and Evidence from Résumé Screening. Available at: https://conference.iza.org/conference_files/MacroEcon_2017/cowgill_b8981.pdf
- Crawford, K. (2017). The Trouble With Bias - NIPS 2017. Video. Available at: <https://www.youtube.com/watch?v=6Uao14eIyGc>
- Cuganesan, S., Hart, A., & Steele, C. (2017). Managing information sharing and stewardship for public-sector collaboration: a management control approach. Public Management Review, 19, 862–879.
- Curts, R.J., & Campbell, D.E. (2001). Avoiding Information Overload Through the Understanding of OODA Loops, A Cognitive Hierarchy and Object-Oriented Analysis and Design. Available at: http://dodccrp.org/events/6th_ICCRTS/Tracks/Papers/Track4/018_tr4.pdf (17/10/2020)

- Danks, D. & London, A.J. (2017). Algorithmic Bias In Autonomous Systems. Proceedings Of The Twenty-Sixth International Joint Conference On Artificial Intelligence (IJCAI-17), 4691-4698.
- Davenport, T. H. (2018). From analytics to artificial intelligence. *Journal of Business Analytics*, 1–8.
- Dawson, P., Clausen, C. &Nielsen, K. (2000). Political Processes In Management, Organization And The Social Shaping Of Technology. *Technology Analysis & Strategic Management*, 12 (1), 5-15.
- Day, D., Fleenor, J., Atwater, L., Sturm, R., &McKee, R. (2014). Advances in leader and leadership development: a review of 25 years of research and theory. *Leadership Quart.*, 25, 63–82.
- De George, R. T. (2003). *The ethics of information technology and business*. Malden, MA: Blackwell Publishing.
- De Mauro, A., Greco, M., Grimaldi, M., &Ritala, P. (2018). Human resources for Big Data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, 54(5), 807–817.
- De Montjoye, Y.A., Farzanehfar, A., Hendrickx, J., & Rocher, L. (2017). Solving artificial intelligence’s privacy problem. *Field Actions Science Reports*. The Journal of Field Actions Special Issue, 17, 80–83.
- Deeks, A. (2019). The judicial demand for explainable artificial intelligence. *Columbia Law Review*, 119(7), 1829-1850.
- Deng, B. (2015). Machine ethics: The robot’s dilemma. *Nature*, 523(7558), 24–26.
- Denrell, J., Fang, C., & Liu, C. (2015). Change Explanations in Management Science. *Organization Science*, 26(3), 923-940.
- Desouza, K. C., Dawson, G. S., &Chenok, D. (2019). Designing, developing, and deploying artificial intelligence systems: Lessons from and for the public sector. *Business Horizons*.

- Devlin, H. (2017). AI programs exhibit racial and gender biases, research reveals. Available at: <https://www.theguardian.com/technology/2017/apr/13/ai-programs-exhibit-racist-and-sexist-biases-research-reveals> (7/10/2020)
- Dheap, V. (2017). IBM QRadar advisor with Watson: Revolutionizing the way security analysts work. Available at: <https://securityintelligence.com/ibm-qradar-advisor-with-watson-revolutionizing-the-way-security-analysts-work/> (5/10/2020)
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2016). Overcoming algorithm aversion: People will use imperfect algorithms if they can (even slightly) modify them. *Management Science*, 64(3), 1155-1170.
- Ding, J. (2018). *Deciphering China's AI Dream. Governance Of AI Program*. Oxford: University of Oxford.
- Ditto, P. H., & Liu, B. (2012). Deontological dissonance and the consequentialist crutch. In M. Mikulincer & P. R. Shaver (Eds.), *The social psychology of morality: Exploring the causes of good and evil* (pp. 51–70). Washington: American Psychological Association.
- Dodds, T. J., Mohler, B. J., & Bülthoff, H. H. (2011). Talk to the virtual hands: Self-animated avatars improve communication in head-mounted display virtual environments. *PLOS ONE*, 6(10), 1-12.
- Domingos, P. (2015). *The Master Algorithm: how the quest for the ultimate learning machine will remake our world*. New York: Basic Books.
- Donahue, J.D., & Zeckhauser, R.J. (2011). *Collaborative governance: private roles for public goals in turbulent times*. Princeton: Princeton University Press.
- Doneda, D., & Almeida, V. A. F. (2016). What Is Algorithm Governance? *IEEE Internet Computing*, 20(4), 60–63.
- Doshi-Velez, F., & Kortz, M. (2017). Accountability of AI Under the Law: The Role of Explanation. Berkman Klein Center Working Group on Explanation and the Law, Berkman Klein Center for Internet & Society working paper, 1-15.
- Dosi, G., & Virgillito, M. E. (2019). Whither the evolution of the contemporary social fabric? New technologies and old socio-economic trends. *International Labour Review*, 158(4), 593-625.

- Dressel, J. & Farid, H. (2018). The accuracy, fairness, and limits of predicting recidivism. *Science Advances*, 4(1).
- Duettmann, A., Afanasjeva, O., Armstrong, S., Braley, R., Cussins, J., Ding, J., Eckersley, P., Guan, M., Vance, A. & Yampolskiy, R. (2018). Artificial General Intelligence: Coordination & Great Powers. 2018 Foresight Institute Strategy Meeting On AGI. Palo Alto: Foresight Institute.
- Durlak, J. A. & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3– 4), 327–350.
- Dutton, T. (2018). An Overview Of National AI Strategies. Medium, Politics + AI. Available at: <https://medium.com/politics-ai/an-overview-of-national-ai-strategies-2a70ec6edfd>
- Ducatel, K., Bogdanowicz, M., Scapolo, F., Leitjen, J., & Burgelman, J.-C. (2005). That's what friends are for - Ambient intelligence (AmI) and the information society in 2010. In A. Zerdick, A. Picot, K. Schrape, J.-C. Burgelman, R. Silverstone, V. Feldmann, C. Wernick & C. Wolff (Eds.), *E-merging media: Communication and the media economy of the future* (pp. 181–200). Berlin: Springer.
- Duncan, G. (2017). Trust, distrust, and the end of politics-as-we-knew-it: the mood of the nation prior to election 2017. *Kōtuitui: New Zealand Journal of Social Sciences Online*, 1–18.
- Eaglin, G.M. (2017). Constructing Recidivism Risk. *Emory Law Journal*, 67(59).
- Ebsworth, J., Johns, S., & Dodson, M. (2021). Surveillance capitalism: the hidden costs of the digital revolution. Available at: <https://evangelicalfocus.com/jubilee-centre/13158/surveillance-capitalism-the-hidden-costs-of-the-digital-revolution> (10/11/2021)
- Eggers, W. D., Fishman, T., & Kishnani, P. (2017). AI-augmented human services: Using cognitive technologies to transform program delivery. Available at:

https://www2.deloitte.com/content/dam/insights/us/articles/4152_AI-human-services/4152_AI-human-services.pdf (26/10/2020)

Eggers, W., & Bellman, J. (2015). The journey to government's digital transformation. Deloitte. Available at:

https://www2.deloitte.com/content/dam/insights/us/articles/digital-transformation-in-government/DUP_1081_Journey-to-govt-digital-future_MASTER.pdf (18/2/2021)

Eisenstein, E.L. (1980). The Printing Press as an Agent of Change. Cambridge: Cambridge University Press.

Elish, M. C., & boyd, danah. (2017). Situating methods in the magic of Big Data and AI. *Communication Monographs*, 85(1), 57–80.

EPEC. (2018). A Guide to Preparing and Procuring a PPP Project Public-Private Partnerships in the Western Balkans. Available at: <https://www.wbif.eu/storage/app/media/Library/8.%20Public%20Private%20Partnership/3.%203-PPP-Preparation-and-Procurement-Guide-FINAL-310818.pdf> (9/3/2021)

Eren, O. & Mocan, N. (2018). Emotional Judges and Unlucky Juveniles. *American Economic Journal: Applied Economics*, 10(3), 171-205.

Ethiraj, S.K., & Levinthal, D. (2009). Hoping for A to Z while rewarding only A: complex organizations and multiple goals. *Organizational Science*, 20, 4–21.

Etzioni, A., & Etzioni, O. (2017). The ethics of robotic caregivers. *Interaction Studies. Social Behaviour and Communication in Biological and Artificial Systems*, 18(2), 174–190.

Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics*, 21(4), 403-418.

European Commission. (2015). Special Eurobarometer 427: Autonomous systems. Report. Available at: http://ec.europa.eu/commfrontoffice/publicopinion/archives/ebs/ebs_427_en.pdf (2/10/2020)

- European Commission. (2018a). Communication Artificial Intelligence for Europe. Available at: <https://ec.europa.eu/digital-single-market/en/news/communication-artificial-intelligence-europe> (23/2/2020)
- European Commission.(2018b). Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions - Coordinated Plan on Artificial Intelligence COM/2018/795 final. Available at: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM:2018:795:FIN> (23/2/2020)
- European Parliament. (2017). Motion for a European Parliament Resolution. Available at: https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html (5/20/2020)
- European Parliament. (2017). European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)). Available at: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051_EN.html (23/2/2020)
- EY. (2018). The growing impact of AI on business. Cambridge,; Massachusetts Institute of Technology. Available at: <https://www.technologyreview.com/s/611013/the-growing-impact-of-ai-on-business/> (4/10/2020)
- Fan, C.Y., Fan, P.S., Chan, T.Y., & Chang, S.H. (2012). Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. *Expert Systems with Applications*, 39(10), 8844–8851.
- Farnham, D., & Horton, S. (1993). *Managing the new public services*. London: Macmillan.
- Fast, N. J., & Schroeder, J. (2019). Power and Decision Making: New Directions for Research in the Age of Artificial Intelligence. *Current Opinion in Psychology*, 33, 172-176.

- Fethi, M. D., & Pasiouras, F. (2010). Assessing bank efficiency and performance with operational research and artificial intelligence techniques: A survey. *European Journal of Operational Research*, 204(2), 189–198.
- Fjeld, J., & Nagy, A. (2015). Principled Artificial Intelligence. Available at: <https://cyber.harvard.edu/publication/2020/principled-ai> (7/10/2020)
- FLI. (2019). AI Policy - China - Future Of Life Institute. Future Of Life Institute. Available at: <https://futureoflife.org/ai-policy/>
- FLI. (2019α). AI Policy Challenges And Recommendations. AI Policy. Future of Life Institute. Available at: <https://futureoflife.org/ai-policy-challenges-and-recommendations/>
- Foote, K. (2016). A Brief History of Artificial Intelligence. Available at: <https://www.dataversity.net/brief-history-artificial-intelligence/> (9/2/2020)
- Foundation for Responsible Robotics. (2020). Available at: <https://responsiblerobotics.org> (4/20/2020)
- Frey, C. B., & Osborne, M. A. (2017). The future of employment. *How Susceptible are Jobs to Computerisation. Technological Forecasting and Social Change*, 114, 254– 280.
- Future of Life Institute. (2015). Autonomous weapons: An open letter from AI & robotics researchers. Available at: <https://futureoflife.org/open-letter-autonomous-weapons/> (2/10/2020)
- Future of Life Institute. (2015). Research priorities for robust and beneficial artificial intelligence. Available at: <https://futureoflife.org/ai-open-letter> (2/10/2020)
- Future of Humanity Institute. (2020). Centre for the Governance of AI. Available at: <https://www.fhi.ox.ac.uk/govai/> (4/10/2020)
- Future of Life. (2020). Asilomar AI Principles. Available at: <https://futureoflife.org/ai-principles/?cn-reloaded=1> (20/2/2020)
- Future of Life. (2020). Beneficial AI 2017. Available at: <https://futureoflife.org/bai-2017/?cn-reloaded=1> (6/10/2020)

- Gagné, J.F. (2018). The global artificial intelligence talent report: 2018. Available at: <https://jfgagne.ai/talent/> (19/10/2020)
- Garcia, D. (2018). Lethal Artificial Intelligence and Change: The Future of International Peace and Security. *International Studies Review*, 20(2), 334–341.
- Gasser, U. (2016). Recoding Privacy Law: Reflections on the Future Relationship among Law, Technology, and Privacy. *Harvard Law Review Forum*, 130(2), 64–65.
- Gasser, U. (2017). AI and the law: Setting the stage. Berkman Klein Center for Internet & Society at Harvard University. Available at: <https://medium.com/berkman-klein-center/ai-and-the-law-setting-the-stage-48516fda1b11> (13/10/2020)
- Gasser, U., & Almeida, V. A. F. (2017). A layered model for AI governance. *IEEE Internet Computing*, 21(6), 58–62.
- Gauglitz, G. (2019). Artificial vs. human intelligence in analytics. *Analytical and Bioanalytical Chemistry*, 1-2.
- Geis, J. R., Brady, A. P., Wu, C. C., Spencer, J., Ranshaert, E., Jaremko, J. L., & Kohli, M. (2019). Ethics of Artificial Intelligence in Radiology: Summary of the Joint European and North American Multisociety Statement. *Canadian Association of Radiologists Journal*, 70, 329-334.
- Gerbert, P., Hecker, M., Steinhäuser, S., & Ruwolt, P. (2017). Putting artificial intelligence to work. Edited by BCG Henderson Institute. Munich, Germany: The Boston Consulting Group. Available at: <https://www.bcg.com/publications/2017/technology-digital-strategy-putting-artificial-intelligence-work> (19/10/2020)
- Giest, S. (2015). Network capacity-building in high-tech sectors: opening the black box of cluster facilitation policy. *Public Administration*, 93, 471–489.
- Goldsmith, S. (2017). Artificial Intelligence Will Help Create a More Responsive Government. Available at: <https://www.govtech.com/opinion/artificial-intelligence-will-help-create-a-more-responsive-government.html> (4/11/2021)
- Gillespie, T. (2016). Algorithm. Available at: <http://culturedigitally.org/2014/06/algorithm-draft-digitalkeyword/> (9/2/2020)

- Gómez de Ágreda, Á. (2020). Ethics of autonomous weapons systems and its applicability to any AI systems. *Telecommunications Policy*, 44(6), 1-15.
- Grosz, B. J., Mackworth, A., Altman, R., Horvitz, E., Mitchell, T., Mulligan, D., & Shoham, Y. (2016). Artificial intelligence and life in 2030: One hundred years study on artificial intelligence. Available at: https://ai100.stanford.edu/sites/default/files/ai_100_report_0831fml.pdf (13/10/2020)
- Gunkel, D. J. (2017). The other question: can and should robots have rights? *Ethics and Information Technology*, 20(2), 87–99.
- Habermas, J., Lennox, S., & Lennox, F. (1974). The Public Sphere: An Encyclopedia Article (1964). *New German Critique*, (3), 49-55.
- Hagendijk, R., & Irwin, A. (2006). Public Deliberation And Governance: Engaging With Science And Technology In Contemporary Europe. *Minerva*, 44 (2), 167-184.
- Hajian, S., & Domingo-Ferrer, J. (2013). A methodology for direct and indirect discrimination prevention in data mining. *IEEE Transactions on Knowledge and Data Engineering*, 25(7), 1445–1459.
- Hall, B. H. (2002). The Assessment: Technology Policy. *Oxford Review Of Economic Policy*, 18(1), 1-9.
- Hallevy, G. (2010). The Criminal Liability of Artificial Intelligence Entities: From Science Fiction to Legal Social Control. *Akron Intellectual Property Journal*, 4(2), 171–201.
- Hameed, I. A., Ten, Z.H., Thomsen, N. B., & Xiaodong, D. (2016). User acceptance of social robots. *The Ninth International Conference on Advances in Computer-Human Interactions*, 274–279.
- Harcourt, B.E. (2005). Against Prediction: Sentencing, Policing, and Punishing in an Actuarial Age. *University of Chicago Law School*, 94.
- Hawksworth, J., Kupelian, B., Berriman, R., & Mckellar, D. (2017). UK economic outlook. Available at: <https://www.pwc.co.uk/economic-services/ukeo/pwc-uk-economic-outlook-full-report-march-2017-v2.pdf> (5/10/2020)

- Heen, H. (2009). One size does not fit all. *Public Management Reviews*, 11, 235–253.
- Hemken, T. & Gray, C. (2016). Smart move: Technologies make their mark on public service. Available at: https://www.accenture.com/t20170224T040950Z__w__/us-en/_acnmedia/PDF-36/Accenture-Public-Service-Emerging-Technologies-Slideshare.pdf (2/9/2020)
- Henman, P. (2020). Improving public services using artificial intelligence: possibilities, pitfalls, governance. *Asia Pacific Journal of Public Administration*, 1–13.
- Herman, J. (2017). Federal pilot to integrate public services into intelligent personal assistants. Available at: <https://digital.gov/2017/04/24/federal-pilot-to-integrate-public-services-into-intelligent-personal-assistants/> (3/10/2020)
- Herman, J. (2017). Opening public services to artificial intelligence assistants. Available at: <https://www.gsa.gov/blog/2017/06/06/Opening-Public-Services-to-Artificial-Intelligence-Assistants> (2/10/2020)
- Herrera-Viedma, E., Vale, Z., Nielsen, P., Martin Del Rey, A., & Casado Vara, R. (2019). Distributed Computing and Artificial Intelligence, 16th International Conference, Special Sessions. Springer.
- Higbie, T. (2013). Why Do Robots Rebel? The Labor History of a Cultural Icon. *Labor Studies in Working-Class History of the Americas*, 10(1), 99–121.
- High-Level Expert Group on AI (HLEG AI). (2019). Ethics guidelines for trustworthy AI. Brussels: European Commission. Available at: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (13/9/2020)
- Hilpert, U. (1991). *State Policies And Techno-Industrial Innovation*. London: Routledge.
- HM Government. (2017). Industrial strategy: building a Britain fit for the future. Industrial Strategy Artificial Intelligence Sector Deal. White Paper.
- HM Treasury. (2013). Private Finance Initiative Projects: 2013 Summary Data. Report number: PU 1587. Available at: <https://www.gov.uk/government/publications/private-finance-initiative-projects-2013-summary-data> (8/3/2021)

- Hodge GA, Greve C. 2007 Public–private partnerships: an international performance review. *Public Administration Review*, 67, 545–558.
- Hoffmann, C. H., & Hahn, B. (2019). Decentered ethics in the machine era and guidance for AI regulation. *AI & Society*, 35, 635-644.
- Hogg, P. (2019). Artificial intelligence: HR friend or foe? *Strategic HR Review*, 12(2), 47-51.
- Holdren, J., & Smith, M. (2016). Preparing for the future of artificial intelligence. Edited by Executive Office of the President National Science and Technology Council Committee on Technology. Available at: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (19/10/2020)
- Holmberg, L. (2020). Human in Command Machine Learning. Available at: <https://www.diva-portal.org/smash/get/diva2:1426167/FULLTEXT01.pdf> (2/10/2020)
- Holmquist, L. E. (2017). Intelligence on tap: Artificial intelligence as a new design material. *Interactions*, 24(4), 28–33.
- House of Lords. 2018 AI in the UK: ready, willing and able? Select Committee on Artificial Intelligence. Report number: HL Paper 100.
- Hovik S, Hanssen G. 2015 The impact of network management and complexity on multi-level coordination. *Public Administration*, 93, 506–523.
- Huang, S.L. & Xiang, J. (2013). Big Data Light up Wisdom of Human Resources Management System. *Science & Technology for Chinese Mass Media*, 12, 76-78.
- Huang M, Kao K, Lu T, Pu W. (2009). Control mechanisms as enhancers of international joint venture performance. Paper presented at the Academy of Management Annual Meeting, Chicago.
- Hunt, A. (2018). The Supreme Court Could Use a Few Good Politicians. Available at: <https://www.bloomberg.com/opinion/articles/2018-07-15/the-supreme-court-could-use-a-few-good-politicians> (31/1/2020)

- Huxham, C. (2003). Theorizing collaboration practice. *Public Management Review* 3, 401–423.
- Ice, B. (2015). Gesture technology, invisible user interface and the evolution of human-to-machine interaction. Available at: <https://www.marketingmag.com.au/hubs-c/gesture-technology-invisible-user-interface-evolution-human-machine-interaction/> (22/10/2020)
- ICF. (2014). Colorado Child Welfare County Workload Study. Available at: https://leg.colorado.gov/sites/default/files/1354s_highlights.pdf (4/11/2021)
- IEEE. (2017). Ethically aligned design: A vision for prioritizing human well-being with autonomous and intelligent systems. Edited by The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available at: http://standards.ieee.org/develop/indconn/ec/ead_v2.pdf (6/10/2020)
- IEEE. (2020). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available at: <https://standards.ieee.org/industry-connections/ec/autonomous-systems.html> (4/10/2020)
- IEEE. (2017). IEEE Global Initiative for Ethical Considerations in Artificial Intelligence (AI) and Autonomous Systems (AS) Drives, together with IEEE Societies, New Standards Projects; Releases New Report on Prioritizing Human Well-Being. Available at: https://standards.ieee.org/news/2017/ieee_p7004.html (21/2/2020)
- IEEE. (2020a). The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems. Available at: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (20/2/2020)
- IEEE. (2020b). Ethically Aligned Design-First Edition. Available at: <https://ethicsinaction.ieee.org> (20/2/2020)
- IEEE. (2020c). Ethically Aligned Design – Version II Request for Input. Available at: https://standards.ieee.org/content/dam/ieee-standards/standards/web/documents/other/ead_v2.pdf (20/2/2020)
- Igbe, O., Darwish, I., & Saadawi, T. (2016). Distributed Network Intrusion Detection Systems: An Artificial Immune System Approach. 2016 IEEE First International

- Conference on Connected Health: Applications, Systems and Engineering Technologies (CHASE), 101-106.
- Ihde, D. (1975). The Experience of Technology: Human-Machine Relations. *Cultural Hermeneutics*, 2(3), 267–279.
- Innes J, Kaplan L, Connick S, Booher D. 2006 Collaborative governance in the CALFED program: adaptive policy making for California water. *Ecological Societies*, 15, 35.
- Innes J, Connick S, Booher D. 2007 Informality as a planning strategy: collaborative water management in the CALFED bay-delta program. *Journal of American Planning Association*, 73, 195–210.
- ITI. (2017). ITI Unveils First Industry-Wide Artificial Intelligence Policy Principles. Available at: <https://www.itic.org/news-events/news-releases/iti-unveils-first-industry-wide-artificial-intelligence-policy-principles> (21/2/2020)
- Jamshidi, P., Camara, J., Schmerl, B., Kaestner, C., & Garlan, D. (2019). Machine Learning Meets Quantitative Planning: Enabling Self-Adaptation in Autonomous Robots. 2019 IEEE/ACM 14th International Symposium on Software Engineering for Adaptive and Self-Managing Systems (SEAMS), 39-50.
- Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. *Government Information Quarterly*, 37(3), 1-8.
- Jefferies, D. (2016). The automated city: Do we still need humans to run public services? Available at: <https://www.theguardian.com/cities/2016/sep/20/automated-city-robots-run-public-services-councils> (3/10/2020)
- Jensen, M.C. (2001). Value maximization, stakeholder theory, and the corporate objective function. *Journal of Application of Corporate Finance*, 14, 8–21.
- Joh, E. (2017). The Undue Influence of Surveillance Technology Companies on Policing. 91 N.Y.U. L. Review Online 101.
- Johnson, D. G. (2015). Technology with no human responsibility? *Journal of Business Ethics*, 127(4), 707–715.

- Johnson, D. G., &Verdicchio, M. (2017). AI anxiety. *Journal of the Association for Information Science and Technology*, 68(9), 2267–2270.
- Jones, M.L. & Levy, K. (2018). Sporting Chances: Robot Referees and the Automation of Enforcement. *We Robot* 2017, 1-42.
- Junqué de Fortuny, E., Martens, D., & Provost, F. (2013). Predictive modeling with big data: is bigger really better, *Big Data*, 1(4), 215-226.
- K & L Gates Hub. (2020). Artificial Intelligence and Machine Learning. Available at: <https://www.klgates.com/artificial-intelligence-machine-learning#LangCode=en-US> (4/10/2020)
- Kahan, D. (2013). Ideology, motivated reasoning, and cognitive reflection. *Judgment and Decision Making*, 8(4), 407–424.
- Kalpouzos, I. (2020). Double elevation: Autonomous weapons and the search for an irreducible law of war. *Leiden Journal of International Law*, 33(2), 289-312.
- Kankanhalli, A., Charalabidis, Y., &Mellouli, S. (2019). IoT and AI for Smart Government: A Research Agenda. *Government Information Quarterly*, 36(2), 304–309.
- Kaplan, A. M., &Haenlein, M. (2009). The fairyland of Second Life: About virtual social worlds and how to use them. *Business Horizons*, 52(6), 563-572.
- Kaplan, A., &Haenlein, M. (2019). Rulers of the world, unite! The challenges and opportunities of artificial intelligence. *Business Horizons*, 63(1), 1-14.
- Katsh, E. (1993). Law in a Digital World: Computer Networks and Cyberspace. *Vill. L. Rev.*, 38(2), 1-84.
- Kaufman, E. M. (2020). Reprogramming consent: implications of sexual relationships with artificially intelligent partners. *Psychology & Sexuality*, 1–12.
- Keating, J., &Nourbakhsh, I. (2018). Teaching artificial intelligence and humanity. *Communications of the ACM*, 61(2), 29-32.
- Kehl, D., Guo, P., & Kessler, S. (2017). Algorithms in the Criminal Justice System: Assessing the Use of Risk Assessments in Sentencing. *Responsive Communities*

- Initiative, Berkman Klein Center for Internet & Society, Harvard Law School, 2-35.
- Kenis, P., & Provan, K. (2009). Towards an exogenous theory of public network performance. *Public Administration*, 87, 440–456.
- Kerikmäe, T., & Pärn-Lee, E. (2020). Legal dilemmas of Estonian artificial intelligence strategy: in between of e-society and global race. *AI & Society*, 1-12.
- Kim, W., Choi, B., Hong, E., Kim, S. & Doheon, L. (2003). Data Mining And Knowledge Discovery, 7(1), 81-99.
- King, T. C., Aggarwal, N., Taddeo, M., & Floridi, L. (2019). Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions. *Science and Engineering Ethics*, 26, 89-120.
- Klein, H.K., & Kleinman, D.L. (2002). The Social Construction Of Technology: Structural Considerations. *Science, Technology, & Human Values*, 27 (1), 28-52.
- Klijn, E., & Teisman, G. (2003). Institutional and strategic barriers to public–private partnerships—an analysis of Dutch cases. *Public Money Management*, 23, 137–146.
- Kline, R. (2011). Cybernetics, Automata Studies, and the Dartmouth Conference on Artificial Intelligence. *IEEE Annals of the History of Computing*, 33(4), 5–16.
- Knight, W. (2017). The dark secret at the heart of AI. *MIT Technology Review*, 120(3), 54–61.
- Kouziokas, G. N. (2017). The application of artificial intelligence in public administration for forecasting high crime risk transportation areas in urban environment. *Transportation Research Procedia*, 24, 467–473.
- Koza, J. R., Bennett, F. H., Andre, D., & Keane, M. A. (1996). Automated Design of Both the Topology and Sizing of Analog Electrical Circuits Using Genetic Programming. *Artificial Intelligence in Design '96*, 151–170.
- Krausová, A. (2017). Intersections between law and artificial intelligence. *International Journal of Computer*, 27(1), 55–68.

- Kreinczes, C. (2016). Artificial Intelligence Innovation Report. Available at: <https://www2.deloitte.com/content/dam/Deloitte/de/Documents/Innovation/Artificial-Intelligence-Innovation-Report-2018-Deloitte.pdf> (24/10/2020)
- Kroll, J.A., Huey, J., Barocas, S., Felten, E.W., Reidenberg, J.R., Robinson, D.G. & Yu, H. (2017). Accountable Algorithms. *Pa. L. Rev.*, 165 (3).
- Kuhlmann, S., &Edler, J. (2003). Scenarios Of Technology And Innovation Policies In Europe: Investigating Future Governance. *Technological Forecasting And Social Change*, 70(7), 619-637.
- Kunze, L., Hawes, N., Duckett, T., Hanheide, M., &Krajnik, T. (2018). Artificial Intelligence for Long-Term Robot Autonomy: A Survey. *IEEE Robotics and Automation Letters*, 3(4), 4023–4030.
- Kuziemski, M., &Misuraca, G. (2020). AI governance in the public sector: Three tales from the frontiers of automated decision-making in democratic settings. *Telecommunications Policy*, 1-13.
- Latonero, M. (2020). Governing Artificial Intelligence:Upholding Human Rights & Dignity. Available at: https://datasociety.net/wp-content/uploads/2018/10/DataSociety_Governing_Artificial_Intelligence_Upholding_Human_Rights.pdf (4/10/2020)
- Lee, M. K., Kusbit, D., Metsky, E., &Dabbish, L. (2015). Working with machines: The impact of algorithmic, data-driven management on human workers. New York: ACM Press.
- Lee, S. G., &Sathikh, P. (2013). A framework for effective human-to-machine communication for artificial interactive systems. In U. Lindemann, S. Venkataraman, Y. S. Kim, S. W. Lee, P. Badke-Schaub, & K. Sato (Eds.), *DS 75-7: Proceedings of the 19th international conference on engineering design (ICED13), Design for harmonies, Vol.7: Human behaviour in design*. Korea: Design Society.
- Legg, S., &Hutter, M. (2006). A formal measure of machine intelligence. In *Proc. 15th Annual Machine Learning Conference of Belgium and The Netherlands (Benelearn2016)*, 73–80.

- Lehr, D. & Ohm, P. (2017). Playing with the Data: What Legal Scholars Should Learn About Machine Learning. *University of California, Davis*, 51, 651-717.
- Leipold, A. (1996). Rethinking Jury Nullification. *Virginia Law Review*, 82(2), 253-324.
- Lever, R. (2017). Tech world debate on robots and jobs heats up. Available at: <https://phys.org/news/2017-03-tech-world-debate-robots-jobs.html> (5/10/2020)
- Leverhulme Centre for the Future of Intelligence. (2020). Available at: <http://lcfi.ac.uk> (4/10/2020)
- Leveringhaus, A. (2018). What's so bad about killer robots? *Journal of Applied Philosophy*, 35(2), 341-358.
- Lewis, O. (1969). Discretionary Justice: A Preliminary Inquiry, by Kenneth Culp Davis. *Case Western Reserve Law Review*, 21(1), 164-173.
- Lieto, A., Bhatt, M., Oltramari, A., & Vernon, D. (2018). The role of cognitive architectures in general artificial intelligence. *Cognitive Systems Research*, 48, 1-3.
- Lin, S.Y., Shanafelt, T.D. & Asch, S.M. (2018). Reimagining clinical documentation with artificial intelligence. *Mayo Clinic Proceedings*, 93(5), 563-565.
- Lind, E. & Van den Bos, K. (2002). When Fairness Works: Toward a General Theory of Uncertainty Management. *Research in Organizational Behavior*, 24, 181-223.
- Liu, C., & Denrell, J. C. (2018). Performance persistence through the lens of chance models. *Academy of Management Proceedings*, 2018(1), 1-4.
- Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., & Lee, I. (2018). Artificial Intelligence in the 21st Century. *IEEE Access*, 6, 34403-34421.
- Liu, H.Y., & Zawieska, K. (2017). From responsible robotics towards a human rights regime oriented to the challenges of robotics and artificial intelligence. *Ethics and Information Technology*, 22, 321-333.
- Livingston, S., & Risse, M. (2019). The Future Impact of Artificial Intelligence on Humans and Human Rights. *Ethics & International Affairs*, 33(02), 141-158.

- Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2017). Brain Intelligence: Go beyond Artificial Intelligence. *Mobile Networks and Applications*, 23(2), 368–375.
- Machine Intelligence Research Institute. (2020). A Guide to MIRI's Research. Available at: <https://intelligence.org/research-guide/> (4/20/2020)
- Madhok, A. (1995). Opportunism and trust in joint venture relationships: an exploratory study and a model. *Scandinavian Journal of Management*, 11, 57–74.
- Magrani, E. (2017). Threats of the internet of things in a techno-regulated society. *ACM SIGCAS Computers and Society*, 47(3), 124–138.
- Magyar, G., Vircikova, M., & Sincak, P. (2016). Interactive Q-learning for social robots that learn from the Wizard: A pilot study. 2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 9-12.
- Makridakis, S. (2017). The forthcoming Artificial Intelligence (AI) revolution: Its impact on society and firms. *Futures*, 90, 46–60.
- Malhotra, C., Kotwaf, V., & Dalai, S. (2018). Ethical Framework for Machine Learning. 2018 ITU Kaleidoscope: Machine Learning for a 5G Future (ITU K), 1-8.
- Malinsky, D., & Danks, D. (2017). Causal discovery algorithms: A practical guide. *Philosophy Compass*, 13(1), 1-11.
- Malli, N., Jacobs, M. & Villeneuve, S. (2018). Intro To AI For Policymakers: Understanding The Shift. March 2018 AI + Public Policy Conference. Ontario: Brookfield Institute.
- Mandel, G. (2017). *Legal Evolution in Response to Technological Change*. England: Oxford University Press.
- Mandell, M.P. (2001) *Getting results through collaboration: networks and network structures for public policy and management*. Westport: Quorum Books.
- Mankins, P. (2015). Can we design trust between humans and artificial intelligence. Available at: <https://www.fastcompany.com/3047500/can-we-design-trust-between-humans-and-artificial-intelligence> (9/10/2020)

- Manyika, J., Chui, M., Miremadi, M., Bughin, J., George, K., Willmott, P., & Dewhurst, M. (2017). A future that works: Automation, employment, and productivity. Available at: https://www.mckinsey.com/~media/McKinsey/Featured%20Insights/Digital%20Disruption/Harnessing%20automation%20for%20a%20future%20that%20works/MGI-A-future-that-works_Full-report.ashx (2/10/2020)
- March, J., & Simon, H. (1993). *Organizations*. Oxford: Blackwell.
- Marchant, G.E. (2011). The Growing Gap between Emerging Technologies and the Law. In Gary E. Marchant et al.(eds) *The Growing Gap between Emerging Technologies and Legal-Ethical Oversight: The Pacing Problem* (pp.19-33). UK: Springer.
- Matthias, A. (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology*, 6(3), 175–183.
- Mavroeidis, V., & Bromander, S. (2017). Cyber Threat Intelligence Model: An Evaluation of Taxonomies, Sharing Standards, and Ontologies within Cyber Threat Intelligence. 2017 European Intelligence and Security Informatics Conference (EISIC), 91-98.
- Mayson, S. (2018). Bias In, Bias Out. *The Yale Law Journal*, 35, 2218-2300.
- McCarthy, J., Minsky, M.L, Rochester, N. & Shannon, C.E. (2006). A proposal for the Dartmouth summer research project on artificial intelligence, August 31, 1955. *AI Magazine*, 27(4), 12–14.
- McCorduck, P. (2004). *Machines who think: A personal inquiry into the history and prospects of artificial intelligence*. Natick, MA: AK Peters Ltd.
- McGarity, T.O. (1992). Some Thoughts on Deossifying the Rulemaking Process. *Duke Law Journal*, 41, 1385-1462.
- McGuire, M. (2006). Collaborative public management: assessing what we know and how we know it. *Public Administration Reviews*, 66, 33–43.
- McIntosh, S. (2011). The Wingman-Philosopher of MiG Alley: John Boyd and the OODA Loop. *Air Power History*, 58(4), 24-33.

- McKinsey Global Institute. 2017 Artificial Intelligence: The next digital frontier? Report. Available at: <https://www.mckinsey.com/~/media/mckinsey/industries/advanced%20electronics/our%20insights/how%20artificial%20intelligence%20can%20deliver%20real%20value%20to%20companies/mgi-artificial-intelligence-discussion-paper.ashx> (11/3/2021)
- Mehr, H. (2017). Artificial intelligence for citizen services and government. Cambridge, MA: Harvard Kennedy School, Ash Center for Democratic Governance And Innovation. Available at: https://ash.harvard.edu/files/ash/files/artificial_intelligence_for_citizen_services.pdf (2/10/2020)
- Meng, X.F. & Ci, X. (2013). Big Data Management: Concepts, Techniques and Challenges. *Journal of Computer Research and Development*, 50, 146-169.
- Metz, C. (2018). Mark Zuckerberg, Elon Musk and the Feud Over Killer Robots. Available at: <https://www.nytimes.com/2018/06/09/technology/elon-musk-mark-zuckerberg-artificial-intelligence.html> (30/1/2020)
- Meurs, P. (1990). Technology, Society And Democracy: The Social Impact Of, And Democratic Control Over Technology, With Special Reference To Information Technology And Data Protection. Ph.D., London School of Economics and Political Science.
- Miall, N. (2017). The policy challenges of automation. *Field Actions Science Reports. The Journal of Field Actions*, 17, 66–71.
- Michel, M. C., & King, M. C. (2018). Categorization of Discoverable Cyber Attributes for Identity Protection, Privacy, and Analytics. *SoutheastCon 2018*, 1-5.
- Micklitz, H.W. (2013). Do Consumers and Businesses Need a New Architecture of Consumer Law? A Thought Provoking Impulse. *Yearbook of European Law*, 32(1), 266–367.
- Microsoft. (2018). Machine translation. Available at: <https://www.microsoft.com/en-us/translator/mt.aspx> (4/10/2020)
- Miller, K. W., Wolf, M. J., & Grodzinsky, F. S. (2017). W

- hy We Should Have Seen That Coming: Comments on Microsoft's tay Experiment, and Wider Implications. *ORBIT Journal*, 1(2), 54-64.
- Minchev, Z. (2016). Challenges to human factor for advance persistent threats proactive identification in modern social networks. *Information & Security: An International Journal*, 34, 57–70.
- Misri, M. (2020). Building Entrepreneurship Culture in Madrasah Actor-Network Theory Bruno Latour Perspective. *Basic and Applied Educational Research Journal*, 1(1), 24-36.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., &Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21.
- Mosser, C. (2018). Mechanical dreams of humanity: When machines long for humanity. *Interdisciplinary Humanities*, 35(1), 78-90.
- Müller, V. (2014). Risks of general artificial intelligence. *Journal of Experimental & Theoretical Artificial Intelligence*, 26(3), 297– 301.
- Müller, M., &Schurr, C. (2016). Assemblage thinking and actor-network theory: conjunctions, disjunctions, cross-fertilisations. *Transactions of the Institute of British Geographers*, 41(3), 217–229.
- Mulligan, D., & Bamberger, K. (2018). Saving Governance-by-Design. *California Law Review*, 697-784.
- Munoz, C., &Smith, M. (2016). Big risks, big opportunities: the intersection of big data and civil rights. Obama White House Archives. Available at: <https://obamawhitehouse.archives.gov/blog/2016/05/04/big-risks-big-opportunities-intersection-big-data-and-civil-rights> (2/2/2021)
- Nagenborg, M., Capurro, R., Weber, J., &Pingel, C. (2008). Ethical regulations on robotics in Europe. *AI & Society*, 22 (3), 349–366.
- Naghibi, S. A., &Pourghasemi, H. R. (2015). A Comparative Assessment Between Three Machine Learning Models and Their Performance Comparison by Bivariate and Multivariate Statistical Methods in Groundwater Potential Mapping. *Water Resources Management*, 29(14), 5217–5236.

- Nebeker, C., Torous, J., & Bartlett Ellis, R. J. (2019). Building the case for actionable ethics in digital health research supported by artificial intelligence. *BMC Medicine*, 17(1), 1-7.
- Nekrasov, V. (2019). Criminal responsibility of robots: actual problem and areas for further development. *Gosudarstvoipravo*, 5, 134-140.
- Neri, E., Coppola, F., Miele, V., Bibbolino, C., & Grassi, R. (2020). Artificial intelligence: Who is responsible for the diagnosis? *La Radiologia Medica*, 125, 517-521.
- New Scientist. (2007). South Korea creates ethical code for righteous robots. Available at: <https://www.newscientist.com/article/dn11334-south-korea-creates-ethical-code-for-righteous-robots/#ixzz6FcZVDBtt> (25/2/2020)
- Ng, G. W., & Leung, W. C. (2020). Strong Artificial Intelligence and Consciousness. *Journal of Artificial Intelligence and Consciousness*, 07(01), 63–72.
- Niehueser, W., & Boak, G. (2020). Introducing artificial intelligence into a human resources function. *Industrial and Commercial Training*, 52(2), 121–130.
- Nikzad–Khasmakhi, N., Balafar, M. A., & Reza Feizi–Derakhshi, M. (2019). The state-of-the-art in expert recommendation systems. *Engineering Applications of Artificial Intelligence*, 82, 126–147.
- Nilsson, N.J. (2003). *Artificial intelligence: A new synthesis*. San Francisco: Morgan Kaufmann Publishers.
- Norman, D. A., Ortony, A., & Russell, D. M. (2003). Affect and machine design: Lessons for the development of autonomous machines. *IBM Systems Journal*, 42(1), 38–44.
- Nutt, P. (2006). Comparing public and private sector decision-making practices. *Journal of Public Administration Research and Theory*, 16, 289–318.
- Nyholm, S. (2018). Attributing Agency to Automated Systems: Reflections on Human–Robot Collaborations and Responsibility-Loci. *Science and Engineering Ethics*, 24(4), 1201–1219.

- OECD. (2019). Artificial Intelligence In Society. Paris: OECD Publishing. Available at: <https://www.oecd.org/publications/artificial-intelligence-in-society-eedfee77-en.htm>
- OFA Intelligence. (2016). Executive Office of the President National Science and Technology Council. 2016 Preparing for the future of Artificial Intelligence. Available at: https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf (20/2/2021)
- Open AI. (2020). Available at: <https://openai.com/about/> (4/20/2020)
- Orlikowski, W.J., & Gash, C.D. (1994). Technological Frames: Making Sense Of Information Technology In Organizations. *ACM Transactions On Information Systems*, 12 (2), 174-207.
- Osório, A., & Pinto, A. (2019). Information, uncertainty and the manipulability of artificial intelligence autonomous vehicles systems. *International Journal of Human-Computer Studies*, 130, 40-46.
- O'Sullivan, S., Nevejans, N., Allen, C., Blyth, A., Leonard, S., Pagallo, U., & Ashrafian, H. (2018). Legal, Regulatory and Ethical Frameworks or Standards for AI and Autonomous Robotic Surgery. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15, 1-12.
- Oude Mulders, J. (2019). Attitudes About Working Beyond Normal Retirement Age: The Role of Mandatory Retirement. *Journal of Aging & Social Policy*, 1–17.
- Özdemir, V., & Hekim, N. (2018). Birth of Industry 5.0: Making Sense of Big Data with Artificial Intelligence, The Internet of Things and Next-Generation Technology Policy. *OMICS: A Journal of Integrative Biology*, 22(1), 65–76.
- Páez, A. (2021). Negligent Algorithmic Discrimination. *Law and Contemporary Problems*, 84(2), 1-18.
- Pagallo, U. (2013). What Robots Want: Autonomous Machines, Codes and New Frontiers of Legal Responsibility. *Human Law and Computer Law: Comparative Perspectives*, 47–65.

- Pagallo, U. (2017). LegalAIze: Tackling the Normative Challenges of Artificial Intelligence and Robotics Through the Secondary Rules of Law. *Perspectives in Law, Business and Innovation*, 281–300.
- Pagallo, U. (2011). Robots of Just War: A Legal Perspective. *Philosophy & Technology*, 24(3), 307–323.
- Pagallo, U. (2018). Vital, Sophia, and Co.—The Quest for the Legal Personhood of Robots. *Information*, 9(9), 230.
- Page, S. (2003). Entrepreneurial strategies for managing interagency collaboration. *Journal of Public Administration Research and Theory*, 13, 311–339.
- Pasquale, F. (2019). A Rule of Persons, Not Machines: The Limits of Legal Automation. *The George Washington Law Review*, 87(1), 1-55.
- Patrick, E. A., & Fattu, J. M. (1986). Artificial intelligence with statistical pattern recognition. Englewood Cliffs, NJ: Prentice Hall.
- Patrick., L., Abney, K., & Bekey, G. A. (2012). Robot ethics: The ethical and social implications of robotics. Cambridge: MIT Press.
- Pearl, J. (2018). *The Book of Why: The New Science of Cause and Effect*. Basic Books.
- Pencheva, I., Esteve, M., & Mikhaylov, S. J. (2018). Big Data and AI – A transformational shift for government: So, what next for research? *Public Policy and Administration*, 1-21.
- Pereira, D. (2017). Watson helps cities help citizens. Available at: <https://medium.com/@darylp/watson-assists-cities-with-311-3d7d6898d132> (4/11/2021)
- Perry, J.L., & Wise, L.R. (1990). The motivational bases of public service. *Public Administration Reviews*, 50, 367–373.
- Pettersen, K. A., McDonald, N., & Engen, O. A. (2009). Rethinking the role of social theory in socio-technical analysis: a critical realist approach to aircraft maintenance. *Cognition, Technology & Work*, 12(3), 181–191.
- Petit, N. (2018). Artificial intelligence and automated law enforcement: A review paper. *SSRN Journal*, 1-12.

- Petzold, C. (2008). *The Annotated Turing: A Guided Tour Through Alan Turing's Historic Paper on Computability and the Turing Machine*. UK: Wiley.
- Pfeffer, J., & Sutton, R. (2006). Hard Facts, Dangerous Half-Truths and Total Nonsense: Profiting from Evidence-Based Management. *Strategy and Leadership*, 34(2), 35–42.
- Philipps, W. (2015). *This is why we can't have nice things: Mapping the relationship between online trolling and mainstream culture*. Cambridge, MA: The MIT Press.
- Podesta, J., Holdren, J., Zients, J., Moniz, E., & Pritzker, P. (2014). *Big Data: seizing opportunities, preserving values*. Report. Obama White House Archives.
- Ponkin, I.V., & Redkina A.I. (2018). Artificial Intelligence from the Point of View of Law. *Rudn Journal of Law*, 22(1), 91–109.
- Power, D. J. (2016). Big Brother can watch us. *Journal of Decision Systems*, 25(1), 578–588.
- Prentice, C., Dominique Lopes, S., & Wang, X. (2019). Emotional intelligence or artificial intelligence— an employee perspective. *Journal of Hospitality Marketing & Management*, 1–27.
- PwC. (2017). *Sherlock in health: How artificial intelligence may improve quality and efficiency, whilst reducing health-care costs in Europe*. Available at: <https://www.pwc.de/de/gesundheitswesen-und-pharma/studie-sherlock-in-health.pdf> (27/10/2020)
- Qela, B., & Mouftah, H. T. (2012). Observe, Learn, and Adapt (OLA)—An Algorithm for Energy Management in Smart Homes Using Wireless Sensors and Artificial Intelligence. *IEEE Transactions on Smart Grid*, 3(4), 2262–2272.
- Qin, X., Luo, Y., Tang, N., & Li, G. (2018). DeepEye: An automatic big data visualization framework. *Big Data Mining and Analytics*, 1(1), 75–82.
- Queudot, M., & Meurs, M.J. (2018). Artificial Intelligence and Predictive Justice: Limitations and Perspectives. *Recent Trends and Future Technology in Applied Intelligence*, 889–897.

- Qiao, X. (2021). Design of Abnormal Behavior Detection System in the State Grid Business Office. In Sun X., Zhang X., Xia Z., Bertino E. (Eds.) *Advances in Artificial Intelligence and Security* (pp.510-520). Springer.
- Quraishi, F. F., Wajid, S. A., & Dhiman, P. (2017). Social and ethical impact of artificial intelligence on public: A case study of university students. *International Journal of Scientific Research in Science, Engineering and Technology*, 3(8), 463– 467.
- Rahwan, I. (2018). Society-in-the-loop: Programming the algorithmic social contract. *Ethics and Information Technology*, 20(1), 5–14.
- Rainey, H.G. (1989). Public management—recent research on the political context and managerial roles, structures and behaviors. *Journal of Management*, 15, 229–250.
- Ram, N. (2017). Innovating Criminal Justice. *Northwestern University Law Review*, 12.
- Raymond, A. (2006). Electronic Commerce and the New Uncitral Convention. *International Law Practicum*, 19(1), 66-74.
- Re, R.M. & Solow-Niederman, A. (2019). Developing Artificially Intelligent Justice. *Stanford Tech. L. Review*, 242-289.
- Reed, C. (2018). How should we regulate artificial intelligence? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376(2128), 1-12.
- Riccucci, N. (2010). *Public administration: traditions of inquiry and philosophies of knowledge*. Washington: Georgetown University Press.
- Rich, E., Knight, K. & Nair, S.B. (2009). *Artificial intelligence*. Third edition. New Dehli, India: Tata McGraw-Hill.
- Rich, A. S., &Gureckis, T. M. (2019). Lessons for artificial intelligence from the study of natural stupidity. *Nature Machine Intelligence*, 1(4), 174–180.
- Richardson, R., Schultz, J. & Crawford, K. (2019). Dirty Data, Bad Predictions: How Civil Rights Violations Impact Police Data, Predictive Policing Systems,

- AndJustice. New York University Law Review Online, Forthcoming 94 (192), 193-228.
- Rizer, A., & Watney, C. (2019). Artificial Intelligence Can Make Our Jail System More Efficient, Equitable and Just. *Texas Review of Law and Politics*, 23(1).
- Roberts, A. (2017). Five big challenges to AI adoption and success. Available at: <https://www.clickz.com/five-big-challenges-to-ai-adoption-and-success/112795/> (1/10/2020)
- Rosa, A., Feyereisl, J., & Team, T. G. (2016). A framework for searching for general artificial intelligence. Available at: <https://arxiv.org/pdf/1611.00685.pdf> (24/10/2020)
- Rössler, B. (2005). *The value of privacy*. Cambridge, UK: Polity Press.
- Rosenstrauch, M. J., & Kruger, J. (2017). Safe human-robot-collaboration-introduction and experiment using ISO/TS 15066. 2017 3rd International Conference on Control, Automation and Robotics (ICCAR), 740-744.
- Rossi, F. (2019). Building trust in artificial intelligence. *Journal of International Affairs*, 72(1), 127-133.
- Roth, A. (2017). Machine Testimony. *The Yale Law Journal*, 1972-2053.
- Rousseau, D. (2014). *The Oxford Handbook of Evidence-Based Management*. Oxford University Press.
- Russell, S., & Norvig, P. (2016). *Artificial intelligence: A modern approach* (Global Edition). Englewood Cliffs, NJ: Pearson Higher Ed.
- Russel, S. & Norvig, P. (2010). *Artificial intelligence: A modern approach*. Third edition. Englewood Cliffs, NJ: Prentice Hall.
- Russell, S. (2015). Take a stand on AI weapons. *Nature*, 521 (7553), 415–416.
- Saarikko, T., Westergren, U. H., & Blomquist, T. (2017). The internet of things: Are you ready for what's coming? *Business Horizons*, 60(5), 667-676.

- Sagredo-Olivenza, I., Gomez-Martin, P. P., Gomez-Martin, M. A., & Gonzalez-Calero, P. A. (2017). Trained Behavior Trees: Programming by Demonstration to Support AI Game Designers. *IEEE Transactions on Games*, 1–1.
- Santoro, M., Marino, D., & Tamburrini, G. (2008). Learning robots interacting with humans: From epistemic risk to responsibility. *AI & Society*, 22(3), 301–314.
- Sanzogni, L., Guzman, G., & Busch, P. (2017). Artificial intelligence and knowledge management: questioning the tacit dimension. *Prometheus*, 35(1), 37–56.
- Sartor, G. (2009). Cognitive automata and the law: electronic contracting and the intentionality of software agents. *Artificial Intelligence Law* 17(4), 253–290.
- Saurabh, P., & Verma, B. (2016). An efficient proactive artificial immune system based anomaly detection and prevention system. *Expert Systems with Applications*, 60, 311–320.
- Saz-Carranza, A., & Ospina, S. (2011). The behavioural dimension of governing inter-organizational goal-directed networks: managing the unity–diversity tension. *Journal of Public Administration Research and Theory*, 21, 327–365.
- Saz-Carranza, A., & Longo, F. (2012). Managing competing institutional logics in public–private joint ventures. *Public Management Review*, 14, 331–357.
- Scherer, M. U. (2016). Regulating artificial intelligence systems: Risk, challenges, competencies, and strategies. *Harvard Journal of Law & Technology*, 29(2), 353–400.
- Scherz, Z., Goldberg, D., & Fund, Z. (1990). Cognitive Implications of Learning Prolog—Mistakes and Misconceptions. *Journal of Educational Computing Research*, 6(1), 89–110.
- Selbst, A. & Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. *Fordham Law Review*, 87(3), 1085–1139.
- Seo, S.A. (2016). The New Public. *The Yale Law Journal*, 125(6), 1548–1819.
- Shafer, G. (1987). Probability Judgment in Artificial Intelligence and Expert Systems. *Statistical Science*, 2(1), 3–16.

- Shaoul, J., Shepherd, A., Stafford, A., & Stapleton, P. (2013). Losing control in joint ventures: the case of building schools for the future. Report. Institute of Chartered Accountants of Scotland. Available at: https://www.research.manchester.ac.uk/portal/files/32803045/FULL_TEXT.PDF (22/2/2021)
- Shavell, S. (2004). Foundations of economic analysis of law. Cambridge, MA: Belknap Press of Harvard University Press.
- Sholla, S., Mir, R. N., & Chishti, M. A. (2018). Eventuality of an Apartheid State of Things. *International Journal of Technoethics*, 9(2), 62–76.
- Simmons, R. (2018). Big Data, Machine Judges, and the Legitimacy of the Criminal Justice System. *University of California*, 52, 1066-1118.
- Smuha, N.A. (2019). The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence. *Computer Law Review International*, 20(4), 97 - 106.
- Society TR. (2017). Machine learning: the power and promise of computers that learn by example. Available at: <https://royalsociety.org/~media/policy/projects/machine-learning/publications/machine-learning-report.pdf> (28/2/2021)
- Soltanzadeh, S., Galliot, J., & Jevglevskaja, N. (2020). Customizable Ethics Settings for Building Resilience and Narrowing the Responsibility Gap: Case Studies in the Socio-Ethical Engineering of Autonomous Systems. *Science and Engineering Ethics*, 1-16.
- Solum, L. (2004). Procedural Justice. *Georgetown University Law Center*, 78, 181-321.
- Solum, L. B. (1992). Legal personhood for artificial intelligences. *North Carolina Law Review*, 70(4), 1231–1288.
- Sonnenberg, F.K. (1992). Partnering: entering the age of co-operation. *Journal of Business Strategy*, 13, 49–52.
- Soo, Z. (2018). The increasing use of artificial intelligence is stoking privacy concerns in China. Available at: <http://www.scmp.com/business/companies/article/2135713/increasing-use-artificial-intelligence-stoking-privacy-concerns> (9/10/2020)

- Sousa, W. G. de, Melo, E. R. P. de, Bermejo, P. H. D. S., Farias, R. A. S., & Gomes, A. O. (2019). How and where is artificial intelligence in the public sector going? A literature review and research agenda. *Government Information Quarterly*, 1-14.
- Spoladore, D., Arlati, S., & Sacco, M. (2017). Semantic and Virtual Reality-Enhanced Configuration of Domestic Environments: The Smart Home Simulator. *Mobile Information Systems*, 1–15.
- Stanford University. (2020). One Hundred Year Study on Artificial Intelligence (AI100). Available at: <https://ai100.stanford.edu/people> (4/10/2020)
- Stoker, G., Smith, G., & Maloney, W. (2004). Building social capital in city politics: scope and limitations at interorganisational level. *Political Studies*, 52, 508–530.
- Sun, T. Q., & Medaglia, R. (2018). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*.
- Sunstein, C.R. (2017). *Republic: Divided Democracy in the Age of Social Media*. England: Princeton.
- Susar, D., & Aquaro, V. (2019). Artificial Intelligence. *Proceedings of the 12th International Conference on Theory and Practice of Electronic Governance*, 418-426.
- Starr, S. (2014). Evidence-Based Sentencing and the Scientific Rationalization of Discrimination. *Stanford Law Review*, 66(4). Surden, H. (2011). The Variable Determinacy Thesis. *The Columbia Science and Technology Law Review*, 12(1).
- Stefan, O.A., Avbelj, M., Eliantonio, M., Hartlapp, M., Korkea-aho, E. & Rubio, N. (2019). EU Soft Law in the EU Legal Order: A Literature Review. King's College London Law School Research Paper Forthcoming. Available at: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3346629 (22/6/2020)
- Swan, M. (2018). Blockchain for Business: Next-Generation Enterprise Artificial Intelligence Systems. *Blockchain Technology: Platforms, Tools and Use Cases*, 111, 121–162.

- Tambe, P., Cappelli, P., &Yacubovich, V. (2019). Artificial intel- ligence in human resources management: Challenges and a path forward. *California Management Review*, 61(4), 15-42.
- Tanaka, T., & Kobayashi, K. (2015). Developing a dividual model using a modular neural network for human-robot interaction. *Journal of Robotics, Networking and Artificial Life*, 2(1), 34–39.
- Tasioulas, J. (1996). *The Paradox of Equity*. Cambridge University Press, 55(3), 456-469.
- Tavallaey, S. S., & Ganz, C. (2019). Automation to Autonomy. 2019 24th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA), 31-34.
- Tavani, H. (2018). Can Social Robots Qualify for Moral Consideration? Reframing the Question about Robot Rights. *Information*, 9(4), 73.
- Tegmark, M. (2017). *Life 3.0: Being Human in The Age Of Artificial Intelligence*. London: Penguin Books.
- Thierer, A., O’Sullivan Castillo, A., & Russell, R. (2017). Artificial intelligence and public policy. Mercatus research. Mercatus Center at George Mason University. Available at:<https://www.mercatus.org/system/files/thierer-artificial-intelligence-policy-mr-mercatus-v1.pdf> (24/10/2020)
- Thomson, A., &Perry, J. (2006). Collaboration processes: inside the black box. *Public Administration Reviews*, 66, 20–32.
- Tian, H. (2014). The Concept, Thinking of Human Resources Management Based on the Era of Big Data. *Human Resources*, 20, 162-163.
- Tucker, C. (2017). Privacy, Algorithms, and Artificial Intelligence. In Ajay A., Joshua G., and Avi G. (Eds.) *The Economics of Artificial Intelligence: An Agenda* (pp. 19-35). University of Chicago Press.
- Turilli, M. (2007). Ethical protocols design. *Ethics and Information Technology*, 9(1), 49–62.

- Tyler, T & Mitchell, G. (1994). Legitimacy and the Empowerment of Discretionary Legal Authority: The United States Supreme Court and Abortion Rights. *Duke Law Journal*, 43(4), 703-815.
- Tzafestas, S. G. (2016). Artificial intelligence. *Roboethics*, 79, 25-33.
- UK Parliament. (2017). Artificial Intelligence Committee - AI in the UK: ready, willing and able. Available at: <https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/10002.htm> (23/2/2020)
- UNESCO. (2017). Report of COMEST on robotics ethics. Paris: World Commission on the Ethics of Scientific Knowledge and Technology. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000253952> (2/9/2020)
- USC Center of AI in Society. (2020). Available at: <https://cais.usc.edu> (21/10/2020)
- Vallor, S., &Bekey, G. A. (2017). Artificial Intelligence and the Ethics of Self-learning Robots. In P. Lin, K. Abney, & R. Jenkins (Eds.), *Robot Ethics 2.0* (pp. 338–353). Oxford: Oxford University Press.
- Velasquez, M., Andre, C., Shanks, T., Meyer, S. J., & Meyer, M. J. (2010). What is ethics. Available at: <https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/> (6/10/2020)
- Verbeek, P. (2011). *Moralizing Technology: Understanding and Designing the Morality of Things*. London: The University of Chicago Press.
- Vermaas, P., Kroes, P., van de Poel, I., Franssen, M., & Houkes, W. (2011). A Philosophy of Technology: From Technical Artefacts to Sociotechnical Systems. *Synthesis Lectures on Engineers, Technology, and Society*, 6(1), 1-134.
- Villaronga, E. F., Kieseberg, P., & Li, T. (2018). Humans forget, machines remember: Artificial intelligence and the Right to Be Forgotten. *Computer Law & Security Review*, 34(2), 304–313.
- Vladeck, D. C. (2014). Machines without principals: liability rules and artificial intelligence. *Washington Law Review*, 89(1), 117–150.

- VoPham, T., Hart, J. E., Laden, F., & Chiang, Y.Y. (2018). Emerging trends in geospatial artificial intelligence (geoAI): Potential applications for environmental epidemiology. *Environmental Health*, 17(40), 1-6.
- Wade, J., Buenfil, J., & Collopy, P. (2020). A Systems Engineering Approach for Artificial Intelligence: Inspired by the VLSI Revolution of Mead & Conway. *INSIGHT*, 23(1), 41–47.
- Wagner, M. (2014). The dehumanization of international humanitarian law: legal, ethical, and political implications of autonomous weapons systems. *Vanderbilt Journal of Transnatl Law*, 47, 1371–1424.
- Wakefield, J. (2018). The man who was fired by a machine. Available at: <https://www.bbc.com/news/technology-44561838> (30/1/2020)
- Wallach, W. & Allen, C. (2008). *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Oxford University Press.
- Wallach, W. & Marchant, G. (2018). An Agile Ethical/Legal Model for the International and National Governance of AI and Robotics. Available at: https://www.aies-conference.com/2018/contents/papers/main/AIES_2018_paper_77.pdf (27/2/2020)
- Wang, C., Teo, T. S. H., & Janssen, M. (2021). Public and private value creation using artificial intelligence: An empirical study of AI voice robot users in Chinese public sector. *International Journal of Information Management*, 61, 1-15.
- Wang, W., &Siau, K. (2019). Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity. *Journal of Database Management*, 30(1), 61–79.
- Wang, L., Yang, C., & Hu, R. Q. (2019). Autonomous Traffic Offloading in Heterogeneous Ultra-Dense Networks Using Machine Learning. *IEEE Wireless Communications*, 1–8.
- Wang, D.H. (2013). Evidence-Based Human Resource Management in the Era of Big Data. *Chinese and Foreign Entrepreneurs*, 8, 159.

- Wassermann, J., Vonesek, V., & Vick, A. (2018). Distributed Industrial Robot Control Using Environment Perception and Parallel Path Planning Cloud Services. 2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA), 1263-1265.
- Waugh, W.L., & Streib, G. (2006). Collaboration and leadership for effective emergency management. *Public Administration Review*, 66, 131–140.
- Weber, E., & Khademian, A. (2008). Wicked problems, knowledge challenges, and collaborative capacity builders in network settings. *Public Administration Reviews*, 68, 334–349.
- Weitz, K., Schiller, D., Schlagowski, R., Huber, T., & André, E. (2019). Do you trust me? Proceedings of the 19th ACM International Conference on Intelligent Virtual Agents - IVA '19, 7-9.
- Weng, Y.H. (2015). Intersection of Tokku special zone, robots, and the law: a case study on legal impacts to humanoid robots. *International Journal of Social Robot*, 7(5), 841–857.
- West, S., Whittaker, M., & Crawford, K. (2019). *Discriminating Systems: Gender, Race, And Power In AI*. New York: AI Now Institute.
- Wirtz, B. W., Weyerer, J. C., & Geyer, C. (2018). Artificial Intelligence and the Public Sector—Applications and Challenges. *International Journal of Public Administration*, 1–20.
- Wisniewski, M. (2001). Using SERVQUAL to assess customer satisfaction with public sector services. *Managing Service Quality: An International Journal*, 11(6), 380–388.
- Wisskirchen, G., Biacabe, B. T., Bormann, U., Muntz, A., Niehaus, G., Jiménez Soler, G., & Brauchitsch, B. V. (2017). Artificial intelligence and robotics and their impact on the workplace. Available at: <https://www.ibanet.org/Document/Default.aspx?DocumentUid=c06aa1a3-d355-4866-beda-9a3a8779ba6e> (7/10/2020)
- Wolf, M. J., Miller, K. W., & Grodzinsky, F. S. (2017). Why We Should Have Seen That Coming. *The ORBIT Journal*, 1(2), 1–12.

- Wong, C., Yang, E., Yan, X.T., & Gu, D. (2017). Adaptive and intelligent navigation of autonomous planetary rovers — A survey. 2017 NASA/ESA Conference on Adaptive Hardware and Systems (AHS), 237-244.\
- Wong, Z. S. Y., Zhou, J., & Zhang, Q. (2018). Artificial Intelligence for infectious disease Big Data Analytics. *Infection, Disease & Health*.
- Woodside, C. (2013). Insights on hybrid vehicles, making global warming local. Available at: <https://www.yaleclimateconnections.org/2013/10/insights-on-hybrid-vehicles-making-global-warming-local/> (2/2/2020)
- World Economic Forum. (2020). Global Future Council on Artificial Intelligence for Humanity. Available at: <https://www.weforum.org/communities/gfc-on-artificial-intelligence-for-humanity> (4/10/2020)
- Yadegaridehkordi, E., Hourmand, M., Nilashi, M., Shuib, L., Ahani, A., & Ibrahim, O. (2018). Influence of big data adoption on manufacturing companies' performance: An integrated DEMATEL-ANFIS approach. *Technological Forecasting and Social Change*, 1-12.
- Yang, J. (2020). Bridging AI Principles to Practice with ABOUT ML. Available at: <https://www.partnershiponai.org/bridging-ai-principles-to-practice-with-about-ml/> (4/20/2020)
- Yasin, R. (2013). How analytics is making NYC's streets and buildings safer. *GCN Magazine*. Available at: <https://gcn.com/articles/2013/10/04/gcn-award-nyc-databridge.aspx> (20/2/2021)
- Yi Tan, C. (2018). Artificial Intelligence, Artificial Persons, and the Law. *Becoming Human: Artificial Intelligence Magazine*. Available at: <https://becominghuman.ai/artificial-intelligence-artificial-persons-and-the-law-2cce322743b6> (20/10/2020)
- Zhao, Y., Hryniewicki, M. K., Cheng, F., Fu, B., & Zhu, X. (2018). Employee Turnover Prediction with Machine Learning: A Reliable Approach. *Intelligent Systems and Applications*, 737–758.
- Zheng, Y., Han, Y., Cui, L., Miao, C., Leung, C., & Yang, Q. (2018). SmartHS: An AI platform for improving government service provision. *The Thirtieth AAAI*

Conference on Innovative Applications of Artificial Intelligence (IAAI- 18),
7704–7711. Available at:
<https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16041/16369>
(20/10/2020)

Zhou, G.H. (2013). Thinking on Human Resources Management Based on the Value
of Big Data. *Science Mosaic*, 12, 104-107.

Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017). Machine learning on big data:
Opportunities and challenges. *Neurocomputing*, 237, 350–361.