



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ**

**Τμήμα Μηχανικών Πληροφοριακών και
Επικοινωνιακών Συστημάτων**

Διπλωματική εργασία

Προσωπικά δεδομένα και εφαρμογές τεχνητής νοημοσύνης.

Ντινοπούλου Ιορδάνα

Επιβλέπουσα καθηγήτρια: Διαμαντοπούλου Βασιλική, Επίκουρη
Καθηγήτρια Τμήματος Μηχανικών Πληροφοριακών και
Επικοινωνιακών Συστημάτων (Μ.Π.Ε.Σ.)

Ευχαριστίες

Η ολοκλήρωση της παρούσας διπλωματικής, θα ήταν αδύνατη χωρίς την πολύτιμη υποστήριξη της επιβλέπουσας καθηγήτριάς μου, της κυρίας Βασιλικής Διαμαντοπούλου. Για αυτό την ευχαριστώ θερμά για την εμπιστοσύνη που μου έδειξε εξ αρχής, αναθέτοντάς μου το συγκεκριμένο θέμα, για την επιστημονική της καθοδήγηση, τις υποδείξεις της, την επιμονή της, τη συμπαράστασή της, τη συνεχή της υποστήριξη και το αμείωτο ενδιαφέρον που έδειξε από την αρχή μέχρι το τέλος.

Επίσης θα ήθελα να ευχαριστήσω την οικογένειά μου και τους φίλους μου, που στάθηκαν δίπλα μου και τους είμαι ευγνώμων.

Περίληψη

Η Τεχνητή Νοημοσύνη (TN) εξελίσσεται ραγδαία και αναμφισβήτητα αποτελεί μία από τις σημαντικότερες τεχνολογικές επαναστάσεις της εποχής μας. Δίκαια θεωρείται μία από τις πιο σπουδαίες επιστήμες των υπολογιστών, καθώς είναι η πρώτη επιστήμη στον χώρο της πληροφορικής που μιμείται την ανθρώπινη σκέψη. Καθιστά τις μηχανές αυτόματες, προσφέροντας εργαλεία για την κατανόηση της ανθρώπινης σκέψης και της λειτουργίας του εγκεφάλου. Μέσα από τη μελέτη της TN, οι επιστήμονες μπορούν να μάθουν περισσότερα για το πώς λειτουργεί η ανθρώπινη νοημοσύνη και να αναπτύξουν τεχνολογίες που τη μιμούνται.

Η αλματώδης πρόοδος στην υπολογιστική ισχύ, οι νέες αλγοριθμικές μέθοδοι και η συλλογή μεγάλων δεδομένων έχουν επιτρέψει στην TN να διεισδύσει σε πολλούς τομείς της καθημερινότητας και της επιστήμης. Από την ιατρική διάγνωση και την αυτοματοποίηση της παραγωγής μέχρι την ανάλυση δεδομένων και τις έξυπνες πόλεις, η TN διαμορφώνει έναν νέο κόσμο όπου οι διαδικασίες γίνονται πιο αποδοτικές, ακριβείς και προσαρμοσμένες στις ανάγκες του χρήστη. Καθώς η τεχνολογία προχωρά, η παρουσία της TN γίνεται ολοένα και πιο αισθητή στην καθημερινή μας ζωή, επηρεάζοντας τον τρόπο με τον οποίο εργαζόμαστε, επικοινωνούμε και παίρνουμε αποφάσεις. Ωστόσο παρά τα επιτεύγματά της σε διάφορους τομείς, επιφέρει αρκετά ρίσκα.

Στο πλαίσιο αυτό, η παρούσα διπλωματική εργασία εξετάζει τις συνεισφορές της Τεχνητής Νοημοσύνης (TN) σε διάφορους τομείς, με ιδιαίτερη έμφαση στους πιθανούς κινδύνους που σχετίζονται με τη χρήση της, ιδίως σε θέματα προστασίας της ιδιωτικότητας και της κυβερνοασφάλειας.

Αναλυτικότερα, παρουσιάζεται εκτενώς η διαδικασία διαχείρισης δεδομένων, κατά την οποία τα μεταδεδομένα συλλέγονται, επεξεργάζονται και μετατρέπονται σε γνώση, έτοιμη να υποστηρίξει την εκπαίδευση εξελιγμένων συστημάτων TN. Αυτή η διαδικασία επιτρέπει την κατανόηση του τρόπου με τον οποίο εκπαιδεύεται η Τεχνητή Νοημοσύνη, ενώ ταυτόχρονα αναδεικνύει πως ακόμη και μια μικρή ευπάθεια μπορεί να οδηγήσει σε παραβίαση της ιδιωτικότητας των χρηστών.

Επιπλέον, στην παρούσα διπλωματική εργασία, αναδεικνύεται η αξία της προστασίας της ανθρώπινης ιδιωτικότητας και της κυβερνοασφάλειας. Μια τέτοια επιστημονική εξέλιξη δεν θα πρέπει να λειτουργεί εις βάρος της ανθρωπότητας, αλλά προς όφελός της.

Abstract

Artificial Intelligence (AI) is rapidly evolving and undoubtedly constitutes one of the most significant technological revolutions of our time. It is rightly considered one of the most important fields in computer science, as it is the first discipline in the information technology sector that mimics human thought. AI makes machines autonomous, providing tools for understanding human thinking and brain function. Through the study of AI, scientists can learn more about how human intelligence works and develop technologies that mimic it.

The exponential progress in computational power, new algorithmic methods, and the collection of big data have allowed AI to penetrate many fields of everyday life and science. From medical diagnosis and production automation to data analysis and smart cities, AI is shaping a new world where processes become more efficient, accurate, and tailored to user needs. As technology advances, the presence of AI becomes increasingly evident in our daily lives, affecting how we work, communicate, and make decisions. However, despite its achievements in various fields, it also brings significant risks.

In this context, the present thesis examines the contributions of Artificial Intelligence (AI) in various sectors, with a particular focus on the potential risks associated with its use, especially in matters of privacy protection and cybersecurity.

More specifically, the thesis extensively presents the data management process, during which metadata are collected, processed and transformed into knowledge, ready to support the training of advanced AI systems. This process allows an understanding of how AI is trained while simultaneously highlighting how even a small vulnerability can lead to breaches of user privacy.

Furthermore, this thesis emphasizes the importance of protecting human privacy and cybersecurity. Such scientific advancements should not operate to the detriment of humanity but for its benefit.

Περιεχόμενα

ΚΕΦΑΛΑΙΟ 1	7
1.1 ΕΙΣΑΓΩΓΗ	7
1.2 ΑΝΤΙΚΕΙΜΕΝΟ ΔΙΠΛΩΜΑΤΙΚΗΣ	7
1.3 ΜΕΘΟΔΟΛΟΓΙΑ	9
1.4 ΔΟΜΗ	9
ΚΕΦΑΛΑΙΟ 2 – ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ	11
2.1 ΟΡΙΣΜΟΣ	11
2.2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ	13
2.2.1 ΕΠΟΠΤΕΥΟΜΕΝΗ ΜΑΘΗΣΗ	14
2.2.2 ΜΑΘΗΣΗ ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ	14
2.2.3 ΕΝΙΣΧΥΤΙΚΗ ΜΑΘΗΣΗ	15
2.2.4 ΒΑΘΙΑ ΜΑΘΗΣΗ	16
ΚΕΦΑΛΑΙΟ 3 - ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΜΗΧΑΝΕΣ ΑΙ	18
3.1 ΔΙΑΔΙΚΑΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΑ ΣΥΣΤΗΜΑΤΑ ΑΙ	19
3.1.1 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ	19
3.1.2 ΑΠΟΘΗΚΕΥΣΗ ΔΕΔΟΜΕΝΩΝ	27
3.1.3 ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	30
3.1.4 ΔΙΑΝΟΜΗ ΚΑΙ ΚΟΙΝΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ	32
3.1.5 ΑΡΧΕΙΟΘΕΤΗΣΗ ΔΕΔΟΜΕΝΩΝ	34
3.1.6 ΔΙΑΓΡΑΦΗ/ΚΑΤΑΣΤΡΟΦΗ ΔΕΔΟΜΕΝΩΝ	35
3.2 ΠΡΟΚΛΗΣΕΙΣ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ	36
3.3 ΕΡΓΑΛΕΙΑ ΕΠΕΞΕΡΓΑΣΙΑΣ ΠΡΟΣΩΠΙΚΩΝ ΔΕΔΟΜΕΝΩΝ	38
3.4 ΣΥΜΠΕΡΑΣΜΑΤΑ	38
ΚΕΦΑΛΑΙΟ 4 - Η ΕΠΙΔΡΑΣΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΣΥΝΗΣ ΣΤΟΝ ΚΥΒΕΡΝΟΧΩΡΟ	40
4.1 ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ	40
4.1.1 ΟΡΙΣΜΟΣ	40
4.2 ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΙ ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ	42
4.2.1 ΚΙΝΔΥΝΟΙ ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑΣ ΑΠΟ ΕΦΑΡΜΟΓΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ	43
4.2.2 ΤΕΧΝΙΚΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΙΣ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΙΣ ΚΑΙ ΤΗΝ ΚΥΒΕΡΝΟΑΜΥΝΑ	53
ΚΕΦΑΛΑΙΟ 5- ΨΗΦΙΑΚΗ ΙΔΙΩΤΙΚΟΤΗΤΑ ΤΗ ΣΥΧΡΟΝΗ ΕΠΟΧΗ ΤΩΝ ΔΙΑΣΥΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ	57
5.1 ΙΔΙΩΤΙΚΟΤΗΤΑ ΣΤΟ ΔΙΑΔΙΚΤΥΟ	57

5.1.1 ΟΡΙΣΜΟΣ	57
5.1.2 ΣΗΜΑΝΤΙΚΟΤΗΤΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΤΟ ΔΙΑΔΙΚΤΥΟ	59
5.1.3 ΝΟΜΟΘΕΣΙΑ ΚΑΙ ΠΡΟΣΤΑΣΙΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΤΟ ΔΙΑΔΥΚΤΙΟ	60
5.2 ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΙ ΙΔΙΩΤΙΚΟΤΗΤΑ	62
5.2.1 ΑΝΑΔΥΟΜΕΝΕΣ ΠΡΟΚΛΗΣΕΙΣ ΚΑΙ ΚΙΝΔΥΝΟΙ ΠΑΡΑΒΙΑΣΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΑΠΟ ΕΦΑΡΜΟΓΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ	62
ΚΕΦΑΛΑΙΟ 6 - ΣΥΜΠΕΡΑΣΜΑΤΑ	69
ΑΝΑΦΟΡΕΣ	72

ΚΕΦΑΛΑΙΟ 1

1.1 ΕΙΣΑΓΩΓΗ

Στη σύγχρονη εποχή, πολλές νέες εφαρμογές της τεχνητής νοημοσύνης έχουν αναδειχθεί, όπως τα αυτόνομα οχήματα, η προσωποποιημένη ιατρική, οι ευφυείς βοηθοί φωνής και η ανάλυση μεγάλων δεδομένων. Από την αυτοματοποίηση των βιομηχανικών διαδικασιών μέχρι την εξατομικευμένη εμπειρία χρήστη στο διαδίκτυο, η Τεχνητή Νοημοσύνη διαμορφώνει το μέλλον μας με τρόπους πρωτοφανείς. Ωστόσο, παρά τα αδιαμφισβήτητα οφέλη της, η αλματώδης ανάπτυξη της Τεχνητής Νοημοσύνης συνοδεύεται από σοβαρές προκλήσεις και κινδύνους.

Είναι γεγονός ότι, οι κυβερνοεπιθέσεις έχουν σημειώσει κατακόρυφη αύξηση λόγω της χρήσης της τεχνητής νοημοσύνης από κακόβουλους φορείς που εκμεταλλεύονται τις δυνατότητές της για να διαπεράσουν τα συστήματα ασφαλείας και να αποκτήσουν πρόσβαση σε ευαίσθητα δεδομένα. Παράλληλα, η συλλογή και ανάλυση τεράστιων ποσοτήτων δεδομένων από συστήματα τεχνητής νοημοσύνης θέτει σε κίνδυνο την ψηφιακή ιδιωτικότητα των χρηστών, καθώς προσωπικά δεδομένα μπορούν να εκτεθούν ή να χρησιμοποιηθούν χωρίς τη συγκατάθεση των ατόμων.

Επομένως όλες αυτές οι τεχνολογικές καινοτομίες έχουν διαμορφώσει μια νέα πραγματικότητα. Εγείρουν όμως καίρια ερωτήματα σχετικά με το πώς η τεχνητή νοημοσύνη μπορεί να αποτελεί μια επιστημονική ανακάλυψη που θα βελτιώνει την καθημερινότητα και θα λειτουργεί ως πολύτιμο εργαλείο της επιστήμης, χωρίς να παραβιάζει την ιδιωτικότητα και να απειλεί την ανθρώπινη ασφάλεια ή την κυβερνοασφάλεια.

Η ενσωμάτωση της τεχνητής νοημοσύνης σε κρίσιμες εφαρμογές απαιτεί αυστηρούς ελέγχους και κανονισμούς, ώστε να διασφαλιστεί ότι τα οφέλη της θα μεγιστοποιηθούν, ενώ οι κίνδυνοι θα ελαχιστοποιηθούν. Έτσι, διατηρείται η ισορροπία μεταξύ τεχνολογικής προόδου και κοινωνικής υπευθυνότητας, εξασφαλίζοντας ένα μέλλον όπου η καινοτομία υπηρετεί το κοινό καλό.

1.2 ΑΝΤΙΚΕΙΜΕΝΟ ΔΙΠΛΩΜΑΤΙΚΗΣ

Η εν λόγω διπλωματική εργασία εμβαθύνει στην ανάλυση των πιθανών κινδύνων, απόρροια της διαρκούς εξέλιξης της τεχνητής νοημοσύνης και των αναδυόμενων εφαρμογών της. Η σημασία αυτής της ανάλυσης πηγάζει από την ολοένα και αυξανόμενη ανάγκη για προστασία από κυβερνοεπιθέσεις, αλλά και τη συχνή ανησυχία των χρηστών για την παραβίαση της ιδιωτικότητάς τους.

Η επιλογή του θέματος της διπλωματικής εργασίας, προέκυψε μετά από προσεκτική και βαθιά σκέψη, καθώς συνδυάζει δύο τομείς που βρίσκονται στην

αιχμή της τεχνολογικής προόδου: Την ιδιωτικότητα και την κυβερνοασφάλεια στην εποχή της τεχνητής νοημοσύνης. Η συνεχής ανάπτυξη της τεχνητής νοημοσύνης έχει επιφέρει σημαντικές αλλαγές στη χρήση και την προστασία των προσωπικών δεδομένων, καθιστώντας αναγκαία την εξερεύνηση νέων μεθόδων και εργαλείων για την ασφάλεια των πληροφοριών.

Η βαρύτητα του συγκεκριμένου θέματος είναι πολυδιάστατη. Αφενός, το ακαδημαϊκό ενδιαφέρον είναι τεράστιο, καθώς το ζήτημα της ιδιωτικότητας και της ασφάλειας αποτελεί έναν από τους πλέον κρίσιμους παράγοντες στην ανάπτυξη των τεχνολογικών εφαρμογών. Αφετέρου, οι επαγγελματικές προοπτικές που αναδεικνύονται από τη μελέτη και την κατανόηση αυτών των κινδύνων είναι εξαιρετικά σημαντικές, ανοίγοντας νέους δρόμους για καινοτόμες λύσεις και πρακτικές στον τομέα της κυβερνοασφάλειας.

Σκοπός της μελέτης είναι να συμβάλει στην καλύτερη κατανόηση των κινδύνων που συνεπάγεται η χρήση της τεχνητής νοημοσύνης και τα αποτελέσματά τους σε πλαίσιο κοινωνικό, ανθρωπιστικό και επιστημονικό, αλλά και να προταθούν πιθανές λύσεις.

Ακόμη, η μελέτη στοχεύει να βρει μια ορθή και ολοκληρωμένη απάντηση σε βασικά ζητήματα και προβληματισμούς, τα οποία είναι τα εξής:

1. Σε ποιο βαθμό η νομοθεσία προστατεύει τους χρήστες από παράνομες επιθέσεις ή μη εξουσιοδοτημένη πρόσβαση στα προσωπικά τους δεδομένα;
2. Σε ποιο βαθμό μπορούν οι χρήστες να γνωρίζουν αν οι ηλεκτρονικές πλατφόρμες, οι ιστοσελίδες και οι οργανισμοί τηρούν τη νομοθεσία;
3. Γιατί εγείρονται συνεχώς ανησυχίες από ένα μεγάλο ποσοστό ανθρώπων ότι απειλείται η ιδιωτικότητα τους;
4. Πού θα μπορούσε να οδηγήσει η παραβίαση της ιδιωτικότητας;
5. Ποιες είναι οι μέθοδοι τεχνητής νοημοσύνης που χρησιμοποιούνται σε κυβερνοεπιθέσεις;
6. Υπάρχουν μέθοδοι και εφαρμογές τεχνητής νοημοσύνης για την προστασία από σύγχρονες κυβερνοεπιθέσεις που βασίζονται στην τεχνητή νοημοσύνη;
7. Τελικά, η τεχνητή νοημοσύνη είναι μια επιστήμη που ωφελεί τον άνθρωπο ή είναι μια επιστήμη που θέτει την ανθρωπότητα σε μεγάλους κινδύνους;
8. Ποια θα είναι η μελλοντική εξέλιξη της τεχνητής νοημοσύνης και κατά πόσο θα αμβλύνει ή όχι τα προαναφερόμενα προβλήματα στο πλαίσιο ιδιωτικότητας και κυβερνοασφάλειας;

1.3 ΜΕΘΟΔΟΛΟΓΙΑ

Η συγγραφή της διπλωματικής εργασίας ξεκίνησε τον Νοέμβριο του 2023, έπειτα από συζήτηση με την επιβλέπουσα καθηγήτρια για την επιλογή του θέματος. Από εκείνο το σημείο, άρχισε η συλλογή βιβλιογραφικών πηγών. Για την εκπόνηση της μελέτης χρησιμοποιήθηκε ένας μεγάλος αριθμός βιβλιογραφικών πηγών, οι οποίες αξιολογήθηκαν προσεκτικά, λαμβάνοντας υπόψη την εγκυρότητα και τη συμβολή τους στο θέμα της εργασίας.

Η επιλογή των πηγών έγινε με βάση τις τρεις θεματικές ενότητες του αντικείμενου της εργασίας: Την τεχνητή νοημοσύνη, τη διαδικασία συλλογής και επεξεργασίας των δεδομένων, τους τελικούς κινδύνους που προκύπτουν από την ανάπτυξη της τεχνητής νοημοσύνης στην ιδιωτικότητα και την κυβερνοασφάλεια. Αναζητήθηκαν πηγές που κάλυπταν την κάθε θεματολογία ξεχωριστά, καθώς και πηγές που ανέδειξαν τη σύνδεση και τις αλληλεπιδράσεις μεταξύ τους.

Με συστηματική και μεθοδική προσέγγιση, όλες οι βιβλιογραφικές πηγές διερευνήθηκαν, ταξινομήθηκαν και αξιολογήθηκαν. Αφαιρέθηκαν όσες δεν πληρούσαν τα κριτήρια εγκυρότητας και αντικαταστάθηκαν με πιο ακριβείς και έγκυρες πηγές. Όλες οι πηγές που χρησιμοποιήθηκαν αναφέρονται στο τέλος της εργασίας, διευκολύνοντας την αναζήτησή τους.

Στο διάστημα Ιουλίου και Αυγούστου του 2024, πραγματοποιήθηκαν οι τελικές διορθώσεις υπό την επίβλεψη της επιβλέπουσας καθηγήτριας και ολοκληρώθηκε η συγγραφή της εργασίας.

1.4 ΔΟΜΗ

Κεφάλαιο 1 : Η δομή της εργασίας ξεκινάει από το κεφάλαιο 1, που αποτελεί την εισαγωγή της εργασίας. Σε αυτό το κεφάλαιο παρουσιάζεται το αντικείμενο της μελέτης και γίνεται μια ιστορική αναδρομή από το παρελθόν μέχρι το σήμερα. Επίσης, εξηγούνται τα κίνητρα και οι λόγοι που οδήγησαν στην επιλογή αυτής της μελέτης, καθώς και τα ερευνητικά ερωτήματα που θα αναλυθούν στη συνέχεια της εργασίας. Τέλος, περιγράφεται αναλυτικά η μεθοδολογία που ακολουθήθηκε για την ανάπτυξη της εργασίας.

Κεφάλαιο 2 - ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ : Σε αυτό το σημείο της διπλωματικής γίνεται μια γενική αναφορά για την τεχνολογία της τεχνητής νοημοσύνης, της μηχανικής μάθησης και των υποκατηγοριών της οι οποίες είναι η εποπτευόμενη μάθηση, η μάθηση χωρίς επίβλεψη, η ενισχυτική μάθηση και βαθιά μάθηση.

Κεφάλαιο 3 - ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΜΗΧΑΝΕΣ ΑΙ : Στο κεφάλαιο αυτό γίνεται εκτενής ανάλυση του τρόπου με τον οποίο τα συστήματα τεχνητής νοημοσύνης αποκτούν πρόσβαση στα δεδομένα των χρηστών. Αναλύεται όλη η διαδικασία μετατροπής των δεδομένων σε γνώση

για την εκπαίδευση των αλγορίθμων. Αυτό το κεφάλαιο αποτελεί τη βάση για την κατανόηση της ανάλυσης των κινδύνων που θα εξεταστούν στα επόμενα δύο κεφάλαια.

Κεφάλαιο 4 - ΨΗΦΙΑΚΗ ΙΔΙΩΤΙΚΟΤΗΤΑ ΤΗ ΣΥΧΡΟΝΗ ΕΠΟΧΗ ΤΩΝ ΔΙΑΣΥΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ : Σε αυτό το κεφάλαιο, ορίζουμε την έννοια της ιδιωτικότητας στο διαδίκτυο και εξετάζουμε γιατί η προστασία της είναι κρίσιμη και απαραίτητη. Αναλύονται τα κανονιστικά πλαίσια που θεσπίζει η νομοθεσία για την προστασία της ιδιωτικότητας. Τέλος, παρουσιάζονται εκτενώς οι νέες προκλήσεις και οι κίνδυνοι παραβίασης της ιδιωτικότητας από εφαρμογές και συστήματα τεχνητής νοημοσύνης .

Κεφάλαιο 5 : Το Κεφάλαιο 5 αναλύει πώς η άνθιση της τεχνητής νοημοσύνης έχει δημιουργήσει σύγχρονες απειλές στον κυβερνοχώρο. Παρουσιάζονται οι τρέχουσες απειλές, οι ευπάθειες και οι κυβερνο-επιθέσεις που πλήττουν την κυβερνοασφάλεια και έχουν προκύψει από την εξέλιξη των εφαρμογών ΑΙ. Επιπλέον, επισημαίνονται ορισμένες τεχνικές τεχνητής νοημοσύνης που έχουν προταθεί ως λύση για την αντιμετώπιση των κυβερνοεπιθέσεων και την ενίσχυση της κυβερνοάμυνας.

Η διπλωματική εργασία ολοκληρώνεται με την παρουσίαση των συμπερασμάτων που προκύπτουν από όλη τη μελέτη. Παρουσιάζονται επίσης μερικές προτεινόμενες λύσεις για την αντιμετώπιση των προβλημάτων που αναλύθηκαν στα Κεφάλαια 4 και 5, καθώς και οι μελλοντικές προβλέψεις που βασίζονται στη μελέτη. Επιπλέον, παρέχονται αναφορές στις πηγές που χρησιμοποιήθηκαν για την εκπόνηση της εργασίας.

ΚΕΦΑΛΑΙΟ 2 – ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

2.1 ΟΡΙΣΜΟΣ

Η τεχνητή νοημοσύνη είναι ένας κλάδος της επιστήμης των υπολογιστών ο οποίος ασχολείται με τη δημιουργία υπολογιστικών συστημάτων τα οποία εκτελούν διεργασίες που απαιτούν ανθρώπινη νοημοσύνη. Η τεχνητή νοημοσύνη καθιστά τις μηχανές ικανές για να αναπαράγουν τις βασικές γνωστικές λειτουργίες ενός ανθρώπου, όπως είναι η εκμάθηση, η προσαρμοστικότητα, γλωσσική κατανόηση και η αλληλεπίδραση με το περιβάλλον, η λήψη αποφάσεων, άλλα και επίλυση προβλημάτων [1].

Η ύπαρξη της τεχνητής νοημοσύνης επηρεάζει κάθε πτυχή της καθημερινότητας και όλα αυτά επιτυγχάνονται με τη λήψη και την επεξεργασία μεγάλου όγκου δεδομένων από διάφορες πηγές. Τα δεδομένα αυτά αναλύονται, μετατρέπονται σε πληροφορίες και έπειτα χρησιμοποιούνται, με σκοπό την εκπαίδευση των υπολογιστικών συστημάτων, τη βελτίωση της αποτελεσματικότητας αλλά και τη βέλτιστη λήψη αποφάσεων .

Τη σημερινή εποχή υπάρχει αυξημένος πλέον όγκος δεδομένων από πολλαπλές πηγές και αυτό αποτελεί έναν από τους κύριους λόγους που η τεχνητή νοημοσύνη έχει εξελιχθεί περισσότερο τα τελευταία χρόνια, ενώ είναι ένας τομέας επιστήμης των υπολογιστών που υπάρχει από τα μέσα του 20ού αιώνα. Τότε οι πρωτοπόροι της πληροφορικής άρχισαν να εξερευνούν την ιδέα της δημιουργίας μηχανών ικανών να εκτελούν εργασίες που απαιτούν ανθρώπινη νοημοσύνη[1]. Η έννοια της τεχνητής νοημοσύνης εκείνη την περίοδο ήταν στα πρώτα της στάδια και αντιμετωπιζόταν κυρίως ως ένα φιλόδοξο επιστημονικό όραμα παρά ως μια πραγματική τεχνολογία.

Η επίσημη γέννηση της τεχνητής νοημοσύνης τοποθετείται το 1956, όταν μια ομάδα κορυφαίων επιστημόνων, μεταξύ των οποίων οι John McCarthy, Marvin Minsky, Nathaniel Rochester και Claude Shannon, συναντήθηκαν στο περίφημο συνέδριο του Dartmouth [59]. Εκεί τέθηκαν οι βάσεις για την ανάπτυξη συστημάτων που θα μπορούσαν να επιλύουν προβλήματα, να μαθαίνουν και να προσαρμόζονται μέσω αλγορίθμων και υπολογιστικών μοντέλων. Κατά τη διάρκεια του συνεδρίου του Dartmouth επίσης συζητήθηκε για πρώτη φορά την ιδέα ενός μηχανήματος που θα μπορούσε να σκεφτεί όπως ο άνθρωπος.

Ωστόσο παρά το γεγονός πως ο επίσημος όρος της τεχνητής νοημοσύνης κάνει για πρώτη φορά την εμφάνισή του στο ακαδημαϊκό συνέδριο του Dartmouth το 1956, το ερώτημα σχετικά με τη δυνατότητα των μηχανών να προσομοιώσουν την ανθρώπινη λογική, αποτελούσε θέμα των ακαδημαϊκών κύκλων για περίπου δέκα χρόνια πριν την εμφάνισή του. Στο σύγγραμμα με τίτλο «As we may think» [8] ο τότε μηχανικός Vannevar Bush προτείνει ένα σύστημα το οποίο διευρύνει τη γνωστική ικανότητα των ανθρώπων, ακολουθούμενος από τη δημοσίευση του

Alan Turing [74] σχετικά με ένα σύστημα το οποίο δύναται να μιμείται την ανθρώπινη ευφυΐα .

Ως κριτήριο σχετικά με την ικανότητα των μηχανών να σκέφτονται κατά τον ανθρώπινο πάντα ορισμό της έννοιας της σκέψης, δημιουργήθηκε από τον Turing το 1950 η λεγόμενη δοκιμή Turing[74] . Σκοπός της εν λόγω δοκιμασίας ήταν η αξιολόγηση λογισμικών σχετικά με την ικανότητά τους να μιμηθούν την ανθρώπινη σκέψη , διατηρώντας μια συζήτηση με έναν άνθρωπο δυσχεραίνοντάς την ικανότητά του να αποφανθεί σχετικά με το αν οι απαντήσεις που δέχεται προέρχονται από άλλο άνθρωπο ή από πρόγραμμα [74].

Μέχρι στιγμής δεν έχει βρεθεί βέβαια κάποιος αλγόριθμος ικανός να περάσει τη δοκιμή Turing. Απλοποιήσεις της δοκιμής έχουν οδηγήσει σε ορισμένα μοντέλα που μπορούν να παράγουν αφηρημένους άνευ συνοχής διαλόγους που όμως εύκολα γίνεται αντιληπτό πως προέρχονται από πρόγραμμα[74].

Μια από τις αδυναμίες βέβαια της δοκιμής Turing είναι ο διαχωρισμός της απλής λογικής από τη συνείδηση, και το κατά πόσο μια μηχανή είναι ικανή για τη δεύτερη. Η αξιολόγηση Turing χρησιμοποιεί ως μόνο κριτήριο τη συνέχεια ενός διαλόγου και τη νοηματική, και συντακτική του αρτιότητα καθώς και λογική συνέχεια, χωρίς να υπάρχει τρόπος να αξιολογηθεί η έννοια της συνείδησης. Ως αντεπιχείρημα θα μπορούσε να επισημανθεί το παράδειγμα των μωρών τα οποία αδυνατούν να περάσουν τη δοκιμή, δεν θα μπορούσαν όμως με βάση αυτό το κριτήριο να χαρακτηριστούν ως μη συνειδητά όντα [23].

Παρά το μέχρι τώρα απροσπέλαστο της δοκιμής Turing, διάφοροι αλγόριθμοι και μοντέλα μηχανικής μάθησης – τεχνητής νοημοσύνης έχουν χρησιμοποιηθεί σε διάφορες εφαρμογές. Υπάρχουν διάφορες κατηγορίες αλγορίθμων (νευρωνικά δίκτυα, αλγόριθμοι ταξινόμησης, μοντέλα πρόβλεψης χρονοσειρών κ.α) τα οποία βρίσκουν εφαρμογή σε διάφορους επιστημονικούς τομείς από τους οποίους προέρχεται τεχνογνωσία για την επίλυση διαφόρων καθημερινών και άλλων προβλημάτων.

Συνοπτικά απο 1950 και μέχρι σήμερα, η τεχνητή νοημοσύνη έχει διανύσει μια εντυπωσιακή πορεία εξέλιξης. Οι πρώτες δεκαετίες, χαρακτηρίστηκαν από αισιοδοξία και σημαντικές προόδους, όπως τα συστήματα αυτόματης απόδειξης θεωρημάτων και τα πρώτα δειλά δειλά αναπτυγμένα νευρωνικά δίκτυα [2]. Ωστόσο , οι περιορισμοί της υπολογιστικής ισχύος και η πολυπλοκότητα των πραγματικών προβλημάτων οδήγησαν σε περιόδους απογοήτευσης και μείωσης των επενδύσεων, γνωστές ως "χειμώνες της τεχνητής νοημοσύνης" [2].

Παρά τις δυσκολίες, η συνεχής βελτίωση των υπολογιστικών πόρων και οι ανακαλύψεις σε τομείς όπως η μηχανική μάθηση και η επεξεργασία φυσικής γλώσσας, έχουν επαναφέρει την τεχνητή νοημοσύνη στο προσκήνιο, καθιστώντας την αναπόσπαστο μέρος της σύγχρονης τεχνολογίας και επιστήμης όπως γνωρίζουν όλοι σήμερα.

2.2 ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

Ένα από τα σημαντικά τμήματα της τεχνητής νοημοσύνης το οποίο εστιάζει στη χρήση δεδομένων για την εκπαίδευση μοντέλων, όπως είναι οι αλγόριθμοι και τα υπολογιστικά συστήματα, είναι η μηχανική μάθηση. Στη μηχανική μάθηση, τα μοντέλα εκπαιδεύονται από τα δεδομένα και βελτιώνονται με την εμπειρία αντί να προγραμματιστούν εξ αρχής [3].

Τα μοντέλα βρίσκουν μοτίβα και συσχετίσεις σε μεγάλα σύνολα δεδομένων και λαμβάνουν τις καλύτερες αποφάσεις και προβλέψεις βάσει αναλύσεων. Με τη χρήση τεχνικών ανάλυσης όπως η ταξινόμηση, η παλινδρόμηση, η ομαδοποίηση, η ανίχνευση ανωμαλιών, η ανάλυση χρονολογικής σειράς και άλλες δημιουργούνται μοντέλα που μπορούν να αναγνωρίζουν μοτίβα, να επεξεργάζονται φυσική γλώσσα να κάνουν προβλέψεις και να λαμβάνουν αποφάσεις βασισμένες σε δεδομένα, οδηγώντας σε πιο αποτελεσματικές και ακριβείς επιχειρηματικές στρατηγικές.

Καθώς τα μοντέλα εκτίθενται σε περισσότερα δεδομένα, βελτιώνονται, καθιστώντας τις προβλέψεις και τις αποφάσεις τους πιο ακριβείς με την πάροδο του χρόνου. Ωστόσο, η επιτυχία της μηχανικής μάθησης εξαρτάται από την ποιότητα και την ποσότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση των μοντέλων, καθώς και από την επιλογή των κατάλληλων αλγορίθμων και παραμέτρων.

Η μηχανική μάθηση είναι μία από τις πιο σημαντικές εξελίξεις της τεχνολογίας που έχει επηρεάσει πολλούς τομείς της καθημερινής μας ζωής. Συγκεκριμένα, έχει ενσωματωθεί σε εφαρμογές υγείας, όπου μπορεί να χρησιμοποιηθεί για την πρόβλεψη ή τη διάγνωση ασθενειών με μεγαλύτερη ακρίβεια και ταχύτητα. Επίσης, η μηχανική μάθηση έχει εφαρμοστεί επιτυχώς σε αυτόνομα οχήματα, βελτιώνοντας την ασφάλεια και την απόδοσή τους στον δρόμο. Επιπλέον, η μηχανική μάθηση έχει προοπτικές εξέλιξης και εφαρμογής σε πολλούς άλλους τομείς. Με την πρόοδο της τεχνολογίας, αναμένεται να επεκταθεί ακόμα περισσότερο σε κλάδους όπως η γεωπληροφορική, η ρομποτική, η οικονομία και πολλούς ακόμα.

Υπάρχουν τρία βασικά μοντέλα μηχανικής μάθησης τα οποία είναι η εποπτευόμενη μάθηση, η μάθηση χωρίς επίβλεψη και η ενισχυτική μάθηση [3]. Τα μοντέλα αυτά αναλύονται παρακάτω, στις επόμενες ενότητες. Κάθε ένα από αυτά τα μοντέλα έχει διαφορετικές εφαρμογές και προτεραιότητες ανάλογα με τον τομέα χρήσης του.

Επιπλέον, τα μοντέλα μηχανικής μάθησης έχουν σημαντική επίδραση στη διαχείριση δεδομένων και την ασφάλεια στον κυβερνοχώρο. Η ανάλυση δεδομένων μέσω αυτών των μοντέλων μπορεί να βοηθήσει στην ανίχνευση απειλών και στην πρόληψη κυβερνοεπιθέσεων. Ωστόσο, η χρήση τους αναδεικνύει ερωτήματα σχετικά με την ιδιωτικότητα των χρηστών, καθώς οι

αλγόριθμοι λειτουργίας τους μπορεί να επηρεάσουν την προστασία των προσωπικών δεδομένων.

2.2.1 ΕΠΟΠΤΕΥΟΜΕΝΗ ΜΑΘΗΣΗ

Η εποπτευόμενη μάθηση βασίζεται στην εκπαίδευση αλγορίθμων με επισημασμένα δεδομένα. Ο αλγόριθμος μαθαίνει να διαχειρίζεται και να οργανώνει τα δεδομένα με βάση τη σχέση μεταξύ των δεδομένων εισόδου και εξόδου. Σκοπός της εποπτευόμενης μάθησης είναι να εκπαιδεύσει το μοντέλο να προβλέπει σωστά την έξοδο για νέα και άγνωστα δεδομένα εισόδου. Συχνά απαιτείται ανθρώπινη καθοδήγηση για να βοηθήσει τους αλγόριθμους κατά τη διάρκεια της εκπαίδευσης τους. Τα μοντέλα πρόβλεψης συχνά χρησιμοποιούν τεχνικές εποπτευόμενης μηχανικής μάθησης. Για παράδειγμα, στην πρόβλεψη καιρού, στα αυτόματα οχήματα, στην αναγνώριση προτύπων, στην επεξεργασία φυσικής γλώσσας, καθιστώντας τη μία από τις πιο βασικές τεχνικές στον τομέα της τεχνητής νοημοσύνης.

Στην κυβερνοασφάλεια γίνεται χρήση εποπτευόμενης μάθησης για την ανάλυση και ταξινόμηση δεδομένων σε επιβλαβή ή μη, αποτρέποντας μελλοντικές επιθέσεις και ανιχνεύοντας απειλές όπως για παράδειγμα οι επιθέσεις άρνησης εξυπηρέτησης (DDos) [5].

2.2.2 ΜΑΘΗΣΗ ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ

Η μάθηση χωρίς επίβλεψη αποτελεί επίσης ένα μοντέλο μηχανικής μάθησης, ωστόσο σε αντίθεση με την εποπτευόμενη μάθηση, οι αλγόριθμοι μαθαίνουν από τα δεδομένα χωρίς την ύπαρξη ανθρώπινης καθοδήγησης.

Παρόλο που οι υπερπαράμετροι (όπως είναι ο αριθμός των σημείων συμπλέγματος) του μοντέλου καθορίζονται από κάποιον άνθρωπο, ο αλγόριθμος θα επεξεργαστεί μόνος του τα δεδομένα [5].

Επίσης τα δεδομένα είναι ακατέργαστα, δεν διαθέτουν ετικέτες και η εκπαίδευση των αλγορίθμων είναι εντελώς αυτόνομη χωρίς την ύπαρξη υποδείξεων. Αυτό προσφέρει μεγαλύτερη ευελιξία στη χρήση ανεπεξέργαστων δεδομένων για την ανάλυση και την ανακάλυψη νέων πληροφοριών. Επιπλέον οι αλγόριθμοι μη εποπτευόμενης μάθησης είναι χρησιμότεροι για πιο πολύπλοκες εργασίες όπως η ομαδοποίηση μεγάλου όγκου δεδομένων σε ομάδες (cluster). Επίσης οι αλγόριθμοι αυτοί μπορούν να ανακαλύψουν κρυφές συσχετίσεις που δεν είναι προκαθορισμένες από ανθρώπινη καθοδήγηση και να παρέχουν πολύτιμες πληροφορίες για τη δομή των δεδομένων. Συνεπώς, η μη εποπτευόμενη μάθηση είναι κατάλληλη για την ανάδειξη αόρατων τάσεων και σχέσεων μεταξύ των δεδομένων, ωστόσο, λόγω της μικρότερης ανθρώπινης επίβλεψης, η εξήγηση των αποτελεσμάτων απαιτεί επιπλέον προσοχή [5].

Η χρήση μάθησης χωρίς επίβλεψη χρησιμοποιείται κυρίως για την πρόβλεψη αναγκών και προτιμήσεων του πελάτη για συγκριμένα προϊόντα και υπηρεσίες. Υπάρχουν πολλές εφαρμογές που χρησιμοποιούν ταυτόχρονα εποπτευόμενη και μη εποπτευόμενη μάθηση. Ένα παράδειγμα είναι η εφαρμογή Spotify που χρησιμοποιεί τόσο εποπτευόμενη για τη δημιουργία συστάσεων, όπως προτάσεις τραγουδιών βασισμένες στο ιστορικό ακρόασης και τις προτιμήσεις του χρήστη, αλλά και μη εποπτευόμενης μάθησης για τον εντοπισμό προτύπων ή νέων τραγουδιών, ανεξάρτητα από προηγούμενες προτιμήσεις του χρήστη.

Οι ομάδες ασφαλείας χρησιμοποιούν μάθηση χωρίς επίβλεψη για την εκπαίδευση αλγορίθμων που ανιχνεύουν καινούργιες και περίπλοκες επιθέσεις στον κυβερνοχώρο οι οποίες δεν είναι ακόμα καλά μελετημένες [5].

2.2.3 ΕΝΙΣΧΥΤΙΚΗ ΜΑΘΗΣΗ

Η ενισχυτική μάθηση είναι ένα μοντέλο της μηχανικής μάθησης στο οποίο ένα αυτόνομο σύστημα, όπως ένας πράκτορας, εκπαιδεύεται στο να προσαρμοστεί κατάλληλα στο περιβάλλον του, αλληλεπιδρώντας με αυτό και λαμβάνοντας επιβράβευση ή ποινή, ανάλογα αν οι ενέργειές που εκτέλεσε ήταν επιθυμητές ή μη. Αυτό μπορεί να γίνει μέσω της χρήσης αλγορίθμων και μοντέλων που επιτρέπουν στον πράκτορα να εξελίξει τη συμπεριφορά του με την πάροδο του χρόνου, λαμβάνοντας υπόψη τις συνέπειες των ενεργειών του.

Πιο συγκεκριμένα, κατά τη διαδικασία εκπαίδευσης, ο αλγόριθμος προσπαθεί να ελαχιστοποιήσει ή να μεγιστοποιήσει μια αντικειμενική συνάρτηση ανταμοιβής, λαμβάνοντας αποφάσεις και πραγματοποιώντας δράσεις στο περιβάλλον.

Αυτό επιτυγχάνεται μέσω της διαδικασίας δοκιμής και λάθους, όπου ο αλγόριθμος εξερευνά διάφορες ενέργειες, παρατηρεί τα αποτελέσματα και προσαρμόζει τη συμπεριφορά του ανάλογα με την επιβράβευση ή ποινή που λαμβάνει. Κύριος στόχος του αλγορίθμου είναι να μάθει μια πολιτική ενέργειας που θα του επιτρέψει να επιτύχει τη μεγαλύτερη δυνατή ανταμοιβή μελλοντικά.

Η ενισχυτική μάθηση έχει εφαρμογές σε πολλούς τομείς, συμπεριλαμβανομένων των ρομποτικών συστημάτων, της αυτόνομης πτήσης, της πολιτικής λήψης αποφάσεων και άλλων περιβαλλοντικών συστημάτων όπου η ανταμοιβή είναι μη σταθερή και πρέπει να προσαρμοστεί σε διαρκώς μεταβαλλόμενες συνθήκες.

Στην κυβερνοασφάλεια γίνεται χρήση ενισχυτικής μάθησης για την ανίχνευση μιας ευρείας γκάμας επιθέσεων στον κυβερνοχώρο. Αυτό γίνεται με την ανάλυση μεγάλου όγκου δεδομένων, την ανάλυση τους με σκοπό την αναγνώριση μοτίβων που υποδεικνύουν απειλές και την τελική πρόληψή από πιθανές μελλοντικές επιθέσεις.

2.2.4 ΒΑΘΙΑ ΜΑΘΗΣΗ

Η βαθιά μάθηση αποτελεί ένα υποσύνολο της μηχανικής μάθησης, με ειδική έμφαση στη χρήση πολυεπίπεδων μοντέλων, όπως τα βαθιά νευρωνικά δίκτυα, για την εξαγωγή πολύπλοκων αναπαραστάσεων από τα δεδομένα.

Η βαθιά μάθηση αναπαριστά μια εξέλιξη στον τομέα της τεχνητής νοημοσύνης, όπου υπολογιστικά μοντέλα αναλαμβάνουν την αυτόματη εξαγωγή χαρακτηριστικών από πολύπλοκα δεδομένα. Η βαθιά μάθηση συνήθως απαιτεί πολλά επίπεδα επεξεργασίας, καθένα από τα οποία εκπαιδεύεται να αναπαριστά χαρακτηριστικά σε διαφορετικά επίπεδα αφάιρησης [6].

Οι προγενέστερες προσεγγίσεις στη μηχανική μάθηση είχαν περιορισμένη δυνατότητα στο να επεξεργαστούν φυσικά δεδομένα στην αρχική τους μορφή, χωρίς πολύ σύνθετες μετατροπές. Επί σειρά ετών, η κατασκευή ενός συστήματος αναγνώρισης προτύπων ή μηχανικής μάθησης απαιτούσε προσεκτική μηχανική εργασία και σημαντική γνώση στον τομέα, προκειμένου να δημιουργηθεί ένας μετασχηματιστής χαρακτηριστικών που να μπορεί να μετατρέψει τα ακατέργαστα δεδομένα (όπως οι τιμές pixel μιας εικόνας) σε μια κατάλληλη εσωτερική αναπαράσταση ή διάνυσμα χαρακτηριστικών, από το οποίο το υποσύστημα μάθησης συχνά ένας ταξινομητής, θα μπορούσε να αναγνωρίσει ή να ταξινομήσει μοτίβα στην είσοδο[6].

Οι μέθοδοι βαθιάς μάθησης είναι μέθοδοι εκμάθησης αναπαράστασης, που επιτρέπουν σε μια μηχανή να τροφοδοτείται με ακατέργαστα δεδομένα και να ανακαλύπτει αυτόματα τις αναπαραστάσεις που απαιτούνται για τον εντοπισμό ή την ταξινόμηση. Επίσης διαθέτουν πολλαπλά επίπεδα αναπαράστασης, που λαμβάνονται με τη σύνθεση απλών αλλά μη γραμμικών ενοτήτων που καθεμία μετατρέπει την αναπαράσταση σε ένα επίπεδο (ξεκινώντας από την ακατέργαστη είσοδο) σε μια αναπαράσταση σε υψηλότερο, ελαφρώς πιο αφηρημένο επίπεδο. Μέσω της σύνθεσης πολλαπλών τέτοιων μετασχηματισμών, είναι δυνατόν να ανακαλύψουμε πολύπλοκες συναρτήσεις. Αυτό βοηθάει στον εντοπισμό περίπλοκων μοτίβων σε δεδομένα με υψηλές διαστάσεις. Ως εκ τούτου, η μέθοδος αυτή μπορεί να εφαρμοστεί σε πολλούς τομείς, συμπεριλαμβανομένων της επιστήμης, των επιχειρήσεων και της δημόσιας διοίκησης [6].

Η βαθιά μάθηση μπορεί να χρησιμοποιηθεί σε εποπτευόμενες μεθόδους μάθησης, όπου τα δεδομένα εκπαιδεύονται με ετικέτες, αλλά και μεθόδους μη-εποπτευόμενης μάθησης επιτρέποντας στο μοντέλο να εξάγει αυτόματα συμπεράσματα από τα δεδομένα. Αυτό καθιστά τη βαθιά μάθηση χρήσιμο εργαλείο για την ανάλυση μεγάλων δεδομένων ακόμη και όταν τα ανεπεξέργαστα δεδομένα δεν έχουν ετικέτες ή διαμερίσεις σε κατηγορίες [7].

Επομένως η βαθιά μάθηση είναι μια πολύ ισχυρή μέθοδος μηχανικής μάθησης που έχει αποδειχθεί αποτελεσματική σε πολλά προβλήματα. Ιδίως σε περιπτώσεις όπου οι δεδομένες πληροφορίες είναι πολύπλοκες ή όταν θέλουμε να αναγνωρίσουμε πρότυπα που είναι δύσκολα να περιγραφούν με παραδοσιακούς τρόπους προγραμματισμού. Επίσης παράγει εξαιρετικά υποσχόμενα

αποτελέσματα για διάφορες εργασίες στην κατανόηση φυσικής γλώσσας, την ταξινόμηση θεμάτων, την ανάλυση συναισθημάτων, την απάντηση ερωτήσεων και τη μετάφραση γλώσσας.

ΚΕΦΑΛΑΙΟ 3 - ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΥΠΟΛΟΓΙΣΤΙΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΜΗΧΑΝΕΣ ΑΙ

Η πρόοδος της τεχνητής νοημοσύνης βασίζεται στη συλλογή τεράστιων όγκων δεδομένων για την εκπαίδευσή της και την περαιτέρω ανάπτυξή της. Η διαχείριση αυτών των μεγάλων όγκων δεδομένων είναι ουσιώδης για την εξαγωγή χρήσιμης πληροφορίας και την επίτευξη ακρίβειας και αποτελεσματικότητας στις εφαρμογές της τεχνητής νοημοσύνης. Όσο πιο πλούσια και ποιοτικά είναι τα δεδομένα που διατίθενται στα μοντέλα μηχανικής μάθησης, τόσο πιο αποτελεσματικά μπορούν να εκπαιδευτούν και να λειτουργήσουν αυτά τα μοντέλα. Τη σύγχρονη εποχή, η αυξανόμενη ποσότητα δεδομένων αποτελεί ένα αναπόφευκτο φαινόμενο που ενισχύει την έννοια των big data.

Τα "big data" αποτελούν την καίρια πηγή πληροφοριών για την εκπαίδευση και την ανάπτυξη μοντέλων τεχνητής νοημοσύνης, καθώς η πολυπλοκότητα και η ποικιλία των δεδομένων συμβάλλουν στην αύξηση της ακρίβειας και της απόδοσης των μοντέλων. Τα big data χαρακτηρίζονται από σύνολα δεδομένων με εξαιρετική διαστασιμότητα και όγκο καθώς η μεγάλη διαστασιμότητα και ο μεγάλος όγκος δεδομένων που χαρακτηρίζουν πολλά σύγχρονα datasets είναι ένα χαρακτηριστικό της σύγχρονης εποχής της ψηφιακής πληροφορίας. Αυτό προκύπτει από τον αυξανόμενο όγκο των δεδομένων που δημιουργούνται και συλλέγονται κάθε μέρα από διάφορες πηγές όπως αισθητήρες, κινητά τηλέφωνα, διαδίκτυο, κοινωνικά δίκτυα, αισθητήρες IoT και πολλά άλλα. Η μεγάλη διαστασιμότητα αναφέρεται στον αριθμό των διαφορετικών χαρακτηριστικών ή επιπέδων πληροφοριών που περιλαμβάνονται σε ένα dataset. Αυτό μπορεί να σημαίνει ότι τα δεδομένα περιλαμβάνουν πολλές διαφορετικές μετρήσεις, παραμέτρους ή χαρακτηριστικά για κάθε είσοδο. Ο μεγάλος όγκος δεδομένων αναφέρεται στη συνολική ποσότητα των δεδομένων που περιέχει ένα dataset. Αυτό μπορεί να είναι σε μορφή πολλαπλών εγγραφών ή παραδειγμάτων και ο όγκος αυτός μπορεί να είναι τεράστιος, καθιστώντας την ανάλυση, την αποθήκευση και την επεξεργασία τέτοιων δεδομένων πρόκληση.

Αυτή η τάση προς τη μεγάλη διαστασιμότητα και τον μεγάλο όγκο δεδομένων επηρεάζει πολλούς τομείς, συμπεριλαμβανομένης της επιστημονικής έρευνας, της βιομηχανίας, της υγείας, της χρηματοοικονομικής ανάλυσης, της τεχνητής νοημοσύνης και πολλών άλλων. Η ανάλυση και αξιοποίηση αυτών των μεγάλων datasets απαιτεί προηγμένες τεχνικές και εργαλεία, όπως αλγόριθμοι μηχανικής μάθησης, συστήματα αποθήκευσης δεδομένων μεγάλης κλίμακας και πλατφόρμες επεξεργασίας δεδομένων.

Επομένως, τα big data επωφελούν την τεχνητή νοημοσύνη να εξελιχθεί και να προοδεύσει σε νέα επίπεδα απόδοσης και λειτουργικότητας. Ωστόσο, αυτή η ραγδαία αύξηση στον όγκο δεδομένων δημιουργεί ανησυχίες σχετικά με τον τρόπο διαχείρισής τους από τις εφαρμογές τεχνητής νοημοσύνης, καθώς υπάρχει ο κίνδυνος μη πλήρους αξιοποίησης των δεδομένων αυτών ή απώλειας πολύτιμων πληροφοριών.

Για να αποφευχθεί ο κίνδυνος της μη επαρκούς εκμετάλλευσης των προτερημάτων που προσφέρουν τα δεδομένα που συλλέγονται και αξιοποιούνται από τα μοντέλα μηχανικής μάθησης, είναι απαραίτητο να διασφαλιστεί η αποτελεσματική διαχείριση αυτών των δεδομένων και η δημιουργία αποτελεσματικών και αξιόπιστων αλγορίθμων και μοντέλων τεχνητής νοημοσύνης. Ειδικότερα, τη σύγχρονη εποχή, που η τεχνητή νοημοσύνη εξελίσσεται διαρκώς και εφαρμόζεται ευρέως σε διάφορους τομείς, η ανάγκη για αποτελεσματική διαχείριση των μεγάλων όγκων δεδομένων είναι ακόμη πιο επιτακτική. Επιπλέον, η επίτευξη αξιόπιστης και ασφαλούς διαχείρισης δεδομένων συμβάλει επίσης στη διασφάλιση της ιδιωτικότητας και της ακεραιότητας των πληροφοριών.

3.1 ΔΙΑΔΙΚΑΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΑ ΣΥΣΤΗΜΑΤΑ ΑΙ

Ο κύκλος ζωής των δεδομένων ξεκινά με τη φάση συλλογής δεδομένων [7]. Συγκεκριμένα, η συλλογή δεδομένων είναι η αρχική διαδικασία που ορίζει το θεμέλιο για την ανάλυση, την επεξεργασία και την αξιοποίηση των δεδομένων από τα συστήματα τεχνητής νοημοσύνης.



Εικόνα 1-Κυκλός ζωής δεδομένων

3.1.1 ΣΥΛΛΟΓΗ ΔΕΔΟΜΕΝΩΝ

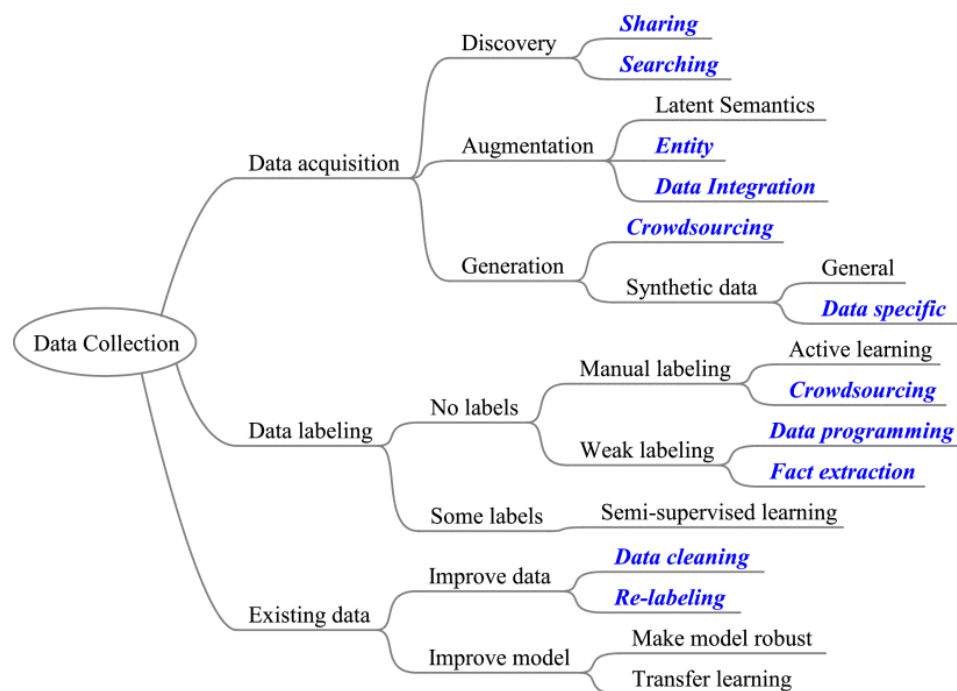
Κατά τη διαδικασία της συλλογής δεδομένων, αποκτούνται δεδομένα από ένα ευρύ φάσμα πηγών [7]. Τα δεδομένα αυτά κατηγοριοποιούνται ως ακολούθως :

- Αρχικά υπάρχουν τα δομημένα δεδομένα τα οποία αποκτώνται από ποικίλες πηγές, όπως βάσεις δεδομένων, φύλλα εργασίας και συστήματα διαχείρισης πόρων (ERP, CRM), προσφέροντας έτσι μια συνεκτική και ευανάγνωστη δομή για ανάλυση.
- Από την άλλη πλευρά, τα αδόμητα δεδομένα περιλαμβάνουν ποικιλία πηγών όπως emails, αναρτήσεις σε κοινωνικά δίκτυα και φωτογραφίες, παρέχοντας πλούσια πληροφορία που αναδεικνύει τις αναδυόμενες ακατέργαστες διαστάσεις των ψηφιακών δεδομένων.
- Τα ημιδομημένα δεδομένα αποτελούν ένα ενδιάμεσο στάδιο, όπου η δομή δεν είναι απολύτως πίνακας, αλλά περιέχουν σήμανση ή ετικέτες για τον διαχωρισμό των βασικών στοιχείων.

- Τέλος, οι ροές δεδομένων αναφέρονται σε δεδομένα που παράγονται πραγματικά κατά τη διάρκεια πραγματικού χρόνου από πηγές όπως, για παράδειγμα, από συσκευές IoT ή ροές τιμών μετοχών, παρέχοντας έτσι ενημερωμένη πληροφορία για άμεσες αναλύσεις και αποφάσεις.

Η διαδικασία της συλλογής δεδομένων πραγματοποιείται κυρίως με 3 τρόπους ανάλογα τον σκοπό που ανακτώνται τα σύνολά δεδομένων [9]. Ο πρώτος τρόπος είναι η χρήση τεχνικών απόκτησης δεδομένων, με τις οποίες ανακαλύπτονται, επεκτείνονται ή δημιουργούνται νέα σύνολα δεδομένων. Αυτές οι τεχνικές εφαρμόζονται όταν ο στόχος είναι η κοινοποίηση και η αναζήτηση νέων δεδομένων. Ο δεύτερος τρόπος είναι η χρήση τεχνικών επισήμανσης διαθέσιμων συνόλων. Τέλος αντί να χρησιμοποιηθούν τεχνικές επισήμανσης δεδομένων, ο τρίτος τρόπος είναι να προσαρμοστούν και βελτιωθούν τα ήδη υπάρχοντα δεδομένων, ή ακόμα γίνει εκπαίδευση βασιζόμενη σε υπάρχοντα εκπαιδευμένα μοντέλα.

Είναι σημαντικό να σημειωθεί ότι οι παραπάνω μέθοδοι δεν είναι απομονωμένες μεταξύ τους και μπορούν να συνδυαστούν εναρμονιστικά, επιτρέποντας έτσι την ολοκληρωμένη προσέγγιση της συλλογής και ανάλυσης δεδομένων στον επιστημονικό χώρο.



Εικόνα 2-Κυκλός ζωής δεδομένων, Βήμα 1:Συλλογή δεδομένων Roh et al. 2024

Όπως απεικονίζεται στην παραπάνω εικόνα, ο πρώτος τρόπος συλλογής είναι η τεχνική της απόκτησης δεδομένων και υλοποιείται με τρεις διαφορετικές μεθόδους [9]:

1. Η ανακάλυψη δεδομένων αποτελεί μια από αυτές τις μεθόδους και επωφελείται από την αυξημένη διαθεσιμότητα δεδομένων στον παγκόσμιο ιστό και στα data lakes τα οποία είναι τεράστια αποθετήρια δεδομένων από διάφορες πηγές, όπως αρχεία, βάσεις δεδομένων και streaming δεδομένα, παρέχουν έναν πλούτο πληροφοριών χωρίς προκαθορισμένη δομή ή σχήμα. Όταν τα συστήματα δεδομένων επικεντρώνονται σε συνεργατική ανάλυση συνόλων δεδομένων και στον κοινό διαμοιρασμό τους, όπως για παράδειγμα το σύστημά DataHub, τα δεδομένα πρέπει να ευρεθούν και να δημοσιευτούν για κοινή χρήση.

Σε αντίθετη περίπτωση όταν έχουμε να κάνουμε με συστήματα που λειτουργούν με πιο ιδιωτικό γνώμονα χωρίς την παροχή ανοιχτών συνόλων δεδομένων στο κοινό, παραδείγματος χάρη σε μεγάλες εταιρίες η στον Ιστό, οι χρήστες μπορεί να έχουν πρόσβαση μόνο σε ένα μέρος των δεδομένων που υπάρχουν στα data lakes, ανάλογα με τις ανάγκες και τους περιορισμούς που ορίζονται. Και αυτό συμβαίνει καθώς η πρόσβαση στα δεδομένα συχνά περιορίζεται και ελέγχεται βάσει των ρόλων, των άδειων πρόσβασης και των επιχειρησιακών περιορισμών για ευνόητους λόγους αλλά και επειδή δεν χρειάζεται όλες οι ομάδες ατόμων να καταβάλλουν περιττές προσπάθειες για την εκ νέου δημιουργία των συνόλων δεδομένων για τις εργασίες μηχανικής εκμάθησης. Σε αυτήν την περίπτωση η συλλογή δεδομένων εξαρτάτε αποκλειστικά από την αναζήτηση τους.

Μερικά εργαλεία που διευκολύνουν την αναζήτηση και τη διαχείριση δεδομένων και συμβάλλουν στη βελτίωση της απόδοσης των συστημάτων μηχανικής μάθησης είναι το Data Civilizer, το Google Data Search (Goods) [10], και η διαδικασία του data wrangling. Το Data Civilizer επιτρέπει στους χρήστες να εξερευνήσουν τις συνδέσεις μεταξύ των δεδομένων μέσω ενός γράφου σύνδεσης δεδομένων, επιτρέποντας την ανακάλυψη νέων μοτίβων και συνδέσεων. Το Google Data Search (Goods) παρέχει έναν κατάλογο μεταδεδομένων για δισεκατομμύρια datasets, επιτρέποντας στους χρήστες να αναζητούν και να εξερευνούν δεδομένα μέσω λέξεων-κλειδιών και να παρακολουθούν την προέλευση των datasets. Η διαδικασία του data wrangling βοηθά στον καθαρισμό, τη μετατροπή και την ενοποίηση δεδομένων, επιτρέποντας τη βελτιστοποίηση των δεδομένων για ανάλυση.

Αυτά τα εργαλεία και διαδικασίες επιτρέπουν στους ερευνητές και επαγγελματίες να ανακαλύπτουν, να προετοιμάζουν και να αναλύουν δεδομένα με μεγαλύτερη αποτελεσματικότητα, βοηθώντας έτσι στη βελτίωση της επίδοσης και της ακρίβειας των μοντέλων μηχανικής μάθησης.

2. Μια ακόμη μέθοδος που συμβάλει στη διαδικασία της τεχνικής απόκτησης δεδομένων είναι η αύξηση των υφιστάμενων συνόλων δεδομένων με εξωτερικά δεδομένα.

Η λανθάνουσα σημασιολογία βοηθά στην αύξηση δεδομένων στον τομέα της επεξεργασίας φυσικής γλώσσας (NLP) και όχι μόνο, διότι επιτρέπει τη δημιουργία νέων δεδομένων ή τη βελτίωση των υπαρχόντων δεδομένων με βάση τις κρυφές σημασιολογικές συσχετίσεις που παρουσιάζονται στα δεδομένα. Με τη χρήση ενσωματώσεων λέξεων, οντοτήτων ή γνώσεων, οι οποίες αποτελούν περιληπτικές αναπαραστάσεις των δεδομένων, μπορούμε να ενισχύσουμε την πλούσια αναπαράσταση των δεδομένων μας και να αυξήσουμε τη διαθέσιμη πληροφορία που περιλαμβάνουν.

Άλλες τέτοιες τεχνικές είναι οι τεχνικές συμπλήρωσης ανεπαρκών δεδομένων που συνεισφέρει στην πληρότητα και την εμπλουτιστική ανάπτυξη των συνόλων δεδομένων. Αυτό επιτυγχάνεται μέσω της ανάκτησης επιπλέον πληροφοριών από διαδικτυακούς πίνακες και την ενσωμάτωσή τους στα υπάρχοντα δεδομένα [11]. Η διαδικασία αυτή εκτελείται μέσω της δημιουργίας ερωτημάτων αναζήτησης στον ιστό και της επεξεργασίας των αποτελεσμάτων προκειμένου να αντληθούν επιπλέον πληροφορίες και να συμπληρωθούν τα απουσιάζοντα δεδομένα.

Η τεχνική ενοποίησης δεδομένων μπορεί επίσης να αντιληφθεί ως διαδικασία αύξησης του συνολικού όγκου δεδομένων, ιδίως όταν επεκτείνουμε υπάρχοντα σύνολα δεδομένων με νέες πληροφορίες που έχουν συλλεχθεί. Επίσης, υπάρχουν προγράμματα και πλατφόρμες λογισμικού που παρέχουν στους εργαζόμενους τη δυνατότητα να συμμετέχουν σε έργα crowdsourcing. Αυτά τα εργαλεία επιτρέπουν την εκτέλεση ποικίλων εργασιών, από την επισήμανση δεδομένων μέχρι την αξιολόγηση και την επίλυση προβλημάτων. Ορισμένα από τα πιο δημοφιλή παραδείγματα περιλαμβάνουν: Οι πλατφόρμες και τα εργαλεία αυτά διευκολύνουν τη συνεργασία μεταξύ εργοδοτών και εργαζομένων από όλο τον κόσμο, παρέχοντας τη δυνατότητα να ολοκληρωθούν μεγάλες ποσότητες εργασίας με αποδοτικό και οικονομικό τρόπο.

3. Τέλος η μέθοδος δημιουργίας δεδομένων αναδεικνύεται πολύ χρήσιμη όταν δεν υπάρχουν καθόλου διαθέσιμες συλλογές δεδομένων για εκπαίδευση μοντέλων μηχανικής μάθησης, και η δημιουργία τους υλοποιείται εξολοκλήρου εκείνη τη στιγμή με αυτόματο ή μη αυτοματοποιημένο τρόπο.

Η δημιουργία συνθετικών δεδομένων (Synthetic Data Generation) αποτελεί την **αυτοματοποιημένη** διαδικασία παραγωγής δεδομένων. Τα συνθετικά δεδομένα μιμούνται τα χαρακτηριστικά και τη δομή των πραγματικών δεδομένων, χωρίς να περιέχουν προσωπικές πληροφορίες ή ευαίσθητα δεδομένα.

Ένας από τους τρόπους δημιουργίας αυτών των δεδομένων είναι η χρήση γνωστών κατανομών πιθανότητας, όπως η κανονική κατανομή, για την παραγωγή δείγματος δεδομένων από αυτήν την κατανομή, με τη βοήθεια εργαλείων όπως το scikit-learn [12].

Ο πιο εξελιγμένος τρόπος δημιουργίας συνθετών δεδομένων ωστόσο είναι η χρήση παραγωγικών αντιπαραθετικών δικτύων γνωστά και ως GAN (Generative adversarial networks), όπου δύο ανταγωνιστικά νευρωνικά δίκτυα (το δίκτυο παραγωγής και το δίκτυο διακρίσεων) ανταγωνίζονται μεταξύ τους για να δημιουργήσουν πιο ρεαλιστικά δεδομένα [13,14]. Αυτό επιτυγχάνεται καθώς το δίκτυο παραγωγής λαμβάνει τυχαίο θόρυβο, συνήθως από μια κανονική ή ομοιόμορφη κατανομή και παράγει συνθετικά δεδομένα, ενώ το διακριτικό Δίκτυο λαμβάνει δεδομένα αυτά και προβλέπει αν τα δεδομένα είναι πραγματικά ή ψεύτικα. Ουσιαστικά τα δυο δίκτυα αλληλοεπιδρούν μεταξύ τους με αποτέλεσμα το διακριτικό δίκτυο μαθαίνει να διακρίνει πραγματικά από ψεύτικα δεδομένα, ενώ το γεννητικό δίκτυο μαθαίνει να παράγει δεδομένα που μοιάζουν όσο το δυνατόν περισσότερο με τα πραγματικά δεδομένα.

Ακόμη μια μέθοδος είναι η συνθετική δημιουργία εικόνων με χρήση δεδομένων. Στον τομέα της ανίχνευσης -αναγνώρισης αντικειμένων, γίνεται χρήση ορισμένων μεθοδολογιών, όπως απόκτηση πληροφοριών από τα χαρακτηριστικά τρισδιάστατων μοντέλων ενός υπό αναγνώριση αντικειμένου το οποίο υπόκειται σε στρέψη περί τους τρεις άξονες [15,16]. Σε περίπτωση ταχέων ακολουθιών στον άξονα του χρόνου, είναι εφικτή η παραδοχή κίνησης του αντικειμένου σε ευθύγραμμη τροχιά για τα τμήματα αυτής μεταξύ ενός στιγμιότυπου και του επόμενου για τα οποία δεν υπάρχει πληροφορία. Για την αναγνώριση κειμένου εντός εικόνων δύναται η μεταβολή των χαρακτηριστικών αυτού, όπως χρώμα, μέγεθος και τύπος γραφής, ώστε να αποκτηθεί η απαραίτητη πληροφορία για την παραγωγή συνθετικών εικόνων κειμένου [17, 18].

Μια ακόμη εναλλακτική μέθοδος από την παραπάνω, για την παραγωγή συνόλων δεδομένων (datasets), αποτελεί η διαλογή και απόρριψη ακατάλληλων εικόνων από σύνολα που περιέχουν εικόνες με θόρυβο, όπως τις περιπτώσεις των [19,20]. Οι Xia et al. [19] ψάχνουν το Διαδίκτυο για εικόνες με θόρυβο και στη συνέχεια χρησιμοποιεί ένα μέτρο βάσει πυκνότητας για να ομαδοποιήσει τις εικόνες και να αφαιρέσει τις εκτός ομάδας. Οι Bai et al. [20] εκμεταλλεύτηκαν μεγάλα αρχεία καταγραφής κλικ, τα οποία περιέχουν ερωτήματα χρηστών και τις εικόνες που επιλέχθηκαν από αυτούς τους χρήστες. Ένα βαθύ νευρωνικό δίκτυο χρησιμοποιείται για να μάθει αναπαραστάσεις των λέξεων και των εικόνων και να υπολογίσει τις ομοιότητες λέξη-λέξη και εικόνα-λέξη.

Ο μη αυτοματοποιημένος τρόπος υλοποιείται με τη διαδικασία crowdsourcing όπου πλήθος ατόμων, ομάδων και εταιριών αναλαμβάνουν επί πληρωμή όλη τη διαδικασία δημιουργίας συλλογών δεδομένων. Υπάρχουν προγράμματα και πλατφόρμες λογισμικού που διευκολύνουν τους εργαζομένους που συμμετέχουν σε έργα crowdsourcing επιτρέποντας την εκτέλεση ποικίλων εργασιών, από την επισήμανση δεδομένων μέχρι την αξιολόγηση και την επίλυση προβλημάτων. Μια από τις πιο γνωστές πλατφόρμες crowdsourcing η Amazon Mechanical Turk, επιτρέπει στους εργαζόμενους να ολοκληρώνουν διεργασίες (HITs - Human Intelligence Tasks) όπως επισήμανση εικόνων, ταξινόμηση δεδομένων και έρευνες [22]. Οι εργοδότες δημοσιεύουν τις εργασίες και οι ενδιαφερόμενοι τις ολοκληρώνουν για μια μικρή αμοιβή. Ακόμη ένα παράδειγμα είναι η πλατφόρμα CrowdFlower, γνωστή ως Figure Eight, που παρέχει ένα περιβάλλον για τη δημιουργία και τη διαχείριση εργασιών crowdsourcing. Οι εργασίες αυτές μπορεί να αφορούν την καθαριότητα δεδομένων, την επισήμανση και την ανάλυση συναισθημάτων. Η πλατφόρμα αυτή επίσης διαχειρίζεται την κατανομή των εργασιών και την ποιότητα των αποτελεσμάτων.

Η διαδικασία crowdsourcing έχει τα εξής στάδια: Αρχικά, οι εργαζόμενοι εξάγουν δεδομένα από τον ιστό με σκοπό να συγκεντρωθούν δεδομένα από ένα μεγάλο πλήθος χρηστών. Αφού συλλεχθούν πολυάριθμα και ποικιλόμορφα σύνολα δεδομένων, τα δεδομένα αυτά προεπεξεργάζονται για αξιοποίηση τους από μοντέλα μηχανικής μάθησης [9]. Συνεπώς οι εργαζόμενοι επιμελούνται και βελτιώνουν τα δεδομένα, επαληθεύοντας και διορθώνοντας τις ετικέτες σε σύνολα δεδομένων, όπως εικόνες ή κείμενα, διασφαλίζοντας την ακρίβεια και την αξιοπιστία των δεδομένων που χρησιμοποιούνται για την εκπαίδευση των αλγορίθμων AI. Επιπλέον αξιολογούν και ελέγχουν το περιεχόμενο που δημοσιεύεται σε διαδικτυακές πλατφόρμες (κοινωνικά δίκτυα, φόρουμ, κλπ.), για να εξασφαλιστεί ότι συμμορφώνεται με τις πολιτικές, και αν όχι, αφαιρούν το περιεχόμενο αυτό είτε πρόκειται για σχόλια, φωτογραφικό υλικό ακόμα και για παραπληροφόρηση. Τέλος, αναλαμβάνουν την ενοποίηση πληροφοριών που αναφέρονται στην ίδια οντότητα (π.χ. πρόσωπο, εταιρεία, τοποθεσία) από διαφορετικές πηγές δεδομένων. Η διαδικασία αυτή ονομάζεται «επίλυση οντοτήτων» και διασφαλίζει ότι οι αναφορές σε μια οντότητα σε διάφορες βάσεις δεδομένων ή έγγραφα αναγνωρίζονται ως αναφορές στην ίδια οντότητα, αποφεύγοντας την ύπαρξη διπλών ή αντικρουόμενων πληροφοριών.

Συνοπτικά, η τεχνική απόκτησης δεδομένων είναι αποτελεσματική για τη δημιουργία νέων συνόλων δεδομένων. Ωστόσο, όταν διαθέτουμε έναν επαρκή και ποικιλόμορφο όγκο δεδομένων, η δεύτερη τεχνική, αυτή της επισήμανσης δεδομένων, καθίσταται πιο χρήσιμη για την αξιοποίησή των δεδομένων από συστήματα μηχανικής μάθησης

Η τεχνική επισήμανσης δεδομένων, δηλαδή η διαδικασία προσθήκης σημάνσεων ή σχολιασμών σε ακατέργαστα δεδομένα, είναι απαραίτητη για τη βελτίωση της ποιότητας και της ακρίβειας των προβλέψεων που πραγματοποιούνται από τα μοντέλα μηχανικής μάθησης. Η επισήμανση δεδομένων επηρεάζει άμεσα την ακρίβεια των μοντέλων, καθώς οι αλγόριθμοι μηχανικής μάθησης βασίζονται σε επισημασμένα δεδομένα για να "μάθουν" να αναγνωρίζουν μοτίβα και να κάνουν προβλέψεις.

Υπάρχουν δυο κατηγορίες μεθόδων υλοποίησης της παραπάνω τεχνικής :

1. Η κυριότερη και πιο δημοφιλής μέθοδος επισήμανσης δεδομένων είναι η αξιοποίηση των ήδη υπάρχοντων ετικετών. Ο τομέας της ημι-επιβλεπόμενης μάθησης εξερευνά εκτενώς αυτή τη μέθοδο, επιδιώκοντας να χρησιμοποιήσει τα διαθέσιμα επισημασμένα δεδομένα για την πρόβλεψη ετικετών στα υπόλοιπα μη επισημασμένα δεδομένα [23].

Η απλούστερη κατηγορία τεχνικών ημι-επιβλεπόμενης μάθησης για επισήμανση συνόλων δεδομένων αποτελεί η εκπαίδευση ενός μοντέλου χρησιμοποιώντας έναν αλγόριθμο εκμάθησης σε ένα σύνολο χαρακτηριστικών. Μία τέτοια τεχνική αποτελεί το Self-training, που αρχικά εκπαιδεύει ένα μοντέλο με τα επισημασμένα παραδείγματα και έπειτα το μοντέλο αυτό χρησιμοποιείται για να προβλέψει ετικέτες στα μη επισημασμένα δεδομένα, προσθέτοντας τις πιο σίγουρες προβλέψεις στο σύνολο των επισημασμένων δεδομένων και επανεκπαιδεύοντας το μοντέλο[24].

Ακόμη μία τεχνική ημι-επιβλεπόμενης μάθησης αποτελεί η εκπαίδευση πολλαπλών ταξινομητών μέσω της δειγματοληψίας των δεδομένων εκπαίδευσης, πολλές φορές και της εκπαίδευσης ενός μοντέλου για κάθε δείγμα δεδομένων [9,25]. Το Tri-training είναι μια αντίστοιχη τεχνική που βασίζεται στη χρήση τριών ταξινομητών για τη βελτίωση της ακρίβειας της επισήμανσης και των προβλέψεων. Αναλυτικότερα, τρεις διαφορετικοί ταξινομητές εκπαιδεύονται ξεχωριστά χρησιμοποιώντας το ίδιο αρχικό σύνολο επισημασμένων δεδομένων. Η μέθοδος Bagging (Bootstrap Aggregating) μπορεί να χρησιμοποιηθεί για τη δημιουργία των διαφορετικών συνόλων εκπαίδευσης για κάθε μοντέλο, προσθέτοντας ποικιλία στα δεδομένα που χρησιμοποιούνται για την εκπαίδευση των μοντέλων. Αφού τα τρία μοντέλα εκπαιδευτούν, χρησιμοποιούνται για να κάνουν προβλέψεις για τα μη επισημασμένα δεδομένα. Για κάθε μη επισημασμένο δείγμα, εάν δύο από τους τρεις ταξινομητές συμφωνούν στην ετικέτα που πρέπει να αποδοθεί, τότε αυτή η ετικέτα θεωρείται αξιόπιστη και το δείγμα προστίθεται στο σύνολο των επισημασμένων δεδομένων με αυτή την ετικέτα.

Η τεχνική Self-training και η τεχνική Tri-training αποτελούν **τεχνικές ταξινόμησης**. Άλλες τεχνικές είναι οι **τεχνικές παλινδρόμησης** αλλά και

οι τεχνικές διάδοσης ετικετών που βασίζονται σε γραφήματα. Μάλιστα οι τελευταίες εκμεταλλεύονται τη δομή του γραφήματος των δειγμάτων ετικετών βάσει των ομοιοτήτων τους, προκειμένου να εξαγάγουν τις ετικέτες των υπόλοιπων ετικετών. Η διάδοση ετικετών που βασίζεται σε γραφήματα βρίσκει εφαρμογές σε πεδία όπως η υπολογιστική όραση, η ανάκτηση πληροφοριών, τα κοινωνικά δίκτυα και η επεξεργασία φυσικής γλώσσας, συμβάλλοντας στην ενίσχυση της ακρίβειας και της αποτελεσματικότητας των αλγορίθμων μηχανικής μάθησης [9] .

2. Όταν δεν υπάρχουν ετικέτες δεδομένων, χρησιμοποιούνται **αυτόματες τεχνικές**. Παρατηρώντας το διάγραμμα της εικόνας 1 μια τέτοια τεχνική αποτελεί η ενεργητική μάθηση (active learning), που λειτουργεί με την επιλογή των δεδομένων που θα επισημανθούν, βασιζόμενη σε ένα αρχικό μοντέλο που προτείνει τα πιο "ενδιαφέροντα" παραδείγματα για επισήμανση, τα οποία αναμένεται να βελτιώσουν την απόδοση του μοντέλου με τον καλύτερο δυνατό τρόπο.

Επιπλέον, οι τεχνικές crowdsourcing, αποτελούν επίσης μια αποτελεσματική τεχνική, όπου μπορούν να συμμετάσχουν πολλοί εργάτες που δεν είναι απαραίτητα ειδικοί στην επισήμανση. Αυτή η προσέγγιση αξιοποιεί τη συλλογική δύναμη του πλήθους για να επιτύχει γρήγορη και αποδοτική επισήμανση δεδομένων, μειώνοντας έτσι την ανάγκη για εξειδικευμένους εμπειρογνώμονες και διαμοιράζοντας το φορτίο της εργασίας. Ωστόσο, όπως είναι αναμενόμενο, υπάρχουν πιθανότητες για περισσότερα σφάλματα κατά τη διαδικασία της επισήμανσης [21,26].

Ωστόσο, με τη ραγδαία ανάπτυξη της τεχνολογίας και την εξέλιξη της βιομηχανίας smart devices είναι δύσκολο να υπάρχουν ετικέτες σε νέες συσκευές, κάτι το οποίο χρήζει **ημιαυτόματη δημιουργία μεγάλων ποσοτήτων ετικετών** αξιοποιώντας την τεχνική αδύναμης εποπτείας. Μία από τις πιο γνωστές μεθόδους υλοποίησης της αδύναμης εποπτείας είναι το data programming [9]. Καθώς η σήμανση δεδομένων σε μεγάλη κλίμακα γίνεται όλο και πιο σημαντική, ιδιαίτερα για εφαρμογές βαθιάς μάθησης, το data programming έχει προταθεί ως λύση για την παραγωγή μεγάλων ποσοτήτων αδύναμων ετικετών χρησιμοποιώντας πολλαπλές συναρτήσεις σήμανσης αντί για ατομικές ετικέτες. Αυτή η προσέγγιση επιτρέπει την αυτοματοποίηση και την κλιμάκωση της διαδικασίας σήμανσης, μειώνοντας ταυτόχρονα το κόστος και τον χρόνο που απαιτείται για την παραγωγή αξιόπιστων δεδομένων.

Μια άλλη μέθοδος για τη δημιουργία αδύναμων ετικετών είναι η εξαγωγή γεγονότων (fact extraction) [9]. Οι βάσεις γνώσεων περιέχουν γεγονότα που εξάγονται από διάφορες πηγές, συμπεριλαμβανομένου του διαδικτύου. Αυτές οι βάσεις δεδομένων μπορούν να αξιοποιηθούν για την αυτόματη παραγωγή ετικετών, ενισχύοντας την ακρίβεια και την

πληρότητα των δεδομένων εκπαίδευσης για τα μοντέλα μηχανικής μάθησης. Συνολικά, οι τεχνικές αδύναμης εποπτείας, όπως το data programming και η εξαγωγή γεγονότων, διαδραματίζουν καθοριστικό ρόλο στην αντιμετώπιση των προκλήσεων που συνδέονται με τη σήμανση μεγάλων ποσοτήτων δεδομένων στις σύγχρονες τεχνολογικές εφαρμογές.

«Η τρίτη τεχνική συλλογής δεδομένων, η οποία αποτελεί μια εναλλακτική προσέγγιση για την απόκτηση νέων δεδομένων και την επισήμανσή τους, συνδυάζει τις δύο πρώτες τεχνικές. Αυτή περιλαμβάνει τη βελτίωση της επισήμανσης οποιωνδήποτε υπαρχόντων συνόλων δεδομένων ή τη βελτίωση της εκπαίδευσης του μοντέλου [9].

Η συγκεκριμένη τεχνική είναι αρκετά ωφέλιμη όταν έχουμε να κάνουμε με μοντέλα και εφαρμογές πρωτότυπες ή μη τετριμμένες και τα διαθέσιμα δεδομένα είναι περιορισμένα. Σε τέτοιες περιπτώσεις, η εύρεση νέων συνόλων δεδομένων μπορεί να είναι δύσκολη ή ακόμα και αδύνατη. Επιπλέον, η απλή προσθήκη περισσότερων δεδομένων δεν εγγυάται πάντα σημαντική βελτίωση στην ακρίβεια του μοντέλου, ειδικά όταν το μοντέλο έχει ήδη μάθει τα βασικά μοτίβα από τα υπάρχοντα δεδομένα. Για την αντιμετώπιση αυτών των προκλήσεων, υπάρχουν δύο κύριες τεχνικές:

1. Η βελτίωση των υπαρχόντων δεδομένων που περιλαμβάνει μεθόδους όπως η αναθεώρηση και η επανασήμανση των δεδομένων και η χρήση τεχνικών καθαρισμού δεδομένων για την αφαίρεση θορύβου και ανακρίβειών. Αυτές οι τεχνικές μπορούν να αυξήσουν την ποιότητα και τη χρησιμότητα των δεδομένων, βελτιώνοντας έτσι την απόδοση των μοντέλων.
2. Από την άλλη πλευρά, η βελτίωση των μοντέλων μπορεί να επιτευχθεί μέσω της χρήσης πιο εξελιγμένων αλγορίθμων, της προσαρμογής των υπαρχόντων μοντέλων με τη χρήση υπερπαραμετροποίησης (hyperparameter tuning), και της εφαρμογής τεχνικών όπως η μεταφορά μάθησης (transfer learning), όπου τα μοντέλα μαθαίνουν από σχετικές, αλλά διαφορετικές, περιοχές προβλημάτων. Συνδυάζοντας αυτές τις τεχνικές, μπορούμε να επιτύχουμε καλύτερη απόδοση και μεγαλύτερη ακρίβεια σε μοντέλα, ακόμα και όταν τα δεδομένα είναι περιορισμένα ή δύσκολα να βρεθούν.

3.1.2 ΑΠΟΘΗΚΕΥΣΗ ΔΕΔΟΜΕΝΩΝ

Εφόσον η συλλογή δεδομένων έχει ολοκληρωθεί το επόμενο σημαντικό στάδιο είναι η αποθήκευση των δεδομένων. Αυτό το στάδιο είναι κρίσιμο για την ασφαλή, αποδοτική και προσβάσιμη διαχείριση των δεδομένων που έχουν συγκεντρωθεί. Η εντοπισμένη αποθήκευση διασφαλίζει όχι μόνο την

ακεραιότητα και την ασφάλεια των δεδομένων, αλλά και τη δυνατότητα ανάκτησης και ανάλυσης τους μελλοντικά.

Στο στάδιο της αποθήκευσης, τα δεδομένα πρέπει να διαχειρίζονται με βάση τον τύπο και τη δομή τους, χρησιμοποιώντας κατάλληλα συστήματα αποθήκευσης. Αναλυτικότερα τα συστήματα AI χρησιμοποιούν μια ποικιλία τεχνολογιών και υποδομών για τη διαχείριση και αποθήκευση των δεδομένων, χρησιμοποιώντας σχεσιακές βάσεις δεδομένων, όπως το MySQL και το PostgreSQL για δομημένα δεδομένα, ενώ για την αποθήκευση μεγάλων ποσοτήτων αδόμητων δεδομένων, χρησιμοποιούνται συχνά NoSQL βάσεις δεδομένων όπως το MongoDB και το Cassandra. Τα Data Warehouses, όπως το Amazon Redshift και το Google BigQuery, είναι μια ακόμη επιλογή, που επιτρέπουν τη συγκέντρωση και ανάλυση μεγάλων ποσοτήτων δεδομένων, όπως και οι λίμνες δεδομένων (Data Lakes), όπως το Apache Hadoop, είναι κατάλληλες για την αποθήκευση μεγάλων ποσοτήτων δεδομένων σε διάφορες μορφές. Ακόμη οι cloud υπηρεσίες, όπως το Amazon Web Services (AWS), το Google Cloud Platform (GCP) και το Microsoft Azure, παρέχουν ολοκληρωμένες λύσεις αποθήκευσης και επεξεργασίας δεδομένων, εξασφαλίζοντας υψηλή διαθεσιμότητα και απόδοση. Όλα αυτά τα συστήματα και τεχνολογίες συνεργάζονται για να παρέχουν πολύτιμες πληροφορίες από τα συλλεγόμενα δεδομένα, βοηθώντας στη λήψη αποφάσεων και στη βελτίωση των επιχειρηματικών στρατηγικών.

Οι διευθυντές επιχειρήσεων και της τεχνολογίας (IT) αντιμετωπίζουν έντονη πίεση για την εκμετάλλευση όσο το δυνατόν περισσότερων δεδομένων σε στρατηγικό επίπεδο, προκειμένου να προωθήσουν την καινοτομία και να διατηρήσουν ανταγωνιστικό πλεονέκτημα. Παράλληλα όμως, οφείλουν να αναπτύξουν στρατηγικές για την αποτελεσματική αποθήκευση, τη δημιουργία αντιγράφων ασφαλείας και την προστασία των δεδομένων από κλοπή, κακή χρήση και επιθέσεις [27].

Η συμμόρφωση με τις απαιτήσεις για την ασφάλεια των αποθηκευμένων δεδομένων είναι ουσιώδης στα υψηλόβαθμα διευθυντικά στελέχη (C-suite) και σε επίπεδο διοικητικού συμβουλίου σε πολλές εταιρείες. Η έλλειψη αποτελεσματικής ασφάλειας για τα αποθηκευμένα δεδομένα θέτει τις επιχειρήσεις σε κίνδυνο, εκθέτοντάς τις σε αγωγές και διακύβευση της επωνυμίας τους.

Υπάρχουν αρκετές προκλήσεις που οι επιχειρήσεις IT πρέπει να αντιμετωπίσουν στο στάδιο αποθήκευσης δεδομένων. Αρχικά η ποικιλόμορφη φύση των δεδομένων, που συλλέγονται από μια ευρεία γκάμα πηγών περιλαμβάνοντας τα μέσα κοινωνικής δικτύωσης, το Διαδίκτυο των πραγμάτων και τους χρήστες ψηφιακών ηλεκτρονικών συσκευών και έξυπνων συσκευών αποτελεί την πρώτη βασική πρόκληση αναγκάζοντας τα τις επιχειρήσεις να ακολουθήσουν αναλυτική και μεθοδευμένη στρατηγική για τη διαδικασία αποθήκευσης των δεδομένων όπως προαναφέρθηκε παραπάνω. Ακόμη οι διεθνείς επιχειρηματικές οργανώσεις εισάγουν τοπικούς κανονισμούς που διαφέρουν από χώρα σε χώρα και από κλάδο σε κλάδο, με περιοριστικές και αυστηρές ρυθμιστικές απαιτήσεις που

ορίζουν αυστηρές κυρώσεις για τις εταιρείες που δεν πληρούν τις κατάλληλες προδιαγραφές για την προστασία και ασφάλεια των δεδομένων των πελατών τους [28]. Αυτό πέρα από την αυξημένη λογοδοσία και τα καθήκοντα που οφείλουν να εκτελέσουν οι επιχειρήσεις, εγείρει και μια ακόμα πρόκληση η οποία είναι η πιθανότητα μη επαρκών γνώσεων και εμπειρογνωμοσύνης ενός εργαζομένου αλλά και τους ίδιους τους χρηστές για τους κανονισμούς και τις πολιτικές που ισχύουν, έχοντας ως απόρροια την επιβολή συχνών κυρώσεων.

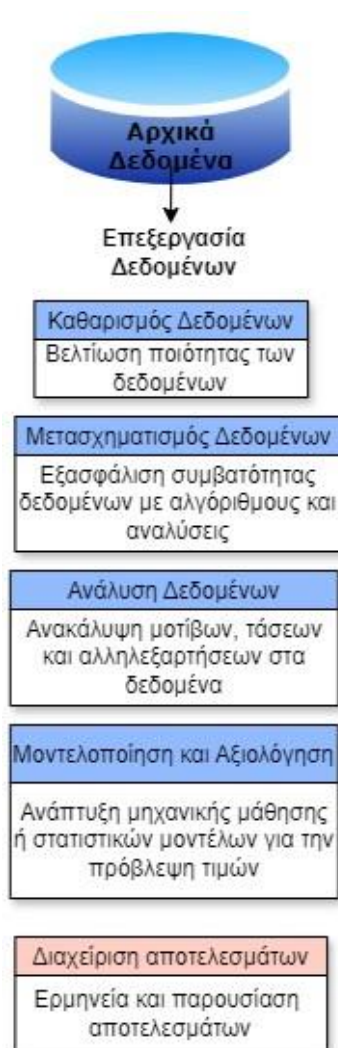
Επομένως η ασφάλεια των δεδομένων είναι κρίσιμη και περιλαμβάνει πρακτικές όπως η κρυπτογράφηση και η εφαρμογή αυστηρών ελέγχων πρόσβαση. Ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR), επιβάλλει στους οργανισμούς την υποχρέωση να λαμβάνουν τα κατάλληλα τεχνικά και οργανωτικά μέτρα για την προστασία των δεδομένων από αθέμιτη πρόσβαση, διαρροή ή καταστροφή. Η μακροχρόνια αποθήκευση και η διαχείριση των δεδομένων απαιτεί μια ολοκληρωμένη και προσεκτικά σχεδιασμένη προσέγγιση που θα εξασφαλίσει τη διατήρηση της ακεραιότητας και της ασφάλειας των δεδομένων καθ' όλη τη διάρκεια του κύκλου ζωής τους, διασφαλίζοντας παράλληλα ότι παραμένουν προσβάσιμα και επαναχρησιμοποιήσιμα όταν απαιτείται.

Οπότε η στρατηγική αποθήκευσης δεδομένων πρέπει να λαμβάνει υπόψη την αναμενόμενη διάρκεια ζωής των αποθηκευτικών μέσων, τις ανάγκες για γρήγορη πρόσβαση και ανάκτηση δεδομένων, την ασφάλεια των δεδομένων και την προστασία τους από απώλεια ή καταστροφή. Η επιλογή και η εφαρμογή των κατάλληλων εργαλείων και υποδομών πρέπει να γίνεται με βάση τις συγκεκριμένες απαιτήσεις του οργανισμού, με έμφαση στην ασφάλεια και τη συμμόρφωση των δεδομένων. Επίσης, είναι σημαντικό να υπάρχει επαρκής εφεδρική κάλυψη και πλάνο ανάκτησης δεδομένων σε περίπτωση καταστροφής.

Τέλος είναι κρίσιμης σημασίας τα συστήματα τεχνητής νοημοσύνης (AI) και οι οργανισμοί που συλλέγουν, επεξεργάζονται και κατέχουν ευαίσθητα και προσωπικά δεδομένα χρηστών να τηρούν αυστηρά τις πολιτικές και τις διατάξεις του Γενικού Κανονισμού για την Προστασία Δεδομένων (GDPR). Αυτό φυσικά συνεπάγεται την αποφυγή εκμετάλλευσης ή παράνομης κατοχής δεδομένων των χρηστών. Συγκεκριμένα, τα δεδομένα των χρηστών δεν πρέπει να χρησιμοποιούνται για άλλους σκοπούς πέραν αυτών για τους οποίους έχουν συλλεχθεί και δεν πρέπει να αποθηκεύονται για χρονικό διάστημα μεγαλύτερο από το απαραίτητο ή καθόλου αν δεν είναι απαραίτητο. Επιπλέον, δεν πρέπει να ζητείται παράνομα συγκατάθεση από τους χρήστες για τη συλλογή δεδομένων (όπως συμβαίνει συχνά με τα cookies), ούτε να συλλέγονται ή να αποθηκεύονται περισσότερα δεδομένα από αυτά που υπόσχεται η ιστοσελίδα ή η εφαρμογή. Η τήρηση αυτών των αρχών είναι απαραίτητη για την προστασία των δικαιωμάτων των χρηστών και την ενίσχυση της εμπιστοσύνης τους.

3.1.3 ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Η επεξεργασία δεδομένων αποτελεί θεμέλιο για τη σύγχρονη επιστήμη δεδομένων και μηχανικής μάθησης, καθώς αφορά τη μετατροπή ραβδωμένων δεδομένων σε πληροφορίες και γνώση. Στην ουσία της, η διαδικασία αυτή περιλαμβάνει την καθαρισμό, τη μετατροπή, την ανάλυση και τη μοντελοποίηση δεδομένων, με σκοπό την εξαγωγή ενδιαφερόντων πληροφοριών και την υποστήριξη αποφάσεων βασισμένων σε δεδομένα. Από τη βιομηχανία έως την ακαδημαϊκή έρευνα, η επεξεργασία δεδομένων είναι κρίσιμη για την επίτευξη ακρίβειας και αξιοπιστίας στην ανάλυση, αλλά και τη βελτίωση της ποιότητας των πρωτογενών δεδομένων, ανοίγοντας νέους ορίζοντες στην κατανόηση και την αξιοποίηση των πολύτιμων αυτών πληροφοριών.



Εικόνα 3-Σταδια επεξεργασίας δεδομένων

Η διαδικασία **καθαρισμού δεδομένων (Data Cleaning)** είναι ένα κρίσιμο βήμα στην επεξεργασία δεδομένων που αποσκοπεί στη βελτίωση της ποιότητας και της αξιοπιστίας των δεδομένων. Σε αυτό το στάδιο γίνεται ανίχνευση και

διόρθωση λαθών ή ανωμαλιών στα δεδομένα, όπως ελλιπείς τιμές, πιθανοί θόρυβοι και ανακολουθίες. Αυτό επιτυγχάνεται με χρήση τεχνικών όπως η αντικατάσταση ή αφαίρεση ελλιπών τιμών, η ανίχνευση και διόρθωση εκτός ορίων τιμών και η συνεπής κωδικοποίηση κατηγοριών [29]. Αυτές οι διαδικασίες είναι απαραίτητες για την εξασφάλιση ότι τα δεδομένα που χρησιμοποιούνται στην ανάλυση είναι αξιόπιστα και ακριβή. Η ύπαρξη πολλαπλών πηγών δεδομένων αυξάνει σημαντικά την πιθανότητα ύπαρξης λανθασμένων και μη ακριβών συνόλων δεδομένων, κάτι το οποίο θα είχε ως αποτέλεσμα την ύπαρξη αναξιόπιστων αλγορίθμων, παρόλο που μπορεί να φαίνονταν άρτιοι στη λειτουργία τους.

Αφού τα δεδομένα έχουν μετατραπεί σε κατάλληλη μορφή με βέλτιστη ποιότητα και ακρίβεια, μετατρέπονται (**Data Transformation**) σε μορφή που είναι κατάλληλη για την ανάλυση ή την εφαρμογή μοντέλων μηχανικής μάθησης. Η διαδικασία μετατροπής δεδομένων είναι απαραίτητη για την εξαγωγή χρήσιμων και αξιόπιστων συμπερασμάτων, καθώς επιτρέπει την ορθή προετοιμασία των δεδομένων, την ελαχιστοποίηση των λαθών και την αύξηση της αποτελεσματικότητας των αναλυτικών μεθόδων. Η μετατροπή μπορεί να περιλαμβάνει διάφορες ενέργειες, όπως η κανονικοποίηση, η τυποποίηση, η συγκέντρωση και η ομαδοποίηση των δεδομένων. Μέσω της κανονικοποίησης, τα δεδομένα προσαρμόζονται σε μια συγκεκριμένη κλίμακα, επιτρέποντας την ομοιόμορφη σύγκριση μεταξύ τους. Η τυποποίηση διασφαλίζει ότι τα δεδομένα ακολουθούν ένα ενιαίο πρότυπο, διευκολύνοντας την επεξεργασία τους από αλγόριθμους ανάλυσης και μοντέλα μηχανικής μάθησης. Επιπλέον, η συγκέντρωση δεδομένων συνενώνει πολλές επιμέρους εγγραφές σε μία συνοπτική αναπαράσταση, ενώ η ομαδοποίηση αναδιατάσσει τα δεδομένα σε λογικές κατηγορίες.

Έχοντας μετατρέψει τα σύνολα δεδομένων κατάλληλα για ανάλυση και χρήση από συστήματα ΑΙ και μοντέλα μηχανικής μάθησης, εφαρμόζονται στατιστικές μεθόδους ή αλγόριθμοι για να εξεταστούν οι σχέσεις μεταξύ των δεδομένων και να εξαχθούν πληροφορίες (**Data Analysis**). Κατά την ανάλυση, τα δεδομένα εξετάζονται σε βάθος για την ανακάλυψη προτύπων, τάσεων ή αλληλεξαρτήσεων που μπορούν να αποκαλυφθούν από αυτά. Οι τεχνικές ανάλυσης μπορεί να περιλαμβάνουν απλή στατιστική ανάλυση, μηχανική μάθηση ή ακόμα και προηγμένες μεθόδους όπως η οπτικοποίηση δεδομένων για την απεικόνιση πολύπλοκων δεδομένων με γραφήματα και διαγράμματα, επιτρέποντας στους αναλυτές να ανακαλύπτουν μοτίβα και τάσεις εύκολα κατανοώντας τη δομή και τις σχέσεις των δεδομένων. Η τελική αξιολόγηση και ερμηνεία των ευρημάτων από την ανάλυση δεδομένων βοηθά στη λήψη αποφάσεων ή στην ανάπτυξη στρατηγικών βασισμένων στα δεδομένα για βελτίωση των αποτελεσμάτων ή την πρόβλεψη μελλοντικών τάσεων.

Το αμέσως επόμενο στάδιο που ακολουθεί είναι η **μοντελοποίηση και αξιολόγηση (Modeling and Evaluation)**. Κατά τη μοντελοποίηση, εφαρμόζονται αλγόριθμοι και μοντέλα για να προβλέψουν τιμές ή να εξηγήσουν συμπεριφορές, αξιολόγηση της απόδοσης των μοντέλων. Αυτό μπορεί να

περιλαμβάνει από απλή στατιστική ανάλυση έως πολύπλοκες μεθόδους μηχανικής μάθησης, όπως νευρωνικά δίκτυα ή μοντέλα βαθιάς μάθησης. Κατά την αξιολόγηση, τα μοντέλα αυτά δοκιμάζονται και αξιολογούνται ως προς την ακρίβειά τους και την ικανότητά τους να γενικεύουν σε νέα δεδομένα. Η διαδικασία αυτή επιτρέπει την επιλογή του καλύτερου μοντέλου για τις συγκεκριμένες ανάγκες και την αποτελεσματική χρήση των αποτελεσμάτων για αποφάσεις ή προβλέψεις που βασίζονται στα δεδομένα.

Μετά την προηγούμενη επεξεργασία και ανάλυση των δεδομένων από την αρχική τους ακατέργαστη μορφή, γίνεται ερμηνεία και παρουσίαση των αποτελεσμάτων από την ανάλυση δεδομένων (**Διαχείριση αποτελεσμάτων (Results Management)**). Αυτή η διαδικασία συχνά στοχεύει στην υποστήριξη αποφάσεων ή την παραγωγή αναφορών, προβάλλοντας και αξιολογώντας κρίσιμες παρατηρήσεις που προέκυψαν κατά την επεξεργασία των δεδομένων. Ο στόχος είναι να κατανοηθούν τα αποτελέσματα και να εφαρμοστούν σε πρακτικές λύσεις ή στρατηγικές, ενισχύοντας έτσι την ικανότητα της οργάνωσης να λαμβάνει ενημερωμένες αποφάσεις και να αντιμετωπίζει προκλήσεις ή ευκαιρίες με βάση τα δεδομένα.

3.1.4 ΔΙΑΝΟΜΗ ΚΑΙ ΚΟΙΝΟΠΟΙΗΣΗ ΔΕΔΟΜΕΝΩΝ

Η διανομή και κοινοποίηση δεδομένων αποτελεί κρίσιμο στάδιο στον κύκλο ζωής των δεδομένων, διασφαλίζοντας ότι οι πληροφορίες είναι διαθέσιμες στους κατάλληλους αποδέκτες σε έγκαιρη και ασφαλή βάση. Αυτό το στάδιο περιλαμβάνει τη μετάδοση δεδομένων σε διάφορους φορείς, οργανισμούς ή συστήματα, διασφαλίζοντας παράλληλα την ακεραιότητα, την εμπιστευτικότητα και την ασφάλεια των δεδομένων. Η ανταλλαγή γνώσεων μεταξύ των ενδιαφερόμενων μερών ενισχύει την αποτελεσματικότητα της ανάλυσης και της λήψης αποφάσεων, καθώς επιτρέπει την πρόσβαση σε διαφορετικές προοπτικές και ειδικότητες, επιταχύνοντας την επίλυση προβλημάτων και την ανάπτυξη καινοτόμων λύσεων.

Έτσι η διανομή και κοινοποίηση δεδομένων προάγει την ανάπτυξη και την εφαρμογή της τεχνητής νοημοσύνης, οδηγώντας σε πιο αποδοτικά, αξιόπιστα και καινοτόμα συστήματα ΑΙ. Αυτό επιτυγχάνεται, καθώς η διαθεσιμότητα μεγάλων και ποικιλόμορφων συνόλων δεδομένων από διαφορετικούς τομείς και περιβάλλοντα καθιστά να μοντέλα ΑΙ ικανά να προσαρμόζονται και να είναι πιο ευέλικτα σε ποικίλες εφαρμογές και καταστάσεις, όπως η αναγνώριση εικόνων, η επεξεργασία φυσικής γλώσσας και η πρόβλεψη συμπεριφορών. Επιπλέον η ανοικτή πρόσβαση σε δεδομένα επιτρέπει σε επιστήμονες και ερευνητές να αναπαράγουν αποτελέσματα και να επιβεβαιώνουν τα ευρήματα άλλων μελετών, προάγοντας τη διαφάνεια και την αξιοπιστία της έρευνας στον τομέα της ΑΙ. Επίσης η διανομή και κοινοποίηση δεδομένων δίνει τη δυνατότητα σύγκρισης διαφορετικών αλγόριθμων στις ίδιες βάσεις δεδομένων, επιτρέποντας την αναγνώριση των πιο αποδοτικών τεχνικών. Ακόμη η εύκολη πρόσβαση σε δεδομένα ενθαρρύνει τη συνεργασία μεταξύ διαφορετικών ερευνητικών ομάδων και οργανισμών, επιτρέποντας την ανάπτυξη νέων ιδεών και τη δημιουργία καινοτόμων λύσεων που μπορεί να μην ήταν εφικτές μέσω της απομονωμένης

εργασίας. Τέλος η συνεχής ροή δεδομένων επιτρέπει την εξέλιξη των μοντέλων AI, καθώς αυτά μπορούν να ενημερώνονται και να βελτιώνονται συνεχώς με νέα δεδομένα, διασφαλίζοντας ότι παραμένουν ακριβή και σχετική με τις τρέχουσες ανάγκες.

Για να επιτευχθούν αυτοί οι παραπάνω στόχοι, χρησιμοποιούνται προηγμένα εργαλεία και τεχνολογίες όπως συστήματα διαχείρισης δεδομένων (DMS), αποθετήρια δεδομένων (data repositories), πλατφόρμες συνεργασίας (collaboration platforms) και τεχνολογίες blockchain για τη διασφάλιση της διαφάνειας και της ανιχνευσιμότητας των δεδομένων. Επιπλέον, οι πολιτικές ανοικτής πρόσβασης και τα πρότυπα FAIR (Findable, Accessible, Interoperable, Reusable) διευκολύνουν την εύρεση και επαναχρησιμοποίηση των δεδομένων από την ευρύτερη επιστημονική κοινότητα, ενισχύοντας τη διαλειτουργικότητα και την αειφορία της επιστημονικής έρευνας.

Συνοπτικά είναι γεγονός ότι η ύπαρξη κοινοποιημένων συνόλων δεδομένων μπορεί να αποφέρει πολλά οφέλη στην κοινωνία αλλά και στην εξέλιξη της επιστήμης της ιατρικής της επιστήμης των υπολογιστών και της τεχνητής νοημοσύνης. Ωστόσο πρέπει να ληφθούν υπόψη κάποια βασικά ζητήματα για τη δημιουργία ενός πλαισίου διακυβέρνησης για την ανταλλαγή δεδομένων. Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR), παρέχει μια κατευθυντήρια γραμμή για την υπεύθυνη κοινής χρήσης δεδομένων, όμως σε πολλά σημεία δεν υπάρχει μια πλήρης και ξεκάθαρη καθοδήγηση για μια συνεκτική εφαρμογή των αρχών και κανόνων, ώστε να επιτευχθεί η επιθυμητή ισορροπία μεταξύ της προώθησης της ανταλλαγής δεδομένων και της διασφάλισης της δημόσιας εμπιστοσύνης.

Είναι σημαντικό, λοιπόν, να ληφθούν υπόψη ορισμένες βασικές αρχές, όπως η μεγιστοποίηση της κοινωνικής ευημερίας, η οποία επιτυγχάνεται με τη διασφάλιση βέλτιστης και ακριβούς ποιότητας συνόλων δεδομένων [30]. Αυτά τα δεδομένα πρέπει να χρησιμοποιούνται αποκλειστικά για επιστημονικούς και ερευνητικούς σκοπούς που ωφελούν το κοινό καλό και να είναι κατάλληλα διαφυλαγμένα, διασφαλίζοντας τη μακροπρόθεσμη προσβασιμότητά τους από ερευνητικές ομάδες και εξειδικευμένους ερευνητές που μπορούν να αιτιολογήσουν τους ερευνητικούς σκοπούς [31,32]. Επιπλέον, πρέπει να συμμορφώνονται με πρότυπα ποιότητας και μηχανισμούς ανάδρασης σε κάθε στάδιο της επεξεργασίας τους [31].

Τα οφέλη από τον διαμοιρασμό δεδομένων θα πρέπει να υπερβαίνουν κατά πολύ τα μειονεκτήματα, ειδικά όταν αναφερόμαστε σε δεδομένα ευαίσθητης φύσης. Οι χρήστες πρέπει να ενημερώνονται πλήρως και να συναινούν με τους σκοπούς και τους όρους για τους οποίους γίνεται η κοινοποίηση των προσωπικών τους δεδομένων. Ακόμη οι χρήστες έχουν μια πληθώρα δικαιωμάτων αναφορικά με τα προσωπικά τους δεδομένα [30], μερικά από τα οποία περιλαμβάνουν το δικαίωμα στην ενημέρωση, για το είδος της έρευνας που διεξάγεται, τις δραστηριότητες, τα αποτελέσματα της έρευνας [31], τη νομική βάση και τους στόχους της επεξεργασίας δεδομένων από τρίτους [31], πώς θα προστατεύονται τα δεδομένα

και η εμπιστευτικότητα αυτών των δεδομένων [33], τα όρια στην ανωνυμία και την εμπιστευτικότητα των δεδομένων [31,33], τις πιθανές αρνητικές συνέπειες της παραβίασης του απορρήτου [33] και τις εξαιρετικές περιστάσεις και συνθήκες υπό τις οποίες οι ερευνητές μπορούν να έχουν πρόσβαση σε δεδομένα που δεν είναι κωδικοποιημένα ή ανώνυμα [33]. Επιπλέον, έχουν το δικαίωμα πρόσβασης στα προσωπικά τους δεδομένα[34], καθώς και να λάβουν αντίγραφα αυτών των δεδομένων, το δικαίωμα στη διόρθωση που εξασφαλίζει ότι οι χρήστες μπορούν να διορθώνουν ανακριβή ή ελλιπή δεδομένα που τους αφορούν και το δικαίωμα στη διαγραφή [34], γνωστό και ως «δικαίωμα στη λήθη», που δίνει στους χρήστες τη δυνατότητα να ζητούν τη διαγραφή των προσωπικών τους δεδομένων υπό ορισμένες προϋποθέσεις. Τέλος, το δικαίωμα στον περιορισμό της επεξεργασίας επιτρέπει στους χρήστες να περιορίζουν την επεξεργασία των δεδομένων τους, ιδιαίτερα σε περιπτώσεις που αμφισβητούν την ακρίβεια των δεδομένων ή έχουν αντιρρήσεις για τη χρήση τους.

3.1.5 ΑΡΧΕΙΟΘΕΤΗΣΗ ΔΕΔΟΜΕΝΩΝ

Εφόσον τα δεδομένα έχουν περάσει από όλα τα στάδια και έχουν αναλυθεί, επεξεργαστεί και μετατραπεί σε γνώση, προχωρούν στη φάση της αρχειοθέτησης και της αποθήκευσής τους.

Η εν λόγω διαδικασία υπόκειται σε σαφείς νομικούς κανόνες και απαιτήσεις συμμόρφωσης που επιβάλλονται στους ενδιαφερόμενους φορείς οι οποίοι συλλέγουν και χρησιμοποιούν αυτά τα σύνολα δεδομένων, όπως επιχειρήσεις, οργανισμοί τεχνολογίας πληροφοριών, ηλεκτρονικές πλατφόρμες κοινωνικής δικτύωσης, εμπορικές εταιρείες που χρησιμοποιούν εργαλεία όπως τα Google Analytics και εφαρμογές τεχνητής νοημοσύνης (π.χ. ChatGPT). Επίσης οι χρήστες των δεδομένων πρέπει να είναι εκτενώς ενημερωμένοι για το χρονικό διάστημα τον λόγο και τον τρόπο που αποθηκεύονται τα δεδομένα τους, αλλά και να έχουν συναινέσει για την όλη τη διαδικασία και τους όρους της αρχειοθέτησης. Με αυτόν τον τρόπο, τα δεδομένα διατηρούνται για μελλοντική χρήση, εκπαίδευση αλγορίθμων, αναφορές ή για νομικούς σκοπούς.

Παρά την ύπαρξη αυστηρών κανονισμών και πολιτικών για την αποθήκευση των δεδομένων με γνώμονα την προστασία και τον σεβασμό της ιδιωτικότητας των χρηστών, δεν μπορεί να υπάρχει απόλυτη διαφάνεια. Σε ορισμένες περιπτώσεις μάλιστα, όπως για λόγους εθνικής ασφάλειας ή για την πρόληψη εγκλημάτων, κυβερνητικοί οργανισμοί μπορεί να συλλέγουν δεδομένα χωρίς την άμεση ενημέρωση των χρηστών. Αυτές οι περιπτώσεις συνήθως ρυθμίζονται από ειδική νομοθεσία και δικαστικές αποφάσεις. Ακόμη δεν είναι λίγες οι περιπτώσεις που πολυεθνικές και μεγάλες εταιρείες ή κυβερνητικοί οργανισμοί αποθήκευσαν παράνομα τα προσωπικά δεδομένα των χρηστών λόγω εσωτερικών διαδικασιών ή παραβιάσεων.

Τέτοια περιστατικά υπογραμμίζουν τη σημασία της τήρησης αυστηρών κανονισμών και βέλτιστων πρακτικών για την ασφάλεια και την προστασία των

δεδομένων σε συστήματα τεχνητής νοημοσύνης και άλλες τεχνολογικές πλατφόρμες.

Το 2024, οι τεχνολογικές εξελίξεις επικεντρώνονται σε συστήματα αρχειοθέτησης που προσφέρουν ολοκληρωμένη διαχείριση δεδομένων, ενσωμάτωση, παγκόσμια πρόσβαση και αποθήκευση πολλαπλών πρωτοκόλλων με οικονομικά αποδοτικούς τρόπους. Αυτά τα συστήματα μετατρέπουν τα αρχεία σε δυναμικές πλατφόρμες που υποστηρίζουν την έρευνα και την ανάπτυξη της τεχνητής νοημοσύνης (AI), επιτρέποντας την ενεργή συμμετοχή τους στην ανάλυση και την αξιοποίηση δεδομένων. Τα συστήματα τεχνητής νοημοσύνης (AI), τα οποία συλλέγουν και επεξεργάζονται μεγάλα σύνολα δεδομένων, βασίζονται σε προηγμένες τεχνολογίες αποθήκευσης για να διαχειριστούν αυτά τα δεδομένα. Τα δεδομένα αποθηκεύονται σε κεντρικά αποθηκευτικά συστήματα ή σε κατανεμημένες πλατφόρμες cloud, με στόχο την ασφάλεια, την προσβασιμότητα και την αποδοτικότητα. Οι μηχανισμοί αρχειοθέτησης περιλαμβάνουν τη συμπίεση δεδομένων, την κρυπτογράφηση και την τακτική δημιουργία αντιγράφων ασφαλείας, εξασφαλίζοντας την ακεραιότητα και την προστασία των δεδομένων. Παράλληλα, οι προκλήσεις που συνοδεύουν την αρχειοθέτηση δεδομένων περιλαμβάνουν τη διαχείριση του όγκου των δεδομένων, την ταξινόμηση και την ταυτοποίηση, τη συμμόρφωση με κανονισμούς διατήρησης δεδομένων, όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR), την εύκολη ανάκτηση και την αντιμετώπιση της τεχνολογικής παλαίωσης. Για την αντιμετώπιση αυτών των προκλήσεων, προτείνονται πρακτικές όπως η καθιέρωση σαφών πολιτικών αρχειοθέτησης, η προτεραιοποίηση των δεδομένων βάσει της αξίας και της ευαισθησίας τους, η εφαρμογή μέτρων ασφαλείας και η χρήση αποδοτικών λύσεων αποθήκευσης.

3.1.6 ΔΙΑΓΡΑΦΗ/ΚΑΤΑΣΤΡΟΦΗ ΔΕΔΟΜΕΝΩΝ

Ο κύκλος ζωής δεδομένων τελειώνει με τη διαδικασία διαγραφής/καταστροφής δεδομένων όπου διασφαλίζεται η πλήρης και ανεπανόρθωτη αφαίρεση των δεδομένων από όλες τις αποθηκευτικές μονάδες και τα συστήματα των επιχειρήσεων και οργανισμών, για να εξαλειφθούν οι πιθανότητες ανεπιθύμητης πρόσβασης σε ευαίσθητα δεδομένα και να διασφαλιστεί η συμμόρφωση στις ισχύουσες νομικές απαιτήσεις για προστασία της ιδιωτικότητας και την ασφάλεια των ατόμων.

Σύμφωνα με τον Γενικό Κανονισμό Προστασίας Δεδομένων (GDPR), όπως αναγράφεται αναλυτικά στο άρθρο 17, τα άτομα έχουν το αναπόσπαστο δικαίωμα της διαγραφής των προσωπικών τους δεδομένων όταν αυτά δεν είναι πλέον απαραίτητα για τους σκοπούς για τους οποίους συλλέχθηκαν, όταν ανακαλούν τη συγκατάθεσή τους ή όταν τα δεδομένα έχουν υποβληθεί σε επεξεργασία παράνομα [35].

Ακόμη το ISO 27001, συγκεκριμένα στο Παράρτημα A.8.3 [36], υποστηρίζει ότι πρέπει πληρούνται όλες οι απαιτήσεις για την ασφαλή διαγραφή των μέσων

αποθήκευσης. Πρέπει να διασφαλιστεί ότι τα δεδομένα σε αφαιρούμενα και μη αφαιρούμενα μέσα να διαγράφονται οριστικά πριν από την απόρριψη ή την εγκατάλειψή τους. Ακόμη πρέπει να υπάρχουν διαδικασίες για την πρόληψη μη εξουσιοδοτημένης πρόσβασης σε ευαίσθητες πληροφορίες κατά τη διάρκεια της διαδικασίας απόρριψης. Για παράδειγμα, οι οργανισμοί μπορεί να χρησιμοποιούν λογισμικό διαγραφής δεδομένων, φυσική καταστροφή των αποθηκευτικών συσκευών ή απομαγνητισμό για να διασφαλίσουν ότι τα δεδομένα δεν μπορούν να ανασυγκροτηθούν ή να ανακτηθούν

3.2 ΠΡΟΚΛΗΣΕΙΣ ΔΙΑΧΕΙΡΙΣΗΣ ΔΕΔΟΜΕΝΩΝ

Η διαχείριση δεδομένων αντιμετωπίζει μια πολυπλοκότητα εκτίνουσας φύσης και έρχεται αντιμέτωπη με μια εκτενή γκάμα προβλημάτων κατά την εκτέλεσή της. Αυτό βεβαίως προκαλείται από την αμέτρητη ποσότητα δεδομένων που παράγονται καθημερινά, λόγω της αύξησης του ρυθμού παραγωγής δεδομένων, σε πραγματικό χρόνο, από πηγές όπως οι συσκευές IoT και τα κοινωνικά μέσα δικτύωσης. Αυτό οδηγεί στην ανάγκη για ανάπτυξη πιο προηγμένων συστημάτων επεξεργασίας δεδομένων προκειμένου να διαχειριστούν αποτελεσματικά αυτήν την τεράστια ροή πληροφοριών σε πραγματικό χρόνο.

Η ποικιλία των μορφών στις οποίες παρουσιάζονται αυτά τα δεδομένα δομημένα, ημιδομημένα και αδόμητα, περιπλέκει ακόμα περισσότερο τη διαδικασία διαχείρισης. Ειδικότερα κατά τη διαδικασία της ενοποίησης των δεδομένων για την εκπαίδευση αλγορίθμων τεχνητής νοημοσύνης, τα δεδομένα παρέχονται από διαφορετικές πηγές και παρουσιάζουν ετερογένεια στη μορφή, κάνοντας τη διαδικασία της ενοποίησης περιπλοκή στην εναρμόνιση των σχημάτων δεδομένων, την επίλυση ασυνεπειών και τη διατήρηση της σειράς δεδομένων [37].

Επίσης η ποιότητα των δεδομένων αποτελεί καίριο παράγοντα που επηρεάζει τη διαχείριση δεδομένων από συστήματα Τεχνητής Νοημοσύνης. Η ανεπαρκής ή μη ακριβής ποιότητα των δεδομένων (δηλαδή δεδομένα με ατέλειες όπως τιμές που λείπουν, λανθασμένες τιμές και τιμές εσφαλμένης μορφοποίησης) μπορεί να οδηγήσει σε αστοχίες και ανακρίβειες κατά την εκπαίδευση και λειτουργία των αλγορίθμων Τεχνητής Νοημοσύνης και να έχει αποτελέσματα παραπλανητικά για τη λήψη αποφάσεων, καθώς τα μη συναφή δεδομένα μπορούν να εισάγουν θόρυβο ή προκατάληψη σε ένα σύστημα ΑΙ, μειώνοντας έτσι την απόδοση και την αποτελεσματικότητά τους [10,37]. Αυτές οι προκαταλήψεις μπορεί να προέρχονται από δημογραφικούς, πολιτισμικούς ή άλλους παράγοντες και μπορούν να οδηγήσουν σε μη αντιπροσωπευτικά ή μεροληπτικά μοντέλα τεχνητής νοημοσύνης [37].

Ακόμη η έλλειψη επισημάνσεων στα δεδομένα αποτελεί εξίσου μια μεγάλη πρόκληση, καθώς οι περισσότερες εταιρείες χρησιμοποιούν εποπτευόμενη μάθηση για την εκπαίδευση των μοντέλων βαθιάς μάθησης (DL) και η απουσία μεταδεδομένων επηρεάζει διαφορετικούς τύπους δεδομένων ποικιλοτρόπως. Για

παράδειγμα, σε δεδομένα χρηματιστηρίου, η έλλειψη μεταδεδομένων δυσκολεύει την αναγνώριση τιμών όπως η τιμή ανοίγματος ή κλεισίματος [7].

Η διαρροή δεδομένων αποτελεί εξίσου σημαντική πρόκληση όταν δεν γίνεται σωστός διαχωρισμός μεταξύ των συνόλων δεδομένων εκπαίδευσης και επικύρωσης/δοκιμής [7]. Αυτό μπορεί να οδηγήσει σε υπερεκτίμηση της απόδοσης του μοντέλου κατά την εκπαίδευση, με αποτέλεσμα μειωμένη ποιότητα πρόβλεψης κατά την παραγωγή. Ακόμη, είναι σημαντικό να λάβουμε υπόψη το γεγονός ότι τα δεδομένα που συλλέγονται κατά τη διάρκεια της λειτουργίας μπορεί να διαφέρουν σημαντικά από αυτά που χρησιμοποιήθηκαν κατά την εκπαίδευση του μοντέλου. Αυτή η απόκλιση μπορεί να επηρεάσει αρνητικά την ακρίβεια και την αξιοπιστία των προβλέψεων. Ένα τέτοιο παράδειγμα αποτελεί το σύστημά Google Drive που εφηύρε η Google το οποίο δεν λειτούργησε άρτια καθώς τα δεδομένα εκπαίδευσης είχαν εξαχθεί μέσω ενός συγκεκριμένου αγωγού, ενώ τα δεδομένα παραγωγής προέρχονταν από ένα API, δημιουργώντας ασυμβατότητες. Αυτό εμπόδισε το μοντέλο να αποδώσει σωστά, υπογραμμίζοντας τη σημασία της συνεπούς και προσεκτικής διαχείρισης των δεδομένων κατά την εκπαίδευση και την παραγωγή.

Η προηγμένη ανάπτυξη της τεχνητής νοημοσύνης αυξάνει επίσης τον κίνδυνο κυβερνοεπιθέσεων κατά την επεξεργασία και ανάλυση δεδομένων από τα συστήματα AI, συντελώντας στην ανάγκη ύπαρξης ισχυρών μηχανισμών ασφαλείας και τη συμμόρφωση με βάση τους κανονισμούς απορρήτου δεδομένων που θέτει ο GDPR.

Τέλος η επιλογή της κατάλληλης υποδομής αποθήκευσης και μηχανισμών ανάκτησης είναι κρίσιμης σημασίας για την απόδοση, την επεκτασιμότητα και τη σχέση κόστους-αποτελεσματικότητας των συστημάτων διαχείρισης δεδομένων. Η σωστή υποδομή αποθήκευσης εξασφαλίζει την αποτελεσματική αποθήκευση και ταχεία ανάκτηση των δεδομένων, ενώ οι κατάλληλοι μηχανισμοί ανάκτησης επιτρέπουν τη γρήγορη πρόσβαση και επεξεργασία των δεδομένων. Επιπλέον, η επεκτασιμότητα της υποδομής είναι σημαντική για να αντιμετωπίσει την αύξηση του όγκου δεδομένων με τον χρόνο, ενώ η σωστή σχέση κόστους-αποτελεσματικότητας εξασφαλίζει ότι η υποδομή αποθήκευσης και ανάκτησης είναι οικονομικά βιώσιμη και αποδοτική για την οργάνωση.

Συνεπώς η διαχείριση δεδομένων απαιτεί εξειδικευμένες γνώσεις και προχωρημένη κατανόηση τεχνικών για την εξόρυξη, την ανάλυση και την επεξεργασία δεδομένων. Επιπλέον απαιτείται να υπάρχουν προηγμένες τεχνολογίες και συστήματα, τα όποια επιλέγονται σωστά, για να ανταποκρίνονται στις πολύπλοκες απαιτήσεις και διαδικασίες της επιστήμης των δεδομένων.

3.3 ΕΡΓΑΛΕΙΑ ΕΠΕΞΕΡΓΑΣΙΑΣ ΠΡΟΣΩΠΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Στον τομέα της επεξεργασίας προσωπικών δεδομένων, η χρήση κατάλληλων εργαλείων είναι κρίσιμη για τη διασφάλιση της ακρίβειας, της συνέπειας και της συμμόρφωσης με τις νομοθετικές απαιτήσεις. Αν και η δημιουργία εργαλείων είναι μια πρόκληση που αντιμετωπίζεται στα περισσότερα στάδια του κύκλου ζωής δεδομένων, πολλά από τα εργαλεία βρίσκονται υπό ανάπτυξη [7].

Στη διαδικασία συλλογής δεδομένων, επικρατούν το Apache Kafka, που διευκολύνει τη ροή δεδομένων σε πραγματικό χρόνο, και το Google Analytics, που προσφέρει ολοκληρωμένες λύσεις για την ανάλυση των δεδομένων ιστού. Η αποθήκευση των δεδομένων επιτυγχάνεται μέσω υπηρεσιών όπως το Amazon S3, το Google BigQuery και το Snowflake, προσφέροντας ασφαλείς και επεκτάσιμες λύσεις cloud. Για τη διαχείριση και τη διακυβέρνηση των δεδομένων, το Collibra και το Informatica παρέχουν ολοκληρωμένες λύσεις για την ποιότητα, τη διακυβέρνηση και την ασφάλεια των δεδομένων. Στον τομέα της ανάλυσης δεδομένων, τα εργαλεία Tableau και Power BI επιτρέπουν την οπτικοποίηση και την ανάλυση δεδομένων με διαδραστικό τρόπο. Για πιο προηγμένες ανάγκες οπτικοποίησης, το QlikView/Qlik Sense προσφέρουν ισχυρές δυνατότητες δημιουργίας διαδραστικών γραφημάτων. Στον τομέα της μηχανικής μάθησης και της τεχνητής νοημοσύνης, οι βιβλιοθήκες TensorFlow και PyTorch είναι αξιόπιστοι σύμμαχοι για την ανάπτυξη προηγμένων μοντέλων. Όσον αφορά τον καθαρισμό και τη μετατροπή των δεδομένων, εργαλεία όπως το TIBCO Clarity και το Data Wrangler προσφέρουν ισχυρές δυνατότητες, ενώ το Great Expectations παρέχει μια αυτόματη επαλήθευση της ποιότητας των δεδομένων. Ενώ τα εργαλεία για τη δοκιμή δεδομένων εξακολουθούν να αναζητούνται, το dbt (data build tool) προσφέρει μια προηγμένη λύση συνδυάζοντας μετατροπή δεδομένων και δοκιμές, ενώ το Apache Griffin προσφέρει ένα πλαίσιο για την επαλήθευση της ποιότητας των δεδομένων σε μεγάλη κλίμακα περιβάλλοντα. Τέλος, εργαλεία όπως το Talend και το Soda SQL ενισχύουν τη διαχείριση ποιότητας δεδομένων, προσφέροντας ελέγχους και επαλήθευση μέσω SQL. Με αυτά τα εργαλεία, οι οργανισμοί μπορούν να διαχειρίζονται αποτελεσματικά τα προσωπικά δεδομένα τους, εξασφαλίζοντας την αξιοπιστία και την ασφάλεια των δεδομένων.

3.4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Συνοπτικά η ανάπτυξη και η ευρεία χρήση της τεχνητής νοημοσύνης (AI) έχει επηρεάσει σημαντικά τον κύκλο ζωής των δεδομένων, αυτοματοποιώντας τη συλλογή δεδομένων από ποικίλες πηγές σε πραγματικό χρόνο, βελτιστοποιώντας την αποθήκευση δεδομένων μέσω προχωρημένων τεχνικών συμπίεσης και οργάνωσης και παρέχοντας αναλυτικά εργαλεία που εξάγουν πολύτιμες πληροφορίες και μοτίβα από τεράστιους όγκους δεδομένων. Αυτές οι δυνατότητες βελτιώνουν δραματικά την αποτελεσματικότητα και την ακρίβεια της επεξεργασίας δεδομένων, επιτρέποντας στους οργανισμούς να λαμβάνουν

καλύτερες αποφάσεις και να παρέχουν εξατομικευμένες υπηρεσίες. Ωστόσο, η αυξημένη χρήση της ΑΙ εγείρει σημαντικά ζητήματα σχετικά με την ιδιωτικότητα και την κυβερνοασφάλεια.

Παραδείγματα όπως το σκάνδαλο της Cambridge Analytica, όπου η συλλογή και η χρήση δεδομένων εκατομμυρίων χρηστών του Facebook έγινε χωρίς τη συγκατάθεση τους, δείχνουν την έκταση της κακής διαχείρισης δεδομένων [39]. Η περίπτωση της Google με την παράνομη συλλογή δεδομένων Wi-Fi μέσω των οχημάτων της υπηρεσίας Street View [40] και το πρόστιμο που επιβλήθηκε στη Microsoft για τη συλλογή δεδομένων χρηστών των Windows 10 χωρίς επαρκή ενημέρωση, υπογραμμίζουν τις προκλήσεις της νόμιμης και ηθικής χρήσης των δεδομένων. Και αυτά είναι μονάχα μερικά από τα σκάνδαλα.

Παρά το γεγονός ότι η τεχνητή νοημοσύνη έχει φέρει επανάσταση στην καθημερινότητά μας και έχει βελτιώσει την αποδοτικότητα και την καινοτομία στις επιχειρήσεις, το κρίσιμο ζήτημα παραμένει κατά πόσο η χρήση της ΑΙ σέβεται τα δικαιώματα των ανθρώπων και λειτουργεί με ηθικό γνώμονα. Είναι αναγκαίο να διασφαλιστεί ότι οι τεχνολογίες αυτές χρησιμοποιούνται με τρόπο που να προστατεύει την ιδιωτικότητα και να ενισχύει την εμπιστοσύνη του κοινού, διασφαλίζοντας έτσι ένα μέλλον όπου η καινοτομία και η ηθική συνυπάρχουν αρμονικά.

Εξάλλου η τεχνολογία δημιουργήθηκε με σκοπό να ωφελήσει την ανθρωπότητα και να προάγει την επιστήμη, παρέχοντας εργαλεία που βελτιώνουν την ποιότητα ζωής και την αποδοτικότητα των δραστηριοτήτων μας. Η ουσιαστική αποστολή της είναι να υπηρετεί την κοινωνία και να ενισχύει την ανάπτυξη και την καινοτομία, όχι να λειτουργεί ως μέσο χειραγώγησης ή να δημιουργεί συγκρούσεις μεταξύ κυβερνήσεων και πολιτών ή μεταξύ οργανισμών και χρηστών. Η ηθική χρήση της τεχνολογίας είναι θεμελιώδης για τη διατήρηση της εμπιστοσύνης και της συνεργασίας, διασφαλίζοντας ότι οι καινοτομίες αξιοποιούνται για το κοινό καλό.

ΚΕΦΑΛΑΙΟ 4 – Η ΕΠΙΔΡΑΣΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΣΥΝΗΣ ΣΤΟΝ ΚΥΒΕΡΝΟΧΩΡΟ

4.1 ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ

4.1.1 ΟΡΙΣΜΟΣ

Στη σημερινή εποχή ο ψηφιακός κόσμος έχει διεισδύσει σε κάθε πτυχή της καθημερινής μας ζωής και οι ανησυχίες για την ασφάλεια των δεδομένων αυξάνονται συνεχώς.

Η αυξανόμενη εξάρτηση της κοινωνίας από τις ψηφιακές τεχνολογίες, σε συνδυασμό με την αυξανόμενη συχνότητα και πολυπλοκότητα των κυβερνοεπιθέσεων και τη διαρκή ανάπτυξη της ψηφιακής οικονομίας, έχουν αναδείξει τη σημασία της κυβερνοασφάλειας. Η κυβερνοασφάλεια αποτελεί τη λύση για την προστασία του ψηφιακού κόσμου και την ασφάλεια των δεδομένων. Είναι εξαιρετικά σημαντική για τους ανθρώπους, τους οργανισμούς και τις παγκόσμιες κυβερνήσεις. Και αυτό επειδή οποιοσδήποτε μπορεί να αποτελεί πιθανό στόχο επίθεσης. Γιαυτό η προστασία των πληροφοριών, η διασφάλιση της ιδιωτικότητας και η αποτροπή κακόβουλων επιθέσεων είναι απαραίτητες για τη διατήρηση της σταθερότητας και της ασφάλειας τόσο σε προσωπικό όσο και σε οργανωτικό επίπεδο.

Ο όρος κυβερνοασφάλεια αναφέρεται σε ένα ευρύ σύνολο τεχνολογιών, πολιτικών, ελέγχων και διαδικασιών που εφαρμόζονται για την προστασία των πληροφοριακών συστημάτων, των δικτύων υπολογιστών και των δεδομένων που περιέχουν, από κακόβουλο λογισμικό και επιθέσεις. Τέτοιου είδους επιθέσεις μπορεί να οδηγήσουν σε μη εξουσιοδοτημένη πρόσβαση, κλοπή ή βλάβη υλικού, λογισμικού ή δεδομένων, καθώς και σε διακοπή ή εσφαλμένη κατεύθυνση των υπηρεσιών που παρέχουν [41].

Επομένως, ο κύριος στόχος της κυβερνοασφάλειας είναι η εξασφάλιση και η διατήρηση της ασφάλειας των περιουσιακών στοιχείων των οργανισμών και των χρηστών έναντι των σχετικών κινδύνων στον κυβερνοχώρο [42]. Οι βασικοί στόχοι ασφαλείας περιλαμβάνουν τη διαθεσιμότητα, την ακεραιότητα και την εμπιστευτικότητα, καθώς και την αυθεντικοποίηση, τη μη αποποίηση (non-repudiation) και την ανθεκτικότητα. Αυτοί οι στόχοι είναι θεμελιώδεις για την προστασία των δεδομένων και των συστημάτων από απειλές και επιθέσεις, διασφαλίζοντας ότι τα συστήματα και οι πληροφορίες παραμένουν ασφαλή, αξιόπιστα και διαθέσιμα όταν απαιτείται. Παράλληλα, εξασφαλίζουν ότι η πρόσβαση περιορίζεται μόνο σε εξουσιοδοτημένους χρήστες, οι ενέργειες δεν μπορούν να αμφισβητηθούν, και ότι τα συστήματα μπορούν να ανακάμπτουν γρήγορα από τυχόν επιθέσεις ή δυσλειτουργίες.

Φυσικά, η επίτευξη της ασφάλειας στον κυβερνοχώρο απαιτεί την υιοθέτηση ενός πολυδιάστατου πλαισίου που συνδυάζει τεχνολογικές υποδομές και οργανωτικές πρακτικές[42]. Βασικοί πυλώνες αυτού του πλαισίου περιλαμβάνουν τα τείχη προστασίας και τα συστήματα ανίχνευσης και

αποτροπής εισβολών (IDS/IPS) για τον έλεγχο και την προστασία των δικτύων, καθώς και τις τεχνολογίες κρυπτογράφησης για την ασφαλή διαχείριση των δεδομένων. Η ενσωμάτωση λογισμικού ανίχνευσης κακόβουλου λογισμικού (antivirus/antimalware) και συστημάτων διαχείρισης πληροφοριών και συμβάντων ασφαλείας (SIEM) παρέχει μια ολοκληρωμένη ανάλυση και ανταπόκριση σε πραγματικό χρόνο. Παράλληλα, η εκπαίδευση του προσωπικού και η ευαισθητοποίηση των χρηστών σχετικά με τις σύγχρονες απειλές είναι κρίσιμη για την ελαχιστοποίηση των ανθρωπίνων λαθών. Τέλος, οι πολιτικές ασφαλείας και τα σχέδια αντιμετώπισης περιστατικών διασφαλίζουν ότι οι οργανισμοί είναι προετοιμασμένοι να ανταποκριθούν αποτελεσματικά σε κάθε πιθανή απειλή, καθιστώντας την κυβερνοασφάλεια μια απαραίτητη προϋπόθεση για την ανθεκτικότητα και την προστασία στον ψηφιακό κόσμο.

Η ευθύνη για την κυβερνοασφάλεια είναι κοινή και καταμερίζεται μεταξύ κυβερνήσεων, όπως δείχνει η πρωτοβουλία GDPR (Γενικός Κανονισμός για την Προστασία Δεδομένων) στην Ευρωπαϊκή Ένωση, και οργανισμών όπως η ISO (Διεθνής Οργανισμός Τυποποίησης) που παρέχει πρότυπα για την ασφάλεια των πληροφοριών. Οι κυβερνήσεις παγκοσμίως επενδύουν σημαντικούς πόρους για την ενίσχυση των συστημάτων κυβερνοασφάλειας, ενώ οι οργανισμοί αναβαθμίζουν συνεχώς τις στρατηγικές και τα μέτρα ασφαλείας τους για να ανταποκριθούν στις νέες προκλήσεις και απειλές. Ακόμη, οι διεθνείς οργανισμοί και οι κυβερνήσεις συνεργάζονται στενά για την καταπολέμηση των κυβερνοαπειλών, ανταλλάσσοντας πληροφορίες, αναπτύσσοντας κοινές πολιτικές και διενεργώντας διπλωματικές προσπάθειες για την προώθηση της διεθνούς ασφαλείας στον κυβερνοχώρο, καθώς και την κατάρτιση συμφωνιών ή συνθηκών που ενισχύουν τη σταθερότητα και την ασφάλεια των πληροφοριακών υποδομών σε παγκόσμιο επίπεδο [43]. Παράλληλα, στρατιωτικές βάσεις διαδραματίζουν κρίσιμο ρόλο στην προστασία των πληροφοριακών δικτύων, διασφαλίζοντας ότι τα εθνικά και διεθνή συστήματα παραμένουν ασφαλή απέναντι σε εξωτερικές απειλές και επιθέσεις [43].

Μερικοί ακόμη οργανισμοί που έχουν συμβάλει στη διευθέτηση θέμάτων κυβερνοασφάλειας είναι ο **NIST** (National Institute of Standards and Technology), που έχει αναπτύξει το NIST Cybersecurity Framework, το οποίο αποτελεί ένα παγκόσμια αναγνωρισμένο πρότυπο που βοηθά οργανισμούς να διαχειρίζονται και να μειώνουν τους κινδύνους της κυβερνοασφάλειας. Ακόμη ο **ENISA** (European Union Agency for Cybersecurity) έχει παίξει κεντρικό ρόλο στην ανάπτυξη πολιτικών και στρατηγικών για την ασφάλεια των δικτύων και των πληροφοριών στην Ευρώπη, προωθώντας τη συνεργασία και την ανταλλαγή πληροφοριών μεταξύ των κρατών μελών της ΕΕ. Επίσης ο **CIS** (Center for Internet Security) έχει προσφέρει εργαλεία και πόρους, όπως οι CIS Controls, που καθοδηγούν τους οργανισμούς στην εφαρμογή βέλτιστων πρακτικών για την προστασία από τις πιο κοινές και επικίνδυνες απειλές. Η **ISACA** (Information Systems Audit and Control Association) έχει βοηθήσει χιλιάδες επαγγελματίες της κυβερνοασφάλειας μέσω της εκπαίδευσης, των πιστοποιήσεων και της καθοδήγησης που παρέχει. Τέλος, το **FIRST** (Forum of Incident Response and Security Teams) έχει ενισχύσει τη συνεργασία σε παγκόσμιο επίπεδο μεταξύ των ομάδων απόκρισης σε περιστατικά ασφαλείας, διευκολύνοντας την ανταλλαγή πληροφοριών και την κοινή αντιμετώπιση των κυβερνοαπειλών. Αυτοί οι

οργανισμοί, με τις πρωτοβουλίες και τα επιτεύγματά τους, έχουν συμβάλει αποφασιστικά στην ενίσχυση της ασφάλειας στον ψηφιακό κόσμο.

Συνοπτικά, η κυβερνοασφάλεια αποτελεί έναν αναπόσπαστο πυλώνα της σύγχρονης ψηφιακής εποχής, η οποία απαιτεί συνεχώς αυξημένη προσοχή και επένδυση. Οι οργανώσεις και οι επιχειρήσεις οφείλουν να ενισχύσουν τα μέτρα προστασίας τους, εφαρμόζοντας προηγμένες τεχνολογίες και στρατηγικές, όπως η ενσωμάτωση τεχνητής νοημοσύνης για την ανίχνευση και αποτροπή επιθέσεων σε πραγματικό χρόνο. Παράλληλα, η εκπαίδευση του προσωπικού και η ανάπτυξη ενός κουλτούρας ασφάλειας αποτελούν κρίσιμα στοιχεία για τη μείωση του ανθρώπινου παράγοντα ως αδύναμο κρίκο.

4.2 ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΙ ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ

Η αλματώδης πρόοδος της τεχνητής νοημοσύνης δεν έχει απλώς ανοίξει νέους ορίζοντες στην τεχνολογία, αλλά έχει επίσης δημιουργήσει σοβαρούς κινδύνους για την κυβερνοασφάλεια. Με την εξέλιξη των αλγορίθμων και την ενσωμάτωση της τεχνητής νοημοσύνης σε ολοένα και πιο σύνθετα συστήματα, οι κυβερνοεπιθέσεις αποκτούν πρωτοφανή επίπεδα πολυπλοκότητας και αποτελεσματικότητας, καθιστώντας τη θωράκιση των πληροφοριακών υποδομών, ιδιαίτερα απαιτητική. Η τεχνητή νοημοσύνη μπορεί να ενισχύσει την κυβερνοασφάλεια, αλλά και να δημιουργήσει νέες προκλήσεις.

Από τη μια πλευρά, η τεχνητή νοημοσύνη μπορεί να αναλύει τεράστια ποσά δεδομένων και να ανιχνεύει ανωμαλίες και απειλές σε πραγματικό χρόνο, ενισχύοντας τις δυνατότητες προστασίας. Από την άλλη, η ίδια τεχνολογία μπορεί να αξιοποιηθεί και από κακόβουλους χρήστες για να δημιουργήσουν πιο εξελιγμένες επιθέσεις. Πιο συγκεκριμένα η τεχνητή νοημοσύνη όχι μόνο διευκολύνει την αυτοματοποίηση των επιθέσεων, αλλά επιτρέπει και τη δημιουργία σύγχρονων εξελιγμένων κακόβουλων λογισμικών ικανών να παρακάμψουν παραδοσιακά μέτρα ασφαλείας.

Επομένως η κλίμακα της απειλής που ενσωματώνει η τεχνητή νοημοσύνη καθίσταται πλέον τόσο εκτεταμένη, που θα μπορούσε σε ακραίες περιπτώσεις να προκαλέσει την κατάρρευση ολόκληρων οργανισμών και επιχειρήσεων. Ακόμη μπορεί να απειλήσει την ελευθερία και την ασφάλεια των χρηστών ψηφιακών συσκευών, τεχνολογιών τεχνητής νοημοσύνης και πλατφορμών κοινωνικής δικτύωσης. Επιπλέον εγείρει ανησυχίες για ενδεχόμενες εθνικές απειλές, καθώς επιθέσεις αυτού του τύπου μπορούν να στοχεύσουν και να διαβρώσουν κρίσιμες εθνικές υποδομές, κυβερνητικούς φορείς και μεγάλες πολυεθνικές επιχειρήσεις, ακόμη και να απειλήσουν την εθνική ασφάλεια και κατά συνέπεια την ασφάλεια των πολιτών.

Η προοπτική αυτή είναι ιδιαίτερα ανησυχητική, καθώς η συνεχής ανάπτυξη της τεχνητής νοημοσύνης ενδέχεται να ενισχύσει περαιτέρω την πολυπλοκότητα και την ένταση των κυβερνοεπιθέσεων, δημιουργώντας πιθανούς κινδύνους για την παγκόσμια σταθερότητα και ασφάλεια. Ως εκ τούτου, η κατανόηση και η

αντιμετώπιση αυτών των κινδύνων είναι επιτακτική για την προστασία των πληροφοριακών υποδομών στο σύγχρονο, ψηφιακό περιβάλλον.

4.2.1 ΚΙΝΔΥΝΟΙ ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑΣ ΑΠΟ ΕΦΑΡΜΟΓΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

Είναι γεγονός ότι, η ενσωμάτωση της τεχνητής νοημοσύνης στις επιχειρηματικές διαδικασίες έχει αυξηθεί σημαντικά. Πλέον, το 65% των εταιρειών χρησιμοποιούν την τεχνητή νοημοσύνη με κάποιο τρόπο εσωτερικά, ενώ το 74% δοκιμάζει τεχνολογίες τεχνητής νοημοσύνης. Επιπλέον, ένα σημαντικό 82% των ηγετών επιχειρήσεων πιστεύει ότι η τεχνητή νοημοσύνη είναι απαραίτητη για τη μελλοντική τους επιτυχία, υπογραμμίζοντας τη διαρκώς αυξανόμενη σημασία της ΑΙ στις επιχειρηματικές στρατηγικές τα επόμενα χρόνια. Ωστόσο ενώ οι ανησυχίες για τις αρνητικές επιπτώσεις που μπορεί να επιφέρει η τεχνητή νοημοσύνη στον κυβερνοχώρο είναι αντιληπτές από πολλούς οργανισμούς και εταιρίες ιδίως από αυτών που κατέχουν υψηλή θέση στον κλάδο της πληροφορικής, υπάρχει μια ομάδα εταιριών που υστερεί στα μέτρα ασφαλείας και στην επαγγελματική και εκπαιδευμένη αντιμετώπιση των κυβερνεπιθέσεων [43].

Στις Ηνωμένες Πολιτείες, το ετήσιο κόστος που προκύπτει από μια κακόβουλη παραβίαση ιδιωτικών δεδομένων ανέρχεται κατά μέσο όρο στα 3,9 εκατομμύρια δολάρια, ενώ η συνολική ετήσια επίδραση στην παγκόσμια οικονομία από τις κυβερνεπιθέσεις εκτιμάται περίπου στα 400 δισεκατομμύρια δολάρια [44].

Οι αλγόριθμοι μηχανικής μάθησης και ένα μεγάλο ποσοστό συστημάτων τεχνητής νοημοσύνης αναπτύσσονται μέσω εκπαίδευσης, με σκοπό τη βέλτιστη απόδοση τους και την ανταπόκριση σε διάφορες νέες καταστάσεις. Επομένως για τη διαδικασία της εκπαίδευσης τους χρησιμοποιούν επανειλημμένα νέα δεδομένα με σκοπό να βελτιωθούν σταδιακά [43]. Αυτό ωστόσο μπορεί να αποτελέσει απειλή επειδή τα συστήματα τεχνητής νοημοσύνης είναι αυτόνομα και παίρνουν αποφάσεις με αυτοματοποιημένο τρόπο χωρίς καθημερινή ανθρώπινη επίβλεψη. Αυτό μπορεί να τα καταστήσει ευάλωτα σε κακόβουλες επιθέσεις που παραμένουν αόρατες για μεγάλα χρονικά διαστήματα [43]. Επιπλέον υπάρχει μία έλλειψη διαφάνειας στους μηχανισμούς λήψης αποφάσεων των συστημάτων τεχνητής νοημοσύνης η οποία καθιστά δύσκολη την ανίχνευση και κατανόηση των παραβιάσεων, ακόμη και όταν αυτές ανιχνεύονται. Αυτό σημαίνει ότι ακόμη και αν εντοπιστεί παραβίαση, ο σκοπός της μπορεί να παραμείνει αδιαφανής [43].

Ο κυριότερος λόγος ωστόσο που η τεχνητή νοημοσύνη μπορεί να αποτελέσει κίνδυνο για τον κυβερνοχώρο είναι η δυνατότητά να αναλύει τη συμπεριφορά των χρηστών, κάτι που επιτρέπει να εντοπιστούν και να εκμεταλλευτούν συγκεκριμένα μοτίβα για να διεισδύσουν σε συστήματα ασφαλείας. Η τεχνητή νοημοσύνη έχει την ικανότητα να επεξεργάζεται και να αναλύει τεράστιους όγκους δεδομένων, ανακαλύπτοντας μοτίβα που οι άνθρωποι αναλυτές

πιθανόν να μην παρατηρούσαν [45] . Με αυτόν τον τρόπο, μπορεί να βοηθήσει στην ανάπτυξη εξειδικευμένων επιθέσεων που ξεπερνούν τους παραδοσιακούς ανθρώπινους περιορισμούς. Αυτές οι εξελιγμένες τεχνικές επιθέσεων θα μπορούσαν να ενσωματώσουν τη χρήση των μοτίβων για να βελτιώσουν την ακρίβεια και την αποτελεσματικότητα των επιθέσεων phishing, malware, και άλλων απειλών στον κυβερνοχώρο. Επομένως οι κακόβουλοι φορείς μπορούν να χρησιμοποιήσουν αυτά τα δεδομένα και τις ανιχνευμένες ανωμαλίες για να σχεδιάσουν στοχευμένες επιθέσεις.

Οι κατηγορίες κυβερνοεπιθέσεων και κακόβουλων λογισμικών υποστηριζόμενων από ΤΝ αναλύονται εκτενώς παρακάτω.

1) Phishing: Όπως προαναφέρθηκε η τεχνητή νοημοσύνη έχει αναβαθμίσει τις επιθέσεις phishing σε επίπεδα αδιανόητα μέχρι πρότινος, καθιστώντας τις πιο εξελιγμένες και δύσκολα ανιχνεύσιμες από τα συστήματα κυβερνοασφάλειας. Ένας από τους πιο ανησυχητικούς τρόπους με τους οποίους οι κακόβουλοι φορείς εκμεταλλεύονται την τεχνητή νοημοσύνη είναι μέσω του αλγορίθμου DeepPhish [46]. Αυτός ο αλγόριθμος μηχανικής μάθησης αναλύει προηγούμενες επιτυχημένες επιθέσεις phishing και μαθαίνει μοτίβα από αυτές, δημιουργώντας νέες συνθετικές διευθύνσεις URL που είναι εξαιρετικά δύσκολο να ανιχνευθούν. Αυτές οι διευθύνσεις URL σχεδιάζονται ώστε να παρακάμπτουν τις παραδοσιακές μεθόδους ανίχνευσης, μιμούμενες τις νόμιμες ιστοσελίδες με εξαιρετική ακρίβεια [46].

Παράλληλα, οι επιθέσεις phishing με χρήση ΑΙ δεν περιορίζονται μόνο στη δημιουργία συνθετικών διευθύνσεων URL. Τεχνικές όπως το deepfake phishing χρησιμοποιούν αλγόριθμους ΑΙ για να δημιουργήσουν ψεύτικες φωνητικές ή βιντεοκλήσεις, οι οποίες μπορούν να πείσουν τον στόχο για την αυθεντικότητα του μηνύματος και να οδηγήσουν σε επιτυχείς επιθέσεις. Αυτός ο τύπος επίθεσης μπορεί να αξιοποιηθεί για την παραπλάνηση στελεχών επιχειρήσεων ή άλλων υψηλού προφίλ ατόμων, κάνοντάς τους να πιστέψουν ότι επικοινωνούν με κάποιον αξιόπιστο συνεργάτη [46].

Επιπλέον, οι επιτιθέμενοι χρησιμοποιούν τεχνητή νοημοσύνη και για την ανάπτυξη πιο στοχευμένων επιθέσεων, όπως το spear phishing. Το spear phishing με ΑΙ περιλαμβάνει την ανάλυση μεγάλου όγκου δεδομένων για να εξατομικεύσουν τα μηνύματα phishing, καθιστώντας τα πιο πειστικά και αυξάνοντας τις πιθανότητες επιτυχίας τους. Με την ικανότητα της ΑΙ να ενσωματώνει προσωπικά δεδομένα στις επιθέσεις, αυτές γίνονται πιο επικίνδυνες και δύσκολα ανιχνεύσιμες, απειλώντας ακόμα και τα πιο προηγμένα συστήματα κυβερνοασφάλειας.

Συνολικά, η ενσωμάτωση της ΑΙ στις επιθέσεις phishing δεν είναι απλώς μια εξέλιξη των υφιστάμενων απειλών, αλλά μια ποιοτική αλλαγή που αναβαθμίζει την επικινδυνότητα και την αποτελεσματικότητα αυτών των επιθέσεων σε πρωτόγνωρα επίπεδα.

2) AI-Driven Password Guessing/ Cracking : Ένας ακόμη τρόπος που οι κακόβουλοι εκμεταλλεύονται την επιθετική τεχνητή νοημοσύνη (TN) είναι χρήση εξελιγμένων αλγορίθμων TN για την παραβίαση και εύρεση κωδικών πρόσβασης. Οι τεχνικές αυτές έχουν αναβαθμιστεί σημαντικά με τη χρήση αλγορίθμων TN καθιστώντας τις επιθέσεις εξαιρετικά αποτελεσματικές και δύσκολα ανιχνεύσιμες.

Μια τέτοια είδους επίθεση αποτελεί το brute-force attack, που ενώ αποτελεί μια από τις πιο κλασικές μεθόδους παραβίασης, έχει αναβαθμιστεί σημαντικά με την ενσωμάτωση της TN. Ενώ παραδοσιακά το brute-force attack απαιτεί την εξαντλητική δοκιμή όλων των δυνατών συνδυασμών χαρακτήρων, οι αλγόριθμοι TN μπορούν να μειώσουν δραστικά τον απαιτούμενο χρόνο με την πρόβλεψη και ιεράρχηση των πιο πιθανών συνδυασμών, αξιοποιώντας μοτίβα που αντλούν από προηγούμενες διαρροές κωδικών και την ανάλυση μεγάλων συνόλων δεδομένων[46]. Έτσι πλέον με βάση προηγμένων μεθόδων ανάλυσης δεδομένων και μηχανικής μάθησης, μπορούν να εκτελούνται εξελιγμένες επιθέσεις brute-force με ταχύτητες που υπερβαίνουν κατά πολύ τις παραδοσιακές μεθόδους. Η δυνατότητα της TN να μαθαίνει από μοτίβα και να προσαρμόζεται σε νέες συνθήκες επιτρέπει τη γρήγορη πρόβλεψη πιθανών συνδυασμών κωδικών, μειώνοντας δραστικά το χρόνο που απαιτείται για την παραβίαση ενός λογαριασμού.

Παράλληλα, οι αλγόριθμοι TN αναδεικνύονται εξαιρετικά ικανοί για την εύρεση κωδικών. Οι αλγόριθμοι αυτοί επιτρέπουν την ανάπτυξη αυτοματοποιημένων τεχνικών που βασίζονται σε γεννητικά αντιπαραγωγικά δίκτυα (GAN) [47]. Μαθαίνοντας την κατανομή των κωδικών πρόσβασης από πραγματικές παραβιάσεις, αναπαράγοντας παρόμοια μοτίβα για την πρόβλεψη νέων κωδικών. Η προσέγγιση αυτή συνδυάζει τη χρήση εξαντλητικών συνδυασμών χαρακτήρων, λεξικών που περιέχουν πιθανές λέξεις-κλειδιά, και κανόνες για τη δημιουργία μετασχηματισμένων κωδικών, όπως η συνένωση λέξεων, για την εξεύρεση του σωστού κωδικού. Αυτός ο συνδυασμός αυξάνει τις πιθανότητες επιτυχούς πρόβλεψης, καθιστώντας τις επιθέσεις αυτές πιο αποτελεσματικές

Επιπροσθέτως, η TN ενισχύει τις επιθέσεις κλοπής κωδικών μέσω της ανάπτυξης κακόβουλου λογισμικού που είναι ικανό να καταγράφει και να αναλύει πληκτρολογήσεις ή να εκμεταλλεύεται ευπάθειες σε συστήματα, αποσπώντας κωδικούς πρόσβασης χωρίς να γίνει αντιληπτή. Οι ερευνητές Lee και Yim εφάρμοσαν μια σειρά από αλγορίθμους μηχανικής μάθησης, όπως οι K-Nearest Neighbors (KNN), logistic regression (LR), decision tree (DT), linear support vector classifier (SVC), random forest (RF), support vector machine (SVM), gradient boosting regression tree, και multilayer perceptron, για την ταξινόμηση δεδομένων πληκτρολογήσεων [48]. Το μοντέλο τους πέτυχε ακρίβεια 96,2% στην κλοπή δεδομένων πληκτρολόγησης, αποδεικνύοντας ότι οι κυβερνοεγκληματίες μπορούν να αποσπών πραγματικά δεδομένα πληκτρολογήσεων σε πραγματικό χρόνο με τεχνικές AI [48]. Επομένως η ενσωμάτωση της AI σε αυτές τις επιθέσεις οδηγεί σε μια νέα εποχή κυβερνοαπειλών, όπου οι παραδοσιακές μέθοδοι ασφάλειας καθίστανται

ανεπαρκείς, απαιτώντας την ανάπτυξη εξίσου προηγμένων συστημάτων ανίχνευσης και άμυνας.

3) *Intelligent Captcha Attack and Manipulation* : Η εξέλιξη των μεθόδων επίθεσης στον τομέα της κυβερνοασφάλειας έχει επίσης καταστήσει την αντιμετώπιση των συστημάτων CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) ολοένα και πιο κρίσιμη. Το Intelligent Captcha Attack and Manipulation (ICAM) αντιπροσωπεύει μια προηγμένη κατηγορία επιθέσεων που αξιοποιούν εξελιγμένους αλγορίθμους μηχανικής μάθησης και τεχνικές τεχνητής νοημοσύνης για την παραβίαση των προστατευτικών μηχανισμών CAPTCHA [49].

Τα CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart) είναι συστήματα που σχεδιάζονται για να διακρίνουν αν ο χρήστης μιας εφαρμογής ή ιστοσελίδας είναι άνθρωπος ή υπολογιστής, με σκοπό την αποτροπή αυτοματοποιημένων επιθέσεων, όπως bots και σάρωση δεδομένων. Τα CAPTCHA λειτουργούν συνήθως ζητώντας από τον χρήστη να ολοκληρώσει ένα τεστ που είναι εύκολο για τους ανθρώπους αλλά δύσκολο για τα αυτόματα συστήματα. Αυτό μπορεί να περιλαμβάνει την αναγνώριση παραμορφωμένων χαρακτήρων, την επιλογή εικόνων που πληρούν συγκεκριμένα κριτήρια ή την επίλυση απλών μαθηματικών προβλημάτων [49].

Το Intelligent Captcha Attack and Manipulation (ICAM) ανατρέπει την αποτελεσματικότητα των παραδοσιακών CAPTCHA μέσω προηγμένων τεχνικών επίθεσης που αξιοποιούν μηχανική μάθηση και τεχνητή νοημοσύνη. Συγκεκριμένα, τα συστήματα ICAM περιλαμβάνουν αλγορίθμους που είναι ικανοί να αναλύουν και να αποκωδικοποιούν τις μεθόδους CAPTCHA. Χρησιμοποιώντας τεχνικές μηχανικής μάθησης, το ICAM μπορεί να εκπαιδεύσει μοντέλα που αναγνωρίζουν πρότυπα και λογικές που χρησιμοποιούνται στα CAPTCHA. Αυτά τα μοντέλα μπορούν να μάθουν πώς να αναγνωρίζουν χαρακτήρες, να αναλύουν παραμορφώσεις ή να κατανοούν τις απαιτήσεις των εικόνων πιο αποτελεσματικά από τα παραδοσιακά εργαλεία [49]. Επίσης οι επιθέσεις ICAM μπορούν να περιλαμβάνουν αυτοματοποιημένες διαδικασίες που προσαρμόζονται συνεχώς για να αναγνωρίσουν νέες παραλλαγές CAPTCHA [49]. Αυτό καθιστά δύσκολη την εφαρμογή στατικών λύσεων ασφαλείας, καθώς το σύστημα ICAM προσαρμόζεται και εξελίσσεται παράλληλα με τις αλλαγές στα CAPTCHA. Τέλος το ICAM μπορεί να εκμεταλλευτεί συγκεκριμένες αδυναμίες ή ευπάθειες στα CAPTCHA, όπως αδύνατους αλγόριθμους παραμόρφωσης χαρακτήρων ή ελαττωματικούς μηχανισμούς επαλήθευσης, για να βρει τρόπους να παρακάμψει τις προστασίες [49].

Συνολικά, το ICAM έχει τη δυνατότητα να μειώσει την αποτελεσματικότητα των CAPTCHA, επηρεάζοντας την προστασία των διαδικτυακών υπηρεσιών από αυτοματοποιημένες επιθέσεις και αναγκάζοντας τους οργανισμούς να αναπτύξουν νέες και πιο εξελιγμένες μεθόδους ασφαλείας για την αντιμετώπιση αυτής της πρόκλησης.

4) *Κακόβουλο λογισμικό ευφυσής απόκρυψης και αποφυγή* : Όπως είναι αναμενόμενο τα νέα κακόβουλα λογισμικά είναι εξελιγμένα αυτόνομα και

παρουσιάζουν ένα είδος ευφυΐας. Πιο συγκεκριμένα έχουν αναπτυχθεί λογισμικά τα οποία μέσω της χρήσης μοντέλων (TN), κατέχουν τη δυνατότητα απόκρυψης των στοιχείων τους (δομή κώδικας) ώστε να αναιρούν την ανιχνευσιμότητα τους από συστήματα ασφαλείας. Αυτό αντιπροσωπεύει μια αυξανόμενη απειλή στον τομέα της κυβερνοασφάλειας [47]. Αυτό το είδος λογισμικού χρησιμοποιεί αλγορίθμους μηχανικής μάθησης και άλλες τεχνολογίες TN για να προσαρμόζει συνεχώς τη συμπεριφορά του και να παρακάμπτει τα σύγχρονα συστήματα ανίχνευσης. Σημαντικές κατηγορίες κακόβουλου λογισμικού που ενσωματώνουν τέτοιες τεχνικές περιλαμβάνουν το **polymorphic και metamorphic malware** [86], το οποίο μπορεί να αλλάζει τη μορφή του ή τον κώδικά του για να αποφεύγει τα αντιϊκά λογισμικά, και το **fileless malware**, το οποίο εκτελείται στη μνήμη του συστήματος χωρίς να αφήνει ίχνη σε αρχεία, καθιστώντας την ανίχνευσή του ιδιαίτερα δύσκολη. Επιπλέον, το κακόβουλο λογισμικό που χρησιμοποιείται σε **advanced persistent threats (APTs)** ενσωματώνει τεχνικές TN για την προσαρμογή και την απόκρυψη της δραστηριότητάς του σε βάθος χρόνου, επιτρέποντας παρατεταμένη πρόσβαση σε κρίσιμα συστήματα χωρίς να γίνεται αντιληπτό. Αυτά τα εξελιγμένα Artificial Intelligence-powered malware αποτελούν σοβαρή πρόκληση για τους επαγγελματίες της κυβερνοασφάλειας, καθώς απαιτούν νέα και πιο εξελιγμένα εργαλεία και στρατηγικές για την αποτελεσματική αντιμετώπισή τους.

5) DeepHack : Ακόμη μια κατηγορία ευφυούς λογισμικού που είναι και λογισμικό αυτοεκμάθησης αποτελεί το μοντέλο μηχανικής μάθησης DeepHack [50]. Το συγκεκριμένο λογισμικό έχει τη δυνατότητα να αποφασίζει και να δρά κακόβουλα, αυτόνομα χωρίς την ανθρώπινη εμπλοκή. Αυτό επιτυγχάνεται καθώς το κακόβουλο λογισμικό εκπαιδεύεται από έναν δικό του αλγόριθμο αυτοεκμάθησης, έτσι ώστε να μπορεί να ξεκινήσει τις δικές του επιθέσεις μειώνοντας το επίπεδο γνώσης που χρειάζεται ο επιτιθέμενος για να επηρεάσει και να εκμεταλλευτεί επιτυχώς το σύστημα-στόχο.

6) Επίθεση DDoS με γνώμονα την τεχνητή νοημοσύνη : Οι επιθέσεις DDoS (Distributed Denial of Service) είναι μια από τις πιο κοινές και επικίνδυνες απειλές στον χώρο της κυβερνοασφάλειας. Ο στόχος αυτών των επιθέσεων είναι η διακοπή της κανονικής λειτουργίας ενός δικτύου, μιας υπηρεσίας ή ενός ιστότοπου, υπερφορτώνοντας τους πόρους του συστήματος με έναν τεράστιο όγκο αιτημάτων.

Οι επιθέσεις DDoS λειτουργούν μέσω της διανομής της επίθεσης σε πολλαπλές πηγές, συνήθως χιλιάδες ή και εκατομμύρια μολυσμένων υπολογιστών ή συσκευών, τα οποία είναι συνήθως γνωστά ως "bots" και αποτελούν μέρος ενός ευρύτερου δικτύου, το λεγόμενο "botnet". Οι επιτιθέμενοι ελέγχουν το botnet και το χρησιμοποιούν για να στέλνουν τεράστιο όγκο αιτημάτων προς τον στόχο, καταναλώνοντας τους πόρους του συστήματος και καθιστώντας το ανίκανο να εξυπηρετήσει νόμιμους χρήστες.

Οι πλέον σύγχρονες επιθέσεις DDoS με ενσωματωμένη την τεχνητή νοημοσύνη είναι ακόμα πιο ισχυρές και αντιπροσωπεύουν μια σημαντική πρόκληση για την κυβερνοασφάλεια, καθώς εξαλείφουν πλήρως την ανάγκη για ανθρώπινη παρέμβαση κατά την εκτέλεσή τους [46].

Οι μηχανές, καθοδηγούμενες από προηγμένους αλγορίθμους AI, επιτίθενται πλέον σε εφαρμογές και υπερσύγχρονες αμυντικές υποδομές με εξαιρετική ακρίβεια και προσαρμοστικότητα. Σε περιπτώσεις όπου μια συγκεκριμένη επιτιθέμενη υπογραφή αποτύχει, η AI μπορεί να αναγνωρίσει την αποτυχία και να προσαρμοστεί αυτόματα, επιλέγοντας διαφορετική υπογραφή για τη συνέχιση της επίθεσης [46]. Αυτό το επίπεδο αυτονομίας και προσαρμοστικότητας καθιστά τις επιθέσεις DDoS με βάση την AI εξαιρετικά δύσκολες να αντιμετωπιστούν, υπογραμμίζοντας την αυξανόμενη ανάγκη για εξελιγμένες στρατηγικές άμυνας.

Ακόμη η τεχνητή νοημοσύνη επιτρέπει στις επιθέσεις αυτές να διαχειρίζονται και να κατευθύνουν τα bots στο botnet με τον πιο αποδοτικό τρόπο. Για παράδειγμα, εντοπίζουν ποια bots έχουν την καλύτερη πρόσβαση στον στόχο και τα αξιοποιούν πιο εντατικά, ελαχιστοποιώντας τις απώλειες και βελτιστοποιώντας τον αντίκτυπο της επίθεσης [46]. Τέλος οι αλγόριθμοι τεχνητής νοημοσύνης μπορούν να αναλύουν συνεχώς τον στόχο, αναζητώντας τρωτά σημεία σε πραγματικό χρόνο. Μόλις εντοπιστούν αυτά τα αδύνατα σημεία, η τεχνητή νοημοσύνη μπορεί να κατευθύνει την επίθεση εκεί, καθιστώντας την πιο αποτελεσματική.

Συνοπτικά οι σύγχρονες επιθέσεις DDoS με γνώμονα την τεχνητή νοημοσύνη μπορούν να επηρεάσουν πολλούς τομείς της κοινωνίας. Αναλυτικότερα μπορούν να πλήξουν επιχειρήσεις όλων των μεγεθών, από μικρές επιχειρήσεις έως πολυεθνικές εταιρείες, καθώς οι επιθέσεις αυτές μπορεί να προκαλέσουν διακοπές στις υπηρεσίες τους, απώλειες εσόδων και ζημιές στη φήμη τους. Επίσης οι κυβερνητικές υπηρεσίες και οι υποδομές εθνικής άμυνας είναι επίσης σημαντικοί στόχοι. Μια επιτυχημένη επίθεση DDoS μπορεί να διαταράξει κρίσιμες λειτουργίες, όπως υπηρεσίες ασφαλείας, συστήματα επικοινωνίας, και πληροφοριακές υποδομές, καθιστώντας τις κυβερνήσεις και τις στρατιωτικές δυνάμεις ανίκανες να ανταποκριθούν αποτελεσματικά σε επείγουσες καταστάσεις. Ακόμη απειλούνται υγειονομικές υπηρεσίες, εταιρείες που διαχειρίζονται υποδομές κρίσιμης σημασίας, όπως ενέργεια, ύδρευση, τηλεπικοινωνίες, μεταφορές και εκπαιδευτικά Ιδρύματα.

Όλα τα παραπάνω καθιστούν την προστασία απέναντι στις σύγχρονες επιθέσεις DDoS κρίσιμη ανάγκη για όλους τους εμπλεκόμενους.

7) Κυβερνοεπιθέσεις στα δίκτυα IoT - Internet of things (IoT) : Η τεχνητή νοημοσύνη και το Διαδίκτυο των Πραγμάτων(IoT) έχουν φέρει στο προσκήνιο μια εποχή ασύγκριτων ευκαιριών, αναδιαμορφώνοντας ριζικά τον τρόπο με τον οποίο αλληλοεπιδρούμε με το περιβάλλον μας και αξιοποιώντας τον απέραντο χώρο των δεδομένων και τη διασυνδεσιμότητα των συσκευών (IoT). Με το IoT, οι βιομηχανίες είναι σε θέση να βελτιστοποιήσουν τις διεργασίες τους, να παρακολουθούν και να ανταποκρίνονται στις ανάγκες των καταναλωτών και να προσφέρουν καινοτόμες λύσεις που εξυπηρετούν την κοινωνία. Με τη συνεχή εξέλιξη του IoT, αναμένεται να δημιουργηθούν περισσότερες ευκαιρίες για την ανάπτυξη και την ανακατασκευή του κόσμου μας προς μια πιο έξυπνη και βελτιστοποιημένη κατεύθυνση.

Ωστόσο οι κίνδυνοι όσον αφορά το IoT διαμοιράζονται και σε όλα τα συστατικά του μέρη.

7.1) Κυβερνοεπιθέσεις σε IoT επιχειρήσεων και οργανισμών : Το Διαδίκτυο των Πραγμάτων (IoT) έχει φέρει επανάσταση στον τρόπο με τον οποίο οι επιχειρήσεις και οι οργανισμοί λειτουργούν, προσφέροντας αυξημένη αυτοματοποίηση, βελτιωμένη παρακολούθηση των διαδικασιών και νέες ευκαιρίες για καινοτομία. Ωστόσο, μαζί με αυτές τις σημαντικές αναβαθμίσεις, έχουν αναδειχθεί και νέες απειλές που θέτουν σε κίνδυνο την ασφάλεια των επιχειρήσεων.

Αφενός, όλοι οι κίνδυνοι διείσδυσης στα δίκτυα οργανισμών και εταιρειών είναι παρόντες και στο περιβάλλον του IoT, μέσω εσφαλμένης αδειοδότησης και εγκατάστασης κακόβουλων λογισμικών [51]. Οι διαφορετικές συσκευές που συνδέονται στο δίκτυο, με ποικίλες αρχιτεκτονικές ηλεκτρονικών κυκλωμάτων, αυξάνουν την τρωτότητα του συστήματος, καθιστώντας το πιο ευάλωτο σε επιθέσεις όσο αυξάνεται ο αριθμός των συνδεδεμένων συσκευών. Στην επιφάνεια παροχής υπηρεσιών, η διαρροή πληροφοριών και η μη εξουσιοδοτημένη πρόσβαση αποτελούν σημαντικούς κινδύνους. Παραδείγματα παραβιάσεων περιλαμβάνουν την αποστολή ανεπιθύμητης αλληλογραφίας (spam), τη διαρροή προσωπικών δεδομένων, την παραβίαση αισθητήρων και καμερών, καθώς και την πλήρη παράλυση των συνδεδεμένων συσκευών, απειλώντας την ακεραιότητα και τη λειτουργικότητα ολόκληρων οργανισμών [51].

Επομένως απειλείται σημαντικά η οικονομική και επαγγελματική ανάπτυξη ενός οργανισμού ή μίας επιχείρησης. Όμως αυτές είναι μερικές μόνο από τις απειλές, που μπορούν να υποστούν οι επιχειρήσεις ή οι οργανισμοί. Πέρα από την κλοπή υλικών πόρων και χρηματικών ποσών, οι επιθέσεις μπορούν να πλήξουν βαθύτερα, με την κλοπή ευρεσιτεχνιών, εφευρέσεων και στρατηγικών σχεδίων που είναι θεμελιώδη για το μέλλον και την ταυτότητα της επιχείρησης. Επίσης οι συνέπειες της άρνησης παροχής υπηρεσιών στους πελάτες, που μπορεί να προκύψουν από κυβερνοεπιθέσεις, καθώς και η κλοπή προσωπικών δεδομένων, έχουν τη δυνατότητα να προκαλέσουν ανεπανόρθωτη ζημία. Η αρνητική δημοσιότητα και η φθορά της φήμης της επιχείρησης μπορούν να οδηγήσουν σε δραματική μείωση της εμπιστοσύνης τόσο των πελατών όσο και των μελλοντικών συνεργατών. Το αποτέλεσμα μπορεί να είναι η υποβάθμιση της επιχείρησης, ο περιορισμός των αναπτυξιακών της δυνατοτήτων και η μακροχρόνια υπονόμευση της θέσης της στην αγορά.

7.2) Κυβερνοεπιθέσεις σε IoT στρατιωτικών μονάδων και υποδομών : Η ταχεία υιοθέτηση του Διαδικτύου των Πραγμάτων (IoT) στον τομέα της εθνικής άμυνας, των πολεμικών μονάδων και των στρατιωτικών υποδομών μιας χώρας, παρέχει τη δυνατότητα ταχέων ενεργειών και συχνά αυτοματοποιημένων. Έτσι παρέχεται η δυνατότητα απομακρυσμένου ελέγχου συσκευών που θα παρέχουν υποστήριξη στο κομμάτι, της ανίχνευσης, παρακολούθησης διαβίβασης πληροφοριών [51], καθώς και υπό περιπτώσεις αυτοματοποιημένων ενεργειών πυροδότησης.

Ωστόσο ελλοχεύουν σημαντικές προκλήσεις και απειλές, οι οποίες υπονομεύουν την ασφάλεια και την αποδοτικότητα των σύγχρονων πολεμικών επιχειρήσεων. Σε ένα περιβάλλον δικτυοκεντρικού πολέμου (NCW), όπου η πληροφορία και η διασύνδεση των στοιχείων μάχης, οπλικών συστημάτων και στρατιωτικού εξοπλισμού παίζουν καθοριστικό ρόλο, η παραβίαση των δικτύων IoT μπορεί να οδηγήσει σε ανυπολόγιστες ζημίες.

Αρχικά, ένας επιτιθέμενος μπορεί να διεισδύσει σε δίκτυα επικοινωνιών και να υποκλέψει ευαίσθητα δεδομένα που αφορούν την εθνική άμυνα της χώρας, θέτοντας σε σοβαρό κίνδυνο την εθνική ασφάλεια και την προστασία των πολιτών. Αυτά τα δεδομένα μπορεί να περιλαμβάνουν θέσεις καταφυγίων για τους πολίτες σε περίπτωση επίθεσης, τοποθεσίες όπου σταθμεύουν στρατεύματα, στρατιωτικές στρατηγικές, καθώς και τοποθεσίες κρυφών βάσεων.

Επιπλέον, υπάρχει η δυνατότητα ο επιτιθέμενος να διαταράξει την ομαλή λειτουργία και επικοινωνία των στρατιωτικών δυνάμεων. Η παραβίαση ενός δικτύου επικοινωνιών, η υποκλοπή ευαίσθητων στοιχείων που αφορούν άτομα, εγκαταστάσεις και πρωτόκολλα ασφαλείας, καθώς και η ψευδής αναπαράσταση φίλιων οντοτήτων – δηλαδή η χρήση κακόβουλων μηχανημάτων και λογισμικών από εχθρικές κυβερνήσεις ή εχθρικές στρατιωτικές μονάδες, που παριστάνουν ότι ανήκουν στις στρατιωτικές μονάδες – μπορεί να οδηγήσει σε κακοπροαίρετη καθοδήγηση των στρατευμάτων. Όλα αυτά μπορούν να οδηγήσουν ακόμα και σε τροπή πυρών προς φίλιες οντότητες (τόσο άμεσα, όσο και έμμεσα με φόρτωση ψευδών στοιχείων και συντεταγμένων ώστε οι φίλιες οντότητες να χαρακτηρίζονται ως εχθρικές) [51].

Ακόμη, ο επιτιθέμενος μπορεί να παρεμβάλει την επικοινωνία μεταξύ στρατιωτικών μονάδων, με σκοπό να εμποδίσει την αποστολή στρατευμάτων εκεί που είναι απαραίτητο και να παραπληροφορήσει τα μέλη των μονάδων, οδηγώντας σε εσφαλμένες ενέργειες και ανεπαρκή διαφύλαξη κρίσιμων περιοχών.

Ωστόσο ο επιτιθέμενος μπορεί να αποδυναμώσει τη μαχητική ικανότητα του αντιπάλου, όχι μόνο υποβαθμίζοντας την αποτελεσματικότητα των μη επανδρωμένων συστημάτων μάχης, αλλά και προκαλώντας σύγχυση μεταξύ των συμμάχων, στρέφοντας τον έναν εναντίον του άλλου. Μέσω της διαστρέβλωσης δεδομένων, ένας επιτιθέμενος μπορεί να αποκτήσει έλεγχο του δικτύου και να προκαλέσει επίθεση από στρατιωτική υποδομή μιας χώρας προς μια άλλη, δημιουργώντας σκόπιμα αναταραχή ή ακόμα και πόλεμο μεταξύ δύο χωρών.

Αυτό, επίσης, θα μπορούσε να αξιοποιηθεί ως εργαλείο για την πρόκληση επιθέσεων από μια χώρα προς μια άλλη, επιτρέποντας στους πραγματικούς δράστες να αποφύγουν την ευθύνη και να επιρρίψουν την κατηγορία σε ένα ανυποψίαστο κράτος, το οποίο δεν φέρει καμία ευθύνη.

Η εκμετάλλευση των ευπαθειών αυτών των συστημάτων, μέσω της παραποίησης δεδομένων και ψευδούς απεικόνισης φίλιων οντοτήτων, μπορεί να έχει καταστροφικές συνέπειες, καθιστώντας τις επιπτώσεις μιας κυβερνοεπίθεσης

στον στρατιωτικό τομέα αισθητά σοβαρότερες από οποιοδήποτε άλλο δίκτυο που περιέχει εφαρμογές IoT.

7.3) Κυβερνοεπιθέσεις σε Συνδεδεμένα και Αυτόματα οχήματα - Connected and Automated Vehicles (CAVs) :Στην εποχή του Διαδικτύου των Πραγμάτων (IoT), μια ακόμη πρωτοπόρα τεχνολογία αποτελούν τα αυτόματα οχήματα. Τα αυτόματα οχήματα δεν αντιπροσωπεύουν απλώς την αιχμή της τεχνολογίας, αλλά αναδεικνύονται ως κεντρικός άξονας της μετασχηματιστικής δύναμης που αλλάζει τη φύση της μετακίνησης και της καθημερινής ζωής. Η συνδυασμένη χρήση προηγμένων αισθητήρων, αναλυτικών αλγορίθμων και δικτύων δεδομένων προσφέρει μια νέα διάσταση στην έννοια της οδήγησης, καθιστώντας την όχι μόνο πιο ασφαλή και αποδοτική, αλλά και αναπόσπαστο τμήμα ενός ευρύτερου οικοσυστήματος πληροφοριών.

Σε αυτό το πλαίσιο, τα αυτόματα οχήματα δεν είναι απλώς τεχνολογικά επιτεύγματα, αλλά πρωτοπόροι στην αναγέννηση των υποδομών μεταφορών και της καθημερινής αλληλεπίδρασης με την τεχνολογία. Ο συνδυασμός της αυτονομίας με τη δυνατότητα συνεχούς επικοινωνίας και ανταλλαγής δεδομένων ανοίγει νέους δρόμους για τη βιωσιμότητα, την ασφάλεια και την ευκολία των μεταφορών, τοποθετώντας τη βιομηχανία σε μια νέα εποχή και καθιστώντας την κινητικότητα ένα ζωντανό κομμάτι του δικτύου του μέλλοντος.

Ωστόσο παρά τις αδιαμφισβήτητες καινοτομίες και τα πλεονεκτήματα που προσφέρουν τα αυτόματα οχήματα στην εποχή του Διαδικτύου των Πραγμάτων, η υιοθέτηση αυτής της τεχνολογίας φέρνει στο προσκήνιο αρκετές προκλήσεις και κινδύνους. Καθώς τα οχήματα γίνονται ολοένα και πιο εξαρτημένα από σύνθετα δίκτυα αισθητήρων, συστημάτων επικοινωνίας και τεχνητής νοημοσύνης, αναδύονται νέες ανησυχίες που χρήζουν προσεκτικής αξιολόγησης και αντιμετώπισης.

Ένας από τους κυριότερους κινδύνους των αυτόματων οχημάτων αποτελεί η ασφάλεια των προσωπικών δεδομένων, καθώς πιθανές κυβερνοεπιθέσεις μπορούν να οδηγήσουν σε διαρροή προσωπικών ευαίσθητων πληροφοριών (Bella et al., 2023) [52]. Τα αυτόματα οχήματα συλλέγουν και αναλύουν τεράστιες ποσότητες προσωπικών δεδομένων και πληροφοριών σχετικά με τις διαδρομές και τις συνήθειες των χρηστών. Αυτή η συγκέντρωση δεδομένων, αν δεν διαχειριστεί σωστά, μπορεί να καταστήσει τα οχήματα ευάλωτα σε επιθέσεις κυβερνοασφάλειας και κλοπή δεδομένων, απειλώντας την ιδιωτικότητα και την ασφάλεια των επιβατών.

Επιπλέον, η εξάρτηση των αυτόματων οχημάτων από τεχνολογικές υποδομές και συνδέσεις με το δίκτυο μπορεί να οδηγήσει σε σοβαρές επιπτώσεις σε περίπτωση τεχνικών αποτυχιών ή διακοπών. Ένα αποτυχημένο σύστημα επικοινωνίας ή μια βλάβη στους αισθητήρες μπορεί να προκαλέσει σοβαρά προβλήματα. Τα πράγματα επιδεινώνονται σημαντικά αν αναλογιστεί κανείς ότι ένα σοβαρό ατύχημα ή μια βλάβη σε ένα αυτόματο όχημα μπορεί να είναι αποτέλεσμα μιας εσκεμμένης κακόβουλης επίθεσης με στόχο την υπονόμηση της ασφάλειας των οχημάτων και των επιβατών τους. Οι επιτιθέμενοι, επομένως, μπορούν να εκμεταλλευτούν τη διασυνδεσιμότητα που προσφέρουν τα οδικά δίκτυα IoT για

να προκαλέσουν κακόβουλες ενέργειες, κρύβοντας όχι μόνο τα στοιχεία τους, αλλά και ρίχνοντας τις ευθύνες σε αθώους τρίτους. Υποβάλλοντας ψευδείς κατηγορίες, μπορούν να ισχυριστούν ότι ευθύνονται άλλοι για τυχόν απόπειρες σωματικής βλάβης ή άλλες σοβαρές κατηγορίες. Ως αποτέλεσμα, πολιτικοί ή απλοί πολίτες ενδέχεται να βρεθούν να διακινδυνεύουν όχι μόνο την προσωπική τους ασφάλεια αλλά και να κατηγορηθούν άδικα για σοβαρά αδικήματα, όπως απόπειρα σωματικής βλάβης, και να στιγματιστούν κοινωνικά ως ανασφαλείς ή επικίνδυνοι οδηγοί.

Όλα αυτά όμως επιτυγχάνονται από πιθανές παραβιάσεις όπως οι εξής :

1. Παραβίαση φυσικών αισθητήρων και οργάνων επικοινωνίας του οχήματος με σκοπό συνήθως την τεχνητή εισροή λανθασμένης πληροφορίας ή υπερφόρτωσής τους [53].
2. Διείσδυση στα εκάστοτε δίκτυα Wi-F, Bluetooth, 4G, με στόχο την εγκατάσταση κακόβουλων λογισμικών απόσπασης πληροφοριών, διατάραξης της οδηγικής πορείας ή ακόμα μεταφοράς του ελέγχου του οχήματος στον κακόβουλο χρήστη. Με τους διάφορους κατασκευαστές να βρίσκονται δίχως μια πλήρη εικόνα σχετικά με τη φύση, τη σοβαρότητα καθώς και την αντιμετώπιση αυτών των απειλών[54].
3. Παραβίαση του εσωτερικού συστήματος ενός αυτόματου οχήματος, δίνοντας σε έναν κακόβουλο χρήστη τη δυνατότητα παροχής εσφαλμένων συνόλων δεδομένων ώστε αυτά να οδηγήσουν σε λανθασμένη εκπαίδευση σχετικά με τη διαφοροποίηση των εκάστοτε εμποδίων [54].

Στο συνέδριο “Black Hat” του έτους 2013 το οποίο αποτελεί συνέδριο σχετικά με την ασφάλεια ηλεκτρονικών συστημάτων, αποδείχθη ότι ορισμένα ηλεκτρικά οχήματα τα οποία διαθέτουν την υποδομή για διασύνδεση στο IoT έχουν τρωτότητες οι οποίες δημιουργούν δυνατότητα παραβίασης μέσω λογισμικών από κακόβουλους χρήστες. Συγκεκριμένα τα λογισμικά ήταν ικανά να παρακάμψουν το σύστημα ελέγχου μεταφέροντας τον τελευταίο στον κακόβουλο χρήστη που είχε πραγματοποιήσει την παραβίαση δίνοντάς του τη δυνατότητα αποτροπής της εκκίνησης ή και του πλήρους ελέγχου των οχημάτων [51].

Συνοπτικά οι κακόβουλες επιθέσεις στα δίκτυα IoT των αυτομάτων και διασυνδεδεμένων οχημάτων μπορεί να έχει σοβαρές επιπτώσεις. Όπως προαναφέρθηκε παραπάνω, κινδυνεύουν οι ίδιοι επιβάτες από πιθανές σωματικές βλάβες, κλοπής των ευαίσθητων προσωπικών τους δεδομένων και παραβίασης της ιδιωτικότητάς τους, μέχρι και προσωπικής τους δυσφήμισης.

Όμως οι κίνδυνοι δεν περιορίζονται μόνο σε ατομικό επίπεδο καθώς σε ακραίες περιπτώσεις μπορούν να πλήξουν την εθνική ασφάλεια και την ασφάλεια των πολιτών. Σίγουρα η παράλυση ή ο έλεγχος ενός ή μερικών οχημάτων από έναν κακόβουλο χρήστη μπορεί να παράξει από μία βραχυπρόθεσμη ενόχληση μέχρι και σημαντικό μέγεθος απωλειών πόρων και χρημάτων. Ωστόσο οι εν λόγω απώλειες είναι δυνατό να κλιμακωθούν και να προσαυξηθούν σημαντικά όταν επεκταθούν στο πλαίσιο ενός οικισμού μιας πόλης ή ενός κράτους. Μία κακόβουλη οργάνωση, ένα εχθρικό κράτος, μία ισχυρή πολυεθνική ή ακόμα και η ίδια η κυβέρνηση ενός κράτους θα μπορούσαν να οργανώσουν μία ευρείας κλίμακας παραβίαση του δικτύου των οχημάτων προκαλώντας γενικευμένη

κυκλοφοριακή παράλυση, οδηγώντας σε απώλεια εσόδων από την αδυναμία πρόσβασης σε εργασία, αδυναμία παροχής αναγκαίων υπηρεσιών όπως προσκόμιση σε νοσοκομεία, διασώσεις, παροχές άμεσης ανάγκης, ή αναλόγως τον χρήστη και τους σκοπούς του μία ευρεία μεταβολή του αλγορίθμου υπεύθυνου για την πλοήγηση των οχημάτων που θα μπορούσε να έχει ως αποτέλεσμα μαζικές λανθασμένες διαδρομές ή ακόμα και μαζικά αυτοκινητιστικά δυστυχήματα. Επίσης οι εχθρικές κυβερνήσεις ξένων χωρών θα μπορούσαν να παρακολουθήσουν προσωπικές συνομιλίες πολιτικών της χώρας, ή να κλέψουν ευαίσθητα εθνικά δεδομένα που αφορούν τη διακυβέρνηση της χώρας και την εθνική άμυνα.

4.2.2 ΤΕΧΝΙΚΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΙΣ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΙΣ ΚΑΙ ΤΗΝ ΚΥΒΕΡΝΟΑΜΥΝΑ

Παρά τις εκτεταμένες ανησυχίες που διατυπώθηκαν και παραπάνω, σχετικά με τους εξελιγμένους κινδύνους στον κυβερνοχώρο, που απορρέουν από την άνηθηση της τεχνητής νοημοσύνης, υπάρχουν στρατηγικές αντιμετώπισης που μπορούν να θωρακίσουν το ψηφιακό οικοσύστημα. Στον ταχέως εξελισσόμενο κόσμο της κυβερνοασφάλειας, η λήψη των κατάλληλων μέτρων προστασίας είναι απαραίτητη για την αντιμετώπιση των συνεχώς αυξανόμενων κυβερνοεπιθέσεων. Είναι αξιοσημείωτο ότι η ίδια η τεχνολογία της τεχνητής νοημοσύνης, η οποία δύναται να αποτελέσει πηγή κινδύνων, μπορεί ταυτόχρονα να λειτουργήσει ως ένας από τους πιο ισχυρούς συμμάχους στην ενίσχυση της κυβερνοασφάλειας. Θα μπορούσε να υποστηριχθεί μάλιστα ότι η πλέον αποτελεσματική λύση για τις προκλήσεις που δημιουργεί η τεχνητή νοημοσύνη στον κυβερνοχώρο είναι η ίδια η εφαρμογή αυτής της τεχνολογίας, όταν αξιοποιείται με στρατηγικό τρόπο.

Η τεχνητή νοημοσύνη (TN) έχει πλέον εισχωρήσει για τα καλά στον τομέα της κυβερνοασφάλειας ενισχύοντας ή αντικαθιστώντας τα υπάρχοντα συστήματα προστασίας δικτύων. Σύμφωνα με τελευταίες εκτιμήσεις προβλέπεται ότι η παγκόσμια αγορά της τεχνητής νοημοσύνης στην ασφάλεια των πληροφοριακών και επικοινωνιακών συστημάτων θα φτάσει τα 46,3 δισεκατομμύρια δολάρια μέχρι το 2027. Αυτό αδιαμφισβήτητα φανερώνει τη σταθερή ενσωμάτωση της στην ασφάλεια των συστημάτων [55].

Η αλλαγή αυτή αποτελεί σημαντική αναβάθμιση καθώς τα συστήματα προστασίας δύνανται να αποκτήσουν επιπλέον επιπρόσθετες δυνατότητες για την αποδοτικότερη λειτουργία τους καθώς και την εξοικονόμηση πόρων. Ένα πρώτο πλεονέκτημα είναι η καταπολέμηση όλων των αποτυχιών που ευθύνονται σε ανθρώπινο σφάλμα, είτε λόγω γνωστικής αποτυχίας είτε λόγω χονδροειδών σφαλμάτων [56], γεγονός που οδηγεί σε απώλεια πολύτιμου χρόνου, πόρων δεδομένων αλλά και κλονισμό της αξιοπιστίας ενός δικτύου.

Μέσω της χρήσης μοντέλων μη επιβλεπόμενης μηχανικής μάθησης για την αναγνώριση αποχών, καθώς και ήδη γνωστών αλλά και νέων απειλών, όσον

αφορά λειτουργία λογισμικού, εξυπηρετητών (servers) [57] και λειτουργικών συστημάτων δημιουργείται η δυνατότητα επεξεργασίας και ελέγχου λογισμικών με μεγάλη ταχύτητα, παρέχοντας τη δυνατότητα λήψης προληπτικών μέτρων, έναντι απλών ενεργειών αντιμετώπισης που συνηθίζεται μέχρι στιγμής. Ένα ακόμα πλεονέκτημα αποτελεί η δυνατότητα εφαρμογής και χρήσης μοντέλων ΑΙ για την παρακολούθηση λειτουργίας και της γενικής κατάστασης του υλικού τμήματος των λειτουργικών συστημάτων [58]. Η εν λόγω χρήση επιτυγχάνεται μέσω συνεχούς παρακολούθησης της λειτουργίας, θερμοκρασίας, των συστημάτων ψύξης, τάση του ρεύματος γεγονός που θα καθιστά τα εν λόγω συστήματα να εντοπίζουν ανωμαλίες στη λειτουργία των εκάστοτε εξαρτημάτων και να προτρέπουν σε στενευμένα προληπτικά μέτρα για την αποκατάσταση της ομαλής λειτουργίας. Μειώνοντας το κόστος συνολικής λειτουργίας και μέσω της μειωμένης απαίτησης για κατανάλωση ενέργειας [59].

Η γενική κατεύθυνση της ΤΝ είναι η ταξινόμηση λογισμικών, η λογαριασμών σε ασφαλείς και μη ασφαλείς. Η παραπάνω διαδικασία επιτυγχάνεται με τη δυνατότητα των συστημάτων που χρησιμοποιούν ΤΝ να επεξεργάζονται ταχέως αφενός μεγάλο όγκο δεδομένων [60] ενώ παράλληλα να έχουν την ικανότητα απόκτησης χαρακτηριστικών όσον αφορά τα δεδομένα αυτά. Υπάρχει μεγάλη ποικιλομορφία αλγορίθμων που χρησιμοποιούνται όπως K-means, Support Vector Machines, Decision Trees, Random Forests, DB – SCAN, logistic regression κ.α καθώς και διαφορετικών αρχιτεκτονικών νευρωνικά δίκτυα. Μια από τις πολλές κατηγοριοποιήσεις των παραπάνω αλγορίθμων είναι αυτή σε επιβλεπόμενη και μη επιβλεπόμενη ταξινόμηση με ενδεικτικά παραδείγματα αυτά των Decision Trees και K-means αντίστοιχα. Ο σκοπός είναι η χαρακτηριστικά ταξινόμηση των προαναφερθέντων λογαριασμών ή λογισμικών, βασιζόμενη σε χαρακτηριστικά τα οποία αποκτώνται και σχετίζονται με τη δομή του κώδικα, τη λειτουργία του, την προέλευση, το ιστορικό δραστηριότητας στο διαδίκτυο κ.α. Ομοιότητα με λογισμικά ή συμπεριφορές που έχουν χαρακτηριστεί ως κακόβουλα καθώς και εντοπισμός αποχών – ανωμαλιών από την καθορισμένη λειτουργία και δραστηριότητα των υπολοίπων λογαριασμών ή λογισμικών ενός οργανισμού, οδηγούν στον εντοπισμό και επισήμανση κακόβουλων οντοτήτων στον κυβερνοχώρο. Ορισμένες μέθοδοι ανίχνευσης και αποτροπής κυβερνοεπιθέσεων μέσω ΤΝ είναι οι ακόλουθες:

- **Μέθοδοι ανίχνευσης ψηφιακής υπογραφής :** Κατά αυτήν τη μέθοδο ο αλγόριθμος συγκρίνει τη δομή του κώδικα ενός λογισμικού με ήδη υπάρχοντα λογισμικά σε βάσεις δεδομένων τα οποία έχουν χαρακτηριστεί ως κακόβουλα [61]. Υψηλή ομοιότητα με κάποια από τα ήδη καταχωρημένα κακόβουλα λογισμικά θα οδηγήσει σε ταξινόμηση του εξεταζόμενου επίσης ως κακόβουλο. Η συγκεκριμένη μέθοδος χαρακτηρίζεται από υψηλή ταχύτητα η οποία είναι απόρροια της αυτοματοποιημένης διαδικασίας λόγω της χρήσης ΤΝ.
- **Προστασία δικτύων :** Ένα μεγάλο μέρος των κυβερνοεπιθέσεων σε επιχειρήσεις και οργανισμούς πραγματοποιούνται χρησιμοποιώντας ως

δίοδο τα δίκτυα που οι πρώτοι χρησιμοποιούν [61]. Η ενίσχυση των τοίχων προστασίας δικτύων με TN δίνει τη δυνατότητα απόκτησης και ταχείας επεξεργασίας μεγάλου όγκου δεδομένων [62] ώστε να εντοπιστούν και να επισημανθούν τυχόν ανωμαλίες σε κόμβους του δικτύου οι οποίες θα μπορούσαν να σημαίνουν παραβίαση. Παράλληλα η ίδια η πρόσβαση στο δίκτυο έχει γίνει πιο δύσκολη χωρίς την άρτια εξουσιοδότηση [63] η οποία μπορεί να περιλαμβάνει διαδοχικές βαθμίδες εξακρίβωσης δεδομένων.

- **Αξιολόγηση και διαχείριση τρωτότητας συστημάτων :** Μοντέλα TN χρησιμοποιούνται για την εύρεση τρωτών σημείων στα λειτουργικά συστήματα των οργανισμών, με τη βιβλιογραφία να επισημαίνει πως παρατηρήθηκε αύξηση τάξης 18% στις συνολικές πηγές τρωτότητας μεταξύ των ετών 2018-2019[64, 65].
- **Αντι σπαμ προστασία :** Κατά την εν λόγω διαδικασία τα μοντέλα εκπαιδεύονται με ένα σετ δεδομένων το οποίο περιλαμβάνει έναν αριθμό από e- mail τα οποία έχουν ήδη κατηγοριοποιηθεί σε σπάμ ή κανονικά. Σκοπός είναι η δημιουργία και εκπαίδευση ενός μοντέλου το οποίο αντλώντας πληροφορίες από χαρακτηριστικά όπως η δομή, το κείμενο, η σύνταξη, ενδεχόμενα επιπρόσθετα στοιχεία όπως url κ.α να αποφανθεί σχετικά με την εγγυρότητα του e-mail, και να το κατατάξει ως κανονικό ή σπάμ [60].
- **Προστασία έναντι PHISING. :** Με τον όρο “Phising” εννοείται η παραπλανητική μη εθελούσια από τον λαμβάνοντα χρήστη, απόσπαση προσωπικών απόρρητων ή μη πληροφοριών, συνήθως ηλεκτρονικής ή οικονομικής φύσεως. Σύμφωνα με τους ερευνητές υπάρχουν τρία επίπεδα προστασίας έναντι του φαινομένου [66]. Αρχικά αυτό της μεθόδου ανίχνευσης κατά την οποία γίνεται η κατάταξη ενός e-mail σε σπάμ ή όχι. Μια άλλη προσέγγιση είναι η αποτρεπτική μέσω, της συνεχούς αναβάθμισης της ασφάλειας των δικτύων, και της αύξησης των αναγκαίων μέτρων διακρίβωσης στοιχείων όσον αφορά χρήστες ή το δίκτυο [60]. Τέλος η διορθωτική δράση είναι αυτή κατά την οποία εντοπίζονται οι πηγές του phising στον κυβερνοχώρο και καταργούνται από αυτόν, (κλείσιμο ιστοσελίδων, διαγραφή λογαριασμών κ.α).
- **Ανάλυση συμπεριφοράς χρηστών :** Η δραστηριότητα χρηστών σε ένα δίκτυο μπορούν να μοντελοποιηθούν ως ακολουθίες μεμονωμένων ενεργειών. Για την εν λόγω χρήση ενδείκνυνται ορισμένες αρχιτεκτονικές νευρωνικών δικτύων όπως παλινδρομικά (RNN), βραχέας-μακράς μνήμης (LSTM), φραγμένης παλινδρομικής μονάδας (GRU). Παρατηρούμενες ανωμαλίες στις εκάστοτε ακολουθίες θα μπορούσαν να αποτελέσουν αντικείμενο διερεύνησης για την κατάταξη ή όχι ως ύποπτη δραστηριότητα. Παράλληλα με τη χρήση γεννητικών δικτύων (GAN) επιτυγχάνεται η δημιουργία λογαριασμών που παράγουν την εν λόγω δραστηριότητα για τη συνεχή εκπαίδευση και βελτίωση των εκάστοτε μοντέλων [60].

Ορισμένες από τις αρνητικές πτυχές της ενσωμάτωσης της TN στην κυβερνοασφάλεια είναι η υψηλή απαίτηση όσον αφορά την υπολογιστική ισχύ των εξαρτημάτων και τη συνεπαγόμενη αναγκαιότητα όσον αφορά τη μνήμη [67] και της αυξημένες δυνατότητες της κεντρικής μονάδας επεξεργασίας(CPU) [68]. Οι γλώσσες προγραμματισμού, τα λειτουργικά συστήματα και οι χρησιμοποιούμενες βάσεις δεδομένων αποτελούν ένα ακόμα εμπόδιο στην ευρείας φύσεως υιοθέτηση του συστήματος καθώς, τα περισσότερα εκ των προαναφερθέντων είναι παλαιωμένα σε σχέση με τις απαιτήσεις της TN [69, 70], ενώ επίσης σημαντικό είναι το πρόβλημα επαρκούς όγκου δεδομένων προς επεξεργασία [71], και η ανάγκη για ένα ευρύ φάσμα εξειδικευμένου προσωπικού (ειδικοί σε TN, Μηχανική Μάθηση, στατιστικολόγοι, μαθηματικοί κ.α.) [72], καθώς και εξειδικευμένων μοντέλων και εργαλείων για τις ανάγκες του εκάστοτε οργανισμού.

Παρά τις προαναφερθείσες προκλήσεις όμως, όσον αφορά την υιοθέτηση της εν λόγω τεχνολογίας έχει αποδειχθεί από τη βιβλιογραφία, η αυξημένη αποτελεσματικότητα της προληπτικής έναντι τις παραδοσιακής- αντιδραστικής αντιμετώπισης των θεμάτων κυβερνοασφαλείας.

Η TN σαφώς, χρησιμοποιείται, εξίσου και από τους εκάστοτε κακόβουλους πράκτορες, καθώς επίσης εργαλιοποιούνται ορισμένα χαρακτηριστικά ώστε αυτά να αποτελέσουν αδυναμίες. Μοντέλα TN είναι σε θέση να βελτιστοποιούν παραμέτρους και μεταβλητές κακόβουλων λογισμικών ώστε να μειώνουν την ανιχνευσιμότητά της δραστηριότητάς τους. Παράλληλα υπάρχει η δυνατότητα μαζικής παραγωγής σετ δεδομένων τα οποία είναι δομημένα ώστε να οδηγούν σε εσφαλμένη εκπαίδευση των μοντέλων, η οποία θα οδηγεί σε λανθασμένες προβλέψεις [73], κατά έναν τρόπο ωφέλιμο για αυτόν που πραγματοποιεί την κυβερνοεπίθεση. Η συνεχής παρακολούθηση της εκπαίδευσης των μοντέλων και η τακτική ανασκόπηση και αναβάθμιση της δομής τους είναι κάποια από τα μέτρα τα οποία μπορούν να μετριάσουν τις επιπτώσεις των κυβερνοεπιθέσεων.

Το επόμενο βήμα στον τομέα της κυβερνοασφαλείας όσον αφορά την TN θα ήταν η συστηματική μείωση της ανθρώπινης παρέμβασης- καθοδήγησης των μοντέλων που χρησιμοποιούν αλγορίθμους TN. Η χρήση αλγορίθμων όπως τα συστήματα ειδικών ("Expert systems") και "Intelligent Actors" [73]. Στόχος θα ήταν η αναβάθμιση των μοντέλων και της δομής των διαφόρων λογισμικών καθώς και η αυτόματη επισήμανση και αντιμετώπιση απειλών χωρίς η απόφαση, αλλά και ο ίδιος ο σαφής ορισμός των απειλών να είναι συνεχώς αναγκαίος από τους ανθρώπους [43]. Στη μελέτη του άρθρου "Artificial intelligence is a threat to cybersecurity. It's also a solution. [43]" παραδείγματος χάρη, γίνεται λόγος για τη δυνατότητα ενός συστήματος ασφαλείας να δημιουργήσει έναν κλώνο του λειτουργικού δικτύου και των ηλεκτρονικών διαδρομών που το απαρτίζουν προκειμένου οι ενδεχόμενοι κακόβουλοι χρήστες να παγιδευτούν, μέσω των ενεργειών τους, και να ιχνηλατηθεί η ηλεκτρονική τους δραστηριότητα. Πολύ σημαντικός θα ήταν επίσης και ο ανεμπόδιστος διαμοιρασμός δεδομένων όσον αφορά την εξέλιξη των διαφόρων απειλών.

ΚΕΦΑΛΑΙΟ 5- ΨΗΦΙΑΚΗ ΙΔΙΩΤΙΚΟΤΗΤΑ ΤΗ ΣΥΧΡΟΝΗ ΕΠΟΧΗ ΤΩΝ ΔΙΑΣΥΔΕΔΕΜΕΝΩΝ ΔΕΔΟΜΕΝΩΝ

5.1 ΙΔΙΩΤΙΚΟΤΗΤΑ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

5.1.1 ΟΡΙΣΜΟΣ

Η ιδιωτικότητα των χρηστών στον διαδίκτυο παρέχει στους χρήστες το δικαίωμα να καθορίζουν το πως και μέχρι ποιο σημείο τα δεδομένα που αφορούν αυτούς συλλέγονται, αποθηκεύονται και κοινοποιούνται σε τρίτους. Το διαδίκτυο αποτελεί έναν παγκόσμιο χώρο διασύνδεσης και ανταλλαγής πληροφοριών, καθιστώντας τη ψηφιακή ιδιωτικότητα αναπόσπαστο τμήμα της καθημερινής μας ζωής. Στις μέρες μας εγείρονται περισσότερες ανησυχίες για την επιτυχή διασφάλιση της προστασίας της ιδιωτικότητας στο διαδίκτυο, λόγω της άνθισης των τεχνολογιών τεχνητής νοημοσύνης και της ψηφιακής διασυνδεσιμότητας.

Πολλοί εμπειρογνώμονες υποστηρίζουν ότι είναι ανέφικτο να υπάρχει πλήρης ιδιωτικότητα. Ο Bruce Schneier υποστήριξε ότι, "Η ιδιωτικότητα μας προστατεύει από καταχρήσεις από όσους βρίσκονται στην εξουσία, ακόμα κι αν δεν κάνουμε τίποτα λάθος κατά την εποχή της παρακολούθησης" [4]. Με αυτή τη δήλωση θέλει να τονίσει ότι παρόλο που η εθνική ασφάλεια της Αμερικής υποστηρίζει ότι η συλλογή ευαίσθητων δεδομένων ή ακόμα και η ύπαρξη καμερών ασφαλείας σε αρκετούς δημοσίους χώρους δεν επηρεάζει πρακτικά τους πολίτες όσο αυτοί δεν διαπράττουν παράνομες συμπεριφορές, υπάρχει κίνδυνος διακύβευσής της ελευθερίας. Για αυτό τον λόγο συμβουλεύει τους πολίτες να μην καθησυχάζονται αλλά να διεκδικούν το δικαίωμα προστασίας της ιδιωτικότητας τους. Επιπλέον υποστηρίζει ότι πολλοί έρχονται σε αντιπαράθεση με τον αν είναι προτιμότερη η ασφάλεια από την ιδιωτικότητα, ενώ το πραγματικό δίλημμά είναι ελευθερία έναντι ελέγχου και τυραννίας.

Δημιουργείται έτσι στους χρήστες η πλασματική αίσθησή της διαφάνειας κάτι που είναι ψευδές. Είναι σημαντικό να σημειώσουμε πως οι περισσότερες ηλεκτρονικές πλατφόρμες, όπως οι μηχανές αναζήτησης και τα κοινωνικά δίκτυα, διατηρούν δεδομένα των χρηστών για αόριστο χρονικό διάστημα. Τα δεδομένα αυτά αφορούν δεδομένα αναζητήσεων και δραστηριότητας του χρήστη σε μηχανές αναζήτησης όπως η Google και το Bing, αλλά και κοινοποιήσεις, συνομιλίες, φωτογραφικό υλικό και βίντεο και άλλες προσωπικές πληροφορίες σε μέσα κοινωνικής δικτύωσης όπως το twitter, το Facebook, το Instagram e.t.c. Ακόμα και εάν ο χρήστης εκκαθαρίσει τα δεδομένα αναζήτησής του, διαγράψει παλιές συνομιλίες και δημοσιεύσεις, αλλάξει τα προσωπικά του στοιχεία ή επιλέξει ανώνυμη πλοήγηση, αυτά τα δεδομένα ενδέχεται να παραμένουν υπαρκτά στο διαδίκτυο για πάντα και μπορούν να ανασυρθούν ανά πάσα στιγμή καθώς υπάρχουν αρκετοί εμπειρογνώμονες που έχουν τη δυνατότητα να αποκτήσουν πρόσβαση σε ευαίσθητα δεδομένα ή κρυπτογραφημένα δεδομένα. Αυτοί είναι εμπειρογνώμονες πληροφορικής και κυβερνοασφάλειας στη δίωξη

ηλεκτρονικού εγκλήματος, οι οποίοι έχουν και εξελιγμένες γνώσεις και παραπάνω δικαιώματα να ανασύρουν δεδομένα για την αντιμετώπιση υποθέσεων ηλεκτρονικού εγκλήματος ή ακόμα και κακόβουλοι χάκερ. Ωστόσο δεν χρειάζεται να είναι κανείς αρκετά εμπειρογνώμων για να ανακτήσει δεδομένα όπως διαγραμμένο ιστορικό αναζήτησης στο Google ή διαγραμμένες δημοσιεύσεις σε πλατφόρμες όπως το Twitter. Πρέπει να σημειωθεί ότι μερικές φορές οι ίδιοι οι χρήστες συντελούν αρκετά στο θάνατο της ιδιωτικότητας, καθώς δημοσιεύουν από μόνοι τους πληροφορίες και δεδομένα ευαίσθητα στα μέσα κοινωνικής δικτύωσης διατηρώντας μια ανεύθυνη στάση και εκθέτοντας τον εαυτό τους. Τέλος πολύ συχνά παραβιάζεται η ιδιωτικότητα των χρηστών στο διαδίκτυο από ιστότοπους που συλλέγουν παράνομα δεδομένα χωρίς τη συγκατάθεσή των χρηστών με διάφορες τεχνικές, όπως η μέθοδος fingerprint canvas ή η τεχνική web scalping. Εξάλλου δεν είναι λίγα τα σκάνδαλα που έχουν γίνει.

Πιο συγκεκριμένα μπορεί να ανατρέξει κάποιος στο 2018 στο σκάνδαλο Facebook-Cambridge Analytica, όπου αποκαλύφθηκε ότι η δημοφιλής εφαρμογή Facebook διέρρευσε προσωπικά δεδομένα 87 εκατομμυρίων προφίλ χρηστών του Facebook χωρίς τη συγκατάθεσή τους στη βρετανική εταιρεία συμβούλων Cambridge Analytica, τα οποία προορίζονταν για πολιτική διαφήμιση.

Το 2019 η κινέζικη εταιρεία Huawei, που είναι η μεγαλύτερη τηλεπικοινωνιακή εταιρεία στον κόσμο κατηγορήθηκε για παράνομη συνεργασία με αφρικανικές κυβερνήσεις [38]. Τεχνικοί της εταιρείας βοήθησαν τις αφρικανικές κυβερνήσεις να κατασκοπεύσουν τους πολιτικούς τους αντιπάλους, παρακολουθώντας κρυπτογραφημένες επικοινωνίες, δεδομένα μέσων κοινωνικής δικτύωσης και δεδομένα κινητής τηλεφωνίας.

Το 2013 ο πρώην αναλυτής στην Εθνική Υπηρεσία Ασφαλείας (NSA) των Ηνωμένων Πολιτειών, ο Edward Snowden αποκάλυψε κάτι που συγκλόνισε τον όλο τον κόσμο και εγείρε ανησυχίες για την προστασία της ιδιωτικότητας στο διαδίκτυο, με κράτη και οργανισμούς να αμφισβητούν τις ενέργειες των ΗΠΑ [76]. Πιο συγκεκριμένα αποκαλύφθηκε ότι η NSA παρακολουθούσε και σύλλεγε δεδομένα που αφορούσαν όχι μόνο τους πολίτες των ΗΠΑ αλλά και πολίτες σε όλο τον κόσμο. Το πρόγραμμα PRISM έδινε στη NSA πρόσβαση σε δεδομένα από μεγάλες τεχνολογικές εταιρείες όπως Google, Microsoft, Apple, Facebook, κ.ά., χωρίς τη συγκατάθεση των χρηστών. Μετά από αυτό το σκάνδαλο ορισμένες εταιρείες προχώρησαν σε αλλαγές στις πολιτικές τους για να αντιμετωπίσουν τις ανησυχίες των χρηστών και των εθνικών αρχών.

Αυτά είναι απλά μερικά από τα σκάνδαλα που έχουν συμβεί που δείχνουν ότι η ιδιωτικότητα στο διαδίκτυο είναι νεκρή και ότι η μη εξουσιοδοτημένη παρακολούθηση των χρηστών φαντάζει απίθανο σενάριο. Ακόμη χειρότερο είναι ότι η νόμοι και τα μέτρα που λαμβάνονται δεν τηρούνται επαρκώς μέχρι και από αυτούς που τα θέσπισαν.

5.1.2 ΣΗΜΑΝΤΙΚΟΤΗΤΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΤΟ ΔΙΑΔΙΚΤΥΟ

Η διαφύλαξη της ιδιωτικότητας στο διαδίκτυο αποτελεί σημαντικό ζήτημα στη σύγχρονη ψηφιακή εποχή για πολλούς λόγους. Είναι σημαντικό οι χρήστες να νιώθουν ότι το διαδίκτυο είναι ένα ασφαλές περιβάλλον, που σέβεται τη ανθρώπινη υπόσταση και αξιοπρέπεια, προστατεύει τα δεδομένα τους και ότι η διαδικτυακή τους παρουσία διαμορφώνεται με σεβασμό προς τα δικαιώματα και την ιδιωτικότητά τους, χωρίς να υπονομεύτε η ατομική τους αυτονομία.

Καθώς η τεχνολογία εξελίσσεται και ο ψηφιακός κόσμος διαδραματίζει σημαντικό ρόλο στην έκφραση της κοινής γνώμης, η προστασία των προσωπικών δεδομένων είναι σημαντική για την ελευθέρια του λόγου. Οι χρήστες πρέπει να έχουν την ελευθερία να εκφράζουν τις απόψεις τους χωρίς τον φόβο μη εξουσιοδοτημένης παρακολούθησης ή πιθανής διαρροής της ταυτότητάς τους όταν αυτοί επιλέγουν να εκφράσουν την άποψή τους ανώνυμα σε ηλεκτρονικές πλατφόρμες .

Επίσης προφυλάσσει τους χρήστες από πιθανές παραβιάσεις εξουσίας από την κυβέρνησή. Η έλλειψη ιδιωτικότητας θα μπορούσε να οδηγήσει σε υπερβολική εξουσία και έλεγχο από τις αρχές, θέτοντας σε κίνδυνο την ελευθερία των χρηστών αλλά και να απειλήσει τη δημοκρατία. Ακόμη υπάρχει ο κίνδυνος αλλοίωσης των δεδομένων από τις αρχές, καθώς μπορούν να παραβιάσουν την ιδιωτικότητα για να διαστρεβλώσουν πληροφορίες θέλοντας να παραπληροφορήσουν και να χειραγωγήσουν τους πολίτες. Ένα παράδειγμα που αναδεικνύει τη σημασία της προστασίας της ιδιωτικότητας είναι η Βόρεια Κορέα. Η κυβέρνηση της βόρειας Κορέας ασκεί εκτεταμένο έλεγχο στους πολίτες της, περιορίζοντας σημαντικά τις ελευθερίες τους και παραβιάζοντας την προσωπική τους ιδιωτικότητα. Οι πολίτες εκεί βρίσκονται υπό στενή παρακολούθηση και οποιαδήποτε προσπάθεια για ελεύθερη έκφραση ή πρόσβαση σε ανεξάρτητες πληροφορίες μπορεί να έχει σοβαρές συνέπειες.

Είναι γνωστό ότι πολλές εταιρείες όπως οι πλατφόρμες κοινωνικών μέσων, ιστότοποι που πραγματοποιούν τις αγορές τους οι χρήστες, αλλά και άλλες ηλεκτρονικές πλατφόρμες, αποθηκεύουν προσωπικά δεδομένα των χρηστών τους, για λόγους κυρίως εμπορικούς αλλά και λόγους ασφαλείας. Εάν οι εταιρίες δεν καταφέρουν να διαφυλάξουν τα δεδομένα αυτά από τρίτους, μπορεί να προκύψουν σημαντικοί κίνδυνοι. Αρχικά απειλείται η ασφάλεια των χρηστών καθώς η ιδιωτικότητα στο διαδίκτυο λειτουργεί ως ασπίδα προστασίας για πιθανούς κινδύνους, όπως η κλοπή ταυτότητας και ηλεκτρονικές απάτες. Επίσης προστατεύει τους χρήστες από δυσφημίσεις από τρίτους στο διαδίκτυο που μπορούν να χρησιμοποιήσουν ευαίσθητα δεδομένα ως εργαλείο ελέγχου για να προκαλέσουν μεγάλο κακό στο πρόσωπό κάποιου η ακόμα να τον βλάψουν σωματικά. Επιπλέον πέρα από την ασφάλεια των ανθρώπων επηρεάζονται άμεσα οι επιχειρήσεις και οι εταιρίες καθώς η δημοσίευση κωδικών πρόσβασης και άλλων ευαίσθητων δεδομένων θα έθετε σε κίνδυνο την ασφάλεια τους από κυβερνοεπιθέσεις, κάτι που θα οδηγούσε σε απώλεια εμπιστοσύνης πελατών, δυσφήμιση και οικονομική καταστροφή. Ακόμη η πιθανή έκθεση κυβερνητικών

δεδομένων θα μπορούσε να αποτελέσει σοβαρή απειλή για την εθνική ασφάλεια. Δυστυχώς ωστόσο έχει αποδειχθεί ότι οι διαρροές δεδομένων παραμένουν ένα σημαντικό πρόβλημα ακόμα και όταν οι εταιρίες φροντίζουν πολύ για την προστασία τους. Μάλιστα βρέθηκαν εταιρίες η οποίες εσκεμμένα διέρρεαν δεδομένα των χρηστών τους σε τρίτους.

Επομένως η αποτυχία διασφάλισης της ιδιωτικότητας στο διαδίκτυο έχει σοβαρές επιπτώσεις, στους ανθρώπους, οργανισμούς και τη συνολική σταθερότητα του ψηφιακού οικοσυστήματος ακόμα και σε κυβερνήσεις.

5.1.3 ΝΟΜΟΘΕΣΙΑ ΚΑΙ ΠΡΟΣΤΑΣΙΑ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΣΤΟ ΔΙΑΔΥΚΤΙΟ

Υπάρχουν πολλοί κανονισμοί και νομοθεσίες για την προστασία της ιδιωτικότητας στο διαδίκτυο οι οποίοι ρυθμίζουν τη διαχείριση και επεξεργασία των δεδομένων. Ωστόσο κάθε περιοχή έχει τους δικούς της κανονισμούς και κανονιστικά πλαίσια που διέπουν την προστασία των δεδομένων και την ιδιωτικότητα, καθώς έχουν διαφορετικές πολιτικές και νομικές απαιτήσεις. Οι νομοθέτες που θεσπίζουν αυτούς τους νόμους, πρέπει να επανεξετάζουν συνεχώς την ισχύ τους, με σκοπό τη θέσπιση νέων νόμων που ενισχύουν την προστασία από τους κινδύνους που προκύπτουν από τις τεχνολογικές αναβαθμίσεις ή την πιθανή αναπροσαρμογή τους με βάση τις νέες εξελίξεις της τεχνητής νοημοσύνης.

Στην Ελλάδα και σε όλη την Ευρώπη, ισχύει ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) [77], ο οποίος αποτελεί τον πλέον εξειδικευμένο κανονισμό προστασίας της ιδιωτικής ζωής παγκοσμίως. Αυτός καθορίζει τους κανόνες για την προστασία δεδομένων και το απόρρητο για όλα τα άτομα εντός της Ευρωπαϊκής Ένωσης και του Ευρωπαϊκού Οικονομικού Χώρου, παρέχοντας εκτενή δικαιώματα στους κατόχους των δεδομένων. Ακόμη απαιτεί από τους υπεύθυνους επεξεργασίας δεδομένων, να τηρούν τις αρχές προστασίας δεδομένων και να συμμορφώνονται σύμφωνα με αυστηρά πρότυπα. Ο GDPR έχει εφαρμογή και σε εταιρίες που δραστηριοποιούνται εκτός των ορίων της ΕΕ, όμως παρέχουν αγαθά και υπηρεσίες σε χρήστες εντός αυτής [78]. Μετά την εφαρμογή του GDPR τα πρόστιμα για παραβίαση προσωπικών δεδομένων μπορούν να ανέρθουν στο ποσό των 20 εκ ευρώ, [78] με τον αντίστοιχο νόμο να έχει λάβει παρόμοια διάσταση με αυτήν του αντιμονοπωλιακού νόμου “anti trust law” οποίος βρίσκεται σε ισχύ εδώ και μία δεκαετία.

Ο κανονισμός GDPR έχει επιφέρει σημαντικές αλλαγές στον τρόπο που διαχειριζόμαστε τα προσωπικά δεδομένα και τα cookies στον ψηφιακό κόσμο. Ένας από τους κύριους στόχους του GDPR είναι η προστασία της ιδιωτικής ζωής των πολιτών της Ευρωπαϊκής Ένωσης, καθώς και η διασφάλιση της ασφαλούς και διαφανούς χρήσης των προσωπικών τους δεδομένων. Όσον αφορά τη συλλογή δεδομένων, ο GDPR ορίζει αυστηρούς κανόνες για την επεξεργασία προσωπικών δεδομένων. Οι εταιρείες πρέπει να έχουν σαφείς και διαφανείς πολιτικές για το πώς συλλέγουν, αποθηκεύουν και χρησιμοποιούν τα προσωπικά δεδομένα των χρηστών. Οι χρήστες πρέπει να δίνουν ενεργή συγκατάθεση για τη

συλλογή και επεξεργασία των προσωπικών τους δεδομένων και να ενημερώνονται για τους σκοπούς χρήσης των δεδομένων τους.

Όσον αφορά τα cookies, ο GDPR επιβάλλει την υποχρέωση ενημέρωσης των χρηστών για τη χρήση τους και την αποτελεσματική δυνατότητα ελέγχου της αποδοχής ή απόρριψής τους. Οι ιστότοποι πρέπει να ζητούν τη συγκατάθεση των χρηστών πριν από την τοποθέτηση cookies στη συσκευή τους, εκτός αν τα cookies είναι απαραίτητα για τη λειτουργία του ιστότοπου. Επιπλέον, οι χρήστες έχουν το δικαίωμα να ανακαλέσουν τη συγκατάθεσή τους ανά πάσα στιγμή.

Το θέμα είναι ότι όταν αναφερόμαστε σε μεγάλες εταιρίες όπως το Facebook, που έχουν μονοπώλιο στον κλάδο τους και δισεκατομμύρια καταναλωτές χρησιμοποιούν τις υπηρεσίες τους παγκοσμίως, η κατάσταση περιπλέκεται. Ακόμη και αν οι χρήστες μελετήσουν εκτενώς ολόκληρη την πολιτική απορρήτου και διαφωνήσουν με ορισμένες πτυχές της, αναγκάζονται να συμβιβαστούν με αυτές και να συνεχίσουν να χρησιμοποιούν την υπηρεσία αυτή, λόγω της ευρείας χρήσης της και της έλλειψης συγκρίσιμων εναλλακτικών.

Οι πρόοδοι στον τομέα της τεχνητής νοημοσύνης επισημαίνουν την ανάγκη για ηθικές κατευθυντήριες γραμμές και βέλτιστες πρακτικές προκειμένου να μειωθούν πιθανοί κίνδυνοι που θα έβλαπταν την ιδιωτικότητα. Για τον λόγο αυτό δημιουργήθηκαν αρκετά σωματεία και οργανώσεις για να διασφαλίσουν ότι οι πρόοδοι της τεχνητής νοημοσύνης κινούνται σε ηθικά πλαίσια και με σεβασμό προς την ανθρωπότητα [79].

Η πρωτοβουλία « Partnership on AI » είναι μία τέτοια οργάνωση που αντιμετωπίζουν ηθικές προκλήσεις που προκαλεί η τεχνητή νοημοσύνη. Πρόκειται για μια πρωτοβουλία πολλαπλών εμπλεκόμενων φορέων της τεχνητής νοημοσύνης που συγκεντρώνει ηγέτες της βιομηχανίας, ερευνητές ακαδημαϊκών κύκλων, οργανώσεις και άλλους φορείς συνεργασίας με στόχο την προώθηση των τεχνολογιών AI με τρόπο που θα δίνει προτεραιότητα στις ηθικές σκέψεις, τη διαφάνεια, τη δικαιοσύνη και την ευθύνη. Ένας από τους βασικούς στόχους του Partnership on AI είναι η θέσπιση ηθικών οδηγιών και βέλτιστων πρακτικών για την ανάπτυξη και την εφαρμογή των συστημάτων AI. Αυτές οι οδηγίες στοχεύουν στο να διασφαλίσουν ότι οι τεχνολογίες AI αναπτύσσονται και χρησιμοποιούνται με τρόπο που προάγει την ευημερία του ανθρώπου, προστατεύει τα θεμελιώδη δικαιώματα και αντιμετωπίζει τυχόν κινδύνους και βλάβες [34].

Μία ακόμη οργάνωση που δημιουργήθηκε για την προώθηση αυτού του σκοπού είναι η Παγκόσμια Πρωτοβουλία IEEE για την Ηθική των Αυτόνομων και Ευφυών Συστημάτων με κύριο σκοπό “να διασφαλίσει ότι κάθε ενδιαφερόμενος που εμπλέκεται στο σχεδιασμό και την ανάπτυξη αυτόνομων και ευφυών συστημάτων είναι εκπαιδευμένο και εξουσιοδοτημένος να δίνει προτεραιότητα σε ηθικούς λόγους, ώστε αυτές οι τεχνολογίες να προωθούνται προς όφελος της ανθρωπότητας”.

Το Συμβουλευτικό Σώμα των Ηνωμένων Εθνών για την Τεχνητή Νοημοσύνη (United Nations Advisory Board on Artificial Intelligence) είναι μια ακόμη

οργάνωση με ειδικούς και εμπειρογνώμονες, για να παρέχει συμβουλές και καθοδήγηση σε θέματα που αφορούν την τεχνητή νοημοσύνη και τις επιπτώσεις της στην κοινωνία και τον κόσμο[80]. Το Συμβουλευτικό Σώμα λειτουργεί ως φόρουμ συζήτησης και ανάλυσης για θέματα όπως η ρύθμιση της τεχνητής νοημοσύνης, η διασφάλιση της δικαιοσύνης και των ανθρωπίνων δικαιωμάτων στον τομέα της τεχνητής νοημοσύνης, καθώς και την προώθηση της ηθικής χρήσης και ανάπτυξης της τεχνολογίας. Οι συστάσεις και οι προτάσεις που παράγονται από το Συμβουλευτικό Σώμα μπορούν να χρησιμοποιηθούν ως κατευθυντήριες γραμμές από κυβερνήσεις, διεθνείς οργανισμούς και άλλα ενδιαφερόμενα μέρη για τη διαμόρφωση πολιτικών και πρακτικών που αφορούν την τεχνητή νοημοσύνη.

5.2 ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΚΑΙ ΙΔΙΩΤΙΚΟΤΗΤΑ

Η άνθιση της τεχνητής νοημοσύνης έχει αναδείξει έναν παγκόσμιο μετασχηματισμό, επηρεάζοντας τις καθημερινές μας δραστηριότητες και συμβάλλοντας στην ανάπτυξη προηγμένων και αυτοματοποιημένων υπολογιστικών συστημάτων. Ωστόσο οι αυξημένες δυνατότητες που προσφέρει δεν έρχονται χωρίς σημαντικούς κινδύνους. Η συλλογή και επεξεργασία μεγάλου όγκου προσωπικών δεδομένων, με σκοπό την εκπαίδευση των αλγορίθμων, ανοίγει τον δρόμο για πιθανές διαρροές και καταχρήσεις πληροφοριών. Είναι, λοιπόν, απαραίτητο να διασφαλιστεί όχι μόνο η τεχνολογική πρόοδος, αλλά και η προστασία των ατομικών δικαιωμάτων και της ιδιωτικότητας στο διαδίκτυο.

5.2.1 ΑΝΑΔΥΟΜΕΝΕΣ ΠΡΟΚΛΗΣΕΙΣ ΚΑΙ ΚΙΝΔΥΝΟΙ ΠΑΡΑΒΙΑΣΗΣ ΙΔΙΩΤΙΚΟΤΗΤΑΣ ΑΠΟ ΕΦΑΡΜΟΓΕΣ ΚΑΙ ΣΥΣΤΗΜΑΤΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

1) Απειλή ψηφιακής ιδιωτικότητας από τρωτότητες στο Διαδίκτυο των Πραγμάτων (IoT): Η άνθιση της Τεχνητής Νοημοσύνης έχει επιδράσει σημαντικά στον τομέα του Διαδικτύου των Πραγμάτων και έχει ενισχύσει τις δυνατότητές του, προσδίδοντας μια πιο έξυπνη λειτουργικότητα και αποτελεσματικότητα στις εφαρμογές του. Επιπλέον μέσω της άνθισης της τεχνητής νοημοσύνης, το Διαδίκτυο των πραγμάτων έχει εξελιχθεί περαιτέρω, αναδεικνύοντας νέες δυνατότητες και λύνοντας προβλήματα που προηγουμένως ήταν δύσκολο να αντιμετωπιστούν. Το μεγαλύτερο πλεονέκτημα είναι η αύξηση του όγκου των δεδομένων που παράγονται από τις συσκευές IoT. Οι προηγμένοι αλγόριθμοι της ΤΝ μπορούν να αναλύουν αυτά τα δεδομένα με μεγαλύτερη ακρίβεια και ταχύτητα, εξάγοντας πολύτιμες πληροφορίες και προβλέποντας μελλοντικές τάσεις. Αυτό δημιουργεί μια ευκαιρία για τη βελτίωση της λειτουργικότητας και της αποτελεσματικότητας των συσκευών IoT, καθώς και για τη δημιουργία νέων υπηρεσιών και εφαρμογών που επιτρέπουν την πιο έξυπνη διαχείριση και χρήση των δεδομένων. Επιπλέον, η διασύνδεση των συσκευών IoT επιτυγχάνεται με

μεγαλύτερη ομαλότητα και αποδοτικότητα λόγω της εξελιγμένης τεχνολογίας της TN. Αυτό συμβάλλει στη δημιουργία ενός πιο συνδεδεμένου και ενιαίου περιβάλλοντος, όπου οι συσκευές μπορούν να επικοινωνούν και να συνεργάζονται μεταξύ τους με μεγαλύτερη ευκολία και αποτελεσματικότητα. Είναι γεγονός ότι η χρήση συσκευών IoT κλιμακώνεται εκθετικά, με σχετικές εκτιμήσεις ότι έως το 2025 θα υπάρχουν 55,7 δισεκατομμύρια συνδεδεμένες συσκευές IoT, οι οποίες θα είναι σε θέση να παράγουν 80 δισεκατομμύρια ζέταμπάιτ (ZB) δεδομένων [81].

Ωστόσο, όλα αυτά εγείρουν ανησυχίες σχετικά με το απόρρητο. Συχνά, ο τρόπος συλλογής και ανάλυσης δεδομένων από τις συσκευές IoT δεν υλοποιείται με διαφάνεια και δεν διαθέτουν επαρκή μέτρα προστασίας, θέτοντας σε κίνδυνο το ιδιωτικό απόρρητο των χρηστών. Επιπλέον οι συσκευές IoT, συχνά, διαχειρίζονται δεδομένα χωρίς να λαμβάνουν υπόψη τη συνάφεια των δεδομένων. Το εύρος των δεδομένων που υπόκεινται σε επεξεργασία περιλαμβάνει ευαίσθητες και προσωπικές πληροφορίες, συμπεριλαμβανομένων ακόμη και βιομετρικών δεδομένων, οπότε η πιθανή διαρροή δεδομένων σε τρίτους θα έθετε σε μεγάλο κίνδυνο τους χρήστες. Ακόμη, οι χρήστες μπορεί να μην έχουν γνώση της πλήρους έκτασης και του τρόπου συλλογής δεδομένων λόγω της έλλειψης μιας σαφούς και διαφανούς πολιτικής απορρήτου [75]. Ως εκ τούτου, οι χρήστες ενδέχεται να μην κατανοούν πλήρως ποια δεδομένα συλλέγονται, για ποιον λόγο συλλέγονται και πώς προστατεύονται. Ιδιαίτερος ανησυχητικό είναι επίσης το γεγονός ότι πολλές συσκευές IoT έχουν ορίσει προκαθορισμένα στοιχεία πρόσβασης των χρηστών, τα οποία είναι εύκολο να εντοπιστούν καθώς δεν είναι πολύπλοκα. Οι κατασκευαστές συχνά παραμελούν τη σημασία της αλλαγής αυτών των προκαθορισμένων ρυθμίσεων κατά την αρχική διαμόρφωση, πράγμα που διευκολύνει τους εισβολείς στην απόκτηση μη εξουσιοδοτημένης πρόσβασης.

2) Ψηφιακοί βοηθοί (Digital assistants): Οι ψηφιακοί βοηθοί αντιπροσωπεύουν μια σημαντική εξέλιξη στον κόσμο της τεχνολογίας, προσφέροντας έναν εντελώς νέο τρόπο αλληλεπίδρασης με τις συσκευές και τις υπηρεσίες. Με τη δύναμη της φωνής, μπορούν πλέον οι χρήστες να διαχειριστούν το περιβάλλον στο οποίο βρίσκονται, να αναζητήσουν πληροφορίες, να αναπαράγουν μουσική και πολλά άλλα, χρησιμοποιώντας απλά τη φωνή τους.

Αυτή η δυνατότητα αλληλεπίδρασης διαμορφώνει ένα πιο φυσικό και ομαλό τρόπο ζωής, καθιστώντας την τεχνολογία πιο προσιτή και φιλική. Για τον παραπάνω λόγο, ολοένα και περισσότεροι χρήστες χρησιμοποιούν εφαρμογές ψηφιακών βοηθών, ενώ οι φωνητικά ελεγχόμενες διεπαφές χρήστη έχουν ενσωματωθεί σε μια ευρεία γκάμα συσκευών του Διαδικτύου των Πραγμάτων (IoT). Επιπλέον με την εξέλιξη της τεχνητής νοημοσύνης, οι ψηφιακοί βοηθοί γίνονται όχι μόνο πιο έξυπνοι αλλά και πιο προσαρμοστικοί, προβλέποντας τις ανάγκες και παρέχοντας εξατομικευμένες λύσεις για κάθε χρήση.

Παρά τη χρησιμότητα και την αυτοματοποίηση που προσφέρουν αυτά τα λογισμικά, πολλοί έχουν αρνητική στάση απέναντί τους, εκφράζοντας αμφιβολίες και ανησυχίες σχετικά με το πόσο ασφαλής είναι όσον αφορά το απόρρητο.

Το βασικό πρόβλημα με αυτήν τη ψηφιακή υπηρεσία είναι ότι το μικρόφωνο της συσκευής πρέπει να είναι συνεχώς ενεργοποιημένο, έτοιμο να ανταποκριθεί σε μια εντολή εκκίνησης και οι χρήστες δεν γνωρίζουν ότι οι συνομιλίες τους πιθανώς καταγράφονται. Πολλές φορές μάλιστα οι ψηφιακοί βοηθοί μπορεί να παρερμηνεύσουν μια λέξη ως εντολή εκκίνησης και να ενεργοποιηθούν, καταγράφοντας ομιλίες των χρηστών. Οι εν λόγω υπηρεσίες, ενεργοποιούνται από συγκεκριμένες εντολές εκκίνησης ("Hi Siri", "Hi Google"), τις οποίες μπορούν επιπλέον να ενσωματώσουν στα διαφημιστικά βίντεο τους πολλές διαφημιστικές εταιρίες. Αυτό επιτυγχάνεται ενεργοποιώντας τον ψηφιακό βοηθό, ώστε να αντλεί πληροφορίες για τις διαφημίσεις που παρακολουθεί ο χρήστης, σκιαγραφώντας έτσι το κατάλληλο καταναλωτικό προφίλ με σκοπό τη στοχευμένη μελλοντική προώθηση διαφημίσεων [82].

Ένας ακόμη κίνδυνος είναι η κυβερνοεπίθεση "Dolphin Attack", που μπορεί να εκμεταλλευτεί τους ψηφιακούς βοηθούς με σκοπό να προκαλέσει ανεπιθύμητες ενέργειες ή αποκαλύψει πληροφορίες. Πιο συγκεκριμένα, η επίθεση προκύπτει όταν ένας εισβολέας δημιουργεί μια συσκευή που εκπέμπει υψηλή συχνότητα ήχου, η οποία δεν είναι αντιληπτή από τον ανθρώπινο ακουστικό φάσμα, αλλά μπορεί να αναγνωριστεί από τους ψηφιακούς βοηθούς. Έτσι, ακούγοντας μια προγραμματισμένη σειρά από αυτούς τους ήχους, ένας εισβολέας μπορεί να ενεργοποιήσει έναν ψηφιακό βοηθό χωρίς την άμεση γνώση του χρήστη και να παρακολουθούν ιδιωτικές συνομιλίες του χρήστη [82].

Επιπλέον τέτοιου είδους συστήματα μπορούν να παραβιαστούν από τρίτους με κυβερνοεπιθέσεις man-in-the-middle (MITM) [82]. Οι επιθέσεις αυτές μπορούν να εκμεταλλευτούν πιθανές αδυναμίες στην κρυπτογράφηση ή στην αυθεντικοποίηση των επικοινωνιών μεταξύ του ψηφιακού βοηθού και των χρηστών. Αυτό συμβαίνει αν ένας ψηφιακός βοηθός δεν χρησιμοποιεί κρυπτογράφηση ή αν η κρυπτογράφηση του διακόπτεται ή αλλοιώνεται κατά τη μετάδοση των δεδομένων, με αποτέλεσμα να είναι ευάλωτος σε MITM επιθέσεις. Κάτι τέτοιο θα έθετε τον χρήστη σε μεγάλο κίνδυνο, καθώς ο επιτιθέμενος θα είχε πρόσβαση σε προσωπικά δεδομένα και θα μπορούσε να τροποποιήσει ή ακόμα και να αποκτήσει έλεγχο επάνω στις επικοινωνίες που διαχειρίζεται ο ψηφιακός βοηθός [82].

Η χρήση ψηφιακών βοηθών εγείρει επίσης σημαντικά θέματα ιδιωτικότητας σχετικά με τα ηχητικά δεδομένα που δημιουργούνται από τους χρήστες. Η ακριβής διαχείριση και χρήση των εν λόγω δεδομένων παραμένει ασαφής, δημιουργώντας ανησυχίες για τον τρόπο που χειρίζονται οι εταιρείες την ιδιωτικότητα των χρηστών. Δεν υπάρχει διαφάνεια στο κατά πόσο οι εταιρείες διεισδύουν στα ηχητικά δεδομένα αυτά και για ποιον σκοπό.

Οι κορυφαίες εταιρίες Google και Amazon αναπτύσσουν και διαθέτουν στην αγορά μια πληθώρα ψηφιακών βοηθών που βασίζονται σε προηγμένες

τεχνολογίες. Αυτοί οι βοηθοί καταγράφουν διάφορες δραστηριότητες, όπως αναζητήσεις στο διαδίκτυο ή ελέγχους έξυπνων συσκευών, με στόχο να παρέχουν εξατομικευμένες υπηρεσίες στους χρήστες. Και οι δύο εταιρίες δίνουν έμφαση στην ασφάλεια και την προστασία των δεδομένων των χρηστών, κρυπτογραφώντας τα δεδομένα κατά τη μεταφορά και κατά την αποθήκευσή τους. Ωστόσο, είναι γνωστό ότι σε περίπτωση που κυβερνητικές αρχές ζητήσουν πρόσβαση σε δεδομένα από αυτούς τους βοηθούς, μπορεί να γίνει αναφορά σε λογαριασμούς που σχετίζονται με τους χρήστες [80]. Επιπλέον, η πρόσβαση και η διαγραφή των δεδομένων αποτελεί δύσκολη διαδικασία για τους χρήστες, οι οποίοι δυσκολεύονται να ανακτήσουν πληροφορίες σχετικά με τη διαχείριση των προσωπικών τους δεδομένων και την προστασία της ιδιωτικότητάς τους [82].

Ενόψει του συνεχούς αυξανόμενου ενδιαφέροντος και των ανησυχιών που εκφράζουν οι τελικοί χρήστες σχετικά με την ιδιωτικότητα τους, διάφοροι ψηφιακοί βοηθοί, παρέχουν στους τελικούς χρήστες τη δυνατότητα να ρυθμίζουν τη συλλογή και την επεξεργασία των δεδομένων τους. Θέλοντας έτσι να διασφαλίσουν τη προστασία του ιδιωτικού απορρήτου των χρηστών και να εξασφαλιστεί ότι οι τελικοί χρήστες θα είναι βέβαιοι για την ασφάλεια του. Πιο συγκεκριμένα οι τελικοί χρήστες μπορούν να προσαρμόσουν τις ρυθμίσεις απορρήτου σύμφωνα με τις προτιμήσεις τους να κάνουν ελέγχους και να διαγράφουν προηγούμενες φωνητικές εγγραφές. Ωστόσο, πολλοί τελικοί χρήστες συχνά δεν γνωρίζουν την ύπαρξη αυτών των μεθόδων ή ακόμη θεωρούν ότι τέτοιες μέθοδοι δεν είναι αρκετές για να διασφαλιστεί η απόλυτη ασφάλεια. Αυτό έχει ως αποτέλεσμα, οι διαθέσιμοι μηχανισμοί προστασίας της ιδιωτικότητας να είναι σχεδόν άγνωστοι ή μη χρησιμοποιημένοι από τους περισσότερους τελικούς χρήστες [82].

3) AI profiling : Η τεχνική του AI profiling αποτελεί μια σημαντική επίτευξη στον τομέα της τεχνητής νοημοσύνης (TN), η οποία εστιάζει στην ανάλυση και κατασκευή προφίλ ατόμων, βάσει πολυάριθμων δεδομένων. Μέσω προηγμένων αλγορίθμων και τεχνικών TN, το AI profiling αναζητά συνδέσεις και προτεραιότητες εντός των δεδομένων, με στόχο τη δημιουργία πλήρους και λεπτομερούς προφίλ με προσωπικές πτυχές που αφορούν ένα φυσικό πρόσωπο. Όλα τα παραπάνω συμβάλλουν στην ανάλυση και πρόβλεψη πτυχών που σχετίζονται με την απόδοσή του φυσικού προσώπου στην εργασία, την οικονομική του κατάσταση, την υγεία, τις προσωπικές του προτιμήσεις, τα ενδιαφέροντά του, την αξιοπιστία, τη συμπεριφορά, την τοποθεσία ή τις μετακινήσεις του [84]. Το AI profiling εφαρμόζεται σε πολλούς τομείς, συμπεριλαμβανομένων κυρίως του μάρκετινγκ, αλλά και της ασφάλειας, της υγείας και της κοινωνιολογίας. Ωστόσο, αυτή η πρακτική συχνά συνοδεύεται από ηθικά και νομικά ζητήματα, καθώς η χρήση δεδομένων για την κατασκευή προφίλ μπορεί να παραβιάζει την ιδιωτικότητα οδηγώντας σε κοινωνικό στιγματισμό.

Πρόσφατες επιστημονικές έρευνες υποδεικνύουν η τεχνική AI profiling, αποτελεί πρωτοποριακό μέσο για την αναγνώριση παραμέτρων της ψυχικής κατάστασης, καθώς τα μοντέλα μηχανικής μάθησης μπορούν να προβλέπουν με ακρίβεια την κατάθλιψη και άλλες καταστάσεις που σχετίζονται με τη ψυχική υγεία, όπως το

άγχος, τη διπολικότητα, την οριακή διαταραχή προσωπικότητας, χρησιμοποιώντας και αναλύοντας δεδομένα μέσω κοινωνικής δικτύωσης, όπως φωτογραφίες από το Instagram και αναρτήσεις στο Twitter [84].

Ενώ η τεχνική AI profiling φαίνεται να προσφέρει προοπτικές για τη βελτίωση της διάγνωσης και της παρακολούθησης της ψυχικής υγείας, η χρήση της εγείρει σημαντικά ηθικά και νομικά ζητήματα.

Μία από τις προκλήσεις είναι η προστασία της ιδιωτικότητας των ατόμων, καθώς η συλλογή και η ανάλυση προσωπικών δεδομένων μπορεί να οδηγήσει σε παραβίαση της ατομικής ιδιωτικότητας. Επιπλέον, η χρήση αυτής της τεχνολογίας μπορεί να ενισχύσει τον κοινωνικό στιγματισμό, επειδή οι ετικέτες που απορρίπτουν τα άτομα ως "ψυχικά ασθενή" μπορεί να επικολλώνται με ευκολία, βασιζόμενες σε αυτοματοποιημένες αναλύσεις δεδομένων. Ακόμη εγείρει σημαντικές ανησυχίες, όχι μόνο στον κοινωνικό τομέα, αλλά και στον επαγγελματικό, απειλώντας την επαγγελματική ανέλιξη του ατόμου. Αυτό θα μπορούσε να συμβεί καθώς τα δεδομένα αυτά θα επηρέαζαν αρνητικά πιθανές προσλήψεις σε θέσεις εργασίας, από εργοδότες που θα μπορούσαν να απορρίψουν αυτά τα άτομα ως "ψυχικά ασθενή" ή ανάξια για υψηλές θέσεις ευθύνης. Επιπλέον, η επιβολή αρνητικών ετικετών σε άτομα με βάση αυτές τις αναλύσεις θα μπορούσε να οδηγήσει σε δυσφήμιση ή ακόμα και να χρησιμοποιηθεί ως εργαλείο για πολιτική ή κοινωνική χειραγώγηση. Εν τέλει, η κοινωνική πίεση που θα ασκηθεί μπορεί να αναγκάσει άτομα να μην λαμβάνουν αυτονόμως αποφάσεις, επιτρέποντας σε άλλους να επηρεάζουν τις αποφάσεις τους.

Εν τέλει, όλες αυτές οι κατηγοριοποιήσεις ενδέχεται να επηρεάζουν τα άτομα και την ιδιωτική τους ζωή, θέτοντας τους ανθρώπους σε κίνδυνο σε πολλούς τομείς της ζωής τους, και το χειρότερο όλων είναι ότι πολλές φορές οι κατηγοριοποιήσεις αυτές δεν είναι καν ακριβείς. Βασίζονται δηλαδή σε λανθασμένες ερμηνείες ή παραβλέπουν σημαντικούς παράγοντες που επηρεάζουν τη ψυχική υγεία, καταλήγοντας σε εσφαλμένους προσδιορισμούς και κατηγοριοποιήσεις. Οι επιπτώσεις αυτές ενισχύουν την ανάγκη για προσεκτική επίβλεψη και ρύθμιση της χρήσης της τεχνολογίας AI profiling στον τομέα της ψυχικής υγείας.

4) Συστήματα αναγνώρισης προσώπου : Τα συστήματα αναγνώρισης προσώπου αποτελούν μια εντυπωσιακή και καινοτόμο εφαρμογή της εξέλιξης της τεχνητής νοημοσύνης [85]. Τα συστήματα αναγνώρισης προσώπου εφαρμόζονται σε αεροδρόμια και πόλεις, σε πολλές ευρωπαϊκές χώρες και σε όλη την Αμερική.

Όστοςο η χρήση συστημάτων αναγνώρισης προσώπου προκαλεί ανησυχίες για την ιδιωτικότητα, καθώς τα ευαίσθητα δεδομένα που χρησιμοποιούνται για τη διαδικασία της αναγνώρισης, υπάρχουν σε πολλές βάσεις δεδομένων και πηγές, όπως κοινωνικά δίκτυα μέχρι καταλόγους διπλωμάτων οδήγησης. Αυτό θα μπορούσε να συντελέσει σε σοβαρό κίνδυνο για πιθανή διαρροή προσωπικών

δεδομένων σε τρίτους, αλλά και να χρησιμοποιηθεί ως εργαλείο παρακολούθησης από τις κυβερνητικές αρχές [86].

Ένα παράδειγμα αποτελεί η κυβέρνηση της Κίνας χρησιμοποίησε τα συστήματα αναγνώρισης προσώπου για να παρακολουθεί και να ελέγχει τους Ουιγούρους, μια εθνική μειονότητα. Αυτές οι πρακτικές προκάλεσαν σοβαρές ανησυχίες όσον αφορά την παραβίαση των ανθρωπίνων δικαιωμάτων και την εθνική διάκριση, οδηγώντας σε νομοθεσίες απαγόρευσης χρήσης συστημάτων αναγνώρισης προσώπου σε πολλές περιοχές και κράτη.

5) Προγνωστική αστυνόμευση: Ακόμη ένα σημαντικό εργαλείο της τεχνητής νοημοσύνης που αναδεικνύουν την ισχύ της σύγχρονης τεχνολογίας στην ανάλυση δεδομένων και την πρόβλεψη μελλοντικών γεγονότων και συμπεριφορών είναι οι προγνωστικοί αλγόριθμοι. Οι προγνωστικοί αλγόριθμοι αναλύουν δεδομένα και εφαρμόζουν μαθηματικά μοντέλα και αλγορίθμους που επιτρέπουν την ανίχνευση προτύπων και την πρόβλεψη του μελλοντικού συμβάντος ή της τάσης.

Η προγνωστική αστυνόμευση αποτελεί έναν τομέα όπου οι προγνωστικοί αλγόριθμοι εφαρμόζονται ευρέως για την ανάλυση και πρόβλεψη εγκληματικών συμβάντων.

Με τη χρήση αλγορίθμων και μοντέλων μηχανικής μάθησης, η προγνωστική αστυνόμευση αποσκοπεί στον προσδιορισμό περιοχών και χρονικών σημείων με υψηλότερο κίνδυνο εγκληματικής δραστηριότητας, παρέχοντας έτσι στις αρχές τη δυνατότητα προληπτικών μέτρων. Οι αλγόριθμοι αυτοί βασίζονται στη συλλογή και ανάλυση διαφόρων δεδομένων, όπως ιστορικά δεδομένα εγκληματικής δραστηριότητας, πληθυσμιακές πληροφορίες, δεδομένα κίνησης και χρήσης οχημάτων, κοινωνικά δίκτυα και άλλες πηγές δεδομένων που σχετίζονται με την εγκληματική δραστηριότητα.

Παρόλα αυτά, η χρήση τέτοιων δεδομένων εγείρει σοβαρά ζητήματα ιδιωτικότητας, καθώς μπορεί να παραβιάζει το προσωπικό απόρρητο και να δημιουργήσει προβλήματα όπως ανεπιθύμητη παρακολούθηση ή διακρίσεις και παραβιάσεις των δικαιωμάτων. Επιπλέον, η εφαρμογή αλγορίθμων μηχανικής μάθησης μπορεί να βασίζεται σε στερεοτυπικές αντιλήψεις ή βεβαιασμένες προσεγγίσεις, με αποτέλεσμα να ενισχύονται υφιστάμενες διακρίσεις ή να δημιουργούνται νέες. Ένα επίκαιρο παράδειγμα αποτελεί το πρόγραμμα COMPAS, το οποίο χρησιμοποιείται για την πρόβλεψη ανάπτυξης εγκληματικής δραστηριότητας και έχει κατηγορηθεί για ανισομερή μεταχείριση μεταξύ διαφορετικών ομάδων ανθρώπων [87].

Ακόμη ένα παράδειγμα που αναδεικνύει την ευαισθησία των αλγορίθμων πρόβλεψης όσον αφορά την ιδιωτικότητα είναι ο αποτυχημένος αλγόριθμος πρόσληψης της Amazon [88]. Η Amazon, μία από τις μεγαλύτερες και πιο επιδραστικές εταιρείες στον κόσμο της τεχνολογίας, είχε αναπτύξει ένα σύστημα πρόβλεψης προσλήψεων, που αποδείχτηκε ότι υπέβαλε τις γυναίκες σε διάκριση. Ο αλγόριθμος χρησιμοποίησε ιστορικά δεδομένα πρόσληψης από την Amazon, τα

οποία δυστυχώς αντανακλούσαν υπάρχουσες προκαταλήψεις ή προτιμήσεις στις πρακτικές πρόσληψης της εταιρείας κατά των γυναικών.

Επομένως, οι αλγόριθμοι πρόβλεψης μπορούν να επιφέρουν σοβαρές επιπτώσεις στην ιδιωτικότητα στο διαδίκτυο, καθώς εξαρτώνται από ευαίσθητα δεδομένα που τους τροφοδοτούν και η ύπαρξη προκαταλήψεων ή ανισοτήτων στα δεδομένα εκπαίδευσης μπορεί να οδηγήσει σε ανεπιθύμητες αποτελέσματα, ενισχύοντας ανισότητες και διακρίσεις. Αυτό οδηγεί σε απουσία διαφάνειας και έλλειψη ισότητας και δικαιοσύνης για όλους τους χρήστες του διαδικτύου.

ΚΕΦΑΛΑΙΟ 6 - ΣΥΜΠΕΡΑΣΜΑΤΑ

Είναι αδιαμφισβήτητο ότι η τεχνητή νοημοσύνη κατατάσσεται ανάμεσα στις πιο εμβληματικές επιστημονικές εξελίξεις της ανθρωπότητας, λόγω της ανατρεπτικής της επίδρασης. Ο αντίκτυπός της είναι συγκρίσιμος με αυτόν άλλων σημαντικών επιστημονικών και τεχνολογικών επιτευγμάτων, δεδομένου ότι έχει τη δυνατότητα να επηρεάσει και να μετασχηματίσει κάθε πτυχή της ανθρώπινης δραστηριότητας. Ωστόσο, η εκτενής και δυναμική φύση της τεχνητής νοημοσύνης ενέχει τον κίνδυνο να εξελιχθεί σε ένα ανεξέλεγκτο φαινόμενο, εφόσον η απεριόριστη δύναμή της δεν υπόκειται σε κατάλληλη ρύθμιση και έλεγχο.

Η τεχνητή νοημοσύνη έχει τη δυνατότητα να συμβάλλει ουσιαστικά στην πρόοδο της ιατρικής, της επιστήμης και της γενικής εξέλιξης της ανθρωπότητας, οδηγώντας σε ένα πιο αυτοματοποιημένο και ασφαλές μέλλον. Μέσω της εφαρμογής της τεχνητής νοημοσύνης, μπορούμε να μειώσουμε τα ατυχήματα στους δρόμους με την ενσωμάτωσή της σε συστήματα Διαδικτύου των Πραγμάτων (IoT), να ενισχύσουμε την προστασία από πυρκαγιές και να βελτιώσουμε την ασφάλεια από εγκληματικές δραστηριότητες, όπως κλοπές και επιθέσεις, μέσω της χρήσης καμερών σε δημόσιους χώρους. Αυτές οι εφαρμογές μπορούν να οδηγήσουν σε σημαντική βελτίωση της ποιότητας ζωής και της ασφάλειας, εφόσον χρησιμοποιηθούν με υπευθυνότητα και σωστό σχεδιασμό.

Θα ήταν πραγματικά λυπηρό αν η ανθρωπότητα δεν εκμεταλλευόταν πλήρως τα οφέλη μιας επιστήμης με τέτοιο ευεργετικό δυναμικό για κάθε πτυχή της καθημερινότητας. Όπως στο παρελθόν ορισμένες σημαντικές επιστημονικές ανακαλύψεις, όπως η πυρηνική ενέργεια και η ραδιενέργεια, χρησιμοποιήθηκαν κακόβουλα ή υπήρξαν ατελείς στις επιδιώξεις τους, προκαλώντας σοβαρά προβλήματα αντί να προάγουν την ανθρωπότητα, έτσι και η τεχνητή νοημοσύνη κινδυνεύει να μην αξιοποιηθεί κατάλληλα. Εάν χρησιμοποιηθεί με τη σωστή στρατηγική και προσοχή, η τεχνητή νοημοσύνη έχει τη δυνατότητα να προσφέρει στην ανθρωπότητα οφέλη που δεν έχει δει ποτέ προηγουμένως από καμία άλλη επιστήμη.

Η ανθρωπότητα οφείλει να αξιοποιήσει την τεχνητή νοημοσύνη με τον κατάλληλο τρόπο, προκειμένου να αποφευχθεί η μετατροπή της σε ένα καταστροφικό σενάριο.

Τα οφέλη της τεχνητής νοημοσύνης είναι τόσο εκτενή και καινοτόμα, που με την εφαρμογή σωστής στρατηγικής, κατάλληλων μέτρων, και τη θέσπιση πολιτικών, νομοθεσιών και οργανισμών υπεύθυνων για τη διασφάλιση της ασφάλειας (όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων - GDPR και άλλες συναφείς ρυθμίσεις), η τεχνητή νοημοσύνη μπορεί να αξιοποιηθεί για να προσφέρει σημαντικά οφέλη. Εντούτοις, χωρίς την κατάλληλη ρύθμιση και επιτήρηση, υπάρχει κίνδυνος να μετατραπεί από ένα ευεργετικό εργαλείο σε μια πηγή δυνητικών κινδύνων.

Επομένως η νομοθεσία παίζει τον πιο καθοριστικό ρόλο στην εξασφάλιση της ισορροπίας στον τομέα της τεχνητής νοημοσύνης και της κυβερνοασφάλειας. Τα

νομοθετικά μέτρα και οι κανονισμοί πρέπει να αναθεωρούνται και να επιβλέπονται διαρκώς για να ανταποκριθούν στις εξελίξεις. Επίσης, η ίδρυση επιπλέον σωμάτων και οργανισμών που θα προωθούν τη συνεχή παρακολούθηση και αναθεώρηση της νομοθεσίας είναι απαραίτητη για την αποτελεσματική διαχείριση των κινδύνων.

Επιπλέον, είναι κρίσιμο οι πολίτες και οι χρήστες να έχουν το δικαίωμα όχι μόνο να εκφράζουν τη γνώμη τους, αλλά και να συμμετέχουν ενεργά μέσω ψηφοφοριών σε ζητήματα που σχετίζονται με την τεχνητή νοημοσύνη, τις σχετικές νομοθεσίες και την προστασία των προσωπικών τους δεδομένων. Η συμμετοχή τους σε αυτές τις διαδικασίες διασφαλίζει ότι οι πολιτικές και οι κανονισμοί που επηρεάζουν τη ψηφιακή τους ζωή αντικατοπτρίζουν τις ανάγκες και τις ανησυχίες τους, προωθώντας έτσι μια δικαιότερη και πιο διαφανή διαχείριση της τεχνολογίας.

Είναι απαραίτητο να θεσπιστούν οι κατάλληλες διαδικασίες και να εφαρμοστούν τα σωστά μέτρα για να διασφαλιστεί ότι η τεχνητή νοημοσύνη θα συνεισφέρει θετικά στην κυβερνοασφάλεια αντί να προκαλέσει βλάβες. Παράλληλα, απαιτείται η ανάπτυξη σύγχρονων μεθόδων και τεχνολογιών για την αποτελεσματική αντιμετώπιση των εξελιγμένων και καινοτόμων κυβερνοεπιθέσεων. Επίσης, οι κυβερνήσεις και οι οργανισμοί πρέπει να συνεργαστούν για την κατάρτιση πολιτικών και κανονισμών, καθώς και για την επένδυση σε πρόσθετα κονδύλια. Αν και ήδη καταβάλλονται προσπάθειες προς αυτήν την κατεύθυνση, η συνεχής ενίσχυση και διεύρυνση των επενδύσεων είναι κρίσιμη για την αποτελεσματική θωράκιση της κυβερνοασφάλειας.

Ακόμη θα μπορούσαν να ιδρυθούν παγκόσμιοι οργανισμοί που θα αναλαμβάνουν την ευθύνη για την προστασία της ανθρώπινης και κρατικής ασφάλειας από ψηφιακούς κινδύνους, παρόμοιοι με τον Οργανισμό Ηνωμένων Εθνών (ΟΗΕ). Αυτοί οι οργανισμοί θα μπορούσαν να ενισχύσουν τη διεθνή συνεργασία και συντονισμό στην αντιμετώπιση των ψηφιακών και φυσικών απειλών.

Χωρίς καμία αμφιβολία ένας ακόμη κρίσιμος σύμμαχος στη διαχείριση των προκλήσεων που συνεπάγεται η τεχνητή νοημοσύνη είναι και η εκπαίδευση, καθώς και η συνεχής ενημέρωση. Επενδύοντας λοιπόν στην εκπαίδευση των χρηστών και των πολιτών στην ενημέρωση για τις νέες απειλές, μπορεί να ενισχυθεί η αντίσταση της κοινωνίας απέναντι σε ψηφιακές επιθέσεις και κακόβουλες πρακτικές. Η εκτενή φύση του διαδικτύου μπορεί να αξιοποιηθεί ως εργαλείο ενημέρωσης κατά της παραπληροφόρησης και των ψηφιακών απειλών. Η δημιουργία επίσημων ιστότοπων που θα παρέχουν ανάλυση και ενημέρωση για τις τελευταίες εξελίξεις και απειλές στον ψηφιακό κόσμο είναι αναγκαία για τη διασφάλιση αξιόπιστης πληροφόρησης.

Όσον αφορά τα μέτρα ασφαλείας B2C, οι οργανισμοί και οι επιχειρήσεις οφείλουν να συνεργάζονται με εξειδικευμένους εμπειρογνώμονες στον τομέα της ασφάλειας δικτύων και των υπολογιστικών συστημάτων. Αυτοί οι ειδικοί θα αναπτύξουν και θα υλοποιήσουν ολοκληρωμένα πλάνα ασφαλείας για την προστασία των δεδομένων και των υποδομών.

Επιπλέον, οι κατασκευαστές εφαρμογών και οι φορείς του ψηφιακού εμπορίου πρέπει να λογοδοτούν για τη συμμόρφωση με τα πρότυπα προστασίας της ιδιωτικότητας και της ασφάλειας των δεδομένων. Είναι αναγκαίο να υιοθετηθούν αποτελεσματικά μέτρα, όπως η ψευδωνυμοποίηση και η ανωνυμοποίηση των δεδομένων, προκειμένου να προστατευτούν τα προσωπικά στοιχεία από μη εξουσιοδοτημένη πρόσβαση και κακή χρήση. Επιπλέον, η ενίσχυση της κρυπτογράφησης των δεδομένων είναι κρίσιμη για την εξασφάλιση της ακεραιότητας και της εμπιστευτικότητάς τους, τόσο κατά τη μεταφορά όσο και κατά την αποθήκευσή τους. Αυτά τα μέτρα είναι απαραίτητα για τη δημιουργία ενός ασφαλούς ψηφιακού περιβάλλοντος που προστατεύει τους χρήστες και ενισχύει την εμπιστοσύνη στο ψηφιακό οικοσύστημα.

Τέλος, θα μπορούσαν να δοθούν σημαντικά κονδύλια, βραβεύσεις και να δημιουργηθούν ειδικά προγράμματα για να ενθαρρυνθούν οι επιστήμονες και οι νέοι άνθρωποι να επενδύσουν στην αμυντική κυβερνοασφάλεια. Αυτές οι πρωτοβουλίες θα μπορούσαν να περιλαμβάνουν:

1. **Χρηματοδοτικά Κονδύλια:** Επενδύσεις σε έρευνα και ανάπτυξη νέων τεχνολογιών και μεθόδων για την αντιμετώπιση των αυξανόμενων προβλημάτων στον κυβερνοχώρο. Τα κονδύλια θα μπορούσαν να διατεθούν σε ερευνητικά κέντρα, πανεπιστήμια και ιδιωτικούς οργανισμούς που εργάζονται στον τομέα της κυβερνοασφάλειας.
2. **Βραβεύσεις και Διακρίσεις:** Ετήσια βραβεία για καινοτόμες λύσεις και εξαιρετικές επιδόσεις στον τομέα της κυβερνοασφάλειας. Αυτά τα βραβεία θα μπορούσαν να αναγνωρίζουν πρωτοπόρους και νέες ιδέες που συνεισφέρουν σημαντικά στην προστασία των ψηφιακών υποδομών και δεδομένων.
3. **Ειδικά Προγράμματα και Ενθάρρυνση:** Δημιουργία προγραμμάτων υποστήριξης και επιβράβευσης για νέους επιστήμονες και επαγγελματίες στον τομέα της κυβερνοασφάλειας. Αυτά τα προγράμματα θα μπορούσαν να περιλαμβάνουν υποτροφίες, εκπαιδευτικά σεμινάρια, και ευκαιρίες συνεργασίας με κορυφαίους οργανισμούς και ερευνητικά κέντρα.
4. **Δημόσιες Καμπάνιες και Ενημέρωση:** Εκστρατείες ευαισθητοποίησης και ενημέρωσης για την αναγκαιότητα της κυβερνοασφάλειας, προκειμένου να προσελκύσουν περισσότερους νέους ανθρώπους και επιστήμονες στον τομέα. Η ενίσχυση της κατανόησης της σημασίας της κυβερνοασφάλειας μπορεί να δημιουργήσει μεγαλύτερο ενδιαφέρον και κίνητρα για συμμετοχή.

Με αυτές τις πρωτοβουλίες, μπορεί να επιτευχθεί ενίσχυση της καινοτομίας και της ανθεκτικότητας στον τομέα της κυβερνοασφάλειας, συμβάλλοντας στη δημιουργία ασφαλέστερων ψηφιακών περιβαλλόντων για όλους.

Συνοπτικά το να μετατραπεί η τεχνητή νοημοσύνη από μια δυνητικά αρνητική δύναμη σε μια ευεργετική επίτευξη είναι αναμφίβολα δύσκολο, αλλά όχι αδύνατο. Με την κοινή προσπάθεια όλων των ενδιαφερόμενων φορέων του χώρου, καθώς και με τη δέσμευση και την ατομική προσπάθεια του καθενός, μπορούμε να διασφαλίσουμε ότι η τεχνητή νοημοσύνη θα εξελιχθεί σε έναν θετικό παράγοντα που θα προσφέρει απτές ωφέλειες για την κοινωνία.

ΑΝΑΦΟΡΕΣ

1. Xu, Y.; Liu, X.; Cao, X.; Huang, C.; Liu, E.; Qian, S.; Liu, X.; Wu, Y.; Dong, F.; Qiu, C.-W.; et al. Artificial Intelligence: A Powerful Paradigm for Scientific Research. *Innovation (Camb)* **2021**, 2, 100179, doi:10.1016/j.xinn.2021.100179.
2. Turing, A.M. Computing Machinery and Intelligence. In *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*; Epstein, R., Roberts, G., Beber, G., Eds.; Springer Netherlands: Dordrecht, 2009; pp. 23–65 ISBN 978-1-4020-6710-5.
3. Machine Learning: A Probabilistic Perspective [Book] / auth. Murphy Kevin P.. - Cambridge, Massachussets : MT Press, 2012.
4. Schneier on Security: Privacy and Control // <https://journalprivacyconfidentiality.org/index.php/jpc/article/view/579/562>. - April 6th, 2010.
5. Seldon Supervised vs Unsupervised Learning Explained Available online: <https://www.seldon.io/supervised-vs-unsupervised-learning-explained> (accessed on 1 July 2024).
6. LeCun, Y.; Bengio, Y.; Hinton, G. Deep Learning. *Nature* **2015**, 521, 436–444, doi:10.1038/nature14539.
7. Data Management for Production Quality Deep Learning Models: Challenges and Solutions. *Journal of Systems and Software* **2022**, 191, 111359, doi:10.1016/j.jss.2022.111359.
8. Bush, Vannevar. 1945. As We May Think. *The Atlantic Monthly*. July 1945.
9. Roh, Y.; Heo, G.; Whang, S.E. A Survey on Data Collection for Machine Learning: A Big Data - AI Integration Perspective. *IEEE Transactions on Knowledge and Data Engineering* **2021**, 33, 1328–1347, doi:10.1109/TKDE.2019.2946162.
10. Halevy, A.; Korn, F.; Noy, N.F.; Olston, C.; Polyzotis, N.; Roy, S.; Whang, S.E. Goods: Organizing Google’s Datasets. In *Proceedings of the Proceedings of the 2016 International Conference on Management of Data*; Association for Computing Machinery: New York, NY, USA, March 14 2016; pp. 795–806.
11. Cafarella, M.J.; Halevy, A.; Khoussainova, N. Data Integration for the Relational Web. *Proc. VLDB Endow.* **2009**, 2, 1090–1101, doi:10.14778/1687627.1687750.
12. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-Learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.
13. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks 2014.
14. Goodfellow, I. NIPS 2016 Tutorial: Generative Adversarial Networks 2017.

15. Peng, X.; Sun, B.; Ali, K.; Saenko, K. Learning Deep Object Detectors from 3D Models. In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV); September 2015; pp. 1278–1286.
16. Kim, S.-H.; Choe, G.; Ahn, B.; Kweon, I.S. Deep Representation of Industrial Components Using Simulated Images. In Proceedings of the 2017 IEEE International Conference on Robotics and Automation (ICRA); February 2017; pp. 2003–2010.
17. Jaderberg, M.; Simonyan, K.; Vedaldi, A.; Zisserman, A. Synthetic Data and Artificial Neural Networks for Natural Scene Text Recognition 2014.
18. Gupta, A.; Vedaldi, A.; Zisserman, A. Synthetic Data for Text Localisation in Natural Images.; 2016; pp. 2315–2324.
19. Bai, Y.; Yang, K.; Yu, W.; Xu, C.; Ma, W.-Y.; Zhao, T. Automatic Image Dataset Construction from Click-through Logs Using Deep Neural Network. In Proceedings of the Proceedings of the 23rd ACM international conference on Multimedia; Association for Computing Machinery: New York, NY, USA, July 13 2015; pp. 441–450.
20. Mallinson, J.; Sennrich, R.; Lapata, M. Paraphrasing Revisited with Neural Machine Translation. In Proceedings of the Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers; Lapata, M., Blunsom, P., Koller, A., Eds.; Association for Computational Linguistics: Valencia, Spain, December 2017; pp. 881–893.
21. Marcus, A.; Parameswaran, A. Crowdsourced Data Management: Industry and Academic Perspectives. *DBS* **2015**, *6*, 1–161, doi:10.1561/19000000044.
22. Amazon Mechanical Turk Available online: <https://www.mturk.com/> (accessed on 1 July 2024).
23. Triguero, I.; García, S.; Herrera, F. Self-Labeled Techniques for Semi-Supervised Learning: Taxonomy, Software and Empirical Study. *Knowl Inf Syst* **2015**, *42*, 245–284, doi:10.1007/s10115-013-0706-y.
24. Yarowsky, D. Unsupervised Word Sense Disambiguation Rivaling Supervised Methods. In Proceedings of the Proceedings of the 33rd annual meeting on Association for Computational Linguistics; Association for Computational Linguistics: USA, March 26 1995; pp. 189–196.
25. Tri-Training: Exploiting Unlabeled Data Using Three Classifiers | IEEE Journals & Magazine | IEEE Xplore Available online: <https://ieeexplore.ieee.org/document/1512038> (accessed on 1 July 2024).
26. Allahbakhsh, M.; Benatallah, B.; Ignjatovic, A.; Motahari-Nezhad, H.R.; Bertino, E.; Dustdar, S. Quality Control in Crowdsourcing Systems: Issues and Directions. *IEEE Internet Computing* **2013**, *17*, 76–81, doi:10.1109/MIC.2013.20.
27. Facilitating Compliance With Data Lifecycle Management Available online: <https://www.techtarget.com/searchstorage/IronMountainCloud/Facilitating-Compliance-With-Data-Lifecycle-Management> (accessed on 1 July 2024).

28. Facilitating Compliance With Data Lifecycle Management Available online: <https://www.techtarget.com/searchstorage/IronMountainCloud/Facilitating-Compliance-With-Data-Lifecycle-Management> (accessed on 1 July 2024).
29. Fan, C.; Chen, M.; Wang, X.; Wang, J.; Huang, B. A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. *Front. Energy Res.* **2021**, *9*, doi:10.3389/fenrg.2021.652801.
30. Kalkman, S.; Mostert, M.; Gerlinger, C.; van Delden, J.J.M.; van Thiel, G.J.M.W. Responsible Data Sharing in International Health Research: A Systematic Review of Principles and Norms. *BMC Medical Ethics* **2019**, *20*, 21, doi:10.1186/s12910-019-0359-9.
31. Knoppers, B.M. Framework for Responsible Sharing of Genomic and Health-Related Data. *HUGO J* **2014**, *8*, 3, doi:10.1186/s11568-014-0003-1.
32. UK National Data Guardian for Health and Care's Review of Data Security: Trust, Better Security and Opt-Outs on Portico Available online: <https://access.portico.org/Portico/auView?auld=ark:%2F27927%2Fphw1p060z03> (accessed on 1 July 2024).
33. OECD Principles and Guidelines for Access to Research Data from Public Funding Available online: https://www.jstage.jst.go.jp/article/dsj/6/0/6_0_OD4/_article/-char/ja/ (accessed on 1 July 2024).
34. Annexe 2. WMA Declaration of Taipei on ethical considerations regarding health databases and biobanks. Adopted by the 53rd WMA General Assembly, Washington, DC, USA, October 2002 and revised by the 67th WMA General Assembly, Taipei, Taiwan, October 2016. *Journal international de bioéthique et d'éthique des sciences* **2017**, *28*, 113–117, doi:10.3917/jib.283.0113.
35. Kelly, M.; Furey, E.; Blue, J. GDPR Article 17: Eradicating Personal Identifiable Information & Achieving Compliance in a Hybrid Cloud. In Proceedings of the 2019 30th Irish Signals and Systems Conference (ISSC); June 2019; pp. 1–6.
36. Costin, D.; Militaru, C. ASSET MANAGEMENT TOWARDS ISO/IEC 27001:2005 ACCREDITATION OF AN INFORMATION SECURITY MANAGEMENT SYSTEM.
37. Applied Sciences | Free Full-Text | Re-Thinking Data Strategy and Integration for Artificial Intelligence: Concepts, Opportunities, and Challenges Available online: <https://www.mdpi.com/2076-3417/13/12/7082> (accessed on 1 July 2024).
38. HEINONLINE / auth. Woodhams Samuel // China, Africa, and the Private Surveillance Industry.
39. SPINGEL LINK / auth. Felipe González-Pizarro Andrea Figueroa, Claudia López & Cecilia Aragon // Regional Differences in Information Privacy Concerns After the Facebook-Cambridge Analytica Data Scandal. - February 14 , 2022. -<https://link.springer.com/article/10.1007/s10606-021-09422-3>.
40. Kiss, J. Google Admits Collecting Wi-Fi Data through Street View Cars. *The Guardian* 2010.

41. X.Sup20 : ITU-T X.1205 - Supplement on Framework of Security Information Sharing Negotiation Available online: <https://www.itu.int/rec/T-REC-X.Sup20-201304-I> (accessed on 27 August 2024).
42. Kask, R.J. Cyberspace Policy Review: Assuring a Trusted and Resilient Information and Communications Infrastructure. In; 2011; pp. 49–127.
43. Goosen, R.; Rontojannis, A.; Deutscher, S.; Rogg, J.; Bohmayr, W.; Mkrtchian, D. Artificial Intelligence Is a Threat to Cybersecurity. It's Also a Solution.
44. Fischer, E.A. Cybersecurity Issues and Challenges: In Brief.
45. Kaloudi, N.; Li, J. The AI-Based Cyber Threat Landscape: A Survey. *ACM Comput. Surv.* **2020**, *53*, 20:1-20:34, doi:10.1145/3372823.
46. Guembe, B.; Azeta, A.; Misra, S.; Osamor, V.C.; Fernandez-Sanz, L.; Pospelova, V. The Emerging Threat of Ai-Driven Cyber Attacks: A Review. *Applied Artificial Intelligence* **2022**, *36*, 2037254, doi:10.1080/08839514.2022.2037254.
47. *Automated Machine Learning: Methods, Systems, Challenges*; Hutter, F., Kotthoff, L., Vanschoren, J., Eds.; The Springer Series on Challenges in Machine Learning; Springer International Publishing: Cham, 2019; ISBN 978-3-030-05317-8.
48. Applied Sciences | Free Full-Text | Cybersecurity Threats Based on Machine Learning-Based Offensive Technique for Password Authentication Available online: <https://www.mdpi.com/2076-3417/10/4/1286> (accessed on 27 August 2024).
49. Noury, Z.; Rezaei, M. Deep-CAPTCHA: A Deep Learning Based CAPTCHA Solver for Vulnerability Assessment 2020.
50. Chung, K.; Kalbarczyk, Z.T.; Iyer, R.K. Availability Attacks on Computing Systems through Alteration of Environmental Control: Smart Malware Approach. In Proceedings of the Proceedings of the 10th ACM/IEEE International Conference on Cyber-Physical Systems; Association for Computing Machinery: New York, NY, USA, December 16 2019; pp. 1–12.
51. Yang, J.-S.; Lee, H.-J.; min-woo, P.; Eom, J. Security Threats on National Defense ICT Based on IoT.; June 23 2015; pp. 94–98.
52. A Review of Cyber Attacks on Sensors and Perception Systems in Autonomous Vehicle. *Journal of Economy and Technology* **2023**, *1*, 242–258, doi:10.1016/j.ject.2024.01.002.
53. Kumar, A.D.; Chebrolu, K.N.R.; R, V.; KP, S. A Brief Survey on Autonomous Vehicle Possible Attacks, Exploits and Vulnerabilities 2018.
54. Saez-Perez, J.; Wang, Q.; Alcaraz-Calero, J.M.; Garcia-Rodriguez, J. Design, Implementation, and Empirical Validation of a Framework for Remote Car Driving Using a Commercial Mobile Network. *Sensors* **2023**, *23*, 1671, doi:10.3390/s23031671.
55. Dymicka, A. Cybersecurity from the Perspective of a New Technology User. *Cybersecurity and Law* **2022**, *7*, 27–36, doi:10.35467/cal/151810.

56. Alhayani, B.; Mohammed, H.J.; Chalooob, I.; Ahmed, J. Effectiveness of Artificial Intelligence Techniques against Cyber Security Risks Apply of IT Industry. *Materials Today: Proceedings* **2021**, doi:10.1016/j.matpr.2021.02.531.
57. Zhan, K. Design of Computer Network Security Defense System Based on Artificial Intelligence and Neural Network. *Journal of Intelligent & Fuzzy Systems* **2021**, 1–13, doi:10.3233/JIFS-189794.
58. Sensors | Free Full-Text | Tackling Faults in the Industry 4.0 Era—A Survey of Machine-Learning Solutions and Key Aspects Available online: <https://www.mdpi.com/1424-8220/20/1/109> (accessed on 27 August 2024).
59. Siri, Siri, in My Hand: Who's the Fairest in the Land? On the Interpretations, Illustrations, and Implications of Artificial Intelligence. *Business Horizons* **2019**, 62, 15–25, doi:10.1016/j.bushor.2018.08.004.
60. Bharadiya, J. Machine Learning in Cybersecurity: Techniques and Challenges. **2023**, 7, doi:10.47672/ejt.1486.
61. Kravets, V. Comparative Analysis of the Cybersecurity Indices and Their Applications. *Theoretical and Applied Cybersecurity* **2019**, 1, doi:10.20535/tacs.2664-2913201[1.169090.
62. Daily, J.; Gardiner, B. Cybersecurity Considerations for Heavy Vehicle Event Data Recorders. *SAE International Journal of Transportation Cybersecurity and Privacy* **2018**, 1, 113–143, doi:10.4271/11-01-02-0006.
63. Ansari, M.F.; Dash, B.; Sharma, P.; Yathiraju, N. The Impact and Limitations of Artificial Intelligence in Cybersecurity: A Literature Review. *IJARCCCE* **2022**, 11, 81–90, doi:10.17148/IJARCCCE.2022.11912.
64. Zorz, Z. The Benefits and Limitations of AI in Cybersecurity. *Help Net Security* 2018.
65. Vostoupal, J. The Cybersecurity Qualifications as the Prerequisite for the Cybersecurity Certification of Entities. *Jusletter-IT* **2021**, doi:10.38023/2029e2f5-bd30-4757-aef5-01b27ae61962.
66. Bharadiya, J. A Comprehensive Survey of Deep Learning Techniques Natural Language Processing. *European Journal of Technology* **2023**, 7, 58–66, doi:10.47672/ejt.1473.
67. Dawson, M. Cybersecurity Impacts for Artificial Intelligence Use within Industry 4.0. *Scientific Bulletin* **2021**, 26, 24–31, doi:10.2478/bsaft-2021-0003.
68. Is Artificial Intelligence a Help or Hindrance? *Network Security* **2018**, 2018, 18–19, doi:10.1016/S1353-4858(18)30046-1.
69. Cucu, C.; Bologa, R.; Gavrioloaia, G.; Cazacu, M. CURRENT TECHNOLOGIES AND TRENDS IN CYBERSECURITY AND THE IMPACT OF ARTIFICIAL INTELLIGENCE.; April 11 2019; pp. 208–214.
70. Arasada, S. Council Post: These Four Challenges In Adopting Machine Learning Can Lower Your ROI And Sabotage Success Available online: <https://www.forbes.com/councils/forbestechcouncil/2021/08/31/these-four-challenges-in->

adopting-machine-learning-can-lower-your-roi-and-sabotage-success/ (accessed on 27 August 2024).

71. The Impact of Artificial Intelligence on Organisational Cyber Security: An Outcome of a Systematic Literature Review. *Data and Information Management* **2024**, 8, 100063, doi:10.1016/j.dim.2023.100063.
72. Attaran, M.; Deb, P. Machine Learning: The New “big Thing” for Competitive Advantage. *IJKEDM* **2018**, 5, 277, doi:10.1504/IJKEDM.2018.095523.
73. Das, R.; Sandhane, R. Artificial Intelligence in Cyber Security. *J. Phys.: Conf. Ser.* **2021**, 1964, 042072, doi:10.1088/1742-6596/1964/4/042072.
74. Turing, Alan. 1950. Computing Machinery and Intelligence. *Mind* 49, 433 – 460.
75. Haris, M.A.H.B.; Yahya, M.I.H.B.; Ibrahim, M.N.B. Security and Privacy Issues in Internet of Things (IoT).
76. Seumas Miller Patrick Walsh // The NSA leaks, Edward Snowden, and the ethics and accountability of intelligence collection. - 2016. -<https://www.taylorfrancis.com/chapters/edit/10.4324/9781315743912-20/nsa-leaks-edward-snowden-ethics-accountability-intelligence-collection-seumas-miller-patrick-walsh>.
77. Art. 4 GDPR – Definitions. *General Data Protection Regulation (GDPR)*.
78. Batikas, M.; Bechtold, S.; Kretschmer, T.; Peukert, C. European Privacy Law and Global Markets for Data 2020.
79. Activism by the AI Community: Analysing Recent Achievements and Future Prospects [Online] / auth. Belfield, Haydn // ACM DIGITAL LIBRARY. - February 07, 2020. - <https://dl.acm.org/doi/10.1145/3375627.3375814>.
80. Artificial Intelligence | Office of the Secretary-General’s Envoy on Technology Available online: <https://www.un.org/techenvoy/content/artificial-intelligence> (accessed on 1 July 2024).
81. Activism by the AI Community: Analysing Recent Achievements and Future Prospects [Online] / auth. Belfield, Haydn // ACM DIGITAL LIBRARY. - February 07, 2020. - <https://dl.acm.org/doi/10.1145/3375627.3375814>.
82. A Survey on Privacy Issues and Solutions for Voice-Controlled Digital Assistants. *Pervasive and Mobile Computing* **2022**, 80, 101523, doi:10.1016/j.pmcj.2021.101523.
84. Ploug, T. The Right Not to Be Subjected to AI Profiling Based on Publicly Available Data—Privacy and the Exceptionalism of AI Profiling. *Philos. Technol.* **2023**, 36, 14, doi:10.1007/s13347-023-00616-9.85
85. Tippavajhala Sundar Srinivas Thota Goutham, Dr. M. Senthil Kumaran // Face Recognition based Smart Attendance System Using IoT. - International Research Journal of Engineering and Technology (IRJET) , Mar 2022. -<https://www.irjet.net/archives/V9/i3/IRJET-V9I333.pdf>.

86. One Month, 500,000 Face Scans: How China Is Using A.I. to Profile a Minority - The New York Times Available online: <https://www.nytimes.com/2019/04/14/technology/china-surveillance-artificial-intelligence-racial-profiling.html> (accessed on 1 July 2024).
87. Spielkamp, M. Inspecting Algorithms for Bias. *Technology Review* **2017**, *120*, 96–98.
88. Insight - Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women | Reuters Available online: <https://www.reuters.com/article/world/insight-amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK0AG/> (accessed on 1 July 2024).