Πανεπιστήμιο Δυτικής Αττικής

Σχολή Μηχανικών

Τμήμα Μηχανικών Πληροφορικής και Υπολογιστών

Πρόγραμμα Μεταπτυχιακών Σπουδών: ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ

Διπλωματική Εργασία

Τεχνητή Νοημοσύνη και Κυβερνοασφάλεια

Φοιτητής: Κίκκος Δημοσθένης
ΑΜ: cscyb21011

Επιβλέπων Καθηγητής

Δρ. Παναγιώτης Ηρ. Γιαννακόπουλος

Αιγάλεω, Μάϊος 2023

Τεχνητή Νοημοσύνη και Κυβερνοασφάλεια

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η μεταπτυχιακή διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

Α/Α	ΟΝΟΜΑ- ΕΠΩΝΥΜΟ	ΥΠΟΓΡΑΦΗ
1	**Δρ. Παναγιώτης Γιαννακόπουλος**	Panagiotis Giannakopoulos Digitally signed by Panagiotis Giannakopoulos Date: 2023.07.25 11:13:15 +03'00'
2	**Δρ. Στέφανος Γκρίτζαλης**	STEFANOS GKRITZALIS 25/07/2023 15:48
3	**Δρ. Κωνσταντίνος Μαυρομμάτης**	Konstantinos Mavrommatis Digitally signed by Konstantinos Mavrommatis Date: 2023.07.25 12:13:54 +03'00'

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος **ΚΙΚΚΟΣ ΔΗΜΟΣΘΕΝΗΣ** του **ΔΗΜΗΤΡΙΟΥ**, με αριθμό μητρώου **cscyb21011** φοιτητής του Προγραμμάτος Μεταπτυχιακών Σπουδών της Κυβερνοασφάλειας του **ΠΑΝΕΠΙΣΤΗΜΙΟΥ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ** της Σχολής ΜΗΧΑΝΙΚΩΝ του Τμήματος **ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ**,

δηλώνω υπεύθυνα ότι:

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

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

Επιθυμώ την απαγόρευση πρόσβασης στο πλήρες κείμενο της εργασίας μου, μέχρι την βαθμολόγηση της και έγκριση του επιβλέποντος καθηγητή.»

Ο Δηλών
ΚΙΚΚΟΣ ΔΗΜΟΣΘΕΝΗΣ του **ΔΗΜΗΤΡΙΟΥ**

(Υπογραφή φοιτητή)

Ευχαριστίες

Η απονομή ευχαριστιών αποτελεί ύψιστη εσωτερικής ανάγκης, για το σύνολο του Εκπαιδευτικού προσωπικού του Προγράμματος Μεταπτυχιακών Σπουδών «ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑ» του τμήματος Μηχανικών Πληροφορικής και Υπολογιστών του Πανεπιστημίου Δυτικής Αττικής και των συμφοιτητών μου, καθώς με ενέπνευσαν στην αναζήτηση της γνώσεως με γνώμονα τις αρχές και τις αξίες που διέπουν την Επιστημονική Κοινότητα στο σύνολο της. Βαρύνουσα και εξέχουσα θέση, έχει ο Επιβλέπων και Διευθυντής του Προγράμματος Μεταπτυχιακών Σπουδών «ΚΥΒΕΡΝΟΑΣΦΑΛΕΙΑΣ» ο Δρ. Καθηγητής κύριος Παναγιώτης Ηρ. Γιαννακόπουλος, τον οποίο ευχαριστώ για την καθοριστική και τεκμηριωμένη συνδρομή του, στην διεκπεραίωση της παρούσας διπλωματικής εργασίας με τίτλο Τεχνητή Νοημοσύνη και Κυβερνοασφάλεια.

Αιγάλεω, 2023

Δημοσθένης Κίκκος

Περίληψη

Η εποχή μας χαρακτηρίζεται από την έρευνα και την αξιοποίηση της Τεχνητής Νοημοσύνης. Εργαλεία τεχνητής νοημοσύνης είτε κατά μόνας είτε ως ολιστική λύση αποσκοπούν στην θωράκιση της κυβερνο-ασφάλειας και την αντιμετώπιση του κυβερνοεγκλήματος. Στη κατεύθυνση αυτή χρειάζεται ένα συγκεκριμένο πλαίσιο δράσης και ανάπτυξης, νομοθετικά εργαλεία καθώς και προτυποποίηση των διαδικασιών και της μεθοδολογίας ανάπτυξης των διάφορων εργαλείων. Υπάρχουν αρκετές κατηγορίες οπού δίνονται εμπορικές λύσεις για την αντιμετώπιση του κυβερνοεγκλήματος. Η τεχνητή νοημοσύνη από την μια δίνει λύσεις σε ζητήματα κυβερνοασφάλειας από την άλλη δημιουργεί νέα ζητήματα. Δίνει την δυνατότητα σε επιτιθέμενους για την δημιουργία νέων επιθέσεων. Τέλος τα ίδια τα συστήματα τεχνητής νοημοσύνης έχουν ζητήματα ευαλωτότητας ως προς την ασφάλεια τους.

Λέξεις – κλειδιά

Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Κυβερνοασφάλεια

Abstract

Our era is characterized by research and the exploitation of artificial intelligence. AI tools either singly or as a holistic solution are aimed at shielding cyber-security and tackling cybercrime. This will require a specific framework for action and development, legislative tools and standardization of the procedures and methodology for the development of the various tools. There are several categories where commercial solutions are provided to tackle cybercrime. AI on the one hand provides solutions to cybersecurity issues on the other hand and poses new issues. It enables attackers to create new attacks. Finally, AI systems themselves have issues of safety vulnerability.

Keywords

Artificial Intelligence, Machine Learning, Cybersecurity

Περιεχόμενα

Ευχαριστίες 5
Περίληψη 6
1.1 Αντικείμενο της πτυχιακής εργασίας 11
1.2 Μεθοδολογία 11
1.3 Δομή 12

Κεφάλαιο 2 – Κυβερνο-χώρος (Cyber-space) 13
2.1 Κυβερνο-έγκλημα (Cyber crime) 13
2.2 Κυβερνο-εγκληματίες (Cyber criminals) 15
2.3 Κυβερνοαπειλές (Cyberthreats) 15
2.3.1 Σημαντικότερες Κυβερνο-απειλές (Types of Cyber Threats) 16
2.4 Κυβερνο-ασφάλεια (Cyber security) 19
2.4.1 Πιστοποίηση (Certification) - Προτυποποίση (Stadarization) 19
2.4.2 Αρμόδιοι Φορείς 20

Κεφάλαιο 3 - Τεχνητή Νοημοσύνη (Artificial Intelligence –AI) 22
3.1 Ορισμός 22
3.2 Η αναπτυξιακή πορεία ενός συστήματος Τεχνητής Νοημοσύνης 22
3.3 Τεχνητός Νευρώνας (Artificial Neuron, AN) 26
3.4 Μηχανική Μάθηση (Machine Learning, ML) 28
3.4.1 Εποπτευόμενη Μάθηση ή Μάθηση με Επίβλεψη (Supervised Learning, SL) 28
3.4.2 Χωρίς Επίβλεψη Μάθηση (Unsupervised Learning, UL) 30
3.4.3 Ενισχυτική Μάθηση (Reinforcement Learning, RL) 32
3.5 Ελληνική Νομοθεσία [64] 33
3.6 Δεοντολογικά Ζητήματα Τεχνητής Νοημοσύνης 33

ΜΕΡΟΣ Α': Κεφάλαιο 4 – Τεχνητή Νοημοσύνη στην Κυβερνοασφάλεια 36
4.1 Κατηγορίες Τεχνητής Νοημοσύνης στην Κυβερνο-ασφάλεια 38
4.2 Ανίχνευση και Ανάλυση Κακόβουλου Λογισμικού (Malware Detection - Analysis) 42
4.3 Ανεπιθύμητη Ηλεκτρονική Αλληλογραφία [54] 46
4.4 Ανίχνευση Δικτυακών απειλών (Network threats) 48
4.5 Ανίχνευση Ψευδοτυχαίων Ονομάτων Τομέα (Detect Generated Domains, DGA) 48
4.6 Ανίχνευση χειραγωγούμενου μολυσμένου δικτύου υπολογιστών (Botnet) 49
4.7 Ανίχνευση Κακόβουλων λογισμικών στο Διαδίκτυο των Πραγμάτων. 49
4.8 Ανίχνευση Οικονομικής Απάτης (Fraud Detection) 50
4.9 Διαχείριση Ευπαθειών (Vulnerability Management) 50

4.10 Ασφάλεια 51
4.10.1 Ασφάλεια κωδικών μέσω μηχανικής μάθησης 51
4.10.2 Ασφάλιση των δεδομένων 51

ΜΕΡΟΣ Β': Κεφάλαιο 5 – Κυβερνοασφάλεια της Τεχνητής Νοημοσύνης 53
5.1 Απειλές 53
5.2 Κακόβουλη χρήση δεδομένων εισόδου 56
5.3 Δημιουργία Πλαστών - Ψευδών Ειδήσεων (DeepFakes) 57
5.4 Παρακάμπτοντας τα CAPTHAS 59
5.5 Επίθεση με μολυσμένα Δεδομένα (Data Poisoning) 60
5.6 Εμκετάλευση πορτών επικοινωνίας 61
5.7 Ευπάθειες Μηχανικής Μάθησης 61
5.8 Απειλές κατά την Μοντελοποίηση 62
5.9 Προτεινόμενοι μηχανισμοί ελέγχου και ασφάλειας των συστημάτων τεχνητής νοημοσύνης 62
Αναφορές 65

Εικόνες

Εικόνα 1 – Κυβερνο-απειλές 2022 έρευνα από την SonicWall [133] 17

Εικόνα 2 – Κυριότερες απειλές από τον ENISA Threat Landscape, ETL 2022 [27] 17

Εικόνα 3 – Απειλές 2030 [53] 19

Εικόνα 4 – Γενική Απεικόνιση Κύκλου ζωής ενός μοντέλου AI [61] 23

Εικόνα 5 – Απαραίτητα μέρη του Συστήματος Τεχνητής Νοημοσύνης [61] 24

Εικόνα 6 – Ανάλυση των μερών της Τεχνητής Νοημοσύνης [61] 25

Εικόνα 7 – Πορεία δεδομένων στην Τεχνητή Νοημοσύνη [61] 26

Εικόνα 8 – Βιολογικός Νευρώνας [52] 27

Εικόνα 9 – Τεχνητός Νευρώνας [52] 27

Εικόνα 10 – Ομαδοποίηση ομοειδών μοτίβων καιρού [68] 31

Εικόνα 11 – Ομαδοποίηση των μετεωρολογικών μοτίβων, χιόνι, βροχή, χιονόνερο, ανομβρία [68] 31

Εικόνα 12 – Απεικόνιση Σχέσεων Βαθιά Μηχανικής Μάθησης – Μηχανική Μάθηση – Τεχνητή Μάθηση [54] 32

Εικόνα 13 – Σχέσεις Εξάρτησης μεταξύ Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης [59] 33

Εικόνα 14 – Δεοντολογικά Ζητήματα Τεχνητής Νοημοσύνης [72] 35

Εικόνα 15 – Τομείς Χρήσης Τεχνητής Νοημοσύνης στην αγορά της Κυβερνοασφάλειας [69] 36

Εικόνα 16 – Αξιοποιήση εργαλείων Τεχνητής Νοημοσύνης για τον εντοπισμό απειλών και επιθέσεων ανά τομέα απασχόλησης [73] 37

Εικόνα 17 – Εργαλεία Τεχνητής Νοημοσύνης στη Κυβερνοασφάλεια ανά χώρα [73] 38

Εικόνα 18 – Χρήση ΤΝ ανά τομέα [73] 39

Εικόνα 19 – Περιοχές χρήσης ΤΝ στην Κυβερνο-ασφάλεια για Ανίχνευση – Πρόβλεψη – Αντιμετώπιση [73] 39

Εικόνα 20 – Εταιρίες Τεχνητής Νοημοσύνης για την Κυβερνοασφάλεια ανα κατηγορία [74] 40

Εικόνα 21 – 10 πιο πλητόμενες χώρες με κακόβουλο λογισμικό σε Παγκόσμιο Επίπεδο, από τη Sonicwall [133]. 43

Εικόνα 22 – Παγκόσμια Αποτύπωση Επιθέσεων από κακόβουλο λογισμικό 2021 -2022, από τη Sonicwall [133]. 43

Εικόνα 23 – Στατιστική Ανάλυση Κακόβουλων Εισβολών, από τη Sonicwall [133] 44

Εικόνα 24 – Ιεράρχιση κακόβουλων λογισμικών [134] 44

Εικόνα 25 – Ιεράρχιση διαδικασίας ανίχνευσης κακόβουλου λογισμικού [70] 45

Εικόνα 26 – Ιεράρχιση διαδικασίας ανίχνευσης κακόβουλου λογισμικού [135] 46

Εικόνα 27 – SVMs → Perceptron 47

Εικόνα 28 – Δενδροειδής Λήψη Απόφασης για το Iris dataset [54] 48

Εικόνα 29 – Παγκόσμιο Επίπεδο, από τη Sonicwall [133] 49

Εικόνα 30 – Μέθοδοι Αυθεντικοποίησης [145] 51

Εικόνα 31 – Κύριες Απειλές Συστημάτων Τεχνητής Νοημοσύνης [61]..... 54

Εικόνα 32 – Επεξήγηση Κύριων Απειλών Συστημάτων Τεχνητής Νοημοσύνης [61] 55

Εικόνα 33 – Επίθεση με βάση τα στοιχεία εισόδου ενός συστήματος Τεχνητής Νοημοσύνης [59] 56

Εικόνα 34 – Deepfakes εικόνες [59] 57

Εικόνα 35 – Διάγραμμα ροής για παραγωγή Deepfakes [152]..... 58

Εικόνα 36 – Υποστήριξη (Backpropagation) της εκπαίδευσης του διακριτή (discriminator) [152] 58

Εικόνα 37 – Υποστήριξη (Backpropagation) της εκπαίδευσης της γεννήτριας (generator) [152] ... 59

Εικόνα 38 – Υποστήριξη (Backpropagation) της εκπαίδευσης της γεννήτριας (generator) [154] ... 60

Εικόνα 39 – Σύγκριση ορθής και κακόβουλης μάθησης ενός συστήματος [59]. 61

Πίνακες

Πίνακας 1 – Ενδεικτική Απεικόνιση Κατηγοριών κυβερνοεγκληματιών από το INTEL TAL [26] 15

Πίνακας 2 – Ενδεικτικές Κυβερνοαπειλές [26] 16

Εισαγωγή

Η ραγδαία τεχνολογική εξέλιξη της τεχνητής νοημοσύνης με ταχύτητα φωτός, επηρεάζει ένα ευρύ φάσμα σε υπάρχουσες εργασιακές δομές και τομείς, συμπεριλαμβανομένου και της κυβερνο-ασφάλειας. Ο κλάδος της κυβερνο-ασφάλειας εξελίσσεται με ταχείς ρυθμούς αποσκοπώντας στην αντιμετώπιση του κυβερνο-εγκλήματος, την εύρεση ευπαθειών, τον εντοπισμό και την αποτροπή επιθέσεων που πλήττουν δεδομένα και πληροφορίες. Οι κυβερνο-εγκληματίες πολλαπλασιάζονται εφευρίσκοντας νέους τρόπους και μεθοδολογίες επιθέσεων ή προβαίνουν στην επέκταση των ήδη υπαρχόντων επιθέσεων με τη χρήση νέων τεχνολογιών όπως της τεχνητής νοημοσύνης.

Αρκετές διαδικασίες και μεθοδολογίες στο κλάδο της ασφάλειας στο κυβερνο-χώρο, που μέχρι τώρα πραγματοποιούνται με βάση κάποια λογισμικά εργαλεία και τις δυνατότητες και δεξιότητες εξειδικευμένων μηχανικών στο αντίστοιχο κλάδο της πληροφορικής, αυτοματοποιούνται και διεκπεραιώνονται με την χρήση εργαλείων και λογισμικών τεχνητής νοημοσύνης. Σε παγκόσμιο επίπεδο δραστηριοποιούνται ήδη εμπορικές εταιρείες σε υψηλό επίπεδο στην παροχή υπηρεσιών κυβερνο-ασφάλειας που βασίζονται στην τεχνητή νοημοσύνη. Τα εργαλεία που βασίζονται στην τεχνητή νοημοσύνη δίνουν λύσεις σε αρκετά ζητήματα που άπτονται με την κυβερνο-ασφάλεια αλλά εγκυμονούν κινδύνους που αφορούν την τεχνολογία της τεχνητής νοημοσύνης. Οπότε, η μια πλευρά του νομίσματος αφορά την ευνοϊκότερη αντιμετώπιση ζητημάτων κυβερνοασφάλειας στον κυβερνο-χώρο με την χρήση εργαλείων τεχνητής νοημοσύνης, από την άλλη όμως υπάρχουν τεχνικά και δεοντολογικά ζητήματα που αφορούν την χρήση της.

1.1 Αντικείμενο της πτυχιακής εργασίας

Βασική επιδίωξη της πτυχιακής εργασίας είναι η ανάδειξη των γενικών πτυχών της τεχνολογίας της τεχνητής νοημοσύνης και της κυβερνο-ασφάλειας. Παρουσιάζεται ένα γενικό πλαίσιο, των απειλών, των αρμόδιων φορέων, της κείμενης νομοθεσίας που αφορούν την κυβερνο-ασφάλεια στον κυβερνοχώρο, καθώς γίνεται και μια γενική επισκόπηση σχετικά με τα διαθέσιμα εμπορικά σύγχρονα εργαλεία - λογισμικών τεχνητής νοημοσύνης, δεοντολογικών ζητημάτων αλλά και το κατά πόσο ασφαλή είναι τα ίδια τα συστήματα τεχνητής νοημοσύνης.

1.2 Μεθοδολογία

Από τον Σεπτέμβριο του 2022 άρχισε η βιβλιογραφική συλλογή πληροφοριών που αφορούν την Τεχνητή Νοημοσύνη από το πρίσμα της Κυβερνο-ασφάλειας. Στην συνέχεια τα βιβλιογραφικά ευρήματα αξιολογήθηκαν με τον μέγιστο δυνατό τρόπο, με σκοπό την αξιοποίηση αυτών για την κατανόηση επί της αρχής της αλληλοσυνύπαρξης της «Τεχνητής Νοημοσύνης και της Κυβερνο-αφάλειας». Η επιλογή μελέτης του συγκεκριμένου ζητήματος, πραγματοποιήθηκε ύστερα από ενδελεχή και εμπεριστατωμένη σκέψη, καθότι αποτέλεσε το εφαλτήριο για την εξερεύνηση νέων τεχνολογιών που σηματοδοτούν την τεκτονική αλλαγή χρήσης εργαλείων για την θωράκιση της κυβερνο-ασφάλειας. Είναι βαρύνουσας σημασίας το συγκεκριμένο θέμα της μεταπτυχιακής διπλωματικής εργασίας, τόσο λόγω της ακαδημαϊκής βαρύτητας που έχει αλλά και λόγο των νέων επαγγελματικών προοπτικών που αναδεικνύονται από την χρήση νέων δομών τεχνολογίας. Στην πορεία και περί τα μέσα Δεκεμβρίου, καθορίστηκε το πλαίσιο για την εκπόνηση της εν λόγω εργασίας, τα αντίστοιχα γνωστικά αντικείμενα που πρόκειται να αναδειχθούν, τα οποία είναι η συνεισφορά της τεχνητής Νοημοσύνης για την επίτευξη της κυβερνο-ασφάλειας και αντίστροφα, καθώς και τα εμπορικά λογισμικά και εργαλεία που έχουν αναδειχθεί.

Σκοπός της εργασίας είναι η βιβλιογραφική περιγραφική ανάδειξη των προηγούμενων γνωστικών αντικειμένων. Διερευνήθηκαν με συστηματικό και μεθοδικό τρόπο όλες οι

βιβλιογραφικές πηγές, οι οποίες ταξινομήθηκαν, αξιολογήθηκαν και μελετήθηκαν. Στο διάστημα από το Δεκέμβριο έως και το Μάϊο άρχισε και τελείωσε η συγγραφή της εργασίας. Χρησιμοποιήθηκαν πληθώρα βιβλιογραφικών πηγών για την συλλογή και την εκπόνηση της εν λόγω εργασίας. Διερευνήθηκαν διαφορετικές πηγές για την εύρεση πληροφοριών που σχετίζονται με το θέμα της εργασίας. Το σύνολο των δεδομένων, των πληροφοριών που βρέθηκαν αναφέρονται αντίστοιχα στο τέλος της εργασίας, ώστε να είναι εφικτή η αναζήτηση τους.

1.3 Δομή

Η παρούσα εργασία έχει ως βασική επιδίωξη την ανάδειξη των λύσεων της τεχνητής νοημοσύνη για την αποτελεσματικότερη κυβερνοασφάλεια, τις εμπορικές και διαθέσιμες λύσεις λογισμικών προς αυτή την κατεύθυνση, τα δεοντολογικά ζητήματα που υφίστανται, την κείμενη νομοθεσία, πρωτυποποίηση διαδικασιών μεθοδολογιών αλλά και τους κινδύνους που έχουν τα συστήματα τεχνητής νοημοσύνης.

Συγκεκριμένα, στο Κεφάλαιο 1 έχουμε την εισαγωγή για το θέμα, προβάλλοντας την επιλογή τους θέματος και την μεθοδολογία ανάπτυξης της.

Στο Κεφάλαιο 2 – Κυβερνο-χώρος (Cyberspace), παρουσιάζονται επιγραμματικά έννοιες που αφορούν τον κυβερνοχώρο (cyberspace), το κυβερνο-έγκλημα (cybercrime), τις κατηγορίες κυβερνο-εγκληματιών (cybercriminals), τις διάφορες κυβερνο-απειλές (cyberthreats), τα είδη των κυβερνο-απειλών (types of cyber threats), τις κυβερνο-επιθέσεις (cyberthreats) και τις επιπτώσεις αυτών, την κυβερνο-ασφάλεια (cybersecurity), την πιστοποίηση της κυβερνοασφάλειας (cybersecurity certification), την προτυποποίηση αυτής (stadarization) καθώς και τους αρμόδιους φορείς.

Στο Κεφάλαιο 3 – Τεχνητή Νοημοσύνη, γίνεται μια γενική αναφορά για την τεχνολογία της τεχνητής νοημοσύνης (artificial intelligence), του τεχνητού νευρώνα (artificial neuron), της μηχανικής μάθησης (machine learning), της εποπτευόμενης μάθησης ή μάθηση με επίβλεψη (Supervised Learning, SL), την χωρίς επίβλεψη μάθηση (Unsupervised Learning), της ενισχυτικής μάθησης της Ελληνικής Νομοθεσίας και των δεοντολογικών ζητημάτων που αφορούν την τεχνητή Νοημοσύνη. Στη συνέχεια γίνεται ο διαχωρισμός της εργασίας σε δυο (2) μέρη με αντίστοιχα Κεφάλαια.

Στο ΜΕΡΟΣ Α: Κεφάλαιο 4 – Τεχνητή Νοημοσύνη στην Κυβερνο-ασφάλεια, όπου αναπτύχθηκαν ενδεικτικά κάποιες εμπορικές λύσεις και οι κατηγορίες τεχνητής νοημοσύνης στην κυβερνο-ασφάλεια. Στη συνέχεια παρουσιάστηκε η χρησιμότητα της τεχνητής νοημοσύνης στην ανίχνευση και την ανάλυση κακόβουλου λογισμικού (Malware Detection – Analysis), της ανεπιθύμητης ηλεκτρονικής αλληλογραφίας, των ψευδοτυχαίων Ονομάτων Τομέα (Detect Generated Domains Detection), στον εντοπισμό μολυσμένου δικτύου υπολογιστών (Botnet Detection), των κακόβουλων λογισμικών στο διαδίκτυο των πραγμάτων και στη διερεύνηση οικονομικής απάτης (Fraud Detection). Επίσης επισημάνθηκε η αξιοποίηση της τεχνητής νοημοσύνης για τη διαχείριση ευπαθειών και γενικότερα για την ασφάλεια, τόσο για τους κωδικούς πρόσβασης όσο και για τα δεδομένα.

Στο ΜΕΡΟΣ Β: Κεφάλαιο 5 – Κυβερνοασφάλεια της Τεχνητής Νοημοσύνης. Περιέχονται σε αυτό το κεφάλαιο, οι απειλές, η κακόβουλη χρήση δεδομένων εισόδου, η δημιουργία πλαστών – ψευδών ειδήσεων. Επιπρόσθετα επισημαίνονται επιθέσεις που παρακάμπτουν μεθόδους ταυτοποίησης captchas, επιθέσεων με μολυσμένα δεδομένα (Data Poisoning), εκμετάλλευσης ανοικτών πορτών επικοινωνίας, ευπάθειες μηχανικής μάθησης καθώς και απειλές που δημιουργούνται κατά την μοντελοποίηση.

Κεφάλαιο 2 – Κυβερνο-χώρος (Cyber-space)

Ο σύγχρονος τεχνολογικά καθημερινός τρόπος ζωής μας είναι άρρηκτα συνδεδεμένος με την χρήση των τεχνολογιών πληροφορικής και επικοινωνιών με κύρια έκφραση του το διαδίκτυο. Σε κάθε έκφανση της καθημερινότητας, όπως για την επικοινωνία, την αναζήτηση πληροφοριών, την διεκπεραίωση οικονομικών συναλλαγών, την εργασία, καθώς και για πλήθος άλλων δραστηριοτήτων, όπου ο μέσος άνθρωπος, πολίτης χρησιμοποιεί τις υπηρεσίες και τις εφαρμογές που χρειάζεται μέσω του διαδικτύου. Το διαδίκτυο είναι ένα μέρος του λεγόμενου κυβερνο-χώρου.

Κατά κοινή ομολογία, ως κυβερνο-χώρος ορίζεται ο χώρος εντός του οποίου περιλαμβάνονται το σύνολο των φυσικών, των λογικών και των οργανωτικών, κοινωνικών διασυνδέσεων, ως απόρροια της διασύνδεσης, των τηλεπικοινωνιακών δικτύων, των δικτύων υπολογιστών, γενικότερα των πληροφοριακών συστημάτων και υπηρεσιών τόσο μεταξύ τους όσο και με τα φυσικά συστήματα, είτε είναι άμεσα συνδεδεμένα με το διαδίκτυο είτε έμμεσα [1] [2].

2.1 Κυβερνο-έγκλημα (Cyber crime)

Στον κυβερνο-χώρο, λαμβάνουν χώρα, με δόλο ή χωρίς, επιθέσεις, οι οποίες εμπεριέχουν νέους τρόπους και μεθοδολογίες, πλήττοντας, είτε την ιδιοκτησία, είτε την ζωή και την προσωπικότητα του θύματος. Σε παγκόσμιο επίπεδο σημείο αναφοράς αποτελεί η Σύμβαση της Βουδαπέστης, η οποία υπεγράφη το 2001, από τα Κράτη Μέλη του Συμβουλίου της Ευρώπης και κυρώθηκε στο Ελληνικό Δίκαιο με τον νόμο 4411/2016 ενσωματώνοντας τις αρχές και διατάξεις της Ευρωπαϊκής Οδηγίας 2013/40/ΕΕ [3] [4].

Ως κυβερνο-έγκλημα ορίζεται η τέλεση της αντικειμενικής υπόστασης των νομοθετημένων εγκλημάτων, όπου κατά την πράξη χρησιμοποιήθηκε, τουλάχιστον μια ηλεκτρονική συσκευή, είτε ως κύριο μέσο (κατά αποκλειστικότητα), είτε ως βοηθητικό και απαραιτήτως υπήρχε και η χρήση του Διαδίκτυου για την τέλεση του [4]. Σε περίπτωση που απουσιάζει από το προηγούμενο ορισμό ο παράγοντας Διαδίκτυο, τότε αναφέρεται ως έγκλημα το οποίο πραγματώθηκε με την αξιοποίηση ηλεκτρονικού υπολογιστή (Computer Crime) [25]. Συνεπώς η γενική κατηγορία είναι το Ηλεκτρονικό Έγκλημα (Electronic Crime) η οποία περιλαμβάνει το Κυβερνό-έγκλημα (Cyber Crime) και το έγκλημα με χρήση ηλεκτρονικού υπολογιστή (Computer Crime).

Ως προς το περιεχόμενο τους κατατάσσονται στις ακόλουθες κατηγορίες, τα εγκλήματα [4]:

- Εναντίον της Εμπιστευτικότητας, της Ακεραιότητας και της Διαθεσιμότητας.
  Για παράδειγμα:
  - Παράνομη πρόσβαση (Hacking).
  - Υποκλοπή δεδομένων με την χρήση τεχνικών μέσων.
  - Δημιουργία παρεμβολών σε δεδομένα.
  - Δημιουργία παρεμβολών σε συστήματα
  - Κακόβουλη χρήση συσκευών.
- Με υπολογιστή
  - Απάτη (Αλλοίωση δεδομένων, διαγραφή, καταστροφή, η άλλη δόλια παρέμβαση)
- Συνυφασμένα με το περιεχόμενο.
  - Παιδική Πορνογραφία
  - Συκοφαντική δυσφήμιση
  - Εκδικητική Πορνογραφία
- Πνευματικών, συγγραφικών και συγγενικών δικαιωμάτων.

Στο Ελληνικό Ποινικό Δίκαιο μπορούν να αναφερθούν επιγραμματικά τα ακόλουθα νομοθετημένα αδικήματα του Ποινικού Κώδικα :

- Άρθρο 292^{A}– Εγκλήματα κατά της ασφάλειας των τηλεφωνικών επικοινωνιών [5].
- Άρθρο 292^{B}– Παρακώλυση λειτουργίας πληροφοριακών συστημάτων [6].
- Άρθρο $292^{Γ}$ – Παραγωγή, πώληση, προμήθεια για χρήση, εισαγωγή, κατοχή, διανομή ή διακίνηση τόσο συσκευών όσο και λογισμικών σχεδιασμένων για την διάπραξη των εγκλημάτων του 292^{B} ή για την παράνομη απόκτηση δεδομένων, κωδικών πρόσβασης, ή άλλων παρεμφερή δεδομένων από ένα πληροφοριακό σύστημα [7].
- Άρθρο $292^{Δ}$ – Προσβολές του απορρήτου των τηλεπικοινωνιών του κοινού [8].
- Άρθρο 292^{E} – Παρακώλυση των τηλεπικοινωνιών [9].
- Άρθρο 346 – Εκδικητική πορνογραφία [10]
- Άρθρο 348 – Πορνογραφία ανηλίκων [11].
- Άρθρο 348^{B} – Προσέλκυση παιδιών για γενετήσιους λόγους [12].
- Άρθρο 363 – Συκοφαντική δυσφήμιση [13]
- Άρθρο 370 – Παραβίαση του απορρήτου των επιστολών. Στο συγκεκριμένο άρθρο υπάρχει αναφορά στην έννοια του εγγράφου [14]. Σύμφωνα με το άρθρο 13 του Ποινικού Κώδικα, ως έγγραφο ορίζεται κάθε γραπτό όπου είναι χρήσιμο για την νομική διαδικασίας όπως και κάθε μέσο το οποίο χρησιμοποιείται για αυτό, όπως ηλεκτρονικός υπολογιστής, περιφερειακές μνήμες και οποιοδήποτε ηλεκτρονικό ή άλλο μέσο στο οποίο εγγράφεται πληροφορία που είναι χρήσιμη στην απόδειξη γεγονότων κατά την νομική έννοια [15].
- Άρθρο 370^{A} - Παραβίαση του απορρήτου τηλεφωνικής επικοινωνίας και προφορικής συνομιλίας [16].
- Άρθρο 370^{B} – Παράνομη πρόσβαση σε σύστημα πληροφοριών ή σε δεδομένα [17]
- Άρθρο $370^{Γ}$ – Παράνομη πρόσβαση σε πληροφοριακό σύστημα [18]
- Άρθρο $370^{Δ}$ – Παράνομη αντιγραφή ή χρήση λογισμικών προγραμμάτων [19]
- Άρθρο 370^{E} – Παραβίαση μη δημόσιων διαβιβάσεων δεδομένων ή ηλεκτρομαγνητικών εκπομπών [20]
- Άρθρο $370^{ΣΤ}$ – Απαγόρευση διακίνησης λογισμικών, συσκευών παρακολούθησης και άλλων δεδομένων [21]
- Άρθρο 381^{A} – Φθορά ηλεκτρονικών δεδομένων [22]
- Άρθρο 381^{B} – Παράνομη παραγωγή, κατοχή, πώληση συσκευών ή λογισμικών για την διάπραξη αδικημάτων του 381^{A} καθώς και για την απόκτηση πρόσβασης στο σύνολο ή μέρος πληροφοριακού συστήματος [23]
- Άρθρο 386^{A} – Απάτη με υπολογιστή [24]

Τα παραπάνω αδικήματα αποτελούν μια ενδεικτική απεικόνιση του Νομοθετικού Πλαισίου που ισχύει στην Χώρα μας για το ηλεκτρονικό έγκλημα και τελούνται από φυσικά πρόσωπα τα οποία μπορούν να ταυτοποιηθούν και να τους αποδοθούν οι αντίστοιχες κατηγορίες. Τέλος, όσα από τα παραπάνω πληρούν δυο προϋποθέσεις, δηλαδή την χρήση ηλεκτρονικού μέσου με επεξεργαστική ισχύ και την χρήση του διαδικτύου, αποτελούν κυβερνο-εγκλήματα.

2.2 Κυβερνο-εγκληματίες (Cyber criminals)

Ως κυβερνο-εγκληματίες χαρακτηρίζονται, οι φυσικοί ή ηθικοί αυτουργοί που τελούν την αντικειμενική υπόσταση των αδικημάτων, εγκλημάτων τα οποία έχει προβλέψει ο νομοθέτης και τεκμαίρονται ως κυβερνο-εγκλήματα. Το πληροφοριακό υλικό, υλικό ή άυλο καθώς και οτιδήποτε είναι σε ηλεκτρονική ψηφιακή μορφή, αναγκαίο για την αποτύπωση και την διαπίστωση τέλεσης της αξιόποινης πράξης θεωρείται ψηφιακό πειστήριο.

Η εταιρία της Intel (Threat Agent Library , TAL) προέβη στον διαχωρισμό και την κατηγοριοποίηση των κυβερνο-εγκληματιών με βάση το τρόπο δράσης που ακολουθούν συνήθως οι επιτιθέμενοι. Η διάκριση των κυβερνο-εγκληματιών αποτυπώνεται ενδεικτικά στον ακόλουθο πίνακα [26 , 28]:

Κυβερνο-εγκληματίες	**Συνηθισμένη Μεθοδολογία Δράσης**
Ακτιβιστής (Civil Activist)	Διαταραχή ή διακοπή των διαθέσιμων επιχειρηματικών υπηρεσιών
Αποθησαυριστής Δεδομένων (Data Miner)	Κλοπή των περιουσιακών στοιχείων ή άλλων αξιοποιήσιμων στοιχείων , δεδομένων και πληροφοριών.
Ανταγωνιστής (Competitor)	Κλοπή των περιουσιακών στοιχείων ή / και άλλων κρίσιμων επιχειρηματικών / προσωπικών πληροφοριών που θα έχουν αντίκτυπο και κόστος στο κάτοχο ή στην επιχείρηση.
Κλέφτης (Thief)	
Κυβερνο-Βανδαλιστής (Cyber-Vandal)	Προβαίνει σε διατάραξη της εύρυθμης λειτουργίας ή, σε διακοπή λειτουργίας του δικτύου, υπολογιστών με αντίστοιχη εξαπόληση επιθέσεων χρησιμοποιώντας είτε κακόβουλο λογισμικό (malware) είτε συνδυαστικά προβαίνοντας σε επιθέσεις που αφορούν τις υπηρεσίες του διαδικτύου (web hijacking)
Κυβερνο-Πολεμιστής (Government Cyberwarrior)	
Εγκληματίας (Mobster)	Παράνομη ιδιοποίησης, δηλαδή κλοπή των περιουσιακών στοιχείων, δεδομένων ή/και επιχειρηματικών/προσωπικών πληροφοριών, με άσκηση βίας.
Δημιουργός Εντυπώσεων (Sensationalist)	Δημόσιες ανακοινώσεις που αφορούν ζητήματα του δημόσιου βίου, τρόπου ζωής, των δημοσίων σχέσεων ή/και με κλοπή επιχειρηματικών πληροφοριών.

Πίνακας 1 – Ενδεικτική Απεικόνιση Κατηγοριών κυβερνοεγκληματιών από το INTEL TAL [26]

Οι κυβερνο-εγκληματίες εκμεταλλεύονται τις σύγχρονες τεχνολογίες για την τέλεση υφιστάμενων αδικημάτων ή για τη δημιουργία νέων. Για παράδειγμα οι επιτιθέμενοι προβαίνουν σε κακόβουλη χρήση των νέων τεχνολογιών, όπως της τεχνητής νοημοσύνης για την δημιουργία ψευδών ειδήσεων και παραπληροφόρησης (AI – enabled disinformation and deepfakes).

Οι αρμόδιοι φορείς που ασχολούνται με την ασφάλεια πληροφοριακών συστημάτων και πληροφοριών είτε του ιδιωτικού είτε του δημόσιου τομέα πραγματοποιούν σειρά ενεργειών με σκοπό την αποτύπωση σε ετήσια βάση των καταγεγραμμένων και γνωστών κυβερνοαπειλών.

2.3 Κυβερνοαπειλές (Cyberthreats)

Ο Ευρωπαϊκός Οργανισμός για την Κυβερνοασφάλεια (European Union Agency for Cybersecurity) δημοσιεύει τις κυβερνοαπειλές που καταγράφηκαν από τα αντίστοιχα συστήματα

σε τακτικά χρονικά διαστήματα [27]. Οι κυβερνοαπειλές θα μπορούσαν να ταξινομηθούν στον ακόλουθο πίνακα [26 , 29]:

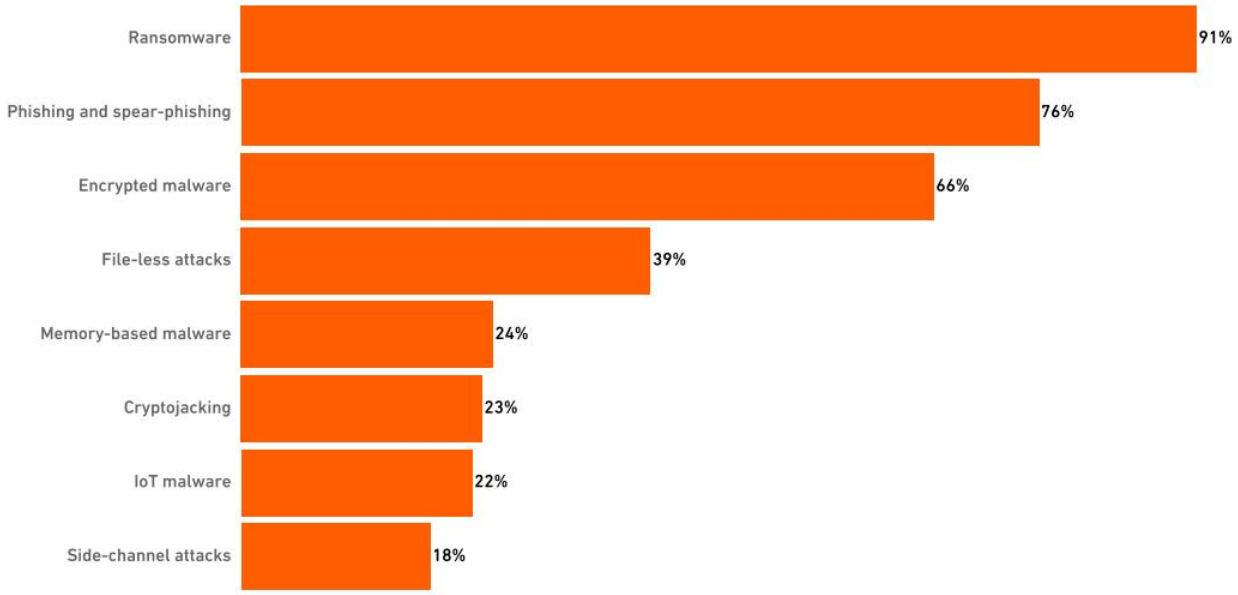
Απειλές	Περιγραφή
Σκόπιμες	Φυσικές: οικονομική εξαπάτηση (fraud), κλοπή, παράνομη ιδιοποίηση (ψηφιακών μέσων δεδομένων, πληροφοριών), διαρροή δεδομένων, εκβιασμός, διαφθορά, τρομοκρατικές επιθέσεις.
	Λαθρακρόαση (eaves dropping), υποκλοπή (interception), υφαρπαγή (hijacking), επιθέσεις με κινούμενο όχημα (war driving), υποκλοπή εκπομπής πληροφορίας, τεχνικές αναγνώρισης δικτύου και συλλογής πληροφοριών.
	Κακόβουλη δραστηριότητα: Κλοπή ταυτότητας (identity theft), αποστολή ανεπιθύμητης ηλεκτρονικής αλληλογραφίας (spam), επιθέσεις άρνησης υπηρεσίας (denial of service), κακόβουλο λογισμικό (malware), τεχνικές κοινωνικής μάθησης (social engineering), εκμετάλλευση αδυναμιών στη διαχείριση της πληροφορίας, πλαστογραφία ψηφιακών πιστοποιητικών, χειραγώγηση υλισμικού/λογισμικού/πληροφορίας (π.χ παραποίηση υπηρεσιών Domain Name Service – DNS), αλλοίωση λογισμικού, μη εξουσιοδοτημένες ενέργειες, botnets (δίκτυα μολυσμένων υπολογιστών με κακόβουλου λογισμικό), cryptojacking, ransomware.
	Νομοθετικό πλαίσιο: Παραβίαση νομοθετικού ή/και κανονιστικού πλαισίου, μη τήρηση όρων επιχειρηματικών συμβολαίων, κατάχρηση δεδομένων προσωπικού χαρακτήρα, κακόβουλες ενέργειες κατά δικαστικών αρχών και διαδικασιών.
Μη σκοπιμες	Διαρροή πληροφοριών
	Κακές πρακτικές διαχείριση συστημάτων
	Χρήση δεδομένων εισόδου από αναξιόπιστες πηγές
	Μη ηθελημένη αλλοίωση δεδομένων
	Παραλείψεις κατά τη διαδικασία σχεδιασμού, ανάλυσης και υλοποίησης ενός πληροφοριακού συστήματος
	Ζημιές από προμηθευτές/συνεργάτες
	Απώλεια της Εμπιστευτικότητας και Ακεραιότητας πληροφοριακών συστημάτων.
Περιβαλλοντικές	Διακοπές λειτουργίας κεντρικών συστημάτων (outages)
	Αστοχίες/δυσλειτουργίες

Πίνακας 2 – Ενδεικτικές Κυβερνοαπειλές [26]

2.3.1 Σημαντικότερες Κυβερνο-απειλές (Types of Cyber Threats)

Στο σημείο αυτό γίνεται μια προσπάθεια αποτύπωσης των κυβερνο-απειλών που απασχόλησαν τη διεθνή κοινότητα το προηγούμενο έτος. Οι σημαντικότερες απειλές οι οποίες καταγράφηκαν αφορούσαν κατά κύριο λόγο απάτες μέσω διαδικτύου. Κύριο μέλημα των επιτιθέμενων σε αυτή τη περίπτωσης είναι η αξία που κατέχουν τα δεδομένα τα οποία έχουν ιδιαίτερο βάρος για τον κάτοχο τους. Στόχος των επιτιθέμενων κατά κύριο λόγο είναι το οικονομικό όφελος για αυτούς, κλειδώνοντας ουσιαστικά υπό ομηρία συστήματα, με δεδομένα και

πληροφορίες χρήσιμες για επιχειρήσεις ή οργανισμούς, έχοντας ως ενέχυρο κάποιο οικονομικό τίμημα. Στην ακόλουθη εικόνα αποτυπώνονται ενδεικτικά κάποια ποσοστά απειλών σε παγκόσμιο επίπεδο.



Εικόνα 1 – Κυβερνο-απειλές 2022 έρευνα από την SonicWall [133]

Εικόνα 2 – Κυριότερες απειλές από τον ENISA Threat Landscape, ETL 2022 [27]

Σύμφωνα με την ετήσια αναφορά του 2022 που συντάσει ο Ευρωπαϊκός Οργανισμός ασφάλειας (ENISA Threat Landscape), οι κυριότερες καταγεγραμμένες απειλές αφορούν [27]:

- **Λογισμικό Λύτρων (Ransomware):** Πρόκειται για μια από τις δημοφιλέστερες απειλές. Θέτει σε ομηρία τα διάφορα λειτουργικά πληροφοριακά συστήματα καθώς και τα κρίσιμα δεδομένα που είναι αποθηκευμένα σε αυτά. Οι επιτιθέμενοι έχοντας υπό ομηρία κρίσιμους υπολογιστικούς πόρους που αποτελούν σημαντικά περιουσιακά στοιχεία για τους κατόχους τους επιδιώκουν την είσπραξη υψηλών οικονομικών λύτρων ώστε να προβούν σε επιστροφή της διαθεσιμότητας τους.
- **Κακόβουλο Λογισμικό (Malicious Software, Malware)**: Είναι μια απειλή που περιλαμβάνει οποιοδήποτε υλισμικό ή λογισμικό, όπου πρόκειται να εκτελέσει μη

εξουσιοδοτημένη διαδικασία έχοντας αρνητικό αντίκτυπο στην εμπιστευτικότητα, την ακεραιότητα και την διαθεσιμότητα ενός συστήματος. Μερικά παραδείγματα κακόβουλων λογισμικών είναι ιοί (virus), σκουλήκια (worms), δούρειοι ίπποι (trojan horses), λογισμικά υποκλοπών (spywares, adwares) ή οποιαδήποτε άλλη οντότητα η οποία βασίζεται σε κώδικα που σκοπό έχει να πλήξει την λειτουργία ενός υπολογιστικού συστήματος ή τα δεδομένα αυτού.

- **Κοινωνική Μηχανική (Social Engineering)**: Η κοινωνική μηχανική περιέχει ένα σύνολο δραστηριοτήτων και ενεργειών, όπου προσβλέπουν στην εκμετάλλευση των ανθρώπινων αδυναμιών, ώστε ο επιτιθέμενος να αποκτήσει πρόσβαση σε κρίσιμες και σημαντικές πληροφορίες για τον κάτοχο τους. Μια από την συχνότερη και συνηθέστερη μεθοδολογία που ακολουθείται είναι η επίμονη παρότρυνση που γίνεται προς τους χρήστες να ανοίξουν αρχεία ή ιστοτόπους όπου φαινομενικά είναι επίσημων φορέων, τραπεζικών ιδρυμάτων, άρχων αλλά δεν είναι έχοντας ως αποτέλεσμα την υποκλοπή κωδικών και άλλων στοιχείων προσωπικών, διαπιστευτηρίων που μπορούν να θίξουν την κοινωνική και οικονομική ζωή του θύματος. Οι πιο διαδεδομένοι μέθοδοι που χρησιμοποιούνται σε αυτή την κατηγορία είναι για παράδειγμα, το ηλεκτρονικό ψάρεμα (phishing), μέσω ηλεκτρονικής αλληλογραφίας επιχειρήσεων (business e-mail compromise, BEC), απάτη (fraud) και πλαστοπροσωπία (impersonation and counterfeit).
- **Διακινδύνευση Δεδομένων (Threats against data)**: Στόχος των απειλών αυτών είναι η μη εξουσιοδοτημένη πρόσβαση στις πηγές δεδομένων. Οι απειλές αυτές αποτελούν τον θεμέλιο λίθο για άλλες απειλές, όπως η άρνηση εξυπηρέτησης με λογισμικό λύτρων (ransomware denial of service, rdos) και η κατανεμημένη άρνηση εξυπηρέτησης (distributed denial of service, ddos).
- **Διακινδύνευση της Διαθεσιμότητας – Επιθέσεις Άρνησης Υπηρεσίας (Threats against availability: Denial of Service) :** Οι απειλές εναντίον της διαθεσιμότητας είναι αρκετές, η πιο διαδεδομένη είναι η άρνηση εξυπηρέτησης συστημάτων (DDoS). Στόχος αυτής της επίθεσης είναι η εξάντληση της χρήσης των υπολογιστικών πόρων, ώστε να μην υπάρχει η δυνατότητα από τους χρήστες να χρησιμοποιήσουν την υπηρεσία αυτή ή το δίκτυο μιας υποδομής.
- **Διακινδύνευση της Διαθεσιμότητας – Απειλές του Διαδικτύου (Internet Threats)** : Η αναγκαία και καθημερινή χρήση του διαδικτύου από κάθε άνθρωπο της κοινωνίας αποτελεί στόχο για τους επιτιθέμενους. Απειλές του Διαδικτύου που έχουν αντίκτυπο στην διαθεσιμότητα είναι για παράδειγμα το highjacking BGP (Border Gateway Protocol). Επιπρόσθετα οι επιθέσεις μέσω του διαδικτύου έχουν σκόπο την κακόβουλη εκμετάλλευση των φυλλομετρητών των στόχων[29].
- **Σκοπούμενη Παραπληροφόρηση – Μη Σκοπούμενη Παραπληροφόρηση (Disinformation – Misinformation)** : Ένα φαινόμενο που αυξάνεται εκθετικά είναι οι μεθοδευμένες εκστρατείες παραπληροφόρησης μέσω των κοινωνικών δικτύων και άλλων δικτυακών κοινωνικών μέσων, για τον επηρεασμό κοινωνικών ομάδων έχοντας διαφορετικά κίνητρα κάθε φορά.
- **Επιθέσεις στην Εφοδιαστική Αλυσίδα (Supply-chain attacks)** : Πρόκειται για επιθέσεις στην εφοδιαστική αλυσίδα, όπου οι στόχοι της επίθεσης είναι τόσο ο προμηθευτής όσο και ο πελάτης.

Ο Ευρωπαϊκός Οργανισμός ENISA παρουσίασε τις δέκα σημαντικότερες απειλές που θα μας απασχολήσουν έως το 2030, όπως φαίνεται στην ακόλουθη εικόνα.

Εικόνα 3 – Απειλές 2030 [53]

Οι Κυβερνο-επιθέσεις απασχολούν τόσο σε Παγκόσμιο και Ευρωπαϊκό επίπεδο όσο και σε Εθνικό. Για αυτό τον λόγο η Εθνική Αρχή Κυβερνοασφάλειας του Υπουργείου Ψηφιακής Διακυβέρνησης έχει δημοσιοποιήσει την «Εθνική Στρατηγική Κυβερνοασφάλειας 2020-2025», εναρμονίζοντας και την σχετική Ευρωπαϊκή Οδηγία.

2.4 Κυβερνο-ασφάλεια (Cyber security)

Με τον όρο Κυβενο-ασφάλεια (cyber security), νοείται το σύνολο των αναγκαίων προπαρασκευαστικών ενεργειών για την ασφάλεια της πληροφορίας, δηλαδή την προστασία των επιμέρους στοιχείων της, την εμπιστευτικότητα (Confidentiality), την Ακεραιότητα (Integrity) και την Διαθεσιμότητα (Availability) [3 , 30]. Γεγονός είναι ότι η Κυβερνο-ασφάλεια θα πρέπει να εντάσσεται μέσα σε ένα θεσμικό πλαίσιο λειτουργίας. Για τον λογο αυτό σε Ευρωπαϊκο και στην συνέχεια σε Εθνικό Επίπεδο ενσωματώντας σχετικές οδηγίες, που αφορούν τον τρόπο πιστοποίησης ως προς την Κυβερνο-ασφάλεια.

2.4.1 Πιστοποίηση (Certification) - Προτυποποίση (Stadarization)

Η πιστοποίηση αφορά προϊόντα, υπηρεσίες και διαδικασίες προβλέπονται από τον Ευρωπαϊκό Κανονισμό 881/2019 και γίνεται ιδιαίτερη αναφορά (στο άρθρο 24 του NIS2), και στην Ευρωπαϊκή Οδηγία 2555/2022 [30, 31, 32]. Με βάση το Ευρωπαϊκό πλαίσιο πιστοποίησης δίνεται η δυνατότητα λειτουργίας ενός μηχανισμού με βάση τον οποίο θεσπίζονται Ευρωπαϊκά Συστήματα Πιστοποίησης ως πρός την Κυβερνοασφάλεια. Στην Χώρα μας σύμφωνα με το άρθρο 15 του Νόμου 6961/2022, η Γενική Διεύθυνση Κυβερνοασφάλειας της Γενικής Γραμματείας Τηλεπικοινωνιών και Ταχυδρομείων του Υπουργείου Ψηφιακής Διακυβέρνησης οριζεται ως Εθνική Αρχή Πιστοποίησης της Κυβερνοασφάλειας [49]. Με την διαδικασία αυτή υπάρχει έμπρακτη και επίσημη πιστοποίηση, ότι τα προϊόντα, οι υπηρεσίες και οι διαδικασίες της Τεχνολογίας Πληροφοριών και Επικοινωνωνίων είναι σύμφωνα με κάποιες αρχές κοινά αποδεκτές

ως προς την ασφάλεια τους [30]. Η πιστοποίηση και η προτυποποίηση, είναι δράσεις που αφορούν την πρόληψη δίνοντας εγγυήσεις, ως προς την ασφάλεια των πληροφοριακών αγαθών και την θωράκιση ενός πληροφοριακού συστήματος [48].

Στην περίπτωση που μια κυβερνο-επίθεση βρει στόχο, καταστρέφοντας και πλήττοντας την ιδιοκτησία, τότε θα πρέπει να δοθεί η δυνατότητα αποζημίωσης του κόστους που επήλθε από αντίστοιχη ασφαλιστική εταιρεία του χώρου. Στις μέρες μας είναι όλο και μεγαλύτερη η ανάγκη ασφάλισης των αγαθών, πληροφοριών και συστημάτων, που είναι κρίσιμα και αναγκαία ανάλογα με τις συνθήκες και τις προτεραιότητες του κάθενα. Είναι ευνόητο, ότι για να δεχθεί μια ασφαλιστική εταιρεία να ασφαλίσει μια υποδομή, θα πρέπει να έχουν ικανοποιηθεί οι προϋποθέσεις και τα αναγκαία μέτρα που αφορούν τα πρότυπα και τις πιστοποιήσεις κυβερνοασφάλειας.

Συνεπώς θα πρέπει να υφίστανται και να λειτουργούν φορείς κυβερνο-ασφάλειας στην κατεύθυνση της αντιμετώπισης των κυβερνο-εγκλημάτων και της δημιουργίας προτύπων, διαπιστεύσεων, πιστοποιήσεων, πολιτικών ασφάλειας, που θα πρέπει να ακολουθούνται ως γενική αρχή από όλους.

2.4.2 Αρμόδιοι Φορείς

Σε παγκόσμιο, Ευρωπαϊκό και Εθνικό επίπεδο δημιουργούνται και ιδρύονται αρχές και αρμόδιοι φορείς που έχουν ως αποστολή τους την κυβερνο-ασφάλεια. Επιδίωξη όλων η θωράκιση των πληροφοριακών δομών των οργανισμών και επιχειρήσεων έναντι ενδεχόμενων απειλών και επιθέσεων μέσω του κυβερνοχώρου. Ιδιαίτερη έμφαση στη θωράκιση και την προστασία των προσωπικών και ευαίσθητων δεδομένων των χρηστών και η ομαλή εξέλιξη των πληροφοριακών συστημάτων με έμφαση στην ασφάλεια αυτών από το σχεδιασμό τους. Στο σημείο αυτό αξίζει να αναφερθούν επιγραμματικά κάποιους από τους φορείς που είναι αρμόδιοι για την Κυβερνοασφάλεια:

- Γενική Διεύθυνση Κυβερνοασφάλειας – Εθνική Αρχή Κυβερνοασφάλειας (National Cybesecurity Authority) [33].
- Εθνική Αρχή Αντιμετώπισης Ηλεκτρονικών Επιθέσεων – Εθνικό CERT (Ε.Υ.Π.) [34].
- Διεύθυνση Κυβερνοάμυνας (Υπουργείο Εθνικής Άμυνας) (ΓΕ.ΕΘ.Α./Ε5/ΔΙ.ΚΥΒ.) – CSIRT [35]
- Αρχή Διασφάλισης Απορρήτου Επικοινωνιών (Α.Δ.Α.Ε.) [36].
- Αρχή Προστασίας Δεδομένων Προσωπικού Χαρακτήρα (Α.Π.Δ.Π.Χ.) [37].
- Εθνική Επιτροπή Τηλεπικοινωνιών και Ταχυδρομίων (Ε.Ε.Τ.Τ.) [38].
- Διεύθυνση Δίωξης Ηλεκτρονικού Εγκλήματος (ΔΙ.Δ.Η.Ε. ΕΛ.ΑΣ.) [39].
- Διεύθυνση Εγκληματολογικών Ερευνών / Τμήμα Εξέταση Ψηφιακών Πειστηριών [40].
- Κέντρο Μελετών Ασφάλειας (ΚΕ.ΜΕ.Α.) [41].

- Επιτροπή Εποπτείας και Ελέγχου Παιγνίων [42].
- ENISA [43]
- Ευρωπαϊκό Κέντρο για το Έγκλημα στον Κυβερνοχώρο European Cybercrime Center (EC3) [44]
- Joint Cybercrime Taskforce (J-CAT) [45]
- Intellectual Property Crime Coordination Centre (IPC3) [46]
- European Financial and Economic Crime Centre (EFECC) [47]

Συμπερασματικά οι αρμόδιοι φορείς έχουν ως αποστολή την ενημέρωση και την συνεισφορά τους τόσο στην πρόληψη των κυβερνο-επιθέσεων όσο και στην αντιμετώπιση τους. Το ζήτημα του κυβερνο-εγκλήματος είναι παγκόσμιο και η γεωμετρικά εξελισσόμενη τεχνολογική πρόοδος γεννά νέου είδους επιθέσεων ή αντιμετώπιση αυτών με νέα τεχνολογικά εργαλεία.

Στις μέρες μας η τεχνητή νοημοσύνη προβληματίζει το παγκόσμιο γίγνεσθαι, ως προς την προοπτική για την ευρεία χρήσης της, την αξιοποίηση της για την αντιμετώπιση υπαρχόντων κινδύνων και παθογενειών αλλά και των κινδύνων που ελλοχεύουν από αυτήν, όπως και των δεοντολογικών και ηθικών ζητημάτων σχετικά με τις ελευθερίες και τα δικαιώματα του ανθρώπου. Για το λόγο αυτό είναι αναγκαίο και απαραίτητο να τεθούν βάσεις και αρχές σε παγκόσμιο επίπεδο που θα ορίζουν με σαφήνεια τόσο τους τομείς που θα επιτρέπεται η χρήση εργαλείων τεχνητής νοημοσύνης, όσο και ο τρόπος και η διαδικασία που θα ακολουθείται για την ανάπτυξη τους με βάση διεθνή πρότυπα ασφάλειας και ποιότητας. Οι αρχές αυτές που θα αφορούν την Τεχνητή Νοημοσύνη θα πρέπει να εκφράζονται από κάποιο παγκόσμιο φορέα, όπου θα συνυπολογίσει όλες τις παραμέτρους που αφορούν τις ελευθερίες του ανθρώπου, την εργασιακή απασχόληση, την ανάπτυξη του, τα θεμελιώδη πανανθρώπινα δικαιώματα του, τα ηθικά και τα δεοντολογικά ζητήματα καθώς και τις αρχές ασφάλειας κατά την ανάπτυξη συστημάτων ώστε να είναι ασφαλή τα εργαλεία και τα συστήματα τεχνητής νοημοσύνης.

Κεφάλαιο 3 - Τεχνητή Νοημοσύνη (Artificial Intelligence –AI)

Η απαρχή της τεχνολογίας που ακούει στον όρο Τεχνητή Νοημοσύνη, καταγράφεται το έτος 1956 όπου αποτελεί ορόσημο για την γέννηση της από τον Marvin Minsky, τον Claude Shannong και άλλους [50]. Αν και εξήντα επτά (67) ετών, η Τεχνητή Νοημοσύνη έγινε ιδιαίτερα δημοφιλής τα τελευταία πέντε (5) χρόνια, με την ανάπτυξη εφαρμογών και εργαλείων δίνοντας επαναστατικές λύσεις για τα έως τώρα υφιστάμενα εργαλεία. Ίσως άργησε να φανεί η αξία της, καθώς σημαντικός παράγοντας για την ανάπτυξη συστημάτων Τεχνητής Νοημοσύνης είναι η κατοχή πληθώρα πρωτογενών δεδομένων. Γεγονός όπου αναδείχθηκε αρκετά την τελευταία δεκαετία με την ευρεία διαχείριση μεγάλου όγκου δεδομένων, είτε από ανοικτές πηγές είτε από άλλες πηγές. Συνεπώς υπάρχουν περισσότερα δεδομένα και η ανάγκη για ταχύτερη και αυτοματοποιημένη συλλογή, ανάλυση για την αντιμετώπιση και την διεκπεραίωση ζητημάτων και λήψης αποφάσεων έφερε στο προσκήνιο την αξιοποίηση και την ανάπτυξη συστημάτων τεχνητής νοημοσύνης.

3.1 Ορισμός

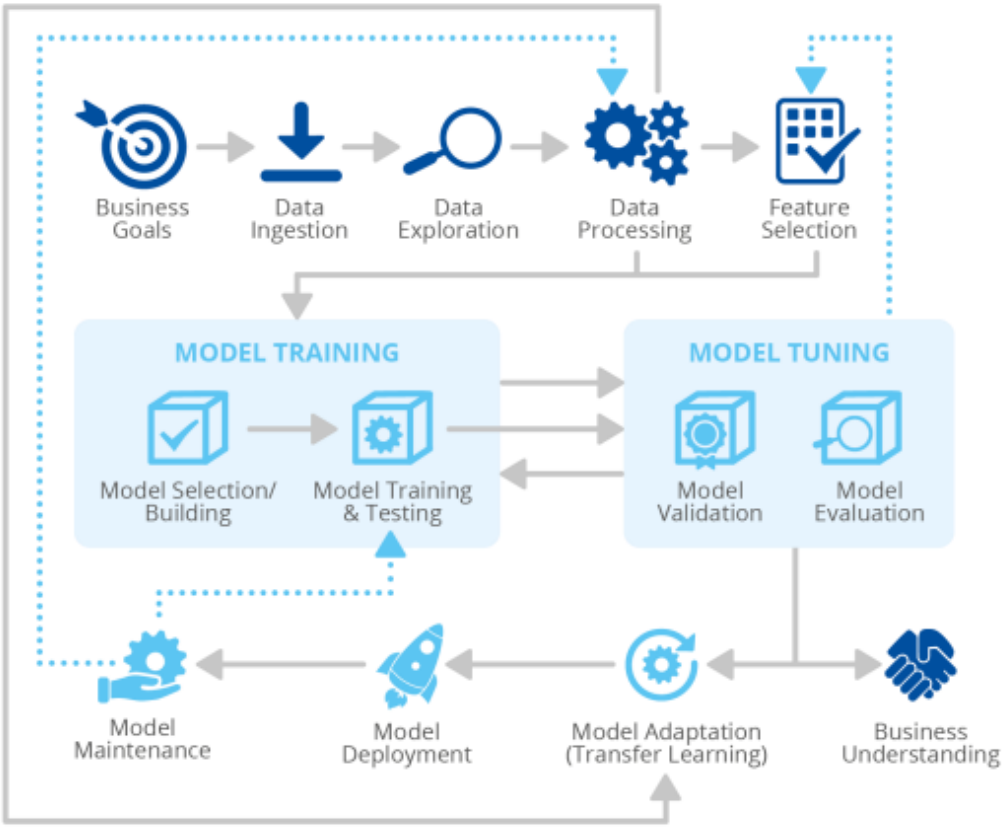
Έως και σήμερα αποδίδονται διαφορετικοί, εμπεριστατωμένοι και τεκμηριωμένοι ορισμοί που αφορούν την Τεχνητή Νοημοσύνη. Έχουν διατυπωθεί αρκετοί ορισμοί, αρχικά από τους Barr και Feigenbaum, από τον Haugeland, τον Winston, των Rich και Knight και άλλων [50]. Ένας προτεινόμενος και ολοκληρωμένος, πλήρης ορισμός θα μπορούσε να είναι ο ακόλουθος:

«*Ο τομέας της Επιστήμης των Υπολογιστών που μελετά την σχεδίαση για την υλοποίηση των αντίστοιχων υπολογιστικών συστημάτων, όπου έχουν την δυνατότητα μίμησης των ανθρώπινων γνωστικών ικανοτήτων ορίζεται ως Τεχνητή Νοημοσύνη*» [50].

Όπως στον άνθρωπο έτσι και στα υπολογιστικά συστήματα τεχνητής νοημοσύνης είναι απαραίτητη η ύπαρξη ενός νευρωνικού δικτύου και τεχνητών νευρώνων που προσομοιάζουν τους βιολογικούς νευρώνες. Η ανάπτυξη ενός συστήματος Τεχνητής Νοημοσύνης είναι μια ακολουθία καλά σχεδιασμένων σταδιών μέχρι την επίτευξη της δημιουργίας του συστήματος.

3.2 Η αναπτυξιακή πορεία ενός συστήματος Τεχνητής Νοημοσύνης

Όπως συμβαίνει σε κάθε πληροφοριακό σύστημα, λογισμικό, λειτουργικό σύστημα υπάρχει μια συγκεκριμένη πορεία για την ανάπτυξη του, το ίδιο επί της αρχής ισχύει και σε ένα σύστημα τεχνητής νοημοσύνης. Θα πρέπει κατά το σχεδιασμό να καθοριστούν τα δεδομένα, τα ζητούμενα, ο στόχος υλοποίησης και ο τρόπος που θα υλοποιηθεί με βάση πάντα την ασφάλεια από τον σχεδιασμό του. Η καταγραφή των απαραίτητων στοιχείων που απαρτίζουν ένα σύστημα τεχνητής νοημοσύνης είναι άρρηκτα και αμφίδρομα συνδεδεμένα μεταξύ τους επιδιώκοντας την ανάπτυξη, την εξέλιξη και την λειτουργία σε βάθος χρόνου με βάση πεπερασμένες αρχικές υπολογιστικές δυνατότητες. Για την ανάπτυξη ενός συστήματος Τεχνητής Νοημοσύνης έχει διατυπωθεί από τον ENISA η παρακάτω γενική απεικόνιση του κύκλου ζωής του, καθώς και τα αντίστοιχα «αγαθά», μέρη (assets) του [61].



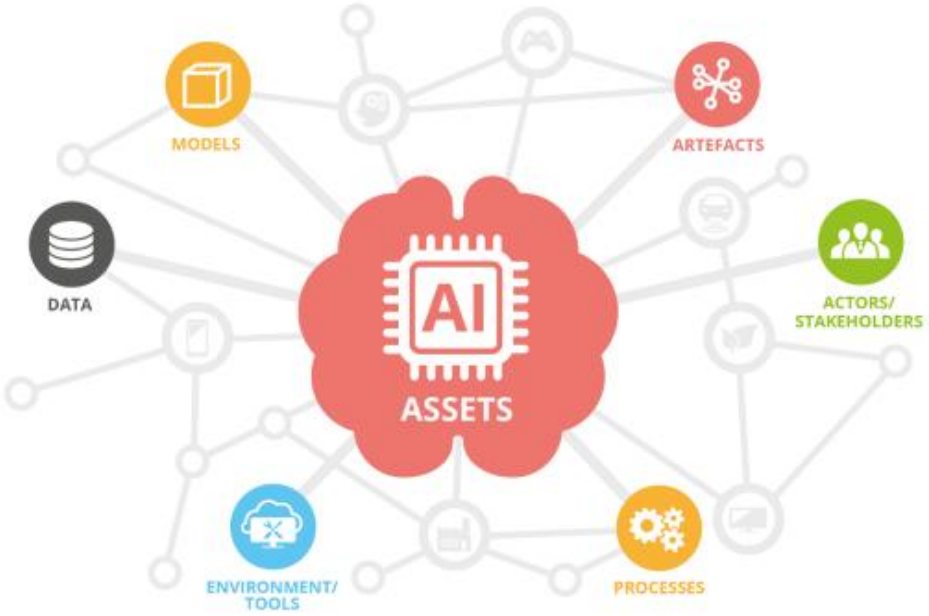
Εικόνα 4 – Γενική Απεικόνιση Κύκλου ζωής ενός μοντέλου AI [61]

Ως σημείο μηδέν για ένα σύστημα τεχνητής νοημοσύνης θεωρείτε ο καθορισμός των στόχων και των επιδιώξεων που έχουν τεθεί. Δηλαδή η αποστολή που θα έχει το μοντέλο του συστήματος τεχνητής νοημοσύνης. Στη συνέχεια έχουμε το κυριότερο και ουσιαστικότερο ζήτημα των δεδομένων, που θα αναζητηθούν, θα βρεθούν, θα χρησιμοποιηθούν και θα επεξεργαστούν από το μοντέλο του συστήματος. Ο παράγοντας δεδομένα είναι σπουδαίος, τόσο κατά τον σχεδιασμό όσο και στην διαδικασία της ανάπτυξης του μοντέλου του συστήματος της τεχνητής νοημοσύνης. Είναι παράγοντας όπου επηρεάζει σε σημαντικό βαθμό την ασφάλεια του συστήματος τεχνητής νοημοσύνης, για αυτό θεωρείτε κρίσιμος για την ανάπτυξη του.

Τα δεδομένα είναι αυτά τα οποία θα εκπαιδεύσουν το μοντέλο εκπαίδευσης του συστήματος τεχνητής νοημοσύνης. Ύστερα από μια σειρά ενεργειών, αποσφαλμάτωσης λαθών και εύρεση των διαδικασιών που εξυπηρετούν τους αρχικούς στόχους και σχεδιασμό, υιοθετείτε το μοντέλο του συστήματος τεχνητής νοημοσύνης το οποίο ανατροφοδοτείτε και συνεχώς εκπαιδεύετε. Η συνεχής εκπαίδευση και ανάπτυξη αυτού εξαρτάται από το είδος μηχανικής μάθησης που έχει επιλεγεί.

Ένα σύστημα τεχνητής νοημοσύνης είναι ένα εξελισσόμενο και συνεχώς τροφοδοτούμενο και ανατροφοδοτούμενο σύστημα με δεδομένα όπου επιτελεί ένα ανώτερο σκοπό για επιτυχή διεκπεραίωση συγκεκριμένης διαδικασίας για την εξαγωγή συμπερασμάτων με ασφάλεια. Στις περιπτώσεις που υπάρχει κάποια αστοχία, στο αποτέλεσμα το ίδιο το σύστημα θα πρέπει να είναι σχεδιασμένο με τέτοιο τρόπο όπου θα του επιτρέπει την αυτοβελτίωση μέσω της αλγοριθμικής λήψης αποφάσεων και της μηχανικής μάθησης. Προφανώς η δυναμική αυτή διαδικασία προϋποθέτει μια συνεχή επίβλεψη και θωράκιση των ενεργειών με τρόπο που δεν θα αποκλίνει από τον αρχικό σχεδιασμό του συστήματος.

Σε ένα σύστημα τεχνητής νοημοσύνης υπάρχουν κάποια συστατικά μέρη, αγαθά, που το απαρτίζουν έχοντας ζωτική σημασία για την λειτουργία του.



Εικόνα 5 – Απαραίτητα μέρη του Συστήματος Τεχνητής Νοημοσύνης [61]

Βαρύνουσας σημασίας σε ένα σύστημα τεχνητής νοημοσύνης είναι:

- τα δεδομένα (data), όπως τα ακατέργαστα δεδομένα (raw data), τα ανοικτά ή δημόσια δεδομένα (public data set), τα δεδομένα εκπαίδευσης (testing data), προ-επεξεργάσιμα δεδομένα (pre-processed data set), δεδομένα αξιολόγησης (evaluation data), επικύρωσης δεδομένων (data validation), αντίγραφα δεδομένων με μικρές διαφοροποιήσεις (augmented data) και δεδομένα με επισήμανση (labelled data).
- το μοντέλο του συστήματος τεχνητής νοημοσύνης (models), οι αλγόριθμοι, αλγόριθμοι προεπεξεργασίας δεδομένων (data pre-processing algorithms), αλγόριθμοι εκπαίδευσης (training algorithms), οι παράμετροι που αφορούν το μοντέλο (model parameters), παράμετροι που αφορούν την εκπαίδευση του μοντέλου (training parameters), τα εκπαιδευμένα μοντέλα (trained models) και τα συντονισμένα μοντέλα (tuned models).
- τα τεχνουργήματα (artefacts), δηλαδή η διαδικασία περιγραφής που ακολουθείτε για την απόρροια του αποτελέσματος που εξάγεται από την διαδικασία της μάθησης. Ενδεικτικά περιλαμβάνονται, οι λίστες ελέγχου πρόσβασης (access control lists), η περίπτωση χρήσης (use case), το επιχειρηματικό μοντέλο έχοντας συμπεριλάβει την αξία που έχει (value proposition and business model), τα δεδομένα διακυβέρνησης (data governance), την αποτύπωση των δεδομένων με αντίστοιχα γραφήματα (data display and plots), την αρχιτεκτονική των μοντέλων (model architecture), το σχεδιαστικό μοντέλο υλικού εξοπλισμού (model hardware design), και τα διάφορα αξιοποιήσιμα σχήματα δεδομένων και μεταδεδομένων (data and metadata schemata)
- οι εμπλεκόμενοι – ενδιαφερόμενοι (actors – stakeholders), όπως ο κάτοχος των δεδομένων (data owner), ο επιστήμονας των δεδομένων – ο προγραμματιστής της τεχνητής νοημοσύνης (data scientist – ai developer), οι μηχανικοί των δεδομένων (data engineers), οι τελικοί χρήστες (end users), παροχή των δεδομένων (data provide – broker), πάροχος υπολογιστικού νέφους (cloud provider), πάροχος του μοντέλου (model provider), και διαθέιμές υπηρεσίες – χρήστες μοντέλου (service consumer – model users).
- διεργασίες (processes), για παράδειγμα ενέργειες που αφορούν τα δεδομένα, αναζήτηση (data exploration), προεπεξεργασια, αποθήκευση δεδομένων (data storage) , συλλογή δεδομένων (data collection), καθώς και ενέργειες που αφορούν την μοντελοποίηση

- και τα διαθέσιμα εργαλεία του γενικότερου περιβάλλοντος του συστήματος (environment - tools), όπως είναι τα δίκτυα επικοινωνιών (communication networks), τα πρωτόκολλα επικοινωνιών (communication protocols), το υπολογιστικό νέφος (cloud), οι πλατφόρμες συλλογής δεδομένων (data ingestion platforms), οι πλατφόρμες εξερεύνησης δεδομένων (data exploration platforms), το σύστημα διαχείρισης βάσης δεδομένων (data base management system dbms), το σύστημα κατανεμημένων αρχείων (distributed file system), κάθε αξιοποιήσιμο υπολογιστικό σύστημα (computational platforms), το περιβάλλον για ολοκληρωμένη ανάπτυξη (integrated development environment), τα εργαλεία παρακολούθησης (monitoring tools), το λειτουργικό σύστημα – λογισμικό (operating system – software), τεχνικές βελτιστοποίησης (optimization techniques), πλατφόρμες μηχανικής μάθησης (machine learning platforms), επεξεργαστές (processors) και τα εργαλεία οπτικοποίησης (visualization tools).

Εικόνα 6 – Ανάλυση των μερών της Τεχνητής Νοημοσύνης [61]

Τα παραπάνω συστατικά μέρη της τεχνητής νοημοσύνης αποτελούν μια ενδεικτική απεικόνιση, καθώς η τεχνολογία συνεχώς αναπτύσσεται και εξελίσσεται. Σε μια προσπάθεια αποτύπωσης των σταδίων για την ανάπτυξη ενός συστήματος τεχνητής νοημοσύνης αξίζει να σημειωθεί, ότι τα δύο (2) κυριότερα στάδια επί της αρχής που μπορούν να διατυπωθούν για την δημιουργία ενός συστήματος τεχνητής νοημοσύνης είναι τα ακόλουθα [58, 62, 63]:

1ο Σχεδιασμός

Σε όλα τα έργα πληροφορικής ο σχεδιασμός αποτελεί το θεμέλιο λίθο του πληροφοριακού συστήματος, που πρόκειται να δημιουργηθεί. Για αυτό στο στάδιο αυτό τίθενται όλα τα δεδομένα επί τάπητος, οι οικονομικοί και υπολογιστικοί πόροι, τα ηθικά, τα νομικά και τα τεχνικά ζητούμενα, οι στόχοι, οι προσδοκίες, η επίτευξη των στόχων σε συγκεκριμένες χρονικές στιγμές ανά φάση, δηλαδή η εν γένη φύση υλοποίησης του προσδοκώμενου αποτελέσματος, από την λειτουργία ενός συστήματος Τεχνητής Νοημοσύνης λαμβάνοντας υπόψιν το σύνολο των παραμέτρων που αφορούν την Ασφάλεια του Συστήματος αυτού από τον σχεδιασμό. Συνεπώς, κατά τον σχεδιασμό είναι απαραίτητη η σαφής διατύπωση του προβλήματος, ορίζοντας το μοντέλο με βάση, το οποίο θα επιλυθεί αξιοποιώντας και αναλύοντας τα αντίστοιχα δεδομένα [59].

Εικόνα 7 – Πορεία δεδομένων στην Τεχνητή Νοημοσύνη [61]

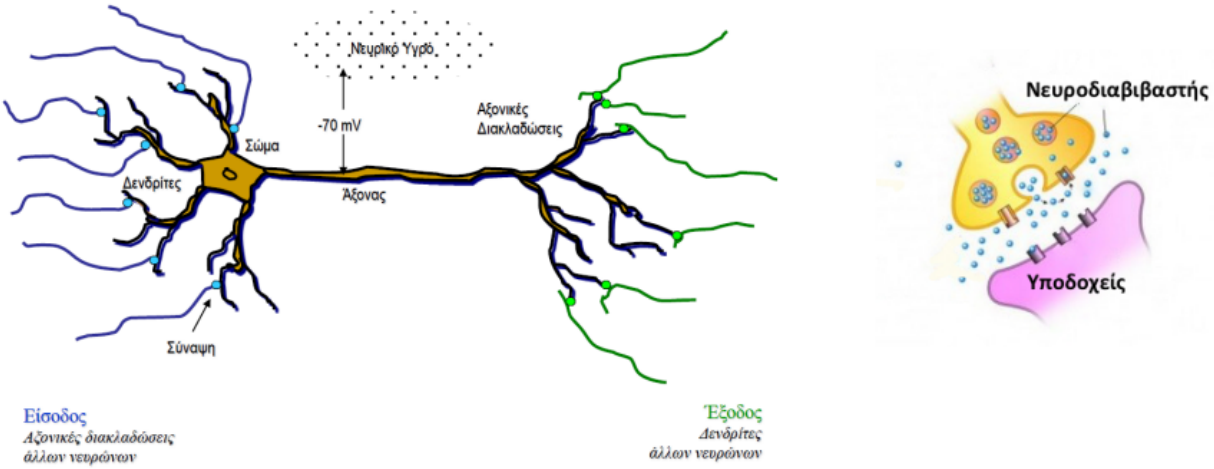
2ο Ανάπτυξη

Στη συνέχεια μετά το σχεδιασμό είναι η έμπρακτη ανάπτυξη του συστήματος. Στο στάδιο αυτό γίνονται όλες οι απαραίτητες ενέργειες για την εκμάθηση του μοντέλου που θα χρησιμοποιηθεί, ελέγχοντας την καταλληλότητα του, με την αντίστοιχη τεκμηρίωση του, ώστε να πραγματοποιηθεί η υλοποίηση του. Επιπρόσθετα πραγματοποιούνται όλα τα απαραίτητα επιμορφωτικά σεμινάρια για την εκπαίδευση των διαχειριστών και των χρηστών του εν λόγο συστήματος. Σημαντικό σημείο η εποπτεία καλής λειτουργίας του συστήματος προβαίνοντας στις απαραίτητες αναβαθμίσεις.

Τα προαναφερθέντα στάδια, αποτελούν την βασική προσέγγιση για την διερεύνηση δημιουργίας ενός συστήματος τεχνητής νοημοσύνης. Υπάρχουν και επιμέρους καθοριστικά στοιχεία εντός αυτών των σταδιών που είναι αναγκαία για την κατανόηση λειτουργίας και ανάπτυξης του επιθυμητού συστήματος τεχνητής νοημοσύνης.

3.3 Τεχνητός Νευρώνας (Artificial Neuron, AN)

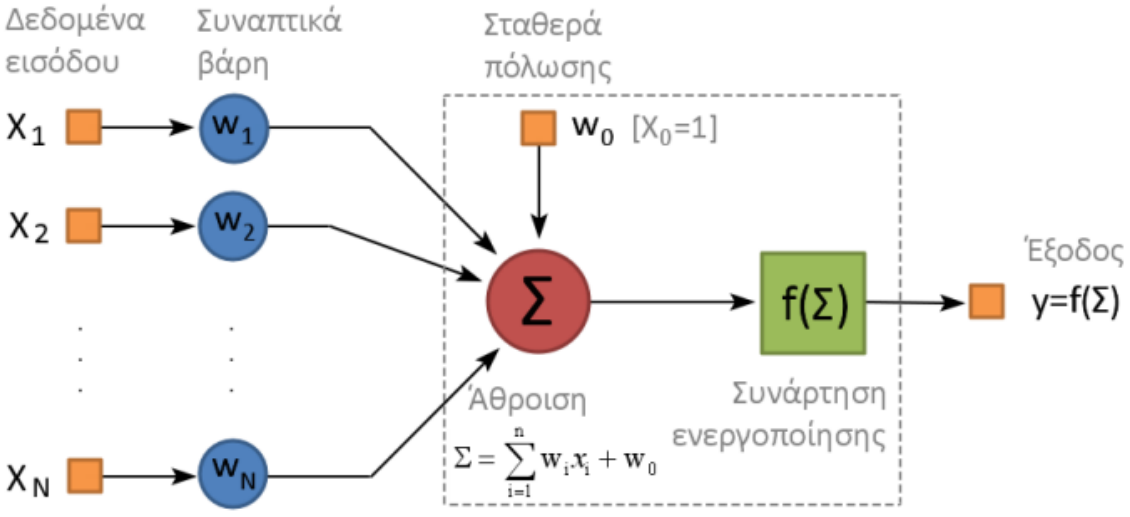
Ένα από τα κυριότερα χαρακτηριστικά του ανθρώπινου είδους η ικανότητα που έχει στην απομνημόνευση καταστάσεων, στον προβληματισμό και την επίλυση ζητημάτων, δηλαδή στο φυσικό χαρακτηριστικό της μαθήσεως. Βασικό στοιχείο για την ικανότητα αυτή αποτελεί η δομική μονάδα του εγκεφάλου ο νευρώνας, σε συνδυασμό με το σώμα και τον άξονα. Με την σειρά του το σώμα εμπεριέχει το πυρήνα και τους δενδρίτες, όπου ως αγωγοί μεταβιβάζουν τα ηλεκτρικά σήματα στον νευρώνα. Σε κάθε δενδρίτη υπάρχει μια σύνδεση η οποία καλείτε ως σύναψη [50 , 51 ,52].



Εικόνα 8 – Βιολογικός Νευρώνας [52]

Κατά αντιστοιχία τα διάφορα στοιχεία, που συγκροτούν τον βιολογικό νευρώνα μπορούν να αντικατοπτριστούν και στον τεχνητό νευρώνα (artificial neuron). Όπως φαίνεται στην Εικόνα 3, ο τεχνητός νευρώνας έχει ως είσοδο κάποια σήματα $x_1,x_2..x_n$, τα οποία διαμορφώνονται από την τιμή του βάρους wi, (ο ρόλος της σύναψης). Επίσης ο τεχνητός νευρώνας δέχεται ως είσοδο και ένα επιπλέον σήμα του ονομάζεται πόλωση (bias), η οποία είναι μια σταθερή τιμή $x_0=1$, μια βηματική συνάρτηση δηλαδή, όπου προκύπτει από το βάρος w_0 [50].

Ο κύριο τεχνητός νευρώνας διαθέτει έναν αθροιστή S, όπου υπολογίζει το άθροισμα των επιμέρους w_i και x_i, δηλαδή $\boldsymbol{S = \sum_1^n x_i \cdot w_i + 1}$ και έχει ακόμη μια συνάρτηση ενεργοποίησης, η οποία δέχεται την τιμή του S και δίνει την έξοδο y_i [50 , 51]. Να συμπληρωθεί ότι υπάρχουν δύο τύποι για την συνάρτηση ενεργοποίησης, η συνάρτηση κατωφλίου και η σημοειδής συνάρτηση [51].



Εικόνα 9 – Τεχνητός Νευρώνας [52]

Κάνοντας μια ιστορική αναδρομή παρατηρείται, ότι το αντίληπτρο (Perceptron) του Frank Rosenblatt, όπου το 1958 αποτέλεσε το πρώτο απλό νευρωνικό δίκτυο ενός γραμμικού ταξινομητή, το οποίο επιδέχεται αλγοριθμικής ερμηνείας [51].

Με τις πρώτες προσπάθειες δημιουργίας συστημάτων Τεχνητής Νοημοσύνης, προέκυψαν δυσκολίες, ως προς την επίλυση πολύπλοκων προβλημάτων, τα οποία δεν βασίζονται σε μια στατική και πεπατημένη διαδικασία λύσης, αλλά στην ικανότητα για μια δυναμική αλληλουχία γνώσης, εξαγωγής προτύπων, διαδικασιών και μοντέλων με βάση τα πηγαία δεδομένα. Η ύπαρξη αυτής της ικανότητας είναι η μηχανική μάθηση [50].

3.4 Μηχανική Μάθηση (Machine Learning, ML)

Ως μηχανική μάθηση, νοείται η τεχνολογική ικανότητα ενός υπολογιστικού συστήματος εκτέλεσης μια λειτουργίας, δημιουργώντας το αντίστοιχο μοντέλο ή πρότυπο, όπου είναι απόρροια της ατομικής μάθησης του από τα δεδομένα, δίχως να υπάρχει εκ των προτέρων συγκεκριμένη προγραμματιστική ακολουθία [50, 51, 57]. Το υπολογιστικό σύστημα εκπαιδεύεται για μια νέα λειτουργία αξιοποιώντας την περιγραφή ενός μοντέλου [50]. Ο όρος μοντέλο στην μηχανική μάθηση είναι η μαθηματική σχέση που συνδέει τα δεδομένα με την δυνατότητα που δίνεται στο σύστημα για την πραγματοποίηση προβλέψεων. Οι αλγόριθμοι μηχανικής μάθησης είναι απαραίτητοι για την τεχνολογία της τεχνητής νοημοσύνης. Η χρήση των αλγορίθμων μηχανικής μάθησης εξαρτώνται από την μέθοδο που θα χρησιμοποιηθεί [50, 51, 57].

Τα βασικότερα είδη μηχανικής μάθησης είναι τρία, η μάθηση με επίβλεψη ή επιτηρούμενη (Supervised learning), η μάθηση χωρίς επίβλεψη (Unsupervised learning) και η ενισχυτική μάθηση (Reinforcement Learning) [50].

3.4.1 Εποπτευόμενη Μάθηση ή Μάθηση με Επίβλεψη (Supervised Learning, SL)

Βασικός σκοπός στην Μηχανική Μάθηση είναι η εκμάθηση των μοντέλων μάθησης έχοντας, ως πρωτογενές υλικό, συγκεκριμένα δεδομένα. Ύστερα από την τροφοδότηση των δεδομένων σε συνδυασμό με τα ορθά επιδιωκόμενα αποτελέσματα, πραγματοποιείται η μοντελοποίηση αυτής της διαδικασίας εκφρασμένη με μαθηματικές σχέσεις [68]. Με τον τρόπο αυτό γίνεται μια σύγκριση μεταξύ δεδομένων και αποτελεσμάτων, αξιοποιώντας βασικές λειτουργίες της εποπτευόμενης μάθησης την κατηγοριοποίηση ή ταξινόμηση (classification) και την παρεμβολή (regression), με στόχο την πρόβλεψη ενός γεγονότος [58]. Οπότε οι επιμέρους ιδιότητες που αφορούν την μάθηση με επίβλεψη ή την εποπτευόμενη ή την επιτηρούμενη μάθηση είναι η [50]:

- Κατηγοριοποίηση ή Ταξινόμηση (Classification), όπου αναφέρεται στην κατασκευή προγνωστικών μοντέλων για διακριτές τάξεις. Αποσκοπεί στην πρόβλεψη και την τοποθέτηση ενός γεγονότος στην αντίστοιχη κατηγορία του με βάση συγκεκριμένα γνωρίσματα που διαθέτει. Χαρακτηριστικό παράδειγμα ο διαχωρισμός της ηλεκτρονικής αλληλογραφίας. Σε περίπτωση που εντοπιστούν λέξεις, όπως «συγχαρητήρια κερδίσατε...», τότε τα μηνύματα αυτά ηλεκτρονικής αλληλογραφίας κατηγοριοποιούνται ως ανεπιθύμητα [58]. Τα μοντέλα που βασίζονται στην ταξινόμηση - κατηγοριοποίηση διακρίνονται σε δύο είδη, την δυαδική και την πολλαπλή ταξινόμηση [68]. Στην περίπτωση της δυαδικής ταξινόμησης υπάρχει η αποτύπωση μίας τιμής, κατάστασης από μα κλάση, κατηγορία στην οποία περιλαμβάνονται αποκλειστικά δύο τιμές [68]. Η δυαδική ταξινόμηση θα μπορούσε να αναφέρεται στην πρόγνωση του καιρού

σχετικά με το αν θα βρέξει ή όχι [68]. Ενώ στην περίπτωση των εναλλακτικών ή πολλαπλών μοντέλων ταξινόμησης το αποτέλεσμα που εξάγεται, πρόεχεται από μια κλάση και μια έξοδος που εμπεριέχει τουλάχιστον δυο διαφορετικές τιμές. Αυτό σημαίνει, ότι θα μπορούσε να γίνει μετεωρολογική πρόβλεψη μεταξύ διαφορετικών αποτελεσμάτων, δηλαδή αν θα βρέξει, αν θα χιονίσει, αν θα έχει ηλιοφάνεια [68].

- Παρεμβολή ή Παλινδρόμηση (Regression), αποσκοπώντας στην κατασκευή προγνωστικών μοντέλων για την ανάδειξη πρόβλεψη συγκεκριμένων αριθμητικών τιμών. Ένα παράδειγμα από την αγορά ακινήτων, έστω ότι εισάγουμε ως δεδομένα στο σύστημα όλες τις διαθέσιμες πληροφορίες που αφορούν ένα ακίνητο όπως τα τετραγωνικά μέτρα, ο αριθμός των υπνοδωματίων, αποθηκευτικός χώρος – αποθήκη, θέση στάθμευσης, ο όροφος, η χρονολογία κτίσης του, η περιοχή με τον ταχυδρομικό του κώδικα, το επιτόκιο των διαθέσιμων στεγαστικών δανείων, ο ενιαίος φόρος ιδιοκτησίας ακινήτων, ο ειδικός φόρος ακινήτων, το κόστος κατασκευής, ο συνολικός αριθμός των κατοικιών που είναι διαθέσιμα προς πώληση στην περιοχή κτλ., με σκοπό την πρόβλεψη της τιμής πώλησης του ακινήτου [68].

Αρχικά η βασική επιδίωξη είναι η εκμάθηση των απαραίτητων εννοιών από το υπολογιστικό σύστημα. Αυτό επιτυγχάνεται με την μέθοδο μάθησης διά της επαγωγής, όπου μια έννοια περιλαμβάνεται μέσα σε ένα γενικότερο σύνολο [50]. Για την μάθηση των εννοιών στο σύστημα αξιοποιείται ο αλγόριθμος απαλοιφής υποψηφίων. Χρήσιμο και αξιοποιήσιμο εργαλείο στο είδος αυτό είναι τα δέντρα κατηγοριοποίησης (classification) αξιοποιώντας τον αλγόριθμο id3, προβλέποντας με σχετική ακρίβεια την τιμή της μεταβλητής του μοντέλου [50].

Επίσης μια άλλη κατηγορία που χρησιμοποιείται, ώστε να προσεγγιστεί η τιμή της μεταβλητής είναι με τους κανόνες κατηγοριοποίησης. Οι βασικότερες κατηγορίες κανόνων είναι οι προτασιακοί, οι οποίοι βασίζονται στην σειριακή αλγοριθμική κάλυψη και οι κατηγορηματικοί πρώτης τάξεως, όπου προκύπτουν ύστερα από την εκτέλεση των αλγοριθμικών κανόνων της πρώτης τάξεως [50].

Επιπρόσθετα μια άλλη μέθοδος είναι η κατά περίπτωση μάθηση, όπου επιτυγχάνεται η μάθηση με επίβλεψη. Σε αυτή την περίπτωση το σύνολο των δεδομένων που είναι διαθέσιμα και αξιοποιήσιμα για την διαδικασία της μάθησης διατηρούν την πρωτογενή υπόσταση τους, δίχως να κωδικοποιηθούν, όπως γίνεται σε άλλες μεθόδους [50 , 51 , 52]. Ο τρόπος λειτουργίας του βασίζεται στην λήψη απόφασης, από το υπολογιστικό σύστημα για την επιλογή της κατηγορίας, που αφορά μια καινούργια περίπτωση εν σύγκριση με τις διαθέσιμες και καταχωρημένες αντίστοιχες περιπτώσεις [50, 51, 52]. Στην περίπτωση αυτή χρησιμοποιείται ο αντίστοιχος αλγόριθμος των k πλησιέστερων γειτόνων [50, 51, 52].

Επιπλέον είναι διαθέσιμη και η χρήση της μάθησης με την αξιοποίηση μια απλουστευμένης περίπτωσης κατηγοριοποίησης Bayesian. Σύμφωνα με την οποία θεωρείτε, ως δεδομένη η αυτοτέλεια και η αυθυπαρξία μεταξύ των ιδιοτήτων των τιμών που είναι αναγκαία για το σύστημα μάθησης [50 , 51].

Στην περίπτωση της παρεμβολής, κατά κύριο λόγο εκτελείται η δομή του γραμμικού μοντέλου κατά το οποίο η έξοδος συναρτάται είτε από κάποιου είδους γραμμικής συνάρτησης είτε από το συνολικό άθροισμα το οποίο έχει σταθμιστεί με βάση τις παραμέτρους εισόδου [50].

Η σπουδαιότερη μεθοδολογία, που αποτελεί και θεμέλιο λίθο για την παρεμβολή και την κατηγοριοποίηση με πληθώρα εφαρμογών έρχεται από την αξιοποίηση των Διανυσματικών Μηχανών Υποστήριξης (Support Vector Machines, SVMs). Βασικός πυλώνας της μεθοδολογίας αυτής είναι το θεωρητικό πλαίσιο που αφορά την στατιστική μάθηση και την χρήση των νευρωνικών δικτύων [50].

Σε ότι αφορά την επιτηρούμενη μάθηση τόσο για την κατηγοριοποίηση όσο και για την παρεμβολή χρησιμοποιούνται και τα νευρωνικά δίκτυα, όπου είναι πιο εύχρηστα για την κατανόηση και την εκμάθηση συναρτήσεων, τόσο σε αριθμητικά, όσο και σε συνεχή ή και σε διακριτά διανυσματικά μεγέθη [50].

Κάποιοι από τους κυριότερους αλγορίθμους που χρησιμοποιούνται στην επιτηρούμενη μάθηση είναι [54]:

- k – Πλησιέστεροι Γείτονες (k- Nearest Neighbors, k-NNs)
- Διανυσματικές Μηχανές Υποστήριξης (Support Vector Machines, SVMs)
- Νευρωνικά Δίκτυα (Neural Networks, NNs)
- Δενδροειδές σύστημα απόφασης-κατηγοριοποίησης (Decision Trees-classificaiton)
- Παλινδρόμησης (Regression, linear and logistic)

Για τον τομέα της Κυβερνοασφάλειας ιδιαίτερο ενδιαφέρον παρουσιάζουν οι αλγόριθμοι που αφορούν την κατηγοριοποίηση (classification), για την αποφυγή κακόβουλων λογισμικών ή την αποφυγή ανεπιθύμητων μηνυμάτων (spam classification). Η αρχή λειτουργίας της κατηγοριοποίησης των ανεπιθύμητων μηνυμάτων βασίζεται στην εκμάθηση του συστήματος με βάση συγκεκριμένα δεδομένα που εισάγονται στο σύστημα και των διαθέσιμων χαρακτηριστικών που αντιπροσωπεύουν τα δεδομένα, τις αντίστοιχες ετικέτες τους, ώστε να είναι σε θέση να κατηγοριοποιεί τα νέα μηνύματα σε σύγκριση με τα παλιά μηνύματα ως ανεπιθύμητα. Συντελεί θετικά η χρήση των νευρωνικών δικτύων, όπου βασίζουν την εκμάθηση τους στις ετικέτες των δεδομένων δημιουργώντας ένα αντίστοιχο μοτίβο που αφορά την ανίχνευση εν δυνάμει επιθέσεων ψαρέματος (phishing), κακόβουλο λογισμικών, κτλ [55].

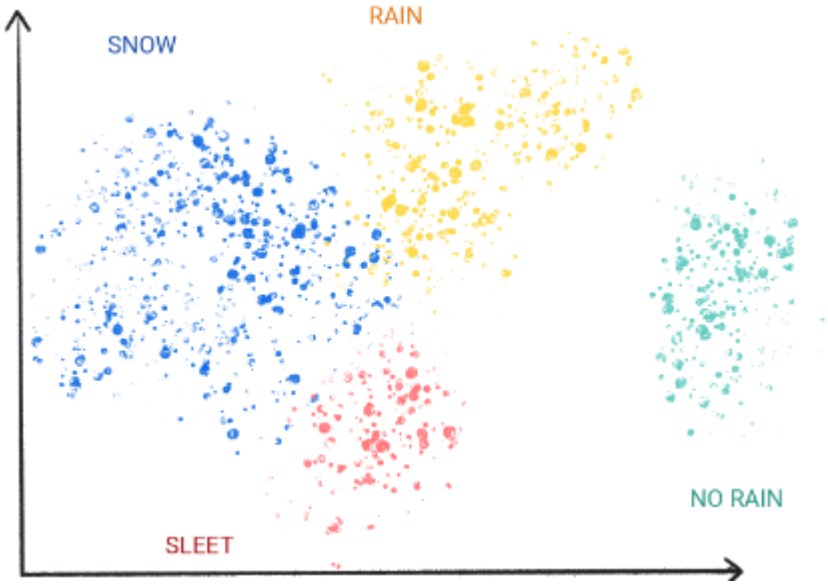
3.4.2 Χωρίς Επίβλεψη Μάθηση (Unsupervised Learning, UL)

Στο δεύτερο είδος το υπολογιστικό σύστημα αναλαμβάνει την δημιουργία προτύπων μέσω συσχετίσεων ή ομάδων, μέσα από ένα σύνολο δεδομένων με βάση τις ιδιότητες τους, δίχως να έχει οποιαδήποτε προγενέστερη γνώση για τα πρότυπα [50 , 51]. Τα αποτελέσματα των εν λόγω μοντέλων είναι προβλέψεις οι οποίες δεν βασίζονται σε δεδομένα με ορθές απαντήσεις. Σκοπός αυτού του είδους είναι ο εντοπισμός των μοτίβων που υπάρχουν στα δεδομένα. Αυτό σημαίνει ότι, δίχως να υπάρχει προγενέστερη μάθηση του μοντέλου για τον τρόπο και την μεθοδολογία που οφείλει να ακολουθήσει για τον διαχωρισμό των δεδομένων ανά κατηγορία, διαμορφώνει από μόνο του την σειρά των κανόνων που χρειάζονται για να πραγματώσει το προηγούμενο αποτέλεσμα την κατηγοριοποίηση [68]. Κάποιες περιπτώσεις που αφορούν, τα εν λόγω πρότυπα και είναι άξια λόγου είναι οι κανόνες συσχέτισης (association rules), καθώς και οι συστάδες, δηλαδή οι ομάδες (clusters) που προκύπτουν από την αντίστοιχη διαδικασία [50 , 52].

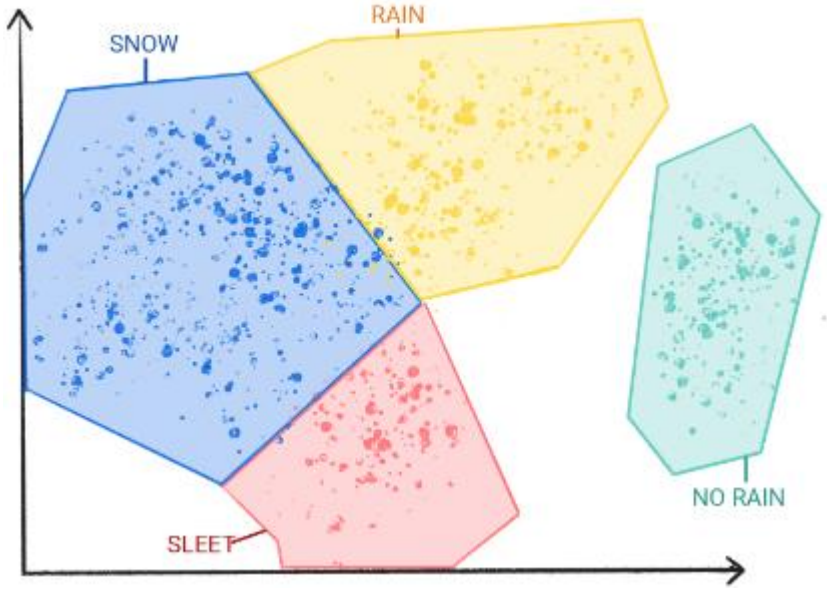
Οι κανόνες συσχέτισης είναι αποτέλεσμα ερευνητικών, κυρίως προγραμμάτων που αφορούσαν τις βάσεις δεδομένων. Η αρχή λειτουργίας τους βασίζεται στην συσχέτιση των δεδομένων εισόδων, με αντίστοιχη ανάλυση τους, με σκοπό την πρόβλεψη της πιθανής εξόδου. Όπως είναι λογικό η αξιοπιστία αυτών των κανόνων συσχέτισης συναρτάται, από την ποσοτική και ποιοτική αξιολόγηση τους [50]. Για αυτό υπάρχουν κάποια σημαντικά μεγέθη, που συντελούν στην κατεύθυνση αυτή, είναι η κάλυψη (coverage) και η ακρίβεια (accuracy). Ο αλγόριθμος, που χρησιμοποιείται για τους κανόνες συσχέτισης είναι ο Apriori [50 , 52, 55]. Ο τρόπος δράσης του αλγορίθμου, στηρίζεται κατά πρώτον στην ολοένα κατασκευή κανόνων που αφορούν τα αντικείμενα και δεύτερον στην μετέπειτα διαμόρφωση και σύσταση των αντίστοιχων κανόνων για την συσχέτιση [50].

Σε ότι αφορά την διαδικασία της ομαδοποίησης για την δημιουργία του αντίστοιχου πρότυπου, τότε κατά κύριο λόγο πραγματοποιείται διαμοιρασμός ανάμεσα, από ένα σύνολο δεδομένων με βάση, είτε την μέγιστη ομοιότητα, ως προς τα στοιχεία που περιέχονται εντός της ίδιας ομάδας, είτε την μέγιστη ανομοιότητα, ως προς τα στοιχεία που περιέχονται σε διαφορετικές ομάδες [50].

Στο σημείο αυτό δεν θα πρέπει να συγχέεται η έννοια της ομαδοποίησης με την ταξινόμηση ή την κατηγοριοποίηση. Η ειδοποιός διαφορά τους είναι ότι στην ομαδοποίηση η διαμόρφωση και η σύσταση των διάφορων κατηγοριών δεν είναι προκαθορισμένη και καλώς ορισμένης κατάσταση από τον μηχανικό του συστήματος. Στην περίπτωση που ένα σύστημα βασίζεται στο μοντέλο χωρίς επίβλεψη και έχει ως δεδομένα των σύνολο των καιρικών θερμοκρασιών είναι σε θέση να τις τοποθετήσει σε κατηγορίες αναδεικνύοντας τις τέσσερις εποχές του χρόνου [68].



Εικόνα 10 – Ομαδοποίηση ομοειδών μοτίβων καιρού [68]



Εικόνα 11 – Ομαδοποίηση των μετεωρολογικών μοτίβων, χιόνι, βροχή, χιονόνερο, ανομβρία [68]

Οι αντίστοιχοι αλγόριθμοί που αξιοποιούνται είναι [50, 51, 54]:

- Ανάλυση Κυρίων Συνιστωσών (Principal Component Analysis, PCA)
- Μεθοδολογία PCA (PCA Kernel)
- k-μέσων (k-means)

- Ιεραρχική Ανάλυση Συστάδων (Hierrarchical cluster analysis, HCA)

Ενδεικτικές περιπτώσεις χρήσης των αλγορίθμων χωρις επίβλεψη ή των μη επιτηρούμενων, για τον τομέα της Κυβερνοασφάλειας αποτελούν, ο εντοπισμός κακόβουλων λογισμικών (malware), οικονομικές απάτες (fraud) και ανεπιθύμητα μηνύματα ηλεκτρονικού ταχυδρομείου (email spamming) [54].

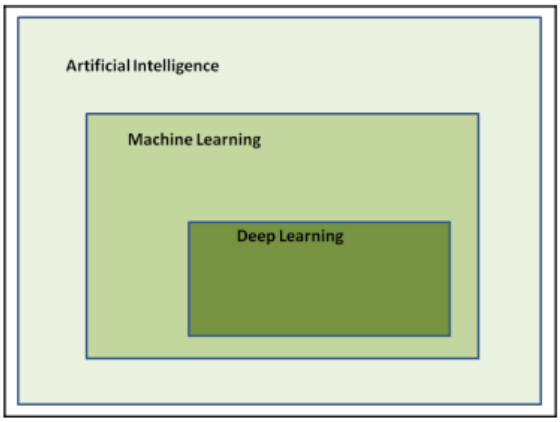
3.4.3 Ενισχυτική Μάθηση (Reinforcement Learning, RL)

Στην ενισχυτική μάθηση, η εκμάθηση του συστήματος βασίζεται στην λογική δοκιμής - λάθους (trial – error). Για την λήψη αποφάσεων αξιολογούνται τα δοθέντα στοιχεία και πληροφορίες στην εν εξελίξει διάρκεια μάθησης, έχοντας ως δεδομένο το μέγεθος και το πλήθος των ορθών απαντήσεων που έχουν προκύψει από τον αλγόριθμο [54]. Στην πραγματικότητα ο τρόπος λειτουργίας του βασίζεται στην καταγραφή τόσο των θετικών αποκρίσεων στα διάφορα αποτελέσματα που προήλθαν από τις δοκιμές όσο και των αρνητικών αποκρίσεων [54]. Πρόκειται για ένα δυναμικό τρόπο μάθησης, όπου αντικατοπτρίζει σε μεγαλύτερο βαθμό την αποστολή της τεχνητής νοημοσύνης [54].

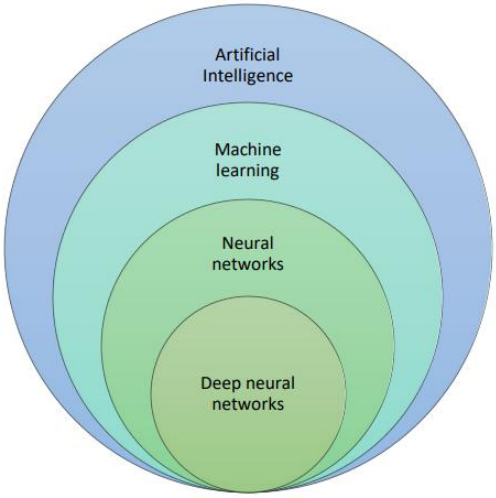
Χρήσιμοι αλγόριθμοι για αυτό το είδος μάθησης είναι [54]:

- Διαδικασία Markov (Markov process)
- Μέθοδος Q-learning
- Χρονικής Διαφοράς , ΧΔ (Temporal Difference, TD)
- Μόντε Κάρλο (Monte Carlo)

Αξίζει να σημειωθεί ότι, ως βαθιά μηχανική μάθηση (deep learning) ορίζεται η εξολοκλήρου επιλογή των δεδομένων που είναι χρήσιμα και αξιοποιήσιμα για το υπολογιστικό σύστημα οδηγώντας στην λύση μέσα από την δημιουργία του κατάλληλου μοντέλου λύσεως [50]. Συνδυάζει διαφορετικά χαρακτηριστικά των δεδομένων εισόδου αξιοποιόντας τους κατάλληλους αλγορίθμους για την αυτοματοποιημένη διαδικάσια εκμάθησης.



Εικόνα 12 – Απεικόνιση Σχέσεων Βαθιά Μηχανικής Μάθησης – Μηχανική Μάθηση – Τεχνητή Μάθηση [54]



Εικόνα 13 – Σχέσεις Εξάρτησης μεταξύ Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης [59]

3.5 Ελληνική Νομοθεσία [64]

Τον Ιούλιο του 2022 ψηφίστηκε από την Ελληνική Βουλή ο Νόμος 4961/2022, όπου αφορά τις αναδυόμενες τεχνολογίες πληροφορικής, συμπεριλαμβανομένης και της τεχνητής νοημοσύνης. Σκοπός του συγκριμένου Νόμου είναι η διαμόρφωση του κατάλληλου νομοθετικού πλαισίου για τον καθορισμό λειτουργίας των τεχνολογιών τεχνητής νοημοσύνης στην Ελληνική Επικράτεια καθώς και η θωράκιση των δημόσιων φορέων και υποδομών έναντι απειλών του κυβερνοχώρου.

Αρχικά όσο αφορά την ψηφιακή αναβάθμιση της δημόσιας διοίκησης, αποτυπώθηκε το ρυθμιστικό πλαίσιο για την ανάπτυξη συστημάτων τεχνητής νοημοσύνης. Συμπεριλήφθηκαν άρθρα, που αφορούν τον τρόπο επεξεργασίας των προσωπικών δεδομένων, από τα συστήματα τεχνητής νοημοσύνης, την εν γένη χρήση τους από τους δημόσιους φορείς, την υποχρέωση εκπόνησης αλγοριθμικής εκτίμησης αντικτύπου, πριν από την έναρξη λειτουργίας τους, την εγγραφή των συστημάτων τεχνητής νοημοσύνης στο αντίστοιχο μητρώο ακολουθώντας συγκεκριμένη πολιτική δεοντολογικής χρήσης των δεδομένων καθώς και μια σειρά υποχρεώσεων ιδιωτικών ή δημόσιων φορέων με την σύσταση αντίστοιχων συντονιστικών, εποπτικών επιτροπών και παρατηρητηρίου.

Επιπρόσθετα υπήρξε μια σειρά θεσμικών ενεργειών για την ασφάλεια πληροφοριών και την προστασία των δεδομένων προσωπικού χαρακτήρα. Συγκεκριμένα ορίστηκε η Γενική Διεύθυνση Κυβερνοασφάλειας της Γενικής Γραμματείας Τηλεπικοινωνιών και Ταχυδρομείων του Υπουργείου Ψηφιακής Διακυβέρνησης, ως Εθνική Αρχή Πιστοποίησης της Κυβερνοασφάλειας και ως γενικό εθνικό κέντρο συντονισμού για ζητήματα Κυβερνοασφάλειας. Ακόμη συστήνεται Παρατηρητήριο Ανάλυσης Υβριδικών Απειλών το οποίο υπάγεται στο Υπουργείο Ψηφιακής Διακυβέρνησης. Επίσης, αποτυπώθηκαν οι υποχρεώσεις που αφορούν τους φορείς με κρίσιμες ψηφιακές υποδομές, τον ορισμό στον αντίστοιχο ηλεκτρονικό μητρώο των υπευθύνων προστασίας δεδομένων, την σύσταση με τις αντίστοιχες αρμοδιότητες της Επιτροπής των υπευθύνων προστασίας δεδομένων.

3.6 Δεοντολογικά Ζητήματα Τεχνητής Νοημοσύνης

Είναι ηλίου φαεινότερο ότι η χρήση της τεχνητής νοημοσύνης βασίζεται σε αλγορίθμους σε δεδομένα και σε λήψη απόφασης για ενέργεια από το σύστημα. Επίσης η έξοδος που παράγεται ως απόκριση του συστήματος είναι αποτέλεσμα διεργασιών που μπορεί να έχουν καθοριστεί με αδιαφανή και μεροληπτικό τρόπο. Επιπρόσθετα η ραγδαία αύξηση της τεχνητής νοημοσύνης με

γεωμετρική πρόοδο, είναι βέβαιο, ότι θα επιφέρει μείωση των ανθρωποωρών εργασίας. Ιδιαίτερα στην περίπτωση που οι χειρωνακτικές εργασίες αντικατασταθούν από Συσκευές τεχνητής νοημοσύνης , αντικαθιστώντας σε μεγάλο βαθμό την ανθρώπινη εργασία [55].

Η ιστορία μας έχει διδάξει ότι η τεχνολογία δυστυχώς χρησιμοποιείται και με κακόβουλο σκοπό. Η περίπτωση χρήσης από κακοποιούς είναι ένα στοιχείο που μέχρι ένα βαθμό είναι αναμενόμενο και προβλέψιμο. Όμως υπάρχει και η περίπτωση η λήψη αποφάσεων από αυτοματοποιημένα συστήματα να οδηγήσουν σε όλεθρο και ακραίες συμπεριφορές. Για παράδειγμα, έστω ότι τα μηχανήματα με τεχνητή νοημοσύνη έχουν ως στόχο την εξάλειψη σε παγκόσμιο επίπεδο του καρκίνου, καταλήγοντας ως έξοδο – λύση στον θάνατο όλων στο πλανήτη, από την μια το μηχάνημα τεχνητής νοημοσύνης έδωσε την λύση από την ανθρώπινη όμως πλευρά αυτό φαντάζει ως όλεθρος [55]. Επιπρόσθετα τα αυτοοδηγούμενα οχήματα τα οποία βασίζονται στην χρήση νοημοσύνης έχουν αρκετά ζητήματα ηθικής. Σε περίπτωση που ένα αυτοοδηγούμενο όχημα αντιληφθεί ότι ένα φορτηγό πρόκειται να συγκρουστεί μετωπικά με το όχημα σε μια γέφυρα άλλα οι εναλλακτικές του είναι, είτε να πέσει από την γέφυρα, είτε να χτυπήσει ένα άλλο διερχόμενο όχημα, τότε όποια απόφαση και αν λάβει το σύστημα ποιο θα είναι το κριτήριο, ο καταλογισμός ευθυνών και τα ηθικά ζητήματα; Έχοντας ως πρωταρχικό στόχο την ανθρώπινη ζωή, όπου είναι το ίδιο σημαντική για όλους, χωρίς να υπόκεινται σε αλγοριθμική εξίσωση.

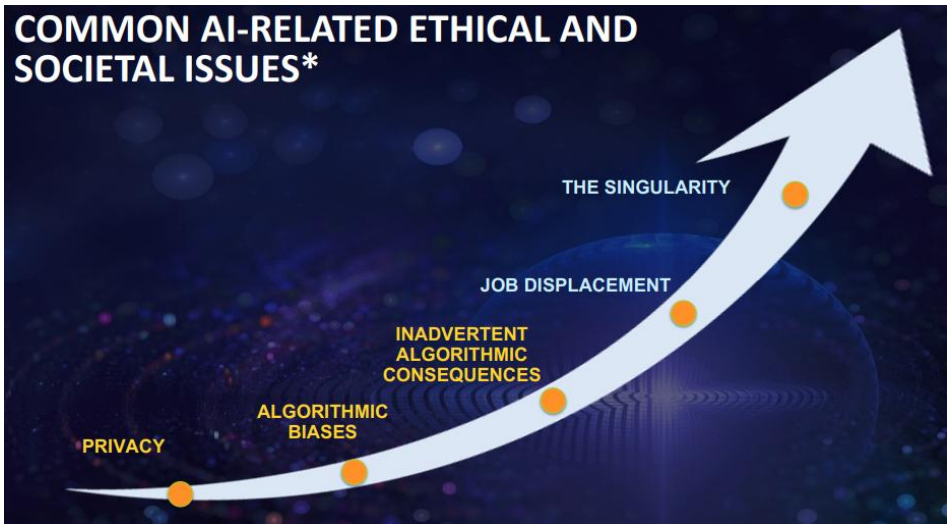
Η κατάσταση αυτή εγείρει πολλούς προβληματισμούς για δεοντολογικά ηθικά ζητήματα καθώς και για την προστασία των υπέρτατων αγαθών που ορίζονται από το Σύνταγμα και τους Νόμους. Οι Ελευθερίες του Ανθρώπου, τα Ανθρώπινα Δικαιώματα οι Δημοκρατικοί θεσμοί είναι έννοιες άρρηκτα συνδεδεμένες με αυτό που αναφέρεται συχνά, το Κράτος Δικαίου. Γεγονός το οποίο αναδεικνύει την αναγκαιότητα διαφύλαξης τους δίχως να δίνεται η δυνατότητα παραβίασης τους [58].

Διεξάγονται έρευνες, κυρίως σε Ευρωπαϊκό επίπεδο για τα ζητήματα ηθικής χρήσης των τεχνολογιών τεχνητής νοημοσύνης και των Αυτόματων Έξυπνων Συστημάτων. Γίνονται ερευνητικές προσπάθειες, όπως από την IEEE με το ερευνητικό πρόγραμμα πιστοποίησης Ηθικής Δεοντολογίας για Αυτόνομα και Ευφυή Συστήματα (Certification Program for Autonomous and Intelligent Systems, ECPAIS). Βασική επιδίωξη είναι η δημιουργία συγκεκριμένων διαδικασιών και προδιαγραφών πιστοποίησης με την αντίστοιχη χρήση ειδικής σήμανσης για την προάσπιση της διαφάνειας και την ελαχιστοποίηση της μεροληπτικής αντιμετώπισης από τα αυτόνομα και Ευφυή Συστήματα, τα όποια χρησιμοποιούν προφανώς την Τεχνητή Νοημοσύνη. [65, 66].

Επιγραμματικά οι εννιά (9) κατευθυντήριοι άξονες για την αποφυγή μεροληπτικών στάσεων και αντιλήψεων και αντίστοιχα τον ορθολογικό σχεδιασμό και ανάπτυξη συστημάτων τεχνητής νοημοσύνης αποτυπώθηκαν από την Ευρωπαϊκή Επιτροπή το 2020 και είναι [58]:

- **Ανθρώπινη Αξιοπρέπεια**: Η εν γένει αντιμετώπιση για οιοδήποτε ζήτημα των ανθρώπων λογίζεται από την ηθική υπόσταση και όχι από την διαχείριση του ανθρώπου ως αντικείμενο το οποίο αξιολογείται, βαθμολογείται ή διαμορφώνεται υπό το αλγοριθμικό πρίσμα.
- **Ανθρώπινη Ελευθερία**: Τα συστήματα Τεχνητής Νοημοσύνης θα ήταν χρήσιμο να αξιοποιηθούν για την βελτίωση των δυνατοτήτων των ανθρώπων και όχι για τον έλεγχο και τον περιορισμό των Ελευθεριών του Ανθρώπου.
- **Πρόληψη έναντι απειλών**: Είναι αδιαμφισβήτητα αναγκαία η προστασία της σωματικής και της ψυχικής ακεραιότητας των ανθρώπων και του περιβάλλοντος από την χρήση της τεχνολογίας τεχνητής νοημοσύνης.

- **Ισότητα Φύλων – Δικαιοσύνη – Διαφορετικότητα**: Οτιδήποτε βασίζεται στην Τεχνητή Νοημοσύνη κατά τον σχεδιασμό είναι απαραίτητες οι δικλίδες ασφαλείας ως προς την ισότητα των ανθρώπων δίχως την ύπαρξη στοιχείων διαχωρισμού τους.
- **Διαφάνεια Συστημάτων Τεχνητής Νοημοσύνης**: Επεξήγηση του τρόπου λειτουργίας και αποσαφήνιση του σκοπού εξαγωγής των αποτελεσμάτων ενός συστήματος τεχνητής νοημοσύνης στα άτομα που επηρεάζονται από αυτό.
- **Ιδιωτικότητα και Προστασία Δεδομένων**: Κατά τον σχεδιασμό των αντίστοιχων συστημάτων τεχνητής νοημοσύνης, είναι σημαντικό να λαμβάνονται υπόψιν όλες οι πτυχές διασφάλισης των προσωπικών δεδομένων του ατόμου.
- **Απόδοση Ευθύνης**: Το σύνολο των εμπλεκομένων στον σχεδιασμό και την ανάπτυξη ενός συστήματος τεχνητή νοημοσύνης, έχουν την υποχρέωση λογοδοσίας με βάση την ισχύουσα νομοθεσία όταν και αν προκαλείται βλάβη στους χρήστες ή σε άλλους.
- **Δημοκρατία**: Κύριο συστατικό στοιχείο του εποπτικού μηχανισμού, όπου διαμορφώνεται από τον σχεδιασμό κιόλας ενός συστήματος τεχνητής νοημοσύνης, είναι η διαφάνεια και η προστασία των δημοκρατικών διαδικασιών λήψης απόφασης, ο πλουραλισμός και η καθολική δημόσια πρόσβαση σε πληροφορίες ανοικτών δεδομένων.
- **Νομοθεσία**: Θωράκιση των συστημάτων τεχνητής νοημοσύνης με διαδικασίες που διασφαλίζουν την διαφάνεια, την ακεραιότητα, και την ορθολογική επεξεργασία των δεδομένων.



Εικόνα 14 – Δεοντολογικά Ζητήματα Τεχνητής Νοημοσύνης [72]

Συμπερασματικά, η τεχνητή νοημοσύνη έχει μια δισυπόστατη συμπεριφορά, όπως ισχύει με τα περισσότερα εργαλεία, από την μια δίνει λύσεις σε πολλά ζητήματα με αποδοτικότερη και βέλτιστη λύση τους και από την άλλη καλείται να αντιμετωπίσει ένα πλήθος νέων προκλήσεων που η ίδια η τεχνητή νοημοσύνη φέρει.

Στο κυβερνοχώρο εργαλεία της τεχνητής νοημοσύνης συντελούν στην θωράκιση της ασφάλειας, ενισχύοντας την ανθεκτικότητα και την καλή λειτουργία ενός πληροφοριακού συστήματος. Κακόβουλες ενέργειες και επιθέσεις αντιμετωπίζονται με αποδοτικότερο και ταχύτερο τρόπο.

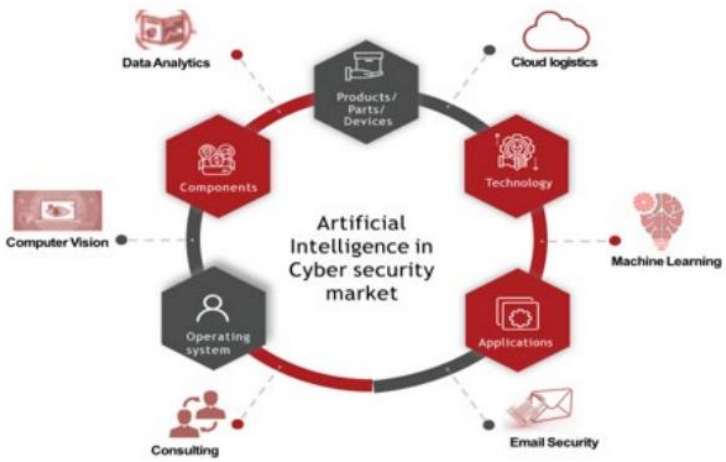
Από την άλλη οι επιτιθέμενοι που θα έχουν την γνώση της Μηχανικής και Βαθιάς Μάθησης, αναζητούν την πραγμάτωση νέων εκσυχρονισμένων τρόπων επίθεσης πλήττοντας μαζικότερους στόχους, με ασύλληπτα ταχύτερους ρυθμούς με καταστροφικότερες συνέπειες αλλά και με συγκεκριμένη στόχευση.

Επίσης η καθολική χρήση της τεχνητής νοημοσύνης και της μηχανικής μάθησης εγείρει προβληματισμού τόσο για το πόσο θωρακισμένα είναι τα συστήματα τεχνητής νοημοσύνης έναντι ενδεχόμενων απειλών σε αυτά. Επιπρόσθετα πολύ σημαντικός παράγοντας και τα δεοντολογικά και ηθικά ζητήματα και την διαδικασία αλγοριθμικής λήψης απόφασης από τα αυτοματοποιημένα συστήματα.

ΜΕΡΟΣ Α': Κεφάλαιο 4 – Τεχνητή Νοημοσύνη στην Κυβερνοασφάλεια

Η αθρόα και απαραίτητη χρήση των υπηρεσιών διαδικτύου αυξήθηκε με γεωμετρική πρόοδο από το 2017 έως το 2023 [73]. Αυτό έχει ως άμεση συνέπεια την αύξηση των επιθέσεων με αντίστοιχη δυσχέρεια στην αποτελεσματική αντιμετώπιση τους [73]. Για τον λόγο αυτό επιχειρήσεις και οργανισμοί, στα πλαίσια ενός ευρύτερου ψηφιακού μετασχηματισμού χρησιμοποιούν εργαλεία τεχνητής νοημοσύνης ώστε να δώσουν μια βέλτιστη λύση ως προς την κυβερνο-ασφάλεια [73].

Η κυβερνο-ασφάλεια σε οποιοδήποτε χώρο αποσκοπεί στην διατήρηση της ασφάλειας των πληροφοριακών συστημάτων και στην προστασία των προσωπικών και ψηφιακών δεδομένων έναντι οποιασδήποτε ενδεχόμενης απειλής [69]. Η χρήση της τεχνητής νοημοσύνης στην Κυβερνο-ασφάλεια προσφέρει λύσεις σε διαφορετικούς τομείς.



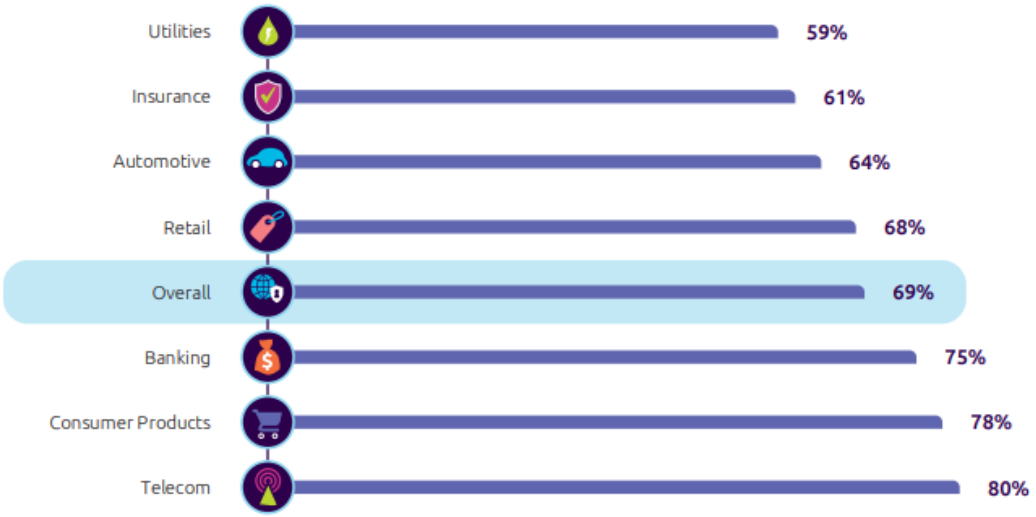
Εικόνα 15 – Τομείς Χρήσης Τεχνητής Νοημοσύνης στην αγορά της Κυβερνοασφάλειας [69]

Ένας σημαντικός τομέας που καταλαμβάνει νέα εδάφη διαρκώς είναι η δημιουργία μοτίβων, με βάση την συλλογή των δεδομένων από υπολογιστικά συστήματα, δικτυακή κίνηση, ανοικτά δεδομένα για γνωστές επιθέσεις στη διεθνή κοινότητα και αλλά στοιχεία όπου στη συνέχεια πραγματοποιείται η αντίστοιχη ανάλυση τους (data analytics) [63]. Με τον τρόπο αυτό αποτρέπονται ενδεχόμενες επιθέσεις [73]. Πάντα υπό την προϋπόθεση της ασφάλειας από το σχεδιασμό, καθώς μια λαθεμένη υλοποίηση και χρήση της τεχνητής νοημοσύνης μπορεί να δημιουργήσει αδυναμίες και νέα ζητήματα ασφάλειας παρόλο που φαινομενικά θα έχει λύσει το αρχικό ζητούμενο [73].

Επιπρόσθετα, λειτουργίες της τεχνητής νοημοσύνης στη κυβερνοασφάλεια είναι η οπτικοποίηση ενός συστήματος, συμβουλευτική λειτουργία ως προς την ομαλή ή όχι ομαλή λειτουργία ενός λειτουργικού συστήματος, σχετικά με την ασφάλεια χρήσης εφαρμογών, ηλεκτρονικής αλληλογραφίας καθώς και τομείς που άπτονται με την μηχανική μάθηση και την

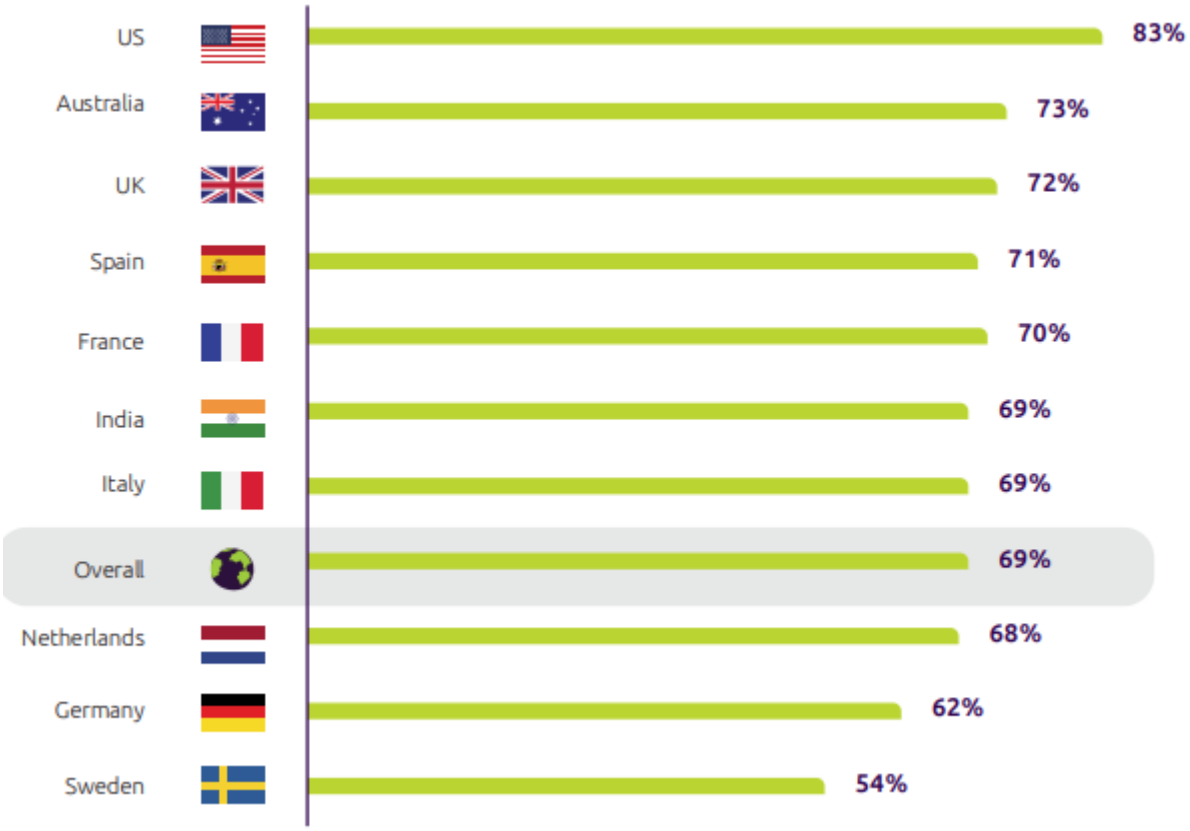
αλληλοσυμπλήρωση των ζητούμενων για την διαμόρφωση ολοκληρωμένης λύσης ανάλογα με τις ισχύουσες ανάγκες οργανισμών ή επιχειρήσεων.

Επιχειρήσεις και τομείς που έχουν ιδιαίτερο επενδυτικό ενδιαφέρον σε τεχνολογίες κυβερνο-ασφάλειας, που βασίζονται σε τεχνητή νοημοσύνη είναι αρκετές και γενικά στο σύνολο της επιχειρηματικής δραστηριότητας όλες οι επιχειρήσεις επιθυμούν να διασφαλίζουν τα δεδομένα τους είτε είναι οικονομικά, είτε δεδομένα πελατών ενδεχομένως και προϊόντων παραγωγής. Ενδεικτικά, οι τηλεπικοινωνιακοί πάροχοι, οι εμπλεκόμενοι σε διαδικτυακές πωλήσεις καταναλωτικών προϊόντων, τραπεζικών προϊόντων αλλά και το λιανεμπόριο, η αυτοκινητοβιομηχανίες, οι ασφαλιστικοί φορείς επιδιώκουν να έχουν τις βέλτιστες λύσεις κυβερνοασφάλειας για τα πληροφοριακά τους συστήματα, με την αξιοποίηση εργαλείων ή ολοκληρωμένων λύσεων τεχνητής νοημοσύνης.



Εικόνα 16 – Αξιοποιήση εργαλείων Τεχνητής Νοημοσύνης για τον εντοπισμό απειλών και επιθέσεων ανά τομέα απασχόλησης [73]

Φυσικά, η επένδυση σε τεχνολογίες τεχνητής νοημοσύνης για την θωράκιση της ασφάλειας των πληροφοριακών συστημάτων των επιχειρήσεων εξαρτάται, από το οικονομικό τους μέγεθος, από τη συνάρτηση κόστους – οφέλους που είναι η βασική οικονομική αρχή για την επίτευξη του κέρδους καθώς από την γεωγραφική τους δραστηριότητα. Το ίδιο ισχύει και τους οργανισμούς των κρατών ανά τον κόσμο. Σε έναν ψηφιακό και ενοποιημένο κόσμο μέσω διαδικτύου, οι απειλές αφορούν τους πάντες και κυρίως τους πιο ευάλωτους. Χώρες οι οποίες δεν έχουν επενδύσει στην κουλτούρα της ασφάλειας των πληροφοριακών συστημάτων, στην ασφάλεια από το σχεδιασμό ενός ολοκληρωμένου πληροφοριακού συστήματος για την παροχή υπηρεσιών και την εξυπηρέτηση της αποστολής του κάθε σύστηματος, αποτελούν στόχο για τους επιτιθέμενους. Σε αντίθεση με προηγμένες τεχνολογικά χώρες που έχουν υψηλά στην ιεραρχία ανάπτυξης και λειτουργίας των συστημάτων ζητήματα που αφορούν την ασφάλεια των δεδομένων και εν γένει των πληροφοριακών συστημάτων και των εφαρμογών που ολοένα και πιο πολύ είναι αναγκαία για την καθημερινή μας ζωή σε κάθε έκφανση της.



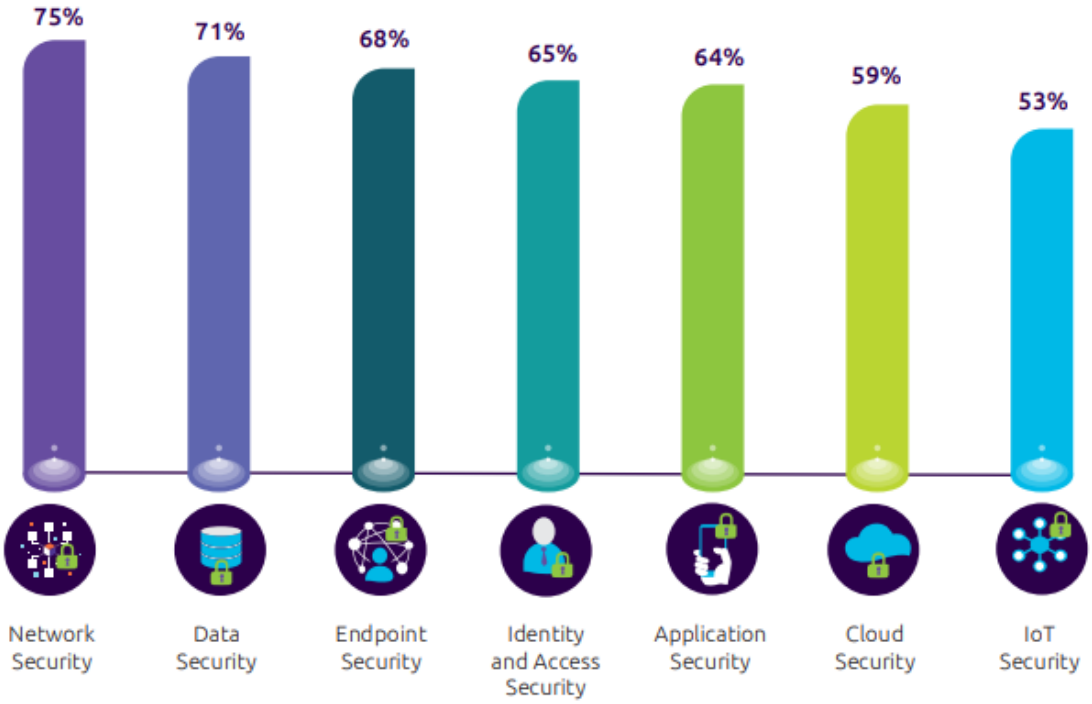
Εικόνα 17 – Εργαλεία Τεχνητής Νοημοσύνης στη Κυβερνοασφάλεια ανά χώρα [73]

Για αυτό οι Ηνωμένες Πολιτείες Αμερικής, η Αυστραλία, το Ηνωμένο Βασίλειο, η Ισπανία, η Γαλλία, η Ινδία, η Ιταλία είναι κάποιες από τις χώρες που έχουν κατανοήσει την αναγκαιότητα χρήσης τεχνολογιών τεχνητής νοημοσύνης για την προάσπιση των συμφερόντων τους και των πληροφοριακών τους συστημάτων. Οι επιθέσεις κοστίζουν και σε κάποιες περιπτώσεις οι επιπτώσεις δεν είναι σε οικονομικά ανεκτό βαθμό αλλά πλήττουν ανεπανόρθωτα την λειτουργία επιχειρήσεων και οργανισμών.

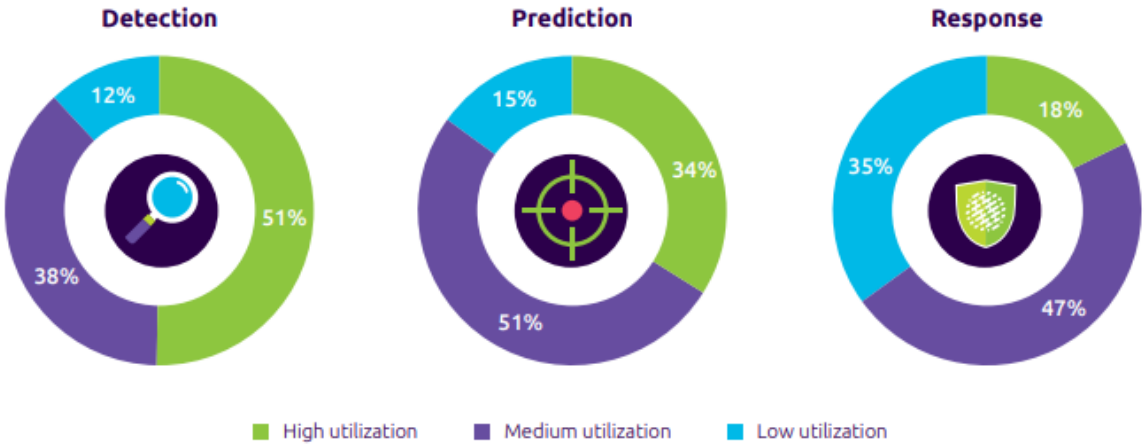
4.1 Κατηγορίες Τεχνητής Νοημοσύνης στην Κυβερνο-ασφάλεια

Οι κυριότερες κατηγορίες τεχνητής νοημοσύνης στην κυβερνο-ασφάλεια είναι οι παρακάτω:

- Ασφάλεια Δικτύου (Network Security),
- Ασφάλεια Δεδομένων (Data Security),
- Ασφάλεια τελικού σημείου (Endpoint Security),
- Ταυτοποίηση και Ασφάλεια Πρόσβασης (Identity and Access Security),
- Ασφάλεια Εφαρμογών (Application Security),
- Ασφάλεια νέφους (Cloud Security) και
- Ασφάλεια στο διαδίκτυο των πραγμάτων (Iot Security)



Εικόνα 18 – Χρήση ΤΝ ανά τομέα [73]



Εικόνα 19 – Περιοχές χρήσης ΤΝ στην Κυβερνο-ασφάλεια για Ανίχνευση – Πρόβλεψη – Αντιμετώπιση [73]

Η χρήση της τεχνητής νοημοσύνης στην Κυβερνοασφάλεια αξιοποιείται σε διαφορετικές κατηγορίες και τομείς ενδιαφέροντος. Στην παρούσα εργασία υπάρχει μια διττή διάκριση, σε γενικές, όπου υπάρχει μια γενική περιγραφή κάποιων αρχών λειτουργίας για την αντιμετώπιση ενδεικτικών ζητημάτων ασφάλειας στο κυβερνοχώρο και σε εμπορικές όπου δίνονται έτοιμες λύσεις από εταιρίες.

- **Γενικές**

Η ιεράρχιση των κατηγοριών που έχουν αναπτυχθεί για την προστασία των αγαθών και υπηρεσιών έναντι απειλών είναι τρεις γενικές κατηγορίες, όπου ισχύουν στην κυβερνοασφάλεια. [55]:

1. Ανίχνευση (Detection): Εποπτική λειτουργία της δικτυακής δραστηριότητας για την διαπίστωση τυχόν μη αναμενόμενων αλλαγών με την αντίστοιχη προειδοποίηση των διαχειριστών του δικτύου για την λήψη της αντίστοιχης απόφασης για την αντιμετώπιση του κάθε περιστατικού.
2. Πρόληψη (Prevention): Χρήση αλγορίθμων μάθησης για την βελτίωση των δυνατοτήτων πρόβλεψης των επιθέσεων.
3. Αντιμετώπιση (Response): Έχοντας προβεί στις αντίστοιχες προγνώσεις των επιθέσεων ή των κακόβουλων λογισμικών δίνεται η δυνατότητα αντιμετώπισης και τερματισμού ενός κακόβουλου λογισμικού με αυτόνομη διαδικασία.

- **Εμπορικές**

Οι τεχνολογίες τεχνητής νοημοσύνης στην Κυβερνοασφάλεια είναι χρήσιμες και αξιοποιήσιμες σε επιχειρήσεις και οργανισμούς. Είναι χρήσιμες καθώς βοηθούν στην ασφάλεια των υπηρεσιών τους δίνοντας την αντίστοιχη αξιοπιστία με αύξηση του εισοδήματος και της βιοσιμότητας τους.

Εικόνα 20 – Εταιρίες Τεχνητής Νοημοσύνης για την Κυβερνοασφάλεια ανα κατηγορία [74]

Οι κυριότερες για τις οποίες έχουν δημιουργηθεί εμπορικά προϊόντα είναι για την [74]:

i. Καταπολέμηση Απάτης & Διαχείριση Ταυτότητας (Anti Fraud & Identity Management)

Ένα από τα σημαντικότερα ζητήματα που απασχολούν τις σύγχρονες οικονομίες είναι η προστασία των διαδικτυακών συναλλαγών. Οι απάτες στον χώρο αυτό είναι πολλές και το κόστος μεγάλο για επιχειρήσεις, τραπεζικούς ομίλους και φυσικά πρόσωπα. Το γεγονός αυτό αποτέλεσε το

έναυσμα για την δημιουργία ενός ασφαλές περιβάλλοντος για τις διαδικτυακές συναλλαγές. Ο τρόπος λειτουργίας τους βασίζεται στον εντοπισμό ενδεχόμενων ύποπτων οικονομικών συναλλαγών, καθώς και ψεύτικών ιστοτόπων που προσομοιάζουν πραγματικούς δικτυακούς χώρους τραπεζικών ιδρυμάτων, αξιοποιώντας στο έπακρο τους αντίστοιχους αλγορίθμους μηχανικής μάθησης.

Κάποιες εμπορικές λύσεις τεχνητής νοημοσύνης αυτής της κατηγορίας είναι: Agari Data [75], Bangsun technology [76] ,Castle Intelligence [77], Cybertonica [78], DataVisor [79], FeedZai [80], GreatHorn [81] , IDwall [82], Pulse ID [83], Ravelin [84], Rippleshot [85], Shift Technology [86], Shumei Technology [87], Sift Science [88], Simility [89], Pathmind [90], Socure [91], και το VU Security [92].

ii. Ασφάλεια Εφαρμογών (App Security)

Η κατηγορία αυτή απευθύνεται κυρίως στους προγραμματιστές εφαρμογών αποσκοπώντας στην εύρεση τυχόν ευπαθειών ασφάλειας που έχουν προβαίνοντας στις αντίστοιχες διορθώσεις αυτών. Ενδεικτικά στην αγορά μεταξύ άλλων υπάρχουν το Authbase [93] και το Cryptosense [94].

iii. Αυτοματοποιημένα Συστήματα (Automated Systems)

Με την λειτουργία των εν λόγω συστημάτων ο χρήστης χρησιμοποιώντας, την φυσική του ομιλία ενημερώνεται σχετικά με την ασφάλεια των τερματικών του συσκευών στο σύνολο του δικτύου της επιχείρησης. Με αυτοματοποιημένο τρόπο συλλέγονται όλα τα δεδομένα και πραγματοποιείται ο συσχετισμός με την αξιοποίηση της τεχνητής νοημοσύνης σχετικά με την κατάσταση ασφάλειας των εμπλεκόμενων συστημάτων. Εμπορικά προϊόντα σε αυτή την κατηγορία είναι: Cortex [95], GoSecure [96], Devo [97], Tanium [98] και OCI [99].

iv. Ασφάλεια έναντι Παραπλάνησης (Deception Security)

Μια άλλη οπτική αντιμετώπισης μιας επίθεσης είναι η παραπλάνηση του επιτιθέμενου, με ένα εικονικό νευρωνικό δίκτυο με αντίστοιχους τερματικους υπολογιστές, υπηρεσίες, εφαρομογές δηλαδή προσομοιάζοντας το πραγματικό δίκτυο. Θυμιζεί την λειτουργία του cloud υπό την λειτουργία και την επίβλεπεψτη των αντίστοιχων αλγορίθμων μηχανικής μάθησης. Δυο τυχαία προϊοντα είναι το Cyberfog [100] και το illusive networks [101].

v. Ασφάλεια του Διαδικτύου των πραγμάτων (IoT Security)

Αναπτύσσονται τεχνολογίες τεχνητής νοημοσύνης, ώστε να υπάρχει προστασία του διαδικτύου των πραγμάτων εντοπίζοντας ενδεχόμενες απειλές. Κάποιες λύσεις είναι το Bastille Networks [102], το CUJO [103] και το SparkCognition [104].

vi. Ασφάλεια Κινητής Τηλεφωνίας (Mobile Security)

Η κατηγορία αυτή περιλαμβάνει τεχνολογίες που βασίζονται στο cloud και έχουν ως στόχο τον εντοπισμό μολυσματικών λογισμικών και προγραμμάτων σε εφαρμογές κινητών τηλεφώνων. Συνεπώς λειτουργεί ως φίλτρο στον εντοπισμό απειλών ενισχύοντας την ασφάλεια των κινητών τηλεφώνων. Αντίστοιχα παραδείγματα εμπορικών προϊόντων είναι: Appthority [105], Zimperium [106], Sentegrity [107].

vii. Πρόβλεψη μέσω ανάλυσης (Predictive Analysis)

Σκοπός της κατηγορίας αυτής είναι η ανάλυση του τρόπου δράσης των επιτιθέμενων (hackers), ώστε να δημιουργηθεί το αντίστοιχο μοντέλο πρόβλεψης ενδεχόμενων απειλών και

επιθέσεων στο Κυβερνοχώρο. Με την μεθοδολογία αυτή επιτυγχάνεται η ασφάλεια συσκευών και δικτύων. Παραδείγματα εμπορικών λύσεων: Cylance BlackBerry [108], Deep Instinct [109], Indeni [110], Innefu Labs [111], Cyr3con [112], sumo logic [113], Log Rhythm [114], Protenus [115], Seclytics [116], SentinelOne [117], και Anomali [118].

viii. Διαχείριση των κινδύνων στο Διαδίκτυο (Cyber-Risk Management)

Ο συγκεκριμένος χώρος αναφέρεται σε εταιρίες που αναλαμβάνουν την διεξαγωγή αντίστοιχων ερευνών και εργασιών που αφορούν την διαχείριση κινδύνου στον Κυβερνοχώρο. Γίνεται χρήση των αντίστοιχων εργαλείων για την προστασία της πνευματικής ιδιοκτησίας, της εμπιστευτικότητας των δεδομένων, την μέτρηση αντικτύπου και τον υπολογισμό των εν δυνάμει απειλών έναντι της ακεραιότητας των πληροφοριών έως και στην συμμόρφωση τήρησης της πολιτικής ασφάλειας. Για παράδειγμα κάποιες έτοιμες πλατφόρμες που αξιοποιούνται είναι τα ακόλουθα: CyberSaint Security [119], Cytora [120], Haystax Technology [121], MetaCert [122] και Aware [123].

ix. Εντοπισμός Δικτυακών παρεγκλίσεων (Anomaly Detection)

Σε αυτή την κατηγορία περιλαμβάνονται εργαλεία τεχνητής νοημοσύνης με προηγμένη μηχανική μάθηση για τον εντοπισμό μη φυσιολογικών λειτουργιών που υπάρχουν σε δίκτυα ή συστήματα. Κάποια εργαλεία αυτή της κατηγορίας είναι: BehavioSec [124], Darktrace [125], Exabeam [126], Intensity Analytics [127], PerimeterX [128], PrismaCloud [129], Spherical Defence [130] , StackPath [131], TwoSense [132].

Στο κυβερνοασφάλειας οι πρώτες προσπάθειες αξιοποίησης της τεχνητής νοημοσύνης βασίστηκαν στην τεχνική των υπογραφών, δημιουργώντας τα αντίστοιχα μοτίβα των επιθέσεων. Με την διαδικασία αυτή έγινε αποδοτικότερος και ταχύτερος ο εντοπισμός απειλών, κακόβουλων λογισμικών και προγραμμάτων. Επιπρόσθετα ο τεράστιος όγκος δεδομένων διαχειρίζεται ευκολότερα με την χρήση της τεχνολογίας της τεχνητής νοημοσύνης.

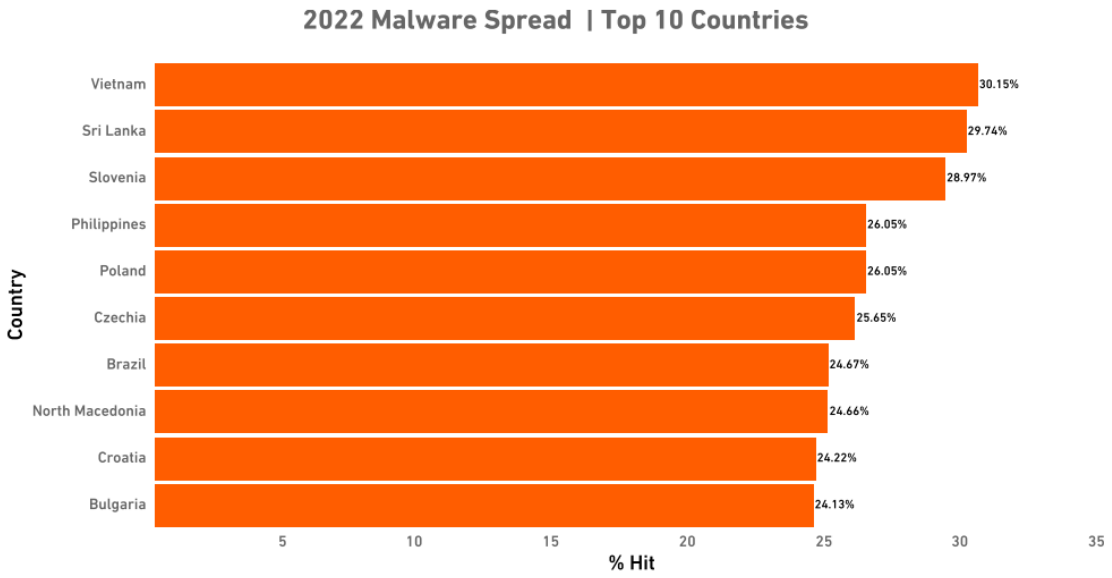
Η ταχύτατη ανάλυση των αρχείων καταγραφής, της δικτυακής συμπεριφοράς , των μολυσμένων προγραμμάτων και των συνεπειών που έχουν επέλθει από μια επίθεση σε ένα πληροφοριακό σύστημα, έδωσαν την δυνατότητα τόσο για την στοχευμένη πρόβλεψη ενδεχόμενων απειλών και επιθέσεων όσο και στον εντοπισμό αυτών. Ύστερα από τον εντοπισμό των επιθέσεων υπάρχει η δυνατότητα για αντιμετώπιση τους [63].

Στη συνέχεια πρόκειται να παρουσιαστούν κάποιες περιπτώσεις που άπτονται με την κυβερνο-ασφάλεια και τον τρόπο που διαχειρίζονται με την χρήση τεχνητής νοημοσύνης. Κυρίως αναφέρονται στο πεδίο της ανίχνευσης και της πρόβλεψης των απειλών, αξιοποιώντας αποδοτικά τους αλγορίθμους μηχανικής μάθησης καθώς οι απειλές στον κυβερνοχώρο μεταβάλλονται ή δημιουργούνται νέες [55].

4.2 Ανίχνευση και Ανάλυση Κακόβουλου Λογισμικού (Malware Detection - Analysis)

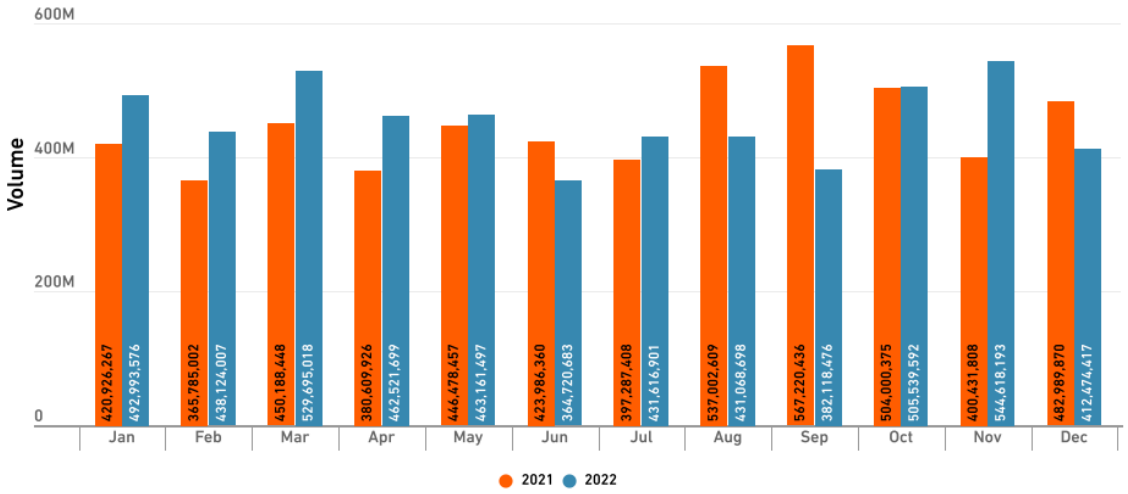
Τα κακόβουλα λογισμικά πλήττουν το σύνολο των λειτουργικών συστημάτων ανά τον κόσμο θέτοντας σε κίνδυνο την διαθεσιμότητα υπηρεσιών αλλά και την ασφάλεια των πληροφοριών και των δεδομένων των χρηστών. Αναφορικά οι 10 χώρες που το 2022 επλήγησαν σε μεγαλύτερο βαθμό είναι το Βιετνάμ (Vietnam), η Σρι Λάνκα (Sri Lanka), η Σλοβενία (Slovenia), οι Φιλιππίνες

(Philippines), η Πολωνία (Poland), η Τσεχία (Czechia), η Βραζιλία (Brazil), η Βόρεια Μακεδονία (North Macedonia), η Κροατία (Croatia) και η Βουλγαρία (Bulgaria).



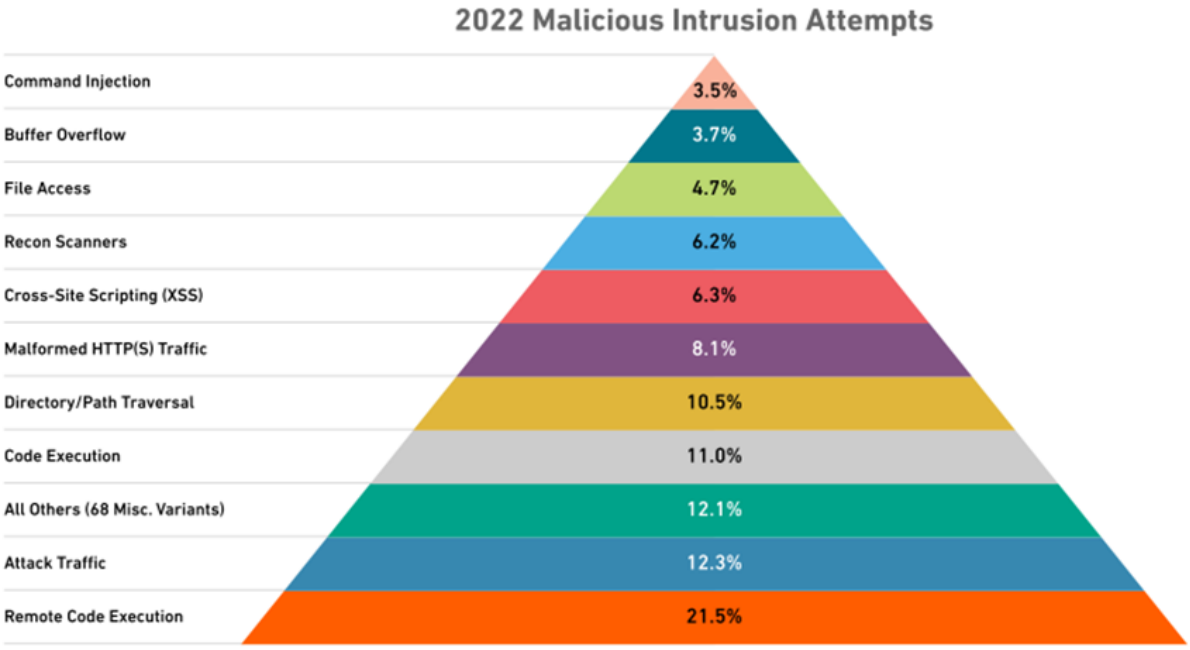
Εικόνα 21 – 10 πιο πλητόμενες χώρες με κακόβουλο λογισμικό σε Παγκόσμιο Επίπεδο, από τη Sonicwall [133].

Ως μια γενική αποτύπωση μπορούν να εξαχθούν τα εξής γενικά συμπεράσματα ως αξιόλογα πρώτον φαίνεται ότι κυμαίνονται σε υψηλά ποσοστά σε παγκόσμιο επίπεδο οι επιθέσεις από κακόβουλο λογιμικό, αντίστοιχα όμως αποτυπώνεται πως η Βόρεια Αμερική έχει αντιμετωπίσει σε σημαντικό βαθμό τις επιθέσεις [133].



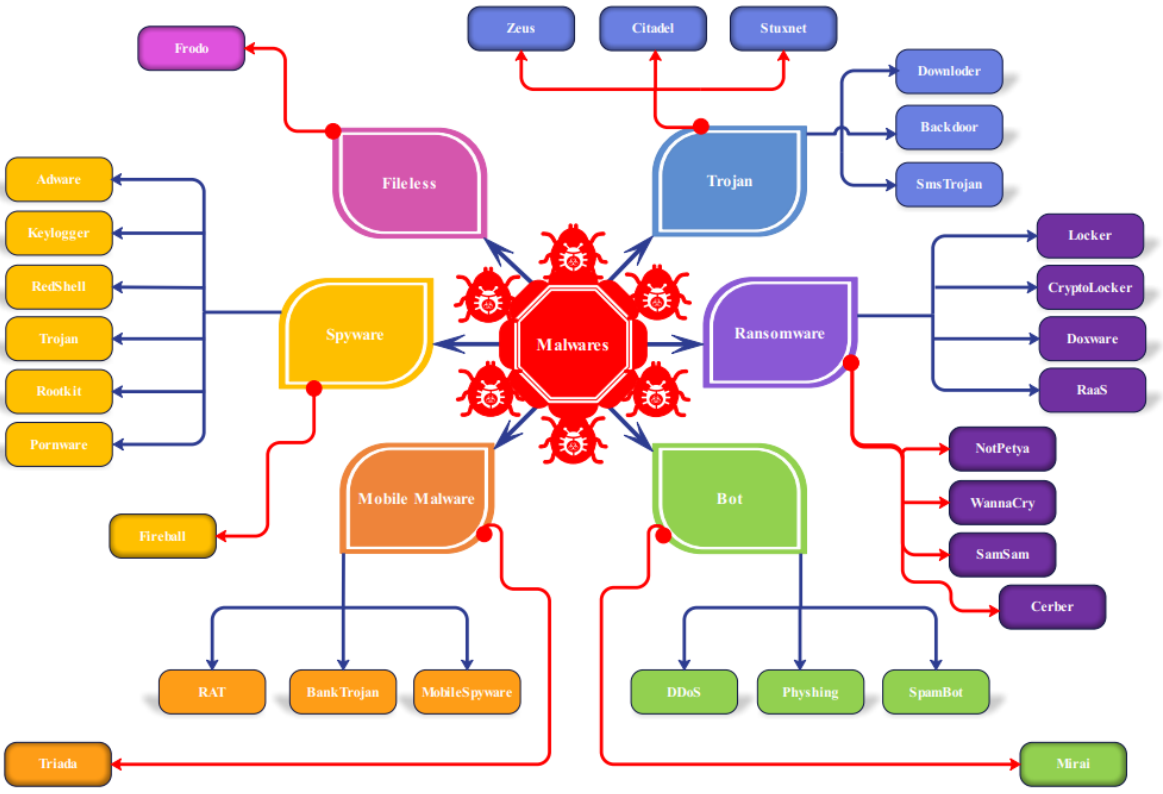
Εικόνα 22 – Παγκόσμια Αποτύπωση Επιθέσεων από κακόβουλο λογισμικό 2021 -2022, από τη Sonicwall [133].

Οι επιθέσεις που πραγματοποιήθηκαν αποσκοπούσαν σε εγκατάσταση μολυσματικού λογισμικού ήταν μέσω απομακρυσμένης εκτέλεσης κώδικα (remote code execution), επίθεσης σε επίπεδο δικτύου (attack traffic), με εκτέλεση κώδικα (code execution), παρεμβαίνοντας στον κατάλογο της διαδρομής (directory – path traversal), μεσω μολυσμένης κίνησης από το διαδίκτυο (malformed http(s) traffic), με εγκάρσια δέσμη ενεργειών (cross-side scripting (xss)), με σαρωτές ανανέωσης (recon scanners), με πρόσβαση σε συστημάτων αρχείων (file access), στοχεύοντας σε υπερχείλιση των προσωρινά αποθηκευμένων δεδομένων (buffer overflow) και μέσω εκτελέσιμων εντολών (command injection).



Εικόνα 23 – Στατιστική Ανάλυση Κακόβουλων Εισβολών, από τη Sonicwall [133]

Στην συνέχεια στην Εικόνα 24, παρουσιάζεται ένα είδος κατηγοριοποίησης και κάποια γνώστα κακόβουλα λογισμικά (malwares).



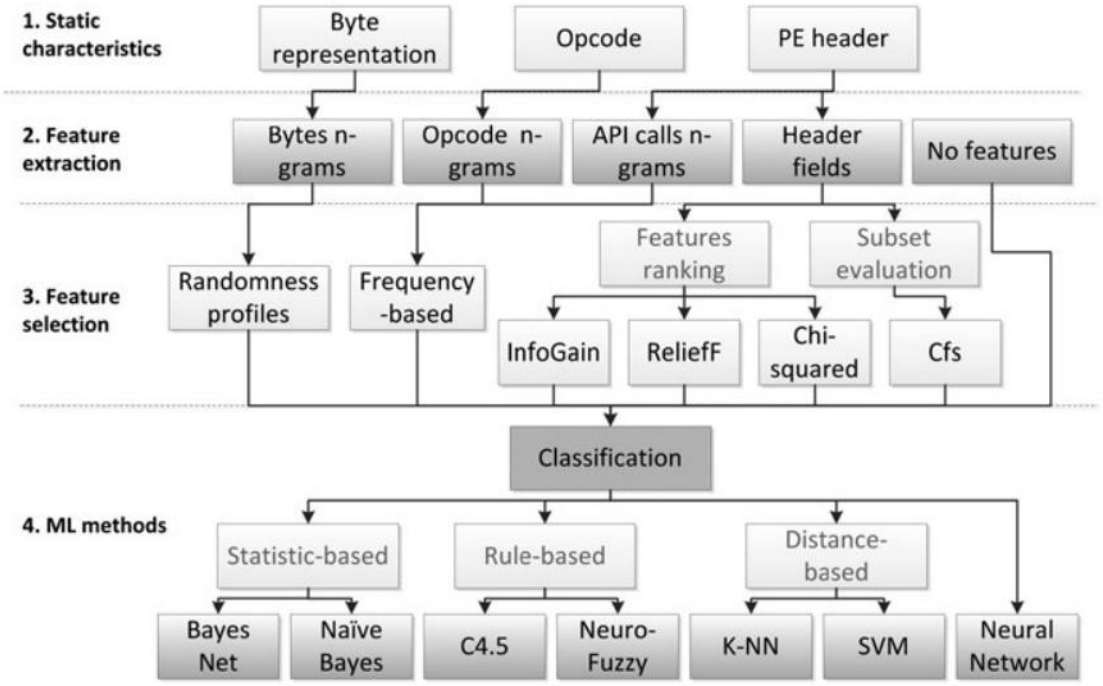
Εικόνα 24 – Ιεράρχιση κακόβουλων λογισμικών [134]

Για την ανίχνευση ενός κακόβουλου λογισμικού, μεταξύ άλλων υπάρχουν κάποιες γενικές αρχές που ακολουθούνται στην κατεύθυνση αυτή. Όπως [54, 55]:

- **Υπολογισμός της τιμής κατακερματισμού (hash value) στα αρχεία**: Ακολουθώντας την πεπατημένη διαδικασία έχοντας βασικές γνώσεις που αφορούν την ανίχνευση και τον εντοπισμό απειλών, από την διαδικασία υπολογισμού των κατακερματισμένων τιμών στα αρχεία.
- **Παρακολούθηση Συστήματος**: Εντοπισμός μη φυσιολογικής δραστηριότητας των υπολογιστικών πόρων είτε σε επίπεδο υλικού είτε σε επίπεδο λειτουργικού συστήματος.
- **Παρακολούθηση Δικτύου**: Ανίχνευση μη φυσιολογικών συνδέσεων σε ένα δίκτυο.

Οι βασικοί τρόποι ανάλυσης είναι δυο, η στατική και η δυναμική [54]:

- **Στατική Ανάλυση**: Αρχικά επιλέγοντας την κατάλληλη τεχνική, πραγματοποιείται η δέουσα αξιολόγηση του εξεταζόμενου προγράμματος εξάγοντας το στην αντίστοιχη δυαδική μορφή, χωρίς να εκτελεστεί. Στην συνέχεια αποτυπώνεται η διαδικασία και ο τρόπος που ακολουθείται σε περίπτωση εκτέλεσης του προγράμματος. Τέλος, συγκρίνεται το εν λόγω μοτίβο με άλλα ευρέως γνωστά μοτίβα που αντιπροσωπεύουν κακόβουλα λογισμικά [54].



Εικόνα 25 – Ιεράρχιση διαδικασίας ανίχνευσης κακόβουλου λογισμικού [70]

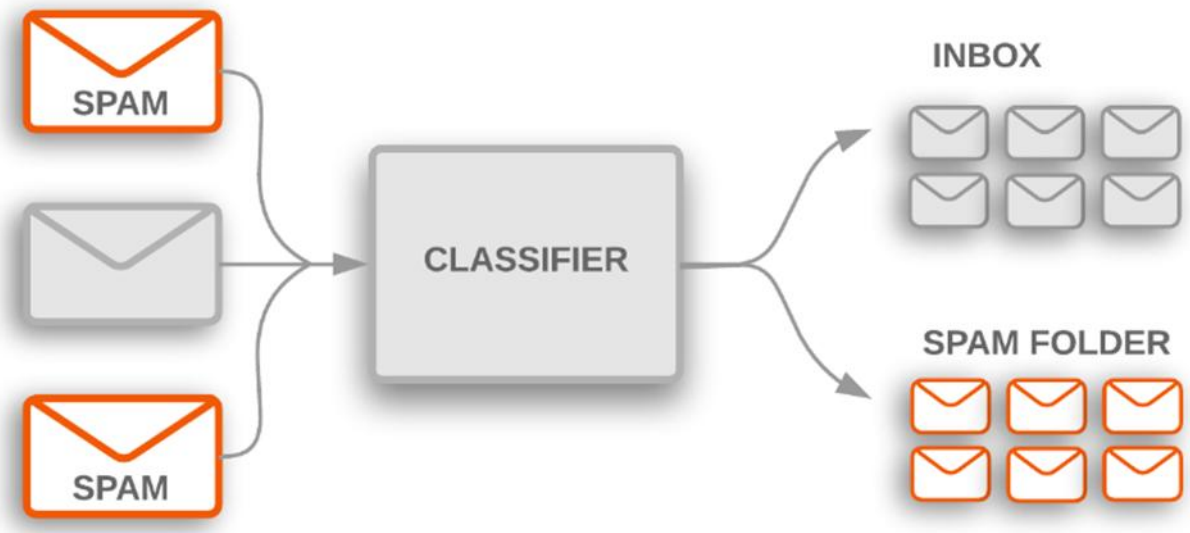
- **Δυναμική Ανάλυση**: Σε αντίθεση με τη στατική ανάλυση, στην δυναμική το εξαγόμενο σε δυαδική μορφή προς εξέταση πρόγραμμα εκτελείται, σε διαμορφωμένα ασφαλή και κατάλληλο περιβάλλον, όπως για παράδειγμα σε ένα εικονικό. Με την διαδικασία αυτή ο αναλυτής ερμηνεύει την δραστηριότητα του προγράμματος κατά την εκτέλεση του, δηλαδή για το αν πραγματοποιείτε, για παράδειγμα λήψη και αποθήκευση κακόβουλων βιβλιοθηκών που δύναται να πλήξουν ένα πληροφοριακό σύστημα [54].

Η ανάγκη για ανάλυση των απειλών που προέρχονται από κακόβουλα λογισμικά ήταν το έναυσμα για την αξιοποίηση της Μηχανικής Μάθησης στην κατεύθυνση αυτή. Η χρήση της μηχανική μάθησης βοηθά τους αναλυτές, ώστε να γίνονται αποδοτικότερα και ταχύτερα οι

πρώτες ενέργειες. Η αυτοματοποίηση των ενεργειών επιτυγχάνεται με την αξιοποίηση των κατάλληλων αλγορίθμων, έχοντας ως σκοπό την ανίχνευση ενός κακόβουλου λογισμικού. Βασική προϋπόθεση για την επίτευξη του ανωτέρου σκοπού είναι η καταγραφή του τρόπου λειτουργίας και των χαρακτηριστικών του λογισμικού σε αντίστοιχες κατηγορίες, ώστε να είναι εφικτή η σύγκριση του με ομοειδή λογισμικά [56].

Με τη μέθοδο της σύγκρισης, εντοπίζονται τα κοινά στοιχεία, οι ομοιότητες που χαρακτηρίζουν κάθε εξεταζόμενο πρόγραμμα. Κατά την διαδικασία εύρεσης των ομοιοτήτων αξιοποιούνται αντιπροσωπευτικά δείγματα χρησιμοποιώντας τους αντίστοιχους αλγορίθμους συμπλέγματος (clustering algorithms), όπως των πλησιέστερων γειτόνων, k – Nearest Neighbors , ή αναδεικνύονται οι ομοιότητες από την χρήση του ίδιου του αλγορίθμου, όπως του k-Means. Επιπρόσθετα υπάρχει και η μεθοδολογία υπολογισμού των αποστάσεων εφαρμόζοντας αντίστοιχες μαθηματικές σχέσεις για την διαδικασία της κατηγοριοποίησης , όπως ο Density – Based Spatial Clustering of Applications with Noise (DBSCAN). Αξίζει να σημειωθεί ότι στην ταξινόμηση για τα κακόβουλα λογισμικά δύναται να χρησιμοποιηθεί και η κατηγοριοποίηση με δενδροειδή λήψη απόφασης (decision trees classification) [54].

4.3 Ανεπιθύμητη Ηλεκτρονική Αλληλογραφία [54]



Εικόνα 26 – Ιεράρχιση διαδικασίας ανίχνευσης κακόβουλου λογισμικού [135]

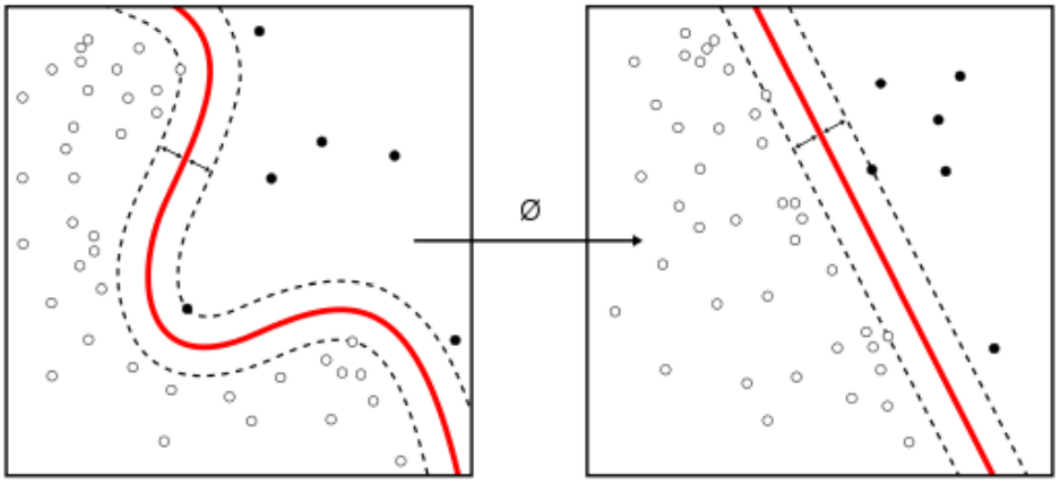
Ο επιτυχής εντοπισμός των ανεπιθύμητων μηνυμάτων αποτέλεσε έναν από τους πρωταρχικούς εφαρμοστέους τομείς χρήσης της Τεχνητής Νοημοσύνης. Το δημοφιλέστερο λογισμικό το οποίο είναι ανοικτού κώδικα ονομάζεται Spam Assasin .

Η αρχή λειτουργίας του εντοπισμού των ανεπιθύμητων μηνυμάτων, βασίστηκε αρχικά σε ένα στατιστικό μοντέλο ανίχνευσης. Με βάση το μοντέλο αυτό, δημιουργείται ένα συνολικό αριθμητικό αποτέλεσμα με βάση, ορισμένες ως ύποπτες λέξεις σε ένα ηλεκτρονικό μήνυμα. Αντίστοιχα έχει καθοριστεί μια κατώτατη τιμή κατωφλίου, οπότε αν το συνολικό αριθμητικό αποτέλεσμα που έχει υπολογιστεί με βάση τον κατάλληλο μηχανισμό, συνάρτηση, ξεπερνά την τιμή κατωφλίου τότε χαρακτηρίζεται, ως ανεπιθύμητο ηλεκτρονικό μήνυμα. Ο μηχανισμός που

δημιουργείται βασίζεται στην διαμόρφωση της αντίστοιχης συνάρτησης υπολογισμού με βάση την γραμμική άλγεβρα. Η μεθοδολογία αυτή ακολουθεί το μοντέλο του Rosenbatt Perceptron.

Σύμφωνα το μοντέλο του Perceptron, το οποίο κατά αναλογία λειτουργεί όπως ο ανθρώπινος νευρώνας, ενεργοποιείται την στιγμή ανίχνευσης του ερεθίσματος που έχει οριστεί. Σε κανονική λειτουργία είναι αδρανής. Άρα όταν το γινόμενο μεταξύ του συντελεστή, k με την τιμή εισόδου είναι ίση ή μεγαλύτερη της τιμής κατωφλίου θ, τότε υπάρχει θετικός σκανδαλισμός και καταχωρείτε στην έξοδο f(ψ) η τιμή +1. Αντίστοιχα αν το γινόμενο είναι μικρότερο από την τιμή κατωφλίου τότε καταχωρείτε το -1. Το μοντέλο αυτό είναι αποδοτικό όταν τα δεδομένα εισόδου χαρακτηρίζονται από γραμμικό διαχωρισμό, έχοντας την αντίστοιχη βαρύτητα εισόδου δεδομένων. Σε αντίθετη περίπτωση, αν τα δεδομένα εισόδου δεν έχουν αντίστοιχο συντελεστή βαρύτητας δεν θα είναι εφικτή η αποδοτική λειτουργία του perceptron.

Για να αντιμετωπιστεί το ζήτημα αυτό αξιοποιήθηκαν τα Support Vector Machines (SVMs) , το οποίο δεν επιδέχεται τους περιορισμούς της γραμμικότητας. Συγκεκριμένα ενώ με το Perceptron ελαχιστοποιούνται τυχόν σφάλματα κατά την κατηγοριοποίηση, ταξινόμηση, στην περίπτωση του SVM μεγιστοποιείται το εύρος.



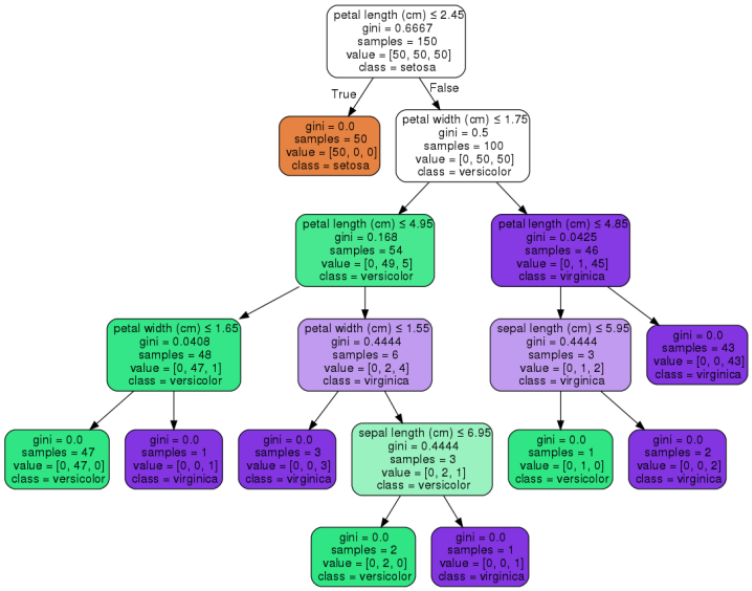
Εικόνα 27 – SVMs → Perceptron

Επιπρόσθετα μπορεί να επιτευχθεί η ανίχνευση ανεπιθύμητης αλληλογραφίας με την χρήση της δενδροείδους λήψης απόφασης, δηλαδή αξιοποιόντας την δυαδική ανάλυση των δεδομένων. Στη πορεία της δενροειδούς ανάπτυξης επιτυγχάνεται η πρόβλεψη ανάμεσα σε δυο ενδεχόμενα, εκφρασμένα με διακριτή και αδιαμφισβήτη αριθμητική τιμή. Η μεθοδολογία που ακολουθείτε για την δενδροειδή λήψη απόφασης μπορεί να αποτυπωθεί στα ακόλουθα στάδια:

1ο. Η διχοτόμηση του αρχικού και κάθε παραγόμενου συνόλου σε υποσύνολα δυαδικής λήψης απόφασης.

2ο. Η διχοτόμηση επαναλαμβάνεται για κάθε υποσύνολο διατηρόντας την αλληλουχία, σύνδεσης - διαδρομής με το προηγούμενο και το επόμενο υποσύνολο.

3ο. Το σύνολο έχει υποδιαιρεθεί σε υποσύνολα τα οποία συνδέονται μεταξύ τους, με τους αντίστοιχους κόμβους. Κάθε υποσύνολο εξαρτιέται από το προηγούμενο κληρονομόντας βασικές ιδιότητες, έχοντας ταξινομημένα τα αντίστοιχα δεδομένα εισόδου.



Εικόνα 28 – Δενδροειδής Λήψη Απόφασης για το Iris dataset [54]

Τέλος αξίζει να σημειωθεί ότι η δενδροειδής λήψη απόφασης είναι ιδιαίτερα αποτελεσματική στην επεξεργασία μεγάλου όγκου δεδομένων.

4.4 Ανίχνευση Δικτυακών απειλών (Network threats)

Τα εργαλεία τεχνητής νοημοσύνης συμβάλουν θετικά στην αναγνώριση και την ερμηνεία της δικτυακής συμπεριφοράς με αυτοματοποιημένο σχεδόν τρόπο. Η δυνατότητα αυτή είναι καθοριστική για την εμπεριστατωμένη αποτύπωση της δικτυακής λειτουργίας και τοπολογίας ενός δικτύου , καθώς και των ενδεχόμενων ευπαθειών και απειλών που μπορεί να εντοπιστούν [70].

Η δικτυακή επίθεση, ως επι το πλείστον έχει στόχο την παρέμβαση στην δρομολόγηση και την κυκλοφορία των δεδομένων, πλήττοντας τις διευθυνσιοδοτήσεις του δικτύου και του αντίστοιχου πρωτοκόλλου του διαδικτύου (Internet Protocol, IP). Σε παγκόσμιο επίπεδο, για την ορθή και αποδοτική επίτευξη της επικοινωνίας και των δρομολογήσεων για την αποστολή των πακέτων δεδομένων μέσω διαδικτύου χρησιμοποιείται το πρωτόκολλο Border Gateway Protocol (BGP). Το εν λόγω πρωτόκολλο έχει ζητήματα ασφάλειας με αποτέλεσμα να είναι εφικτές οι επιθέσεις ανακατευθύνοντας την ροή δεδομένων [59].

Η ενσωμάτωση τεχνολογιών τεχνητής νοημοσύνης στα συστήματα εντοπισμού απειλών σε ένα δίκτυο είναι αποτελεσματικές για την αποτροπή επιθέσεων σε ένα πληροφοριακό σύστημα καθώς και για την διακοπή επιθέσεων που λαμβάνουν χώρα σε αυτό. Συνεπώς λειτουργεί κατά κύριο λόγο προληπτικά αλλά και για την αντιμετώπιση μιας εν εξελίξει επίθεσης [63]. Τα συστήματα ανίχνευσης απειλών με τεχνολογία τεχνητής νοημοσύνης, χρησιμοποιούν στοχευμένα τους ανάλογους αλγορίθμους μάθησης με βάση τις αντίστοιχες διαθέσιμες βάσεις δεδομένων για μεγαλύτερη ακρίβεια στον εντοπισμό ενδεχόμενων απειλών [69].

4.5 Ανίχνευση Ψευδοτυχαίων Ονομάτων Τομέα (Detect Generated Domains, DGA)

Μια διαδεδομένη τακτική των κυβερνο-εγκληματιών για την αποφυγή του εντοπισμού τους, είναι η χρήση ψευδοτυχαίων ονομάτων τομέα. Η τακτική αυτή διασφαλίζει την ανωνυμία στους

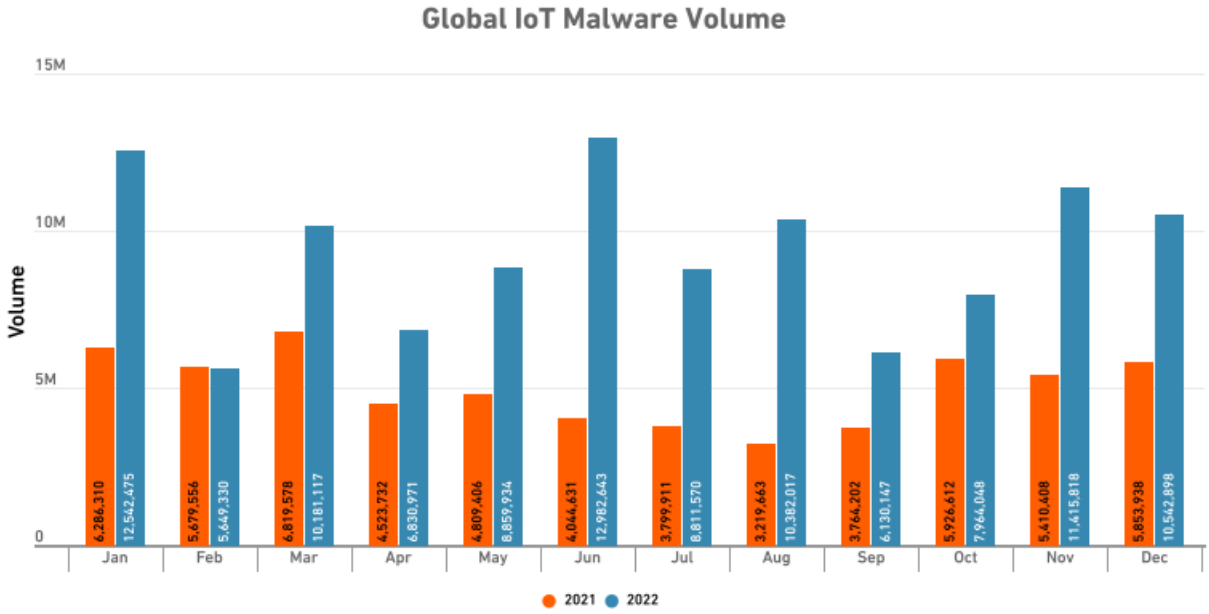
επιτιθέμενους. Η αρχή λειτουργίας τους βασίζεται σε αλγορίθμους που δίνουν την δυνατότητα δημιουργίας χιλιάδων τυχαίων και διαφορετικών ονομάτων τομέα. Η τεχνητή νοημοσύνη, τα νευρωνικά δίκτυα καθώς και η μηχανική μάθηση συμβάλουν στην αντιμετώπιση και την εύρεση των ψευδοτυχαίων ονομάτων τομέα. Οπότε είναι δυνατή η αποτροπή χρήσης αυτών και ένταξη του σε λίστες χαρακτηρίζοντας τα ως επικίνδυνα [71].

4.6 Ανίχνευση χειραγωγούμενου μολυσμένου δικτύου υπολογιστών (Botnet)

Τα χειραγωγούμενα μολυσμένα δίκτυα υπολογιστών (botnet), είναι ένα πλήθος υπολογιστικών πόρων που χρησιμοποιούνται για την διεξαγωγή κακόβουλων επιθέσεων σε ευρεία κλίμακα, όπως για παράδειγμα επιθέσεις άρνησης εξυπηρέτησης (DDos), υποκλοπής προσωπικών δεδομένων και πληροφοριών, η εκτεταμένη επίθεση σε όλο και περισσότερα πληροφοριακά συστήματα, η ανάπτυξη κακόβουλων λογισμικών και αρκετές άλλες κακόβουλες ενέργειες. Σε παγκόσμιο επίπεδο μπορεί να καταταχθεί, ως μια από τις πιο σημαντικές απειλές για οργανισμούς και επιχειρήσεις και μια από τις πιο κερδοφόρες τακτικές των κυβερνοεπιτιθέμενων. Παραδείγματα επιθέσεων αυτού του είδους, mariposa botnet, mirai bot και methbot.

Οι αναλυτές ασφάλειας πληροφοριακών συστημάτων έχουν αναπτύξει διαφορετικές μεθοδολογίες ανίχνευσης των χειραγωγούμενων μολυσμένων δικτύων υπολογιστών. Η ταχεία εναλλαγή των στοιχείων ταυτότητας τους, δικτύωσης τους και εν γένει όλων των χαρακτηριστικών ταυτοποίησης τους, δυσχεραίνει σε σημαντικό βαθμό την αντιμετώπιση αυτού του είδους της απειλής. Στο σημείο αυτό η τεχνητή νοημοσύνη έχει την δυνατότητα να αποτελέσει μέρος της λύσης. Αξιοποιώντας την Μηχανική μάθηση, τα νευρωνικά δίκτυα, αναπτύσσοντας εποπτικούς μηχανισμούς τεχνητής νοημοσύνης, με συνεχή εκμάθηση τους είναι εφικτός ο εντοπισμός τους σε ταχύτερο χρονικό διάστημα και η λήψη αποτρεπτικών μέτρων για την αντιμετώπιση του [151].

4.7 Ανίχνευση Κακόβουλων λογισμικών στο Διαδίκτυο των Πραγμάτων.



Εικόνα 29 – Παγκόσμιο Επίπεδο, από τη Sonicwall [133]

Η ταχεία ανάπτυξη των δικτύων 5G, έξυπνων πόλεων καθώς και η αύξηση των συσκευών που είναι στο διαδίκτυο γενούν νέες δυνατότητες αλλά και σοβαρά ζητήματα ασφάλειας. Ο μέγαλος όγκος των δεδομένων που δημιουργούνται από την χρήση αυτών των συσκευών καθώς και τα στοιχεία δικτύωσης είναι αδύνατο να εποπτευθούν μόνο από τον ανθρώπινο παράγοντα. Για να υπάρξει αυξημένη ασφάλεια και για τον θωρακισμό αυτών είναι απαραίτητη η χρήση της τεχνητής νοημοσύνης και των εργαλείων της [83, 100].

Στην κατεύθυνση μπορούν να αξιοποιηθούν λύσεις που υπάρχουν για την αντιμετώπιση των κακόβουλων λογισμικών, για την ανίχνευση ασυνήθιστης συμπεριφοράς σε ενσύρματα ή και ασύρματα δίκτυα. Η ταχύτατη αξιολόγηση αρχείων καταγραφής και η λήψης απόφασης για το αν είναι κρίσιμα ή μη, σημαντικά ή όχι, αποτελέι εφαλτήριο για την εύρεση πιθανόν ζητημάτων ασφάλειας. Οι λειτουργίες ελέγχου και χρήσης τεχνητής νοημοσύνης είναι απαραίτητες από το σχεδιασμό των δικτύων που υποστηρίζουν το διαδίκτυο των πραγμάτων ώστε να αποδοθεί στον ύψιστο βαθμό η ασφάλεια αυτών [147].

4.8 Ανίχνευση Οικονομικής Απάτης (Fraud Detection)

Οι οικονομικές απάτες μέσω διαδικτύου, είναι σε έξαρση τα τελευταία χρόνια. Οικονομικές απάτες μεγαλύτερης ή μικρότερης κλίμακας αποτελούν ένα παγκόσμιο ζήτημα για την κοινωνικό-οικονομική ισορροπία. Οργανωμένο έγκλημα, διαδικτυακές συναλλαγές, ψηφιακά νομίσματα, μεταφορά χρημάτων, πληθώρα συναλλαγών, υποκλοπή στοιχείων και κωδικών ταυτοποίησης πιστωτικών καρτών, λογαριασμών διαδικτυακών συναλλαγών χρηματοπιστωτικών οργανισμών είναι μερικά μόνο παραδείγματα από ένα ευρύ φάσμα επιθέσεων και εμπράγματων απειλών που στοχεύουν στην οικονομική ωφέλεια ή ζημία φυσικών προσώπων, εταιριών ή οργανισμών.

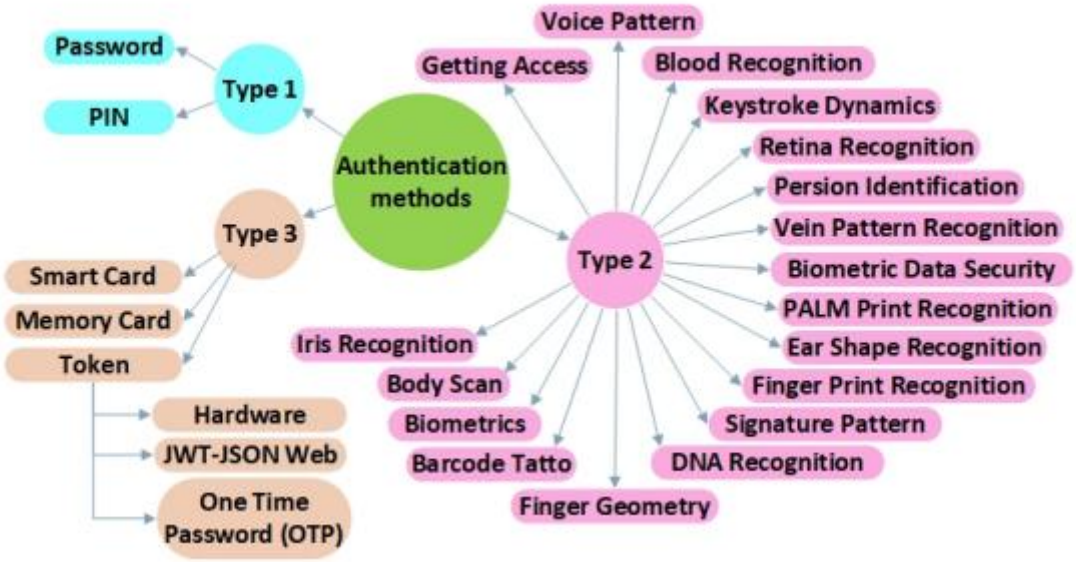
Η ανίχνευση και ο εντοπισμός των παράνομων αυτών οικονομικών δραστηριοτήτων σε πραγματικό χρόνο, αποτελεί μείζων ζήτημα για τις Αρχές Επιβολής του Νόμου σε παγκόσμιο επίπεδο. Συνεπώς, ο υπέρογκος αριθμός αρχείων καταγραφής και συναλλαγών, είναι αποδοτικότερο να τεθεί υπο την διαχείριση και την ανάλυση τεχνολογιών τεχνητής νοημοσύνης και μηχανικής μάθησης. Με τον τρόπο αυτό μπορούν να εντοπισθούν, να αναλυθούν και να αποδοθούν ευθύνες σχετικά με σοβαρά οικονομικά εγκλήματα [54].

4.9 Διαχείριση Ευπαθειών (Vulnerability Management)

Η διαχείριση των ευπαθειών σε ένα πληροφοριακό σύστημα αποτελεί τον ακρογωνιαίο λίθο για την ασφάλεια του. Σε καθημερινή σχεδόν βάση πραγματοποιούνται έλεγχοι για τον εντοπισμό ευπαθειών τόσο από τους επιτιθέμενους όσο και από τους υπερασπιστές των πληροφοριακών συστημάτων. Εκτός από τις κλασικές μεθόδους για την εύρεση ευπαθειών, η καταγραφη της συμπεριφοράς των χρηστών,των διακιμιστών, των εφαρμογών χρήσης με εργαλεία τεχνητής νοημοσύνης συμβάλει στην εξαγωγή σημαντικών συμπερασμάτων με την δημιουργία αντίστοιχων μοτίβων [136].

Σύμφωνα με έρευνα της εταιρίας της IBM, τα συστήματα τεχνητής νοημοσύνης υπερτερούν στην καταγραφή και την διαχείριση των ευπαθειών σε πραγματικό χρόνο. Η αποδοτικότητα αυτών των συστημάτων λειτουργεί αντίστροφως ανάλογα από την αποτελεσματικότητα των επιθέσεων από τους κυβερνοεγκληματίες [63].

4.10 Ασφάλεια



Εικόνα 30 – Μέθοδοι Αυθεντικοποίησης [145]

Η ασφάλεια των πληροφοριακών συστημάτων εξαρτάται και από την μέθοδο αυθεντικοποίησης που ακολουθείται. Προφανώς ανάλογα με το αγαθό και το αντίκτυπο που θα έχει πιθανή εκμετάλλευση αδυναμίας, κάποιας τρωτότητας που έχει, εφαρμόζονται τα αντίστοιχα μέτρα, μέθοδοι αυθεντικοποίησης. Για παράδειγμα, διαφορετικό αντίκτυπο έχει ο τρόπος αυθεντικοποίησης ενός χρήστη που εργάζεται σε πληροφοριακό σύστημα σε πυρηνικό εργοστάσιο και διαφορετικό ενός χρήστη που εργάζεται σε πληροφοριακό σύστημα ενός πανεπιστημίου.

Ως προς την μέθοδο αυθεντικοποίησης υπάρχουν τρεις διαφορετικοί τύποι όπως φαίνεται στην Εικόνα 30, οι οποίες μπορούν είτε να συνυπάρχουν είτε να λειτουργούν κατά μόνας [145].

4.10.1 Ασφάλεια κωδικών μέσω μηχανικής μάθησης

Ένα από τα καίρια ζητήματα στο χώρο της Κυβερνο-ασφάλειας είναι η διασφάλιση των κωδικών ώστε να μην μπορεί κάποιος να τους παρακάμψει. Μια εφαρμόσιμη λύση είναι η πολλαπλή χρήση διαφορετικών τύπων αυθεντικοποίησης ενός χρήστη. Ακόμη καλύτερη η χρήση τεχνολογιών τεχνητής νοημοσύνης ώστε να είναι υπάρχει μεγαλύτερη αξιοπιστία κατά την αυθεντικοποίηση. Χαρακτηριστικό παράδειγμα χρήσης είναι η ανίχνευση προσώπου (Face id) της εταιρίας Apple, όπου αξιοποιώντας τους αισθητήρες υπερύθρων, αντικατοπτρίζονται και αποτυπώνονται οι ιδιότητες, τα χαρακτηριστικά, που υπάρχουν στο πρόσωπο του χρήστη. Στη συνέχεια με βάση το αποτέλεσμα της διαδικασίας μεταγλώττισης του προσώπου σε μια έξοδο, καθίσταται εφικτή ο εντοπισμός ομοιοτήτων με βάση της τεχνολογία τεχνητής νοημοσύνης της Apple [136].

4.10.2 Ασφάλιση των δεδομένων

Στην εποχή μας τα δεδομένα και η πληροφορία έχουν υψηλή αξία για αυτό και γίνονται μήλον της έριδος από τους επιτιθέμενους. Τα κέντρα δεδομένων (Data Centers), αποτελούν ένα κρίσιμο παράγοντα και χρίζουν ιδιαίτερης προστασίας και ασφάλειας. Για τον λόγο αυτό, διαδικασίες ελέγχου που αφορούν τα κέντρα δεδομένων, έχουν αυτοματοποιηθεί. Στην κατεύθυνση αυτή αξιοποιείται και η τεχνολογία τεχνητής νοημοσύνης. Ζητήματα που άπτονται με την

λειτουργία των κέντρων δεδομένων, όπως η κατανάλωση ενέργειας, η επεξεργαστική ισχύς, οι θερμοκρασίες που αναπτύσσονται, γενικά οι περιβαλλοντικές συνθήκες, διαχειρίζονται με εργαλεία τεχνητής νοημοσύνης. Με το πέρας των ετών, δημιουργούνται τεχνολογικές λύσεις τεχνητής νοημοσύνης που βοηθούν στην προστασία και την κυβερνο-ασφάλεια στο σύνολο του κύκλου ζωής των πληροφοριακών συστημάτων καθώς και των εμπλεκομένων με αυτά [63].

Η τεχνητή νοημοσύνη είναι μια επανάσταση της σύγχρονης τεχνολογικής εποχής, η οποία επανακαθορίζει μεθοδολογίες, στρατηγικές, διαδικασίες που ακολουθούνται στο χώρο της Κυβερνο-ασφάλειας. Συνέχως δημιουγούνται εργαλεία που βοηθούν στην κατεύθυνση της προστασίας των δεδομένων, των υπολογιστικών συστημάτων και υποδομών. Αυτοματοποιούνται ανθρώπινες ενέργειες, υπάρχει μεγαλύτερη αποδοτικότητα στην πρόβλεψη, την ανίχνευση και την αντιμετώπιση ευπαθειών, απειλών, επιθέσεων. Θωρακίζεται η κυβερνοασφάλεια με νέες δυνατότητες διαχείρισης κρίσεων εξασφαλίζοντας την Εμπιστευτικότητα, την Ακεραιότητα, την Αυθεντικοποίηση, την Μη αποποίηση ευθυνών και την Διαθεσιμότητα των υπηρεσιών, των υπολογιστικών συστημάτων.

Όμως, όπως όλα τα εργαλεία έτσι και στην τεχνητή νοημοσύνη, δίνεται η δυνατότητα κακόβουλης χρήσης αυτών. Λόγω της δυναμικής εξέλιξης είναι δύσκολο να αποτυπωθούν στο σύνολο τους είτε το σύνολο των εργαλείων, των τομέων όπου επιλύονται ζητήματα κυβερνοασφάλειας όπως αντίστοιχα δεν είναι εφικτή η εν λόγω αποτύπωση των αρνητικών συνεπειών και νέων ζητημάτων κυβερνοασφάλειας που απορρέουν από την κακόβουλη χρήση της Τεχνητής Νοημοσύνης. Στο επόμενο Μέρος Β’ θα γίνει μια προσπάθεια αποτύπωσης αυτών.

ΜΕΡΟΣ Β': Κεφάλαιο 5 – Κυβερνοασφάλεια της Τεχνητής Νοημοσύνης

Η Τεχνητή Νοημοσύνη δίνει λύσεις σε ζητήματα Κυβερνοασφάλειας. Η ασφάλεια ενός πληροφοριακού συστήματος, ή κυβερνο-ασφάλεια στο χώρο του Διαδικτύου επηρεάζεται κατά ένα σημαντικό ποσοστό από το σχεδιασμό του. Κατά το σχεδιασμό και την ανάπτυξη μιας διαδικτυακής εφαρμογής, ενός πληροφοριακού συστήματος, θα πρέπει να λαμβάνεται υπόψιν η ασφάλεια αυτού.

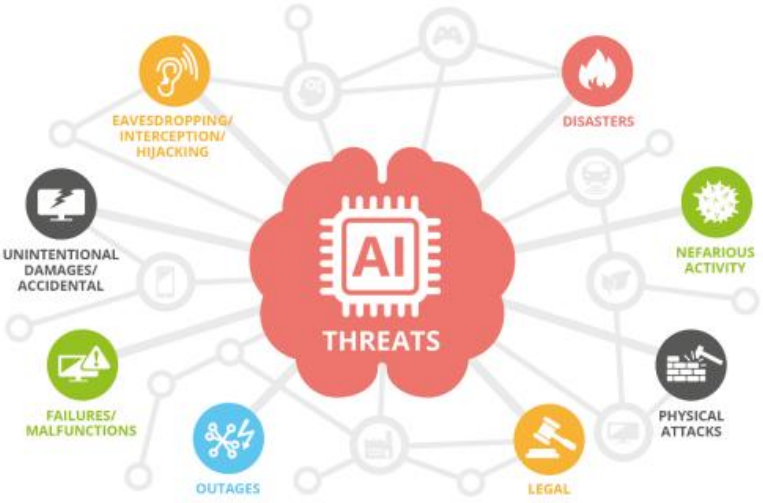
Ένα σύστημα τεχνητής νοημοσύνης από τον σχεδιασμό του θα πρέπει να έχει το ίδιο ένα αυξημένο επίπεδο ασφάλειας. Κατά την ανάπτυξη ενός συστήματος τεχνητής νοημοσύνης μπορούν να διακριθούν κάποια επίπεδα λειτουργίας, όπως του υλικού – της υποδομής, των Δεδομένων , της Μοντελοποίησης και των αλγορίθμων μηχανικής μάθησης. Η ασφάλεια του συστήματος αφορά τόσο στο είδος και την χρήση των αλγορίθμων μάθησης όσο και στα δεδομένα εισόδου που διαχειρίζεται, επεξεργάζεται και μέσα από την εκτέλεση αλγορίθμων και της μηχανικής μάθησης εξάγει ένα αποτέλεσμα. Αρκετά εργαλεία τεχνητής νοημοσύνης βασίστηκαν σε γραμμικούς αλγορίθμους και συναρτήσεις. Σύντομα διαπιστώθηκε όπως ο συντελεστής βαρύτητας που αντιπροσωπεύει τον τρόπο της αλγοριθμικής διαδικασίας λήψης απόφασης έδινε λανθασμένα αποτελέσματα. Οπότε κάθε περίπτωση θα πρέπει να αξιολογείται διαφορετικά ως προς το είδος της τεχνητής νοημοσύνης που δύναται να υλοποιηθεί και να χρησιμοποιηθεί. Για παράδειγμα, υπάρχουν τεχνολογίες τεχνητής νοημοσύνης τα οποία εξάγουν ένα συγκεκριμένο αποτέλεσμα με βάση ένα σύνολο καθορισμένων κανόνων και δεδομένων δίχως να έχουν την δυνατότητα προσαρμογής τους. Επίσης υπάρχουν συστήματα τεχνητής νοημοσύνης τα οποία προσαρμόζονται και συνεχώς βελτιώνονται και εξελίσσονται με βάση τα νέα δεδομένα που προκύπτουν.

Οι διαφορετικές τεχνολογίες, αλγόριθμοι μηχανικής μάθησης, δεδομένα εκμάθησης που αφορούν την τεχνητή νοημοσύνη εγείρουν προβληματισμούς ως προς την ασφάλεια τους συνδυαστικά με τους νέους κινδύνους που μπορούν να επιφέρουν με την κακόβουλη εκμετάλευση τους. Επιπρόσθετα ο εντοπισμός των συστημάτων τεχνητής νοημοσύνης που έχουν δεχθεί επίθεση ή διακυβεύεται η ασφάλεια τους κατά την λειτουργία είναι αρκετά δύσκολο να εντοπιστούν. Για αυτό θα πρέπει να υπάρχει ένα πλαίσιο ασφάλειας από τον σχεδιασμό και την δημιουργία των συστημάτων καθώς και αξιολόγηση αυτών ως προς την ανθεκτικότητα τους [59, 134].

5.1 Απειλές

Κάθε τεχνολογικό επίτευγμα, σύστημα, υλικό, λογισμικό, εφαρμογή θα πρέπει να ελέγχεται ως προς τις απειλές, τις επιπτώσεις και του αντικτύπου που μπορεί να έχουν. Σε ότι αφορά τις απειλές υπάρχουν οι απειλές που οφείλονται από την κακή χρήση των νέων τεχνολογιών που ανακαλύπτονται είτε τις απειλές που έχουν ως στόχο την εκμετάλευση των αδυναμιών των νέων αυτών τεχνολογιών.

Οι απειλές των συστημάτων τεχνητής νοημοσύνης, σύμφωνα με τον Ευρωπαϊκό Οργανισμό ENISA, μπορούν να αποτυπωθούν στις ακόλουθες κατηγορίες [61]:



Εικόνα 31 – Κύριες Απειλές Συστημάτων Τεχνητής Νοημοσύνης [61]

- **Έκνομες Δραστηριότητες (Nefarious Activity):** Κακόβουλες ενέργειες με σκοπό την πρόκληση δυσλειτουργιών, βλαβών σε πληροφοριακά συστήματα, δεδομένα και υποδομές δικτύου. Κάποια ενδεικτικά παραδείγματα:
  - Μη εξουσιοδοτημένη πρόσβαση σε δεδομένα και στα δίκτυα μεταφοράς τους.
  - Οριοθέτηση των αποτελεσμάτων τεχνητής νοημοσύνης.
  - Κατευθυνόμενη εξαγωγή συμπεράσματος με βάση τον επηρεασμό των δεδομένων ή των αλγορίθμων μηχανικής μάθησης.
  - Μολυσμένα Δεδομένα εισόδου.
  - Παραβίαση Δεδομένων εισόδου.
  - Αναβάθμιση Δικαιωμάτων χρήστη.
  - Απειλή εκ των έσω.
  - Επιθέσεις άρνησης υπηρεσιών.

- **Επιθέσεις σε Φυσικό Επίπεδο (Physical Attack):** Επιθετικές ενέργειες εις βάρος φυσικών υποδομών, εξοπλισμού, δικτύων με σκοπό την φθορά, την κλοπή και εν γένει την μη εξουσιοδοτημένη πρόσβαση σε αυτά.
  Κάποια ενδεικτικά παραδείγματα:
  - Δυσλειτουργίες και σφάλματα λόγω χρήσης αμφιλεγόμενων υποδομών και δεδομένων
  - Στοχευμένη επίθεση στις φυσικές υποδομές δικτύων, φυσικών μηχανημάτων (Διακομιστών, Μεταγωγέων, καλωδίων κτλ.)

- **Καταστροφή (Disaster)**: Ολική ή μερική απώλεια υποδομών ως αποτέλεσμα ολέθριου ατυχήματος ή φυσικής καταστροφής. Όπως είναι οι οι φυσικές καταστροφές, δηλαδή σεισμός, πλημμύρα, πυρκαγιά κτλ.

- **Νομοθεσία (Legal)**: Νομικές δεσμεύσεις, όπου καθορίζονται με σαφήνεια οι απαγορεύσεις που ισχύουν και θα πρέπει να ακολουθούνται.
  Κάποια ενδεικτικά παραδείγματα:
  - Παραποίηση ή καταστροφή των χαρακτηριστικών των δεδομένων.
  - Ζητημάτα ιδιωτικότητας κατά την διαχείριση των δεδομένων.
  - Δημιουργία προφίλ χρηστών

- **Συνέπειες (Outages)**: Υποβάθμιση των παρεχόμενων υπηρεσιών ποιότητας δεδομένων.
- **Δεισλειτουργίες (Failures malfunctions):** Ανεπαρκής λειτουργία εξοπλισμού ή λογισμικού.
- **Εξ αμελείας καταστροφές (Unintentional Damages / Accidental):** Εξ αμελείας διάπραξη ατυχημάτων με πρόκληση βλάβης, ολικής ή μερικής καταστροφής εξοπλισμού ή εφαρμογών ή λογισμικών που χρησιμοποιούνται κατά την ομαλή λειτουργία ενός συστήματος.
- **Παρέμβαση στην επικοινωνία δεδομένων (Eaves Dropping / Interception / Hijacking):** Κακόβουλες ενέργειες με στόχο τη διακοπή, την ακρόαση, την υποκλοπή δεδομένων επικοινωνίας.

Στην επόμενη εικόνα παρουσίαζονται και παραδείγματα των απειλών των παραπάνω κατηγοριών.

Εικόνα 32 – Επεξήγηση Κύριων Απειλών Συστημάτων Τεχνητής Νοημοσύνης [61]

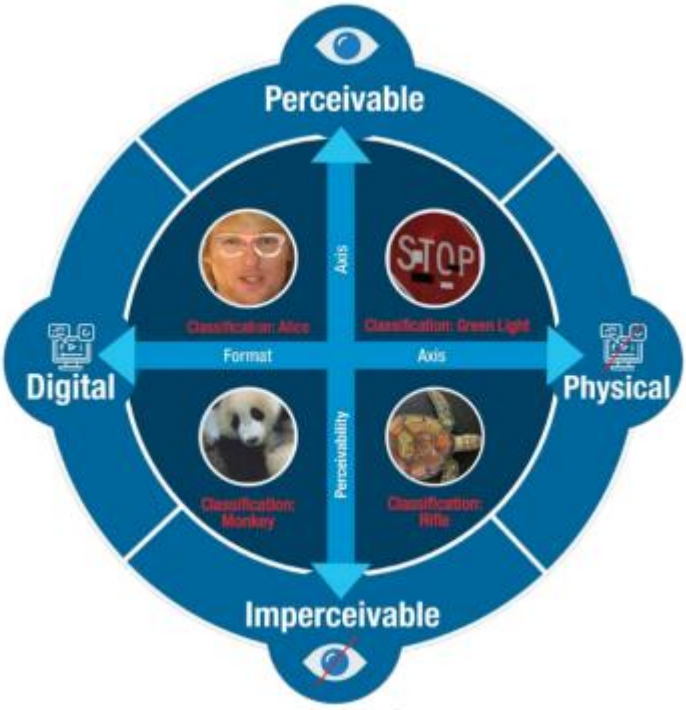
Οι παραπάνω κατηγορίες απειλών με τα σχετικά παραδείγματα που αφορούν την τεχνητή νοημοσύνη είναι είτε γνήσιες απειλές αυτόνομες, όπου η αυθυπαρξία τους σχετίζεται αποκλειστικά και μόνο λόγω της τεχνολογίας της τεχνητής νοημοσύνης ή υβριδικές κατηγορίες όπου ήδη υπάρχοντες απειλές που υφίστανται στο κυβερνοχώρο και τα πληροφοριακά συστήματα ισχύουν και στις τεχνολογίες της τεχνητής νοημοσύνης.

Ακολούθως γίνεται μια επισκόπηση μιας σειρά ενδεικτικών επιθέσεων που αφορούν την τεχνητή νοημοσύνη.

5.2 Κακόβουλη χρήση δεδομένων εισόδου

Όπως συμβαίνει σε κάθε πληροφοριακό σύστημα, έτσι και στην περίπτωση ενός συστήματος τεχνητής νοημοσύνης είναι εν δυνάμει στόχος, δηλαδή μπορεί να δεχθεί οποιοδήποτε είδος επίθεσης υπό κανονικές καθημερινές συνθήκες λειτουργίας. Κατά τη καθημερινή λειτουργία, το εν λόγω σύστημα δέχεται ως πηγή ένα σύνολο δεδομένων. Ο επιτιθέμενος προσβλέπει στην τροφοδότηση του συστήματος τεχνητής νοημοσύνης με αποδεκτά και αναμενόμενα δεδομένα εισόδου τα οποία έχουν διαφοροποιηθεί σε απειροελάχιστο βαθμό σε σημείο που αναγνωρίζονται ως ακίνδυνα για το σύστημα όμως δημιουργούν δυσλειτουργίες και άλλους κινδύνους ως προς την ασφάλεια του συστήματος [70].

Για παράδειγμα αν το σύστημα τεχνητής νοημοσύνης τροφοδοτείται στην είσοδο του με κάποια εικόνα, τότε ίσως να αρκεί ένα διαφορετικό εικονοστοιχείο (pixel) για την επίτευξη της επίθεσης [59]. Όπως γίνεται κατανοητό υπάρχουν αρκετές δυσκολίες για την επίτευξη της ασφάλειας του ίδιου του συστήματος τεχνητής νοημοσύνης, ιδιαίτερα στο ζήτημα που αφορά την πηγή εισόδου του. Οι περιπτώσεις πιθανόν επιθέσεων αυτού του είδους, μπορούν μέχρι ενός βαθμού να αποτυπωθούν με αντίστοιχες μελέτες περιπτώσεων, παρόλα αυτά υπάρχουν δυσκολίες για την απόλυτη πρόβλεψη ενδεχόμενων αδυναμιών που δυνητικά μπορεί να αποτελέσουν την δίοδο για την διεξαγωγή της αντίστοιχης επίθεσης στο σύστημα, μέσω τροποποιημένων δεδομένων εισόδου.



Εικόνα 33 – Επίθεση με βάση τα στοιχεία εισόδου ενός συστήματος Τεχνητής Νοημοσύνης [59]

Μια κατηγοριοποίηση, ταξινόμηση των διαφορετικών περιπτώσεων επίθεσης εισόδου είναι η ακόλουθη [59]:

- Αντιληπτές (Perceivable), όπου οι διαφοροποιήσεις είναι αντιληπτές και ανθρωπίνως ορατές

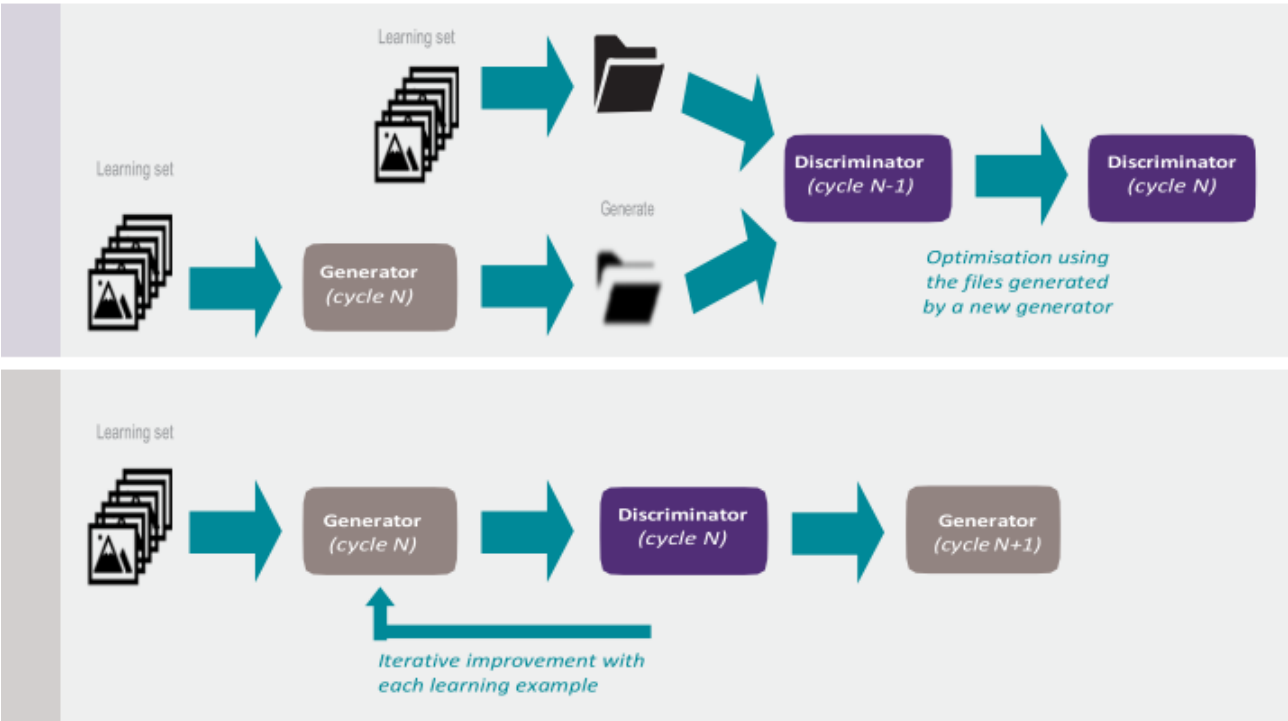
- Μη αντιληπτές (Imperceivable), όπου αλλοιώνονται τα δεδομένα τα οποία διανέμονται δίχως να επηρεάζεται το περιεχόμενο τους.

Συνεπώς θα πρέπει να υπάρχει ένα συνεχής έλεγχος τόσο των δεδομένων εισαγωγής όσο και των αποτελεσμάτων που προκύπτουν. Χρειάζεται ένα είδος ασφάλειας και αυτοματισμού μεταξύ της αλληλουχίας των δεδομένων και των προσδοκώμενων αποτελεσμάτων.

5.3 Δημιουργία Πλαστών - Ψευδών Ειδήσεων (DeepFakes)

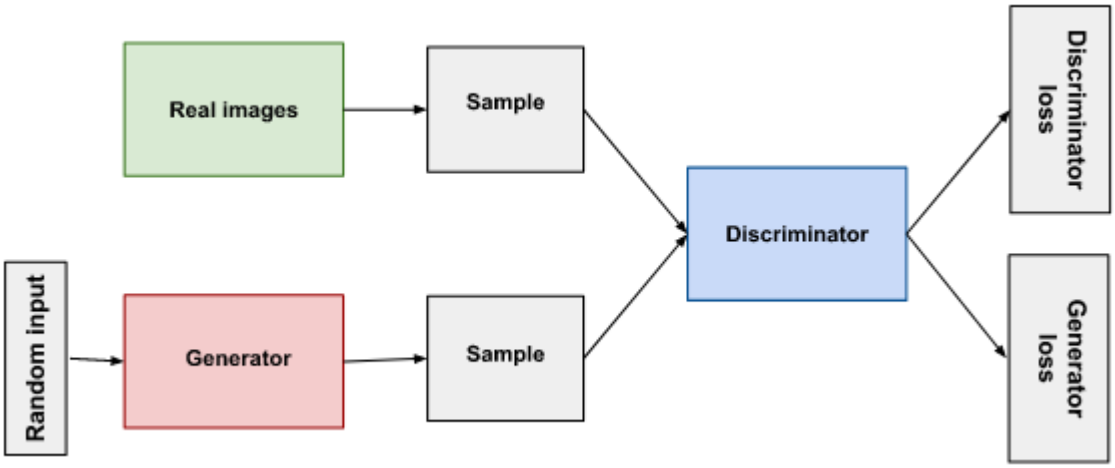
Μια άκρως επικίνδυνη χρήση της τεχνητής νοημοσύνης είναι η δημιουργία ψευδών ειδήσεων, βίντεο, φωτογραφία, πίνακα, ψεύτικων προσώπων, τεχνητής μόδας κτλ. με την χρήση της βαθιάς μάθησης, τα λεγόμενα εις την Αγγλική Γλώσσα 'DeepFakes'. Από το 2017 έως και σήμερα υπάρχει μια εκθετική αύξηση δημιουργίας 'deepfakes'. Το συνηθέστερο συμβάν είναι η εναλλαγή και η τοποθέτηση της εικόνας ενός προσώπου σε ένα βίντεο και αντίστοιχα το άλλο πρόσωπο στο άλλο, υποδύοντας ενέργειες και πράξεις που δεν έχουν υλοποιηθεί από τον ίδιο [59]. Επίσης είναι εφικτή η από το μηδέν δημιουργία νέων εικόνων, έχοντας τυχαία δεδομένα εισόδου οπού στην συνέχεια πραγματοποιείται και έλεγχος για το αν τα δεδομένα και οι εικόνες αντιστοιχούν σε αναγνωρίσιμα πρόσωπα.

Η υλοποίηση αυτού του είδους βασίζεται σε αλγόριθμο κωδικοποίησης, όπου εντοπίζει τα κοινά χαρακτηριστικά των εμπλεκόμενων προσώπων και τα προσομοιάζει στο αντίστοιχο βίντεο με την χρήση του αλγορίθμου του αποκωδικοποιητή.



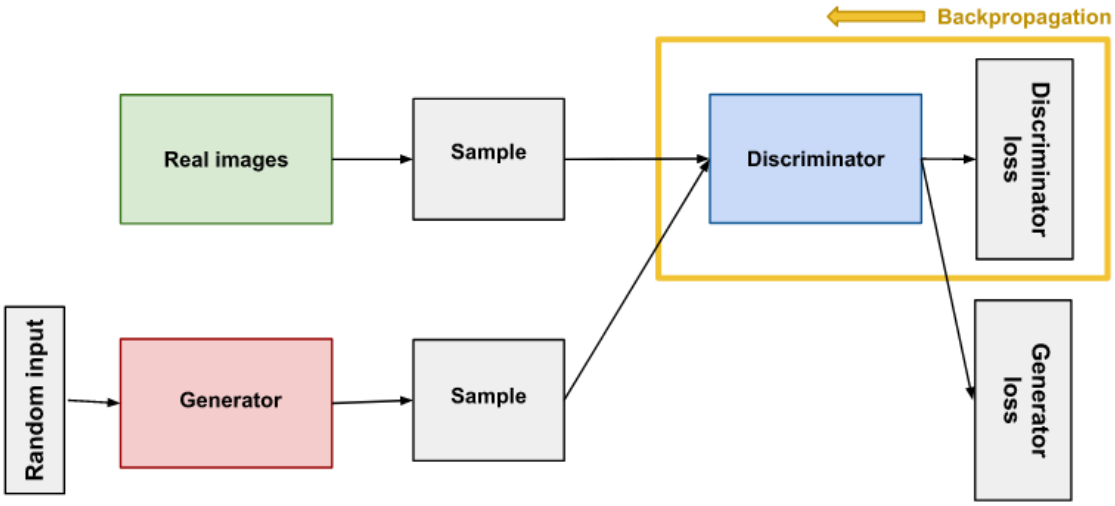
Εικόνα 34 – Deepfakes εικόνες [59]

Συγκεκριμένα υπάρχει μια γεννήτρια δικτύωσης αντίρροπων οντοτήτων (generative adversial network, GAN), όπου αρχικά τροφοδοτείται με τα αναμενόμενα δεδομένα, για παράδειγμα εικόνες. Οι έξοδοι της γεννήτριας (generator), σε συνδυασμό με άλλα δεδομένα που θέλουμε να αντιστραφούν, εισάγονται σε έναν διακριτή (discriminator). Έπειτα οπτικοποιείται η έξοδος των αρχείων με μια νέα εκ νέου γεννήτρια παραγωγής εξόδου, υλοποιώντας τον σκοπό της επίθεσης.



Εικόνα 35 – Διάγραμμα ροής για παραγωγή Deepfakes [152]

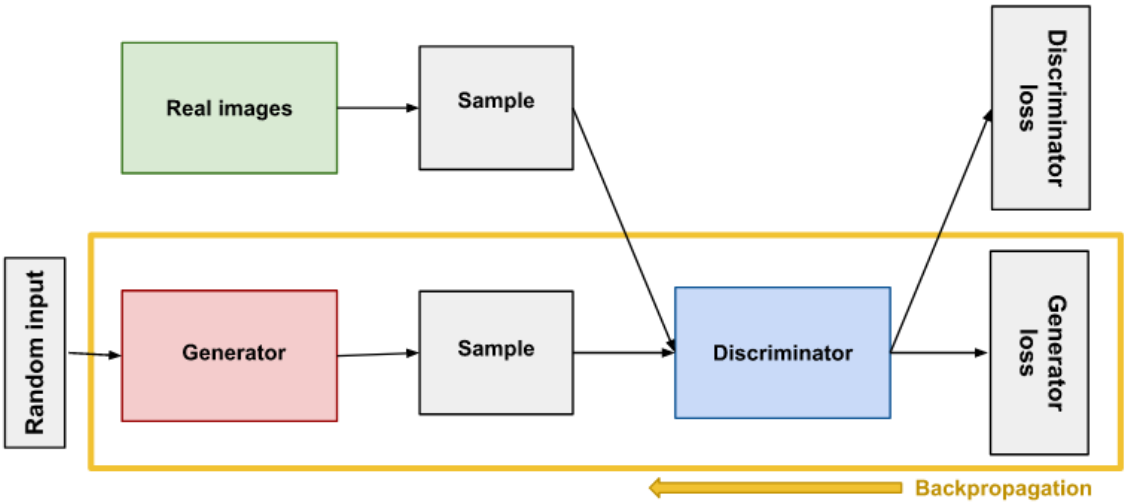
Στην παραπάνω εικόνα φαίνεται ένα διάγραμμα ροής για την δημιουργία των deepfakes. Όπως φαίνεται από τα αριστερά προς τα δεξιά, διακρίνονται τα τυχαία δεδομένα εισόδου (random input), στη συνέχεια έχουμε τη γεννήτρια (generator) και την αναπαράσταση των τυχαίων ή των ψεύτικων δεδομένων που δημιουργήθηκαν ως δείγμα αυτού αντίστοιχα και παράλληλα από την άλλη μεριά υπάρχουν οι αληθινές εικόνες όπου αναπαριστώνται αντίστοιχα από άλλο ένα δείγμα (sample). Στη συνέχεια εισάγονται στο διακριτή (discriminator) τόσο τα δεδομένα που προέρχονται από την αληθινές φωτογραφίες όσο και από την γεννήτρια που είναι οι ψεύτικες, αυτές που θα πλαστογραφήσουν μια αληθινή [152].



Εικόνα 36 – Υποστήριξη (Backpropagation) της εκπαίδευσης του διακριτή (discriminator) [152]

Ο ρόλος του διακριτή είναι η διάκριση μεταξύ των πραγματικών δεδομένων και των ψεύτικων δεδομένων που γεννήθηκαν από την αντίστοιχη γεννήτρια. Ο διακριτής έχει ως είσοδο δεδομένα που προέρχονται και από τις δυο πηγές αληθινών και ψεύτικων και όπως φαίνεται παραπάνω, υπάρχει και η αντίστοιχη υποστήριξη εκπαίδευσης του διακριτή.

Σε ότι αφορά τη ροή της δημιουργίας των ψεύτικών δεδομένων από τη γεννήτρια, όπως φαίνεται στην ακόλουθη φωτογραφία λαμβάνει μέρος στη διαδικασία της παραγωγής των ψεύτικων εξόδων με αντίστοιχη εκπαίδευση της γεννήτριας.



Εικόνα 37 – Υποστήριξη (Backpropagation) της εκπαίδευσης της γεννήτριας (generator) [152]

Όπως γίνεται αντιληπτό οι κίνδυνοι που ελλοχεύουν είναι πολλοί και επηρεάζουν εκ βάθρων τις ζωές των ανθρώπων έχοντας ανεπανόρθωτες συνέπειες σε περίπτωση που δεν εντοπιστούν εγκαίρως. Αξίζει να αναφερθούν κάποιες συγκεκριμένες περιπτώσεις επί του ζητήματος αυτού.

Για παράδειγμα έχουν ήδη πραγματοποιηθεί διάφορες επιθετικές ενέργειες με σκοπό τη κακόβουλη, καθώς και τη ψευδή ταυτοποίηση ενός προσώπου από τους διάφορους εποπτικούς φορείς και συστήματα ελέγχου, όπως είναι τα συστήματα ελέγχου των διαβατηρίων. Επιπρόσθετα έχουν καταγραφεί περιπτώσεις που έγκειται σε ζητήματα πνευματικής ιδιοκτησίας, καθώς ανοικτά προς το κοινό συστήματα τεχνητής νοημοσύνης, χρησιμοποιούνται για την δημιουργία ψεύτικων κειμένων, ποιημάτων, εικόνων με την χρήση της τεχνολογίας τεχνητής νοημοσύνης, όπως το GPT-3.

Όλες αυτές οι κακόβουλες ενέργειες ακόμη και άλλες που δεν έχουν εμφανιστεί ακόμη δίνουν την δυνατότητα στους κυβερνοεγκληματίες να υποδυθούν άλλους, ενισχύοντας την λεγόμενη κοινωνική μηχανική. Με τον τρόπο αυτό προβαίνουν σε υποκλοπή των προσωπικών στοιχείων και δεδομένων ανυποψίαστων ατόμων με σκοπό να τους πλήξουν οικονομικά και ως προσωπικότητες. Επίσης δύναται να αξιοποιηθούν κακόβουλα δίνοντας την ευκαιρία, την δυνατότητα χειραγώγησης των πληθυσμών διασπείροντας ψευδείς ειδήσεις [59].

Συνεπώς δημιουργούνται νέοι κίνδυνοι σε διαφορετικούς τομείς της κοινωνίας, σε στρατιωτικούς, στην δημόσια ασφάλεια, στο γενικό κοινωνικό γίγνεσθαι και αντίστοιχα θα πρέπει να αναπτυχθούν οι αντίστοιχες λύσεις για την αντιμετώπιση αυτών των κακόβουλων ενεργειών [59].

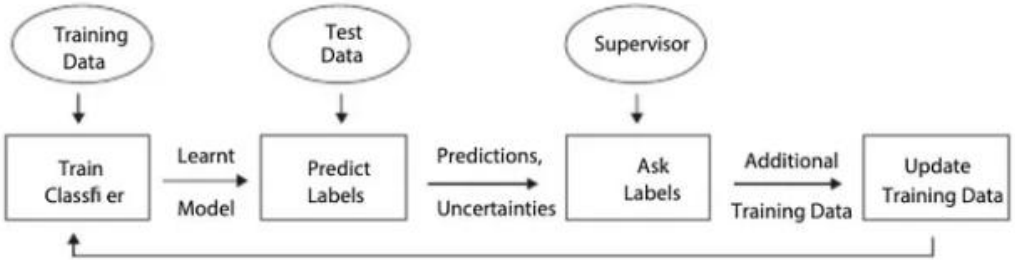
5.4 Παρακάμπτοντας τα CAPTHAS

Στις μέρες μας κάποιες διαδικασίες που δημιουργήθηκαν για την αποτροπή χρήσης των bot, όπως τα captchas είναι δυνατό να παρακαμφθούν. Αξιοποιώντας τις λύσεις της τεχνητής νοημοσύνης μπορούν να παρακαμφθούν αυτού του είδους οι έλεγχοι σε λιγότερο από 0,005 του δευτερολέπτου [59]. Έχουν ήδη δημιουργηθεί ολοκληρωμένες λύσεις ώστε να παρακάμπτονται τα captchas με την χρήση τεχνολογιών τεχνητής νοημοσύνης. Υπάρχουν αρκετές εμπορικές λύσεις που ικανοποιούν την ανάγκη υπεκφυγής των captchas [153].

Η αξιοποίηση της μηχανικής μάθησης, χρησιμοποιεί την μηχανική όραση (computer vision) και τα νευρωνικά δίκτυα με σκοπό την εκπαίδευση του συστήματος τεχνητής νοημοσύνης για την

εύρεση των ζητούμενων αριθμών, ψηφίων, γραμμάτων, επιλογής ζητούμενων εικόνων στα captchas [154].

Ένα χαρακτηριστικό παράδειγμα το οποίο βασίζεται στη μηχανική μάθηση βασίζεται στο OpenCV. Η αρχή λειτουργίας του βασίζεται στην μετατροπή όλων των εμπλεκόμενων εικόνων σε απρόμαυρες. Με τον τρόπο αυτό όλες οι εικόνες επιδέχονται ένα είδος προεπεξεργασίας ώστε να μεταφραστούν σε μαθηματικές τιμές συγκεκριμένα σε μια διακριτή τιμή κατωφλίου, ώστε να μπορεί να χρησιμοποιηθεί η αντίστοιχη μαθηματική τιμή ανά περίπτωση. Στη συνέχεια και με τη χρήση της κατάλληλης λειτουργίας, συνάρτησης, [OpenCV findContour ()] η εικόνα του Captcha διαιρείται στα ανάλογα γράμματα. Συνεπώς οι εικόνες είναι πλέον χαρακτήρες και ψηφία. Έπειτα γίνεται η τροφοδότηση του αντίστοιχου μοντέλου CNN με σκοπό της εκπαίδευση του και την επίλυση, παρακάμψει των CAPTCHAs [154].



Εικόνα 38 – Υποστήριξη (Backpropagation) της εκπαίδευσης της γεννήτριας (generator) [154]

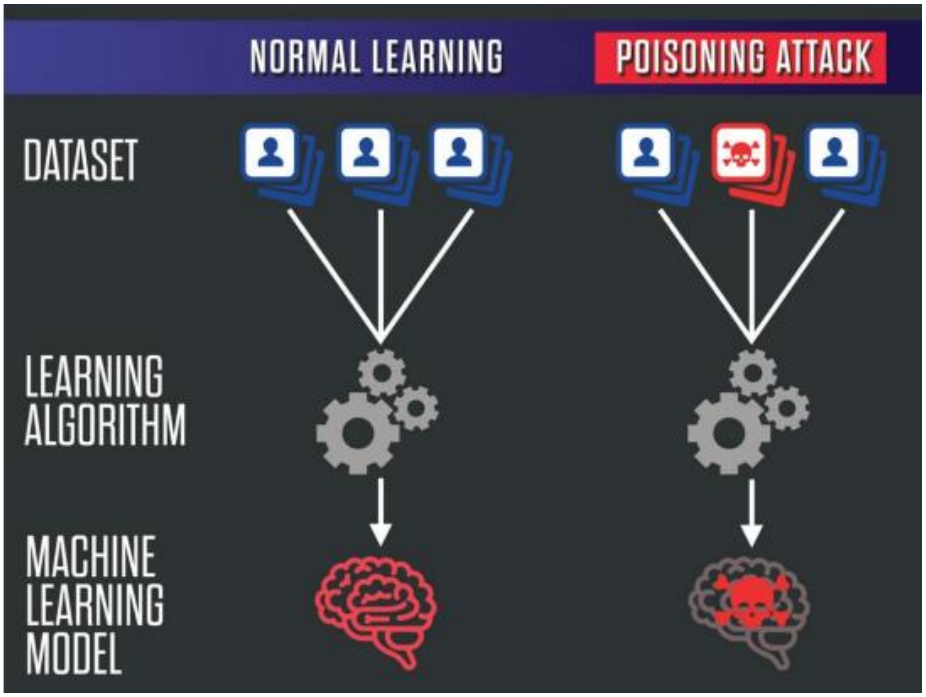
5.5 Επίθεση με μολυσμένα Δεδομένα (Data Poisoning)

Έχοντας ως προϋπόθεση ότι ο επιτιθέμενος έχει πρόσβαση σε ένα μέρος των δεδομένων που χρησιμοποιούνται για την εκμάθηση ενός συστήματος τεχνητής νοημοσύνης, τότε τροφοδοτεί το μέρος αυτό με δεδομένα που μπορούν να μολύνουν το σύστημα [60]. Στόχος του επιτιθέμενου είναι η δημιουργία στοχευμένων σφαλμάτων κατά την λειτουργία και ανάπτυξη του συστήματος, μειώνοντας την συνολική απόδοση του.

Ο επιτιθέμενος μπορεί να τροφοτήσει τουλάχιστον μια φόρα το σύστημα με μολυσμένα δεδομένα επιδιώκοντας τον έλεγχο, ή την καταστροφή ή την μείωση της αποδοτικότητας του συστήματος σε μεταγενέστερο χρονικό διάστημα και όχι αμέσως.

Χαρακτηριστικό παράδειγμα η λειτουργία διαχωρισμού της ηλεκτρονικής αλληλογγραφίας και η κατηγοριοποίηση και η επισήμανση της ανεπιθύμητης αλληλογραφίας. Θα μπορούσε ο επιτιθέμενος να χρησιμοποιεί χαρακτηριστικά, λέξεις, φράσεις που έχουν οριστεί για τον διαχωρισμό των ανεπιθύμητων μηνυμάτων, ώστε να επέμβει στην απόδοση του συστήματος και να εντάσσετε ηλεκτρονική αλληλογραφία στα ανεπιθύμητα ενώ δεν θα έπρεπε [54].

Η κακόβουλη χρήση δεδομένων εισόδου, αποσκοπεί στην παρεμπόδιση ορθής λειτουγίας ενός λειτουργικού συστημάτος τεχνητής νοημοσύνης και μηχανικής μάθησης. Ενώ η κακόβουλη χρήση των μολυσμένων δεδομένα στοχεύει απευθείας το σύστημα κατά την ανάπτυξη του. Συνεπώς κατά την λειτουργία ενός συστήματος, τροφοδοτείται μαζί με τα ορθά και κάποια μολυσμένα δεδομένα, ύποκεινται την αλγοριθμική επεξεργασία και στη συνέχεια μολύνουν το σύστημα –μοντέλο μηχανικής μάθησης της τεχνητής νοημοσύνης [59].



Εικόνα 39 – Σύγκριση ορθής και κακόβουλης μάθησης ενός συστήματος [59].

5.6 Εμκετάλευση πορτών επικοινωνίας

Σε περίπτωση που ο επιτιθέμενος έχει υλοποιήσει την προηγούμενη επίθεση με μολυσμένα δεδομένα τότε στη συνέχεια μπορεί να εκμεταλλευτεί διόδους, πόρτες (backdoors) στο μοντέλο του συστήματος της τεχνητής νοημοσύνης. Οπότε κατά την εκπαίδευση κάποιου μοντέλου μηχανικής μάθησης, είναι δυνατό να εκπαιδευτεί με δεδομένα, όπου οι εν δυνάμει επιτιθέμενοι μπορούν με κεκαλυμμένο τρόπο να εισάγουν ανοικτές πόρτες "backdoors", χρησιμοποιώντας ανάλογα τροποποιημένα και αποδεκτά από το σύστημα τεχνητής νοημοσύνης κακόβουλης υπόστασης δείγματα. Ο επιτιθέμενος με αυτή τη διαδικασία μπορεί να αποκτήσει πρόσβαση σε ένα νευρωνικό δίκτυο και να προβεί σε κακόβουλες ενέργειες που θα πλήξουν την διαθεσιμότητα, την ακεραιότητα και την ασφάλεια του συστήματος [154].

5.7 Ευπάθειες Μηχανικής Μάθησης

Ένας σημαντικός τομέας που αφορά την ασφάλεια των συστημάτων Τεχνητής Νοημοσύνης είναι ο βαθμός ευαλωτότητας και γενικότερα οι ευπάθειες της μηχανικής μάθησης που μπορεί να αποδειχθούν, ως γόνιμο έδαφος για επιθέσεις. Εκτός από τους γνωστούς τρόπους επιθέσεων που έχουν ως συνέπεια να πλήξουν πληροφοριακά συστήματα ένα σημαντικός και κρίσιμος παράγοντας για τα συστήματα τεχνητής νοημοσύνης και μηχανικής μάθσησης είναι τα δεδομένα.

Οι αλγόριθμοι και τα συστήματα μηχανικής μάθησης θα πρέπει να είναι θωρακισμένα για να αντιμετωπίσουν τα ενδεχόμενα σφάλματα που προκύπτουν μετά από υπολογισμούς ή λόγω πιθανής κωδικοποίησης των δεδομένων με στόχο την κακόβουλη αξιοποίησης τους κατά την ανάπτυξη του συστήματος. Όπως γίνεται εύκολα αντιληπτό ένα σύστημα τεχνητής νοημοσύνης είναι πιο ευάλωτο και επιρρεπές σε επιθέσεις. Μέσο του επιτιθέμενου είναι η εκμετάλευση αλγορίθμων μηχανικής μάθησης με στόχο την υποκλοπή δεδομένων, ή την πλήρη αποτύπωση λειτουργίας του συστήματος μηχανικής μάθησης για την αποτύπωση της αρχής λειτουργιας του συστήματος. Αυτό επιτυγχάντεται με την παράνομο εισβολή στο σύστημα, πιθανό με άλλο σύστημα τεχνητής νοημοσύνης για την αποτύπωσης και την καταγραφή της συμπεριφοράς του συστήματος ανάλογα

με τις διάφορες ενέργειες που επιβάλλονται από τον επιτιθέμενο και τα αποτελέσματα που προκύπτουν [59].

5.8 Απειλές κατά την Μοντελοποίηση

Σε συνέχεια των ενδεχόμενων απειλών, συγκαταλέγεται οτιδήποτε και οποιαδήποτε κακόβουλη αλληλουχία δομημένων ενεργειών με στόχο την εύρεση ενδεσχόμενων απειλών σε συστήματα Μηχανικής Μάθησης της Τεχνητής Νοημοσύνης. Κατά τον αρχικό σχεδιασμό ενός πληροφοριακού συστήματος ο σημαντικότερος πλέον παράγοντας είναι η ασφάλεια από τον σχεδιασμό του. Συνεπώς ο σχεδιασμός, το είδος, ο τρόπος, η διαδικασία, το σύστημα ελέγχου που θα επιλεγεί, διαφέρει και είναι άμεσα συνδεδεμένο με την κατηγορία μάθησης που πρόκειται να χρησιμοποιηθεί στο εν λόγω σύστημα. Όπως έχει αναφερθεί η λήψη και η επιλογή των δεδομένων που θα χρησιμοποιηθούν και θα τροφοτούν το σύστημα θα πρέπει να προσδίδουν και διαθέτουν και την αντίστοιχη ασφάλεια για την μηχανική μάθηση.

Οι κατηγορίες των επιθέσεων που αφορούν την μηχανική μάθηση κατηγοριοποιούνται στις ακόλουθες τέσσερις (4) κατηγορίες, έχοντας ως βάση το κίνητρο του επιτιθέμενου, όπως φαίνεται παρακάτω έχουμε τις επιθέσεις εναντίον [59]:

- Της Εμπιστευτικότητας (Confidentiality): Ο επιτιθέμενος επιτυγχάνει τον προσδιορισμό των δεδομένων εισόδου αξιοποιόντας τα δεδομένα που χρησιμοποιήθηκαν για την ανάπτυξη του μοντέλου.
- Της Ακεραιότητας (Integrity): Ο επιτιθέμενος προβαίνει σε δολιοφθορά στην ανάπτυξη και την συμπεριφορά του μοντέλου μέσω επιθέσεων που βασίζονται στο δίκτυο αλλά στόχο έχουν τα δεδομένα μάθησης. Μέσω των δεδομένων μεταβάλλεται και η συμπεριφορά του αντίστοιχου μοντέλου.
- Της Διαθεσιμότητας (Availability): Επιδίωξη του επιτιθέμενου είναι η έλλειψη της διαθεσιμότητας ενός μοντέλου, το οποίο πραγματοποιείται με την εισαγωγή δεδομένων. Τα δεδομένα αυτά μπορεί να φαίνονται ορθά και αποδεκτά αλλά η επίθεση αυτή βασίζεται σε μια μικρή διαφοροποίηση που δημιουργεί μια εσφαλμένη λειτουργία στη διαθεσιμότητα του μοντέλου. Ενώ η μεθοδολογία της επίθεσης συμπίπτει με αυτή της ακεραιότητας, η τεχνική διαφέρει.

Επίσης υπάρχει και η επίθεση που βασίζεται στην αντίστροφη μηχανική διαδικασία, κατά την οποία ο επιτιθέμενος επιδιώκει να αντιγράψει, να αποτυπώσει τον μηχανισμό που έχει αναπτυχθεί. Σημαντικός παράγοντας και η χρήση του επιπέδου των αλγορίθμων. Μέσω αυτών ο επιτιθέμενος μπορεί να λάβει σημαντικές πληροφορίες για το μοντέλο του συστήματος τεχνητής νοημοσύνης.

5.9 Προτεινόμενοι μηχανισμοί ελέγχου και ασφάλειας των συστημάτων τεχνητής νοημοσύνης

Η ασφάλεια ενός πληροφοριακού συστήματος, όπως και ενός συστήματος τεχνητής νοημοσύνης θα πρέπει να έχει διαδικασίες και μηχανισμούς ελέγχου που θα εγγυώνται την ασφάλεια του. Οι μηχανισμοί ελέγχου θα μπορούσαν να περιλαμβάνουν τους [137]:

- **Διοικητικούς**, δηλαδή την δημιουργία και την τήρηση στη πράξη πολιτικών ασφάλειας, με βάση το σχεδιασμό του συστήματος και των προτύπων ασφάλειας που πρέπει να υλοποιούνται

σε αυτό. Έχουν βαρύνουσα σημασία καθώς επιλέγονται οι λειτουργικοί και τεχνικοί μηχανισμοί ελέγχου του συστήματος, των διαχειριστών και των χρηστών αυτού. Στους φορείς, που αξιοποιούν συστήματα τεχνητής νοημοσύνης, εκτός από τα ζητήματα προστασίας προσωπικών δεδομένων που εδράζουν από το Γενικό Κανονισμό Προσωπικών Δεδομένων θα πρέπει να τηρούνται και η κείμενη νομοθεσία που αφορά στην χρήση και λειτουργία συστημάτων τεχνητής νοημοσύνης σύμφωνα με το Νόμο 4961/2022. Μια από τις σημαντικότερες ενέργειες είναι η αλγοριθμική εκτίμηση αντικτύπου, που σε καμιά περίπτωση η πραγμάτωση της προηγούμενης δεν απαλλάσσει από την υποχρέωση διενέργειας και αποτύπωσης εκτίμησης αντικτύπου προστασίας προσωπικών δεδομένων όπως ορίζεται στο Νόμου 4624/2019.

Η αλγοριθμική εκτίμηση αντικτύπου θα πρέπει να προηγείται χρονολογικά πριν από την έναρξη λειτουργίας του συστήματος. Στη μελέτη, αλγοριθμική εκτίμηση θα πρέπει να συγκαταλέγονται ο σκοπός, η αποστολή που έχει να επιτελέσει το σύστημα, τα χαρακτηριστικά που το απαρτίζουν, οι δυνατότητες του, οι παράμετροι του, το είδος των δεδομένων που θα διαχειρίζεται, τα οφέλη που θα υπάρχουν καθώς και οι κίνδυνοι που εγκυμονούν από τη χρήση του.

- **Λειτουργικούς μηχανισμούς,** όπου έχουν ως κύριο μέλημα την ορθή εφαρμογή των πολιτικών και προτύπων ασφάλειας και των
- **Τεχνικούς μηχανισμούς,** που αφορούν την σωστή, ορθή αξιοποίηση των δυνατοτήτων ασφάλειας που αποτελεί το σύστημα τεχνητής νοημοσύνης.

Οι υπεύθυνοι των οργανισμών θα πρέπει να έχουν αναπτύξει τους κατάλληλους μηχανισμούς ελέγχου με αντίστοιχη παρακολούθηση των κινδύνων που ελλοχεύουν στην λειτουργία του συστήματος τεχνητής νοημοσύνης. Επίσης η θέσπιση ενός οργανωτικού και διοικητικού πλαισίου που θα αποσκοπεί στην οργάνωση και την ασφάλεια των πληροφοριών σε ένα οργανισμό ή μια επιχείρηση αποτελεί βαρόμετρο για την λειτουργία και την καλλιέργεια αντίστοιχης κουλτούρας ασφάλειας από τους εργαζόμενους. Οι εργαζόμενοι καθώς και τα εμπλεκόμενα μέρη θα πρέπει να διακατέχονται από υψηλό αίσθημα ευθύνης και υπευθυνότητας, έχοντας πλήρη γνώση σχετικά με τα καθήκοντα και τις αρμοδιότητες τους, που αφορούν την ασφάλεια των πληροφοριών και των πληροφοριακών συστημάτων τεχνητής νοημοσύνης. Όλοι οι διαθέσιμοι υπολογιστικοί και άλλοι πόροι είναι δόκιμο να χαρακτηρίζονται από το προσδοκώμενο και το κατάλληλο επίπεδο προστασίας ώστε να να αποτρέπεται η μη εξουσιοδοτημένη πρόσβαση σε αυτούς. Για αυτό θα πρέπει να υπάρχει ο αντίστοιχος έλεγχος κατά την πρόσβαση των χρηστών ανάλογα με το ρόλο τους σε πληροφορίες και υπολογιστικά συστήματα, εφαρμογές και υπηρεσίες με αντίστοιχη απαγόρευση όσων δεν προβλέπεται να έχουν την αντίστοιχη πρόσβαση. Τα δεδομένα με τα οποία τροφοδοτείται το σύστημα τεχνητής νοημοσύνης, θα πρέπει να έχουν πιστοποιηθεί ως προς την ορθότητα και την ασφάλεια τους. Για τον λόγο αυτό θα πρέπει να υπάρχουν τα αντίστοιχα φίλτρα τα οποία θα εγγυώνται την αυθεντικότητα τους, πιθανό σύμφωνα με την ταυτότητα που θα τα χαρακτηρίζει τα δεδομένα εισόδου τα οποία αξιοποιούνται. Αντίστοιχα λόγω της προγενέστερης αλγοριθμικής εκτίμησης αντικτύπου, είναι σαφές η χρήση κατάλληλων και αμερόληπτων αλγορίθμων για την λήψη απόφασης ανάλογα με το σκοπό του συστήματος τεχνητής νοημοσύνης και της αποστολής του. Επίσης θα πρέπει να υπάρχει μέριμνα ώστε να υλοποιείται επαρκής κρυπτογράφηση τόσο για τα δεδομένα όσο και για τα αποτελέσματα αυτών. Παράλληλα και κατά αντιστοιχία θα πρέπει να διασφαλίζονται ως προς το επίπεδο ασφάλειας και τα υπόλοιπα μέρη του συστήματος, όπως το είδος μηχανικής μάθησης, το μοντέλο που ακολουθείται και οι ενδεχόμενες ευπάθειες αυτών. Σημαντικό ζήτημα είναι και η ασφάλεια των επικοινωνίων και των πληροφοριών

που διακινούνται εντός των δικτυακών υποδομών καθώς και της κατάλληλης και ασφαλής λειτουργία του συστήματος τεχνητής νοημοσύνης.

Επιπρόσθετα θα πρέπει να λαμβάνονται όλα τα μέτρα για τη διαχείριση ενδεχόμενων περιστατικών ασφάλειας, με συγκεκριμένη τήρηση διαδικασιών και ομάδων ασφάλειας που οφείλουν να δράσουν για την αντιμετώπιση τους. Ακόμη θα πρέπει να υπάρχει πρόβλεψη και μέριμνα ώστε να διασφαλίζεται η διαθεσιμότητα των υπηρεσιών που παρέχονται από το σύστημα τεχνητής νοημοσύνης.

Συμπεράσματα

Η εκθετική αύξηση των επιθέσεων με γεωμετρική πρόοδο καθιστούν αναγκαία την χρήση της Τεχνητής Νοημοσύνης στο χώρο της Κυβερνο-ασφάλειας. Επιχειρήσεις, Οργανισμοί και Αρχές Επιβολής του Νομού έχουν την ανάγκη θωράκισης των συστημάτων τους με εργαλεία τεχνητής νοημοσύνης. Όπως παρουσιάστηκε η αξιοποίηση των αλγορίθμων μηχανικής μάθησης και η ενσωμάτωση τους σε συστήματα ασφαλείας είναι αποδοτικότερα στην ανίχνευση κυβερνοαπειλών σε συνδυασμό με τον ανθρώπινο παράγοντα. Είναι ηλίου φαεινότερο, ότι δημιουργείται μια νέα οπτική για τα εμπλεκόμενα μέρη, στην προστασία ζωτικής σημασίας κρίσιμων ή απόρρητων και ευαίσθητων δεδομένων, καθώς και στην θωράκιση των υποδομών και των συμφερόντων τους. Δίνεται η δυνατότητα διαχείρισης και ανάλυσης πληθώρα δεδομένων, εξάγοντας ασφαλή συμπεράσματα για τη μη φυσιολογική συμπεριφορά ενός δικτύου, την αποφυγή και την αντιμετώπιση ενδεχόμενων απειλών, την πρόβλεψη πιθανόν επικείμενων κινδύνων παραβίασης συστημάτων και εν γένει συμβάλει αποδοτικά στην ασφάλεια των δεδομένων, των υπηρεσιών, των εφαρμογών, των πληροφοριακών συστήματων.

Η ανάπτυξη των κοινά αποδεκτών και πιστοποιημένων αλγορίθμων μάθησης, έχοντας λάβει υπόψιν όλες τις νομικές, δεοντολογικές και ηθικές πτυχές είναι το πρώτο βήμα κατά τον σχεδιασμό ενός τέτοιου συστήματος ή εργαλείου. Η δημιουργία ενός καλώς ορισμένου πλαισίου σε Παγκόσμιο, Ευρωπαϊκό και Εθνικό Επίπεδο αποτελεί τον θεμέλιο λίθο για την ορθή λειτουργία των εν λόγω συστημάτων έχοντας ως βασική προϋπόθεση τις Πανανθρώπινες Αξίες δίχως να υπάρχουν μεροληπτικές τάσεις, την Διαφάνεια, την Αξιοκρατία, την Αντικειμενικότητα, τους Θεσμούς και εν τέλει την Δημοκρατία.Η ανάπτυξη των εργαλείων τεχνητής νοημοσύνης για την κυβερνο-ασφάλεια χρειάζεται να έχουν τις αναγκαίες πιστοποιήσεις ασφάλειας με βάση διεθνή πρότυπα. Η προτυποποίηση και η πιστοποίηση των συστημάτων κυβερνο-ασφάλειας είναι απαραίτητη καθώς το ίδιο το εργαλείο από την μια προστατεύει έναντι απειλών ή επιθέσεων από την άλλη θα πρέπει να είναι και το ίδιο ασφαλές, δίχως να τίθεται ζήτημα ασφάλειας για το ίδιο.

Το τεχνολογικό άλμα στην Ακαδημαϊκή Έρευνα θα πρέπει να είναι σε αντιστοιχία με τα κοινωνικά κεκτημένα και τις πανανθρώπινες αξίες που διέπουν τον παγκόσμιο ανθρώπινο πολιτισμό μας. Συνεπώς το χρέος της επιστημονικής κοινότητας είναι η διασφάλιση των πληροφοριών που αφορούν την τεχνητή νοημοσύνη και η αυστηρή χρήση τους έχοντας ως βάση τα ηθικά και δεοντολογικά ζητήματα, με στόχο την αντιμετώπιση των Κυβερνοεγκλημάτων και την αποτροπή τέλεσης αυτών. Η κυβερνοασφάλεια αποτελεί ένα απο τα παγκόσμια και διεθνή ζητήματα όπου απασχολεί τους πάντες από άκρη σε άκρη της εμφιλύου. Τα παράνομα έσοδα που προκύπτουν από κυβερνοεπιθέσεις , ηλεκτρονικές απάτες και κάθε μορφής παράνομης δραστηριότητας στο διαδίκτυο είναι πλέον μεγαλύτερα από πολλές άλλες εκληματικές δραστηριότητες. Το ζήτημα είναι ύψιστης σημασίας, η ορθολογική χρήση της τεχνητής νοημοσυνής μπορεί να δώσει λύσεις για την πάταξη του Κυβερνοεγκλήματος στο διαδίκτυο καθώς

και για την θωράκιση των Πληροφοριακών Συστημάτων Δημόσιων Υποδομών και Δικτύων, Επιχειρήσεων και Οργανισμών.

Η Νέα Εποχή που διανύουμε είναι μια ακόμη πρόκληση για όλους, η γνώση είναι δύναμη και κάθε εργαλείο μπορεί να ενισχύσει την εξελικτική πορεία του ανθρώπου δίνοντας ριξικέλευθρες λύσεις σε μείζονα παγκόσμια ζητήματα, εν προκειμένω στην Κυβερνοασφάλεια. Η προετοιμασία και η ανάπτυξη των εργαλείων τεχνητής νοημοσύνης, θα πρέπει να υλοποιείται σε σωστές βάσεις με σεβασμό στα Δικαιώματα των Ανθρώπων έχοντας, ως κοινή επιδίωξη την Ασφάλεια και την Διαφύλαξη των Ελευθεριών και των Δικαιωμάτων του Ανθρώπου.

Αναφορές

Αναφορά σε βιβλίο:

[1]. «Ασφάλεια Πληροφοριών & Συστημάτων στον Κυβερνοχώρο», (2021), Στέφανος Γκρίτζαλης, Σωκράτης Κάτσικας, Κωνσταντίνος Λαμπρινουδάκης, Εκδόσεις: Νέων Τεχνολογιών, NewTech Pub., ISBN: 978-960-578-064-7, σελίδες 30-32.

Αναφορά σε ιστοσελίδα (link):

[2]. «Τι είναι το Κυβερνοέγκλημα», https://cyberalert.gr/cybercrime/ , προσπελάστηκε 25/01/2023

Αναφορά σε βιβλίο:

[3]. «Κυβερνοέγκλημα», (2021), Γεώργιος Αθ. Γέρμανος & Νικόλαος Χ. Γεωργίου, ISBN: 978-618-00-2651-1, σελίδες 67-77.

Αναφορά σε ιστοσελίδα (link):

[4]. «Έγκλημα στο διαδίκτυο: Μια γενική επισκόπηση», Κατερίνα Μεζίνη, https://www.homodigitalis.gr/posts/12340#1668768932125-6d8e4c25-312d , προσπελάστηκε στις 29/01/2023.

[5]. «Άρθρο 292^{A} – Ποινικός Κώδικας (Νόμος 4619/2019) – Εγκλήματα κατά της ασφάλειας των τηλεφωνικών επικοινωνιών», https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-292a-poinikos-kodikas-nomos-4619-2019-egklimata , προσπελάστηκε στις 29/01/2023.

[6]. «Άρθρο 292^{B} – Πονικός Κώδικας – Παρακώληση λειτουργίας πληροφοριακών συστημάτων», https://www.lawspot.gr/nomikes-plirofories/nomothesia/pk/arthro-292v-poinikos-kodikas-parakolysi-leitoyrgias-pliroforiakon , προσπελάστηκε στις 29/01/2023.

[7]. «Άρθρο $292^{Γ}$ – Ποινικός Κώδικας (Νόμος 4619/2019)», https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-292g-poinikos-kodikas-nomos-4619-2019 , προσπελάστηκε στις 29/01/2023.

[8]. «Άρθρο $292^{Δ}$ – Ποινικός Κώδικας (Νόμος 4619/2019) – Προσβολές του απορρήτου των τηλεπικοινωνιών του κοινού», https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-292d-poinikos-kodikas-nomos-4619-2019-prosvoles , προσπελάστηκε στις 29/01/2023.

[9]. «Άρθρο 292^{E} – Ποινικός Κώδικας (Νόμος 4619/2019) – Παρακώληση των τηλεπικοινωνιών», https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-292e-poinikos-kodikas-nomos-4619-2019-parakolysi , προσπελάστηκε στις 29/01/2023.

[10]. «Άρθρο 346 – Ποινικός Κώδικας (Άρθρο 38 Νόμος 4947/2022) – Εκδικητική πορνογραφία» https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4947-2022/arthro-38-nomos-4947-2022-ekdikitiki-pornografia, προσπελάστηκε στις 29/01/2023.
[11]. «Άρθρο 348 – Ποινικός Κώδικας (Νόμος 4619/2019) – Διευκόλυνση προσβολών ανηλικότητας» , https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-348-poinikos-kodikas-nomos-4619-2019-dieykolynsi, προσπελάστηκε στις 06/02/2023.
[12]. «Άρθρο 348Β – Ποινικός Κώδικας (Νόμος 4619/2019) – Προσέλκυση παιδιών για γενετήσιους λόγους» https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-348v-poinikos-kodikas-nomos-4619-2019-proselkysi, προσπελάστηκε στις 06/02/2023.
[13]. «Άρθρο 363 – Ποινικός Κώδικας – Συκοφαντική Δυσφήμηση» https://www.lawspot.gr/nomikes-plirofories/nomothesia/pk/arthro-363-poinikos-kodikas-sykofantiki-dysfimisi , προσπελάστηκε στις 06/02/2023.
[14]. «Άρθρο 370 – Ποινικός Κώδικας – Παραβίαση του απορρήτου των επιστολών, https://www.lawspot.gr/nomikes-plirofories/nomothesia/pk/arthro-370-poinikos-kodikas-paraviasi-toy-aporritoy-ton-epistolon , προσπελάστηκε στις 29/01/2023.
[15]. «Άρθρο 13 – Ποινικός Κώδικας (Νόμος 4619/2019) – Έννοια όρων του Κώδικα», https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-13-poinikos-kodikas-nomos-4619-2019-ennoia-oron , προσπελάστηκε στις 29/01/2023.
[16]. «Άρθρο 370^{Α} – Ποινικός Κώδικας (Νόμος 5002/2022) – Παραβίαση του απορρήτου τηλεφωνικής και προφορικής συνομιλίας», https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-5002-2022/arthro-10-nomos-5002-2022-paraviasi-aporritoy-tilefonikis , προσπελάστηκε στις 07/02/2023.
[17]. «Άρθρο 370^{Β} – Ποινικός Κώδικας (Νόμος 4619/2019) – Παράνομη πρόσβαση σε σύστημα πληροφοριών ή σε δεδομένα», https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-370v-poinikos-kodikas-nomos-4619-2019-paranomi, προσπελάστηκε στις 07/02/2023.
[18]. «Άρθρο 370^{Γ} – Ποινικός Κώδικας – Παράνομη πρόσβαση σε πληροφοριακό σύστημα», https://www.lawspot.gr/nomikes-plirofories/nomothesia/pk/arthro-370g-poinikos-kodikas-paranomi-prosvasi-se-pliroforiako, προσπελάστηκε στις 07/02/2023.
[19]. «Άρθρο 370^{Δ} – Ποινικός Κώδικας – Παράνομη Υποκλοπή Ψηφιακών Δεδομένων», https://www.lawspot.gr/nomikes-plirofories/nomothesia/pk/arthro-370d-poinikos-kodikas, προσπελάστηκε στις 07/02/2023.
[20]. «Άρθρο 370^{Ε} – Ποινικός Κώδικας (Νόμος 5002/2022) – Παραβίαση μη δημόσιων διαβιβάσεων δεδομένων ή ηλεκτρομαγνητικών εκπομπών» https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-5002-2022/arthro-11-nomos-5002-2022-paraviasi-mi-dimosion , προσπελάστηκε στις 07/02/2023.
[21]. «Άρθρο 370^{ΣΤ} – Ποινικός Κώδικας (Νόμος 5002/2022) – Απαγόρευση διακίνησης λογισμικών και άλλων δεδομένων, https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-5002-2022/arthro-12-nomos-5002-2022-apagoreysi-diakinisis , προσπελάστηκε στις 07/02/2023.
[22]. «Άρθρο 381^{Α} – Ποινικός Κώδικας – Φθορά ηλεκτρονικών δεδομένων» https://www.lawspot.gr/nomikes-plirofories/nomothesia/pk/arthro-381a-poinikos-kodikas-fthora-ilektronikon-dedomenon , προσπελάστηκε στις 07/02/2023.

[23]. «Άρθρο 381^{B} – Ποινικός Κώδικας» https://www.lawspot.gr/nomikes-plirofories/nomothesia/pk/arthro-381v-poinikos-kodikas, προσπελάστηκε στις 07/02/2023.

[24]. «Άρθρο 386^{A} – Ποινικός Κώδικας (Νόμος4619/2019) – Απατη με υπολογιστή» https://www.lawspot.gr/nomikes-plirofories/nomothesia/n-4619-2019/arthro-386a-poinikos-kodikas-nomos-4619-2019-apati-me , προσπελάστηκε στις 07/02/2023.

[25]. «Νόμος 4411/2016 : Κύρωση της Σύμβασης για το έγκλημα στον Κυβερνοχώρο», https://www.e-nomothesia.gr/nomikes-plirofories/n44112016-kyrosi-tis-symvasis-gia-to-egklima-ston-yvernochoro.html, προσπελάστηκε στις 07/02/2023

Αναφορά σε βιβλίο:

[26]. «Ασφάλεια Πληροφοριών & Συστημάτων στον Κυβερνοχώρο», (2021), Στέφανος Γκρίτζαλης, Σωκράτης Κάτσικας, Κωνσταντίνος Λαμπρινουδάκης, Ιωάννης Στέλλιος, Παναγιώτης Κοτζανικολάου, Νινέτα Πολέμη και Χρήστος Δουληγέρης, Εκδόσεις: Νέων Τεχνολογιών, NewTech Pub., ISBN: 978-960-578-064-7, σελίδες 138-140, 145-147.

[27]. "Enisa Threat Landscape 2022", (2022), Ifigenia Lella, Eleni Tsekmezoglou, Rossen Svetozarov Naydenov, Cosmin Ciobanu, Apostolos Malatras, Marianthi Theocharidou, European Union Agency for Cybersecurity (ENISA), ISBN: 978-92-9204-588-3, pages 4-11.

[28]. "Counterterrorism and Cybersecurity", (2013), Newton Lee, Springer Science and Business Media New York Heidelberg Dordrecht London, ISBN: 978-1-4614-7204-9, pages 125-133.

[29]. «Εθνική Στρατηγική Κυβερνοασφάλειας 2020-2025», (2020), Εθνική Αρχή Κυβερνοασφάλειας, Ελληνικής Δημοκρατίας, Υπουργείου Ψηφιακής Διακυβέρνησης, σελίδες 10-13.

[30]. «Ασφάλεια Πληροφοριών & Συστημάτων στον Κυβερνοχώρο», (2021), Στέφανος Γκρίτζαλης, Σωκράτης Κάτσικας, Κωνσταντίνος Λαμπρινουδάκης, Λίλιαν Μήτρου Εκδόσεις: Νέων Τεχνολογιών, NewTech Pub., ISBN: 978-960-578-064-7, σελίδες 74-76, σελίδες 87-91.

Αναφορά σε ιστοσελίδα (link):

[31]. «Κανονισμός (ΕΕ) 2019/881 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου, της 17ης Απριλίου 2019, σχετικά με τον ENISA («Οργανισμός της Ευρωπαϊκής Ένωσης για την Κυβερνοασφάλεια») και με την πιστοποίηση της κυβερνοασφάλειας στον τομέα της τεχνολογίας πληροφοριών και επικοινωνιών και για την κατάργηση του κανονισμού (ΕΕ) αριθ. 526/2013 (πράξη για την κυβερνοασφάλεια) (Κείμενο που παρουσιάζει ενδιαφέρον για τον ΕΟΧ)», (2019), https://eur-lex.europa.eu/legal-content/EL/ALL/?uri=CELEX%3A32019R0881, προσπελάστηκε στις 08/02/2023.

[32]. «Οδηγία (ΕΕ) 2022/2555 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου της 14ης Δεκεμβρίου 2022 σχετικά με μέτρα για υψηλό κοινό επίπεδο κυβερνοασφάλειας σε ολόκληρη την Ένωση, την τροποποίηση του κανονισμού (ΕΕ) αριθ. 910/2014 και της οδηγίας (ΕΕ) 2018/1972, και για την κατάργηση της οδηγίας (ΕΕ) 2016/1148 (οδηγία NIS 2) (Κείμενο που παρουσιάζει ενδιαφέρον για τον ΕΟΧ)», (2022), https://eur-lex.europa.eu/legal-content/EL/TXT/?uri=CELEX:32022L2555, προσπελάστηκε στις 08/02/2023.

Αναφορά σε ιστοσελίδα (link):

[33]. «Γενική Διεύθυνση Κυβερνοασφάλειας», https://mindigital.gr/dioikisi/kyvernoasfaleia προσπελάστηκε στις 08/02/2023.
[34]. «Εθνικό Cert – Εθνική Αρχή Αντιμετώπισης Ηλεκτρονικών Επιθέσεων» https://www.nis.gr/el/national-cert/ , προσπελάστηκε στις 08/02/2023.
[35]. «Ελληνική Αρμόδια Ομάδα Απόκρισης για συμβάντα που αφορούν την ασφάλεια υπολογιστών», https://csirt.cd.mil.gr/el/home/, προσπελάστηκε στις 08/02/2023.
[36]. «Αρχή Διασφάλισης του Απορρήτου των Επικοινωνιών», http://www.adae.gr/, προσπελάστηκε στις 08/02/2023.
[37]. «Αρχή Προστασίας Δεδομένων», https://www.dpa.gr/, προσπελάστηκε στις 08/02/2023.
[38]. «Εθνική Επιτροπή Τηλεπικοινωνιών & Ταχυδρομείων», https://www.eett.gr/ , προσπελάστηκε στις 08/02/2023.
[39]. «Διεύθυνση Δίωξης Ηλεκτρονικού Εγκλήματος», https://cyberalert.gr/, προσπελάστηκε στις 08/02/2023.
[40]. «Διεύθυνση Εγκληματολογικών Ερευνών», https://www.astynomia.gr/elliniki-astynomia/eidikes-ypiresies/diefthynsi-egklimatologikon-erevnon-d-e-e/, προσπελάστηκε στις 08/02/2023.
[41]. «Κέντρο Μελετών Ασφαλείας», http://www.kemea.gr/el/ , προσπελάστηκε στις 08/02/2023.
[42]. «Επιτροπή Εποπτείας και Ελέγχου Παιγνίων», https://www.gamingcommission.gov.gr/, προσπελάστηκε στις 08/02/2023.
[43]. "ENISA- European Union Agency for Cybersecurity", https://www.enisa.europa.eu/ , προσπελάστηκε στις 08/02/2023.
[44]. "European Cybercrime Centre – EC3", https://www.europol.europa.eu/about-europol/european-cybercrime-centre-ec3, προσπελάστηκε στις 08/02/2023.
[45]. "Joint Cybercrime Action Taskforce (J-CAT)", https://www.europol.europa.eu/operations-services-and-innovation/services-support/joint-cybercrime-action-taskforce , προσπελάστηκε στις 08/02/2023.
[46]. " Intellectual Property Crime Coordinated Coalition – IPC3", https://www.europol.europa.eu/about-europol/european-financial-and-economic-crime-centre-efecc/intellectual-property-crime-coordinated-coalition-ipc3, προσπελάστηκε στις 08/02/2023.
[47]. "European Financial and Economic Crime Centre – EFECC", https://www.europol.europa.eu/about-europol/european-financial-and-economic-crime-centre-efecc , προσπελάστηκε στις 08/02/2023.
[48]. "Trustworthy AI: from certification to standardization", https://incyber.org/en/trustworthy-ai-from-certification-to-standardization/, προσπελάστηκε στις 08/02/2023.
[49]. «Αναδυόμενες τεχνολογίες πληροφορικής και επικοινωνιών, ενίσχυση της ψηφιακής διακυβέρνησης και άλλες διατάξεις, Νόμος 4961/2022», https://www.e-nomothesia.gr/kat-demosia-dioikese/nomos-4961-2022-phek-146a-27-7-2022.html , προσπελάστηκε στις 08/02/2023.

Αναφορά σε βιβλίο:

[50]. «Τεχνητή Νοημοσύνη», (2020) Δ' Έκδοση, Ι. Βλαχάβας, Π. Κεφαλάς, Ν.Βασιλειάδης, Φ. Κόκκορας, Η. Σακελλαρίου, Εκδόσεις: Πανεπιστημίου Μακεδονίας, ISBN: 978-618-5196-44-8, σελίδες 4,12,431-440,547-553

[51]. «Νευρωνικά Δίκτυα και Μηχανική Μάθηση», (2010) 3η Έκδοση, Simon Haykin, Απόδοση: Ελένη Γκαγκάτσιου, Εκδόσεις: Παπασωτηρίου, ISBN: 978-960-7182-64-7, σελίδες 8-52

Αναφορά σε παρουσίαση:

[52]. «Εξελικτικοί Αλγόριθμοι Βελτιστοποίησης – Τεχνητά Νευρωνικά Δίκτυα», Δρ. Χασιακός Αθανάσιος, Παναγιώτης Τσίκας, Πανεπιστημίου Πατρών https://eclass.upatras.gr/modules/document/file.php/CIV1756/8-Artificial%20Neural%20Networks%20%28ANN%29.pdf , προσπελάστηκε 09/02/2023.

Αναφορά σε ιστοσελίδα (link):

[53]. "Top 10 Emerging cyber-security threats for 2030", https://www.enisa.europa.eu/news/cybersecurity-threats-fast-forward-2030 προσπελάστηκε στις 12/02/2023.

Αναφορά σε βιβλίο:

[54]. "Hands-On Artificial Intelligence for Cybersecurity", (2019), Alessandro Parisi, Packt Publishing Ltd., ISBN: 978-1-78980-402-7, pages 8-20, 49-148, 150-175, 177-220

[55]. "How to Compete in the Age of Artificial Intelligence", (2018), Soumendra Mohanty Sachin Vyas, Apress, ISBN : 978-1-4842-3807-3, pages 95-120, 143-153, 206 - 214

[56]. "Cyber Threat Intelligence", (2018), Ali Dehghantanha, Mauro Conti, Tooska Dargahi, Springer International Publishing AG, ISBN 978-3-319-73950-2, pages 7-24, 207-214

[57]. "Machine Learning", (1997), Tom M. Mitchell, ISBN 0070428077, pages 2-5.

Αναφορά σε άρθρο:

[58]. "Artificial Intelligence, Human Rights, Democracy and the Rule of Law", (2021) David Leslie, Cristopher Burr, Mhairi Aitken, Josh Cowls, Mike Katell & Morgan Briggs, The Alan Turing Institute, Council of Europe, pages 9-12.

Αναφορά σε βιβλίο:

[59]. "Artificial Intelligence and Cybersecurity", (2021), Lorenzo Pupillo, Stefano Fantin, Afonso Ferreira, Carolina Polito, Centre for European Policy Studies (CEPS), ISBN: 978-94-6138-785-1,pages 12-21, page 30-44, pages 55-75.

[60]. "Securing Machine Learning Algorithms", (2021), Apostolos Malatras, Ioannis Agrafiotis, Monika Adamczyk, European Union Agency for Cybersecurity (ENISA), ISBN: 978-92-9204-543-2, pages 11-17.

[61]. "AI Cybersecurity Challenges", (2020), Apostolos Malatras, Georgia Dede, European Union Agency for Cybersecurity (ENISA), ISBN: 978-92-9204-462-6, pages 13-23

[62]. "German Standardization Roadmap on Artificial Intelligence", (2020), Wolfgang Wahlster, Cristoph Winterhalter, DKE German Commission for Electrical, Electronic & Information Technologies DIN and DVE, pages 81, 144 – 164.

Αναφορά σε άρθρο:

[63]. "The Impact and Limitations of Artificila Intelligence in Cybersecurity: A Literature Review", (2022), Meraj Farheen Ansari , Bibhu Dash , Pawankumar Sharma, Nikhitha Yathiraju, International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), doi: 10.17148/IJARCCE.2022.11912, page 57.

Αναφορά σε ιστοσελίδα:

[64]. «Νόμος 4961/2022 – Αναδυόμενες τεχνολογίες πληροφορικής και επικοινωνιών, ενίσχυση της ψηφιακής διακυβέρνησης και άλλες διατάξεις», https://www.lawspot.gr/nomikes-plirifories/nomothesia/nomos-4961-2022 , προσπελάστηκε 19/02/2023.

[65]. "AI initiatives", Council Of Europe, https://www.coe.int/en/web/artificial-intelligence/national-initiatives , προσπελάστηκε 20/02/2023.

[66]. "Intelligent Systems – Ethics Certification Program for Autonomous and Intelligent Systems (ECPAIS)", της IEEE, https://standards.ieee.org/industry-connections/ecpais/ , προσπελάστηκε 20/02/2023.

Αναφορά σε άρθρο:

[67]. "Human Rights Impact Assessments for AI: Analysis and Recommendations", (2021), Brandie Nonnecke, Philip Dawson, accessnow (Access Now, accessnow.org defends and extends the digital rights of people and communities at risk. As a grassroots-to-global organization, we partner with local actors to bring a human rights agenda to the use, development, and governance of digital technologies, and to intervene where technologies adversely impact our human rights. By combining direct technical support, strategic advocacy, grassroots grantmaking, and convenings such as RightsCon, we fight for human rights in the digital age) pages 5-21

Αναφορά σε ιστοσελίδα:

[68]. "What is Machine Learning?", https://developers.google.com/machine-learning/intro-to-ml/what-is-ml , προσπελάκηε στις 23/02/2023.

Αναφορά σε άρθρο:

[69]. "The future of Artificial Intelligence in Cybersecurity: A Comprehensive Survey", (2021), Feng Tao, Muhammad Shoaib Akhtar, Zhang Jiayuan, Research Article EAI Endorsed Transactions on Creative Technologies, doi: 10.4108/eai. 7-7-2021.170285

Αναφορά σε βιβλίο:

[70]. "AI in Cybersecurity", (2019), Leslie F. Sikos, Springer Nature Switzerland AG, ISBN: 978-3-319-98841-2, pages 115 - 179

[71]. "Artificial Intelligence & Cybersecurity" (2018), Ted Coombs, John Wiley & Sons, Inc. ISBN:978-1-119-50825-0

Αναφορά σε άρθρο:

[72]. "Artificial Intelligence. Overview of the AI Standards Program and Novel Ecosystem Approach ISO/IEC Workshop Series, Inaugural Event", (May 2022), Wael William Diab, Chair SC 42 –Artificial Intelligence, ISO,JTC1,IECM Information Technology Standards, page 13.

[73]. "The Future of Cybersecurity with Artificial Intelligence (AI) and Machine Learning (ML)", (2023), Hrishitva Patel, https://www.preprints.org/manuscript/202301.0115/v1

Αναφορά σε ιστοσελίδα:

[74]. "Cybersecurity's Next Frontier: 80+ Companies Using Artificial Intelligence to Secure The Future In One Infographic", https://www.cbinsights.com/research/cybersecurity-artificial-intelligence-startups-market-map/ , προσπελάστηκε στις 27/02/2023

[75]. "Agari Email Security", https://www.agari.com/ , προσπελάστηκε στις 27/02/2023

[76]. "Bangsun Technology", https://www.bsfit.com.cn/ , προσπελάστηκε στις 27/02/2023

[77]. "Castle", https://castle.io/ , προσπελάστηκε στις 27/02/2023

[78]. "Cybertonica", https://cybertonica.com/ , προσπελάστηκε στις 27/02/2023

[79]. "Datavisor", https://www.datavisor.com/ , προσπελάστηκε στις 27/02/2023

[80]. “Feedzai”, https://feedzai.com/ , προσπελάστηκε στις 27/02/2023
[81]. “GreatHorn”, https://www.greathorn.com/ , προσπελάστηκε στις 27/02/2023
[82]. “Idwall”, https://idwall.co/ , προσπελάστηκε στις 27/02/2023
[83]. “Pulse iD”, https://www.pulseid.com/ , προσπελάστηκε στις 27/02/2023
[84]. “Ravelin”, https://www.ravelin.com/ , προσπελάστηκε στις 27/02/2023
[85]. “Rippleshot”, https://www.rippleshot.com/ , προσπελάστηκε στις 27/02/2023
[86]. “Shift”, https://www.shift-technology.com/ , προσπελάστηκε στις 27/02/2023
[87]. “Ishumei”, https://ishumei.com/ , προσπελάστηκε στις 27/02/2023
[88]. “Sift”, https://sift.com/ , προσπελάστηκε στις 27/02/2023
[89]. “Simility”, https://simility.com/ , προσπελάστηκε στις 27/02/2023
[90]. “Pathmind”, https://pathmind.com/ , προσπελάστηκε στις 27/02/2023
[91]. “Socure”, https://www.socure.com/ , προσπελάστηκε στις 27/02/2023
[92]. “Vu”, https://www.vusecurity.com/es/ , προσπελάστηκε στις 27/02/2023
[93]. “AuthBase”, https://www.authbasenetworks.com/ , προσπελάστηκε στις 27/02/2023
[94]. “Cryptosense”, https://cryptosense.com/ , προσπελάστηκε στις 27/02/2023
[95]. “Cortex”, https://www.paloaltonetworks.com/cortex , προσπελάστηκε στις 27/02/2023
[96]. “GoSecure”, https://www.gosecure.net/ , προσπελάστηκε στις 27/02/2023
[97]. “Devo”, https://www.devo.com/applications/soar/ , προσπελάστηκε στις 27/02/2023
[98]. “Tanium”, https://www.tanium.com/ , προσπελάστηκε στις 27/02/2023
[99]. “Oracle Cloud”, https://www.oracle.com/cloud/ , προσπελάστηκε στις 27/02/2023
[100]. “Cyber Swarm”, https://cyber-swarm.net/, προσπελάστηκε στις 27/02/2023
[101]. “Illusive”, https://illusive.com/ , προσπελάστηκε στις 27/02/2023
[102]. “Bastille”, https://www.bastille.net/ , προσπελάστηκε στις 27/02/2023
[103]. “Cujo AI”, https://cujo.com/ , προσπελάστηκε στις 27/02/2023
[104]. “Sparkcognition”, https://www.sparkcognition.com/ , προσπελάστηκε στις 27/02/2023
[105]. “MTPhelp”,https://help-mtp.appthority.com/WelcomeToMTP.html, προσπελάστηκε στις 27/02/2023
[106]. “Zimperium”,https://www.zimperium.com/ , προσπελάστηκε στις 27/02/2023
[107]. “Sentegity”,https://sentegrity.com/ , προσπελάστηκε στις 27/02/2023
[108]. “Blackberry”,https://www.blackberry.com/us/en/products/cylance-endpoint-security/cylance-is-blackberry-cybersecurity , προσπελάστηκε στις 27/02/2023
[109]. “Deepinstict”,https://www.deepinstinct.com/ , προσπελάστηκε στις 27/02/2023
[110]. “Ideni”, https://indeni.com/ , προσπελάστηκε στις 27/02/2023
[111]. “Innefu”, https://www.innefu.com/ , προσπελάστηκε στις 27/02/2023
[112]. “Cyr3con”, https://www.cyr3con.ai/ , προσπελάστηκε στις 27/02/2023
[113]. “Sumologic”, https://www.sumologic.com/ , προσπελάστηκε στις 27/02/2023
[114]. “Logrhythm”, https://logrhythm.com/ , προσπελάστηκε στις 27/02/2023
[115]. “Protenus”, https://www.protenus.com/ , προσπελάστηκε στις 27/02/2023
[116]. “Seclytics”, https://seclytics.com/ , προσπελάστηκε στις 27/02/2023
[117]. “Sentinelone”, https://www.sentinelone.com/ , προσπελάστηκε στις 27/02/2023
[118]. “Api threatstream”, https://api.threatstream.com/ , προσπελάστηκε στις 27/02/2023
[119]. “Cybersaint”,https://www.cybersaint.io/ , προσπελάστηκε στις 27/02/2023
[120]. “Cytora”, https://cytora.com/ , προσπελάστηκε στις 27/02/2023
[121]. “Haystax”, https://haystax.com/ , προσπελάστηκε στις 27/02/2023
[122]. “Metacert”, https://metacert.com/ , προσπελάστηκε στις 27/02/2023

[123]. “Awarehq”, https://www.awarehq.com/ , προσπελάστηκε στις 27/02/2023
[124]. “Behaviosec”, https://www.behaviosec.com/ , προσπελάστηκε στις 27/02/2023
[125]. “Darktrace”, https://darktrace.com/ , προσπελάστηκε στις 27/02/2023
[126]. “Exabeam”, https://exabeam.com/, προσπελάστηκε στις 27/02/2023
[127]. “Intensityanaltics”, https://intensityanalytics.com/, προσπελάστηκε στις 27/02/2023
[128]. “Perimeterx”, https://www.perimeterx.com/, προσπελάστηκε στις 27/02/2023
[129]. “Paloaltonetworks”, https://www.paloaltonetworks.com/prisma/cloud, προσπελάστηκε στις 27/02/2023
[130]. “Sphericaldefence”, https://sphericaldefence.com/, προσπελάστηκε στις 27/02/2023
[131]. “Stackpath”, https://www.stackpath.com/, προσπελάστηκε στις 27/02/2023
[132]. “Twosense”, https://www.twosense.ai/, προσπελάστηκε στις 27/02/2023

Αναφορά σε άρθρο:

[133]. “2023 Sonicwall Cyber Threat Report | Charting Cybercrime’s Shifting Frontlines”, https://www.sonicwall.com/2023-cyber-threat-report/ , 06/03/2023
[134]. “Artificial Intelligence – Based Malware Detection, Analysis and Mitigation”, (2023) Amir Djenna, Ahmed Bouridane, Saddaf Rubab and Ibrahim Moussa Marou, Symmetry 2023, 15, 677. https://doi.org/10.3390/sym15030677

Αναφορά σε ιστοσελίδα:

[135]. “Machine Learning”, https://developers.google.com/machine-learning/guides/text-classification , προσπελάστηκε 22/03/2022.

Αναφορά σε άρθρο:

[136]. “Artificial Intelligence in Information and Cyber Security”, (2021), Vamsi Krisma Vedantam,https://www.researchgate.net/publication/349350306_Artificial_Intelligence_in_Information_and_Cyber_Security

Αναφορά σε βιβλίο:

[137]. «Ασφάλεια Υπολογιστών Αρχές και Πρακτικές», (2016) 3η Αμερικανική Έκδοση William Stallings, Lawrie Brown, Μετάφραση: Γιώργος Στάμου, Επιμέλεια κειμένου: Παναγιώτης Αρκουδέας, Εκδόσεις: Κλειδάριθμος, ISBN: 978-960-461-668-8
[138]. “Application of Artificial Intelligence in Cyber Security”, (2018), Dr. Sunil Bhutada,[2]Preeti Bhutada Professor, Department of Information Technology, Sreenidhi Institute of Science and Technology Faculty, Shri Shakti, International Journal of Engineering Research in Computer Science and Engineering (IJERCSE), ISSN (Online) 2394-2320
[139]. Enisa Cybersecurity Market Analysis Framework (ECSMAF), (2022), Louis Marinos, Domenico Ferrara, Silvia Portesi, Eleni Tsekmezoglou, European Union Agency for Cybersecurity (ENISA), ISBN: 978-9204-561-6
[140]. “Security and Privacy of Public DNS Resolvers”, (2022), Evangelos Kantas, Marnix Dekker, European Union Agency for Cybersecurity (ENISA), ISBN: 978-9204-558-6
[141]. “Remote Identity Proofing: Attacks & Countermeasures”, (2022), Viktor Paggio, Evgenia Nikolouzou, Marnix Dekker, European Union Agency for Cybersecurity (ENISA), ISBN: 978-92-9204-549-4
[142]. “Cyber Europe 2022: After Action Report” (2022), Nikolas Cristoforatos, Ifigenia Lella, Evangelos Rekleitis, Cristian Van heurck, Alexandros Zacharis, ISBN: 978-92-9204-606-7

[143]. “Nis Investments”, (2022), Athanasios Drougkas, Viktor Paggio, Javier Gomez Prieto, Patrick Abel, François Gratiolet, Edwin Maaskant, Gartner, European Union Agency for Cybersecurity (ENISA), ISBN: 978-92-9204-585-2

Αναφορά σε άρθρο:

[144]. “Ai in cybersecurity market to be worth $60.6 billion by 2028:Report”, (2022), CIOSEA News, ETCIO SEA

[145]. “Cybersecurity in the AI-Based Metaverse: A Survey”, (2022) Mitra Pooyandeh, Ki-Jin Han and Insoo Sohn, Multidisciplinary Digital Publishing Institute

[146]. “Ai and automation for cybersecurity”, (2022), Dr. Sridhar Muppidi, Lisa Fisher, Gerald Parham, IBM Corporation 2022

[147]. “Iot Threats & Implementation of Ai/Ml to address emerging cyber security issues in iot with cloud computing”, (2023), Sr. SAP Basis Cloud Architect, Raley’s, Sacramento, California, USA, DOI : https://www.doi.org/10.56726/IRJMETS32866

[148]. “Cyber Threat Report”, (2022), Bill Conner, Sonicwall Inc.

[149]. “2023 – 2027 Strategic Technology Roadmap version 5”, (2022), Cybersecurity and Infrastructure Security Agence, CISA, https://www.cisa.gov/publication/strategic-technology-roadmap-overview

[150]. “Detecting Cybersecurity Attacks in internet of things Using Artificial Intelligence Methods: A Systematic Literature Review”, (2022), Mujaheed Abdullahi, Yahia Baashar, Hitham Alhussian, Ayed Alwadain, Norshakirah Aziz, Luiz Fernando Capretz and Said Jadid Abdulkadir, Multidisciplinary Digital Publishing Institute (MDPI), https://doi.org/10.3390/electronics11020198

[151]. “Botnet Detection Using Artificial Intelligence”, (2023), R. Sri Skandha Moorthy, N. Nathiya, International Conference on Machine Learning and Data Engineering, School of Advanced Sciences, Division of Mathematics, Vellore Institute of Technology Chennai,Chennai,600 127,India., Procedia Computer Science 218 (2023) 1405–1413

Αναφορά σε ιστοσελίδα:

[152]. “Overview of GAN Structure”, https://developers.google.com/machine-learning/gan/gan_structure , προσπελάστηκε 18/05/2023.

[153]. “CaptchaAI”, https://captchaai.com/ , προσπελάστηκε στις 18/05/2023.

[154]. “Understanding the AI Model that Broke CAPTHCAs” , https://towardsdatascience.com/understanding-rcns-structure-ec4b51b9c257 , προσπελάστηκε στις 18/05/2023.

[155]. “The Adversarial Robustness Toolbox v0.30.: Closing the Backdoor in AI security” https://www.ibm.com/blogs/research/2018/08/art-v030-backdoor/ , προσπελάστηκε στις 18/05/2023.