ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΙΓΑΙΟΥ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ
ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Η Τεχνητή Νοημοσύνη για την Κυβερνοασφάλεια και η Κυβερνοασφάλεια για την Τεχνητή Νοημοσύνη

-AI for Cybersecurity and Cybersecurity for AI-

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

των

ΤΡΙΑΝΤΑΦΥΛΛΟΥ ΧΡΗΣΤΟΥ

και

ΔΑΜΙΓΟΥ ΓΕΩΡΓΙΟΥ

Επιβλέπουσα: ΔΙΑΜΑΝΤΟΠΟΥΛΟΥ ΒΑΣΙΛΙΚΗ

Σάμος, 2025

Ευχαριστίες

Η παρούσα Διπλωματική Εργασία εκπονήθηκε κατά την χειμερινή περίοδο του Ακαδημαϊκού Έτους 2024-2025, στα πλαίσια του Προπτυχιακού Προγράμματος Σπουδών του Πανεπιστήμιου Αιγαίου του τμήματος Μηχανικών Πληροφοριακών & Επικοινωνιακών Συστημάτων.

Καταρχάς, θα θέλαμε να εκφράσουμε βαθιά ευγνωμοσύνη στην επιβλέπουσα καθηγήτριά, Δρ. Βασιλική Διαμαντοπούλου, για την αμέριστη βοήθειά της, τις πολύτιμες συμβουλές της και την συνεχή ενθάρρυνσή της καθ' όλη την διάρκεια της συγγραφής. Η εμπειρία και η γνώση της αποτέλεσαν φάρο καθοδήγησης σε αυτό το απαιτητικό ταξίδι.

Ιδιαίτερες ευχαριστίες οφείλουμε στις οικογένειές μας και τους φίλους μας, για την αστείρευτη υπομονή, την κατανόηση και την ηθική υποστήριξη που μας παρείχαν σε όλη την διάρκεια των σπουδών μας.

Τέλος, ευχαριστούμε όλους όσους, με οποιοδήποτε τρόπο, συνέβαλαν στην ολοκλήρωση αυτής της εργασίας.

Περιεχόμενα

1. Εισαγωγή 7

1.1 Θεωρητικό Υπόβαθρο και Κίνητρα Έρευνας 7

1.1.1. Επισκόπηση της Τεχνητής Νοημοσύνης (ΤΝ) 8

1.1.2. Επισκόπηση της Κυβερνοασφάλειας 9

1.1.3. Η Σημασία και η Αναγκαιότητα της Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια 10

1.2 Σκοπός, Στόχοι, Πεδίο και Περιορισμοί της Έρευνας 11

1.2.1. Διατύπωση Ερευνητικών Ερωτημάτων και Στόχων 12

1.2.2. Καθορισμός του Πεδίου και των Περιορισμών της Έρευνας 13

1.3 Δομή της Διπλωματικής Εργασίας 15

2: Τεχνητή Νοημοσύνη στην Κυβερνοασφάλεια 18

2.1 Θεμελιώδεις Τεχνολογίες και Μεθοδολογίες 18

2.1.1. Μηχανική Μάθηση (ΜΜ) 18

2.1.2. Βαθιά Μάθηση (ΒΜ) 20

2.1.3. Ανίχνευση Κακόβουλου Λογισμικού μέσω ΒΜ 20

2.1.4. Ανάλυση Δικτύων και Ανίχνευση Ανωμαλιών 21

2.2 Εφαρμογές της Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια 22

2.2.1. Ανίχνευση κακόβουλου λογισμικού (malware) 22

2.2.2. Ανίχνευση εισβολών (Intrusion Detection) 25

2.2.3. Ανάλυση Απειλών (Threat Analysis) 27

2.3 Πλεονεκτήματα και Περιορισμοί της ΤΝ στην Κυβερνοασφάλεια 30

2.3.1 Αναλυτική Παρουσίαση Δυνατοτήτων και Αδυναμιών 31

2.3.2 Δυνατότητες της ΤΝ στην Κυβερνοασφάλεια 31

2.3.3 Περιορισμοί της ΤΝ στην Κυβερνοασφάλεια 32

2.3.4 Μελλοντικές Προοπτικές και Εξελίξεις 33

3. Μελέτη Κυβερνοασφάλειας για Συστήματα Τεχνητής Νοημοσύνης 33
3.1 Κατηγοριοποίηση Κυβερνοεπιθέσεων σε Συστήματα ΤΝ 34
3.1.1 Επιθέσεις σε δεδομένα εκπαίδευσης (data poisoning) 34
3.1.2 Επιθέσεις σε μοντέλα (model attacks) 35
3.1.3 Επιθέσεις σε διαδικασίες αλληλεπίδρασης (interaction process attacks) 37
3.2 Επιπτώσεις των Κυβερνοεπιθέσεων στα Πληροφορικά Συστήματα ΤΝ 39
3.2.1 Αλλοίωση δεδομένων (data tampering) 39
3.2.2 Παραβίαση της ιδιωτικότητας (privacy breaches) 41
3.2.3 Διαταραχή της λειτουργικότητας (functional disruption) 43
3.3 Στρατηγικές Αντιμετώπισης Κυβερνοεπιθέσεων 45
3.3.1 Προληπτικά μέτρα (preventive measures) 45
3.3.2 Τεχνικές αντίδρασης και αποκατάστασης (reaction and recovery techniques) 46
4. Αμυντικές Τεχνικές για την Προστασία των Συστημάτων Τεχνητής Νοημοσύνης 47
4.1 Τεχνικές Αντίστασης σε Κυβερνοεπιθέσεις 48
4.1.1 Ανθεκτική εκπαίδευση μοντέλων (robust model training) 48
4.1.2 Προστασία δεδομένων (data protection) 49
4.2 Σχεδιασμός Ασφαλών Συστημάτων Τεχνητής Νοημοσύνης 51
4.2.1 Διανεμημένη μάθηση με προστασία ιδιωτικότητας (privacy-preserving distributed learning) 51
4.2.2 Κρυπτογραφημένη μάθηση (encrypted learning) 53
4.3 Τεχνολογίες Ασφάλειας για τα Συστήματα Τεχνητής Νοημοσύνης 55
4.3.1 Τεχνολογίες ανίχνευσης απειλών (threat detection technologies) 55
4.3.2 Τεχνολογίες προστασίας από επιθέσεις (attack protection technologies) 57
5. Συνδυαστικές Εφαρμογές Τεχνητής Νοημοσύνης και Κυβερνοασφάλειας 58
5.1 Προληπτική Κυβερνοασφάλεια μέσω Τεχνητής Νοημοσύνης 59

5.1.1 Έξυπνη παρακολούθηση απειλών (intelligent threat monitoring) 59

5.1.2 Προληπτική ανάλυση δεδομένων (proactive data analysis)..... 61

5.2 Ανίχνευση και Αντίδραση σε Κυβερνοεπιθέσεις με Τεχνητή Νοημοσύνη..... 63

5.2.1 Ανάλυση συμπεριφοράς (behavior analysis) 63

5.2.2 Αυτόματη απόκριση σε περιστατικά (automated incident response) 65

5.3 Περιπτωσιολογικές Μελέτες (Case Studies)..... 67

5.3.1 IBM Watson: Ένα επιτυχημένο παράδειγμα Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια 69

5.3.2 Microsoft και η Ασφάλεια στο Cloud..... 70

5.3.3 Darktrace και το Enterprise Immune System 71

5.3.4 PayPal και AI στην Ασφάλεια Συναλλαγών..... 72

6. Μελλοντικές Προοπτικές και Ερευνητικές Προκλήσεις 74

6.1 Αναδυόμενες Τάσεις και Τεχνολογίες..... 74

6.1.1 Ανάλυση Μεγάλων Δεδομένων στην Κυβερνοασφάλεια (Big Data Analysis in Cybersecurity) 74

6.1.2 Ανάπτυξη Νέων Αλγορίθμων ΤΝ για Κυβερνοασφάλεια (Development of New AI Algorithms for Cybersecurity)..... 76

6.2 Προκλήσεις και Προτεινόμενες Λύσεις (Challenges and Proposed Solutions)..... 77

6.2.1 Αντιμετώπιση της πολυπλοκότητας (Complexity Management) 78

6.2.2 Εξασφάλιση της Ιδιωτικότητας και της Ασφάλειας (Ensuring Privacy and Security). 79

6.3 Διεπιστημονικές Ερευνητικές Κατευθύνσεις (Interdisciplinary Research Directions)..... 80

7. Ρυθμιστικά πλαίσια για την Τεχνητή Νοημοσύνη - Ο Κανονισμός AI Act της Ευρωπαϊκής Ένωσης 81

7.1 Κατηγοριοποίηση Κινδύνου των Συστημάτων Τεχνητής Νοημοσύνης..... 81

7.2 Υποχρεώσεις για Παρόχους και Χρήστες Συστημάτων ΤΝ 82

7.3 Διακυβέρνηση και Εποπτεία 83

7.4 Κώδικες Δεοντολογίας και Εθελοντική Συμμόρφωση 83

7.5 Χρονοδιάγραμμα Εφαρμογής του Κανονισμού 83

7.6 Συμπεράσματα για το Ρυθμιστικό Πλαίσιο της Ευρωπαϊκής Ένωσης στην Τεχνητή Νοημοσύνη 84

8. Συμπεράσματα 84

8.1 Κύρια Συμπεράσματα της Έρευνας 84

8.2 Προτάσεις για Μελλοντική Έρευνα 86

Βιβλιογραφία 88

1. Εισαγωγή

Η παρούσα διπλωματική εργασία εστιάζει στην ενσωμάτωση της Τεχνητής Νοημοσύνης (ΤΝ) ως κρίσιμο εργαλείο στην κυβερνοασφάλεια. Η έρευνα διερευνά τις δυνατότητες και τις προκλήσεις που αντιμετωπίζει η χρήση της ΤΝ για την ανίχνευση και την αποτροπή σύγχρονων κυβερνοαπειλών, όπως κακόβουλο λογισμικό, επιθέσεις τύπου zero-day, παραβιάσεις ιδιωτικότητας και άλλες μορφές απειλών στον κυβερνοχώρο. Η χρήση τεχνολογιών μηχανικής μάθησης, βαθιάς μάθησης και ανάλυσης μεγάλων δεδομένων συμβάλλει σημαντικά στην πρόληψη επιθέσεων, καθώς και στην αυτοματοποιημένη απόκριση σε περιστατικά, βελτιώνοντας τη συνολική ασφάλεια των ψηφιακών υποδομών.

Μέσω της ΤΝ, οι οργανισμοί μπορούν να ανιχνεύουν απειλές πιο αποτελεσματικά και να διαχειρίζονται κρίσιμες καταστάσεις σε πραγματικό χρόνο, αποτρέποντας επιθέσεις πριν προκαλέσουν σημαντική ζημιά. Η έρευνα εξετάζει επίσης πώς η ΤΝ μπορεί να ενσωματωθεί με άλλες αναδυόμενες τεχνολογίες, όπως το blockchain, για την ενίσχυση της διαφάνειας και της ακεραιότητας στις διαδικασίες ασφαλείας.

Παράλληλα, παρουσιάζονται οι προκλήσεις που πρέπει να αντιμετωπιστούν, όπως η πολυπλοκότητα των αλγορίθμων ΤΝ και η ανάγκη για μεγάλες ποσότητες δεδομένων για εκπαίδευση. Τέλος, η εργασία προτείνει τομείς για μελλοντική έρευνα, συμπεριλαμβανομένων της ανθεκτικότητας των συστημάτων ΤΝ απέναντι σε επιθέσεις τύπου adversarial και της βελτίωσης της επεξηγησιμότητας των αποφάσεων που λαμβάνονται από τα συστήματα αυτά, ώστε να αυξηθεί η εμπιστοσύνη και η αποδοχή τους από τους οργανισμούς.

1.1 Θεωρητικό Υπόβαθρο και Κίνητρα Έρευνας

Σε αυτό το κεφάλαιο, θα αναλυθούν τα θεωρητικά θεμέλια και τα κίνητρα που οδήγησαν στην πραγματοποίηση αυτής της έρευνας. Θα εξεταστεί η αλληλεπίδραση μεταξύ της Τεχνητής Νοημοσύνης (ΤΝ) και της Κυβερνοασφάλειας, καθώς και ο αντίκτυπος που έχουν οι εξελίξεις στους δύο αυτούς τομείς στις σύγχρονες ψηφιακές υποδομές. Η ανάλυση θα επικεντρωθεί στην ανάγκη για προηγμένες τεχνολογίες ανίχνευσης και αποτροπής απειλών στον κυβερνοχώρο, δεδομένου του αυξανόμενου αριθμού και της πολυπλοκότητας των κυβερνοεπιθέσεων. Το κεφάλαιο θα περιγράψει πώς οι τεχνολογίες ΤΝ, και κυρίως οι αλγόριθμοι μηχανικής μάθησης,

έχουν τη δυνατότητα να προσφέρουν νέες λύσεις για την πρόληψη και την απόκριση σε επιθέσεις, βελτιώνοντας την ασφάλεια και την αξιοπιστία των συστημάτων.

1.1.1. Επισκόπηση της Τεχνητής Νοημοσύνης (ΤΝ)

Η Τεχνητή Νοημοσύνη (ΤΝ) αποτελεί έναν από τους πιο δυναμικούς και πολυσχιδείς τομείς της σύγχρονης επιστήμης των υπολογιστών. Ο βασικός στόχος της ΤΝ είναι η ανάπτυξη συστημάτων που μπορούν να μιμούνται και να υπερβαίνουν τις ανθρώπινες γνωστικές ικανότητες, όπως η μάθηση, η λήψη αποφάσεων και η αναγνώριση προτύπων. Οι πρώτες προσπάθειες στον τομέα αυτό εντοπίζονται στις αρχές του 20ού αιώνα, ωστόσο, η πραγματική επανάσταση στην ΤΝ ξεκίνησε με την ανάπτυξη της μηχανικής μάθησης (Machine Learning - ML) και της βαθιάς μάθησης (Deep Learning - DL). Η ML βασίζεται σε αλγορίθμους που επιτρέπουν στα συστήματα να "μαθαίνουν" από τα δεδομένα, βελτιώνοντας την ακρίβεια και την αποδοτικότητά τους μέσω της συνεχούς ανατροφοδότησης [1].

Η βαθιά μάθηση, ως μια προηγμένη μορφή ML, χρησιμοποιεί πολυστρωματικά νευρωνικά δίκτυα για την επεξεργασία και την ανάλυση μεγάλων ποσοτήτων δεδομένων. Αυτό έχει οδηγήσει σε σημαντικές προόδους σε τομείς όπως η αναγνώριση εικόνας, η φυσική γλώσσα και, κυρίως, η κυβερνοασφάλεια. Στον τομέα της κυβερνοασφάλειας, η ΤΝ έχει καταστεί ένα κρίσιμο εργαλείο, επιτρέποντας την ανίχνευση και την αποτροπή σύνθετων απειλών που δεν θα μπορούσαν να εντοπιστούν με παραδοσιακές μεθόδους [1], [10].

Μια από τις βασικές προκλήσεις που αντιμετωπίζει η ΤΝ είναι η ανάγκη για τεράστιες ποσότητες δεδομένων για την εκπαίδευση των αλγορίθμων της. Καθώς η ποσότητα και η πολυπλοκότητα των δεδομένων συνεχίζουν να αυξάνονται, η ΤΝ εξελίσσεται συνεχώς για να ανταποκριθεί στις νέες προκλήσεις. Αυτό έχει ως αποτέλεσμα τη βελτίωση της απόδοσης και της ακρίβειας των συστημάτων ΤΝ, ιδιαίτερα σε περιβάλλοντα όπου απαιτείται υψηλός βαθμός προσαρμοστικότητας και ευελιξίας [10].

Επιπλέον, η ΤΝ έχει ανοίξει νέες δυνατότητες σε τομείς όπως η προληπτική ανάλυση και η αυτόματη αντίδραση σε απειλές. Τα συστήματα που βασίζονται στην ΤΝ μπορούν να αναλύουν μεγάλα δεδομένα σε πραγματικό χρόνο, επιτρέποντας την άμεση αντίδραση σε κυβερνοεπιθέσεις και την πρόληψη πιθανών επιθέσεων πριν αυτές πραγματοποιηθούν. Αυτό έχει οδηγήσει σε μια σημαντική μετατόπιση από τις παραδοσιακές αντιδραστικές στρατηγικές ασφαλείας προς πιο προληπτικές και προσαρμοστικές προσεγγίσεις [1], [10].

Η τεχνολογία της ΤΝ, παρά την αδιαμφισβήτητη πρόοδό της, αντιμετωπίζει ακόμα προκλήσεις που σχετίζονται με την ασφάλεια και την ηθική. Η συνεχής βελτίωση και η αυτονομία των συστημάτων ΤΝ εγείρουν ερωτήματα σχετικά με την ευθύνη και την ασφάλεια, καθώς και με την ηθική χρήση αυτών των τεχνολογιών σε κρίσιμες εφαρμογές, όπως είναι η εθνική ασφάλεια και η προστασία προσωπικών δεδομένων [1], [10].

Αυτές οι εξελίξεις στην ΤΝ και η εφαρμογή τους σε κρίσιμους τομείς όπως η κυβερνοασφάλεια καταδεικνύουν την ανάγκη για συνεχή έρευνα και ανάπτυξη, προκειμένου να διασφαλιστεί ότι η ΤΝ θα μπορέσει να ανταποκριθεί στις μελλοντικές προκλήσεις και να παρέχει αποτελεσματικές λύσεις σε παγκόσμια κλίμακα [10].

1.1.2. Επισκόπηση της Κυβερνοασφάλειας

Η Κυβερνοασφάλεια αποτελεί έναν από τους πιο κρίσιμους τομείς στην τεχνολογική εποχή μας, καθώς οι ψηφιακές επιθέσεις και οι κυβερνοαπειλές συνεχώς εξελίσσονται και γίνονται όλο και πιο περίπλοκες. Ο όρος "κυβερνοασφάλεια" αναφέρεται σε όλες τις πρακτικές, τις τεχνολογίες και τις διαδικασίες που σχεδιάζονται για την προστασία των υπολογιστικών συστημάτων, των δικτύων και των δεδομένων από μη εξουσιοδοτημένη πρόσβαση, καταστροφή, τροποποίηση ή αποκάλυψη.

Η ανάγκη για αποτελεσματική κυβερνοασφάλεια έχει καταστεί ιδιαίτερα επιτακτική λόγω της αύξησης της συνδεσιμότητας και της εξάρτησης από τα ψηφιακά συστήματα, τόσο στον ιδιωτικό όσο και στον δημόσιο τομέα. Τα τελευταία χρόνια, οι κυβερνοεπιθέσεις έχουν αυξηθεί δραματικά σε αριθμό και σοβαρότητα, με επιθέσεις όπως αυτές των ransomware να απειλούν κρίσιμες υποδομές και επιχειρήσεις σε παγκόσμιο επίπεδο. Οι επιθέσεις αυτές δεν περιορίζονται μόνο στην κλοπή δεδομένων ή τη διατάραξη της λειτουργίας των συστημάτων, αλλά μπορούν να οδηγήσουν σε τεράστιες οικονομικές απώλειες και στην υπονόμευση της εμπιστοσύνης του κοινού στις τεχνολογίες πληροφοριών [9].

Η πολυπλοκότητα της κυβερνοασφάλειας είναι τέτοια που απαιτείται συνεχής ενημέρωση και προσαρμογή στις νέες απειλές. Οι οργανισμοί, για να αντιμετωπίσουν αυτές τις απειλές, επενδύουν σημαντικούς πόρους σε προηγμένες τεχνολογίες και εξειδικευμένο προσωπικό. Παράλληλα, η συνεργασία μεταξύ των κυβερνήσεων, των διεθνών οργανισμών και του ιδιωτικού τομέα είναι απαραίτητη για την αποτελεσματική αντιμετώπιση των κυβερνοεπιθέσεων σε παγκόσμιο επίπεδο [25].

Ένα κρίσιμο στοιχείο της κυβερνοασφάλειας είναι η συνεχής εκπαίδευση και ενημέρωση των χρηστών. Οι περισσότερες κυβερνοεπιθέσεις εκμεταλλεύονται την ανθρώπινη αδυναμία, είτε μέσω επιθέσεων phishing είτε μέσω κακόβουλων προγραμμάτων που εγκαθίστανται κατά λάθος από ανυποψίαστους χρήστες. Η ενημέρωση και η εκπαίδευση των χρηστών σε συνδυασμό με την εφαρμογή ισχυρών τεχνολογικών μέτρων, όπως τα συστήματα ανίχνευσης εισβολών και τα firewalls, είναι απαραίτητη για την προστασία των υποδομών από κακόβουλες επιθέσεις [9], [25].

Με την αύξηση των απειλών, η κυβερνοασφάλεια δεν μπορεί πλέον να θεωρείται απλά ως ένας τομέας της πληροφορικής, αλλά ως ένα απαραίτητο κομμάτι της στρατηγικής διαχείρισης κινδύνου κάθε οργανισμού. Η προσαρμογή στις νέες τεχνολογίες και η ενσωμάτωση της τεχνητής νοημοσύνης (ΤΝ) στην κυβερνοασφάλεια προσφέρει νέες δυνατότητες για την ανίχνευση και την αποτροπή των επιθέσεων σε πραγματικό χρόνο, ενισχύοντας την ικανότητα των οργανισμών να αντιμετωπίζουν τις νέες προκλήσεις του κυβερνοχώρου [9], [25].

1.1.3. Η Σημασία και η Αναγκαιότητα της Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια

Η Τεχνητή Νοημοσύνη (ΤΝ) και η Κυβερνοασφάλεια είναι δύο τομείς που έχουν αποκτήσει τεράστια σημασία στην ψηφιακή εποχή, ιδιαίτερα λόγω της ραγδαίας ανάπτυξης του Διαδικτύου και της αυξανόμενης εξάρτησης των ανθρώπων από τις ψηφιακές τεχνολογίες. Η συνεχής και αδιάλειπτη συνδεσιμότητα έχει οδηγήσει σε μια εποχή όπου τα δεδομένα και οι πληροφορίες αποτελούν τον πυρήνα σχεδόν κάθε ανθρώπινης δραστηριότητας. Ωστόσο, αυτή η συνδεσιμότητα συνοδεύεται από αυξημένους κινδύνους κυβερνοεπιθέσεων, οι οποίες μπορούν να έχουν καταστροφικές συνέπειες τόσο για τους οργανισμούς όσο και για τους μεμονωμένους χρήστες.

Η ΤΝ προσφέρει νέες δυνατότητες για την αντιμετώπιση αυτών των απειλών. Οι αλγόριθμοι μηχανικής μάθησης και βαθιάς μάθησης μπορούν να επεξεργάζονται τεράστιους όγκους δεδομένων με τρόπο που δεν είναι εφικτός από τους ανθρώπους, ανακαλύπτοντας μοτίβα και ανωμαλίες που μπορεί να υποδηλώνουν κυβερνοαπειλές [1]. Αυτό καθιστά δυνατή την ανίχνευση επιθέσεων σε πραγματικό χρόνο, επιτρέποντας ταχύτερες αντιδράσεις και βελτιωμένη προστασία των συστημάτων [1], [10].

Επιπλέον, η συνδυαστική προσέγγιση της ΤΝ και της Κυβερνοασφάλειας δεν περιορίζεται μόνο στην ανίχνευση επιθέσεων. Η ΤΝ μπορεί να χρησιμοποιηθεί για τη δημιουργία προσαρμοστικών συστημάτων που μαθαίνουν από προηγούμενες επιθέσεις και βελτιώνονται συνεχώς, αναπτύσσοντας πιο ανθεκτικές άμυνες [10]. Αυτή η ικανότητα των συστημάτων να εξελίσσονται

καθιστά τη χρήση της ΤΝ αναγκαία για την αντιμετώπιση των νέων και εξελισσόμενων απειλών στον κυβερνοχώρο.

Η σημασία αυτής της συνδυαστικής προσέγγισης έγκειται επίσης στη δυνατότητα της ΤΝ να προβλέπει πιθανά σημεία επίθεσης πριν αυτά εκδηλωθούν. Αυτό επιτρέπει στους οργανισμούς να λαμβάνουν προληπτικά μέτρα και να ελαχιστοποιούν τους κινδύνους πριν αυτοί γίνουν απειλές για την ασφάλεια των δεδομένων και των συστημάτων τους [1].

Η αυξημένη εξάρτηση από τα ψηφιακά συστήματα και η συνεχής παρουσία νέων απειλών καθιστούν τη συνδυαστική μελέτη της ΤΝ και της Κυβερνοασφάλειας όχι απλώς επιθυμητή, αλλά απολύτως αναγκαία. Χωρίς αυτή την προσέγγιση, οι οργανισμοί κινδυνεύουν να βρεθούν αντιμέτωποι με απειλές που δεν μπορούν να αντιμετωπιστούν με τις παραδοσιακές μεθόδους ασφάλειας [10].

Τέλος, η ενσωμάτωση της ΤΝ στην κυβερνοασφάλεια προάγει τη συνεργασία μεταξύ διαφορετικών επιστημονικών πεδίων, όπως η πληροφορική, η στατιστική, και η ψυχολογία, δημιουργώντας νέες διεπιστημονικές προσεγγίσεις που ενισχύουν τη συνολική ασφάλεια των ψηφιακών συστημάτων. Αυτή η προσέγγιση είναι κρίσιμη για την ανάπτυξη ολοκληρωμένων στρατηγικών που μπορούν να αντιμετωπίσουν τις σύγχρονες και μελλοντικές προκλήσεις [1], [10].

1.2 Σκοπός, Στόχοι, Πεδίο και Περιορισμοί της Έρευνας

Αυτή η ενότητα περιλαμβάνει μια αναλυτική παρουσίαση του σκοπού και των ερευνητικών στόχων της μελέτης, επικεντρωμένης στην αξιοποίηση της Τεχνητής Νοημοσύνης (ΤΝ) για την ενίσχυση της κυβερνοασφάλειας. Θα εξεταστούν οι δυνατότητες της ΤΝ να ανταποκρίνεται σε σύγχρονες προκλήσεις στον τομέα της ασφάλειας ψηφιακών συστημάτων, με έμφαση στην ανίχνευση και αποτροπή κυβερνοεπιθέσεων σε πραγματικό χρόνο. Η έρευνα επιδιώκει να απαντήσει σε κεντρικά ερωτήματα σχετικά με την αποτελεσματικότητα των αλγορίθμων ΤΝ και τις μεθόδους βελτιστοποίησης της απόδοσης τους, ενώ αναλύονται και οι περιορισμοί που σχετίζονται με τη συλλογή και διαχείριση δεδομένων, καθώς και τα ζητήματα ηθικής που προκύπτουν από τη χρήση της ΤΝ σε ευαίσθητες εφαρμογές.

1.2.1. Διατύπωση Ερευνητικών Ερωτημάτων και Στόχων

Η έρευνα γύρω από την Τεχνητή Νοημοσύνη (ΤΝ) και την Κυβερνοασφάλεια αποσκοπεί στην κατανόηση και αξιολόγηση της δυνατότητας της ΤΝ να αντιμετωπίσει τις σύγχρονες προκλήσεις που αναδύονται στον κυβερνοχώρο. Σε έναν κόσμο όπου οι κυβερνοαπειλές γίνονται όλο και πιο σύνθετες και επικίνδυνες, η ανάγκη για αποτελεσματικά εργαλεία και τεχνικές είναι πιο επιτακτική από ποτέ. Η ΤΝ, με την ικανότητά της να επεξεργάζεται και να αναλύει τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο, φαίνεται να προσφέρει λύσεις σε αυτά τα προβλήματα. Ωστόσο, η εφαρμογή της ΤΝ στην κυβερνοασφάλεια συνοδεύεται από μια σειρά από επιστημονικά και πρακτικά ερωτήματα που απαιτούν διερεύνηση [3].

Ένα από τα κύρια ερωτήματα της έρευνας είναι πώς οι αλγόριθμοι ΤΝ μπορούν να βελτιώσουν την αποτελεσματικότητα των συστημάτων ανίχνευσης εισβολών. Στη σύγχρονη πρακτική, οι παραδοσιακές τεχνικές ανίχνευσης βασίζονται σε προκαθορισμένα μοτίβα και υπογραφές επιθέσεων, καθιστώντας τις συχνά αναποτελεσματικές απέναντι σε νέες ή μη αναγνωρισμένες απειλές. Η ΤΝ μπορεί να υπερβεί αυτούς τους περιορισμούς μέσω της ικανότητάς της να μαθαίνει από προηγούμενες επιθέσεις και να προσαρμόζεται σε νέες μορφές απειλών. Το ερευνητικό ερώτημα που προκύπτει είναι πώς μπορεί η ΤΝ να ενσωματωθεί στα υπάρχοντα συστήματα ασφαλείας, ώστε να βελτιωθεί η ανίχνευση νέων και εξελιγμένων επιθέσεων [5].

Ένα δεύτερο ερώτημα αφορά την προγνωστική ικανότητα της ΤΝ. Ενώ οι υπάρχουσες τεχνικές ασφαλείας επικεντρώνονται κυρίως στην αντιμετώπιση των επιθέσεων μετά την εμφάνισή τους, η ΤΝ προσφέρει τη δυνατότητα της πρόβλεψης και της πρόληψης. Αυτό σημαίνει ότι οι αλγόριθμοι ΤΝ μπορούν να αναλύουν πρότυπα δεδομένων και να εντοπίζουν προειδοποιητικά σημάδια πιθανών απειλών πριν αυτές εκδηλωθούν. Η αποτελεσματικότητα αυτής της προγνωστικής προσέγγισης και οι προκλήσεις που σχετίζονται με την ανάπτυξη και εφαρμογή τέτοιων συστημάτων αποτελούν κρίσιμα ερωτήματα της έρευνας [3].

Επιπλέον, η έρευνα εξετάζει τους περιορισμούς και τις προκλήσεις που συνοδεύουν τη χρήση της ΤΝ στην κυβερνοασφάλεια. Αν και η ΤΝ προσφέρει σημαντικά πλεονεκτήματα, δεν είναι απαλλαγμένη από προβλήματα. Για παράδειγμα, η ανάπτυξη αλγορίθμων που είναι ανθεκτικοί σε επιθέσεις τύπου "adversarial" (εχθρικές επιθέσεις που προσπαθούν να παραπλανήσουν τα συστήματα ΤΝ) αποτελεί μια σοβαρή πρόκληση. Οι αντίπαλοι μπορεί να επιχειρήσουν να χειραγωγήσουν τα δεδομένα εκπαίδευσης ή να επιτεθούν στα ίδια τα μοντέλα ΤΝ, οδηγώντας σε

λανθασμένες αποφάσεις και μειώνοντας την αποτελεσματικότητα των συστημάτων ασφαλείας. Η έρευνα εξετάζει πώς μπορούν να ξεπεραστούν αυτοί οι περιορισμοί και πώς μπορούν να αναπτυχθούν πιο ανθεκτικοί αλγόριθμοι [5].

Τέλος, τα ηθικά ζητήματα που σχετίζονται με τη χρήση της ΤΝ στην κυβερνοασφάλεια δεν μπορούν να παραβλεφθούν. Η ενίσχυση της ασφάλειας μέσω της ΤΝ ενδέχεται να επιφέρει ακούσιες συνέπειες για την ιδιωτικότητα και τα ανθρώπινα δικαιώματα. Η έρευνα επικεντρώνεται στο πώς μπορούν να επιτευχθούν οι στόχοι της κυβερνοασφάλειας χωρίς να παραβιάζονται οι βασικές αρχές της ιδιωτικότητας και της ηθικής. Το ερώτημα που τίθεται είναι πώς μπορεί να διασφαλιστεί ότι τα συστήματα ΤΝ είναι διαφανή, επεξηγήσιμα και συμμορφωμένα με τις ηθικές αρχές, ενώ ταυτόχρονα παραμένουν αποτελεσματικά στην αντιμετώπιση των απειλών [3], [5].

Η διατύπωση αυτών των ερωτημάτων και στόχων προσφέρει ένα πλαίσιο μέσα στο οποίο η έρευνα θα εξετάσει τις δυνατότητες και τους περιορισμούς της ΤΝ στην κυβερνοασφάλεια, αναζητώντας νέες προσεγγίσεις που θα επιτρέψουν την ανάπτυξη ασφαλέστερων και πιο ανθεκτικών ψηφιακών συστημάτων.

1.2.2. Καθορισμός του Πεδίου και των Περιορισμών της Έρευνας

Η παρούσα έρευνα επικεντρώνεται στην εξερεύνηση της εφαρμογής της Τεχνητής Νοημοσύνης (ΤΝ) στην ενίσχυση της κυβερνοασφάλειας, εστιάζοντας σε συγκεκριμένα σενάρια και εφαρμογές που θεωρούνται κρίσιμα για τη διατήρηση της ακεραιότητας των ψηφιακών συστημάτων. Το πεδίο της έρευνας περιλαμβάνει την ανάπτυξη, εφαρμογή και αξιολόγηση αλγορίθμων ΤΝ που στοχεύουν στη βελτίωση των διαδικασιών ανίχνευσης και πρόληψης κυβερνοεπιθέσεων. Στόχος είναι να εξεταστεί η αποτελεσματικότητα αυτών των αλγορίθμων σε πραγματικά περιβάλλοντα, λαμβάνοντας υπόψη τις ποικίλες προκλήσεις που παρουσιάζει ο σύγχρονος ψηφιακός κόσμος [8], [16].

Ένα από τα βασικά στοιχεία που καθορίζουν το πεδίο της έρευνας είναι η εστίαση στις επιθέσεις που πραγματοποιούνται μέσω κακόβουλων λογισμικών και η χρήση της ΤΝ για την ανίχνευσή τους. Η έρευνα επικεντρώνεται ιδιαίτερα στην ανίχνευση επιθέσεων σε κινητές συσκευές και εφαρμογές Android, όπου η ανάπτυξη κακόβουλου λογισμικού έχει εξελιχθεί σε μια ιδιαίτερα σοφιστικέ απειλή. Η αξιολόγηση των αλγορίθμων γίνεται σε πραγματικά περιβάλλοντα, όπου τα

δεδομένα που συλλέγονται από διάφορες πηγές χρησιμοποιούνται για την εκπαίδευση και τη βελτίωση των μοντέλων ΤΝ [8].

Οι περιορισμοί της έρευνας αποτελούν επίσης ένα σημαντικό στοιχείο της μελέτης. Ένας από τους σημαντικότερους περιορισμούς είναι η ανάγκη για μεγάλα και ποιοτικά δεδομένα εκπαίδευσης. Η αποτελεσματικότητα των αλγορίθμων ΤΝ εξαρτάται σε μεγάλο βαθμό από την ποιότητα και την ποσότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευσή τους. Ωστόσο, η συλλογή τέτοιων δεδομένων μπορεί να είναι δύσκολη και χρονοβόρα, ενώ τα δεδομένα μπορεί να περιέχουν προκαταλήψεις ή να είναι ελλιπή, επηρεάζοντας την απόδοση των μοντέλων. Επιπλέον, η χρήση προσωπικών δεδομένων εγείρει σοβαρά ζητήματα ιδιωτικότητας, τα οποία πρέπει να αντιμετωπιστούν με την εφαρμογή κατάλληλων μέτρων ασφαλείας [16].

Ένας άλλος σημαντικός περιορισμός αφορά την πολυπλοκότητα και την προσαρμοστικότητα των επιθέσεων. Οι κυβερνοεγκληματίες συνεχώς αναπτύσσουν νέες τεχνικές που μπορούν να παρακάμπτουν τις παραδοσιακές άμυνες. Ενώ η ΤΝ έχει τη δυνατότητα να ανιχνεύει και να αντιμετωπίζει εξελιγμένες απειλές, δεν είναι πανάκεια. Οι αλγόριθμοι μπορεί να είναι ευάλωτοι σε "adversarial attacks", όπου οι αντίπαλοι προσπαθούν να παραπλανήσουν τα μοντέλα μέσω επιθέσεων που εκμεταλλεύονται τις αδυναμίες τους. Η έρευνα αυτή αντιμετωπίζει αυτές τις προκλήσεις μελετώντας τρόπους για την ανάπτυξη πιο ανθεκτικών και ευέλικτων αλγορίθμων που μπορούν να προσαρμοστούν στις συνεχώς μεταβαλλόμενες απειλές [8].

Επιπλέον, η έρευνα περιορίζεται στην ανάλυση συγκεκριμένων περιβαλλόντων και τύπων επιθέσεων, γεγονός που μπορεί να επηρεάσει τη γενικευσιμότητα των αποτελεσμάτων. Παρόλο που οι δοκιμές πραγματοποιούνται σε πραγματικά σενάρια, τα ευρήματα ενδέχεται να μην είναι πλήρως εφαρμόσιμα σε διαφορετικά περιβάλλοντα ή άλλους τομείς της κυβερνοασφάλειας. Αυτός ο περιορισμός επισημαίνει την ανάγκη για περαιτέρω έρευνα που να καλύπτει ένα ευρύτερο φάσμα επιθέσεων και σεναρίων για να ενισχυθεί η γενική ισχύς και εφαρμοσιμότητα των αποτελεσμάτων [16].

Επιπλέον, η πολυπλοκότητα της ενσωμάτωσης της ΤΝ στα υφιστάμενα συστήματα ασφαλείας αποτελεί μια άλλη πρόκληση. Τα παραδοσιακά συστήματα ασφαλείας ενδέχεται να μην είναι σχεδιασμένα για να υποστηρίξουν την προσθήκη αλγορίθμων ΤΝ, απαιτώντας σημαντικές

προσαρμογές και βελτιώσεις. Η έρευνα αυτή αναλύει τις τεχνικές και οργανωτικές δυσκολίες που μπορεί να προκύψουν κατά τη διαδικασία της ενσωμάτωσης, εξετάζοντας πώς μπορεί να επιτευχθεί η ομαλή ενσωμάτωση της ΤΝ στα υπάρχοντα συστήματα, χωρίς να διαταραχθεί η λειτουργικότητά τους [8], [16].

Τέλος, ένας περιορισμός που αφορά την ίδια την τεχνολογία της ΤΝ είναι η διαφάνεια και η επεξηγησιμότητα των αποφάσεων που λαμβάνονται από τους αλγορίθμους. Καθώς τα συστήματα ΤΝ γίνονται πιο περίπλοκα, μπορεί να γίνει δύσκολο να κατανοηθούν οι λόγοι πίσω από τις αποφάσεις που λαμβάνουν. Αυτό δημιουργεί ένα σημαντικό εμπόδιο στην εμπιστοσύνη και την υιοθέτηση της ΤΝ από τους χρήστες και τους οργανισμούς. Η έρευνα εξετάζει τις μεθόδους που μπορούν να χρησιμοποιηθούν για να ενισχυθεί η επεξηγησιμότητα των συστημάτων ΤΝ, επιτρέποντας στους χρήστες να κατανοούν καλύτερα τις λειτουργίες και τις αποφάσεις τους [16].

1.3 Δομή της Διπλωματικής Εργασίας

Περιγραφή της Διάρθρωσης των Κεφαλαίων και των Επιμέρους Περιεχομένων τους

Η διπλωματική εργασία αποτελείται από μια σειρά κεφαλαίων που είναι επιμελώς διαρθρωμένα ώστε να παρέχουν μια ολοκληρωμένη και λεπτομερή ανάλυση του θέματος που εξετάζεται. Η οργάνωση της εργασίας είναι κρίσιμη για την παρουσίαση των αποτελεσμάτων και την ανάπτυξη επιχειρημάτων με σαφή και λογική σειρά. Κάθε κεφάλαιο της εργασίας επιλέχθηκε για να προσφέρει μια συγκεκριμένη διάσταση της ερευνητικής διαδικασίας, ξεκινώντας από τις βασικές θεωρητικές αρχές και καταλήγοντας στα πρακτικά ευρήματα και τις μελλοντικές προοπτικές.

Πρώτο Κεφάλαιο: Εισαγωγή

Το πρώτο κεφάλαιο θέτει τα θεμέλια της εργασίας και προετοιμάζει τον αναγνώστη για το περιεχόμενο που θα ακολουθήσει. Εδώ εισάγονται οι βασικές έννοιες της Τεχνητής Νοημοσύνης (ΤΝ) και της Κυβερνοασφάλειας, με σκοπό να παρέχεται το θεωρητικό υπόβαθρο που απαιτείται για την κατανόηση της συνδυαστικής τους μελέτης. Στο κεφάλαιο αυτό αναλύονται επίσης οι λόγοι για τους οποίους η έρευνα επικεντρώνεται στη συνδυαστική εφαρμογή της ΤΝ και της Κυβερνοασφάλειας, τονίζοντας την αυξανόμενη ανάγκη για καινοτόμες λύσεις που μπορούν να αντιμετωπίσουν τις προκλήσεις στον ψηφιακό κόσμο. Επιπλέον, διατυπώνονται τα ερευνητικά

ερωτήματα που καθοδηγούν τη μελέτη, καθώς και οι στόχοι που επιδιώκονται μέσω αυτής της έρευνας. Καθορίζεται επίσης το πεδίο της έρευνας και οι περιορισμοί που τίθενται, ώστε να διασφαλιστεί ότι η εργασία παραμένει εστιασμένη και συγκεκριμένη.

Δεύτερο Κεφάλαιο: Τεχνητή Νοημοσύνη στην Κυβερνοασφάλεια

Το δεύτερο κεφάλαιο εστιάζει στη διερεύνηση των τεχνολογιών ΤΝ που είναι κρίσιμες για την ενίσχυση της κυβερνοασφάλειας. Ξεκινώντας με μια λεπτομερή ανάλυση της μηχανικής μάθησης (ML) και της βαθιάς μάθησης (DL), το κεφάλαιο εξετάζει πώς αυτές οι τεχνολογίες έχουν εξελιχθεί και πώς μπορούν να χρησιμοποιηθούν στην ανίχνευση και πρόληψη κυβερνοεπιθέσεων. Η μηχανική μάθηση, για παράδειγμα, επιτρέπει στα συστήματα να αναγνωρίζουν πρότυπα στα δεδομένα και να προβλέπουν πιθανές απειλές, ενώ η βαθιά μάθηση προσφέρει πιο προηγμένες δυνατότητες, χρησιμοποιώντας πολυστρωματικά νευρωνικά δίκτυα για την αναγνώριση πιο περίπλοκων μοτίβων. Σε αυτό το κεφάλαιο αναλύονται επίσης οι πρακτικές εφαρμογές αυτών των τεχνολογιών στην πραγματική ζωή, όπως η ανίχνευση κακόβουλου λογισμικού, η ανάλυση συμπεριφοράς των χρηστών, και η πρόβλεψη απειλών βάσει ιστορικών δεδομένων. Επιπλέον, το κεφάλαιο εξετάζει τις προκλήσεις που αντιμετωπίζουν οι αλγόριθμοι ΤΝ, όπως η ανάγκη για μεγάλα σύνολα δεδομένων και οι δυσκολίες στην αντιμετώπιση νέων ή άγνωστων απειλών.

Τρίτο Κεφάλαιο: Κυβερνοαπειλές για την Τεχνητή Νοημοσύνη

Το τρίτο κεφάλαιο αναλύει τις απειλές που αντιμετωπίζουν τα συστήματα ΤΝ, ειδικά όταν χρησιμοποιούνται για την ενίσχυση της κυβερνοασφάλειας. Το κεφάλαιο αυτό εξετάζει τις διάφορες μορφές επιθέσεων που μπορεί να στοχεύσουν τα συστήματα ΤΝ, όπως οι επιθέσεις adversarial, όπου οι επιτιθέμενοι προσπαθούν να παραπλανήσουν τα συστήματα ΤΝ με δεδομένα που έχουν σκόπιμα αλλοιωθεί. Επίσης, περιλαμβάνει επιθέσεις σε δεδομένα εκπαίδευσης, όπου κακόβουλα δεδομένα μπορούν να εισαχθούν στο σύστημα με σκοπό να το αποπροσανατολίσουν, καθώς και επιθέσεις που στοχεύουν στην ακεραιότητα των μοντέλων ΤΝ, όπως η αλλοίωση των παραμέτρων των μοντέλων. Το κεφάλαιο αναλύει τον αντίκτυπο αυτών των επιθέσεων και τις συνέπειες που μπορούν να έχουν στην ακεραιότητα και την αποτελεσματικότητα των συστημάτων ασφαλείας. Επιπλέον, προτείνονται στρατηγικές για την άμυνα ενάντια σε αυτές τις επιθέσεις,

όπως η χρήση ανθεκτικής εκπαίδευσης μοντέλων και η ενίσχυση των πρωτοκόλλων ασφαλείας για την προστασία των δεδομένων και των συστημάτων.

Τέταρτο Κεφάλαιο: Αμυντικές Τεχνικές για την Προστασία της Τεχνητής Νοημοσύνης

Το τέταρτο κεφάλαιο επικεντρώνεται στις αμυντικές τεχνικές που μπορούν να χρησιμοποιηθούν για την προστασία των συστημάτων ΤΝ από τις κυβερνοαπειλές. Οι τεχνικές αυτές περιλαμβάνουν την ανθεκτική εκπαίδευση των μοντέλων, η οποία επιτρέπει στα συστήματα να παραμένουν αποτελεσματικά ακόμα και υπό την πίεση εχθρικών επιθέσεων. Το κεφάλαιο εξετάζει, επίσης, τις τεχνικές ανίχνευσης και αποκάλυψης απειλών, όπως η χρήση αλγορίθμων ΤΝ για την ανάλυση δεδομένων και την ανίχνευση ανωμαλιών σε πραγματικό χρόνο. Ένα άλλο κρίσιμο στοιχείο είναι η προστασία των δεδομένων, η οποία περιλαμβάνει τη διασφάλιση ότι τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση των αλγορίθμων είναι αξιόπιστα και ασφαλή από παραβιάσεις. Το κεφάλαιο αυτό αναλύει επίσης τις προκλήσεις που συνδέονται με την εφαρμογή αυτών των τεχνικών, όπως η ανάγκη για συνεχής ενημέρωση και προσαρμογή των συστημάτων στις νέες απειλές.

Πέμπτο Κεφάλαιο: Συνδυαστικές Εφαρμογές Τεχνητής Νοημοσύνης και Κυβερνοασφάλειας

Το πέμπτο κεφάλαιο παρέχει παραδείγματα και περιπτωσιολογικές μελέτες πραγματικών εφαρμογών της ΤΝ στην κυβερνοασφάλεια. Εδώ παρουσιάζονται παραδείγματα όπου η ΤΝ έχει χρησιμοποιηθεί με επιτυχία για την ανίχνευση και αποτροπή κυβερνοεπιθέσεων, όπως στην προστασία δικτύων, την ανάλυση συμπεριφοράς των χρηστών για την πρόληψη απειλών, και την αυτοματοποίηση διαδικασιών απόκρισης σε περιστατικά ασφάλειας. Αυτές οι μελέτες προσφέρουν μια σαφή εικόνα της πρακτικής αξίας της ΤΝ στην κυβερνοασφάλεια και εξετάζουν τις δυνατότητες περαιτέρω ανάπτυξης και εφαρμογής αυτών των τεχνολογιών σε ποικίλα περιβάλλοντα. Επιπλέον, εξετάζονται οι προκλήσεις που προκύπτουν κατά την εφαρμογή της ΤΝ σε διαφορετικά πλαίσια, όπως η ανάγκη για προσαρμογή στις ιδιαιτερότητες του κάθε περιβάλλοντος και η διασφάλιση της επεκτασιμότητας των λύσεων.

Έκτο Κεφάλαιο: Μελλοντικές Προοπτικές και Ερευνητικές Προκλήσεις

Το έκτο κεφάλαιο εξετάζει τις μελλοντικές τάσεις και τις προοπτικές που θα διαμορφώσουν τον τομέα της ΤΝ στην κυβερνοασφάλεια. Αναλύονται οι τεχνολογικές εξελίξεις που αναμένονται τα επόμενα χρόνια, όπως η ενσωμάτωση της ΤΝ σε ολοένα και περισσότερα πεδία της κυβερνοασφάλειας, η ανάπτυξη νέων αλγορίθμων για την ανίχνευση προηγμένων απειλών, και η χρήση της ΤΝ για την πρόβλεψη και πρόληψη απειλών σε παγκόσμιο επίπεδο. Το κεφάλαιο αυτό συζητά επίσης τις προκλήσεις που παραμένουν ανεπίλυτες και προτείνει λύσεις και ερευνητικές κατευθύνσεις για την περαιτέρω βελτίωση της ασφάλειας των ψηφιακών συστημάτων.

Έβδομο Κεφάλαιο: Συμπεράσματα

Το τελευταίο κεφάλαιο συνοψίζει τα βασικά ευρήματα της έρευνας, αναδεικνύοντας τη σημασία της συνδυαστικής χρήσης της ΤΝ στην κυβερνοασφάλεια. Παρουσιάζονται οι κύριες προκλήσεις και οι ευκαιρίες που προκύπτουν από την εφαρμογή αυτών των τεχνολογιών, και παρέχονται προτάσεις για την περαιτέρω έρευνα και την ανάπτυξη νέων προσεγγίσεων που θα βελτιώσουν την ασφάλεια των συστημάτων στον κυβερνοχώρο.

2: Τεχνητή Νοημοσύνη στην Κυβερνοασφάλεια

2.1 Θεμελιώδεις Τεχνολογίες και Μεθοδολογίες

2.1.1. Μηχανική Μάθηση (ΜΜ)

Η Μηχανική Μάθηση (ΜΜ) αποτελεί έναν από τους θεμέλιους λίθους της Τεχνητής Νοημοσύνης (ΤΝ) και παίζει καθοριστικό ρόλο στην ενίσχυση της Κυβερνοασφάλειας. Πρόκειται για έναν τομέα που έχει αναπτυχθεί ραγδαία τις τελευταίες δεκαετίες, επιτρέποντας στα συστήματα υπολογιστών να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους χωρίς ρητή προγραμματιστική καθοδήγηση. Η ικανότητα αυτή είναι ιδιαίτερα χρήσιμη στην Κυβερνοασφάλεια, όπου οι απειλές εξελίσσονται συνεχώς και οι παραδοσιακές τεχνικές ανίχνευσης συχνά αδυνατούν να ανταποκριθούν στις νέες προκλήσεις.

Η ΜΜ στηρίζεται στην ανάλυση μεγάλων όγκων δεδομένων για την αναγνώριση προτύπων που μπορεί να υποδηλώνουν απειλές ασφαλείας. Χρησιμοποιώντας τεχνικές όπως η ταξινόμηση, η

παλινδρόμηση και η συσταδοποίηση, οι αλγόριθμοι ΜΜ μπορούν να εντοπίσουν ύποπτες δραστηριότητες ή ανωμαλίες που μπορεί να υποδηλώνουν μια επικείμενη επίθεση. Η ΜΜ δεν περιορίζεται μόνο στην ανίχνευση απειλών, αλλά μπορεί επίσης να προβλέψει πιθανές απειλές βάσει ιστορικών δεδομένων, προσφέροντας έτσι ένα επιπλέον επίπεδο προστασίας.

Μια από τις πιο κοινές εφαρμογές της ΜΜ στην Κυβερνοασφάλεια είναι η ανίχνευση κακόβουλου λογισμικού. Οι παραδοσιακές τεχνικές ανίχνευσης βασίζονται στη σύγκριση υπογραφών γνωστών απειλών με τα αρχεία του συστήματος, μια διαδικασία που μπορεί να μην είναι αποτελεσματική απέναντι σε νέες ή παραλλαγμένες μορφές κακόβουλου λογισμικού. Αντίθετα, οι αλγόριθμοι ΜΜ μπορούν να εκπαιδευτούν να αναγνωρίζουν μοτίβα συμπεριφοράς που είναι χαρακτηριστικά του κακόβουλου λογισμικού, ακόμα και όταν αυτό δεν έχει προηγουμένως αναγνωριστεί από συστήματα ασφαλείας [4], [7].

Ένα άλλο σημαντικό πεδίο εφαρμογής της ΜΜ είναι η ανίχνευση ανωμαλιών στο δίκτυο. Οι αλγόριθμοι ΜΜ μπορούν να παρακολουθούν τη δραστηριότητα του δικτύου σε πραγματικό χρόνο και να εντοπίζουν αποκλίσεις από την κανονική συμπεριφορά, οι οποίες μπορεί να υποδηλώνουν μια επίθεση. Αυτή η μέθοδος είναι ιδιαίτερα χρήσιμη για την ανίχνευση επιθέσεων τύπου "zero-day", οι οποίες εκμεταλλεύονται άγνωστες ευπάθειες και είναι δύσκολο να εντοπιστούν με παραδοσιακές μεθόδους.

Παρά τα σημαντικά πλεονεκτήματα της ΜΜ στην Κυβερνοασφάλεια, υπάρχουν και προκλήσεις. Η αποτελεσματικότητα των αλγορίθμων ΜΜ εξαρτάται σε μεγάλο βαθμό από την ποιότητα και την ποσότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευσή τους. Δεδομένα που είναι ελλιπή ή προκατειλημμένα μπορούν να οδηγήσουν σε λανθασμένα συμπεράσματα, θέτοντας σε κίνδυνο την ασφάλεια του συστήματος. Επιπλέον, η ανάπτυξη αλγορίθμων που μπορούν να αντέξουν σε επιθέσεις adversarial αποτελεί μια σημαντική πρόκληση στον τομέα της ΜΜ. Οι επιθέσεις αυτές στοχεύουν στη χειραγώγηση των εισροών στα μοντέλα ΜΜ με σκοπό την παραπλάνηση του συστήματος, οδηγώντας το σε εσφαλμένες αποφάσεις [4].

Η ενσωμάτωση της ΜΜ στην Κυβερνοασφάλεια είναι ένα πολυσύνθετο έργο που απαιτεί συνεχή βελτίωση και προσαρμογή των αλγορίθμων στις νέες απειλές που εμφανίζονται στον κυβερνοχώρο. Παρά τις προκλήσεις, η ΜΜ έχει αποδειχθεί ως ένα από τα πιο ισχυρά εργαλεία

για την ενίσχυση της ασφάλειας, επιτρέποντας στους οργανισμούς να παραμένουν ένα βήμα μπροστά από τους επιτιθέμενους.

2.1.2. Βαθιά Μάθηση (ΒΜ)

Η Βαθιά Μάθηση (ΒΜ) αποτελεί έναν πυλώνα της σύγχρονης Τεχνητής Νοημοσύνης (ΤΝ), προσφέροντας σημαντικές δυνατότητες στην ανάλυση μεγάλων και πολύπλοκων δεδομένων. Βασιζόμενη σε βαθιά νευρωνικά δίκτυα, η ΒΜ διακρίνεται για την ικανότητά της να επεξεργάζεται δεδομένα με πολυπλοκότητα και ποικιλομορφία που ξεπερνούν τις δυνατότητες των παραδοσιακών αλγορίθμων Μηχανικής Μάθησης (ΜΜ). Αυτή η μοναδική ικανότητα την καθιστά εξαιρετικά πολύτιμη στην Κυβερνοασφάλεια, όπου οι κυβερνοεπιθέσεις συχνά χαρακτηρίζονται από υψηλή πολυπλοκότητα και δυναμικότητα.

Η αρχιτεκτονική των βαθιών νευρωνικών δικτύων είναι το κλειδί για την επιτυχία της ΒΜ. Τα δίκτυα αυτά αποτελούνται από πολλαπλά επίπεδα τεχνητών νευρώνων, όπου κάθε επίπεδο μαθαίνει να εξάγει χαρακτηριστικά σε διαφορετικά επίπεδα αφαίρεσης από τα δεδομένα εισόδου. Στο πλαίσιο της Κυβερνοασφάλειας, αυτό σημαίνει ότι η ΒΜ μπορεί να αναγνωρίσει λεπτομερείς και σύνθετες σχέσεις μέσα στα δεδομένα, όπως ανωμαλίες στη ροή δικτύου ή μοτίβα κακόβουλης συμπεριφοράς, που θα ήταν αόρατες σε πιο απλοϊκά μοντέλα [2], [4].

2.1.3. Ανίχνευση Κακόβουλου Λογισμικού μέσω ΒΜ

Μια από τις πιο κρίσιμες εφαρμογές της ΒΜ στην Κυβερνοασφάλεια είναι η ανίχνευση κακόβουλου λογισμικού. Ενώ οι παραδοσιακές τεχνικές βασίζονται σε υπογραφές και μοτίβα που έχουν προαποφασιστεί, η ΒΜ χρησιμοποιεί συνελικτικά νευρωνικά δίκτυα (CNNs) για να αναλύσει τα χαρακτηριστικά των εκτελέσιμων αρχείων σε βάθος. Τα CNNs επιτρέπουν την ανίχνευση κακόβουλου λογισμικού με βάση τα πρότυπα συμπεριφοράς και τα δομικά χαρακτηριστικά του κώδικα, καθιστώντας τα ικανά να εντοπίζουν νέες, άγνωστες απειλές που μπορεί να μην έχουν προηγουμένως καταγραφεί. Αυτή η ικανότητα είναι κρίσιμη για την ανίχνευση επιθέσεων τύπου "zero-day", όπου το κακόβουλο λογισμικό εκμεταλλεύεται ευπάθειες που δεν έχουν ακόμα δημοσιοποιηθεί ή επιδιορθωθεί [2].

Επιπλέον, η ΒΜ επιτρέπει την ανάλυση των δυναμικών συμπεριφορών του κακόβουλου λογισμικού, δηλαδή πώς το λογισμικό συμπεριφέρεται όταν εκτελείται σε ένα σύστημα. Αυτή η προσέγγιση βασίζεται σε νευρωνικά δίκτυα αναδρομής (RNNs) που μπορούν να επεξεργάζονται ακολουθίες δεδομένων και να αναγνωρίζουν μοτίβα που σχετίζονται με κακόβουλες δραστηριότητες. Η δυνατότητα των RNNs να διατηρούν πληροφορίες από προηγούμενες καταστάσεις τα καθιστά ιδανικά για την ανίχνευση περίπλοκων επιθέσεων που εκδηλώνονται σε βάθος χρόνου [4].

2.1.4. Ανάλυση Δικτύων και Ανίχνευση Ανωμαλιών

Η ΒΜ διαδραματίζει επίσης κρίσιμο ρόλο στην ανάλυση δικτύων, ειδικά στην ανίχνευση ανωμαλιών και επιθέσεων σε πραγματικό χρόνο. Τα βαθιά νευρωνικά δίκτυα μπορούν να εκπαιδευτούν σε τεράστια σύνολα δεδομένων από τη δικτυακή κυκλοφορία, μαθαίνοντας να αναγνωρίζουν φυσιολογικά πρότυπα και να εντοπίζουν αποκλίσεις που μπορεί να υποδηλώνουν μια κυβερνοεπίθεση. Σε αντίθεση με τις παραδοσιακές μεθόδους, οι οποίες μπορεί να βασίζονται σε κανόνες που ορίζονται από τον χρήστη, η ΒΜ έχει την ικανότητα να προσαρμόζεται και να βελτιώνεται συνεχώς, μαθαίνοντας από τα δεδομένα που επεξεργάζεται [2].

Μια ιδιαίτερα καινοτόμος χρήση της ΒΜ στην ανάλυση δικτύων είναι η εφαρμογή συνελικτικών νευρωνικών δικτύων (CNNs) για την ανίχνευση εισβολών σε δίκτυα. Τα CNNs μπορούν να αναλύουν την δικτυακή κίνηση με τέτοιο τρόπο ώστε να εντοπίζουν ακόμα και τις πιο μικρές ανωμαλίες που θα μπορούσαν να υποδηλώνουν μια κακόβουλη δραστηριότητα. Αυτές οι ανωμαλίες μπορεί να περιλαμβάνουν μικρές αποκλίσεις στη συχνότητα των πακέτων δεδομένων, ασυνήθιστες αλλαγές στα πρότυπα κίνησης ή ακόμα και μη φυσιολογική κατανάλωση πόρων [4].

2.1.5. Προκλήσεις και Περιορισμοί της ΒΜ στην Κυβερνοασφάλεια

Παρά τις πολλές δυνατότητες που προσφέρει η ΒΜ, υπάρχουν σημαντικές προκλήσεις που πρέπει να αντιμετωπιστούν. Μία από τις μεγαλύτερες είναι η ευπάθεια σε επιθέσεις adversarial. Οι επιτιθέμενοι μπορούν να εκμεταλλευτούν τις λεπτές διαφοροποιήσεις στα δεδομένα εισόδου για να παραπλανήσουν τα νευρωνικά δίκτυα, αναγκάζοντάς τα να παράγουν λανθασμένα αποτελέσματα. Για παράδειγμα, με την προσθήκη συγκεκριμένου θορύβου σε μια εικόνα ή ένα

πακέτο δεδομένων, οι επιτιθέμενοι μπορούν να διαταράξουν τη λειτουργία του δικτύου, καθιστώντας το αναποτελεσματικό στην ανίχνευση απειλών [2].

Επιπλέον, η πολυπλοκότητα των βαθιών νευρωνικών δικτύων μπορεί να τα καθιστά δύσκολα στη διαχείριση και τη συντήρηση. Η ανάπτυξη, η εκπαίδευση και η βελτιστοποίηση αυτών των δικτύων απαιτεί εξειδικευμένους πόρους και γνώση, κάτι που μπορεί να είναι περιοριστικός παράγοντας για μικρότερους οργανισμούς. Επίσης, η ανάγκη για μεγάλα σύνολα δεδομένων για την εκπαίδευση των δικτύων μπορεί να δημιουργήσει ζητήματα ιδιωτικότητας και ασφάλειας δεδομένων [4].

Παρά τις προκλήσεις, η ΒΜ συνεχίζει να αποτελεί μία από τις πιο προηγμένες και αποτελεσματικές τεχνολογίες στην Κυβερνοασφάλεια. Με τη συνεχή ανάπτυξη νέων αλγορίθμων και τη βελτίωση των υφιστάμενων μεθόδων, η ΒΜ αναμένεται να διαδραματίσει ακόμη μεγαλύτερο ρόλο στην ανίχνευση και την πρόληψη κυβερνοαπειλών στο μέλλον.

2.2 Εφαρμογές της Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια

Στο παρόν κεφάλαιο θα παρουσιαστεί η εφαρμογή της Τεχνητής Νοημοσύνης (ΤΝ) στον τομέα της κυβερνοασφάλειας, αναλύοντας θεμελιώδεις τεχνολογίες και μεθοδολογίες. Η ενσωμάτωση της ΤΝ στην προστασία των ψηφιακών συστημάτων έχει επιφέρει ριζικές αλλαγές στον τρόπο με τον οποίο εντοπίζονται και αποτρέπονται οι κυβερνοαπειλές. Οι βασικές τεχνολογίες όπως η Μηχανική Μάθηση (ΜΜ) και η Βαθιά Μάθηση (ΒΜ) διαδραματίζουν κεντρικό ρόλο στην αναγνώριση ανωμαλιών και στην αυτόματη ανίχνευση επιθέσεων σε πραγματικό χρόνο. Μέσα από αυτές τις τεχνολογίες, οι οργανισμοί έχουν τη δυνατότητα να ενισχύσουν την προστασία τους και να ανιχνεύσουν ακόμα και τις πιο εξελιγμένες απειλές, προσφέροντας πιο ασφαλή περιβάλλοντα λειτουργίας για κρίσιμες υποδομές.

2.2.1. Ανίχνευση κακόβουλου λογισμικού (malware)

Η ανίχνευση κακόβουλου λογισμικού αποτελεί έναν από τους πιο σημαντικούς τομείς εφαρμογής της Τεχνητής Νοημοσύνης (ΤΝ) στην Κυβερνοασφάλεια. Το κακόβουλο λογισμικό συνεχώς εξελίσσεται και γίνεται πιο περίπλοκο, καθιστώντας τις παραδοσιακές μεθόδους ανίχνευσης

συχνά ανεπαρκείς. Σε αυτό το περιβάλλον, η ΤΝ και ειδικότερα οι τεχνικές Μηχανικής Μάθησης (ΜΜ) και Βαθιάς Μάθησης (ΒΜ) προσφέρουν προηγμένες λύσεις για την αποτελεσματική ανίχνευση και την πρόληψη επιθέσεων που προέρχονται από κακόβουλο λογισμικό.

Η προσέγγιση που ακολουθείται για την ανίχνευση κακόβουλου λογισμικού μέσω της ΤΝ βασίζεται στην ικανότητα των αλγορίθμων να αναλύουν μεγάλους όγκους δεδομένων και να αναγνωρίζουν μοτίβα που υποδηλώνουν κακόβουλη δραστηριότητα. Οι τεχνικές αυτές περιλαμβάνουν τη χρήση νευρωνικών δικτύων για την ανάλυση του κώδικα των προγραμμάτων, καθώς και για τη συμπεριφορική ανάλυση των αρχείων κατά την εκτέλεση τους. Τα νευρωνικά δίκτυα, ειδικά όταν εφαρμόζονται με ΒΜ, μπορούν να αναγνωρίσουν πολύπλοκα μοτίβα που μπορεί να διαφύγουν από τις παραδοσιακές μεθόδους ανίχνευσης που βασίζονται σε υπογραφές [3], [6].

2.2.1.1. Αναγνώριση κακόβουλου λογισμικού μέσω συμπεριφορικής ανάλυσης

Μια από τις πιο αποτελεσματικές εφαρμογές της ΤΝ στην ανίχνευση κακόβουλου λογισμικού είναι η συμπεριφορική ανάλυση. Αντί να βασίζονται μόνο στα χαρακτηριστικά του κώδικα, οι αλγόριθμοι ΜΜ και ΒΜ αναλύουν τον τρόπο με τον οποίο τα προγράμματα αλληλεπιδρούν με το σύστημα κατά την εκτέλεση τους. Αυτή η προσέγγιση επιτρέπει την ανίχνευση κακόβουλου λογισμικού που μπορεί να μην έχει αναγνωριστεί προηγουμένως και δεν έχει υπογραφή σε υπάρχουσες βάσεις δεδομένων. Για παράδειγμα, η ανάλυση της συμπεριφοράς μπορεί να αποκαλύψει κακόβουλες ενέργειες όπως η προσπάθεια πρόσβασης σε απαγορευμένες περιοχές της μνήμης ή η προσπάθεια αλλαγής κρίσιμων αρχείων συστήματος [6].

Η συμπεριφορική ανάλυση μέσω ΤΝ προσφέρει επίσης τη δυνατότητα για έγκαιρη ανίχνευση νέων μορφών κακόβουλου λογισμικού, καθώς μπορεί να εντοπίσει ύποπτες δραστηριότητες ακόμα και πριν αυτές προκαλέσουν ζημιά. Αυτό καθιστά τη διαδικασία ανίχνευσης πιο προληπτική, μειώνοντας τις πιθανότητες επιτυχούς επίθεσης.

2.2.1.2. Ενσωμάτωση της ΒΜ στην ανίχνευση κακόβουλου λογισμικού

Η Βαθιά Μάθηση, με την ικανότητά της να επεξεργάζεται και να αναλύει δεδομένα σε βάθος, χρησιμοποιείται ευρέως στην ανίχνευση κακόβουλου λογισμικού. Ένα χαρακτηριστικό

παράδειγμα είναι η χρήση συνελικτικών νευρωνικών δικτύων (CNNs) για την ανάλυση των χαρακτηριστικών των εκτελέσιμων αρχείων. Τα CNNs μπορούν να αναγνωρίσουν ακόμα και τις πιο λεπτές διαφορές στα χαρακτηριστικά των αρχείων, επιτρέποντας την ανίχνευση νέων ή παραλλαγμένων μορφών κακόβουλου λογισμικού. Αυτή η ικανότητα είναι ιδιαίτερα χρήσιμη για την ανίχνευση επιθέσεων τύπου "zero-day", όπου οι επιτιθέμενοι εκμεταλλεύονται άγνωστες ευπάθειες [3].

2.2.1.3. Προκλήσεις και μελλοντικές εξελίξεις

Παρόλο που η ΤΝ έχει σημειώσει σημαντική πρόοδο στην ανίχνευση κακόβουλου λογισμικού, υπάρχουν προκλήσεις που πρέπει να αντιμετωπιστούν. Μια από τις μεγαλύτερες είναι η ανάπτυξη αλγορίθμων που είναι ανθεκτικοί σε επιθέσεις τύπου adversarial. Οι επιτιθέμενοι μπορούν να προσαρμόσουν το κακόβουλο λογισμικό ώστε να παραπλανήσει τα μοντέλα ΤΝ, καθιστώντας δύσκολη την ανίχνευση. Για να ξεπεραστούν αυτές οι προκλήσεις, η έρευνα συνεχίζεται με στόχο τη βελτίωση των αλγορίθμων και την ενσωμάτωση τεχνικών που αυξάνουν την ανθεκτικότητα των μοντέλων απέναντι σε τέτοιες επιθέσεις.

Επιπλέον, οι αλγόριθμοι ΤΝ πρέπει να εμπλουτίζονται συνεχώς με νέα δεδομένα για να παραμένουν αποτελεσματικοί. Αυτό σημαίνει ότι οι οργανισμοί πρέπει να έχουν πρόσβαση σε μεγάλες βάσεις δεδομένων κακόβουλου λογισμικού και να εφαρμόζουν διαδικασίες που εξασφαλίζουν ότι τα μοντέλα τους ενημερώνονται τακτικά. Η διαχείριση αυτών των δεδομένων, καθώς και η διασφάλιση της ποιότητας και της ακρίβειας τους, αποτελεί κρίσιμο στοιχείο για την επιτυχία των εφαρμογών ΤΝ στην ανίχνευση κακόβουλου λογισμικού [6].

Με τη συνεχή εξέλιξη των κυβερνοαπειλών, η ανάγκη για καινοτόμες λύσεις ανίχνευσης κακόβουλου λογισμικού είναι μεγαλύτερη από ποτέ. Η ΤΝ προσφέρει σημαντικές δυνατότητες για την προστασία των συστημάτων, αλλά απαιτείται συνεχής έρευνα και ανάπτυξη για να αντιμετωπιστούν οι προκλήσεις και να βελτιωθεί η αποτελεσματικότητα των λύσεων αυτών. Στο μέλλον, η περαιτέρω ενσωμάτωση της ΤΝ και η ανάπτυξη νέων τεχνικών θα αποτελέσουν κλειδί για την αντιμετώπιση των εξελισσόμενων απειλών στον κυβερνοχώρο.

2.2.2. Ανίχνευση εισβολών (Intrusion Detection)

Η ανίχνευση εισβολών είναι ένας από τους πιο κρίσιμους τομείς της Κυβερνοασφάλειας, καθώς οι εισβολές μπορούν να προκαλέσουν σοβαρές ζημιές στα δίκτυα και τα συστήματα ενός οργανισμού. Οι μέθοδοι ανίχνευσης εισβολών έχουν εξελιχθεί σημαντικά με την εισαγωγή της Τεχνητής Νοημοσύνης (ΤΝ), παρέχοντας πιο ακριβή και προσαρμοστικά εργαλεία για την αναγνώριση κακόβουλων δραστηριοτήτων. Η χρήση της ΤΝ επιτρέπει την ανίχνευση όχι μόνο γνωστών απειλών αλλά και νέων, άγνωστων τύπων επιθέσεων, βελτιώνοντας έτσι την ασφάλεια των συστημάτων.

2.2.2.1. Συστήματα Ανίχνευσης Κατάχρησης (Misuse Detection Systems)

Τα Συστήματα Ανίχνευσης Κατάχρησης παραδοσιακά βασίζονται σε προκαθορισμένα πρότυπα και υπογραφές για την αναγνώριση επιθέσεων. Αυτά τα συστήματα συγκρίνουν τα εισερχόμενα δεδομένα με γνωστές υπογραφές κακόβουλου λογισμικού ή άλλων απειλών. Αν και αυτή η μέθοδος είναι αποτελεσματική για την ανίχνευση γνωστών απειλών, έχει σημαντικούς περιορισμούς όταν πρόκειται για νέες ή παραλλαγμένες επιθέσεις. Τα προγράμματα κακόβουλου λογισμικού εξελίσσονται συνεχώς και οι επιτιθέμενοι βρίσκουν συνεχώς νέους τρόπους να παρακάμψουν τις υπογραφές, καθιστώντας την ανάγκη για πιο δυναμικές μεθόδους ανίχνευσης επιτακτική [4].

2.2.2.2 Συστήματα Ανίχνευσης Ανωμαλιών (Anomaly Detection Systems)

Τα Συστήματα Ανίχνευσης Ανωμαλιών, τα οποία βασίζονται στην ΤΝ, παρέχουν μια πιο δυναμική και προσαρμοστική προσέγγιση στην ανίχνευση εισβολών. Αυτά τα συστήματα εκπαιδεύονται με τη χρήση Μηχανικής Μάθησης (ΜΜ) για να αναγνωρίσουν τα κανονικά πρότυπα συμπεριφοράς στο δίκτυο ή το σύστημα. Οποιαδήποτε απόκλιση από αυτά τα πρότυπα μπορεί να θεωρηθεί ως πιθανή απειλή. Αυτός ο τύπος ανίχνευσης είναι ιδιαίτερα χρήσιμος για την ανίχνευση επιθέσεων τύπου "zero-day", όπου οι επιτιθέμενοι εκμεταλλεύονται ευπάθειες που δεν έχουν ακόμα αναγνωριστεί από τα παραδοσιακά συστήματα ασφαλείας [8].

Η ΜΜ χρησιμοποιείται για να εκπαιδεύσει τα συστήματα αυτά με δεδομένα που αντιπροσωπεύουν φυσιολογική δραστηριότητα. Μετά την εκπαίδευση, τα συστήματα είναι σε

θέση να εντοπίζουν αποκλίσεις σε πραγματικό χρόνο, οι οποίες μπορεί να υποδεικνύουν κακόβουλες δραστηριότητες. Η διαδικασία αυτή περιλαμβάνει την ανάλυση μεγάλων ποσοτήτων δεδομένων, επιτρέποντας στα συστήματα να αναγνωρίζουν ακόμη και τις πιο λεπτές αλλαγές στα πρότυπα κυκλοφορίας δεδομένων ή στις ενέργειες των χρηστών, που θα μπορούσαν να υποδεικνύουν μια πιθανή εισβολή [4].

2.2.2.3. Βαθιά Μάθηση και Αναγνώριση Εισβολών

Η Βαθιά Μάθηση (ΒΜ) έχει ενισχύσει περαιτέρω την ικανότητα ανίχνευσης εισβολών, ειδικά μέσω της χρήσης βαθιών νευρωνικών δικτύων (DNNs). Αυτά τα δίκτυα είναι ικανά να επεξεργάζονται και να αναλύουν δεδομένα σε πολλαπλά επίπεδα αφαίρεσης, καθιστώντας τα ιδιαίτερα αποτελεσματικά στην αναγνώριση πολύπλοκων προτύπων που μπορεί να υποδεικνύουν κακόβουλη δραστηριότητα. Η ΒΜ επιτρέπει τη δημιουργία μοντέλων που μπορούν να αναγνωρίσουν ανωμαλίες στα δεδομένα δικτύου, ακόμα και όταν αυτές είναι λεπτές ή διακριτικές [8].

Μια σημαντική πρόκληση που αντιμετωπίζει η ΒΜ στην ανίχνευση εισβολών είναι η ανάγκη για τεράστιες ποσότητες δεδομένων εκπαίδευσης. Για να επιτευχθεί υψηλή ακρίβεια, τα μοντέλα ΒΜ πρέπει να εκπαιδευτούν με δεδομένα που αντιπροσωπεύουν τόσο φυσιολογική δραστηριότητα όσο και επιθέσεις. Ωστόσο, η συλλογή και η ετικετοποίηση αυτών των δεδομένων μπορεί να είναι χρονοβόρα και απαιτητική. Παρά την πολυπλοκότητά τους, τα βαθιά νευρωνικά δίκτυα έχουν αποδειχθεί εξαιρετικά αποτελεσματικά στην ανίχνευση και την πρόληψη κυβερνοεπιθέσεων, καθιστώντας τα ένα βασικό εργαλείο για τα σύγχρονα συστήματα ανίχνευσης εισβολών [4].

2.2.2.4 Εξελιγμένες Τεχνικές και Προκλήσεις

Η χρήση ΤΝ στα συστήματα ανίχνευσης εισβολών δεν είναι χωρίς προκλήσεις. Ένας από τους κύριους περιορισμούς είναι η ανάγκη για υψηλή υπολογιστική ισχύ, η οποία απαιτείται για την εκπαίδευση και τη λειτουργία των αλγορίθμων ΒΜ. Αυτό μπορεί να οδηγήσει σε καθυστερήσεις στην ανίχνευση και την απόκριση στις εισβολές, ειδικά σε περιβάλλοντα όπου η ταχύτητα είναι κρίσιμη. Επιπλέον, τα μοντέλα ΒΜ είναι συχνά ευάλωτα σε επιθέσεις adversarial, όπου κακόβουλοι χρήστες μπορούν να εκμεταλλευτούν τις αδυναμίες των μοντέλων για να παρακάμψουν τα συστήματα ασφαλείας [8].

Επιπλέον, η ακρίβεια των συστημάτων αυτών εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευσή τους. Δεδομένα που είναι προκατειλημμένα ή ανεπαρκή μπορούν να οδηγήσουν σε υψηλά ποσοστά ψευδών θετικών (false positives), γεγονός που μπορεί να δημιουργήσει προβλήματα στην απόκριση και την αξιοπιστία του συστήματος. Η διαχείριση αυτών των δεδομένων, καθώς και η συνεχής ενημέρωση των μοντέλων, είναι κρίσιμες για τη διασφάλιση της αποτελεσματικότητας των συστημάτων ανίχνευσης εισβολών.

2.2.2.5. Μελλοντικές Προοπτικές

Οι μελλοντικές προοπτικές για την ανίχνευση εισβολών με τη χρήση ΤΝ περιλαμβάνουν την ανάπτυξη πιο αποδοτικών και ανθεκτικών αλγορίθμων, οι οποίοι θα μπορούν να λειτουργούν με λιγότερους υπολογιστικούς πόρους και να είναι πιο ανθεκτικοί σε επιθέσεις. Επίσης, η αυτοματοποίηση των διαδικασιών ενημέρωσης και προσαρμογής των μοντέλων αναμένεται να βελτιώσει την αποτελεσματικότητα και την ταχύτητα απόκρισης των συστημάτων αυτών. Η συνεχής έρευνα στον τομέα αυτό θα είναι καθοριστική για την ανάπτυξη νέων τεχνολογιών που θα επιτρέψουν την καλύτερη ανίχνευση και αποτροπή κυβερνοεπιθέσεων.

2.2.3. Ανάλυση Απειλών (Threat Analysis)

Η ανάλυση απειλών αποτελεί θεμέλιο λίθο της Κυβερνοασφάλειας, επιτρέποντας στους οργανισμούς να εντοπίζουν και να αποτρέπουν πιθανές κυβερνοεπιθέσεις προτού αυτές προκαλέσουν ζημιές. Η Τεχνητή Νοημοσύνη (ΤΝ) έχει αλλάξει ριζικά τον τρόπο με τον οποίο προσεγγίζεται η ανάλυση απειλών, επιτρέποντας την ανάπτυξη πιο εξελιγμένων και ακριβών μοντέλων για την αναγνώριση και την αντιμετώπιση των κυβερνοαπειλών.

2.2.3.1. Συλλογή και Επεξεργασία Δεδομένων για Ανάλυση Απειλών

Η ανάλυση απειλών μέσω της ΤΝ ξεκινά με τη συλλογή μεγάλων ποσοτήτων δεδομένων από πολλαπλές πηγές. Αυτά τα δεδομένα μπορεί να περιλαμβάνουν αρχεία καταγραφής συστημάτων, δεδομένα από αισθητήρες ασφαλείας, δικτυακή κίνηση, και ακόμα και πληροφορίες από τις κοινωνικές πλατφόρμες. Με τη χρήση αλγορίθμων Μηχανικής Μάθησης (ΜΜ), η ΤΝ είναι σε θέση να επεξεργάζεται αυτά τα δεδομένα σε πραγματικό χρόνο, αναγνωρίζοντας μοτίβα και ανωμαλίες που μπορεί να υποδηλώνουν μια πιθανή απειλή [5], [10].

Οι αλγόριθμοι αυτοί δεν βασίζονται απλά σε στατιστικές μεθόδους, αλλά χρησιμοποιούν πιο εξελιγμένες τεχνικές, όπως η βαθιά μάθηση (Deep Learning), για να εξάγουν πολύπλοκα χαρακτηριστικά από τα δεδομένα. Αυτό σημαίνει ότι η ΤΝ μπορεί να αναγνωρίσει ακόμα και τα πιο λεπτά μοτίβα, τα οποία μπορεί να είναι πρόδρομοι μιας επικείμενης επίθεσης. Η ικανότητα αυτή επιτρέπει στους οργανισμούς να αντιδρούν πιο γρήγορα και να λαμβάνουν προληπτικά μέτρα πριν η απειλή εξελιχθεί σε σοβαρό περιστατικό.

2.2.3.2. Ανάλυση Συμπεριφοράς και Μοτίβων Απειλών

Η ανάλυση των μοτίβων συμπεριφοράς είναι ένας από τους βασικούς τομείς όπου η ΤΝ έχει αποδειχθεί ιδιαίτερα χρήσιμη. Τα συστήματα ΤΝ μπορούν να παρακολουθούν και να αναλύουν τη συμπεριφορά των χρηστών και των συστημάτων σε πραγματικό χρόνο, αναζητώντας ανωμαλίες που θα μπορούσαν να υποδηλώνουν κακόβουλη δραστηριότητα. Για παράδειγμα, αν ένας χρήστης ξαφνικά αρχίσει να αποκτά πρόσβαση σε περιοχές του συστήματος στις οποίες δεν έχει συνήθως πρόσβαση, αυτό μπορεί να θεωρηθεί ως ένδειξη πιθανής παραβίασης.

Οι τεχνικές αυτές χρησιμοποιούν ένα ευρύ φάσμα δεδομένων, από τις καθημερινές ενέργειες των χρηστών μέχρι τις μικροαλλαγές στη δικτυακή κυκλοφορία, για να αναγνωρίσουν σημάδια κακόβουλης δραστηριότητας. Η ΤΝ μπορεί να συνδυάσει αυτά τα δεδομένα με άλλες πληροφορίες, όπως την ώρα της ημέρας ή την τοποθεσία, για να δημιουργήσει ένα ολοκληρωμένο προφίλ της απειλής και να προβλέψει την πιθανή εξέλιξή της [5].

2.2.3.3. Προγνωστική Ανάλυση και Πρόληψη Απειλών

Ένα από τα πιο ισχυρά χαρακτηριστικά της ΤΝ στην ανάλυση απειλών είναι η δυνατότητα της προγνωστικής ανάλυσης. Χρησιμοποιώντας τεχνικές όπως η μηχανική μάθηση και η βαθιά μάθηση, τα συστήματα ΤΝ μπορούν να αναλύουν ιστορικά δεδομένα και να προβλέπουν πιθανούς κινδύνους προτού αυτοί εκδηλωθούν. Η προγνωστική αυτή ανάλυση επιτρέπει στους οργανισμούς να λαμβάνουν προληπτικά μέτρα, μειώνοντας τον αντίκτυπο των επιθέσεων ή αποτρέποντάς τις εντελώς [10].

Για παράδειγμα, τα συστήματα ΤΝ μπορούν να αναγνωρίζουν σημάδια προετοιμασίας μιας επίθεσης, όπως την αυξημένη δραστηριότητα σε συγκεκριμένους servers ή την ανίχνευση δικτύου

από εξωτερικές πηγές. Αυτές οι ενδείξεις μπορούν να αναλυθούν σε συνδυασμό με άλλες πληροφορίες, επιτρέποντας στον οργανισμό να εντοπίσει πιθανές αδυναμίες στο σύστημά του και να τις διορθώσει πριν εκδηλωθεί μια επίθεση.

2.2.3.4. Αναγνώριση και Ανάλυση Αλυσίδων Επιθέσεων

Η ανάλυση απειλών μέσω της ΤΝ περιλαμβάνει επίσης την αναγνώριση και ανάλυση των αλυσίδων επιθέσεων, δηλαδή τη σειρά των βημάτων που ακολουθεί ένας επιτιθέμενος για να παραβιάσει ένα σύστημα. Οι αλγόριθμοι ΤΝ μπορούν να ανιχνεύσουν και να αναλύσουν τα διάφορα στάδια μιας επίθεσης, από την αρχική διείσδυση έως την τελική εκτέλεση της κακόβουλης δραστηριότητας.

Αυτή η ανάλυση επιτρέπει στους ειδικούς ασφαλείας να κατανοήσουν καλύτερα τις τακτικές και τις τεχνικές που χρησιμοποιούν οι επιτιθέμενοι, καθώς και να αναπτύξουν στρατηγικές άμυνας που μπορούν να σταματήσουν την επίθεση σε οποιοδήποτε σημείο της αλυσίδας [5]. Η κατανόηση της αλυσίδας επιθέσεων βοηθά επίσης στον εντοπισμό των πιο κρίσιμων σημείων στο δίκτυο που πρέπει να προστατευθούν.

2.2.3.5. Ενσωμάτωση της ΤΝ σε Συστήματα Διαχείρισης Απειλών

Τα συστήματα διαχείρισης απειλών, όπως τα SIEM (Security Information and Event Management), έχουν ενσωματώσει την ΤΝ για να βελτιώσουν την ικανότητά τους να ανιχνεύουν και να ανταποκρίνονται σε απειλές. Η ΤΝ επιτρέπει σε αυτά τα συστήματα να αναλύουν τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο, να ανιχνεύουν ανωμαλίες και να δημιουργούν αυτοματοποιημένα συναγερμούς.

Η αυτοματοποίηση αυτή μειώνει την ανάγκη για ανθρώπινη παρέμβαση, επιτρέποντας στους οργανισμούς να ανταποκρίνονται γρηγορότερα και αποτελεσματικότερα στις απειλές. Επιπλέον, τα συστήματα αυτά μπορούν να ενσωματώσουν δεδομένα από πολλαπλές πηγές, παρέχοντας μια ολοκληρωμένη εικόνα της κατάστασης ασφαλείας ενός οργανισμού [10].

2.2.3.6. Μελλοντικές Προοπτικές στην Ανάλυση Απειλών με ΤΝ

Καθώς οι απειλές γίνονται πιο περίπλοκες και εξελιγμένες, η ΤΝ θα συνεχίσει να διαδραματίζει κρίσιμο ρόλο στην ανάλυση και την αποτροπή τους. Η ανάπτυξη αλγορίθμων που μπορούν να λειτουργούν με λιγότερα δεδομένα και να είναι πιο ανθεκτικοί σε παραπλανητικές επιθέσεις αναμένεται να βελτιώσει σημαντικά την αποτελεσματικότητα της ανάλυσης απειλών. Επιπλέον, η ενσωμάτωση της ΤΝ με άλλες τεχνολογίες, όπως το blockchain και τα υπολογιστικά συστήματα αιχμής, θα δημιουργήσει νέες δυνατότητες για την προστασία των συστημάτων από εξελισσόμενες και σύνθετες απειλές.

2.3 Πλεονεκτήματα και Περιορισμοί της ΤΝ στην Κυβερνοασφάλεια

Το κεφάλαιο που ακολουθεί παρουσιάζει τις σημαντικές δυνατότητες που προσφέρει η Τεχνητή Νοημοσύνη (ΤΝ) στην ανίχνευση και αντιμετώπιση των κυβερνοαπειλών, καθώς και τους περιορισμούς που πρέπει να ληφθούν υπόψη. Η ΤΝ έχει συμβάλει στην αυτοματοποίηση της ανίχνευσης απειλών, επιτρέποντας την επεξεργασία μεγάλων όγκων δεδομένων σε πραγματικό χρόνο και την πρόβλεψη επιθέσεων. Παράλληλα, προσφέρει δυνατότητες προσαρμογής και μάθησης από νέα δεδομένα, γεγονός που την καθιστά ιδιαίτερα αποτελεσματική σε ένα μεταβαλλόμενο περιβάλλον απειλών.

Ωστόσο, το κεφάλαιο αναλύει και τις αδυναμίες της τεχνολογίας αυτής, όπως η εξάρτηση από δεδομένα υψηλής ποιότητας, η ευπάθεια σε επιθέσεις τύπου *adversarial* και η έλλειψη διαφάνειας στις αποφάσεις που λαμβάνουν οι αλγόριθμοι. Οι προκλήσεις αυτές τονίζονται ως βασικά εμπόδια που πρέπει να ξεπεραστούν για την πλήρη αξιοποίηση της δυναμικής της ΤΝ στην Κυβερνοασφάλεια.

2.3.1 Αναλυτική Παρουσίαση Δυνατοτήτων και Αδυναμιών

Η Τεχνητή Νοημοσύνη (ΤΝ) έχει φέρει επανάσταση στην Κυβερνοασφάλεια, προσφέροντας νέες και καινοτόμες λύσεις για την αντιμετώπιση των ολοένα και πιο εξελιγμένων κυβερνοαπειλών. Η εφαρμογή της ΤΝ έχει επιτρέψει στους οργανισμούς να ανιχνεύουν, να αναλύουν και να αποκρίνονται σε απειλές με τρόπους που ήταν αδιανόητοι πριν από μερικά χρόνια. Ωστόσο, παρά τις εντυπωσιακές δυνατότητες της ΤΝ, υπάρχουν και σημαντικοί περιορισμοί που πρέπει να ληφθούν υπόψη για να αξιοποιηθεί πλήρως η δυναμική αυτής της τεχνολογίας.

2.3.2 Δυνατότητες της ΤΝ στην Κυβερνοασφάλεια

Η αυτοματοποίηση που προσφέρει η ΤΝ είναι ένα από τα μεγαλύτερα πλεονεκτήματά της στην Κυβερνοασφάλεια. Οι αλγόριθμοι ΤΝ μπορούν να επεξεργαστούν τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο, εντοπίζοντας μοτίβα και ανωμαλίες που υποδηλώνουν κακόβουλες δραστηριότητες. Αυτό επιτρέπει στους οργανισμούς να ανιχνεύουν επιθέσεις πολύ πιο γρήγορα και να αντιδρούν σε πραγματικό χρόνο, μειώνοντας σημαντικά τον χρόνο αντίδρασης και την έκταση της ζημιάς που μπορεί να προκληθεί [1].

Η ΤΝ έχει επίσης τη δυνατότητα να βελτιώνει συνεχώς την απόδοσή της μέσω διαδικασιών συνεχούς μάθησης. Οι αλγόριθμοι μηχανικής μάθησης (ΜΜ) μπορούν να ενημερώνονται και να προσαρμόζονται καθώς αποκτούν νέα δεδομένα, βελτιώνοντας συνεχώς την ακρίβεια και την αποδοτικότητά τους στην ανίχνευση και την πρόβλεψη απειλών. Αυτό σημαίνει ότι τα συστήματα ΤΝ μπορούν να προσαρμόζονται σε νέες και εξελισσόμενες απειλές, διατηρώντας την ασφάλεια των συστημάτων σε ένα συνεχώς μεταβαλλόμενο περιβάλλον [9].

Ένα άλλο σημαντικό πλεονέκτημα της ΤΝ είναι η προγνωστική ανάλυση, η οποία επιτρέπει στους οργανισμούς να προβλέπουν πιθανές απειλές προτού αυτές εκδηλωθούν. Οι αλγόριθμοι ΤΝ μπορούν να αναλύουν ιστορικά δεδομένα και να εντοπίζουν μοτίβα που υποδηλώνουν πιθανές μελλοντικές επιθέσεις. Αυτή η δυνατότητα επιτρέπει στους οργανισμούς να λαμβάνουν προληπτικά μέτρα, μειώνοντας την πιθανότητα επιτυχούς επίθεσης και ενισχύοντας τη συνολική στρατηγική ασφαλείας τους [1].

Η ενσωμάτωση της ΤΝ σε υπάρχοντα συστήματα διαχείρισης απειλών, όπως τα SIEM (Security Information and Event Management), έχει επίσης βελτιώσει την αποτελεσματικότητα αυτών των συστημάτων. Με την ΤΝ, τα SIEM μπορούν να αναλύουν δεδομένα από πολλαπλές πηγές σε πραγματικό χρόνο, ανιχνεύοντας ανωμαλίες και δημιουργώντας αυτοματοποιημένους συναγερμούς. Η αυτοματοποίηση αυτή μειώνει την ανάγκη για ανθρώπινη παρέμβαση, επιτρέποντας στους οργανισμούς να ανταποκρίνονται γρήγορα και αποτελεσματικά σε απειλές [9].

2.3.3 Περιορισμοί της ΤΝ στην Κυβερνοασφάλεια

Παρά τις παραπάνω δυνατότητες, η εφαρμογή της ΤΝ στην Κυβερνοασφάλεια συνοδεύεται από σημαντικούς περιορισμούς. Ένας από τους κυριότερους περιορισμούς είναι η εξάρτηση από τα δεδομένα. Οι αλγόριθμοι ΤΝ βασίζονται σε μεγάλα σύνολα δεδομένων για την εκπαίδευσή τους, και η ποιότητα αυτών των δεδομένων είναι κρίσιμη για την απόδοση των συστημάτων. Εάν τα δεδομένα είναι ανεπαρκή ή προκατειλημμένα, οι αλγόριθμοι ενδέχεται να παράγουν ανακριβή ή παραπλανητικά αποτελέσματα. Αυτό μπορεί να οδηγήσει σε ψευδώς θετικά (false positives) ή ψευδώς αρνητικά (false negatives), τα οποία μπορούν να μειώσουν την αποτελεσματικότητα των συστημάτων ασφαλείας και να προκαλέσουν απώλειες σε πόρους και εμπιστοσύνη [1], [9].

Επιπλέον, οι αλγόριθμοι ΤΝ είναι ευάλωτοι σε επιθέσεις adversarial, όπου οι επιτιθέμενοι εισάγουν σκόπιμα παραπλανητικά δεδομένα για να προκαλέσουν σφάλματα στα συστήματα ΤΝ. Οι επιθέσεις αυτές είναι ιδιαίτερα επικίνδυνες, καθώς μπορούν να παρακάμψουν ακόμα και τα πιο εξελιγμένα συστήματα ασφαλείας, επιτρέποντας στους επιτιθέμενους να εκμεταλλευτούν τις αδυναμίες των μοντέλων ΤΝ [9]. Η ανθεκτικότητα των συστημάτων ΤΝ απέναντι σε τέτοιες επιθέσεις αποτελεί ένα από τα μεγαλύτερα ανοιχτά ερευνητικά ζητήματα.

Ένας άλλος σημαντικός περιορισμός είναι η έλλειψη διαφάνειας και εξηγήσιμης λήψης αποφάσεων στα συστήματα ΤΝ, γνωστό ως «μαύρο κουτί». Οι αποφάσεις που λαμβάνονται από πολύπλοκους αλγόριθμους Βαθιάς Μάθησης είναι συχνά δύσκολο να εξηγηθούν ή να κατανοηθούν από τους ανθρώπους. Αυτή η έλλειψη διαφάνειας μπορεί να μειώσει την εμπιστοσύνη στα συστήματα αυτά, ειδικά σε περιπτώσεις όπου απαιτείται η λήψη κρίσιμων

αποφάσεων για την ασφάλεια. Η δυσκολία στην εξήγηση των αποφάσεων μπορεί, επίσης, να δυσκολέψει τον εντοπισμό και τη διόρθωση σφαλμάτων στα μοντέλα ΤΝ [1].

Η εφαρμογή της ΤΝ στην Κυβερνοασφάλεια απαιτεί σημαντικούς πόρους, τόσο σε επίπεδο υπολογιστικής ισχύος όσο και σε επίπεδο ανθρώπινου δυναμικού. Η ανάπτυξη, η εκπαίδευση και η συντήρηση αποτελεσματικών συστημάτων ΤΝ είναι απαιτητικές διαδικασίες που απαιτούν εξειδικευμένες γνώσεις και εμπειρία. Αυτό μπορεί να αποτελέσει πρόκληση για μικρότερους οργανισμούς που δεν διαθέτουν τους απαραίτητους πόρους για να αναπτύξουν και να διαχειριστούν τέτοια συστήματα [9].

2.3.4 Μελλοντικές Προοπτικές και Εξελίξεις

Η συνεχής έρευνα στον τομέα της ΤΝ και της Κυβερνοασφάλειας αναμένεται να οδηγήσει σε βελτιώσεις που θα ξεπεράσουν πολλούς από τους σημερινούς περιορισμούς. Για παράδειγμα, η ανάπτυξη πιο ανθεκτικών αλγορίθμων που μπορούν να αντισταθούν σε επιθέσεις adversarial είναι ένα από τα κύρια ερευνητικά πεδία. Επίσης, η έρευνα στην εξηγήσιμη ΤΝ (Explainable AI) στοχεύει στην ανάπτυξη συστημάτων που μπορούν να παρέχουν σαφείς και κατανοητές εξηγήσεις για τις αποφάσεις τους, βελτιώνοντας την εμπιστοσύνη των χρηστών στα συστήματα αυτά.

Επιπλέον, η μείωση της εξάρτησης από μεγάλα σύνολα δεδομένων μέσω τεχνικών όπως η μεταφορά μάθησης (transfer learning) και η εκπαίδευση με μικρά δείγματα (few-shot learning) αναμένεται να βελτιώσει την απόδοση των συστημάτων ΤΝ ακόμα και σε περιπτώσεις όπου τα δεδομένα είναι περιορισμένα. Καθώς η ΤΝ συνεχίζει να εξελίσσεται, αναμένεται να διαδραματίσει έναν ακόμα πιο κρίσιμο ρόλο στην Κυβερνοασφάλεια, προσφέροντας νέες λύσεις για την αντιμετώπιση των αυξανόμενων απειλών στον κυβερνοχώρο.

3. Μελέτη Κυβερνοασφάλειας για Συστήματα Τεχνητής Νοημοσύνης

Σε αυτή την ενότητα, θα αναλυθούν οι τρόποι με τους οποίους οι κυβερνοεπιθέσεις επηρεάζουν τα συστήματα Τεχνητής Νοημοσύνης (ΤΝ). Οι απειλές που στοχεύουν την εκπαίδευση και τη λειτουργία των μοντέλων ΤΝ αποτελούν μία αυξανόμενη ανησυχία, καθώς οι επιτιθέμενοι εκμεταλλεύονται τις αδυναμίες της διαδικασίας μάθησης των συστημάτων. Θα εξεταστούν συγκεκριμένες μορφές επιθέσεων, όπως οι επιθέσεις "data poisoning", οι επιθέσεις σε μοντέλα και οι επιθέσεις στις διαδικασίες αλληλεπίδρασης. Μέσω αυτών των απειλών, οι κυβερνοεγκληματίες

προσπαθούν να παραπλανήσουν ή να αποσταθεροποιήσουν τα συστήματα ΤΝ, με επιπτώσεις που μπορεί να είναι καταστροφικές για την ασφάλεια των πληροφοριακών συστημάτων.

3.1 Κατηγοριοποίηση Κυβερνοεπιθέσεων σε Συστήματα ΤΝ

Σε αυτό το σημείο της έρευνας, θα αναπτυχθούν οι θεωρητικές έννοιες που σχετίζονται με τις επιθέσεις στα συστήματα Τεχνητής Νοημοσύνης (ΤΝ) μέσω αλλοίωσης των δεδομένων εκπαίδευσης (data poisoning). Η κατηγοριοποίηση αυτών των επιθέσεων έχει κεντρική σημασία, καθώς στοχεύουν στον πυρήνα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση των μοντέλων, θέτοντας σε κίνδυνο την αξιοπιστία των προβλέψεων τους. Οι τεχνικές που χρησιμοποιούνται για την υλοποίηση αυτών των επιθέσεων, καθώς και οι μέθοδοι ανίχνευσης και πρόληψης, αποτελούν κρίσιμα στοιχεία που θα εξεταστούν σε βάθος, με στόχο τη βελτίωση της ανθεκτικότητας των συστημάτων ΤΝ απέναντι σε τέτοιου είδους απειλές.

3.1.1 Επιθέσεις σε δεδομένα εκπαίδευσης (data poisoning)

Οι επιθέσεις τύπου *data poisoning* αποτελούν έναν από τους πιο ανησυχητικούς και επικίνδυνους τύπους κυβερνοεπιθέσεων στα συστήματα Τεχνητής Νοημοσύνης (ΤΝ), καθώς στοχεύουν στον πυρήνα της διαδικασίας εκπαίδευσης των μοντέλων. Κατά τη διάρκεια της εκπαίδευσης, τα μοντέλα βασίζονται σε μεγάλα σύνολα δεδομένων για να "μάθουν" τα μοτίβα και τις σχέσεις που θα τους επιτρέψουν να κάνουν προβλέψεις σε πραγματικά δεδομένα. Όταν τα δεδομένα που χρησιμοποιούνται για αυτή την εκπαίδευση είναι αλλοιωμένα ή κακόβουλα, το αποτέλεσμα μπορεί να είναι η δραματική μείωση της ακρίβειας των προβλέψεων, με επιπτώσεις που μπορεί να είναι καταστροφικές σε κρίσιμες εφαρμογές [3].

Η επίθεση *data poisoning* μπορεί να υλοποιηθεί με διάφορους τρόπους, αλλά ο πιο συνηθισμένος είναι η εισαγωγή σκόπιμα παραμορφωμένων ή ψευδών δεδομένων στο σύνολο εκπαίδευσης του μοντέλου. Αυτά τα δεδομένα είναι σχεδιασμένα έτσι ώστε να φαίνονται φυσιολογικά ή να διαφεύγουν της ανίχνευσης, ενώ παράλληλα εισάγουν προκαταλήψεις στο μοντέλο. Όταν το μοντέλο συναντά πραγματικά δεδομένα, τα οποία θα πρέπει να αναλύσει, η εκπαίδευση με αυτά τα δηλητηριασμένα δεδομένα το οδηγεί σε λανθασμένες αποφάσεις. Αυτός ο τύπος επίθεσης είναι ιδιαίτερα αποτελεσματικός σε εφαρμογές υψηλού κινδύνου, όπως είναι η αναγνώριση κακόβουλου λογισμικού, η αυτόνομη οδήγηση και η ανίχνευση ανωμαλιών σε δίκτυα [3].

Η δυσκολία με την οποία μπορεί να ανιχνευθεί μια επίθεση *data poisoning* συνίσταται στο γεγονός ότι τα δεδομένα εκπαίδευσης μπορεί να μην παρουσιάζουν εμφανείς διαφορές από τα κανονικά δεδομένα. Οι επιτιθέμενοι εισάγουν συχνά μικρές και μη προφανείς αλλαγές, που όμως επηρεάζουν δυσανάλογα τις αποφάσεις του μοντέλου. Ακόμα και ένα μικρό ποσοστό κακόβουλων δεδομένων μπορεί να επιφέρει μεγάλες αποκλίσεις στις προβλέψεις του μοντέλου, οδηγώντας σε κακή απόδοση, η οποία μπορεί να είναι δύσκολο να αναστραφεί χωρίς εκ νέου εκπαίδευση [5].

Ένα από τα πιο χαρακτηριστικά παραδείγματα αυτών των επιθέσεων αφορά συστήματα αναγνώρισης εικόνας ή βίντεο, όπου μικρές αλλοιώσεις στις εικόνες εκπαίδευσης μπορεί να παραπλανήσουν το μοντέλο να αναγνωρίσει λάθος αντικείμενα ή πρόσωπα. Στα συστήματα ανίχνευσης κακόβουλου λογισμικού, η εισαγωγή παραλλαγμένων δειγμάτων στο σύνολο εκπαίδευσης μπορεί να οδηγήσει στην αποτυχία ανίχνευσης νέων ή παραλλαγμένων απειλών. Οι συνέπειες αυτών των επιθέσεων είναι σοβαρές, ειδικά σε τομείς όπως η κυβερνοασφάλεια και η υγειονομική περίθαλψη, όπου η ακρίβεια των προβλέψεων μπορεί να επηρεάσει άμεσα την ασφάλεια και την υγεία των χρηστών [5].

Για την αντιμετώπιση αυτών των επιθέσεων, έχουν προταθεί αρκετές τεχνικές, με κυριότερη την εφαρμογή αλγορίθμων ανίχνευσης ανωμαλιών, που προσπαθούν να εντοπίσουν και να αποκλείσουν τα κακόβουλα δεδομένα πριν αυτά χρησιμοποιηθούν για την εκπαίδευση των μοντέλων. Αυτές οι τεχνικές μπορούν να ανιχνεύσουν ατυπικές ή ύποπτες συμπεριφορές στα δεδομένα και να προφυλάξουν το σύστημα από επιθέσεις *data poisoning*. Επιπλέον, η χρήση πολλαπλών πηγών δεδομένων ή η διασταυρωτική επικύρωση των δεδομένων μπορεί να συμβάλλει στην αποτροπή εισαγωγής κακόβουλων δεδομένων στα σύνολα εκπαίδευσης [3], [5].

3.1.2 Επιθέσεις σε μοντέλα (model attacks)

Οι επιθέσεις σε μοντέλα Τεχνητής Νοημοσύνης (ΤΝ) συνιστούν μία από τις πλέον επικίνδυνες μορφές επιθέσεων στον χώρο της κυβερνοασφάλειας, καθώς στοχεύουν όχι απλώς στα δεδομένα που χρησιμοποιούνται για την εκπαίδευση των μοντέλων, αλλά στα ίδια τα μοντέλα και τις υποκείμενες αρχιτεκτονικές τους. Σε αυτή την κατηγορία επιθέσεων, οι επιτιθέμενοι εκμεταλλεύονται αδυναμίες ή ελλείψεις στον σχεδιασμό ή τη λειτουργία των μοντέλων για να τα παραπλανήσουν ή να προκαλέσουν δυσλειτουργίες, οδηγώντας τα σε λάθος αποφάσεις. Μία από τις πιο γνωστές και εξελιγμένες μορφές αυτών των επιθέσεων είναι οι επιθέσεις *adversarial*, στις

οποίες οι επιτιθέμενοι τροποποιούν ελάχιστα τα δεδομένα εισόδου με στόχο να παραπλανήσουν το μοντέλο, χωρίς η τροποποίηση να είναι εμφανής στον άνθρωπο [4].

Οι επιθέσεις *adversarial* βασίζονται σε πολύ μικρές και σχεδόν ανεπαίσθητες αλλαγές στα δεδομένα εισόδου, οι οποίες ωστόσο αρκούν για να προκαλέσουν δραματικές αλλαγές στις προβλέψεις του μοντέλου. Αυτές οι αλλαγές μπορεί να είναι τόσο μικρές ώστε να μην γίνονται αντιληπτές από τον χρήστη, αλλά επαρκούν για να παραμορφώσουν τα αποτελέσματα του μοντέλου. Ένα χαρακτηριστικό παράδειγμα αυτής της τεχνικής είναι στις εφαρμογές αναγνώρισης εικόνων, όπου η αλλαγή λίγων μόνο pixel σε μια εικόνα μπορεί να οδηγήσει το μοντέλο να αναγνωρίσει λανθασμένα το αντικείμενο ή το πρόσωπο που απεικονίζεται. Μια εικόνα ενός σκύλου, για παράδειγμα, μπορεί μέσω μιας τέτοιας επίθεσης να αναγνωριστεί λανθασμένα ως γάτα ή ακόμα και ως ένα άσχετο αντικείμενο, όπως ένα αυτοκίνητο [4], [6].

Οι συνέπειες των επιθέσεων *adversarial* δεν περιορίζονται μόνο στην αναγνώριση εικόνας. Σε συστήματα αυτόνομης οδήγησης, για παράδειγμα, μια τέτοια επίθεση θα μπορούσε να κάνει το σύστημα να ερμηνεύσει λανθασμένα μια πινακίδα κυκλοφορίας, με επικίνδυνα αποτελέσματα. Σε άλλες περιπτώσεις, όπως οι επιθέσεις σε εκτελέσιμα αρχεία (*executables*), τέτοιες επιθέσεις μπορούν να χρησιμοποιηθούν για να παρακάμψουν τα συστήματα ανίχνευσης κακόβουλου λογισμικού. Μικρές τροποποιήσεις στον κώδικα ενός κακόβουλου προγράμματος μπορεί να είναι αρκετές για να το καταστήσουν μη ανιχνεύσιμο από τα μοντέλα ανίχνευσης, επιτρέποντας έτσι την ευρεία διάδοσή του [6].

Αυτού του είδους οι επιθέσεις είναι εξαιρετικά δύσκολο να ανιχνευθούν, διότι οι αλλαγές στα δεδομένα εισόδου είναι συχνά ανεπαίσθητες για το ανθρώπινο μάτι ή τις παραδοσιακές μεθόδους ανάλυσης δεδομένων. Το μοντέλο μπορεί να συνεχίζει να λειτουργεί κανονικά και να φαίνεται αξιόπιστο, μέχρι τη στιγμή που θα συναντήσει τα επιμελώς τροποποιημένα δεδομένα εισόδου. Η λανθασμένη απόκριση του μοντέλου μπορεί να έχει σοβαρές επιπτώσεις, ιδίως σε κρίσιμους τομείς εφαρμογών, όπως η ασφάλεια πληροφοριών, η ανίχνευση κακόβουλου λογισμικού, η αναγνώριση προσώπου και τα συστήματα αυτόνομης οδήγησης [4], [6].

Για να αντιμετωπιστούν οι επιθέσεις σε μοντέλα, η ερευνητική κοινότητα έχει προτείνει αρκετές αμυντικές στρατηγικές. Μία από τις πιο συνηθισμένες και αποτελεσματικές μεθόδους είναι η *ανθεκτική εκπαίδευση* (*robust training*). Στην προσέγγιση αυτή, τα μοντέλα εκπαιδεύονται να αναγνωρίζουν και να αποφεύγουν παραπλανητικά δεδομένα εισόδου μέσω της χρήσης

παραδειγμάτων που περιλαμβάνουν επιθέσεις *adversarial* κατά την εκπαίδευση. Αυτά τα παραπλανητικά δεδομένα, γνωστά ως *adversarial examples*, ενσωματώνονται στη διαδικασία εκπαίδευσης, ώστε το μοντέλο να μπορεί να αναγνωρίσει και να απορρίψει τις αντίστοιχες επιθέσεις κατά τη διάρκεια της λειτουργίας του [4].

Επιπλέον, άλλες αμυντικές στρατηγικές περιλαμβάνουν την ενσωμάτωση συστημάτων ανίχνευσης επιθέσεων *adversarial* σε πραγματικό χρόνο. Αυτές οι τεχνικές παρακολουθούν συνεχώς την είσοδο δεδομένων και ανιχνεύουν μικρές ή ύποπτες αλλαγές που μπορεί να υποδεικνύουν προσπάθεια επίθεσης. Επίσης, η συνεχής ενημέρωση των μοντέλων με νέα δεδομένα και η παρακολούθηση της απόδοσής τους σε διάφορα σενάρια χρήσης είναι σημαντική για την ανίχνευση αποκλίσεων στις προβλέψεις. Ακόμη, τεχνικές κρυπτογράφησης των δεδομένων κατά τη διάρκεια της επεξεργασίας τους μπορούν να βοηθήσουν στην προστασία των συστημάτων από τέτοιου είδους επιθέσεις [6].

Τέλος, είναι σημαντικό να τονιστεί πως η προστασία των συστημάτων ΤΝ από επιθέσεις σε μοντέλα απαιτεί συνεχείς προσπάθειες και πολυεπίπεδη προσέγγιση. Ο συνδυασμός αμυντικών τεχνικών, ανθεκτικής εκπαίδευσης, ανίχνευσης επιθέσεων και διαρκούς ενημέρωσης των μοντέλων μπορεί να συμβάλλει στην αύξηση της ανθεκτικότητας των συστημάτων και στη μείωση της πιθανότητας επιτυχημένων επιθέσεων [4], [6].

3.1.3 Επιθέσεις σε διαδικασίες αλληλεπίδρασης (interaction process attacks)

Οι επιθέσεις σε διαδικασίες αλληλεπίδρασης (*interaction process attacks*) αποτελούν έναν σημαντικό κίνδυνο για τα συστήματα Τεχνητής Νοημοσύνης (ΤΝ), καθώς εκμεταλλεύονται τις αδυναμίες στις διαδικασίες εισαγωγής, επεξεργασίας και διαχείρισης δεδομένων. Αυτού του είδους οι επιθέσεις στοχεύουν στο να παρεμποδίσουν ή να διαταράξουν τη φυσιολογική ροή των πληροφοριών που επεξεργάζεται το μοντέλο ΤΝ, με απώτερο σκοπό να προκαλέσουν λανθασμένες ενέργειες ή να αποκτήσουν πρόσβαση σε ευαίσθητα δεδομένα. Αυτές οι επιθέσεις είναι ιδιαίτερα επικίνδυνες όταν αφορούν συστήματα που βασίζονται στην αλληλεπίδραση με εξωτερικές πηγές δεδομένων, όπως αισθητήρες, επικοινωνιακά πρωτόκολλα ή βάσεις δεδομένων [10].

Η βασική ιδέα πίσω από τις επιθέσεις αυτές είναι ότι οι επιτιθέμενοι εκμεταλλεύονται την αδυναμία των συστημάτων να ελέγξουν επαρκώς την ακρίβεια και την ακεραιότητα των δεδομένων που εισάγονται στο σύστημα. Οι επιτιθέμενοι μπορούν να εισάγουν κακόβουλα ή αλλοιωμένα δεδομένα σε πραγματικό χρόνο, προκαλώντας τη διαταραχή της κανονικής λειτουργίας του συστήματος ή ακόμη και την παραβίαση της ασφάλειας του. Αυτό μπορεί να επιτευχθεί μέσω της χειραγώγησης της εισόδου δεδομένων από αισθητήρες, της παραποίησης δεδομένων επικοινωνίας ή της εκμετάλλευσης κενών ασφαλείας κατά την αλληλεπίδραση του συστήματος ΤΝ με άλλες εφαρμογές ή δίκτυα [10].

Για παράδειγμα, σε συστήματα αυτόνομης οδήγησης, οι επιτιθέμενοι θα μπορούσαν να εισαγάγουν αλλοιωμένα δεδομένα από τους αισθητήρες του οχήματος, οδηγώντας το σε λανθασμένες αποφάσεις, όπως να αναγνωρίσει λάθος το περιβάλλον ή να αντιδράσει λανθασμένα σε κινδύνους. Αντίστοιχα, σε εφαρμογές ασφαλείας, οι επιτιθέμενοι θα μπορούσαν να αλλοιώσουν τη ροή δεδομένων ενός συστήματος ανίχνευσης απειλών, προκαλώντας είτε την αποτυχία ανίχνευσης ενός κινδύνου είτε την ενεργοποίηση ψευδών συναγερμών, γεγονός που θα μπορούσε να διαταράξει τη λειτουργία ενός οργανισμού [10].

Ένας άλλος τρόπος εκτέλεσης αυτών των επιθέσεων είναι μέσω της εκμετάλλευσης των αδυναμιών στις διαδικασίες επικύρωσης και ελέγχου πρόσβασης που εφαρμόζονται από το σύστημα. Οι επιτιθέμενοι μπορούν να εισαγάγουν κακόβουλα δεδομένα σε σημεία που δεν ελέγχονται επαρκώς ή να εκμεταλλευτούν κενά ασφαλείας σε διαδικασίες αλληλεπίδρασης, αποκτώντας μη εξουσιοδοτημένη πρόσβαση σε ευαίσθητα δεδομένα ή ακόμη και ελέγχοντας τη ροή των πληροφοριών. Αυτή η παραβίαση μπορεί να οδηγήσει σε επικίνδυνες καταστάσεις, όπως η αποκάλυψη εμπιστευτικών δεδομένων, η αχρήστευση βασικών λειτουργιών ενός οργανισμού, ή ακόμα και η διακοπή κρίσιμων υπηρεσιών [25].

Για να αντιμετωπιστούν οι επιθέσεις σε διαδικασίες αλληλεπίδρασης, οι οργανισμοί πρέπει να εφαρμόσουν αυστηρές πολιτικές και τεχνικές για τη διασφάλιση της ακεραιότητας των δεδομένων και των επικοινωνιών τους. Πρώτον, οι διαδικασίες επικύρωσης δεδομένων πρέπει να είναι αρκετά ισχυρές, ώστε να εντοπίζουν κακόβουλες ή μη τυπικές συμπεριφορές κατά την εισαγωγή δεδομένων. Αυτό περιλαμβάνει τη χρήση μεθόδων ελέγχου ταυτότητας των πηγών δεδομένων, καθώς και τη συνεχή παρακολούθηση των ροών δεδομένων για την ανίχνευση ύποπτης

δραστηριότητας. Επιπλέον, είναι απαραίτητο να εφαρμόζονται πολιτικές ελέγχου πρόσβασης που θα αποτρέπουν τη μη εξουσιοδοτημένη εισαγωγή ή τροποποίηση δεδομένων στο σύστημα [25].

Επιπρόσθετα, οι αμυντικές στρατηγικές πρέπει να περιλαμβάνουν τη χρήση τεχνολογιών που διασφαλίζουν την ασφάλεια των δεδομένων κατά τη μεταφορά τους. Για παράδειγμα, η κρυπτογράφηση των δεδομένων κατά τη διάρκεια της μεταφοράς τους μεταξύ του συστήματος ΤΝ και των εξωτερικών πηγών δεδομένων μπορεί να αποτρέψει την υποκλοπή ή αλλοίωση των δεδομένων από κακόβουλους παράγοντες. Παράλληλα, η εφαρμογή ασφαλών πρωτοκόλλων επικοινωνίας, όπως το HTTPS και η προστασία των API που χρησιμοποιούνται για τη σύνδεση του συστήματος με άλλες υπηρεσίες, είναι κρίσιμες για την αποτροπή τέτοιων επιθέσεων [25].

Τέλος, η ευρεία υιοθέτηση πρακτικών ασφάλειας, όπως ο διαρκής έλεγχος των συστημάτων για κενά ασφαλείας και η τακτική ενημέρωση των συστημάτων ασφαλείας, μπορεί να μειώσει τον κίνδυνο επιθέσεων σε διαδικασίες αλληλεπίδρασης. Η ανάπτυξη και η υλοποίηση συστημάτων που μπορούν να ανιχνεύσουν και να αντιδράσουν γρήγορα σε επιθέσεις, καθώς και η ύπαρξη σχεδίων αποκατάστασης σε περίπτωση παραβίασης, αποτελούν βασικές στρατηγικές για την προστασία των συστημάτων ΤΝ από τέτοιου είδους απειλές [10], [25].

3.2 Επιπτώσεις των Κυβερνοεπιθέσεων στα Πληροφορικά Συστήματα ΤΝ

Η ανάλυση των επιπτώσεων των κυβερνοεπιθέσεων στα συστήματα Τεχνητής Νοημοσύνης (ΤΝ) αποκαλύπτει τις σοβαρές απειλές που αντιμετωπίζουν τόσο τα δεδομένα όσο και η λειτουργικότητα αυτών των συστημάτων. Σε αυτό το πλαίσιο, ιδιαίτερη έμφαση δίνεται στην αλλοίωση δεδομένων (data tampering), στην παραβίαση της ιδιωτικότητας (privacy breaches) και στη διατάραξη της λειτουργικότητας (functional disruption). Αυτές οι επιθέσεις όχι μόνο διακινδυνεύουν την αξιοπιστία των προβλέψεων και των αποφάσεων των συστημάτων ΤΝ, αλλά μπορούν να έχουν σοβαρές συνέπειες σε εφαρμογές υψηλής κρισιμότητας, όπως η υγειονομική περίθαλψη και οι χρηματοοικονομικές υπηρεσίες.

3.2.1 Αλλοίωση δεδομένων (data tampering)

Η αλλοίωση δεδομένων (*data tampering*) αποτελεί μία από τις πιο σοβαρές και επικίνδυνες επιπτώσεις των κυβερνοεπιθέσεων στα συστήματα Τεχνητής Νοημοσύνης (ΤΝ), καθώς μπορεί να καταστρέψει την ακεραιότητα του συστήματος και να υπονομεύσει την αξιοπιστία των προβλέψεων του μοντέλου. Αυτή η μορφή επίθεσης συμβαίνει όταν οι επιτιθέμενοι εισάγουν

σκόπιμα λανθασμένα ή παραποιημένα δεδομένα σε ένα σύστημα ΤΝ με στόχο να αλλοιώσουν τις προβλέψεις και τις αποφάσεις του μοντέλου. Το αποτέλεσμα αυτής της παραβίασης μπορεί να έχει σοβαρές συνέπειες, ιδιαίτερα όταν το σύστημα χρησιμοποιείται σε κρίσιμες εφαρμογές, όπως η ανίχνευση κακόβουλου λογισμικού, η ιατρική διάγνωση, ή η αυτόνομη οδήγηση, όπου η ακρίβεια και η ακεραιότητα των δεδομένων είναι υψίστης σημασίας [6].

Η αλλοίωση δεδομένων μπορεί να λάβει χώρα σε διάφορα στάδια της λειτουργίας ενός συστήματος ΤΝ, με τις επιθέσεις να χωρίζονται σε δύο βασικές κατηγορίες: την αλλοίωση δεδομένων κατά τη φάση της εκπαίδευσης και την αλλοίωση δεδομένων κατά τη φάση της κανονικής λειτουργίας. Στην πρώτη περίπτωση, οι επιτιθέμενοι εισάγουν παραποιημένα δεδομένα στο σύνολο εκπαίδευσης του μοντέλου, μεταβάλλοντας τα μοτίβα και τις σχέσεις που το μοντέλο μαθαίνει. Αυτή η παραβίαση μπορεί να προκαλέσει σημαντικές στρεβλώσεις στη διαδικασία μάθησης, οδηγώντας σε μοντέλα που λαμβάνουν λανθασμένες αποφάσεις και είναι ανίκανα να ανταποκριθούν σωστά σε νέα δεδομένα [6].

Η δεύτερη κατηγορία επιθέσεων αφορά την αλλοίωση δεδομένων κατά τη διάρκεια της κανονικής λειτουργίας του συστήματος ΤΝ. Σε αυτή την περίπτωση, οι επιτιθέμενοι εκμεταλλεύονται τα κενά ασφαλείας στις διαδικασίες αλληλεπίδρασης του μοντέλου με το περιβάλλον του, εισάγοντας παραποιημένα δεδομένα σε πραγματικό χρόνο. Τα δεδομένα αυτά μπορεί να προέρχονται από αισθητήρες, εξωτερικές βάσεις δεδομένων ή άλλες πηγές που χρησιμοποιούνται για την εισαγωγή δεδομένων στο σύστημα. Οι επιθέσεις αυτού του είδους μπορεί να προκαλέσουν άμεσες στρεβλώσεις στις προβλέψεις του μοντέλου, οδηγώντας το να λάβει αποφάσεις που βασίζονται σε εσφαλμένες πληροφορίες [6].

Η αλλοίωση δεδομένων μπορεί να έχει καταστροφικές συνέπειες σε τομείς όπου η αξιοπιστία των συστημάτων ΤΝ είναι ζωτικής σημασίας. Για παράδειγμα, σε χρηματοοικονομικά συστήματα που χρησιμοποιούν ΤΝ για την αυτόματη λήψη αποφάσεων, η αλλοίωση δεδομένων μπορεί να οδηγήσει σε λανθασμένες εκτιμήσεις και αποφάσεις, προκαλώντας οικονομικές ζημίες. Στον τομέα της υγείας, η παραποίηση δεδομένων σε συστήματα ιατρικής διάγνωσης μπορεί να οδηγήσει σε λανθασμένη διάγνωση και θεραπεία, θέτοντας σε κίνδυνο τη ζωή των ασθενών. Στα αυτόνομα οχήματα, η εισαγωγή αλλοιωμένων δεδομένων από αισθητήρες μπορεί να οδηγήσει σε αποτυχία να αναγνωριστούν σωστά οι κίνδυνοι στον δρόμο, γεγονός που μπορεί να έχει τραγικές συνέπειες [19].

Για την αντιμετώπιση της αλλοίωσης δεδομένων, οι οργανισμοί και οι κατασκευαστές συστημάτων ΤΝ πρέπει να εφαρμόζουν ολοκληρωμένες τεχνικές ασφαλείας σε όλα τα στάδια της διαδικασίας. Ένας βασικός μηχανισμός είναι η εφαρμογή τεχνικών επικύρωσης δεδομένων, που διασφαλίζουν ότι τα δεδομένα που εισάγονται στο σύστημα είναι αυθεντικά και ακριβή. Αυτές οι τεχνικές μπορούν να περιλαμβάνουν την πιστοποίηση των πηγών δεδομένων, τη χρήση υπογραφών δεδομένων ή την κρυπτογράφηση για να αποτρέπεται η αλλοίωση κατά τη μεταφορά. Επιπλέον, η χρήση τεχνικών ανίχνευσης ανωμαλιών, που εντοπίζουν δεδομένα με ατυπική ή κακόβουλη συμπεριφορά, μπορεί να βοηθήσει στην αποτροπή της εισαγωγής παραποιημένων δεδομένων στο σύστημα [19].

Ένα άλλο σημαντικό μέτρο είναι η χρήση ασφαλών πρωτοκόλλων επικοινωνίας και η παρακολούθηση των συστημάτων σε πραγματικό χρόνο για να ανιχνευθούν τυχόν απόπειρες αλλοίωσης των δεδομένων. Η εφαρμογή αυτών των στρατηγικών μπορεί να συμβάλλει στη διασφάλιση της ακεραιότητας και της αξιοπιστίας των δεδομένων που χρησιμοποιούν τα συστήματα ΤΝ, μειώνοντας έτσι τον κίνδυνο επιτυχημένων επιθέσεων *data tampering* [6], [19].

3.2.2 Παραβίαση της ιδιωτικότητας (privacy breaches)

Μία από τις πιο σοβαρές επιπτώσεις των κυβερνοεπιθέσεων στα συστήματα Τεχνητής Νοημοσύνης (ΤΝ) είναι η παραβίαση της ιδιωτικότητας (*privacy breaches*), η οποία συμβαίνει όταν επιτιθέμενοι αποκτούν μη εξουσιοδοτημένη πρόσβαση σε ευαίσθητα ή προσωπικά δεδομένα που επεξεργάζεται το σύστημα. Τα δεδομένα αυτά μπορεί να περιλαμβάνουν προσωπικές πληροφορίες χρηστών, δεδομένα υγείας, οικονομικές πληροφορίες, ή άλλα εμπιστευτικά στοιχεία, τα οποία, εάν διαρρεύσουν ή χρησιμοποιηθούν από κακόβουλους παράγοντες, μπορεί να οδηγήσουν σε σοβαρές επιπτώσεις για τα άτομα που αφορούν. Αυτές οι επιπτώσεις μπορεί να περιλαμβάνουν την κλοπή ταυτότητας, την εξαπάτηση ή ακόμη και την καταστροφή της φήμης των θιγόμενων ατόμων [7].

Η παραβίαση της ιδιωτικότητας μπορεί να λάβει χώρα όταν οι επιτιθέμενοι εκμεταλλευτούν αδυναμίες στα πρωτόκολλα ασφαλείας που προστατεύουν τα δεδομένα κατά τη φάση της εισαγωγής, αποθήκευσης ή μετάδοσής τους. Τέτοιου είδους επιθέσεις μπορεί να χρησιμοποιούν τεχνικές όπως η υποκλοπή δεδομένων κατά τη μεταφορά, η παραβίαση βάσεων δεδομένων, ή η πρόσβαση σε αποθηκευμένα δεδομένα μέσω κενών ασφαλείας. Μόλις οι επιτιθέμενοι αποκτήσουν πρόσβαση σε αυτά τα δεδομένα, μπορούν να τα χρησιμοποιήσουν για κακόβουλες ενέργειες, όπως

κλοπή χρημάτων, απόκτηση ψευδούς ταυτότητας ή εκβιασμό των θιγόμενων χρηστών. Επιπλέον, τα δεδομένα αυτά μπορούν να πωληθούν στην παράνομη αγορά του κυβερνοχώρου (*dark web*), όπου γίνονται αντικείμενο εκμετάλλευσης από άλλους κακόβουλους χρήστες [20].

Ένα από τα πιο χαρακτηριστικά παραδείγματα παραβίασης της ιδιωτικότητας αφορά τα δεδομένα υγείας. Τα συστήματα ΤΝ που χρησιμοποιούνται για διαγνωστικούς σκοπούς μπορεί να επεξεργάζονται ιδιαίτερα ευαίσθητες πληροφορίες σχετικά με την υγεία των ασθενών. Εάν οι πληροφορίες αυτές πέσουν στα χέρια επιτιθέμενων, μπορεί να αποκαλύψουν ευαίσθητες προσωπικές πληροφορίες που αφορούν την κατάσταση υγείας των ασθενών, τις θεραπείες τους, και άλλες κρίσιμες λεπτομέρειες. Αυτή η παραβίαση μπορεί να προκαλέσει όχι μόνο προσωπικές και επαγγελματικές δυσκολίες για τους ασθενείς, αλλά και να οδηγήσει σε σοβαρές νομικές και οικονομικές συνέπειες για τους οργανισμούς που χειρίζονται τα δεδομένα αυτά [20].

Η ανάγκη για αποτελεσματικά μέτρα προστασίας της ιδιωτικότητας είναι επιτακτική, ιδιαίτερα καθώς οι επιθέσεις αυτού του τύπου γίνονται ολοένα και πιο εξελιγμένες. Για να αποτραπούν οι παραβιάσεις της ιδιωτικότητας, οι οργανισμοί πρέπει να εφαρμόζουν προηγμένα μέτρα ασφαλείας που διασφαλίζουν την προστασία των δεδομένων σε όλα τα στάδια επεξεργασίας τους. Μία από τις πιο αποτελεσματικές μεθόδους είναι η κρυπτογράφηση των δεδομένων κατά τη μεταφορά τους ή την αποθήκευσή τους, ώστε ακόμα και αν τα δεδομένα πέσουν στα χέρια επιτιθέμενων, να μην μπορούν να χρησιμοποιηθούν χωρίς τη σωστή αποκρυπτογράφηση [7].

Επιπλέον, η αυστηρή διαχείριση δικαιωμάτων πρόσβασης είναι απαραίτητη για τη μείωση του κινδύνου παραβίασης. Μόνο εξουσιοδοτημένα άτομα πρέπει να έχουν πρόσβαση στα ευαίσθητα δεδομένα, και οι οργανισμοί πρέπει να εφαρμόζουν συστήματα πολυεπίπεδης ασφάλειας για την επαλήθευση της ταυτότητας των χρηστών που προσπαθούν να αποκτήσουν πρόσβαση. Παράλληλα, η ανωνυμοποίηση των δεδομένων, όπου είναι εφικτό, μπορεί να συμβάλει σημαντικά στη μείωση του αντίκτυπου σε περίπτωση παραβίασης. Όταν τα προσωπικά στοιχεία δεν μπορούν να ταυτοποιηθούν άμεσα, η ζημιά που προκαλείται από την παραβίαση είναι μικρότερη [7].

Η τακτική αναθεώρηση των πολιτικών ασφαλείας, σε συνδυασμό με την εφαρμογή διεθνών προτύπων όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR), είναι επίσης καθοριστικής σημασίας για την προστασία της ιδιωτικότητας των χρηστών. Ο GDPR παρέχει ένα πλαίσιο για την προστασία των προσωπικών δεδομένων και θέτει αυστηρούς κανόνες για το πώς

πρέπει να συλλέγονται, να αποθηκεύονται και να διαχειρίζονται τα δεδομένα, διασφαλίζοντας την ιδιωτικότητα των χρηστών [20].

3.2.3 Διαταραχή της λειτουργικότητας (functional disruption)

Η διαταραχή της λειτουργικότητας (*functional disruption*) αποτελεί μία από τις πιο κρίσιμες επιπτώσεις των κυβερνοεπιθέσεων στα συστήματα Τεχνητής Νοημοσύνης (ΤΝ), καθώς μπορεί να προκαλέσει σοβαρές δυσλειτουργίες στις διαδικασίες που εκτελούν αυτά τα συστήματα. Οι επιτιθέμενοι εκμεταλλεύονται κενά ασφαλείας στα μοντέλα ΤΝ και τις υποδομές που τα υποστηρίζουν, με στόχο να διακόψουν τη λειτουργία τους, να τα παραπλανήσουν ή ακόμη και να τα αχρηστέψουν εντελώς. Αυτού του είδους οι επιθέσεις μπορεί να περιλαμβάνουν τη διακοπή της παροχής υπηρεσιών, την τροποποίηση των αλγορίθμων λειτουργίας ή ακόμη και την πλήρη αδυναμία του συστήματος να ανταποκριθεί στις βασικές του λειτουργίες [16].

Η διαταραχή της λειτουργικότητας προκύπτει συχνά από επιθέσεις που στοχεύουν άμεσα στα μοντέλα ΤΝ ή στη συνεχή ροή δεδομένων που τα υποστηρίζει. Δεδομένου ότι τα συστήματα ΤΝ εξαρτώνται σε μεγάλο βαθμό από την αδιάκοπη επικοινωνία και επεξεργασία δεδομένων σε πραγματικό χρόνο, οποιαδήποτε διακοπή σε αυτή τη ροή μπορεί να οδηγήσει σε σοβαρές επιπτώσεις. Ένα χαρακτηριστικό παράδειγμα αυτής της απειλής είναι τα συστήματα αυτόνομης οδήγησης, όπου η παροχή δεδομένων από αισθητήρες είναι κρίσιμη για τη σωστή λειτουργία του οχήματος. Μια διακοπή ή παραποίηση των δεδομένων που προέρχονται από τους αισθητήρες μπορεί να οδηγήσει σε αποτυχία του οχήματος να αντιδράσει σε πραγματικούς κινδύνους, όπως πεζοί ή άλλα οχήματα στον δρόμο, γεγονός που θέτει σε άμεσο κίνδυνο τη ζωή των επιβατών και των πεζών [21].

Οι συνέπειες των επιθέσεων που διαταράσσουν τη λειτουργικότητα των συστημάτων ΤΝ δεν περιορίζονται μόνο σε αυτόνομες τεχνολογίες. Η διαταραχή της λειτουργικότητας μπορεί επίσης να επηρεάσει κρίσιμες υποδομές, όπως ενεργειακά δίκτυα, συστήματα υγείας, ή χρηματοοικονομικές υπηρεσίες. Σε ενεργειακά δίκτυα, για παράδειγμα, μια διακοπή της λειτουργίας μπορεί να προκαλέσει εκτεταμένα μπλακ άουτ, με σοβαρές επιπτώσεις στην οικονομία και την ασφάλεια. Σε χρηματοοικονομικά συστήματα, η διαταραχή της λειτουργίας μπορεί να οδηγήσει σε λανθασμένες συναλλαγές ή να προκαλέσει σημαντικές απώλειες εσόδων για οργανισμούς και ιδιώτες. Επιπλέον, τέτοιες επιθέσεις μπορούν να προκαλέσουν πανικό,

δημιουργώντας αίσθημα αβεβαιότητας στους χρήστες των συστημάτων και μειώνοντας την εμπιστοσύνη τους σε τεχνολογικές εφαρμογές ΤΝ [16], [21].

Οι επιθέσεις που στοχεύουν στη διατάραξη της λειτουργικότητας των συστημάτων ΤΝ χρησιμοποιούνται επίσης από επιτιθέμενους για να ασκήσουν πίεση ή να εκβιάσουν οργανισμούς. Σε περιπτώσεις όπου η αδιάλειπτη λειτουργία είναι απαραίτητη, οι επιτιθέμενοι μπορούν να απειλήσουν με πλήρη διακοπή της λειτουργίας των συστημάτων, απαιτώντας λύτρα ή άλλα ανταλλάγματα. Αυτό μπορεί να είναι ιδιαίτερα καταστροφικό για οργανισμούς που βασίζονται στη συνεχή παροχή υπηρεσιών, όπως νοσοκομεία, χρηματοπιστωτικά ιδρύματα και πάροχοι ενέργειας [21].

Για την αποτροπή των επιθέσεων που στοχεύουν στη διατάραξη της λειτουργικότητας είναι απαραίτητο να υιοθετηθούν πολυεπίπεδες αμυντικές στρατηγικές. Πρώτον, η εφαρμογή εφεδρικών συστημάτων είναι κρίσιμη για την εξασφάλιση της συνέχισης της λειτουργίας, ακόμη και σε περίπτωση επίθεσης. Τα εφεδρικά συστήματα μπορούν να ενεργοποιηθούν αυτόματα σε περίπτωση που ανιχνευθεί διακοπή της λειτουργίας, διασφαλίζοντας την αδιάλειπτη παροχή των υπηρεσιών που παρέχει το σύστημα ΤΝ [16].

Επιπλέον, η συνεχής παρακολούθηση της λειτουργίας των συστημάτων και η ανίχνευση ανωμαλιών είναι απαραίτητα εργαλεία για την πρόληψη τέτοιων επιθέσεων. Η χρήση τεχνολογιών ανίχνευσης δυσλειτουργιών μπορεί να βοηθήσει στην έγκαιρη ανίχνευση προβλημάτων, επιτρέποντας στους διαχειριστές συστημάτων να αντιδράσουν πριν οι επιθέσεις προκαλέσουν σοβαρή ζημιά. Επιπλέον, η διασφάλιση της ασφαλούς επικοινωνίας μέσω κρυπτογραφημένων καναλιών και η τακτική συντήρηση των συστημάτων είναι βασικά μέτρα για την αποτροπή των κυβερνοεπιθέσεων που στοχεύουν στην παραβίαση της λειτουργικότητας [16], [21].

Τέλος, είναι σημαντικό οι οργανισμοί να ενσωματώνουν προληπτικά μέτρα, όπως δοκιμές ανθεκτικότητας, όπου τα συστήματα δοκιμάζονται σε συνθήκες επιθέσεων προσομοίωσης για να αξιολογηθεί η ανθεκτικότητά τους. Αυτές οι στρατηγικές όχι μόνο συμβάλλουν στην προστασία των συστημάτων ΤΝ από επιθέσεις που στοχεύουν στη λειτουργικότητά τους, αλλά και στη συνολική αύξηση της ασφάλειας και της αξιοπιστίας τους [21].

3.3 Στρατηγικές Αντιμετώπισης Κυβερνοεπιθέσεων

3.3.1 Προληπτικά μέτρα (preventive measures)

Οι στρατηγικές αντιμετώπισης κυβερνοεπιθέσεων σε συστήματα Τεχνητής Νοημοσύνης (ΤΝ) ξεκινούν από την εφαρμογή προληπτικών μέτρων, τα οποία στοχεύουν στη θωράκιση των συστημάτων πριν αυτά γίνουν στόχοι επιθέσεων. Η υιοθέτηση προληπτικών στρατηγικών είναι ζωτικής σημασίας για την αποτροπή των κυβερνοεπιθέσεων, καθώς περιορίζει τις πιθανότητες οι επιτιθέμενοι να εκμεταλλευτούν αδυναμίες των συστημάτων. Η προληπτική ασφάλεια περιλαμβάνει τόσο την ενίσχυση των τεχνικών χαρακτηριστικών των συστημάτων ΤΝ όσο και την ευαισθητοποίηση των χρηστών και διαχειριστών σχετικά με τις απειλές [5].

Μία από τις βασικές προληπτικές στρατηγικές είναι η τακτική ενημέρωση των συστημάτων ΤΝ και των αλγορίθμων τους. Οι αλγόριθμοι και τα μοντέλα ΤΝ εξελίσσονται συνεχώς, και η μη εφαρμογή ενημερώσεων μπορεί να αφήσει τα συστήματα ευάλωτα σε γνωστές αδυναμίες. Η αναβάθμιση των λογισμικών ασφαλείας και των τεχνολογιών προστασίας σε τακτά χρονικά διαστήματα είναι απαραίτητη για την αποτροπή των επιθέσεων. Η εφαρμογή κρυπτογραφημένων πρωτοκόλλων επικοινωνίας και η χρήση τεχνικών ασφαλούς επικύρωσης ταυτότητας είναι επίσης κρίσιμα μέτρα για την προστασία των δεδομένων και των αλληλεπιδράσεων των συστημάτων ΤΝ με το περιβάλλον τους [8].

Επιπλέον, οι οργανισμοί πρέπει να επενδύσουν σε τεχνολογίες ανίχνευσης ανωμαλιών, οι οποίες μπορούν να εντοπίζουν ύποπτες συμπεριφορές σε πραγματικό χρόνο. Οι τεχνικές αυτές χρησιμοποιούν προχωρημένα μοντέλα ανάλυσης δεδομένων για να εντοπίσουν ανωμαλίες στη ροή δεδομένων ή στις προβλέψεις του μοντέλου. Αν και δεν μπορούν πάντα να αποτρέψουν μια επίθεση, η έγκαιρη ανίχνευση ανωμαλιών μπορεί να δώσει τη δυνατότητα στους διαχειριστές των συστημάτων να αντιδράσουν γρήγορα και να περιορίσουν την έκταση της ζημιάς [5].

Ακόμη, ένα σημαντικό προληπτικό μέτρο είναι η εκπαίδευση των ανθρώπων που διαχειρίζονται τα συστήματα ΤΝ. Η ενημέρωση των χρηστών σχετικά με τους κινδύνους των κυβερνοεπιθέσεων και η καλλιέργεια καλών πρακτικών ασφαλείας, όπως η δημιουργία ισχυρών κωδικών πρόσβασης και η εφαρμογή αυστηρών πολιτικών πρόσβασης στα δεδομένα, μπορούν να αποτρέψουν ανθρώπινα λάθη που συχνά οδηγούν σε επιτυχημένες επιθέσεις. Η ευαισθητοποίηση γύρω από τις

απειλές του κυβερνοχώρου μπορεί να συμβάλλει στη μείωση των αδυναμιών των συστημάτων και στην προώθηση μιας κουλτούρας ασφάλειας σε ολόκληρο τον οργανισμό [8].

Ένα ακόμα προληπτικό μέτρο που πρέπει να ληφθεί υπόψη είναι η ανάπτυξη ανθεκτικών μοντέλων ΤΝ. Τα ανθεκτικά μοντέλα έχουν τη δυνατότητα να λειτουργούν αποτελεσματικά ακόμη και όταν δέχονται επιθέσεις ή παραπλανητικά δεδομένα. Η ενσωμάτωση τεχνικών *adversarial training*, όπου το μοντέλο εκπαιδεύεται να αντιμετωπίζει επιθέσεις μέσω παραποιημένων δεδομένων, είναι μια από τις μεθόδους που έχουν αποδειχθεί αποτελεσματικές στην ενίσχυση της ανθεκτικότητας των συστημάτων ΤΝ [5].

3.3.2 Τεχνικές αντίδρασης και αποκατάστασης (reaction and recovery techniques)

Εκτός από τα προληπτικά μέτρα, οι τεχνικές αντίδρασης και αποκατάστασης (*reaction and recovery techniques*) είναι απαραίτητες για την αποτελεσματική αντιμετώπιση των επιθέσεων όταν αυτές καταφέρουν να περάσουν τα προληπτικά συστήματα άμυνας. Οι επιθέσεις που καταφέρνουν να διαπεράσουν τα μέτρα ασφαλείας απαιτούν άμεση και συντονισμένη αντίδραση για να περιοριστεί η ζημιά και να αποκατασταθεί η λειτουργία των συστημάτων ΤΝ. Ο συνδυασμός ταχείας αντίδρασης και διαδικασιών αποκατάστασης είναι καθοριστικός για τη μείωση των επιπτώσεων των κυβερνοεπιθέσεων [14].

Η άμεση αντίδραση σε μία κυβερνοεπίθεση περιλαμβάνει τη γρήγορη ανίχνευση της επίθεσης και την απομόνωση των προσβεβλημένων συστημάτων για να αποτραπεί η περαιτέρω εξάπλωση της ζημιάς. Τα συστήματα πρέπει να διαθέτουν μηχανισμούς ανίχνευσης επιθέσεων σε πραγματικό χρόνο, όπως τεχνολογίες IDS (Intrusion Detection Systems), που επιτρέπουν την αυτόματη ειδοποίηση των διαχειριστών σε περίπτωση ύποπτης δραστηριότητας. Επίσης, η χρήση τεχνικών *rate limiting* και *firewalls* μπορεί να συμβάλλει στην αποτροπή της εξάπλωσης επιθέσεων, όπως οι επιθέσεις DDoS, που στοχεύουν στη διακοπή των υπηρεσιών ενός συστήματος [22].

Η αποκατάσταση μετά από μία επίθεση απαιτεί την αποκατάσταση της κανονικής λειτουργίας του συστήματος, την επιδιόρθωση των κενών ασφαλείας που εκμεταλλεύτηκαν οι επιτιθέμενοι, και την επανεκπαίδευση των μοντέλων ΤΝ, εάν κριθεί απαραίτητο. Για παράδειγμα, εάν η επίθεση επηρέασε τη λειτουργία των δεδομένων εκπαίδευσης, το μοντέλο μπορεί να χρειαστεί επανεκπαίδευση με νέα, καθαρά δεδομένα για να αποκατασταθεί η ακρίβεια των προβλέψεών του.

Επίσης, η διαδικασία αποκατάστασης περιλαμβάνει τη λεπτομερή ανάλυση της επίθεσης για να προσδιοριστεί η πηγή και η μέθοδος που χρησιμοποιήθηκε, επιτρέποντας στους οργανισμούς να ενισχύσουν τις άμυνές τους έναντι παρόμοιων απειλών στο μέλλον [14].

Ένας άλλος σημαντικός παράγοντας στην αποκατάσταση είναι η διατήρηση εφεδρικών αντιγράφων δεδομένων και συστημάτων. Η τακτική δημιουργία εφεδρικών αντιγράφων δεδομένων, καθώς και η ύπαρξη διαδικασιών αποκατάστασης καταστροφών (*disaster recovery plans*), επιτρέπει την ταχύτερη επαναφορά του συστήματος μετά από μια επίθεση. Με αυτόν τον τρόπο, οι οργανισμοί μπορούν να ελαχιστοποιήσουν τον χρόνο διακοπής λειτουργίας και να επιστρέψουν σε κανονική λειτουργία το συντομότερο δυνατόν [22].

Η επικοινωνία με τους εμπλεκόμενους φορείς κατά τη διάρκεια και μετά από μια κυβερνοεπίθεση είναι κρίσιμη. Οι οργανισμοί πρέπει να ενημερώνουν άμεσα τους χρήστες τους για την επίθεση και τα μέτρα που λαμβάνονται για την προστασία των δεδομένων τους. Η διαφάνεια σε αυτές τις περιπτώσεις συμβάλλει στη διατήρηση της εμπιστοσύνης των χρηστών και του κοινού [22].

Τέλος, η συνεχής αναβάθμιση των διαδικασιών αντίδρασης και αποκατάστασης με βάση τις νέες απειλές είναι απαραίτητη. Οι οργανισμοί πρέπει να επανεξετάζουν τακτικά τις στρατηγικές τους, να αξιολογούν την αποτελεσματικότητά τους και να προσαρμόζονται στις εξελίξεις του κυβερνοχώρου, ώστε να διασφαλίζουν ότι τα συστήματά τους είναι επαρκώς προετοιμασμένα να αντιμετωπίσουν μελλοντικές επιθέσεις [14], [22].

4. Αμυντικές Τεχνικές για την Προστασία των Συστημάτων Τεχνητής Νοημοσύνης

Σε αυτή την ενότητα θα παρουσιαστούν οι βασικές αρχές της αντιμετώπισης κυβερνοεπιθέσεων μέσω προληπτικών μέτρων και τεχνικών αποκατάστασης. Τα προληπτικά μέτρα στοχεύουν στη θωράκιση των συστημάτων Τεχνητής Νοημοσύνης πριν από την εκδήλωση κυβερνοεπιθέσεων, ενώ οι τεχνικές αποκατάστασης επιτρέπουν τη γρήγορη ανάκαμψη των συστημάτων σε περίπτωση επίθεσης. Μέσα από την ανάλυση αυτή, θα αναδειχθούν οι κύριες στρατηγικές που εφαρμόζονται για την ενίσχυση της ασφάλειας των συστημάτων, καθώς και οι προκλήσεις που προκύπτουν κατά την εφαρμογή αυτών των μεθόδων.

4.1 Τεχνικές Αντίστασης σε Κυβερνοεπιθέσεις

Σε αυτή την ενότητα θα παρουσιαστούν οι βασικές αρχές των στρατηγικών αντιμετώπισης κυβερνοεπιθέσεων στα συστήματα Τεχνητής Νοημοσύνης (ΤΝ), με έμφαση στα προληπτικά μέτρα και τις τεχνικές αντίδρασης και αποκατάστασης. Οι στρατηγικές αυτές διαδραματίζουν κρίσιμο ρόλο στην ενίσχυση της ασφάλειας των συστημάτων ΤΝ, καθώς οι κυβερνοεπιθέσεις γίνονται ολοένα και πιο πολύπλοκες και συχνές. Οι προληπτικές μέθοδοι στοχεύουν στη θωράκιση των συστημάτων πριν από τις επιθέσεις, ενώ οι τεχνικές αντίδρασης επικεντρώνονται στην άμεση αντιμετώπιση και αποκατάσταση των συστημάτων σε περίπτωση παραβίασης.

4.1.1 Ανθεκτική εκπαίδευση μοντέλων (robust model training)

Η ανθεκτική εκπαίδευση μοντέλων (*robust model training*) αποτελεί μία από τις πιο αποτελεσματικές αμυντικές τεχνικές για την προστασία της Τεχνητής Νοημοσύνης (ΤΝ) από κυβερνοεπιθέσεις. Στόχος αυτής της μεθόδου είναι να ενισχύσει την ικανότητα των μοντέλων ΤΝ να αντέχουν επιθέσεις, διατηρώντας την ακρίβεια και την αξιοπιστία των προβλέψεών τους, ακόμη και όταν εκτίθενται σε κακόβουλα ή παραποιημένα δεδομένα. Οι κυβερνοεπιθέσεις, όπως οι *adversarial attacks*, όπου μικρές τροποποιήσεις στα δεδομένα εισόδου μπορεί να οδηγήσουν σε σημαντικές αποκλίσεις στις προβλέψεις, καθιστούν την ανθεκτική εκπαίδευση μία από τις σημαντικότερες στρατηγικές στην ασφάλεια των μοντέλων [4].

Η ανθεκτική εκπαίδευση βασίζεται στην εισαγωγή παραπλανητικών παραδειγμάτων (*adversarial examples*) κατά τη διαδικασία εκπαίδευσης των μοντέλων. Αυτά τα παραδείγματα είναι σκόπιμα παραποιημένα δεδομένα, σχεδιασμένα για να προκαλέσουν εσφαλμένες προβλέψεις από το μοντέλο. Η εκπαίδευση των μοντέλων με τέτοια δεδομένα έχει ως στόχο να αυξήσει την ικανότητά τους να αναγνωρίζουν και να απορρίπτουν επιθέσεις κατά τη διάρκεια της λειτουργίας τους. Τα μοντέλα που έχουν εκπαιδευτεί με ανθεκτική εκπαίδευση μπορούν να παραμείνουν αξιόπιστα ακόμη και όταν αντιμετωπίζουν *adversarial attacks*, μειώνοντας την πιθανότητα λανθασμένων αποφάσεων [6].

Ένας από τους τρόπους με τους οποίους υλοποιείται η ανθεκτική εκπαίδευση είναι μέσω της χρήσης ειδικών αλγορίθμων, οι οποίοι εκπαιδεύουν το μοντέλο όχι μόνο με καθαρά δεδομένα, αλλά και με παραποιημένα δεδομένα, τα οποία αποτελούν ρεαλιστικές προσομοιώσεις πιθανών επιθέσεων. Αυτή η τεχνική επιτρέπει στο μοντέλο να "μάθει" πώς να αναγνωρίζει και να αποφεύγει παραπλανητικά δεδομένα, βελτιώνοντας την ανθεκτικότητά του. Επιπλέον, σε συνδυασμό με την

ενίσχυση των αλγορίθμων μάθησης, οι ερευνητές αναπτύσσουν αλγορίθμους που μπορούν να εντοπίζουν αυτόματα τα δεδομένα που είναι επιρρεπή σε επιθέσεις και να τα διορθώνουν πριν χρησιμοποιηθούν για προβλέψεις [4].

Ένα χαρακτηριστικό παράδειγμα εφαρμογής της ανθεκτικής εκπαίδευσης είναι στα συστήματα αναγνώρισης εικόνων. Μικρές αλλαγές σε μια εικόνα, οι οποίες μπορεί να μην είναι ορατές στον άνθρωπο, μπορεί να προκαλέσουν λανθασμένες προβλέψεις από το σύστημα, όπως η λανθασμένη αναγνώριση αντικειμένων ή προσώπων. Με την ανθεκτική εκπαίδευση, τα μοντέλα μαθαίνουν να αγνοούν αυτές τις μικρές αλλαγές, μειώνοντας την πιθανότητα επιτυχίας των επιτιθέμενων. Παρόμοιες τεχνικές χρησιμοποιούνται και σε άλλους τομείς, όπως η αυτόνομη οδήγηση και τα χρηματοοικονομικά συστήματα, όπου οι επιθέσεις μπορούν να οδηγήσουν σε σοβαρές οικονομικές ή ακόμη και ανθρώπινες απώλειες [6].

Ωστόσο, η ανθεκτική εκπαίδευση δεν είναι απόλυτα αλάνθαστη. Παρά την αποτελεσματικότητά της, οι επιθέσεις εξελίσσονται συνεχώς, και οι επιτιθέμενοι αναπτύσσουν νέες τεχνικές για να ξεπεράσουν τις άμυνες που παρέχει η ανθεκτική εκπαίδευση. Αυτό σημαίνει ότι η ανθεκτική εκπαίδευση πρέπει να συνδυάζεται με άλλες αμυντικές στρατηγικές για να εξασφαλιστεί η μέγιστη προστασία των συστημάτων ΤΝ. Επίσης, η χρήση αυτής της τεχνικής μπορεί να απαιτεί περισσότερους πόρους για την εκπαίδευση των μοντέλων, κάτι που μπορεί να αυξήσει τον χρόνο και το κόστος ανάπτυξης [4], [6].

Συνολικά, η ανθεκτική εκπαίδευση μοντέλων αποτελεί μία από τις πιο αποτελεσματικές μεθόδους για την προστασία της ΤΝ από κυβερνοεπιθέσεις. Η εισαγωγή παραπλανητικών δεδομένων κατά τη διάρκεια της εκπαίδευσης ενισχύει την ανθεκτικότητα των μοντέλων και μειώνει την πιθανότητα επιτυχίας των επιτιθέμενων. Παρά τις προκλήσεις που μπορεί να αντιμετωπίζει η εφαρμογή της, η ανθεκτική εκπαίδευση αποτελεί αναπόσπαστο μέρος της σύγχρονης άμυνας απέναντι στις απειλές του κυβερνοχώρου.

4.1.2 Προστασία δεδομένων (data protection)

Η προστασία των δεδομένων (*data protection*) αποτελεί έναν άλλο κρίσιμο πυλώνα στην αμυντική στρατηγική για την προστασία της Τεχνητής Νοημοσύνης από κυβερνοεπιθέσεις. Η αποτελεσματική προστασία των δεδομένων είναι ζωτικής σημασίας, καθώς τα δεδομένα αποτελούν τη βάση για την εκπαίδευση και τη λειτουργία των μοντέλων ΤΝ. Οι κυβερνοεπιθέσεις που στοχεύουν στα δεδομένα, όπως οι επιθέσεις αλλοίωσης δεδομένων (*data tampering*) και οι

παραβιάσεις ιδιωτικότητας, μπορούν να οδηγήσουν σε ανακριβείς προβλέψεις, παραβίαση της ιδιωτικότητας των χρηστών και καταστροφή της ακεραιότητας των συστημάτων [1].

Η προστασία των δεδομένων ξεκινά από τη σωστή διαχείριση και αποθήκευσή τους. Ένας από τους βασικούς τρόπους προστασίας των δεδομένων είναι η κρυπτογράφηση, η οποία διασφαλίζει ότι τα δεδομένα που αποθηκεύονται ή μεταδίδονται είναι προσβάσιμα μόνο από εξουσιοδοτημένους χρήστες. Η χρήση ισχυρών αλγορίθμων κρυπτογράφησης, όπως το AES (Advanced Encryption Standard), μπορεί να αποτρέψει την πρόσβαση κακόβουλων χρηστών σε ευαίσθητα δεδομένα, ακόμη και αν αυτά τα δεδομένα υποκλαπούν κατά τη μεταφορά ή αποθήκευσή τους [10].

Επιπλέον, η προστασία των δεδομένων περιλαμβάνει και την εφαρμογή ισχυρών μέτρων ελέγχου πρόσβασης. Οι οργανισμοί που διαχειρίζονται συστήματα ΤΝ πρέπει να εφαρμόζουν πολιτικές αυστηρής διαχείρισης δικαιωμάτων πρόσβασης, διασφαλίζοντας ότι μόνο εξουσιοδοτημένοι χρήστες μπορούν να έχουν πρόσβαση στα δεδομένα ή να τα τροποποιούν. Η χρήση τεχνικών όπως η πολυεπίπεδη ταυτοποίηση (*multi-factor authentication*) και τα συστήματα διαχείρισης ταυτότητας και πρόσβασης (*Identity and Access Management - IAM*) μπορούν να περιορίσουν τον κίνδυνο παραβίασης δεδομένων από μη εξουσιοδοτημένους χρήστες [1].

Επιπλέον, η ανωνυμοποίηση δεδομένων αποτελεί μία ακόμη στρατηγική προστασίας των δεδομένων, ειδικά σε περιπτώσεις όπου τα συστήματα ΤΝ επεξεργάζονται ευαίσθητα προσωπικά δεδομένα, όπως δεδομένα υγείας ή οικονομικές πληροφορίες. Η ανωνυμοποίηση διασφαλίζει ότι τα προσωπικά στοιχεία των χρηστών δεν μπορούν να ταυτοποιηθούν άμεσα, ακόμα και αν τα δεδομένα παραβιαστούν. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη σε συστήματα που λειτουργούν σύμφωνα με κανονισμούς, όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR) [10].

Η τακτική δημιουργία εφεδρικών αντιγράφων ασφαλείας είναι επίσης ζωτικής σημασίας για την προστασία των δεδομένων. Τα εφεδρικά αντίγραφα δεδομένων εξασφαλίζουν ότι, σε περίπτωση παραβίασης ή αλλοίωσης των αρχικών δεδομένων, ο οργανισμός μπορεί να αποκαταστήσει τη λειτουργία του με ελάχιστες απώλειες. Είναι σημαντικό τα αντίγραφα ασφαλείας να αποθηκεύονται με ασφαλή τρόπο, σε απομονωμένα περιβάλλοντα, ώστε να μην είναι προσβάσιμα από κακόβουλους παράγοντες [10].

Τέλος, είναι σημαντικό να υπάρχει διαρκής παρακολούθηση και έλεγχος των δεδομένων για να εντοπίζονται γρήγορα τυχόν παραβιάσεις ή ανωμαλίες. Τα συστήματα παρακολούθησης πρέπει να ενσωματώνουν τεχνολογίες ανίχνευσης επιθέσεων σε πραγματικό χρόνο, ώστε να εντοπίζονται έγκαιρα οι προσπάθειες παραβίασης των δεδομένων. Επιπλέον, οι οργανισμοί πρέπει να εφαρμόζουν τακτικούς ελέγχους ασφαλείας και δοκιμές ευπαθειών για να διασφαλίζουν την ασφάλεια των δεδομένων τους και να προσαρμόζονται στις νέες απειλές [1].

4.2 Σχεδιασμός Ασφαλών Συστημάτων Τεχνητής Νοημοσύνης

4.2.1 Διανεμημένη μάθηση με προστασία ιδιωτικότητας (privacy-preserving distributed learning)

Η διανεμημένη μάθηση με προστασία ιδιωτικότητας αποτελεί μια πρωτοποριακή τεχνολογία που στοχεύει στην ασφαλή επεξεργασία δεδομένων από διαφορετικές πηγές, διατηρώντας παράλληλα την προστασία της ιδιωτικότητας των χρηστών. Σε αυτήν την προσέγγιση, τα δεδομένα δεν χρειάζεται να μεταφερθούν σε έναν κεντρικό διακομιστή ή να συγχωνευθούν σε ένα ενιαίο σύνολο για την εκπαίδευση των μοντέλων. Αντιθέτως, τα δεδομένα παραμένουν στους τοπικούς κόμβους, και η συνεργασία για την εκπαίδευση του μοντέλου επιτυγχάνεται με την ανταλλαγή των παραμέτρων που προκύπτουν από τα τοπικά δεδομένα. Με αυτόν τον τρόπο, αποφεύγεται η άμεση έκθεση των δεδομένων, κάτι που είναι κρίσιμο σε περιπτώσεις που απαιτείται η αυστηρή τήρηση κανόνων ιδιωτικότητας και ασφάλειας [2].

Αυτή η τεχνολογία είναι ιδιαίτερα σημαντική σε περιβάλλοντα όπου τα δεδομένα είναι εξαιρετικά ευαίσθητα, όπως για παράδειγμα στα νοσοκομεία, στις τράπεζες και σε κρατικούς οργανισμούς. Στα νοσοκομεία, η διανεμημένη μάθηση επιτρέπει την εκπαίδευση μοντέλων τεχνητής νοημοσύνης για την πρόβλεψη ασθενειών ή την ανάλυση ιατρικών εικόνων χωρίς να χρειάζεται να μεταφερθούν τα δεδομένα των ασθενών εκτός των νοσοκομείων. Αυτό σημαίνει ότι οι πληροφορίες για την υγεία των ασθενών δεν αποκαλύπτονται σε άλλους φορείς ή συστήματα, διασφαλίζοντας έτσι ότι τα δικαιώματα ιδιωτικότητας των ασθενών παραμένουν σεβαστά [17]. Η διατήρηση της τοπικής φύλαξης των δεδομένων σε συνδυασμό με την κεντρική διαχείριση των παραμέτρων του μοντέλου βελτιώνει σημαντικά την απόδοση των συστημάτων, διότι επιτρέπει την εκπαίδευση σε ένα ευρύτερο φάσμα δεδομένων από πολλαπλές πηγές, χωρίς να εκτίθενται τα ίδια τα δεδομένα.

Ένα από τα πιο κοινά παραδείγματα της διανεμημένης μάθησης με προστασία ιδιωτικότητας είναι η ομοσπονδιακή μάθηση (*federated learning*). Σε αυτό το πλαίσιο, πολλοί τοπικοί κόμβοι εκπαιδεύουν το δικό τους μοντέλο με τα δεδομένα που έχουν στη διάθεσή τους, και στη συνέχεια αποστέλλουν τις ενημερώσεις των παραμέτρων τους σε έναν κεντρικό διακομιστή. Ο κεντρικός διακομιστής συγκεντρώνει αυτές τις ενημερώσεις για να δημιουργήσει ένα κεντρικό μοντέλο, χωρίς ωστόσο να μεταφέρει τα αρχικά δεδομένα. Αυτή η προσέγγιση μειώνει τον κίνδυνο παραβίασης δεδομένων, καθώς τα δεδομένα παραμένουν στους τοπικούς κόμβους και δεν εκτίθενται σε εξωτερικούς κινδύνους [17]. Επιπλέον, η χρήση τεχνικών όπως η *Differential Privacy* διασφαλίζει ότι οι ενημερώσεις των παραμέτρων που αποστέλλονται στον κεντρικό διακομιστή είναι ασφαλείς και δεν μπορούν να αποκαλύψουν ευαίσθητες πληροφορίες για τα αρχικά δεδομένα.

Η διανεμημένη μάθηση με προστασία ιδιωτικότητας έχει ευρεία εφαρμογή σε διάφορους τομείς. Στις χρηματοοικονομικές υπηρεσίες, για παράδειγμα, οι τράπεζες μπορούν να χρησιμοποιήσουν την τεχνολογία αυτή για να αναλύσουν δεδομένα πελατών από πολλαπλά τραπεζικά υποκαταστήματα χωρίς να μοιράζονται τα πραγματικά δεδομένα μεταξύ τους. Στα δίκτυα IoT, όπου πολλές συσκευές παράγουν δεδομένα σε πραγματικό χρόνο, η διανεμημένη μάθηση επιτρέπει τη συνεργασία των συσκευών για την εκπαίδευση ενός κεντρικού μοντέλου, χωρίς την ανάγκη μεταφοράς μεγάλων όγκων δεδομένων μέσω του δικτύου [2].

Ωστόσο, παρά τα πλεονεκτήματα της διανεμημένης μάθησης με προστασία ιδιωτικότητας, υπάρχουν και προκλήσεις που σχετίζονται με την εφαρμογή της. Αρχικά, οι τοπικοί κόμβοι που συμμετέχουν στη διαδικασία πρέπει να διαθέτουν επαρκείς υπολογιστικούς πόρους για να διαχειριστούν την εκπαίδευση των μοντέλων τους. Σε περιπτώσεις όπου οι πόροι είναι περιορισμένοι, η διαδικασία εκπαίδευσης μπορεί να καθυστερήσει σημαντικά ή να οδηγήσει σε ανακριβή αποτελέσματα. Επίσης, υπάρχει η πρόκληση της ανθεκτικότητας απέναντι σε κακόβουλους κόμβους, οι οποίοι μπορεί να εισαγάγουν ψευδείς ή παραποιημένες παραμέτρους στο κεντρικό μοντέλο, υπονομεύοντας έτσι την ακρίβεια και την αξιοπιστία του συνολικού συστήματος [2]. Για την αντιμετώπιση αυτού του κινδύνου, απαιτούνται μέθοδοι ανίχνευσης και απομόνωσης κακόβουλων κόμβων, καθώς και αυστηρότερος έλεγχος των ενημερώσεων παραμέτρων πριν την ενσωμάτωσή τους στο κεντρικό μοντέλο.

Συνοψίζοντας, η διανεμημένη μάθηση με προστασία ιδιωτικότητας προσφέρει μια ισχυρή λύση για την ασφαλή εκπαίδευση μοντέλων ΤΝ με δεδομένα από πολλαπλές πηγές, ενώ ταυτόχρονα διατηρεί την προστασία της ιδιωτικότητας. Παρά τις προκλήσεις που συνοδεύουν την εφαρμογή της, η τεχνολογία αυτή παρουσιάζει σημαντικές προοπτικές για την ευρύτερη υιοθέτησή της σε τομείς όπως η υγεία, οι χρηματοοικονομικές υπηρεσίες και τα δίκτυα IoT.

4.2.2 Κρυπτογραφημένη μάθηση (encrypted learning)

Η κρυπτογραφημένη μάθηση αποτελεί μια από τις πιο καινοτόμες και ασφαλείς τεχνολογίες που εφαρμόζονται στην Τεχνητή Νοημοσύνη (ΤΝ) για την προστασία της ιδιωτικότητας και της ασφάλειας των δεδομένων. Η βασική αρχή πίσω από αυτή την τεχνολογία είναι ότι τα δεδομένα κρυπτογραφούνται πριν τη μεταφορά ή την επεξεργασία τους από τα συστήματα ΤΝ, διασφαλίζοντας ότι κανένα εξουσιοδοτημένο ή κακόβουλο μέρος, είτε διαχειριστές είτε επιτιθέμενοι, δεν μπορεί να έχει πρόσβαση στα αρχικά δεδομένα καθ' όλη τη διάρκεια της διαδικασίας μάθησης ή πρόβλεψης. Με αυτή την προσέγγιση, τα δεδομένα παραμένουν κρυφά ακόμα και όταν υφίστανται επεξεργασία, περιορίζοντας τον κίνδυνο διαρροής ευαίσθητων πληροφοριών [19].

Η πιο προηγμένη μορφή κρυπτογραφημένης μάθησης βασίζεται σε αλγόριθμους όπως η *Fully Homomorphic Encryption* (FHE), μια μέθοδος που επιτρέπει την εκτέλεση υπολογισμών απευθείας σε κρυπτογραφημένα δεδομένα χωρίς να χρειάζεται αποκρυπτογράφηση. Αυτή η τεχνολογία δίνει τη δυνατότητα στα συστήματα ΤΝ να εκπαιδεύονται και να προβλέπουν αποτελέσματα με πλήρη προστασία των δεδομένων, διατηρώντας την ιδιωτικότητα ακόμη και όταν τα δεδομένα επεξεργάζονται σε εξωτερικούς ή απομακρυσμένους κόμβους. Αυτό καθιστά την FHE μια από τις ασφαλέστερες λύσεις για την προστασία δεδομένων κατά τη διάρκεια της επεξεργασίας, αλλά συνοδεύεται από αυξημένο υπολογιστικό κόστος. Το κόστος αυτό οφείλεται στην πολυπλοκότητα της κρυπτογραφίας και μπορεί να περιορίσει την απόδοση των συστημάτων ΤΝ, ειδικά σε εφαρμογές που απαιτούν γρήγορες αποφάσεις και επεξεργασία μεγάλου όγκου δεδομένων σε πραγματικό χρόνο [21].

Ένα κλασικό παράδειγμα εφαρμογής της κρυπτογραφημένης μάθησης είναι στον τομέα των χρηματοοικονομικών υπηρεσιών, όπου οι τράπεζες και οι χρηματοπιστωτικοί οργανισμοί μπορούν να αναλύουν δεδομένα συναλλαγών πελατών για την εκπαίδευση μοντέλων ανάλυσης κινδύνου, χωρίς να διακυβεύεται η ασφάλεια των ευαίσθητων πληροφοριών. Χρησιμοποιώντας

κρυπτογραφημένη μάθηση, οι τράπεζες μπορούν να προστατεύσουν τα δεδομένα τους από πιθανές κυβερνοεπιθέσεις, ενώ παράλληλα οι συνεργάτες τους δεν αποκτούν πρόσβαση στις λεπτομέρειες των πελατών. Ακόμη και σε περίπτωση παραβίασης ασφάλειας, τα κρυπτογραφημένα δεδομένα παραμένουν αδιάβαστα χωρίς την αποκρυπτογράφηση [19]. Αυτή η προσέγγιση προσφέρει ιδιαίτερα υψηλό επίπεδο προστασίας σε κρίσιμους τομείς όπου η διαρροή δεδομένων θα μπορούσε να έχει καταστροφικές συνέπειες.

Επιπλέον, η κρυπτογραφημένη μάθηση μπορεί να συνδυαστεί με άλλες τεχνικές ασφαλείας, όπως η κρυπτογραφία δημόσιου κλειδιού (*public key encryption*), για την ασφαλή μεταφορά των δεδομένων μεταξύ διαφόρων κόμβων ή συστημάτων. Αυτή η τεχνολογία είναι ιδιαίτερα χρήσιμη σε περιβάλλοντα διανεμημένων συστημάτων, όπως τα υπολογιστικά νέφη (*cloud computing*), όπου τα δεδομένα πρέπει να μεταφέρονται συχνά μεταξύ γεωγραφικά απομακρυσμένων κέντρων δεδομένων. Χρησιμοποιώντας κρυπτογράφηση, τα δεδομένα παραμένουν ασφαλή σε όλη τη διάρκεια της μεταφοράς τους, εξασφαλίζοντας ότι οι επιτιθέμενοι δεν μπορούν να αποκτήσουν πρόσβαση στα ευαίσθητα στοιχεία ακόμη και αν καταφέρουν να παραβιάσουν το δίκτυο [21].

Παρά τα σημαντικά οφέλη που προσφέρει η κρυπτογραφημένη μάθηση, υπάρχουν ορισμένες προκλήσεις που αφορούν κυρίως την απόδοσή της. Το υψηλό υπολογιστικό κόστος που απαιτείται για την κρυπτογράφηση και την αποκρυπτογράφηση των δεδομένων μπορεί να επηρεάσει αρνητικά τη δυνατότητα εφαρμογής της σε συστήματα που χρειάζονται άμεσες αποφάσεις ή επεξεργασία μεγάλων ποσοτήτων δεδομένων σε πραγματικό χρόνο. Για παράδειγμα, σε εφαρμογές όπως η αυτόνομη οδήγηση ή τα χρηματοοικονομικά συστήματα υψηλής ταχύτητας, η χρήση κρυπτογράφησης μπορεί να επιβραδύνει σημαντικά την επεξεργασία, κάτι που μπορεί να μειώσει την αποτελεσματικότητα του συστήματος [19].

Ωστόσο, με τη συνεχή εξέλιξη των τεχνολογιών κρυπτογράφησης και την αύξηση της υπολογιστικής ισχύος, οι προκλήσεις αυτές αναμένεται να μειωθούν στο μέλλον. Η χρήση πιο αποδοτικών αλγορίθμων κρυπτογράφησης και η ανάπτυξη ταχύτερων υπολογιστικών συστημάτων θα επιτρέψουν την εφαρμογή της κρυπτογραφημένης μάθησης σε μεγαλύτερη κλίμακα, διατηρώντας παράλληλα την ασφάλεια των δεδομένων και την ιδιωτικότητα των χρηστών. Σε συνδυασμό με άλλες τεχνολογίες προστασίας της ιδιωτικότητας, η κρυπτογραφημένη μάθηση θα διαδραματίσει καθοριστικό ρόλο στην ασφάλεια των συστημάτων ΤΝ, τόσο στον δημόσιο όσο και στον ιδιωτικό τομέα [19], [21].

4.3 Τεχνολογίες Ασφάλειας για τα Συστήματα Τεχνητής Νοημοσύνης

Σε αυτό το σημείο θα παρουσιαστούν οι βασικές αρχές για τον σχεδιασμό ασφαλών συστημάτων Τεχνητής Νοημοσύνης (ΤΝ), με έμφαση στις τεχνικές που προστατεύουν την ιδιωτικότητα και την ασφάλεια των δεδομένων. Η διανεμημένη μάθηση με προστασία ιδιωτικότητας αποτελεί μια πρωτοποριακή προσέγγιση που επιτρέπει την εκπαίδευση μοντέλων ΤΝ χωρίς τη μεταφορά των δεδομένων σε κεντρικούς διακομιστές, διασφαλίζοντας την ιδιωτικότητα των χρηστών. Παράλληλα, η κρυπτογραφημένη μάθηση προσφέρει ένα επιπλέον επίπεδο ασφάλειας μέσω της χρήσης εξελιγμένων κρυπτογραφικών αλγορίθμων, προστατεύοντας τα δεδομένα καθ' όλη τη διάρκεια της επεξεργασίας τους.

4.3.1 Τεχνολογίες ανίχνευσης απειλών (threat detection technologies)

Οι τεχνολογίες ανίχνευσης απειλών αποτελούν έναν από τους πιο σημαντικούς μηχανισμούς προστασίας για την Τεχνητή Νοημοσύνη (ΤΝ), καθώς επιτρέπουν την έγκαιρη αναγνώριση και αποτροπή κακόβουλων δραστηριοτήτων που μπορεί να υπονομεύσουν την ασφάλεια και την αξιοπιστία των συστημάτων. Η έγκαιρη ανίχνευση απειλών είναι ζωτικής σημασίας, διότι επιτρέπει στους οργανισμούς να προλαμβάνουν επιθέσεις πριν αυτές εξελιχθούν σε σοβαρά περιστατικά ασφάλειας. Οι τεχνολογίες ανίχνευσης απειλών χρησιμοποιούν εξελιγμένους αλγορίθμους ανάλυσης δεδομένων και τεχνικές μηχανικής μάθησης για την παρακολούθηση των αλληλεπιδράσεων των συστημάτων ΤΝ με το περιβάλλον τους, εντοπίζοντας οποιεσδήποτε αποκλίσεις από τη φυσιολογική λειτουργία, που θα μπορούσαν να είναι ενδείξεις κακόβουλης δραστηριότητας [4].

Μία από τις πιο διαδεδομένες τεχνολογίες σε αυτό το πλαίσιο είναι τα Συστήματα Ανίχνευσης Εισβολών (*Intrusion Detection Systems* - IDS), τα οποία έχουν ως στόχο τη συνεχή παρακολούθηση της κυκλοφορίας δικτύου και των δεδομένων εισόδου σε πραγματικό χρόνο. Τα IDS χρησιμοποιούν αλγορίθμους μηχανικής μάθησης για να αναλύσουν τεράστιους όγκους δεδομένων, προκειμένου να ανιχνεύσουν πρότυπα ή ανωμαλίες που ενδέχεται να υποδεικνύουν κακόβουλες δραστηριότητες. Για παράδειγμα, τα IDS μπορούν να εντοπίσουν την παραποίηση δεδομένων, την εισαγωγή κακόβουλου λογισμικού ή άλλες ενέργειες που υποδηλώνουν ότι το σύστημα βρίσκεται υπό επίθεση. Αυτό γίνεται με την εκπαίδευση των μοντέλων ΤΝ σε δεδομένα από προηγούμενες επιθέσεις, ώστε να αναγνωρίζουν τα ίδια μοτίβα ή συμπεριφορές στο μέλλον [18].

Το κύριο πλεονέκτημα αυτών των τεχνολογιών είναι η ικανότητά τους να ανιχνεύουν ακόμη και νέες ή άγνωστες απειλές, που δεν έχουν εμφανιστεί στο παρελθόν. Με την εξέλιξη των μεθόδων ανίχνευσης, όπως η *Deep Learning*, τα συστήματα μπορούν να αναγνωρίζουν λεπτές και δύσκολα ανιχνεύσιμες αλλαγές στα δεδομένα εισόδου, που θα μπορούσαν να υποδηλώνουν μια επίθεση. Αυτή η δυνατότητα είναι ιδιαίτερα χρήσιμη στις περιπτώσεις των επιθέσεων *adversarial*, στις οποίες οι επιτιθέμενοι πραγματοποιούν μικρές αλλαγές στα δεδομένα για να παραπλανήσουν τα συστήματα ΤΝ και να προκαλέσουν λανθασμένες προβλέψεις. Οι παραδοσιακές μέθοδοι ανίχνευσης μπορεί να μην επαρκούν για τέτοιου είδους επιθέσεις, γι' αυτό η χρήση πιο εξελιγμένων τεχνικών γίνεται αναγκαία [4].

Η ανάπτυξη τεχνολογιών ανίχνευσης απειλών παραμένει συνεχώς εξελισσόμενη για να μπορεί να ανταπεξέλθει στην ολοένα αυξανόμενη πολυπλοκότητα των επιθέσεων. Οι επιτιθέμενοι υιοθετούν όλο και πιο περίπλοκες μεθόδους για να διαπεράσουν τα συστήματα ασφαλείας, αναγκάζοντας τους ερευνητές να εξελίσσουν συνεχώς τις μεθόδους ανίχνευσης. Μία από τις πιο σύγχρονες προσεγγίσεις περιλαμβάνει την ενσωμάτωση μεθόδων *anomaly detection*, όπου τα μοντέλα μηχανικής μάθησης εντοπίζουν ανωμαλίες στα δεδομένα ή στη συμπεριφορά του συστήματος που δεν αντιστοιχούν σε φυσιολογικές καταστάσεις. Αυτές οι ανωμαλίες μπορεί να αποτελούν ένδειξη ότι βρίσκεται σε εξέλιξη μια κακόβουλη δραστηριότητα [18].

Παρά τα μεγάλα πλεονεκτήματα που προσφέρουν αυτές οι τεχνολογίες, εξακολουθούν να υπάρχουν προκλήσεις. Ένα από τα μεγαλύτερα προβλήματα είναι η παρουσία ψευδών θετικών ή αρνητικών ανιχνεύσεων. Τα ψευδώς θετικά σήματα ενδέχεται να οδηγήσουν σε μη απαραίτητες ενέργειες από τους διαχειριστές των συστημάτων, ενώ τα ψευδώς αρνητικά μπορεί να επιτρέψουν την ανεμπόδιστη διείσδυση κακόβουλων παραγόντων στο σύστημα. Για την αντιμετώπιση αυτών των προβλημάτων, χρησιμοποιούνται τεχνικές *ensemble learning*, οι οποίες συνδυάζουν τα αποτελέσματα πολλαπλών μοντέλων ανίχνευσης για να βελτιώσουν την ακρίβεια και να μειώσουν τις πιθανότητες ψευδών συναγερμών. Τα συστήματα που χρησιμοποιούν τέτοιες τεχνικές είναι πιο αξιόπιστα και μπορούν να ανιχνεύσουν πιο ακριβώς τις πραγματικές απειλές [18].

Συνολικά, οι τεχνολογίες ανίχνευσης απειλών είναι απαραίτητες για την προστασία των συστημάτων ΤΝ από κακόβουλες επιθέσεις. Η χρήση εξελιγμένων αλγορίθμων και η συνεχής ανάπτυξη νέων μεθόδων ανίχνευσης διασφαλίζουν ότι τα συστήματα ΤΝ παραμένουν ανθεκτικά στις νέες απειλές που προκύπτουν. Ωστόσο, είναι σημαντικό να συνεχιστεί η εξέλιξη αυτών των

τεχνολογιών ώστε να μπορούν να προσαρμόζονται στις συνεχώς μεταβαλλόμενες απειλές του κυβερνοχώρου [4], [18].

4.3.2 Τεχνολογίες προστασίας από επιθέσεις (attack protection technologies)

Οι τεχνολογίες προστασίας από επιθέσεις διαδραματίζουν καθοριστικό ρόλο στην ασφάλεια των συστημάτων Τεχνητής Νοημοσύνης (ΤΝ), εστιάζοντας τόσο στην πρόληψη όσο και στον περιορισμό των επιπτώσεων που μπορεί να έχουν οι επιθέσεις στη λειτουργία των συστημάτων. Αυτές οι τεχνολογίες δεν στοχεύουν μόνο στον άμεσο εντοπισμό των επιθέσεων, αλλά και στην ανάπτυξη στρατηγικών που θα αποτρέψουν τις επιθέσεις ή θα ελαχιστοποιήσουν τις επιπτώσεις τους. Πρόκειται για έναν συνδυασμό προληπτικών μέτρων και αμυντικών τεχνικών, που ενεργοποιούνται είτε προτού εντοπιστεί μια επίθεση είτε αμέσως μετά, διασφαλίζοντας ότι το σύστημα μπορεί να συνεχίσει να λειτουργεί ομαλά χωρίς σοβαρές επιπτώσεις [8].

Μία από τις πιο γνωστές τεχνολογίες που χρησιμοποιούνται για την προστασία των συστημάτων ΤΝ είναι τα *firewalls* και τα συστήματα *rate limiting*. Τα *firewalls* είναι ουσιαστικά φίλτρα που εμποδίζουν την είσοδο κακόβουλων δεδομένων ή τη διέλευση ανεπιθύμητης κυκλοφορίας στο σύστημα, αποτρέποντας πιθανές επιθέσεις από εξωτερικούς παράγοντες. Παράλληλα, τα συστήματα *rate limiting* περιορίζουν τον αριθμό των αιτημάτων που μπορούν να γίνουν σε ένα σύστημα σε ένα συγκεκριμένο χρονικό διάστημα, μειώνοντας έτσι την πιθανότητα επιτυχημένων επιθέσεων *DDoS* (Distributed Denial of Service). Σε τέτοιες επιθέσεις, οι επιτιθέμενοι βομβαρδίζουν το σύστημα με τεράστιο αριθμό αιτημάτων, προκαλώντας υπερφόρτωση και τελικά διακοπή της λειτουργίας του. Τα συστήματα *rate limiting* εξασφαλίζουν ότι το σύστημα δεν θα καταρρεύσει υπό το βάρος των αιτημάτων αυτών, διατηρώντας τη λειτουργία του και μειώνοντας τις επιπτώσεις της επίθεσης [23].

Μία άλλη σημαντική τεχνολογία για την προστασία από επιθέσεις, και ειδικά από επιθέσεις τύπου *adversarial*, είναι η τεχνική *adversarial training*. Σε αυτό το πλαίσιο, τα μοντέλα ΤΝ εκπαιδεύονται όχι μόνο με κανονικά δεδομένα αλλά και με παραπλανητικά παραδείγματα, τα οποία ενδέχεται να χρησιμοποιηθούν από επιτιθέμενους για να παραπλανήσουν το μοντέλο. Μέσω της εισαγωγής αυτών των παραδειγμάτων στα σύνολα εκπαίδευσης, το μοντέλο μαθαίνει να αναγνωρίζει και να αποκρούει τις επιθέσεις, αυξάνοντας την ανθεκτικότητά του. Αυτή η μέθοδος είναι ιδιαίτερα αποτελεσματική στην αντιμετώπιση επιθέσεων παραποίησης εισόδου, όπου μικρές αλλαγές στα δεδομένα μπορούν να προκαλέσουν λανθασμένες προβλέψεις από το σύστημα [8].

Με την εφαρμογή αυτής της τεχνικής, τα συστήματα ΤΝ γίνονται πιο ανθεκτικά και έχουν τη δυνατότητα να ανταποκριθούν ακόμη και σε εξελιγμένες μορφές επιθέσεων.

Μια επιπλέον τεχνική που χρησιμοποιείται για την προστασία από επιθέσεις είναι το *sandboxing*, το οποίο απομονώνει τις κρίσιμες διαδικασίες του συστήματος σε εικονικά ή απομονωμένα περιβάλλοντα. Αυτό σημαίνει ότι αν μια επίθεση στοχεύσει και παραβιάσει μια συγκεκριμένη διαδικασία, δεν θα έχει τη δυνατότητα να εξαπλωθεί σε άλλες λειτουργίες του συστήματος. Το *sandboxing* διασφαλίζει ότι ακόμα και σε περίπτωση επιτυχούς επίθεσης, η ζημιά θα παραμείνει περιορισμένη και δεν θα επηρεάσει άλλες, κρίσιμες λειτουργίες. Έτσι, το σύστημα συνεχίζει να λειτουργεί και οι επιπτώσεις της επίθεσης περιορίζονται σε συγκεκριμένες διαδικασίες χωρίς να επηρεαστεί η συνολική απόδοσή του [23].

Παρά τη χρήση αυτών των εξελιγμένων τεχνολογιών προστασίας, είναι σημαντικό να τονιστεί ότι οι επιθέσεις στον κυβερνοχώρο εξελίσσονται συνεχώς. Οι επιτιθέμενοι βρίσκουν διαρκώς νέους τρόπους για να παρακάμπτουν τα υπάρχοντα μέτρα ασφαλείας, με αποτέλεσμα οι οργανισμοί να πρέπει να ενημερώνουν και να αναβαθμίζουν τα συστήματά τους τακτικά. Οι αμυντικές τεχνικές που είναι αποτελεσματικές σήμερα ενδέχεται να μην είναι το ίδιο αποδοτικές αύριο, καθώς οι απειλές εξελίσσονται με ραγδαίο ρυθμό. Για να παραμείνουν ασφαλή, τα συστήματα ΤΝ πρέπει να ενσωματώνουν συνεχείς ενημερώσεις ασφαλείας και να προσαρμόζουν τις αμυντικές τους στρατηγικές για να αντιμετωπίζουν τις νέες μορφές επιθέσεων που προκύπτουν [23].

Τέλος, οι οργανισμοί πρέπει να υιοθετούν μια πολυεπίπεδη προσέγγιση στην ασφάλεια των συστημάτων τους, συνδυάζοντας διαφορετικές τεχνολογίες και αμυντικές στρατηγικές. Μια μεμονωμένη τεχνολογία, όσο ισχυρή και αν είναι, δεν αρκεί για να προστατεύσει πλήρως ένα σύστημα από όλες τις πιθανές απειλές. Ο συνδυασμός τεχνολογιών όπως τα *firewalls*, το *sandboxing*, τα συστήματα *rate limiting* και οι τεχνικές *adversarial training* προσφέρει μια πιο ολοκληρωμένη και ανθεκτική προσέγγιση στην προστασία των συστημάτων ΤΝ από επιθέσεις [8], [23].

5. Συνδυαστικές Εφαρμογές Τεχνητής Νοημοσύνης και Κυβερνοασφάλειας

Σε αυτό το κεφάλαιο εξετάζονται οι βασικές παράμετροι που σχετίζονται με τη συνδυαστική χρήση της Τεχνητής Νοημοσύνης (ΤΝ) και των συστημάτων κυβερνοασφάλειας. Οι σύγχρονες

τεχνολογίες ΤΝ ενισχύουν τις δυνατότητες ανίχνευσης, πρόληψης και αντίδρασης σε κυβερνοεπιθέσεις, προσφέροντας λύσεις που προσαρμόζονται δυναμικά σε νέες απειλές και προσφέρουν σημαντική ασφάλεια σε κρίσιμες υποδομές και συστήματα. Στο πλαίσιο αυτό, θα αναλυθούν οι στρατηγικές προληπτικής κυβερνοασφάλειας μέσω ΤΝ, η ανίχνευση και απόκριση σε απειλές, καθώς και συγκεκριμένες περιπτωσιολογικές μελέτες που αποδεικνύουν τη σημαντικότητα αυτών των εφαρμογών.

5.1 Προληπτική Κυβερνοασφάλεια μέσω Τεχνητής Νοημοσύνης

Σε αυτή την ενότητα θα παρουσιαστούν οι βασικές αρχές της προληπτικής κυβερνοασφάλειας μέσω της Τεχνητής Νοημοσύνης. Η χρήση των τεχνολογιών ΤΝ για την ανίχνευση και πρόβλεψη απειλών πριν αυτές εκδηλωθούν πλήρως, αποτελεί μια από τις πιο σημαντικές και καινοτόμες προσεγγίσεις για την προστασία κρίσιμων συστημάτων από κακόβουλες ενέργειες. Οι αλγόριθμοι ΤΝ που βασίζονται σε μηχανική μάθηση και ανάλυση δεδομένων, μπορούν να εντοπίσουν ανωμαλίες σε πραγματικό χρόνο, προσφέροντας μια εξαιρετικά αποτελεσματική προληπτική στρατηγική.

5.1.1 Έξυπνη παρακολούθηση απειλών (intelligent threat monitoring)

Η έξυπνη παρακολούθηση απειλών αποτελεί έναν από τους πιο καινοτόμους και αποτελεσματικούς τρόπους προστασίας των συστημάτων Τεχνητής Νοημοσύνης (ΤΝ) στον τομέα της προληπτικής κυβερνοασφάλειας. Μέσω αυτής της προσέγγισης, οι οργανισμοί έχουν τη δυνατότητα να ανιχνεύουν πιθανές απειλές σε πραγματικό χρόνο, εντοπίζοντας κακόβουλες δραστηριότητες πριν αυτές καταφέρουν να προκαλέσουν σοβαρή ζημιά στα συστήματα. Η αξιοποίηση τεχνολογιών ΤΝ, όπως οι αλγόριθμοι μηχανικής μάθησης και ανάλυσης δεδομένων, καθιστά εφικτή την παρακολούθηση μεγάλων ροών δεδομένων και τη διαρκή ανάλυση της συμπεριφοράς των χρηστών και των συσκευών. Το σύστημα παρακολουθεί τη συμπεριφορά του δικτύου και τις αλληλεπιδράσεις με το περιβάλλον του, αναζητώντας οποιαδήποτε ανωμαλία που μπορεί να αποτελεί ένδειξη επικείμενης επίθεσης [9].

Η βασική αρχή της έξυπνης παρακολούθησης απειλών έγκειται στη δυνατότητα των συστημάτων ΤΝ να εκπαιδεύονται με τεράστια σύνολα δεδομένων, τα οποία περιλαμβάνουν τόσο φυσιολογική όσο και ανώμαλη συμπεριφορά. Τα συστήματα αυτά βασίζονται σε τεχνικές *machine learning* για να ανιχνεύσουν πρότυπα ή μοτίβα που θα μπορούσαν να συνδεθούν με κακόβουλες δραστηριότητες. Για παράδειγμα, ένας αυξημένος αριθμός αιτημάτων από έναν συγκεκριμένο

χρήστη ή συσκευή μπορεί να υποδεικνύει την πρόθεση για επιθέσεις τύπου *brute force*, ενώ ασυνήθιστες αλληλεπιδράσεις με την υποδομή του δικτύου μπορεί να υποδεικνύουν επικείμενη επίθεση *phishing*. Αυτές οι ανωμαλίες, τις οποίες αναγνωρίζουν τα μοντέλα ΤΝ, είναι εξαιρετικά δύσκολο να εντοπιστούν με παραδοσιακά συστήματα ασφαλείας, ενώ οι τεχνολογίες ΤΝ επιτρέπουν την πρόβλεψη των επιθέσεων πριν αυτές γίνουν αντιληπτές από τον ανθρώπινο παράγοντα [24].

Η έξυπνη παρακολούθηση απειλών έχει τη δυνατότητα να προσαρμόζεται σε νέες μορφές επιθέσεων. Οι παραδοσιακές τεχνικές βασίζονται σε γνωστές υπογραφές κακόβουλου λογισμικού ή σε προκαθορισμένα πρότυπα απειλών. Αντίθετα, οι αλγόριθμοι ΤΝ μπορούν να μαθαίνουν συνεχώς από τα δεδομένα, αναβαθμίζοντας τα μοντέλα τους για να αναγνωρίζουν και νέες μορφές επιθέσεων, όπως οι επιθέσεις τύπου *zero-day*, οι οποίες εκμεταλλεύονται άγνωστα μέχρι στιγμής κενά ασφαλείας. Οι επιθέσεις αυτές είναι ιδιαίτερα επικίνδυνες καθώς δεν καταγράφονται στις βάσεις δεδομένων των παραδοσιακών τεχνολογιών ασφαλείας και μπορούν να προκαλέσουν σοβαρά προβλήματα σε οργανισμούς που δεν διαθέτουν τα κατάλληλα εργαλεία ανίχνευσης. Τα μοντέλα *deep learning,* που χρησιμοποιούνται στη σύγχρονη ΤΝ, επιτρέπουν την ανίχνευση επιθέσεων ακόμα και όταν αυτές δεν ακολουθούν τα γνωστά πρότυπα συμπεριφοράς, ενισχύοντας σημαντικά την ικανότητα ανίχνευσης και πρόληψης [9].

Ένα από τα σημαντικότερα πλεονεκτήματα της έξυπνης παρακολούθησης απειλών είναι η αυτοματοποίηση των διαδικασιών ανίχνευσης και αντιμετώπισης. Η ΤΝ έχει τη δυνατότητα να αναγνωρίζει απειλές και να τις αποκλείει αυτόματα, μειώνοντας την ανάγκη για άμεση ανθρώπινη παρέμβαση. Αυτό μειώνει σημαντικά τον φόρτο εργασίας των διαχειριστών συστημάτων, επιτρέποντάς τους να επικεντρωθούν σε πιο στρατηγικές εργασίες, ενώ η ΤΝ αναλαμβάνει τη διαχείριση καθημερινών περιστατικών. Παράλληλα, η χρήση ΤΝ βελτιώνει και τον χρόνο απόκρισης σε κρίσιμες καταστάσεις, καθώς οι απειλές αναγνωρίζονται και αποκλείονται αμέσως, χωρίς να χρειάζεται η παρέμβαση διαχειριστών [24]. Αυτό σημαίνει ότι ακόμη και σε περιπτώσεις επιθέσεων μεγάλης κλίμακας, τα συστήματα μπορούν να αντιδράσουν άμεσα, περιορίζοντας τις επιπτώσεις πριν αυτές γίνουν ανεξέλεγκτες.

Ωστόσο, η εφαρμογή της έξυπνης παρακολούθησης απειλών απαιτεί συνεχή επανεκπαίδευση των μοντέλων ΤΝ, καθώς οι απειλές στον κυβερνοχώρο εξελίσσονται συνεχώς. Τα μοντέλα πρέπει να προσαρμόζονται σε νέες τεχνικές και να αναβαθμίζονται τακτικά για να είναι σε θέση να

ανιχνεύσουν τις πιο πρόσφατες μορφές κακόβουλων επιθέσεων. Οι οργανισμοί πρέπει να επενδύσουν τόσο σε τεχνολογίες ΤΝ όσο και σε ειδικούς που θα είναι υπεύθυνοι για τη συνεχή βελτίωση αυτών των συστημάτων. Η διαρκής αναβάθμιση είναι απαραίτητη για να διασφαλιστεί ότι τα συστήματα παραμένουν αποτελεσματικά και ανθεκτικά απέναντι στις νέες απειλές [9].

Συμπερασματικά, η έξυπνη παρακολούθηση απειλών μέσω της ΤΝ αποτελεί μια από τις πιο αποτελεσματικές στρατηγικές προληπτικής κυβερνοασφάλειας, προσφέροντας τη δυνατότητα ανίχνευσης απειλών σε πραγματικό χρόνο, αυτόματης απόκρισης και συνεχιζόμενης προσαρμογής σε νέες μορφές επιθέσεων. Η εφαρμογή αυτών των τεχνολογιών επιτρέπει στους οργανισμούς να προστατεύουν καλύτερα τις υποδομές τους, να μειώνουν τον κίνδυνο επιθέσεων και να ανταποκρίνονται πιο γρήγορα σε κρίσιμα περιστατικά [24].

5.1.2 Προληπτική ανάλυση δεδομένων (proactive data analysis)

Η προληπτική ανάλυση δεδομένων αποτελεί έναν από τους πιο σημαντικούς μηχανισμούς της προληπτικής κυβερνοασφάλειας, ειδικά στην εποχή που τα συστήματα Τεχνητής Νοημοσύνης (ΤΝ) καλούνται να επεξεργαστούν τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο. Αυτή η προσέγγιση επικεντρώνεται στην πρόληψη των επιθέσεων πριν αυτές καν εκδηλωθούν, εντοπίζοντας ασυνήθιστες συμπεριφορές ή ανωμαλίες στα δεδομένα που θα μπορούσαν να είναι ενδείξεις επικείμενων απειλών. Η ικανότητα της ΤΝ να αναλύει συνεχώς ροές δεδομένων επιτρέπει στους οργανισμούς να εντοπίζουν επικίνδυνα γεγονότα και να λαμβάνουν μέτρα αποτροπής προτού αυτά εξελιχθούν σε σοβαρές επιθέσεις [5].

Σε αντίθεση με τις παραδοσιακές μεθόδους κυβερνοασφάλειας, οι οποίες λειτουργούν αντιδραστικά αφού έχει ήδη εντοπιστεί μια επίθεση, η προληπτική ανάλυση δεδομένων εστιάζει στην πρόληψη μέσω του εντοπισμού ανωμαλιών προτού αυτές εξελιχθούν σε πραγματικές επιθέσεις. Αυτός ο προληπτικός χαρακτήρας βασίζεται σε αλγορίθμους μηχανικής μάθησης που είναι σε θέση να αναλύουν μεγάλα σύνολα δεδομένων από διαφορετικές πηγές και να εντοπίζουν μοτίβα συμπεριφοράς που δεν συμβαδίζουν με την κανονική ροή λειτουργίας του συστήματος. Για παράδειγμα, η ανάλυση δεδομένων μπορεί να εντοπίσει περιπτώσεις όπου κάποιος αποκτά παράνομη πρόσβαση ή προσπαθεί να κλέψει ευαίσθητα δεδομένα μέσω τεχνικών όπως το *phishing* ή οι επιθέσεις *man-in-the-middle* [25].

Μία από τις σημαντικότερες εφαρμογές της προληπτικής ανάλυσης δεδομένων είναι η αντιμετώπιση των επιθέσεων τύπου *zero-day*, δηλαδή επιθέσεων που εκμεταλλεύονται άγνωστα

κενά ασφαλείας και δεν μπορούν να εντοπιστούν από παραδοσιακά συστήματα ασφαλείας. Αυτές οι επιθέσεις, που παραμένουν αόρατες για μεγάλο διάστημα, είναι ιδιαίτερα επικίνδυνες, καθώς μπορεί να προκαλέσουν σοβαρές βλάβες στα συστήματα πριν εντοπιστούν. Η ΤΝ, μέσω προληπτικής ανάλυσης δεδομένων, μπορεί να εντοπίσει ασυνήθιστη δραστηριότητα στα δεδομένα και να ειδοποιήσει τους διαχειριστές συστημάτων για ενδεχόμενη απειλή προτού αυτή εκδηλωθεί πλήρως. Αυτή η ικανότητα αναδεικνύει την ΤΝ ως βασικό παράγοντα πρόληψης στην κυβερνοασφάλεια, βελτιώνοντας σημαντικά την ταχύτητα απόκρισης και μειώνοντας τον χρόνο που απαιτείται για την ανίχνευση μιας επίθεσης [5].

Μια δημοφιλής προσέγγιση στην προληπτική ανάλυση δεδομένων είναι η χρήση τεχνικών *predictive analytics*, όπου τα συστήματα ΤΝ εκπαιδεύονται σε ιστορικά δεδομένα και αναλύουν μοτίβα συμπεριφοράς που προηγούνται επιθέσεων. Αυτά τα μοντέλα μπορούν να παράγουν προβλέψεις σχετικά με το πώς μπορεί να εξελιχθούν μελλοντικές επιθέσεις και να ειδοποιήσουν τους διαχειριστές προτού οι απειλές προκαλέσουν σοβαρά προβλήματα. Για παράδειγμα, σε ένα περιβάλλον δικτύου, τα δεδομένα που συλλέγονται από τις ροές κυκλοφορίας και τις αλληλεπιδράσεις των χρηστών μπορούν να αναλυθούν σε πραγματικό χρόνο, προκειμένου να εντοπιστούν ασυνήθιστα μοτίβα που θα μπορούσαν να υποδηλώνουν επικείμενη επίθεση *phishing* ή *DDoS* [25].

Ένα από τα κύρια πλεονεκτήματα της προληπτικής ανάλυσης δεδομένων είναι η δυνατότητα εφαρμογής προληπτικών μέτρων ασφαλείας χωρίς την ανάγκη για άμεση ανθρώπινη παρέμβαση. Τα συστήματα ΤΝ μπορούν να αυτοματοποιήσουν τις διαδικασίες ανίχνευσης και αντιμετώπισης απειλών, επιτρέποντας στους οργανισμούς να εντοπίζουν και να περιορίζουν τις επιθέσεις στο πρώιμο στάδιο ανάπτυξής τους. Αυτή η αυτοματοποίηση δεν βελτιώνει μόνο την ταχύτητα απόκρισης, αλλά μειώνει και την εξάρτηση από τον ανθρώπινο παράγοντα, επιτρέποντας στους διαχειριστές συστημάτων να επικεντρώνονται σε πιο στρατηγικές λειτουργίες. Με τη βοήθεια των αυτοματοποιημένων συστημάτων, οι οργανισμοί μπορούν να εφαρμόσουν μέτρα αντιμετώπισης όπως το μπλοκάρισμα της κυκλοφορίας ή την απομόνωση ευάλωτων υποσυστημάτων, προτού οι επιθέσεις αποκτήσουν πρόσβαση σε κρίσιμα δεδομένα ή λειτουργίες [5].

Επιπλέον, η προληπτική ανάλυση δεδομένων προσφέρει στους οργανισμούς τη δυνατότητα λήψης πιο στοχευμένων και ακριβών αποφάσεων σχετικά με τα μέτρα ασφάλειας που πρέπει να ληφθούν. Η ανάλυση των ροών δεδομένων σε πραγματικό χρόνο επιτρέπει στους διαχειριστές να

κατανοήσουν καλύτερα τις αδυναμίες του συστήματος και να εφαρμόσουν τα κατάλληλα μέτρα, βελτιώνοντας συνολικά την ασφάλεια και μειώνοντας τις πιθανότητες ζημιών ή οικονομικών απωλειών. Οι οργανισμοί που χρησιμοποιούν προληπτική ανάλυση δεδομένων έχουν μεγαλύτερη ικανότητα να προσαρμόζονται στις συνεχώς μεταβαλλόμενες συνθήκες του κυβερνοχώρου και να διασφαλίζουν τη μακροχρόνια ασφάλεια των υποδομών τους [25].

5.2 Ανίχνευση και Αντίδραση σε Κυβερνοεπιθέσεις με Τεχνητή Νοημοσύνη

Σε αυτό το υποκεφάλαιο θα αναλύσουμε την εφαρμογή της Τεχνητής Νοημοσύνης (ΤΝ) στην ανίχνευση και αντίδραση σε κυβερνοεπιθέσεις. Συγκεκριμένα, θα εξετάσουμε δύο βασικές προσεγγίσεις που έχουν υιοθετηθεί για την ενίσχυση της κυβερνοασφάλειας: η ανάλυση συμπεριφοράς (behavior analysis) και η αυτόματη απόκριση σε περιστατικά (automated incident response). Η ανάλυση συμπεριφοράς εστιάζει στην αναγνώριση ανωμαλιών και ασυνήθιστων δραστηριοτήτων που θα μπορούσαν να υποδηλώνουν κακόβουλες ενέργειες, ενώ η αυτόματη απόκριση επιτρέπει στα συστήματα να αντιδρούν άμεσα στις επιθέσεις, χωρίς την ανάγκη ανθρώπινης παρέμβασης. Αυτές οι τεχνολογίες ενισχύουν την ικανότητα των οργανισμών να προλαμβάνουν, να ανιχνεύουν και να αντιμετωπίζουν αποτελεσματικά τις σύγχρονες απειλές στον κυβερνοχώρο, μειώνοντας ταυτόχρονα τον χρόνο απόκρισης και τον φόρτο εργασίας των διαχειριστών συστημάτων.

5.2.1 Ανάλυση συμπεριφοράς (behavior analysis)

Η ανάλυση συμπεριφοράς (behavior analysis) αποτελεί έναν από τους θεμελιώδεις πυλώνες της σύγχρονης κυβερνοασφάλειας και συνδυάζει τις δυνατότητες της Τεχνητής Νοημοσύνης (ΤΝ) με προηγμένες τεχνικές μηχανικής μάθησης για τον εντοπισμό και την πρόληψη κυβερνοεπιθέσεων. Η ανάλυση συμπεριφοράς επικεντρώνεται στη συνεχή παρακολούθηση των δραστηριοτήτων των χρηστών, των συσκευών και των δικτύων, με στόχο την αναγνώριση ανωμαλιών και ασυνήθιστων συμπεριφορών που ενδέχεται να υποδηλώνουν κακόβουλες ενέργειες. Αυτή η τεχνολογία έχει γίνει αναπόσπαστο μέρος των συστημάτων ανίχνευσης απειλών (Intrusion Detection Systems - IDS), καθώς προσφέρει ένα δυναμικό και προσαρμοστικό τρόπο εντοπισμού επιθέσεων, ακόμη και όταν αυτές αποκλίνουν από τα γνωστά πρότυπα απειλών [7].

Στην καρδιά της ανάλυσης συμπεριφοράς βρίσκεται η δυνατότητα των συστημάτων ΤΝ να μαθαίνουν από τις αλληλεπιδράσεις των χρηστών και να εντοπίζουν ασυνήθιστες συμπεριφορές που μπορεί να υποδηλώνουν επικείμενες επιθέσεις. Αυτό επιτυγχάνεται μέσω της συλλογής και

ανάλυσης μεγάλων ποσοτήτων δεδομένων σε πραγματικό χρόνο, όπως τα αιτήματα πρόσβασης, οι συνδέσεις δικτύου, η χρήση εφαρμογών και οι αλληλεπιδράσεις με τους πόρους του συστήματος. Για παράδειγμα, εάν ένας χρήστης που συνδέεται συνήθως από μια συγκεκριμένη γεωγραφική περιοχή ξαφνικά επιχειρήσει να συνδεθεί από μια διαφορετική τοποθεσία ή από μια συσκευή που δεν είχε χρησιμοποιήσει ποτέ στο παρελθόν, το σύστημα μπορεί να αναγνωρίσει αυτή τη δραστηριότητα ως ύποπτη και να τη σηματοδοτήσει για περαιτέρω διερεύνηση [8].

Ένα από τα μεγάλα πλεονεκτήματα της ανάλυσης συμπεριφοράς είναι η ικανότητα να ανιχνεύει και να αποτρέπει επιθέσεις που βασίζονται σε σύνθετες στρατηγικές, όπως οι επιθέσεις *phishing* ή *brute force*. Οι παραδοσιακές μέθοδοι ασφαλείας βασίζονται σε προκαθορισμένα πρότυπα ή υπογραφές κακόβουλου λογισμικού για να αναγνωρίζουν απειλές, κάτι που δεν είναι πάντα αποτελεσματικό στις σύγχρονες επιθέσεις που εξελίσσονται με ταχύ ρυθμό. Αντιθέτως, τα συστήματα ανάλυσης συμπεριφοράς βασίζονται στη δυναμική εκμάθηση και προσαρμογή, εντοπίζοντας οποιαδήποτε απόκλιση από την κανονική συμπεριφορά του χρήστη ή του συστήματος. Για παράδειγμα, μια ξαφνική αύξηση των αιτημάτων σύνδεσης από τον ίδιο χρήστη μπορεί να υποδηλώνει μια επίθεση τύπου *brute force*, ενώ η μη φυσιολογική αλληλεπίδραση με εφαρμογές μπορεί να σηματοδοτεί μια προσπάθεια παραβίασης ασφαλείας [7].

Μια σημαντική διάσταση της ανάλυσης συμπεριφοράς αφορά τη χρήση μη επιβλεπόμενης μάθησης (*unsupervised learning*), όπου τα μοντέλα μηχανικής μάθησης μπορούν να εντοπίζουν ανωμαλίες χωρίς την ανάγκη για προκαθορισμένα πρότυπα ή επισημασμένα δεδομένα. Αυτά τα μοντέλα βασίζονται στην αναγνώριση μοτίβων που δεν έχουν παρατηρηθεί στο παρελθόν και τα οποία μπορεί να υποδηλώνουν νέες μορφές απειλών. Αυτός ο τύπος ανάλυσης είναι ιδιαίτερα αποτελεσματικός για την αντιμετώπιση επιθέσεων τύπου *zero-day*, όπου οι επιτιθέμενοι εκμεταλλεύονται άγνωστα κενά ασφαλείας και δεν έχουν αφήσει ίχνη στις βάσεις δεδομένων κακόβουλου λογισμικού. Τα μοντέλα αυτά μπορούν να προσαρμόζονται αυτόματα σε νέα δεδομένα, βελτιώνοντας συνεχώς την ικανότητα ανίχνευσης [8].

Η ανάλυση συμπεριφοράς χρησιμοποιείται συχνά σε συνδυασμό με άλλες τεχνικές κυβερνοασφάλειας, όπως η ανάλυση δικτύου και η παρακολούθηση κυκλοφορίας, για την ενίσχυση της συνολικής ασφάλειας του συστήματος. Τα *behavior-based Intrusion Detection Systems (IDS)* επιτρέπουν στα συστήματα να αντιδρούν άμεσα σε ασυνήθιστες δραστηριότητες, ενώ η χρήση τεχνικών *ensemble learning* ενισχύει την αξιοπιστία των ανιχνεύσεων. Αυτές οι

τεχνικές συνδυάζουν πολλά μοντέλα μάθησης για να μειώσουν τα ψευδώς θετικά ή ψευδώς αρνητικά αποτελέσματα, βελτιώνοντας την ακρίβεια της ανίχνευσης και επιτρέποντας στα συστήματα να διαχειρίζονται αποτελεσματικά τις απειλές [7].

Παρά τα σημαντικά οφέλη που προσφέρει η ανάλυση συμπεριφοράς, υπάρχουν και προκλήσεις που πρέπει να αντιμετωπιστούν. Η διαχείριση μεγάλων ποσοτήτων δεδομένων σε πραγματικό χρόνο μπορεί να απαιτεί σημαντικούς πόρους υπολογιστικής ισχύος και αποθήκευσης, ενώ η σωστή ρύθμιση των αλγορίθμων είναι απαραίτητη για την αποφυγή ψευδών ανιχνεύσεων. Επιπλέον, η συνεχής εξέλιξη των κυβερνοεπιθέσεων απαιτεί από τα συστήματα ανάλυσης συμπεριφοράς να αναβαθμίζονται διαρκώς για να αντιμετωπίζουν τις νέες απειλές που προκύπτουν [8].

Σε γενικές γραμμές, η ανάλυση συμπεριφοράς μέσω ΤΝ προσφέρει στους οργανισμούς ένα ισχυρό εργαλείο για την ανίχνευση και πρόληψη κυβερνοεπιθέσεων, βελτιώνοντας σημαντικά την ασφάλεια των συστημάτων. Με την ικανότητα προσαρμογής σε νέες απειλές και την αυτοματοποιημένη ανάλυση των δεδομένων, η ανάλυση συμπεριφοράς επιτρέπει στους οργανισμούς να ανταποκρίνονται πιο γρήγορα και αποτελεσματικά στις προκλήσεις του σύγχρονου κυβερνοχώρου, διασφαλίζοντας τη συνεχή λειτουργία των συστημάτων τους [7], [8].

5.2.2 Αυτόματη απόκριση σε περιστατικά (automated incident response)

Η αυτόματη απόκριση σε περιστατικά (automated incident response) είναι μία από τις πιο καινοτόμες και χρήσιμες εφαρμογές της Τεχνητής Νοημοσύνης (ΤΝ) στον τομέα της κυβερνοασφάλειας. Επιτρέπει στα συστήματα να αντιδρούν σε περιστατικά κυβερνοεπιθέσεων χωρίς την άμεση ανθρώπινη παρέμβαση, βελτιώνοντας δραματικά τον χρόνο απόκρισης και περιορίζοντας την εξάπλωση των επιθέσεων. Αυτός ο αυτοματισμός είναι ζωτικής σημασίας για οργανισμούς που διαχειρίζονται μεγάλα και περίπλοκα συστήματα, καθώς μειώνει τον φόρτο εργασίας των διαχειριστών και αποτρέπει μεγαλύτερες επιπτώσεις όταν τα περιστατικά κυβερνοεπιθέσεων δεν μπορούν να αντιμετωπιστούν άμεσα με χειροκίνητες μεθόδους [12].

Ένα από τα βασικά χαρακτηριστικά της αυτόματης απόκρισης είναι η δυνατότητα να ενεργοποιείται αμέσως μόλις ανιχνευθεί μια ύποπτη δραστηριότητα. Οι αλγόριθμοι ΤΝ, οι οποίοι εκπαιδεύονται σε προηγούμενα περιστατικά και βασίζονται σε μοτίβα κακόβουλης δραστηριότητας, έχουν τη δυνατότητα να λαμβάνουν αποφάσεις αυτόνομα. Όταν, για παράδειγμα, ανιχνευθεί μια ανωμαλία στην κυκλοφορία του δικτύου ή στην αλληλεπίδραση με κρίσιμες

λειτουργίες, το σύστημα μπορεί να περιορίσει αυτόματα την πρόσβαση του χρήστη, να απομονώσει το δίκτυο ή να τερματίσει ύποπτες διεργασίες. Αυτός ο τύπος απόκρισης επιτρέπει στα συστήματα να αντιδρούν ακαριαία σε επιθέσεις όπως το *ransomware*, οι επιθέσεις *DDoS* ή οι προσπάθειες παραβίασης ευαίσθητων δεδομένων, διασφαλίζοντας ότι οι κυβερνοεπιθέσεις δεν θα διαδοθούν και δεν θα επηρεάσουν την συνολική λειτουργικότητα του συστήματος [18].

Ένα βασικό πλεονέκτημα της αυτόματης απόκρισης είναι η ικανότητα των συστημάτων να προσαρμόζονται στις νέες μορφές απειλών. Οι παραδοσιακές μέθοδοι αντιμετώπισης επιθέσεων βασίζονται σε προκαθορισμένα πρότυπα, τα οποία όμως συχνά δεν επαρκούν για τις σύγχρονες, πιο εξελιγμένες επιθέσεις. Αντιθέτως, η ΤΝ έχει τη δυνατότητα να μαθαίνει από τα δεδομένα σε πραγματικό χρόνο και να λαμβάνει έξυπνες αποφάσεις σχετικά με τις ενέργειες που πρέπει να γίνουν. Αυτή η ευελιξία επιτρέπει στα συστήματα να προσαρμόζονται στις νέες μορφές επιθέσεων, ακόμα και αν αυτές δεν έχουν καταγραφεί στο παρελθόν. Για παράδειγμα, όταν ανιχνευθούν νέες τακτικές επίθεσης που δεν έχουν αντιμετωπιστεί στο παρελθόν, το σύστημα μπορεί να αναλύσει τα δεδομένα και να προσαρμόσει τα πρωτόκολλα απόκρισης, εξασφαλίζοντας έτσι ότι παραμένει προστατευμένο ακόμα και σε απρόβλεπτες καταστάσεις [12].

Η αυτόματη απόκριση σε περιστατικά μπορεί επίσης να προγραμματιστεί για να αναλύει δεδομένα από προηγούμενα περιστατικά και να προλαμβάνει μελλοντικές επιθέσεις. Μέσα από την ανάλυση μεγάλων ποσοτήτων δεδομένων που προκύπτουν από τις κυβερνοεπιθέσεις, οι αλγόριθμοι ΤΝ μπορούν να ανιχνεύουν επαναλαμβανόμενα μοτίβα και να προσαρμόζουν τα μοντέλα ασφαλείας για να αντιμετωπίζουν καλύτερα τις νέες απειλές. Έτσι, βελτιώνεται όχι μόνο η άμεση απόκριση στα περιστατικά, αλλά και η προληπτική ικανότητα του συστήματος να αποτρέψει παρόμοιες επιθέσεις στο μέλλον [18].

Ωστόσο, για να παραμείνουν αποτελεσματικά αυτά τα συστήματα αυτόματης απόκρισης, απαιτείται συνεχής παρακολούθηση και επανεκπαίδευση των αλγορίθμων τους. Καθώς οι επιθέσεις στον κυβερνοχώρο γίνονται πιο περίπλοκες και εξελίσσονται, τα μοντέλα ΤΝ πρέπει να ενημερώνονται τακτικά με νέα δεδομένα και να προσαρμόζονται στις νέες συνθήκες απειλών. Εάν δεν γίνεται σωστή ρύθμιση και εκπαίδευση, τα συστήματα ενδέχεται να παράγουν ψευδώς θετικά αποτελέσματα, δηλαδή να αντιδρούν σε περιστατικά που δεν αποτελούν πραγματική απειλή. Αυτό μπορεί να οδηγήσει σε περιττές διακοπές λειτουργιών ή σε υπερβολική επιβάρυνση του δικτύου.

Γι’ αυτόν τον λόγο, η συνεχής αξιολόγηση των αυτοματοποιημένων συστημάτων είναι κρίσιμη για τη διασφάλιση της ακρίβειας και της αποδοτικότητάς τους [12].

Τελικά, η αυτοματοποίηση της απόκρισης σε περιστατικά μέσω της ΤΝ προσφέρει σημαντικά πλεονεκτήματα στους οργανισμούς που αντιμετωπίζουν μεγάλους όγκους δεδομένων και αντιμετωπίζουν συνεχείς απειλές στον κυβερνοχώρο. Η δυνατότητα των συστημάτων να λαμβάνουν αποφάσεις σε πραγματικό χρόνο, χωρίς την ανάγκη για ανθρώπινη παρέμβαση, αυξάνει σημαντικά την ταχύτητα και την ακρίβεια της αντίδρασης. Αυτή η τεχνολογία επιτρέπει στα συστήματα να παραμένουν προστατευμένα ακόμα και όταν οι επιθέσεις εξελίσσονται ταχύτατα, εξασφαλίζοντας την αδιάλειπτη λειτουργία τους και την αποτροπή σοβαρών επιπτώσεων [18].

5.3 Περιπτωσιολογικές Μελέτες (Case Studies)

Η ανάλυση περιπτωσιολογικών μελετών (case studies) αποτελεί μια ιδιαίτερα σημαντική μέθοδο για την αξιολόγηση της εφαρμογής της Τεχνητής Νοημοσύνης (ΤΝ) στον τομέα της κυβερνοασφάλειας. Μέσα από την εμπειρική ανάλυση επιτυχημένων εφαρμογών, μπορούμε να κατανοήσουμε καλύτερα πώς οι τεχνολογίες ΤΝ μπορούν να ενισχύσουν τα συστήματα ασφάλειας, να αποτρέψουν κυβερνοεπιθέσεις και να διασφαλίσουν την προστασία ευαίσθητων δεδομένων. Οι περιπτωσιολογικές μελέτες προσφέρουν επίσης πολύτιμα μαθήματα για το πώς μπορούν να αντιμετωπιστούν οι προκλήσεις και να βελτιωθούν οι στρατηγικές ασφάλειας μέσω της εφαρμογής ΤΝ [6].

Μια από τις πλέον χαρακτηριστικές περιπτώσεις είναι η χρήση της ΤΝ για την ανίχνευση και αντιμετώπιση κακόβουλου λογισμικού (malware). Σε πραγματικά παραδείγματα εφαρμογών, τα συστήματα ΤΝ έχουν καταφέρει να εντοπίσουν επιθέσεις κακόβουλου λογισμικού που δεν θα μπορούσαν να ανιχνευτούν με τις παραδοσιακές μεθόδους. Για παράδειγμα, εταιρείες ασφαλείας όπως η Symantec και η McAfee έχουν αναπτύξει λύσεις που βασίζονται στη μηχανική μάθηση για την ανάλυση μεγάλων ποσοτήτων δεδομένων και την ανίχνευση άγνωστων τύπων malware μέσω ανάλυσης συμπεριφοράς των αρχείων και των εφαρμογών. Αυτές οι τεχνικές επιτρέπουν στα συστήματα να εντοπίζουν απειλές που δεν βασίζονται σε γνωστές υπογραφές κακόβουλου λογισμικού, αλλά αντίθετα αναλύουν τις αλληλεπιδράσεις τους με το σύστημα για να αναγνωρίσουν ύποπτη συμπεριφορά [6].

Μια ακόμη επιτυχημένη εφαρμογή της ΤΝ στον τομέα της κυβερνοασφάλειας είναι η αντιμετώπιση επιθέσεων *DDoS* (Distributed Denial of Service). Σε περιπτώσεις όπως η προστασία ιστοσελίδων και διαδικτυακών υπηρεσιών από τέτοιες επιθέσεις, η ΤΝ έχει χρησιμοποιηθεί για την ανάλυση των ροών δικτύου σε πραγματικό χρόνο και τον εντοπισμό της αυξημένης κυκλοφορίας που συνδέεται με επιθέσεις *DDoS*. Εταιρείες όπως η Cloudflare και η Akamai έχουν αναπτύξει λύσεις που χρησιμοποιούν αλγορίθμους μηχανικής μάθησης για την ταξινόμηση της δικτυακής κίνησης και τον αποκλεισμό ύποπτης δραστηριότητας, πριν αυτή προλάβει να επηρεάσει την κανονική λειτουργία του συστήματος. Αυτή η δυνατότητα επιτρέπει την αποτροπή των επιθέσεων σε πρώιμο στάδιο, διασφαλίζοντας ότι οι επιχειρήσεις συνεχίζουν να λειτουργούν χωρίς διακοπές [23].

Ένα άλλο σημαντικό πεδίο εφαρμογής της ΤΝ είναι η προστασία από επιθέσεις τύπου *phishing*. Μέσω τεχνικών μηχανικής μάθησης, τα συστήματα μπορούν να αναλύσουν τα χαρακτηριστικά των email ή των ιστοσελίδων και να εντοπίσουν ψεύτικα μηνύματα που στοχεύουν στη συλλογή προσωπικών δεδομένων από τους χρήστες. Για παράδειγμα, η Google έχει αναπτύξει λύσεις βασισμένες στη μηχανική μάθηση που επιτρέπουν την ανίχνευση επιθέσεων *phishing* μέσω της ανάλυσης του περιεχομένου των μηνυμάτων και της αναγνώρισης ύποπτων συνδέσμων ή αρχείων. Αυτές οι λύσεις έχουν αποδειχθεί εξαιρετικά αποτελεσματικές στην αποτροπή των επιθέσεων, καθώς επιτρέπουν στα συστήματα να αναγνωρίζουν και να αποκλείουν κακόβουλα μηνύματα προτού φτάσουν στους χρήστες [6].

Μια άλλη επιτυχημένη περιπτωσιολογική μελέτη αφορά τη χρήση της ΤΝ στην προστασία κρίσιμων υποδομών, όπως οι τράπεζες και τα χρηματοοικονομικά συστήματα. Ειδικά στις περιπτώσεις αυτών των οργανισμών, η χρήση της ΤΝ επιτρέπει την ανίχνευση ασυνήθιστων συναλλαγών που μπορεί να υποδηλώνουν προσπάθεια για οικονομική απάτη. Οι τεχνικές ανάλυσης συμπεριφοράς, όπως αυτές που αναπτύχθηκαν από εταιρείες όπως η Darktrace, επιτρέπουν την ανίχνευση ύποπτων μοτίβων συναλλαγών ή αλληλεπιδράσεων με τα χρηματοοικονομικά συστήματα, ενισχύοντας έτσι την ασφάλεια των συστημάτων αυτών [23].

Παρά τις επιτυχίες αυτές, οι περιπτωσιολογικές μελέτες αποκαλύπτουν επίσης προκλήσεις και περιορισμούς στη χρήση της ΤΝ στην κυβερνοασφάλεια. Ορισμένες επιθέσεις είναι τόσο εξελιγμένες που ακόμα και τα πιο προηγμένα συστήματα ΤΝ δυσκολεύονται να τις αναγνωρίσουν ή να ανταποκριθούν άμεσα. Για παράδειγμα, οι επιθέσεις *adversarial*, όπου οι επιτιθέμενοι

εισάγουν εσκεμμένα μικρές τροποποιήσεις στα δεδομένα εισόδου για να παραπλανήσουν τα μοντέλα ΤΝ, έχουν αποδειχθεί ιδιαίτερα αποτελεσματικές στο να παρακάμπτουν τις υπάρχουσες άμυνες. Οι μελέτες αυτές τονίζουν την ανάγκη για συνεχή βελτίωση των συστημάτων ΤΝ και την προσαρμογή τους στις νέες μορφές επιθέσεων που αναδύονται συνεχώς [6].

Συνοψίζοντας, οι περιπτωσιολογικές μελέτες δείχνουν πώς η ΤΝ μπορεί να προσφέρει σημαντικά πλεονεκτήματα στην κυβερνοασφάλεια, από την ανίχνευση κακόβουλου λογισμικού μέχρι την προστασία από επιθέσεις *DDoS* και *phishing*. Τα επιτυχημένα παραδείγματα εφαρμογής αυτών των τεχνολογιών προσφέρουν πολύτιμα μαθήματα για το πώς μπορούν να βελτιωθούν περαιτέρω οι στρατηγικές ασφάλειας, ενώ οι προκλήσεις που αναδεικνύονται τονίζουν τη σημασία της συνεχούς βελτίωσης και εξέλιξης των συστημάτων αυτών [23].

5.3.1 IBM Watson: Ένα επιτυχημένο παράδειγμα Τεχνητής Νοημοσύνης στην Κυβερνοασφάλεια

Η IBM ανέπτυξε μία από τις πιο ισχυρές εφαρμογές Τεχνητής Νοημοσύνης (ΤΝ) για την ενίσχυση της κυβερνοασφάλειας, χρησιμοποιώντας την πλατφόρμα Watson. Το Watson συνδυάζει τεχνικές μηχανικής μάθησης και επεξεργασίας φυσικής γλώσσας για την ανάλυση μεγάλων όγκων δεδομένων σε πραγματικό χρόνο. Μέσω της λύσης *QRadar Advisor with Watson*, η IBM έχει επιτύχει σημαντική βελτίωση στην ταχύτητα και την ακρίβεια ανίχνευσης και απόκρισης σε κυβερνοαπειλές, μετατρέποντας τα δεδομένα ασφαλείας σε χρήσιμη πληροφορία με μεγάλη ταχύτητα.

Ένα εξαιρετικό παράδειγμα επιτυχημένης εφαρμογής του Watson είναι η περίπτωση της Cargills Bank στη Σρι Λάνκα. Η τράπεζα χρησιμοποίησε την πλατφόρμα *QRadar Advisor with Watson* για να αντιμετωπίσει τις αυξανόμενες κυβερνοαπειλές. Με την εφαρμογή αυτής της τεχνολογίας, η Cargills Bank μείωσε τον χρόνο που απαιτείται για την ανίχνευση και ανάλυση περιστατικών από ώρες σε λίγα λεπτά, βελτιώνοντας δραστικά την απόκριση της ομάδας ασφαλείας. Το Watson ανέλυσε δεδομένα από πολλές πηγές, συμπεριλαμβανομένων ιστολογίων και επιστημονικών άρθρων, και προσέφερε στους αναλυτές πληροφόρηση που υπό άλλες συνθήκες θα απαιτούσε ώρες αναζήτησης και συσχέτισης [26].

Επιπλέον, η IBM μέσω του Watson κατάφερε να μειώσει τα ψευδώς θετικά κατά 30%, επιτρέποντας στις ομάδες ασφαλείας να εστιάσουν περισσότερο στις πραγματικές απειλές, αυξάνοντας την αποτελεσματικότητα και την ακρίβεια των ενεργειών τους [26]. Η τεχνολογία

αυτή απέδειξε ότι μπορεί να προσαρμόζεται γρήγορα σε νέες μορφές επιθέσεων, καθώς η συνεχής εκπαίδευση του Watson σε νέες απειλές το καθιστά πιο ευέλικτο και προδραστικό στη διαχείριση κυβερνοεπιθέσεων [25].

Ένα σημαντικό στοιχείο της επιτυχίας του Watson είναι η δυνατότητά του να ενσωματώνεται στα υπάρχοντα συστήματα ασφαλείας μιας εταιρείας, διασφαλίζοντας τη συνεχή και αποτελεσματική προστασία των δεδομένων. Το Watson είναι σε θέση να ελέγχει και να επεξεργάζεται χιλιάδες σήματα από διάφορες συσκευές και εφαρμογές σε πραγματικό χρόνο, προσαρμόζοντας τις στρατηγικές ασφαλείας ανάλογα με την απειλή [25]. Αυτό το σύστημα έχει επιτρέψει σε οργανισμούς όπως η Cargills Bank να αυξήσουν σημαντικά την ασφάλεια των διαδικτυακών τους συναλλαγών και να μειώσουν τον κίνδυνο παραβιάσεων.

5.3.2 Microsoft και η Ασφάλεια στο Cloud

Η Microsoft έχει εφαρμόσει προηγμένες λύσεις Τεχνητής Νοημοσύνης (ΤΝ) για την ενίσχυση της κυβερνοασφάλειας στα περιβάλλοντα cloud, με έμφαση στην άμεση ανίχνευση απειλών και την προληπτική άμυνα. Μέσω του **Microsoft Intelligent Security Graph**, ένα σύστημα που επεξεργάζεται πάνω από 6,5 τρισεκατομμύρια σήματα κάθε μέρα, η Microsoft καταφέρνει να παρέχει διαρκή προστασία σε όλες τις cloud υπηρεσίες της. Αυτό το σύστημα χρησιμοποιεί αλγόριθμους μηχανικής μάθησης και μεγάλα δεδομένα για την αναγνώριση ανωμαλιών και πιθανών απειλών προτού αυτές μπορέσουν να προκαλέσουν σοβαρές βλάβες [27].

Ένα βασικό επίτευγμα του συστήματος αυτού ήταν η μείωση του χρόνου ανίχνευσης από 24 ώρες σε λιγότερο από μία ώρα, κάτι που οδήγησε στην αύξηση της αποτελεσματικότητας ανίχνευσης απειλών κατά 40%. Αυτή η ταχύτητα είναι κρίσιμη για την αντιμετώπιση σύνθετων επιθέσεων, όπως οι επιθέσεις phishing και κακόβουλου λογισμικού (malware), που μπορούν να πλήξουν σοβαρά εταιρείες και οργανισμούς που βασίζονται σε υπηρεσίες cloud για την αποθήκευση και επεξεργασία των δεδομένων τους [27].

Στο πλαίσιο της χρήσης της ΤΝ, η Microsoft έχει ενσωματώσει την **Microsoft Defender for Cloud**, ένα σύστημα που προσφέρει πλήρη προστασία από κυβερνοαπειλές σε περιβάλλοντα cloud. Αυτό το εργαλείο βοηθά στη συνεχή παρακολούθηση της ασφάλειας των εφαρμογών και των δικτύων, παρέχοντας άμεσες ειδοποιήσεις σε περίπτωση που εντοπιστούν πιθανές απειλές. Επιπλέον, η ΤΝ ενισχύει την ικανότητα πρόβλεψης επιθέσεων, επιτρέποντας στις εταιρείες να αναγνωρίζουν τάσεις και μοτίβα που υποδεικνύουν επικείμενες επιθέσεις [27].

Ένα χαρακτηριστικό παράδειγμα αποτελεσματικής εφαρμογής των λύσεων της Microsoft είναι η χρήση της **Microsoft Defender for Endpoint**, που κατάφερε να εντοπίσει και να αποτρέψει επιθέσεις brute force σε περιβάλλον cloud. Μέσα από την ανάλυση των δικτύων σε πραγματικό χρόνο, η ΤΝ ήταν σε θέση να εντοπίσει επικίνδυνες ανωμαλίες στη ροή δεδομένων, αποτρέποντας την επέκταση της επίθεσης σε κρίσιμα συστήματα. Παράλληλα, τα **Security Information and Event Management (SIEM)** εργαλεία που χρησιμοποιεί η Microsoft επιτρέπουν τη συνεχή αξιολόγηση και την προσαρμογή των συστημάτων ασφάλειας, καθιστώντας τα πιο ανθεκτικά απέναντι σε νέες μορφές κυβερνοεπιθέσεων [27].

Η εφαρμογή αυτών των τεχνολογιών έχει βοηθήσει σημαντικά στη μείωση των περιστατικών παραβίασης δεδομένων σε πολυνεφικές (multicloud) υποδομές, προσφέροντας ταυτόχρονα καλύτερη κατανόηση των απειλών και πιο αποδοτική απόκριση από τις ομάδες ασφάλειας. Η δυνατότητα ανάλυσης τεράστιων ποσοτήτων δεδομένων και η αυτοματοποιημένη απόκριση σε απειλές καθιστούν το cloud της Microsoft από τα πιο ασφαλή στην αγορά, προστατεύοντας τόσο τις υποδομές της εταιρείας όσο και τους πελάτες της [27].

5.3.3 Darktrace και το Enterprise Immune System

Η Darktrace, μία από τις κορυφαίες εταιρείες στην εφαρμογή της Τεχνητής Νοημοσύνης (ΤΝ) για την αντιμετώπιση κυβερνοεπιθέσεων, έχει αναπτύξει το **Enterprise Immune System**, ένα σύστημα βασισμένο στη μηχανική μάθηση που λειτουργεί όπως το ανθρώπινο ανοσοποιητικό σύστημα, ανιχνεύοντας απειλές και ανωμαλίες σε πραγματικό χρόνο. Αυτή η τεχνολογία δεν απαιτεί προηγούμενη γνώση ή προκαθορισμένα δεδομένα για να ανιχνεύσει μια απειλή, γεγονός που την καθιστά ιδιαίτερα αποτελεσματική σε σύνθετα περιβάλλοντα, όπως η ενεργειακή βιομηχανία [28].

Ένα χαρακτηριστικό παράδειγμα επιτυχούς εφαρμογής του **Enterprise Immune System** ήταν στην εταιρεία **Drax Group**, έναν από τους μεγαλύτερους προμηθευτές ηλεκτρικής ενέργειας στο Ηνωμένο Βασίλειο. Η Drax αντιμετώπισε σημαντικούς κινδύνους, λόγω της συνδεσιμότητας των συστημάτων της και της μετάβασής της στη χρήση βιομάζας, γεγονός που την κατέστησε στόχο επιθέσεων. Το **Enterprise Immune System** ανίχνευσε επιθέσεις που δεν είχαν εντοπιστεί από τα παραδοσιακά συστήματα ασφαλείας, παρέχοντας πλήρη ορατότητα των λειτουργικών και πληροφοριακών συστημάτων της εταιρείας [28].

Η εφαρμογή της τεχνολογίας αυτής προσέφερε στη **Drax** τη δυνατότητα να προσαρμόζεται άμεσα σε νέες απειλές, όπως επιθέσεις zero-day, που δεν είχαν προηγουμένως καταγραφεί. Το σύστημα προσαρμόζει τα μοντέλα του με βάση την «κανονική» συμπεριφορά κάθε συσκευής και χρήστη στο δίκτυο και ανιχνεύει ασυνήθιστες δραστηριότητες ή ανωμαλίες, παρέχοντας άμεση προειδοποίηση και ενεργοποίηση αμυντικών μέτρων. Η χρήση της λύσης αυτής όχι μόνο βελτίωσε τη συνολική ασφάλεια της Drax αλλά και ενίσχυσε την ανθεκτικότητά της απέναντι σε κυβερνοεπιθέσεις, διασφαλίζοντας κρίσιμες λειτουργίες [28].

Η **Darktrace** αξιοποίησε επίσης το **Antigena**, ένα σύστημα αυτοματοποιημένης απόκρισης, που χρησιμοποιείται για τον περιορισμό ή την απομόνωση των επιθέσεων αμέσως μόλις αυτές εντοπιστούν. Αυτό το εργαλείο έχει αποδειχθεί αποτελεσματικό στην αντιμετώπιση επιθέσεων, όπως το **WannaCry**, που επηρέασε παγκόσμιους οργανισμούς το 2017. Η τεχνολογία αυτή έδωσε στη Drax τη δυνατότητα να προλαμβάνει επιθέσεις πριν προκαλέσουν σημαντικές ζημιές, ενισχύοντας την προληπτική της ασφάλεια [28].

5.3.4 PayPal και AI στην Ασφάλεια Συναλλαγών

Η PayPal, ως μία από τις κορυφαίες πλατφόρμες ηλεκτρονικών πληρωμών παγκοσμίως, αξιοποιεί την Τεχνητή Νοημοσύνη (ΤΝ) για να ενισχύσει την ασφάλεια των χρηματοοικονομικών συναλλαγών της. Το σύστημα αυτόματης ανάλυσης της PayPal βασίζεται σε αλγορίθμους μηχανικής μάθησης, οι οποίοι έχουν τη δυνατότητα να αναγνωρίζουν και να αποτρέπουν ύποπτες συναλλαγές σε πραγματικό χρόνο, μειώνοντας σημαντικά την πιθανότητα οικονομικών απωλειών για τους πελάτες της.

Ένα από τα βασικά στοιχεία του συστήματος ασφαλείας της PayPal είναι η παρακολούθηση της συμπεριφοράς των χρηστών και των συναλλαγών τους μέσω των μεγάλων δεδομένων που συλλέγει από τις καθημερινές συναλλαγές. Τα μοτίβα συμπεριφοράς των χρηστών αναλύονται σε βάθος, προκειμένου να ανιχνευτούν ανωμαλίες που μπορεί να υποδεικνύουν πιθανές απάτες. Το σύστημα αξιοποιεί επίσης την ανάλυση δεδομένων σε πραγματικό χρόνο, επιτρέποντας την άμεση αντίδραση σε οποιεσδήποτε ύποπτες δραστηριότητες, μειώνοντας έτσι τον κίνδυνο εξαπάτησης των πελατών [29].

Η PayPal έχει επενδύσει σημαντικά στην ανάπτυξη των δυνατοτήτων μηχανικής μάθησης και έχει επίσης προχωρήσει στην εξαγορά εταιρειών, όπως η Simility, η οποία εστιάζει στην πρόληψη απάτης μέσω της μηχανικής μάθησης [29]. Αυτές οι επενδύσεις της PayPal οδήγησαν στην αύξηση της ακρίβειας της ανίχνευσης απάτης και στη μείωση των ψευδών συναγερμών, βελτιώνοντας την εμπειρία των χρηστών. Το σύστημα της PayPal καταφέρνει να εντοπίζει απάτες με τρόπο που δεν διακόπτει τις νόμιμες συναλλαγές, εξασφαλίζοντας παράλληλα την προστασία των πελατών από κακόβουλες ενέργειες, όπως η κλοπή ταυτότητας και η νομιμοποίηση εσόδων από παράνομες δραστηριότητες [29].

Επιπλέον, η PayPal αξιοποιεί την ανάλυση δεδομένων για να βελτιώσει την πρόληψη απάτης σε μεγαλύτερη κλίμακα, εξετάζοντας γεωγραφικά δεδομένα, μοτίβα συναλλαγών και άλλες παραμέτρους που ενδέχεται να υποδηλώνουν ύποπτη δραστηριότητα. Αυτή η τεχνολογία επιτρέπει στον οργανισμό να είναι μπροστά από τις εξελίξεις, διασφαλίζοντας ότι τα συστήματα του είναι σε θέση να ανιχνεύουν νέες μορφές απάτης καθώς αυτές προκύπτουν, προσφέροντας παράλληλα ασφάλεια και αξιοπιστία στους χρήστες της πλατφόρμας.

6. Μελλοντικές Προοπτικές και Ερευνητικές Προκλήσεις

Σε αυτό το κεφάλαιο, θα αναλυθούν οι μελλοντικές προοπτικές και οι ερευνητικές προκλήσεις στον τομέα της κυβερνοασφάλειας, με ιδιαίτερη έμφαση στις τεχνολογίες που προάγουν την ασφάλεια και στην αντιμετώπιση των σύγχρονων προκλήσεων που προκύπτουν. Εξετάζονται αναδυόμενες τάσεις, όπως η ανάλυση μεγάλων δεδομένων (Big Data) και η ανάπτυξη νέων αλγορίθμων Τεχνητής Νοημοσύνης (ΤΝ), οι οποίες συμβάλλουν σημαντικά στην ενίσχυση των συστημάτων ασφαλείας. Η αυξανόμενη πολυπλοκότητα των επιθέσεων απαιτεί τη χρήση καινοτόμων τεχνικών ανίχνευσης και απόκρισης, καθώς και την ανάπτυξη νέων εργαλείων που επιτρέπουν την προσαρμογή των συστημάτων στις εξελισσόμενες απειλές. Μέσα από αυτήν την ανάλυση, θα αναδειχθούν τα οφέλη των νέων τεχνολογιών, αλλά και οι προκλήσεις που πρέπει να αντιμετωπιστούν, όπως η πολυπλοκότητα των συστημάτων και η διασφάλιση της ιδιωτικότητας.

6.1 Αναδυόμενες Τάσεις και Τεχνολογίες

Η υποενότητα, που ακολουθεί θα επικεντρωθεί στις αναδυόμενες τάσεις και τεχνολογίες που αναμένεται να διαμορφώσουν το μέλλον της κυβερνοασφάλειας. Η συνεχής εξέλιξη των απειλών και η αύξηση της πολυπλοκότητας των συστημάτων καθιστούν αναγκαία την υιοθέτηση νέων τεχνολογιών, όπως η ανάλυση μεγάλων δεδομένων (Big Data) και η ανάπτυξη προηγμένων αλγορίθμων Τεχνητής Νοημοσύνης (ΤΝ). Η ικανότητα επεξεργασίας τεράστιων όγκων δεδομένων σε πραγματικό χρόνο και η αξιοποίηση καινοτόμων αλγορίθμων ενισχύουν τις δυνατότητες ανίχνευσης και πρόληψης απειλών, ενώ παράλληλα δημιουργούν νέες προκλήσεις που πρέπει να αντιμετωπιστούν. Μέσα από την ανάλυση αυτών των τεχνολογιών, θα εξετάσουμε τις δυνατότητες που προσφέρουν για την προστασία των συστημάτων και τα νέα δεδομένα που δημιουργούν για την ασφάλεια στον κυβερνοχώρο.

6.1.1 Ανάλυση Μεγάλων Δεδομένων στην Κυβερνοασφάλεια (Big Data Analysis in Cybersecurity)

Η ανάλυση μεγάλων δεδομένων (Big Data) αποτελεί ένα από τα πιο σημαντικά εργαλεία στη σύγχρονη κυβερνοασφάλεια, καθώς η ικανότητα επεξεργασίας τεράστιων όγκων δεδομένων σε πραγματικό χρόνο επιτρέπει την ανίχνευση και την πρόληψη απειλών με μεγαλύτερη ακρίβεια και ταχύτητα. Στην εποχή του ψηφιακού μετασχηματισμού, η ποσότητα των δεδομένων που παράγεται αυξάνεται με ραγδαίους ρυθμούς, και τα συστήματα ασφαλείας πρέπει να είναι σε θέση

να αναλύουν τα δεδομένα αυτά για να εντοπίσουν τυχόν ανωμαλίες που θα μπορούσαν να υποδηλώνουν κυβερνοεπιθέσεις [3].

Η ανάλυση μεγάλων δεδομένων στην κυβερνοασφάλεια χρησιμοποιεί διάφορες πηγές δεδομένων, όπως κίνηση δικτύου, αλληλεπιδράσεις χρηστών, καταγραφές συμβάντων και δεδομένα από συσκευές IoT. Μέσω αυτών των δεδομένων, τα συστήματα Τεχνητής Νοημοσύνης (ΤΝ) έχουν τη δυνατότητα να αναγνωρίζουν πρότυπα που θα μπορούσαν να υποδεικνύουν επικείμενες επιθέσεις. Για παράδειγμα, η ανίχνευση απότομης αύξησης στην κίνηση δικτύου ή η παρακολούθηση δραστηριοτήτων από άγνωστες διευθύνσεις IP μπορούν να θεωρηθούν δείκτες ύποπτων ενεργειών που σχετίζονται με επιθέσεις τύπου DDoS ή phishing [12].

Η ανάλυση μεγάλων δεδομένων έχει την ικανότητα να παρέχει στους οργανισμούς μια πιο περιεκτική εικόνα για τις απειλές που τους αφορούν, επιτρέποντας την καλύτερη προσαρμογή των μέτρων ασφαλείας. Με τη χρήση αλγορίθμων μηχανικής μάθησης, τα συστήματα ΤΝ μπορούν να εντοπίζουν ανωμαλίες στα δεδομένα, καθώς και να προβλέπουν επιθέσεις που δεν ακολουθούν τα γνωστά πρότυπα ή τα παραδοσιακά μοντέλα απειλών, όπως επιθέσεις zero-day [3], [12].

Ένα από τα πιο εντυπωσιακά παραδείγματα της εφαρμογής της ανάλυσης μεγάλων δεδομένων είναι σε περιβάλλοντα cloud computing, όπου ο τεράστιος όγκος δεδομένων που παράγεται από τους διακομιστές και τις συνδεδεμένες συσκευές απαιτεί συνεχή παρακολούθηση και ανάλυση. Οι αλγόριθμοι ΤΝ που εφαρμόζονται σε cloud περιβάλλοντα είναι σε θέση να εντοπίζουν ασυνήθιστες ενέργειες, όπως η ξαφνική αύξηση της πρόσβασης από συγκεκριμένα IP ή ασυνήθιστες αλλαγές σε αρχεία, και να προειδοποιούν τις ομάδες ασφαλείας για πιθανές επιθέσεις πριν αυτές προχωρήσουν σε καταστροφικές ενέργειες [12].

Μια σημαντική πτυχή της ανάλυσης μεγάλων δεδομένων είναι η χρήση εργαλείων όπως το Apache Hadoop και το Apache Spark, τα οποία επιτρέπουν την ταυτόχρονη ανάλυση τεράστιων ποσοτήτων δεδομένων. Αυτά τα εργαλεία είναι σε θέση να επεξεργαστούν δεδομένα σε κλίμακα και να προσφέρουν άμεσες πληροφορίες για τις τρέχουσες απειλές. Επιπλέον, οι αλγόριθμοι ΤΝ μαθαίνουν από προηγούμενα δεδομένα, επιτρέποντας τη συνεχή βελτίωση και την αυτοπροσαρμογή των συστημάτων ασφαλείας για την αντιμετώπιση νέων και πιο εξελιγμένων απειλών [12].

Η ανάλυση μεγάλων δεδομένων παίζει κρίσιμο ρόλο στη μείωση των ψευδών θετικών ανιχνεύσεων, καθώς τα συστήματα ΤΝ μπορούν να εντοπίσουν φυσιολογικές αποκλίσεις στη δραστηριότητα των χρηστών και να τις ξεχωρίσουν από πραγματικές απειλές. Αυτό μειώνει τον χρόνο απόκρισης των διαχειριστών ασφαλείας και τους επιτρέπει να εστιάσουν σε πραγματικές και σοβαρές επιθέσεις, βελτιώνοντας τη συνολική αποδοτικότητα των διαδικασιών κυβερνοασφάλειας [3].

6.1.2 Ανάπτυξη Νέων Αλγορίθμων ΤΝ για Κυβερνοασφάλεια (Development of New AI Algorithms for Cybersecurity)

Η ανάπτυξη νέων αλγορίθμων Τεχνητής Νοημοσύνης (ΤΝ) αποτελεί μία από τις πιο σημαντικές προκλήσεις και ευκαιρίες στον τομέα της κυβερνοασφάλειας, καθώς οι επιθέσεις στον κυβερνοχώρο γίνονται όλο και πιο πολύπλοκες. Η ταχύτητα με την οποία εξελίσσονται οι απειλές απαιτεί νέες προσεγγίσεις και καινοτόμους αλγορίθμους που μπορούν να αντιμετωπίσουν τις σύγχρονες μορφές επιθέσεων, όπως οι επιθέσεις zero-day και οι επιθέσεις στα δεδομένα εκπαίδευσης. Οι νέοι αλγόριθμοι ΤΝ έχουν σχεδιαστεί για να μαθαίνουν και να προσαρμόζονται συνεχώς σε νέες απειλές, προσφέροντας τη δυνατότητα πρόληψης επιθέσεων σε πραγματικό χρόνο, καθώς και αυτόματης απόκρισης στις επιθέσεις [15].

Ένα από τα σημαντικότερα παραδείγματα νέων αλγορίθμων ΤΝ είναι οι τεχνικές adversarial machine learning. Αυτές οι τεχνικές επικεντρώνονται στην εκπαίδευση των μοντέλων ΤΝ ώστε να ανιχνεύουν και να αντιστέκονται σε επιθέσεις που εκμεταλλεύονται αδυναμίες στα δεδομένα εισόδου. Στις επιθέσεις αυτού του τύπου, οι επιτιθέμενοι τροποποιούν ελάχιστα τα δεδομένα με σκοπό να ξεγελάσουν το σύστημα και να προκαλέσουν λάθος αποφάσεις ή προβλέψεις. Οι τεχνικές adversarial learning βοηθούν τα συστήματα να γίνουν πιο ανθεκτικά, μειώνοντας την πιθανότητα επιτυχούς επίθεσης από παραπλανητικές εισόδους και εξασφαλίζοντας μεγαλύτερη ασφάλεια σε κρίσιμες εφαρμογές, όπως οι οικονομικές συναλλαγές και η διαχείριση δεδομένων υγείας [16].

Επιπλέον, οι αλγόριθμοι deep learning παίζουν σημαντικό ρόλο στη βελτίωση των συστημάτων κυβερνοασφάλειας. Χρησιμοποιώντας βαθιά νευρωνικά δίκτυα, οι νέες τεχνικές είναι σε θέση να ανιχνεύουν ανωμαλίες στα δεδομένα και να προβλέπουν με μεγαλύτερη ακρίβεια τις επιθέσεις. Η ικανότητα των αλγορίθμων deep learning να αναλύουν τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο, και να αναγνωρίζουν πολύπλοκα μοτίβα που δεν είναι ορατά με παραδοσιακές

μεθόδους, καθιστά αυτές τις τεχνολογίες απαραίτητες για την προστασία των σύγχρονων πληροφοριακών συστημάτων. Αυτή η προσέγγιση ενισχύει την ικανότητα ανίχνευσης και αντίδρασης σε επιθέσεις που δεν έχουν καταγραφεί στο παρελθόν, όπως οι zero-day επιθέσεις, οι οποίες εκμεταλλεύονται άγνωστα κενά ασφαλείας [16].

Μια άλλη καινοτόμος τεχνολογία είναι οι αλγόριθμοι reinforcement learning, οι οποίοι επιτρέπουν στα συστήματα ΤΝ να μαθαίνουν μέσω αλληλεπιδράσεων με το περιβάλλον τους και να προσαρμόζονται αυτόματα σε νέες καταστάσεις. Η χρήση αυτών των αλγορίθμων επιτρέπει στα συστήματα να λαμβάνουν καλύτερες αποφάσεις σε πραγματικό χρόνο, μειώνοντας τον χρόνο απόκρισης σε επιθέσεις και επιτρέποντας την αυτόνομη προστασία των δικτύων και των συστημάτων από κακόβουλες δραστηριότητες [15].

Η ενσωμάτωση αυτών των αλγορίθμων ΤΝ στις διαδικασίες κυβερνοασφάλειας δεν παρέχει μόνο καλύτερη ανίχνευση και αντίδραση στις επιθέσεις, αλλά και τη δυνατότητα πρόβλεψης απειλών με βάση προηγούμενα μοτίβα. Οι σύγχρονοι αλγόριθμοι ΤΝ δεν βασίζονται μόνο στην ανίχνευση γνωστών απειλών, αλλά αναλύουν συνεχώς τα δεδομένα για την πρόβλεψη μελλοντικών επιθέσεων. Με την ανάλυση των μοτίβων δραστηριότητας και τη συνεχή μάθηση από τις αλληλεπιδράσεις με τους χρήστες και τα δίκτυα, οι αλγόριθμοι αυτοί βοηθούν στην πρόληψη επιθέσεων πριν καν αυτές εκδηλωθούν, ενισχύοντας την συνολική ασφάλεια των συστημάτων [16].

6.2 Προκλήσεις και Προτεινόμενες Λύσεις (Challenges and Proposed Solutions)

Στην ενότητα αυτή θα αναλύσουμε τις βασικές προκλήσεις που αντιμετωπίζει ο τομέας της κυβερνοασφάλειας καθώς και τις προτεινόμενες λύσεις που αναδύονται για την αντιμετώπισή τους. Καθώς τα συστήματα γίνονται ολοένα και πιο σύνθετα και τα δεδομένα αυξάνονται ραγδαία, οι απειλές στον κυβερνοχώρο γίνονται πιο εξελιγμένες, προκαλώντας την ανάγκη για νέες τεχνολογικές λύσεις και στρατηγικές. Οι δύο κύριες προκλήσεις που εξετάζονται είναι η διαχείριση της πολυπλοκότητας των συστημάτων και η διασφάλιση της ιδιωτικότητας και της ασφάλειας των δεδομένων. Μέσα από τις αναδυόμενες τεχνολογίες, όπως η αρχιτεκτονική μηδενικής εμπιστοσύνης (Zero Trust) και οι τεχνικές κρυπτογράφησης, επιχειρείται η αντιμετώπιση αυτών των ζητημάτων, διατηρώντας την κυβερνοασφάλεια σε υψηλά επίπεδα.

6.2.1 Αντιμετώπιση της πολυπλοκότητας (Complexity Management)

Η αντιμετώπιση της πολυπλοκότητας αποτελεί μια από τις μεγαλύτερες προκλήσεις στον τομέα της κυβερνοασφάλειας, καθώς τα σύγχρονα πληροφοριακά συστήματα γίνονται ολοένα και πιο πολύπλοκα. Η πολυπλοκότητα δεν αφορά μόνο την αρχιτεκτονική των συστημάτων, αλλά και τον όγκο των δεδομένων που αυτά επεξεργάζονται, τον αριθμό των χρηστών που αλληλεπιδρούν με αυτά και τον πολλαπλασιασμό των συνδεδεμένων συσκευών και εφαρμογών. Η διαχείριση της πολυπλοκότητας καθίσταται κρίσιμη, καθώς οι επιτιθέμενοι εκμεταλλεύονται συχνά κενά ασφαλείας που προκύπτουν από την πολυπλοκότητα των συστημάτων, όπως μη ενημερωμένα στοιχεία ή ελλιπείς μηχανισμοί ασφαλείας [9].

Μια προτεινόμενη λύση για την αντιμετώπιση αυτής της πρόκλησης είναι η υιοθέτηση μιας αρχιτεκτονικής μηδενικής εμπιστοσύνης (Zero Trust Architecture), όπου κάθε πρόσβαση σε πόρους του δικτύου ελέγχεται και πιστοποιείται ανεξάρτητα από την τοποθεσία του χρήστη ή του πόρου. Αυτή η προσέγγιση μειώνει την πιθανότητα επιθέσεων που εκμεταλλεύονται την πολυπλοκότητα των συστημάτων, καθώς οι επιτιθέμενοι δεν μπορούν να κινηθούν ελεύθερα μέσα στο δίκτυο αν αποκτήσουν πρόσβαση [9].

Επιπλέον, η χρήση τεχνητής νοημοσύνης (ΤΝ) και προηγμένων αλγορίθμων ανάλυσης δεδομένων μπορεί να συμβάλει στη διαχείριση της πολυπλοκότητας, καθώς επιτρέπει την αυτοματοποίηση πολλών διαδικασιών ανίχνευσης και απόκρισης σε επιθέσεις. Οι αλγόριθμοι αυτοί μπορούν να εντοπίσουν ανωμαλίες σε πραγματικό χρόνο και να προτείνουν άμεσα λύσεις, μειώνοντας την ανάγκη για ανθρώπινη παρέμβαση και περιορίζοντας τα σημεία αδυναμίας που προκύπτουν από την πολυπλοκότητα των συστημάτων [13].

Επίσης, η απλοποίηση των διαδικασιών ασφαλείας μέσω της χρήσης ενιαίων συστημάτων διαχείρισης ταυτοτήτων και πρόσβασης (IAM) μπορεί να βοηθήσει στην αντιμετώπιση της πολυπλοκότητας. Με τα IAM, οι οργανισμοί μπορούν να ενοποιήσουν και να αυτοματοποιήσουν τον έλεγχο πρόσβασης σε συστήματα, εφαρμογές και δεδομένα, διευκολύνοντας τη διαχείριση των χρηστών και μειώνοντας τον κίνδυνο ανθρώπινων σφαλμάτων [9].

6.2.2 Εξασφάλιση της Ιδιωτικότητας και της Ασφάλειας (Ensuring Privacy and Security)

Η εξασφάλιση της ιδιωτικότητας και της ασφάλειας αποτελεί μια διαρκή πρόκληση στην εποχή του ψηφιακού μετασχηματισμού, καθώς η ποσότητα των ευαίσθητων δεδομένων που διακινούνται συνεχώς αυξάνεται. Τα συστήματα τεχνητής νοημοσύνης (ΤΝ) καλούνται να επεξεργάζονται δεδομένα προσωπικού χαρακτήρα για να παρέχουν πιο αποτελεσματικές υπηρεσίες, όμως αυτό δημιουργεί κινδύνους σχετικά με την παραβίαση της ιδιωτικότητας. Η κύρια πρόκληση έγκειται στην επίτευξη ισορροπίας μεταξύ της πρόσβασης σε δεδομένα και της διασφάλισης της ιδιωτικότητας των χρηστών [11].

Μια βασική τεχνολογία για την προστασία της ιδιωτικότητας είναι η κρυπτογράφηση, η οποία διασφαλίζει ότι τα δεδομένα παραμένουν απόρρητα κατά τη μεταφορά και την επεξεργασία τους. Ιδιαίτερα χρήσιμες είναι οι τεχνικές πλήρως ομομορφικής κρυπτογράφησης (Fully Homomorphic Encryption – FHE), οι οποίες επιτρέπουν την εκτέλεση υπολογισμών σε κρυπτογραφημένα δεδομένα χωρίς να απαιτείται η αποκρυπτογράφησή τους. Αυτό διασφαλίζει ότι ακόμη και αν τα δεδομένα αποκτήσουν μη εξουσιοδοτημένοι χρήστες, δεν θα είναι δυνατό να τα εκμεταλλευτούν [14].

Μια άλλη τεχνολογία που μπορεί να συμβάλει στην προστασία της ιδιωτικότητας είναι η τεχνολογία διαφοροποιημένης ιδιωτικότητας (Differential Privacy). Η διαφοροποιημένη ιδιωτικότητα επιτρέπει την ανάλυση μεγάλων συνόλων δεδομένων χωρίς να αποκαλύπτονται προσωπικά στοιχεία, μειώνοντας τον κίνδυνο ταυτοποίησης ατόμων μέσα από τα δεδομένα. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη για οργανισμούς που επεξεργάζονται μεγάλα ποσά δεδομένων, όπως χρηματοοικονομικά ιδρύματα και πάροχοι υγειονομικών υπηρεσιών [14].

Επιπλέον, η εφαρμογή αυστηρών πρωτοκόλλων πρόσβασης στα δεδομένα, όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR), ενισχύει τη διασφάλιση της ιδιωτικότητας και της ασφάλειας. Η συμμόρφωση με τα διεθνή πρότυπα και τους κανονισμούς βοηθά τους οργανισμούς να διασφαλίσουν ότι τα δεδομένα των χρηστών τους προστατεύονται, ενώ ταυτόχρονα αποφεύγουν νομικές κυρώσεις [11].

6.3 Διεπιστημονικές Ερευνητικές Κατευθύνσεις (Interdisciplinary Research Directions)

Οι διεπιστημονικές ερευνητικές κατευθύνσεις αποτελούν έναν από τους πιο σημαντικούς τομείς για την ανάπτυξη καινοτόμων λύσεων στην κυβερνοασφάλεια, καθώς απαιτούν τη συνεργασία μεταξύ διαφορετικών επιστημονικών κλάδων, όπως η πληροφορική, η μηχανική μάθηση, η νομική επιστήμη και η διοίκηση επιχειρήσεων. Οι κυβερνοαπειλές είναι πολυδιάστατες και η αποτελεσματική αντιμετώπισή τους απαιτεί συνδυασμό γνώσεων και μεθοδολογιών από διαφορετικά επιστημονικά πεδία, καθώς και την ανάπτυξη διεπιστημονικών συνεργασιών για τη δημιουργία πιο ολοκληρωμένων και ανθεκτικών συστημάτων [16].

Η συνεργασία μεταξύ ερευνητικών κέντρων και πανεπιστημίων είναι απαραίτητη για την προώθηση της έρευνας στον τομέα της κυβερνοασφάλειας και της τεχνητής νοημοσύνης (ΤΝ). Τα ερευνητικά κέντρα έχουν τη δυνατότητα να παρέχουν πρόσβαση σε εξειδικευμένο εξοπλισμό και μεγάλες ποσότητες δεδομένων, ενώ τα πανεπιστήμια διαθέτουν την απαραίτητη θεωρητική γνώση και την ικανότητα να αναπτύσσουν νέες τεχνολογίες και αλγορίθμους. Αυτές οι συνεργασίες ενισχύουν την ανάπτυξη καινοτόμων λύσεων που αντιμετωπίζουν τις αυξανόμενες κυβερνοαπειλές και προσφέρουν τη δυνατότητα δοκιμών σε πραγματικά περιβάλλοντα [24].

Ένα χαρακτηριστικό παράδειγμα αυτών των συνεργασιών είναι η σύμπραξη μεταξύ ακαδημαϊκών ιδρυμάτων και βιομηχανικών ερευνητικών εργαστηρίων. Αυτές οι συνεργασίες έχουν οδηγήσει στην ανάπτυξη νέων μεθόδων για την ανίχνευση και αντιμετώπιση επιθέσεων, όπως η χρήση deep learning και advanced analytics για την ανίχνευση επιθέσεων τύπου zero-day. Οι βιομηχανίες επωφελούνται από την εφαρμογή αυτών των τεχνολογιών στα πληροφοριακά τους συστήματα, ενώ τα πανεπιστήμια και τα ερευνητικά κέντρα αποκτούν πολύτιμα δεδομένα και εμπειρία από τον πραγματικό κόσμο [16].

Επιπλέον, οι συνεργασίες αυτές προωθούν την ανταλλαγή γνώσεων και τεχνολογίας μεταξύ ερευνητικών φορέων και επιχειρήσεων, δημιουργώντας μια ισχυρή βάση για την ανάπτυξη πρακτικών λύσεων σε θέματα κυβερνοασφάλειας. Οι ερευνητικές ομάδες από πανεπιστήμια και βιομηχανίες μπορούν να συνεργαστούν για την εφαρμογή τεχνολογιών τεχνητής νοημοσύνης σε διαφορετικά πεδία, από τη χρηματοοικονομική ανάλυση μέχρι τη δημόσια ασφάλεια, προσφέροντας πολυδιάστατες λύσεις για την αντιμετώπιση κυβερνοεπιθέσεων [24].

Οι διεπιστημονικές κατευθύνσεις δεν περιορίζονται μόνο στη συνεργασία μεταξύ των τεχνολογικών τομέων, αλλά επεκτείνονται και στη διαμόρφωση πολιτικών και στρατηγικών. Η συνεργασία με νομικούς και πολιτικούς αναλυτές συμβάλλει στην κατανόηση των νομοθετικών πλαισίων και των κανονισμών που διέπουν την κυβερνοασφάλεια, ενώ ταυτόχρονα διασφαλίζεται ότι οι τεχνολογικές λύσεις συμμορφώνονται με τα διεθνή πρότυπα ασφαλείας και ιδιωτικότητας [16].

7. Ρυθμιστικά πλαίσια για την Τεχνητή Νοημοσύνη - Ο Κανονισμός AI Act της Ευρωπαϊκής Ένωσης

Η Τεχνητή Νοημοσύνη (ΤΝ) έχει γίνει ένα από τα πιο καινοτόμα εργαλεία της ψηφιακής εποχής, με την ικανότητα να αλλάξει ριζικά τον τρόπο με τον οποίο λειτουργούν οι διάφοροι τομείς της οικονομίας, όπως η υγεία, οι μεταφορές, η εκπαίδευση και οι χρηματοοικονομικές υπηρεσίες. Παράλληλα με τα οφέλη της ΤΝ, όμως, προκύπτουν και σοβαρά ζητήματα δεοντολογίας και ασφάλειας, τα οποία καθιστούν απαραίτητη τη ρύθμιση της. Σε αυτό το πλαίσιο, η Ευρωπαϊκή Ένωση ανέλαβε την πρωτοβουλία να δημιουργήσει το **AI Act** (Artificial Intelligence Act), το πρώτο ολοκληρωμένο ρυθμιστικό πλαίσιο παγκοσμίως, με στόχο τη ρύθμιση της ανάπτυξης, της χρήσης και της εμπορικής διάθεσης συστημάτων ΤΝ.

Το AI Act είναι σχεδιασμένο για να διασφαλίσει την προστασία των θεμελιωδών δικαιωμάτων και την ασφάλεια των πολιτών, προσφέροντας παράλληλα το έδαφος για την ανάπτυξη και καινοτομία στην ΤΝ εντός της Ευρωπαϊκής Ένωσης. Σε αντίθεση με άλλες προσεγγίσεις, που επικεντρώνονται στην αυτορρύθμιση της βιομηχανίας, ο AI Act εισάγει δεσμευτικούς κανονισμούς και απαιτήσεις για τα συστήματα ΤΝ, κατηγοριοποιώντας τα σε κλίμακα κινδύνου. Αυτός ο κανονισμός επιδιώκει να εξισορροπήσει τα πλεονεκτήματα της ΤΝ με την ανάγκη για διαφάνεια και προστασία των δικαιωμάτων των πολιτών [30].

7.1 Κατηγοριοποίηση Κινδύνου των Συστημάτων Τεχνητής Νοημοσύνης

Μία από τις βασικές αρχές του AI Act είναι η κατηγοριοποίηση των συστημάτων ΤΝ με βάση τον κίνδυνο που ενέχουν. Οι κατηγορίες αυτές κυμαίνονται από υψηλό έως ελάχιστο κίνδυνο, και κάθε κατηγορία συνοδεύεται από συγκεκριμένες υποχρεώσεις και περιορισμούς.

Απαγορευμένες Πρακτικές

Περιλαμβάνουν συστήματα ΤΝ που θεωρούνται επικίνδυνα για τα θεμελιώδη δικαιώματα των πολιτών. Ένα χαρακτηριστικό παράδειγμα είναι η κοινωνική βαθμολόγηση από κυβερνητικές οντότητες, όπου οι πολίτες κατατάσσονται ή αξιολογούνται με βάση την κοινωνική τους συμπεριφορά. Αυτή η πρακτική θεωρείται ασυμβίβαστη με τα ευρωπαϊκά πρότυπα, καθώς θέτει σε κίνδυνο την ιδιωτικότητα, την αυτονομία και την ελευθερία των ατόμων [31].

Υψηλού Κινδύνου Συστήματα:

Οι εφαρμογές που εμπίπτουν στην κατηγορία υψηλού κινδύνου είναι αυτές που ενσωματώνονται σε κρίσιμους τομείς, όπως η ιατροφαρμακευτική περίθαλψη, οι διαδικασίες πρόσληψης, η αξιολόγηση της πιστοληπτικής ικανότητας, και η επιβολή του νόμου. Τα συστήματα αυτά απαιτούν εκτενείς αξιολογήσεις κινδύνου, αυστηρούς ελέγχους και τεχνικές συμμόρφωσης πριν και μετά την κυκλοφορία τους στην αγορά. Οι πάροχοι αυτών των συστημάτων υποχρεούνται να διασφαλίζουν την ποιότητα των δεδομένων, να εφαρμόζουν διαφανείς διαδικασίες και να επιτρέπουν την ανθρώπινη εποπτεία σε κρίσιμες λειτουργίες [30].

Περιορισμένου Κινδύνου Συστήματα

Τα συστήματα αυτά απαιτούν συγκεκριμένες υποχρεώσεις διαφάνειας. Ένα παράδειγμα τέτοιας υποχρέωσης είναι τα συστήματα ΤΝ που αλληλεπιδρούν άμεσα με τους χρήστες, όπως τα chatbots. Οι χρήστες θα πρέπει να ενημερώνονται ρητά ότι αλληλεπιδρούν με μία μηχανή και όχι με άνθρωπο. Αυτό μειώνει τον κίνδυνο εξαπάτησης των χρηστών και διασφαλίζει ότι τα συστήματα αυτά χρησιμοποιούνται για τους δηλωμένους σκοπούς τους [32].

Ελάχιστου Κινδύνου Συστήματα

Περιλαμβάνει τις περισσότερες εφαρμογές ΤΝ, όπως φίλτρα ανεπιθύμητης αλληλογραφίας (spam filters), τα οποία δεν απαιτούν πρόσθετες ρυθμιστικές απαιτήσεις ή υποχρεώσεις για τους παρόχους και τους χρήστες τους. Αυτές οι εφαρμογές θεωρούνται χαμηλού κινδύνου, και έτσι είναι ελεύθερες να χρησιμοποιούνται χωρίς επιπρόσθετη εποπτεία ή περιορισμούς [31].

7.2 Υποχρεώσεις για Παρόχους και Χρήστες Συστημάτων ΤΝ

Το AI Act επιβάλλει μια σειρά υποχρεώσεων στους παρόχους και χρήστες συστημάτων ΤΝ. Οι πάροχοι, ιδιαίτερα όσοι προσφέρουν συστήματα υψηλού κινδύνου, υποχρεούνται να διασφαλίζουν ότι τα συστήματα τους είναι διαφανή, αξιόπιστα και ασφαλή. Αυτοί οι πάροχοι

πρέπει να εξασφαλίζουν τη συνεχή παρακολούθηση και την αξιολόγηση της λειτουργίας των συστημάτων τους, να διασφαλίζουν την ποιότητα των δεδομένων και να διατηρούν αναλυτικά αρχεία για τον τρόπο εκπαίδευσης και δοκιμών των μοντέλων τους [33]. Επίσης, απαιτείται να ενημερώνουν και να καθοδηγούν τους χρήστες σχετικά με τη σωστή χρήση και τις δυνατότητες των συστημάτων τους, ενώ οι χρήστες οφείλουν να παρακολουθούν τη λειτουργία αυτών των συστημάτων και να τα χρησιμοποιούν σύμφωνα με τις προβλεπόμενες οδηγίες.

7.3 Διακυβέρνηση και Εποπτεία

Για την εξασφάλιση της ορθής εφαρμογής του κανονισμού, το AI Act προβλέπει τη δημιουργία του Ευρωπαϊκού Γραφείου Τεχνητής Νοημοσύνης (European Artificial Intelligence Board). Το γραφείο αυτό θα συντονίζει και θα εποπτεύει την εφαρμογή του AI Act σε επίπεδο Ευρωπαϊκής Ένωσης, ενισχύοντας τη συνεργασία και τη συνέπεια μεταξύ των κρατών μελών. Επίσης, τα κράτη μέλη υποχρεούνται να ορίσουν εθνικές αρχές υπεύθυνες για την επιβολή του κανονισμού, οι οποίες θα διασφαλίζουν τη συμμόρφωση των παρόχων και χρηστών με τις απαιτήσεις του κανονισμού [30].

7.4 Κώδικες Δεοντολογίας και Εθελοντική Συμμόρφωση

Ο κανονισμός προτρέπει την ανάπτυξη εθελοντικών κωδίκων δεοντολογίας για τα συστήματα ΤΝ που δεν κατατάσσονται στις κατηγορίες υψηλού κινδύνου. Αυτοί οι κώδικες δεοντολογίας προσφέρουν κατευθυντήριες γραμμές για την υπεύθυνη ανάπτυξη και χρήση της ΤΝ, καλώντας τις εταιρείες να τηρούν βέλτιστες πρακτικές, ακόμα και όταν δεν υπόκεινται σε αυστηρές νομικές απαιτήσεις. Ο στόχος είναι να ενθαρρύνει τη βιομηχανία να αναλάβει πρωτοβουλίες που θα ενισχύσουν την εμπιστοσύνη του κοινού στη χρήση της ΤΝ [31].

7.5 Χρονοδιάγραμμα Εφαρμογής του Κανονισμού

Το AI Act τέθηκε σε ισχύ την 1η Αυγούστου 2024, και οι περισσότερες από τις διατάξεις του θα αρχίσουν να εφαρμόζονται επίσημα στις 2 Αυγούστου 2026. Το μεταβατικό αυτό διάστημα προσφέρει στις εταιρείες και τους παρόχους τη δυνατότητα να προσαρμοστούν στα νέα δεδομένα και να διασφαλίσουν ότι πληρούν τις νομικές απαιτήσεις του κανονισμού. Η Ευρωπαϊκή Ένωση ενθαρρύνει τη συμμόρφωση και παρέχει καθοδήγηση για να διευκολύνει την υιοθέτηση αυτών των νέων προτύπων [33].

7.6 Συμπεράσματα για το Ρυθμιστικό Πλαίσιο της Ευρωπαϊκής Ένωσης στην Τεχνητή Νοημοσύνη

Το AI Act αποτελεί σημαντική καινοτομία στην πολιτική ρύθμισης της ΤΝ, αναγνωρίζοντας τόσο τις δυνατότητες όσο και τους κινδύνους της τεχνολογίας αυτής. Η Ευρωπαϊκή Ένωση, μέσα από αυτό το ρυθμιστικό πλαίσιο, επιδιώκει να εξισορροπήσει την ανάπτυξη της καινοτομίας με την προστασία των δικαιωμάτων των πολιτών, δημιουργώντας ένα περιβάλλον ασφάλειας και διαφάνειας. Η εισαγωγή του AI Act αναμένεται να διαμορφώσει τα πρότυπα για τις ρυθμίσεις της ΤΝ σε παγκόσμιο επίπεδο, αποτελώντας σημείο αναφοράς για άλλες χώρες και οργανισμούς που ενδιαφέρονται για τη ρύθμιση της τεχνητής νοημοσύνης [32].

Το συγκεκριμένο ρυθμιστικό πλαίσιο παρέχει ένα πρότυπο για την υπεύθυνη και ασφαλή ανάπτυξη της ΤΝ, ενθαρρύνοντας παράλληλα τις εταιρείες να ενσωματώσουν προληπτικές πρακτικές προστασίας και να ενισχύσουν την εμπιστοσύνη των χρηστών στα συστήματα ΤΝ που χρησιμοποιούν.

8. Συμπεράσματα

8.1 Κύρια Συμπεράσματα της Έρευνας

Η παρούσα έρευνα ασχολήθηκε διεξοδικά με την εφαρμογή της Τεχνητής Νοημοσύνης (ΤΝ) στην κυβερνοασφάλεια, αναδεικνύοντας τις δυνατότητές της στην ανίχνευση και αντιμετώπιση σύγχρονων κυβερνοαπειλών. Η ΤΝ έχει ήδη επιφέρει ριζικές αλλαγές στον τομέα της κυβερνοασφάλειας, κυρίως μέσω της αυτοματοποίησης διαδικασιών ανίχνευσης, ανάλυσης, και απόκρισης σε επιθέσεις, ενισχύοντας την αποτελεσματικότητα των συστημάτων ασφαλείας και μειώνοντας το χρόνο αντίδρασης.

Ένα από τα σημαντικότερα ευρήματα της έρευνας είναι η ικανότητα της ΤΝ να επεξεργάζεται τεράστιες ποσότητες δεδομένων σε πραγματικό χρόνο. Αυτή η δυνατότητα ενισχύεται από τεχνικές μηχανικής μάθησης και βαθιάς μάθησης, οι οποίες επιτρέπουν στα συστήματα να αναγνωρίζουν μοτίβα και ανωμαλίες που μπορεί να υποδηλώνουν την ύπαρξη κακόβουλων δραστηριοτήτων. Μέσω της ανάλυσης αυτών των δεδομένων, τα συστήματα ΤΝ μπορούν να προσφέρουν προληπτική προστασία, αποτρέποντας επιθέσεις πριν αυτές εκδηλωθούν. Αυτό επιτρέπει τη βελτίωση της ασφάλειας των συστημάτων και των υποδομών, ενώ ταυτόχρονα μειώνει την ανάγκη για ανθρώπινη παρέμβαση [1], [5].

Η ΤΝ διαδραματίζει επίσης κρίσιμο ρόλο στην ανίχνευση απειλών που δεν ακολουθούν τα παραδοσιακά πρότυπα επιθέσεων. Ένα χαρακτηριστικό παράδειγμα είναι η ανίχνευση επιθέσεων τύπου zero-day, οι οποίες εκμεταλλεύονται άγνωστα κενά ασφαλείας. Αυτές οι επιθέσεις, που παραδοσιακά ήταν δύσκολο να εντοπιστούν, μπορούν τώρα να ανιχνευθούν μέσω της χρήσης τεχνικών βαθιάς μάθησης, οι οποίες εξετάζουν δεδομένα σε διαφορετικά επίπεδα αφαίρεσης και εντοπίζουν ανεπαίσθητες αποκλίσεις στη συμπεριφορά των χρηστών ή των συστημάτων [5]. Η δυνατότητα της ΤΝ να εντοπίζει αυτές τις επιθέσεις πριν προκληθεί ζημιά, την καθιστά ζωτικής σημασίας για τη διασφάλιση της ακεραιότητας των συστημάτων ασφαλείας.

Ένα άλλο σημαντικό συμπέρασμα είναι η αυτοματοποίηση που προσφέρει η ΤΝ στην αντιμετώπιση των απειλών. Τα συστήματα ΤΝ έχουν τη δυνατότητα να αναγνωρίζουν αυτόματα ύποπτες δραστηριότητες και να λαμβάνουν τα κατάλληλα μέτρα αποτροπής ή απομόνωσης της επίθεσης. Αυτό περιλαμβάνει τη δυνατότητα περιορισμού της πρόσβασης, την απομόνωση μολυσμένων συστημάτων, ή την εφαρμογή πρόσθετων ασφαλιστικών δικλίδων. Αυτή η αυτοματοποίηση όχι μόνο μειώνει τον χρόνο απόκρισης, αλλά και τις πιθανότητες επιτυχούς ολοκλήρωσης μιας επίθεσης [5]. Επιπλέον, η ικανότητα των συστημάτων να μαθαίνουν από τα δεδομένα τους και να προσαρμόζονται σε νέες απειλές, διασφαλίζει ότι τα μοντέλα ΤΝ παραμένουν ενημερωμένα και ικανά να ανιχνεύουν ακόμα και εξελιγμένες επιθέσεις.

Παρόλα αυτά, η έρευνα ανέδειξε και ορισμένες προκλήσεις που συνοδεύουν την εφαρμογή της ΤΝ στην κυβερνοασφάλεια. Ένας από τους πιο βασικούς περιορισμούς είναι η ανάγκη για μεγάλα και ποιοτικά σύνολα δεδομένων για την εκπαίδευση των αλγορίθμων ΤΝ. Η συγκέντρωση και διαχείριση αυτών των δεδομένων μπορεί να είναι απαιτητική και να εγείρει σημαντικά ζητήματα προστασίας της ιδιωτικότητας [1]. Επιπλέον, τα μοντέλα ΤΝ, αν και αποτελεσματικά, είναι ευάλωτα σε επιθέσεις τύπου adversarial, κατά τις οποίες οι επιτιθέμενοι εκμεταλλεύονται παραποιημένα δεδομένα για να παραπλανήσουν τα συστήματα ασφαλείας. Αυτές οι επιθέσεις υπογραμμίζουν την ανάγκη για περαιτέρω έρευνα στην ανάπτυξη ανθεκτικών αλγορίθμων και τεχνικών που θα ενισχύσουν την ασφάλεια των μοντέλων ΤΝ [5].

Συνολικά, η παρούσα έρευνα επιβεβαιώνει ότι η ΤΝ προσφέρει ανυπολόγιστα πλεονεκτήματα στον τομέα της κυβερνοασφάλειας, αλλά και επισημαίνει την ανάγκη για συνεχή ανάπτυξη και εξέλιξη των τεχνολογιών αυτών, ώστε να παραμείνουν αποτελεσματικές ενάντια στις νέες μορφές κυβερνοαπειλών.

8.2 Προτάσεις για Μελλοντική Έρευνα

Η παρούσα έρευνα ανέδειξε τις δυνατότητες της Τεχνητής Νοημοσύνης (ΤΝ) στην κυβερνοασφάλεια, αλλά παραμένουν πολλά πεδία για μελλοντική έρευνα που θα μπορούσαν να βελτιώσουν περαιτέρω την αποτελεσματικότητα και την ανθεκτικότητα των συστημάτων ασφαλείας. Οι νέες απειλές και οι προκλήσεις που αντιμετωπίζουν τα πληροφοριακά συστήματα απαιτούν συνεχή προσαρμογή και καινοτομία, ώστε να διατηρηθεί η εμπιστοσύνη των οργανισμών και των χρηστών στην ασφάλεια των ψηφιακών υποδομών τους.

Ένας από τους βασικότερους τομείς για μελλοντική έρευνα είναι η ανάπτυξη νέων τεχνικών που θα είναι ανθεκτικές στις επιθέσεις τύπου adversarial. Οι επιθέσεις αυτές, κατά τις οποίες οι επιτιθέμενοι προσπαθούν να παραπλανήσουν τα συστήματα ΤΝ μέσω παραποιημένων δεδομένων, αποτελούν μία από τις πιο σημαντικές προκλήσεις στην εφαρμογή της ΤΝ στην κυβερνοασφάλεια. Η ανάπτυξη αλγορίθμων που θα μπορούν να ανιχνεύουν τέτοιες επιθέσεις και να διατηρούν την αποτελεσματικότητά τους απέναντι σε παραποιημένα δεδομένα είναι απαραίτητη για την περαιτέρω εξέλιξη των συστημάτων αυτών. Οι τεχνικές adversarial training, όπου τα μοντέλα εκπαιδεύονται με παραπλανητικά παραδείγματα, αποτελούν ένα βασικό σημείο εκκίνησης για την ανθεκτικότητα των συστημάτων, αλλά απαιτούν περαιτέρω βελτιώσεις και προσαρμογές [2], [12].

Επιπλέον, η ενοποίηση της ΤΝ με άλλες αναδυόμενες τεχνολογίες, όπως το blockchain, αποτελεί ένα σημαντικό πεδίο για περαιτέρω μελέτη. Το blockchain, με την αποκεντρωμένη αρχιτεκτονική του και την ικανότητα του να διασφαλίζει την ακεραιότητα και τη διαφάνεια των δεδομένων, μπορεί να ενισχύσει τη λειτουργία των συστημάτων ΤΝ, προσφέροντας μεγαλύτερη προστασία από επιθέσεις. Αυτή η συνεργασία θα μπορούσε να προσφέρει νέες λύσεις για την αποτροπή επιθέσεων και την προστασία των δεδομένων σε κατανεμημένα περιβάλλοντα, ενισχύοντας την ασφάλεια σε κρίσιμα πληροφοριακά συστήματα [12].

Ένας τρίτος τομέας που απαιτεί περισσότερη έρευνα είναι η βελτίωση της επεξηγησιμότητας των αλγορίθμων ΤΝ. Παρόλο που τα συστήματα ΤΝ προσφέρουν εξαιρετικές δυνατότητες ανίχνευσης και αντίδρασης σε κυβερνοεπιθέσεις, ένα από τα σημαντικότερα εμπόδια για την ευρύτερη υιοθέτησή τους είναι η έλλειψη διαφάνειας στις αποφάσεις που λαμβάνουν. Η ανάπτυξη μεθόδων

που επιτρέπουν την κατανόηση των εσωτερικών διεργασιών της ΤΝ, όπως η επεξηγησιμότητα και η διαφάνεια των αποφάσεων, θα ενισχύσει την εμπιστοσύνη των οργανισμών και των χρηστών στη χρήση αυτής της τεχνολογίας. Η ενσωμάτωση εργαλείων που θα παρέχουν καλύτερη κατανόηση των αποφάσεων που λαμβάνονται από τα συστήματα ΤΝ θα μπορούσε να αποτελέσει έναν σημαντικό καταλύτη για την ευρύτερη αποδοχή της τεχνολογίας στον τομέα της κυβερνοασφάλειας [2].

Παράλληλα, σημαντική έρευνα απαιτείται για την ενσωμάτωση της ΤΝ σε νέα περιβάλλοντα και τεχνολογίες, όπως τα συστήματα Internet of Things (IoT) και τα δίκτυα 5G. Η πολυπλοκότητα αυτών των περιβαλλόντων και η ευπάθεια σε νέες μορφές επιθέσεων καθιστούν απαραίτητη την ανάπτυξη καινοτόμων λύσεων που θα μπορούν να λειτουργήσουν αποτελεσματικά σε αυτά τα πλαίσια. Η έρευνα σε αυτούς τους τομείς μπορεί να οδηγήσει σε βελτιώσεις στην ανίχνευση και την αντιμετώπιση κυβερνοαπειλών, επιτρέποντας στα συστήματα ΤΝ να διαδραματίσουν ακόμα μεγαλύτερο ρόλο στην προστασία των μελλοντικών δικτύων και υποδομών [12].

Συνοψίζοντας, η μελλοντική έρευνα στην ΤΝ για την κυβερνοασφάλεια πρέπει να επικεντρωθεί στην αντιμετώπιση των επιθέσεων adversarial, στη συνεργασία με τεχνολογίες όπως το blockchain, στη βελτίωση της επεξηγησιμότητας των αλγορίθμων, καθώς και στην εφαρμογή της σε νέες τεχνολογικές υποδομές. Οι προκλήσεις αυτές προσφέρουν εξαιρετικές ευκαιρίες για καινοτομία και εξέλιξη, ενισχύοντας την προστασία των πληροφοριακών συστημάτων σε ένα συνεχώς εξελισσόμενο τοπίο κυβερνοαπειλών.

Βιβλιογραφία

[1] H. Hashemi, A. Azmoodeh, A. Hamzeh, and S. Hashemi, "Graph embedding as a new approach for unknown malware detection," *Journal of Computer Virology and Hacking Techniques*, vol. 13, 2017. doi: 10.1007/s11416-016-0278-y.

[2] M. J. Hossain Faruk *et al.*, "Malware detection and prevention using artificial intelligence techniques," in *2021 IEEE International Conference on Big Data (Big Data)*, 2021. doi: 10.1109/BigData52589.2021.9671434.

[3] J. Johns, "Representation learning for malware classification," FireEye, 2017. [Διαθέσιμο στο διαδίκτυο]: https://www.fireeye.com/content/dam/fireeye-www/blog/pdfs/malware-classification-slides.pdf.

[4] B. Kolosnjaji, A. Demontis, B. Biggio, D. Maiorca, G. Giacinto, C. Eckert, and F. Roli, "Adversarial malware binaries: evading deep learning for malware detection in executables," *arXiv preprint arXiv:1803.04173*, 2018.

[5] F. Kreuk *et al.*, "Deceiving end-to-end deep learning malware detectors using adversarial examples," *arXiv preprint arXiv:1802.04528*, 2018.

[6] B. Luo and J. Xia, "A novel intrusion detection system based on feature generation with visualization strategy," *Expert Systems with Applications*, vol. 41, pp. 4139–4147, 2014. doi: 10.1016/j.eswa.2013.12.048.

[7] R. Mahajan and I. Siddavatam, "Phishing website detection using machine learning algorithms," *International Journal of Computer Applications*, vol. 181, pp. 45–47, 2018. doi: 10.5120/ijca2018918026.

[8] N. McLaughlin *et al.*, "Deep android malware detection," in *2017 IEEE International Conference on Software Engineering Companion (ICSE-C)*, 2017. doi: 10.1145/3029806.3029823.

[9] L. Molina Valdiviezo, A. Furfaro, G. Malena, and A. Parise, "A simulation model for the analysis of DDOS amplification attacks," in *2015 UKSim-AMSS 17th International Conference on Computer Modelling and Simulation*, 2015. doi: 10.1109/UKSim.2015.52.

[10] B. Obotivere and A. Nwaezeigwe, "Cyber security threats on the internet and possible solutions," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 9, pp. 92–97, 2020. doi: 10.17148/IJARCCE.2020.9913.

[11] T. Peng, I. Harris, and Y. Sawa, "Detecting phishing attacks using natural language processing and machine learning," in *2018 IEEE International Conference on Semantic Computing (ICSC)*, 2018. doi: 10.1109/ICSC.2018.00056.

[12] M. O. F. Rokon, R. Islam, A. Darki, E. Papalexakis, and M. Faloutsos, "Sourcefinder: finding malware source-code from publicly available repositories in GitHub," 2020.

[13] L. Bilge and T. Dumitraş, "Before we knew it: an empirical study of zero-day attacks in the real world," in *Proceedings of the 2012 ACM Conference on Computer and Communications Security*, 2012, pp. 833–844.

[14] A. Moghimi, J. Wichelmann, T. Eisenbarth, and B. Sunar, "Memjam: a false dependency attack against constant-time crypto implementations," *International Journal of Parallel Programming*, vol. 47, no. 4, pp. 538–570, 2019.

[15] M. Warkentin and R. Willison, "Behavioral and policy issues in information systems security: the insider threat," *European Journal of Information Systems*, vol. 18, no. 2, pp. 101–105, 2009.

[16] M. Ohm, A. Sykosch, and M. Meier, "Towards detection of software supply chain attacks by forensic artifacts," in *Proceedings of the 15th International Conference on Availability, Reliability and Security*, 2020, pp. 1–6.

[17] S. Eggers, "A novel approach for analyzing the nuclear supply chain cyber-attack surface," *Nuclear Engineering and Technology*, vol. 53, no. 3, pp. 879–887, 2021.

[18] D. Kügler, "'Man in the middle' attacks on bluetooth," in *International Conference on Financial Cryptography and Data Security*, Springer, 2003, pp. 149–161.

[19] A. Shaw, "Data breach: from notification to prevention using PCI DSS," *Columbia Journal of Law and Social Problems*, vol. 43, pp. 517, 2009.

[20] S. Hong, "Survey on analysis and countermeasure for hacking attacks to cryptocurrency exchange," *Journal of Korea Convergence Society*, vol. 10, no. 10, pp. 1–6, 2019.

[21] S. W. Boyd and A. D. Keromytis, "Sqlrand: preventing sql injection attacks," in *International Conference on Applied Cryptography and Network Security*, Springer, 2004, pp. 292–302.

[22] F. Tong and Z. Yan, "A hybrid approach of mobile malware detection in Android," *Journal of Parallel and Distributed Computing*, vol. 103, pp. 22–31, 2017.

[23] V. G. Shankar, M. Jangid, B. Devi, and S. Kabra, "Mobile big data: malware and its analysis," in *Proceedings of the First International Conference on Smart System, Innovations and Computing*, Springer, 2018, pp. 831–842.

[24] L. Davi, A. Dmitrienko, A.-R. Sadeghi, and M. Winandy, "Privilege escalation attacks on Android," in *International Conference on Information Security*, Springer, 2010, pp. 346–360.

[25] N. Virvilis and D. Gritzalis, "The big four-what we did wrong in advanced persistent threat detection," in *2013 International Conference on Availability, Reliability and Security*, IEEE, 2013, pp. 248–254.

[26] IBM, "Cargills Bank Ltd. Case Study - QRadar Advisor with Watson." IBM. [Διαθέσιμο στο διαδίκτυο]: https://www.ibm.com/cargillsbank.

[27] Microsoft, "Cyber Signals: Navigating cyberthreats and strengthening defenses in the era of AI," *Microsoft Switzerland News Center*, 2024.

[28] Darktrace, "Case Study: Drax Group - Cybersecurity in the Energy Sector," 2024.

[29] H2O.ai, "Driving Away Fraudsters at PayPal," 2018.

[30] European Commission, “Regulatory framework proposal on artificial intelligence,” Digital Strategy of the EU, May 2021. [Διαθέσιμο στο διαδίκτυο]: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai

[31] European Commission, “European Commission News on AI Act,” European Commission, Aug. 1, 2024. [Διαθέσιμο στο διαδίκτυο] : https://commission.europa.eu/news/ai-act-enters-force-2024-08-01_en

[32] Council of the European Union, “Artificial intelligence act: Council gives final green light to the first worldwide rules on AI,” Press Release, May 2024. [Διαθέσιμο στο διαδίκτυο]: https://www.consilium.europa.eu/en/press/press-releases/2024/05/21/artificial-intelligence-ai-act-council-gives-final-green-light-to-the-first-worldwide-rules-on-ai/

[33] “AI Act,” DIGI.gov.gr, [Διαθέσιμο στο διαδίκτυο]: https://digi.gov.gr/ai-act