



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**ΕΠΙΔΡΑΣΗ ΑΛΓΟΡΙΘΜΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ
ΠΡΑΓΜΑΤΟΠΟΙΗΣΗ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΩΝ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΓΕΩΡΓΙΑΔΗ ΑΝΤΩΝΙΟΥ

Επιβλέπουσα: ΔΙΑΜΑΝΤΟΠΟΥΛΟΥ ΒΑΣΙΛΙΚΗ

Τριμελής Επιτροπή: ΔΙΑΜΑΝΤΟΠΟΥΛΟΥ ΒΑΣΙΛΙΚΗ, ΓΕΩΡΓΙΟΣ ΣΠΑΘΟΥΛΑΣ,
ΜΑΡΙΑ ΚΑΡΥΔΑ

Πρόλογος και ευχαριστίες

Η παρούσα Διπλωματική Εργασία εκπονήθηκε κατά την εαρινή περίοδο του Ακαδημαϊκού Έτους 2026-2027, στο πλαίσιο του Προπτυχιακού Προγράμματος Σπουδών του τμήματος Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων.

Η εργασία πραγματοποιήθηκε υπό την επίβλεψη της κα. Διαμαντοπούλου Βασιλικής, Καθηγήτρια του Τμήματος.

Αντικείμενο της εργασίας αποτελεί η επίδραση αλγορίθμων τεχνητής νοημοσύνης στην πραγματοποίηση κυβερνοεπιθέσεων.

Στο σημείο αυτό, οφείλω να εκφράσω τις θερμές μου ευχαριστίες, προς την επιβλέπουσα της εργασίας, καθηγήτρια κα. Διαμαντοπούλου Βασιλική, για την καθοδήγησή της, και την πολύτιμη βοήθεια που προσέφερε σε κάθε στάδιο εκπόνησης της διπλωματικής εργασίας μου.

Τέλος, ευχαριστώ θερμά την οικογένεια και τους φίλους μου, για την κατανόηση και συμπαράσταση που έδειξαν ολόκληρη την περίοδο εκπόνησης της εργασίας αυτής.

© 2026

του

ΓΕΩΡΓΙΑΔΗ ΑΝΤΩΝΙΟΥ

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

ΠΕΡΙΛΗΨΗ.....	6
1. ΕΙΣΑΓΩΓΗ	8
1.1 Παρουσίαση του θέματος και της σημασίας του	8
1.2 Σημασία της έρευνας	9
1.3. Σκοπός και ερευνητικά ερωτήματα	10
1.4. Στόχοι της έρευνας.....	11
1.5. Η μεθοδολογία της εργασίας	12
1.6. Η δομή της εργασίας.....	13
2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	15
2.1. Βασικές αρχές της Τεχνητής Νοημοσύνης.....	15
2.2. Βασικοί αλγόριθμοι / Αλγόριθμοι μηχανικής μάθησης και βαθιάς μάθησης.....	17
2.2.1. Επιβλεπόμενη Μάθηση.....	18
2.2.2. Μη Επιβλεπόμενη Μάθηση	18
2.2.3. Νευρωνικά Δίκτυα	19
2.3. Βασικές έννοιες κυβερνοεπιθέσεων / Ευπάθειες συστημάτων και κυβερνοεπιθέσεις.....	20
2.3.1. Malware	21
2.3.2. Phishing.....	22
2.3.3. Denial-of-Service Attacks.....	22
2.3.4. Σύγχρονες Τάσεις Κυβερνοεπιθέσεων.....	23
2.4. Σύγχρονες τεχνικές κυβερνοεπιθέσεων	23
2.4.1. Επιθέσεις Εκμετάλλευσης Ευπαθειών	24
2.4.2. Επιθέσεις Ταυτότητας και Υποκλοπή Διαπιστευτηρίων	24
2.4.3. Σύγχρονες Τεχνικές Phishing και Social Engineering	25
2.4.4. Επιθέσεις Ransomware και Κακόβουλου Λογισμικού.....	26
2.4.5. Επιθέσεις σε Υποδομές Νέφους και Εφοδιαστική Αλυσίδα	26
3. Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΩΣ ΕΡΓΑΛΕΙΟ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΩΝ	27
3.1. Αυτοματοποίηση Επιθέσεων	27
3.2. Μείωση του κόστους και των τεχνικών απαιτήσεων για τους επιτιθέμενους	29
3.3. Κλιμάκωση επιθέσεων (mass attacks)	32
3.4. Artificial Intelligence στη συλλογή πληροφοριών (reconnaissance)	36
3.5. Παραγωγή κακόβουλου λογισμικού	39
3.6. Social engineering με AI.....	42
4. ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΩΝ ΒΑΣΕΙ ΠΟΡΩΝ ΚΑΙ ΠΟΛΥΠΛΟΚΟΤΗΤΑΣ	48
4.1. Επιθέσεις χαμηλού κόστους vs επιθέσεις υψηλής πολυπλοκότητας (Low-cost vs high-complexity attacks)	48

4.1.1. Επιθέσεις χαμηλού κόστους (Low-cost attacks).....	48
4.1.2. Επιθέσεις υψηλής πολυπλοκότητας (High-complexity attacks).....	49
4.2. Ρόλος υπολογιστικών πόρων και πώς η διαθεσιμότητα πόρων επηρεάζει τη δυναμική του επιτιθέμενου	52
4.3. Συγκριτική ανάλυση πριν και μετά την τεχνητή νοημοσύνη.....	57
4.4. Αυτοματοποιημένες επιθέσεις με τεχνητή νοημοσύνη και ο ρόλος του Cloud Computing (Cloud & AI-as-a-Service).....	62
5. ΕΝΔΕΙΚΤΙΚΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΕ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΙΣ.....	67
5.1. AI-driven phishing campaigns	67
5.2. Automated malware generation	69
5.3. Επιθέσεις με χρήση automation tools	71
6. Η ΧΡΗΣΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΟ ΚΥΒΕΡΝΟΕΓΚΛΗΜΑ	75
6.1. AI-as-a-Service	75
6.2. Dark web tools	79
6.3 Οικονομία κυβερνοεγκλήματος.....	84
6.4. Προκλήσεις προς αρχές επιβολής του νόμου	89
7. Ο ΡΟΛΟΣ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ ΚΥΒΕΡΝΟΑΜΥΝΑ.....	95
7.1. Χρήση αλγορίθμων Τεχνητής Νοημοσύνης για την ανίχνευση επιθέσεων.....	95
7.2. Intrusion Detection Systems (IDS)	99
7.3. Αυτοματοποιημένες στρατηγικές απόκρισης σε κυβερνοεπιθέσεις	104
7.4. Attack AI vs Defence AI.....	109
8. ΝΟΜΙΚΕΣ ΚΑΙ ΗΘΙΚΕΣ ΔΙΑΣΤΑΣΕΙΣ ΤΗΣ ΧΡΗΣΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΕ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΙΣ	116
8.1 Ηθικά διλήμματα της χρήσης Τεχνητής Νοημοσύνης σε κυβερνοεπιθέσεις	116
8.2 Νομοθετικό πλαίσιο για την Τεχνητή Νοημοσύνη και την κυβερνοασφάλεια	121
8.3. Διακρατικές διαφορές και παγκόσμιες στρατηγικές πρόληψης	128
9. ΟΙΚΟΝΟΜΙΚΕΣ ΚΑΙ ΚΟΙΝΩΝΙΚΕΣ ΕΠΙΠΤΩΣΕΙΣ ΤΩΝ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΩΝ ΜΕ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ	135
9.1. Οικονομικές απώλειες από AI-driven κυβερνοεπιθέσεις	135
9.2. Κοινωνικές επιπτώσεις (σε εταιρείες, κυβερνήσεις και άτομα)	140
9.3. Αλλαγές στη στρατηγική κυβερνοασφάλειας λόγω AI	145
10. ΜΕΛΛΟΝΤΙΚΕΣ ΤΑΣΕΙΣ ΚΑΙ ΠΡΟΚΛΗΣΕΙΣ.....	149
10.1. Εξέλιξη αλγορίθμων και νέες απειλές	149
10.1.1. Αυτόνομες επιθετικές ροές	149
10.1.2. Ταχύτερη εκμετάλλευση ευπαθειών	150
10.1.3. Επιθέσεις εναντίον AI συστημάτων	150
10.1.4. Επιθέσεις adversarial machine learning.....	150
10.1.5. AI-powered social engineering επιθέσεις	151

10.2. Χρήση Τεχνητής Νοημοσύνης στην κυβερνοάμυνα	151
10.2.1. Security Operations Centers	152
10.2.2. Threat intelligence	152
10.3. Πιθανές εξελίξεις στο ρυθμιστικό πλαίσιο και στην ασφάλεια πληροφοριών	153
11. Συμπεράσματα και προτάσεις.....	155
11.1 Απαντήσεις στα ερευνητικά ερωτήματα.....	155
11.2 Σύνοψη βασικών ευρημάτων της έρευνας.....	156
11.3 Προτάσεις για μελλοντική έρευνα	157
12. ΒΙΒΛΙΟΓΡΑΦΙΑ	159
13. ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ – ΞΕΝΟΓΛΩΣΣΟΙ ΟΡΟΙ	185

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία εξετάζει την επίδραση των αλγορίθμων τεχνητής νοημοσύνης στην πραγματοποίηση κυβερνοεπιθέσεων, με έμφαση στον τρόπο με τον οποίο οι σύγχρονες τεχνολογίες Machine Learning, Deep Learning και Generative AI μεταβάλλουν τη φύση, το κόστος, την πολυπλοκότητα και την κλίμακα των επιθέσεων. Αρχικά, παρουσιάζεται το θεωρητικό υπόβαθρο της τεχνητής νοημοσύνης και των κυβερνοεπιθέσεων, ώστε να αποσαφηνιστούν οι βασικές τεχνολογικές και επιχειρησιακές έννοιες που στηρίζουν την ανάλυση.

Στη συνέχεια, η εργασία αναλύει την AI ως εργαλείο ενίσχυσης επιθετικών ενεργειών. Εξετάζονται η αυτοματοποίηση επιθέσεων, η μείωση των τεχνικών απαιτήσεων για τους επιτιθέμενους, η δυνατότητα μαζικής κλιμάκωσης, η χρήση AI στη συλλογή πληροφοριών, η παραγωγή κακόβουλου λογισμικού και η ενίσχυση τεχνικών social engineering. Ιδιαίτερη έμφαση δίνεται στο γεγονός ότι η AI δεν δημιουργεί απαραίτητα εξ ολοκλήρου νέες μορφές κυβερνοεπιθέσεων, αλλά ενισχύει ήδη υπάρχουσες τεχνικές, καθιστώντας τις ταχύτερες, πιο πειστικές, πιο προσαρμοστικές και πιο προσβάσιμες.

Παράλληλα, η εργασία εξετάζει την κατηγοριοποίηση των κυβερνοεπιθέσεων με βάση το κόστος, τους απαιτούμενους πόρους και την πολυπλοκότητα, καθώς και τη διαφοροποίηση του τοπίου πριν και μετά την ενσωμάτωση της τεχνητής νοημοσύνης. Μέσα από case studies αναδεικνύονται πρακτικά παραδείγματα AI-driven phishing, αυτοματοποιημένης παραγωγής κακόβουλου λογισμικού και χρήσης εργαλείων αυτοματοποίησης σε επιθετικά σενάρια.

Επιπλέον, αναλύεται η χρήση Τεχνητής Νοημοσύνης στο κυβερνοέγκλημα, δίνοντας έμφαση στα dark web tools, στην οικονομία του Cybercrime-as-a-Service και στις προκλήσεις που δημιουργούνται για τις αρχές επιβολής του νόμου. Από την άλλη πλευρά, παρουσιάζεται η αμυντική αξιοποίηση της Τεχνητής Νοημοσύνης, μέσω της χρήσης της σε συστήματα ανίχνευσης, Intrusion Detection Systems (IDS), αυτοματοποιημένες στρατηγικές απόκρισης και σύγχρονες πρακτικές κυβερνοάμυνας. Η εργασία αναδεικνύει την AI ως τεχνολογία dual-use, καθώς μπορεί να ενισχύσει τόσο τους επιτιθέμενους όσο και τους αμυνόμενους.

Τέλος, εξετάζονται οι ηθικές, νομικές, κοινωνικές και οικονομικές επιπτώσεις των AI-enabled cyberattacks, καθώς και οι μελλοντικές τάσεις και προκλήσεις. Η εργασία καταλήγει ότι η

αποτελεσματική αντιμετώπιση των κινδύνων που συνδέονται με την τεχνητή νοημοσύνη στην κυβερνοασφάλεια απαιτεί συνδυασμό τεχνικής προστασίας, οργανωτικής ετοιμότητας, θεσμικής προσαρμογής, διεθνούς συνεργασίας, ηθικής ευθύνης και συνεχούς εκπαίδευσης.

1. ΕΙΣΑΓΩΓΗ

1.1 Παρουσίαση του θέματος και της σημασίας του

Η ταχεία εξέλιξη της τεχνητής νοημοσύνης έχει επιφέρει σημαντικές μεταβολές στο σύγχρονο ψηφιακό περιβάλλον, επηρεάζοντας κρίσιμους τομείς της τεχνολογικής, οικονομικής και κοινωνικής δραστηριότητας. Η αξιοποίηση αλγορίθμων μηχανικής μάθησης και βαθιάς μάθησης έχει συμβάλει στην ανάπτυξη ευφυνών συστημάτων, ικανών να επεξεργάζονται μεγάλους όγκους δεδομένων, να αναγνωρίζουν σύνθετα πρότυπα και να υποστηρίζουν αυτοματοποιημένες διαδικασίες λήψης αποφάσεων. Παρά τις θετικές δυνατότητες που προσφέρει η τεχνητή νοημοσύνη, η ευρεία διάδοσή της δημιουργεί παράλληλα νέες προκλήσεις στο πεδίο της κυβερνοασφάλειας.

Η σύνδεση της τεχνητής νοημοσύνης με την πραγματοποίηση κυβερνοεπιθέσεων αποτελεί ένα ιδιαίτερα κρίσιμο ζήτημα, καθώς μεταβάλλει τον τρόπο με τον οποίο σχεδιάζονται, εκτελούνται και εξελίσσονται οι σύγχρονες επιθέσεις. Οι κυβερνοεπιθέσεις πλέον έχουν αποκτήσει χαρακτηριστικά αυτοματοποίησης και προσαρμοστικότητας. Μέσω της χρήσης αλγορίθμων τεχνητής νοημοσύνης, ένας επιτιθέμενος μπορεί να επιταχύνει διαδικασίες όπως η συλλογή πληροφοριών, η ανίχνευση ευπαθειών, η επιλογή στόχων και η προσαρμογή της επίθεσης στις εκάστοτε συνθήκες του πληροφοριακού συστήματος.

Η σημασία του θέματος ενισχύεται από το γεγονός ότι η τεχνητή νοημοσύνη μπορεί να λειτουργήσει με διττό τρόπο στο πεδίο της κυβερνοασφάλειας. Από τη μία πλευρά, αξιοποιείται για την ενίσχυση της άμυνας, την ανίχνευση απειλών, την πρόβλεψη κακόβουλων ενεργειών και την αυτοματοποιημένη απόκριση σε περιστατικά ασφαλείας. Από την άλλη πλευρά, οι ίδιες τεχνολογικές δυνατότητες μπορούν να χρησιμοποιηθούν από κακόβουλους δρώντες για την αύξηση της αποτελεσματικότητας, της ταχύτητας και της κλίμακας των επιθέσεων. Το γεγονός αυτό καθιστά την τεχνητή νοημοσύνη και εργαλείο τεχνολογικής προόδου και παράγοντα που διαμορφώνει το σύγχρονο τοπίο απειλών [1], [2].

Οι κυβερνοεπιθέσεις που αξιοποιούν τεχνητή νοημοσύνη μπορούν να πλήξουν οργανισμούς, επιχειρήσεις, κρατικές υπηρεσίες, κρίσιμες υποδομές και φυσικά πρόσωπα, προκαλώντας οικονομικές απώλειες, διαρροή ευαίσθητων δεδομένων, διατάραξη υπηρεσιών και απώλεια

εμπιστοσύνης στα ψηφιακά συστήματα. Παράλληλα, δημιουργούνται ζητήματα σχετικά με την ευθύνη χρήσης της τεχνητής νοημοσύνης, την ανάγκη ρύθμισής τους και την αποτελεσματικότητα των υπαρχόντων μηχανισμών προστασίας.

Η παρούσα διπλωματική εργασία εξετάζει την επίδραση των αλγορίθμων τεχνητής νοημοσύνης στην πραγματοποίηση κυβερνοεπιθέσεων. Η ανάλυση εστιάζει στον τρόπο με τον οποίο οι τεχνολογίες αυτές επηρεάζουν το κόστος, την πολυπλοκότητα, την κλίμακα, την αυτοματοποίηση και τη δυναμική των επιθέσεων, καθώς και στις ευρύτερες συνέπειες που προκύπτουν για την κυβερνοασφάλεια, την κοινωνία και το θεσμικό πλαίσιο προστασίας των πληροφοριακών συστημάτων.

1.2 Σημασία της έρευνας

Η σημασία της παρούσας έρευνας απορρέει από τη ραγδαία μεταβολή του περιβάλλοντος κυβερνοαπειλών και από την αυξανόμενη δυνατότητα αξιοποίησης της τεχνητής νοημοσύνης από κακόβουλους δρώντες. Η ενσωμάτωση αλγορίθμων τεχνητής νοημοσύνης σε επιθετικές διαδικασίες μεταβάλλει ουσιώδεις παραμέτρους των κυβερνοεπιθέσεων, καθώς επηρεάζει την ταχύτητα εκτέλεσης, τον βαθμό αυτοματοποίησης, την ακρίβεια επιλογής στόχων και τη δυνατότητα προσαρμογής στις αμυντικές αντιδράσεις ενός συστήματος [2], [3].

Ιδιαίτερη σημασία παρουσιάζει το γεγονός ότι η τεχνητή νοημοσύνη μπορεί να μειώσει τον χρόνο και την προσπάθεια που απαιτούνται για τον σχεδιασμό και την εκτέλεση μιας επίθεσης. Διαδικασίες οι οποίες παραδοσιακά απαιτούσαν υψηλή τεχνική εξειδίκευση και σημαντική ανθρώπινη παρέμβαση δύνανται πλέον να αυτοματοποιηθούν σε μεγάλο βαθμό. Η εξέλιξη αυτή ενδέχεται να διευρύνει το φάσμα των επιτιθέμενων, καθώς τα εργαλεία τεχνητής νοημοσύνης μπορούν να καταστήσουν απαιτητικές τεχνικές περισσότερο προσβάσιμες ακόμη και σε άτομα με μικρή τεχνική κατάρτιση.

Παράλληλα, η τεχνητή νοημοσύνη αυξάνει την κλίμακα και την πολυπλοκότητα των επιθέσεων. Μέσω αυτοματοποιημένων μηχανισμών ανάλυσης και προσαρμογής, οι επιθέσεις μπορούν να

εκτελούνται σε μεγαλύτερο εύρος στόχων, να εξατομικεύονται με ακρίβεια και να είναι δυσκολότερα ανιχνεύσιμες από συμβατικά συστήματα προστασίας.

Οι επιθέσεις με χρήση τεχνητής νοημοσύνης δεν επηρεάζουν μόνο την τεχνική ακεραιότητα των πληροφοριακών συστημάτων, αλλά έχουν ευρύτερες συνέπειες στην οικονομική λειτουργία οργανισμών, στην προστασία προσωπικών δεδομένων και στην εμπιστοσύνη των πολιτών προς τις νέες ψηφιακές τεχνολογίες.

1.3. Σκοπός και ερευνητικά ερωτήματα

Η εργασία επιδιώκει να αναλύσει τον τρόπο με τον οποίο οι τεχνικές machine learning και deep learning επηρεάζουν τη δυναμική της επιτιθέμενης πλευράς, μεταβάλλοντας βασικές παραμέτρους όπως το κόστος υλοποίησης, το επίπεδο απαιτούμενων τεχνικών γνώσεων, την πολυπλοκότητα, την κλίμακα και την αποτελεσματικότητα των επιθέσεων.

Στο πλαίσιο αυτό, η εργασία εξετάζει ένα κεντρικό ερευνητικό πρόβλημα, το οποίο αφορά τον βαθμό και τον τρόπο με τον οποίο η τεχνητή νοημοσύνη αναδιαμορφώνει το σύγχρονο περιβάλλον κυβερνοεπιθέσεων. Το πρόβλημα αυτό αναλύεται μέσα από επιμέρους θεματικές διαστάσεις, οι οποίες περιλαμβάνουν την τεχνολογική (ρόλος αλγορίθμων, αυτοματοποίηση, machine learning), επιχειρησιακή (κόστος, πολυπλοκότητα, κλιμάκωση επιθέσεων), αμυντική (κυβερνοάμυνα και χρήση AI για ανίχνευση), εγκληματολογική (κυβερνοέγκλημα, Dark Web, AI-as-a-Service), νομική και ηθική (κανονιστικά πλαίσια, ευθύνη, GDPR, AI Act), κοινωνικοοικονομική (επιπτώσεις σε οργανισμούς και κοινωνία) και μελλοντική διάσταση του φαινομένου (εξέλιξη απειλών και τεχνολογιών). Με τον τρόπο αυτό, η μελέτη δεν περιορίζεται αποκλειστικά στην τεχνική πλευρά των επιθέσεων, αλλά επιχειρεί να αποδώσει μια ολοκληρωμένη εικόνα των συνεπειών που επιφέρει η χρήση τεχνητής νοημοσύνης στο πεδίο της κυβερνοασφάλειας.

Ειδικότερα, η παρούσα εργασία επιχειρεί να απαντήσει στα ακόλουθα ερευνητικά ερωτήματα:

1. Πώς επηρεάζει η χρήση αλγορίθμων τεχνητής νοημοσύνης το κόστος σχεδιασμού και εκτέλεσης κυβερνοεπιθέσεων;

2. Σε ποιο βαθμό η τεχνητή νοημοσύνη συμβάλλει στην αύξηση της πολυπλοκότητας και της αποτελεσματικότητας των κυβερνοεπιθέσεων;
3. Κατά πόσο η αξιοποίηση εργαλείων τεχνητής νοημοσύνης μειώνει το απαιτούμενο επίπεδο τεχνικών γνώσεων για την υλοποίηση επιθέσεων;
4. Με ποιον τρόπο επηρεάζεται η κλίμακα των κυβερνοεπιθέσεων και η δυνατότητα μαζικής εκτέλεσής τους;
5. Πώς μεταβάλλεται η ισορροπία μεταξύ επιτιθέμενων και αμυνόμενων στο σύγχρονο περιβάλλον κυβερνοασφάλειας; [1], [4]

1.4. Στόχοι της έρευνας

Οι στόχοι της έρευνας διαμορφώνονται με τρόπο που να εξυπηρετεί την ολοκληρωμένη κατανόηση του φαινομένου και να συνδέεται άμεσα με τα ερευνητικά ερωτήματα που έχουν τεθεί.

Πρωταρχικός στόχος της εργασίας είναι η ανάλυση του ρόλου της τεχνητής νοημοσύνης στην υλοποίηση κυβερνοεπιθέσεων.

Δεύτερος βασικός στόχος είναι η κατηγοριοποίηση των κυβερνοεπιθέσεων με βάση το απαιτούμενο κόστος υλοποίησης, την τεχνική πολυπλοκότητα, το βαθμό αυτοματοποίησης και τους υπολογιστικούς πόρους που απαιτούνται. Έτσι αναδεικνύονται οι διαφοροποιήσεις ανάμεσα σε παραδοσιακές μορφές κυβερνοεπιθέσεων και σε επιθέσεις που αξιοποιούν τεχνητή νοημοσύνη. Με τον τρόπο αυτό διαμορφώνεται ένα ξεκάθαρο πλαίσιο κατανόησης των σύγχρονων απειλών, το οποίο μπορεί να υποστηρίξει τόσο την επιστημονική ανάλυση όσο και την πρακτική αξιολόγηση των κινδύνων.

Τρίτος στόχος της έρευνας είναι η ανάδειξη των μεταβολών που επιφέρει η τεχνητή νοημοσύνη στα χαρακτηριστικά των κυβερνοεπιθέσεων. Ειδικότερα, εξετάζεται κατά πόσο η χρήση ευφών αλγορίθμων επηρεάζει την ταχύτητα εκτέλεσης, την κλίμακα διάδοσης, την ακρίβεια στόχευσης, την ανθεκτικότητα της επίθεσης εναντίον αμυντικών μηχανισμών και τη δυνατότητα προσαρμογής της επίθεσης σε πραγματικό χρόνο.

Επιπλέον, η εργασία αναλύει το κενό μεταξύ επιτιθέμενων και αμυνόμενων. Η τεχνητή νοημοσύνη παρέχει σημαντικές δυνατότητες και στις δύο πλευρές, γεγονός που καθιστά αναγκαία την εξέταση του τρόπου με τον οποίο οι επιθετικές και αμυντικές εφαρμογές της αλληλεπιδρούν. Στο πλαίσιο αυτό, η έρευνα επιδιώκει να αναδείξει κατά πόσο η χρήση τεχνητής νοημοσύνης ενισχύει τη δυναμική των επιτιθέμενων ή, αντιθέτως, τη δυναμική των αμυνόμενων.

1.5. Η μεθοδολογία της εργασίας

Η ανάλυση που ακολουθεί βασίζεται σε ποιοτική ερευνητική προσέγγιση, η οποία εστιάζει στη συστηματική μελέτη και ανάλυση πληροφοριών σχετικά με την επίδραση της τεχνητής νοημοσύνης στις κυβερνοεπιθέσεις.

Ως βασική μέθοδος συλλογής δεδομένων χρησιμοποιείται η εκτενής βιβλιογραφική ανασκόπηση, η οποία περιλαμβάνει επιστημονικά άρθρα και συγγράμματα, ακαδημαϊκές δημοσιεύσεις, τεχνικές αναφορές και μελέτες σχετικές με την τεχνητή νοημοσύνη, την κυβερνοασφάλεια και τις σύγχρονες μορφές κυβερνοεπιθέσεων. Μέσω της ανασκόπησης αυτής επιδιώκεται η καταγραφή των βασικών εννοιών, η αποτύπωση των κυρίαρχων τάσεων και η κατανόηση των μηχανισμών μέσω των οποίων οι αλγόριθμοι τεχνητής νοημοσύνης αξιοποιούνται σε επιθετικά πλαίσια.

Παράλληλα, η εργασία αξιοποιεί συγκριτική ανάλυση, προκειμένου να εξετάσει τις διαφοροποιήσεις που προκύπτουν μεταξύ παραδοσιακών κυβερνοεπιθέσεων και επιθέσεων που ενσωματώνουν τεχνολογίες τεχνητής νοημοσύνης. Η προσέγγιση αυτή επιτρέπει την ανάδειξη μεταβολών σε κρίσιμες παραμέτρους, όπως το κόστος υλοποίησης, η τεχνική πολυπλοκότητα, η ταχύτητα εκτέλεσης, η δυνατότητα κλιμάκωσης και ο βαθμός αυτοματοποίησης των επιθέσεων.

Επιπρόσθετα, γίνεται μεθοδική εξέταση χαρακτηριστικών παραδειγμάτων κυβερνοεπιθέσεων που σχετίζονται με τη χρήση τεχνητής νοημοσύνης. Η ανάλυση των περιπτώσεων αυτών συμβάλλει στην πρακτική κατανόηση της εφαρμογής των θεωρητικών εννοιών, καθώς και στην αποτύπωση των πραγματικών επιπτώσεων που έχουν οι εν λόγω τεχνολογίες στο πεδίο της κυβερνοασφάλειας [2], [3], [4].

1.6. Η δομή της εργασίας

Η παρούσα διπλωματική εργασία οργανώνεται σε έντεκα κύρια κεφάλαια, ενώ συνοδεύεται από αναλυτική βιβλιογραφία και πίνακα συντομογραφιών - ξενόγλωσσων όρων, κεφάλαια υπ' αριθμών δώδεκα και δεκατρία αντίστοιχα. Τα κεφάλαια διαρθρώνονται με τρόπο που επιτρέπει τη σταδιακή ανάπτυξη και εις βάθος ανάλυση του εξεταζόμενου αντικειμένου.

Στο πρώτο κεφάλαιο παρουσιάζεται η εισαγωγή της εργασίας, όπου περιγράφεται το αντικείμενο μελέτης, αναδεικνύεται η σημασία του, διατυπώνονται ο σκοπός και τα ερευνητικά ερωτήματα, καθορίζονται οι στόχοι της έρευνας και παρουσιάζεται η μεθοδολογική προσέγγιση που ακολουθείται.

Στο δεύτερο κεφάλαιο αναπτύσσεται το θεωρητικό υπόβαθρο της εργασίας, το οποίο περιλαμβάνει βασικές έννοιες της τεχνητής νοημοσύνης, ανάλυση των κυριότερων αλγορίθμων μηχανικής και βαθιάς μάθησης (machine και deep learning), καθώς και εισαγωγή στις έννοιες των κυβερνοεπιθέσεων και των ευπαθειών πληροφοριακών συστημάτων.

Στο τρίτο κεφάλαιο εξετάζεται η τεχνητή νοημοσύνη ως εργαλείο κυβερνοεπιθέσεων, με έμφαση σε ζητήματα όπως η αυτοματοποίηση επιθέσεων, η μείωση του απαιτούμενου κόστους και τεχνικής εξειδίκευσης, η δυνατότητα κλιμάκωσης, η αξιοποίηση τεχνικών συλλογής πληροφοριών και η παραγωγή κακόβουλου λογισμικού.

Στο τέταρτο κεφάλαιο πραγματοποιείται κατηγοριοποίηση των κυβερνοεπιθέσεων με βάση το κόστος, την πολυπλοκότητα και τους απαιτούμενους υπολογιστικούς πόρους, ενώ παράλληλα εξετάζεται η επίδραση της τεχνητής νοημοσύνης στη διαμόρφωση της δυναμικής των επιτιθέμενων και αναλύεται ο ρόλος του υπολογιστικού νέφους.

Στο πέμπτο κεφάλαιο παρουσιάζονται μελέτες περίπτωσης κυβερνοεπιθέσεων που αξιοποιούν τεχνητή νοημοσύνη, με στόχο την ανάδειξη πρακτικών εφαρμογών και την κατανόηση των μεταβολών που επιφέρει η χρήση των τεχνολογιών αυτών.

Στο έκτο κεφάλαιο εξετάζεται η χρήση της τεχνητής νοημοσύνης στο πλαίσιο του κυβερνοεγκλήματος, συμπεριλαμβανομένων των υπηρεσιών τεχνητής νοημοσύνης ως υπηρεσία, των εργαλείων του dark web και της οικονομίας του κυβερνοεγκλήματος.

Στο έβδομο κεφάλαιο αναλύεται ο ρόλος της τεχνητής νοημοσύνης στην κυβερνοάμυνα, με έμφαση σε τεχνικές ανίχνευσης επιθέσεων, συστήματα ανίχνευσης εισβολών και αυτοματοποιημένες στρατηγικές απόκρισης.

Στο όγδοο κεφάλαιο αναλύονται οι νομικές και ηθικές διαστάσεις της χρήσης τεχνητής νοημοσύνης σε κυβερνοεπιθέσεις, με αναφορά σε κανονιστικά πλαίσια και διεθνείς προσεγγίσεις.

Στο ένατο κεφάλαιο παρουσιάζονται οι οικονομικές και κοινωνικές επιπτώσεις των κυβερνοεπιθέσεων που αξιοποιούν τεχνητή νοημοσύνη, τόσο σε επίπεδο οργανισμών όσο και σε επίπεδο κοινωνίας.

Στο δέκατο κεφάλαιο εξετάζονται οι μελλοντικές τάσεις και προκλήσεις που σχετίζονται με την εξέλιξη της τεχνητής νοημοσύνης και τη χρήση της στο πεδίο της κυβερνοασφάλειας.

Τέλος, στο ενδέκατο κεφάλαιο διατυπώνονται τα συμπεράσματα της έρευνας, συνοψίζονται τα βασικά ευρήματα και προτείνονται κατευθύνσεις για μελλοντική μελέτη του αντικειμένου.

2. ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1. Βασικές αρχές της Τεχνητής Νοημοσύνης

Η τεχνητή νοημοσύνη αποτελεί έναν από τους σημαντικότερους και ταχύτερα εξελισσόμενους κλάδους της επιστήμης των υπολογιστών, με αντικείμενο τη δημιουργία συστημάτων που μπορούν να εκτελούν λειτουργίες οι οποίες παραδοσιακά θεωρούνταν πως απαιτούσαν ανθρώπινη σκέψη και ευφυΐα. Τέτοιες λειτουργίες περιλαμβάνουν την επεξεργασία πληροφοριών, τη λήψη αποφάσεων, την κατανόηση φυσικής γλώσσας, την ανάλυση εικόνων και την προσαρμογή σε νέα δεδομένα. Η τεχνητή νοημοσύνη δεν συνιστά μία ενιαία τεχνική, αλλά ένα ευρύτερο επιστημονικό και τεχνολογικό πεδίο, το οποίο περιλαμβάνει διαφορετικές μεθόδους, αλγόριθμους και υπολογιστικά μοντέλα [6].

Σε γενικό επίπεδο, η τεχνητή νοημοσύνη μπορεί να οριστεί ως η ικανότητα ενός υπολογιστικού συστήματος να επεξεργάζεται δεδομένα, να εξάγει συμπεράσματα και να εκτελεί ενέργειες με τρόπο παρόμοιο με την ανθρώπινη νοημοσύνη. Η λειτουργία αυτή μπορεί να βασίζεται είτε σε προκαθορισμένους κανόνες και προγραμματισμένες διαδικασίες είτε σε μεθόδους μάθησης από δεδομένα [12].

Κεντρική έννοια στο πεδίο της τεχνητής νοημοσύνης αποτελεί η μηχανική μάθηση. Η μηχανική μάθηση (machine learning) αναφέρεται σε μεθόδους μέσω των οποίων ένας αλγόριθμος δεν ακολουθεί αποκλειστικά έναν εκ των προτέρων καθορισμένο τρόπο επίλυσης ενός προβλήματος, αλλά εκπαιδεύεται με βάση δεδομένα που του δίνονται. Με τον τρόπο αυτό, το σύστημα μπορεί να βελτιώνει την απόδοσή του, χωρίς να απαιτείται προκαθορισμένη μεθοδολογία επίλυσης [8], [12].

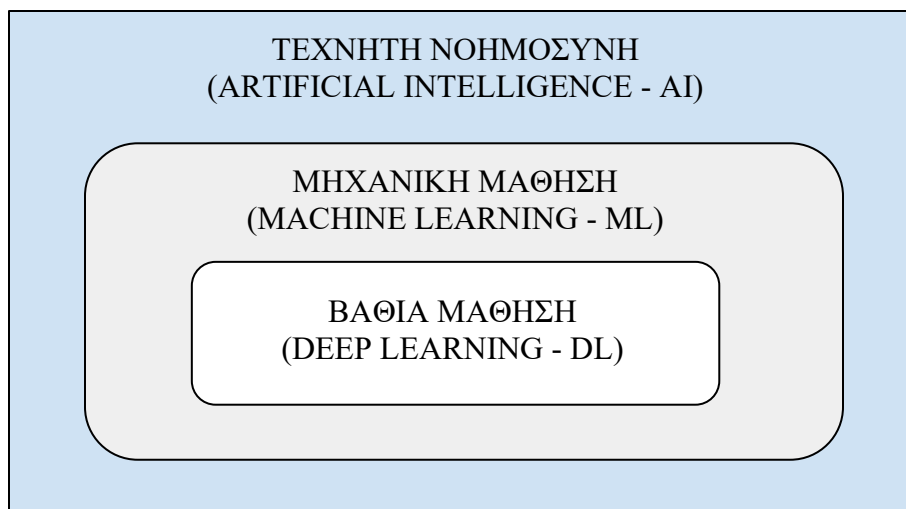
Η μηχανική μάθηση διαφέρει από τις παραδοσιακές προγραμματιστικές προσεγγίσεις, διότι το αποτέλεσμα δεν προκύπτει μόνο από προκαθορισμένους κανόνες, αλλά από τη σταδιακή προσαρμογή του μοντέλου στα δεδομένα εκπαίδευσης. Κατά τη διαδικασία αυτή, το μοντέλο λαμβάνει δεδομένα εισόδου, τα επεξεργάζεται και τροποποιεί εσωτερικές παραμέτρους, ώστε να επιτυγχάνει καλύτερη απόδοση σε συγκεκριμένη εργασία. [10], [12].

Η μηχανική μάθηση μπορεί να εφαρμοστεί σε διαφορετικά είδη προβλημάτων, όπως η ταξινόμηση, η πρόβλεψη, η αναγνώριση ανωμαλιών, η ομαδοποίηση δεδομένων και η εξαγωγή συσχετίσεων. Ανάλογα με το είδος των διαθέσιμων δεδομένων και τον στόχο της ανάλυσης, οι μέθοδοι μηχανικής μάθησης μπορούν να οργανωθούν σε κατηγορίες, όπως η επιβλεπόμενη μάθηση, η μη επιβλεπόμενη μάθηση και άλλες μορφές μάθησης που αξιοποιούνται σε πιο σύνθετα περιβάλλοντα. [8].

Ιδιαίτερη θέση στο πεδίο της μηχανικής μάθησης κατέχει η βαθιά μάθηση. Η βαθιά μάθηση (deep learning) αποτελεί υποπεδίο της μηχανικής μάθησης και βασίζεται κυρίως στη χρήση τεχνητών νευρωνικών δικτύων με πολλαπλά επίπεδα επεξεργασίας. Τα νευρωνικά δίκτυα αποτελούνται από διασυνδεδεμένους κόμβους, οι οποίοι επεξεργάζονται πληροφορίες και μεταβιβάζουν αποτελέσματα στα επόμενα επίπεδα του δικτύου [7], [12].

Η βασική ιδιαιτερότητα της βαθιάς μάθησης έγκειται στην ύπαρξη πολλών ενδιάμεσων επιπέδων, μέσω των οποίων το σύστημα μπορεί να εξάγει σταδιακά όλο και πιο σύνθετες αναπαραστάσεις των δεδομένων. Σε απλούστερα επίπεδα μπορεί να αναγνωρίζονται βασικά χαρακτηριστικά, ενώ σε ανώτερα επίπεδα το μοντέλο μπορεί να συσχετίζει τα χαρακτηριστικά αυτά και να διαμορφώνει πιο σύνθετες αναλύσεις. Η δυνατότητα αυτή καθιστά τη βαθιά μάθηση ιδιαίτερα κατάλληλη για πεδία όπως η αναγνώριση εικόνας, η επεξεργασία φυσικής γλώσσας, η αναγνώριση ομιλίας και η ανάλυση μεγάλων συνόλων δεδομένων [7], [14].

Η αυξανόμενη πολυπλοκότητα των συστημάτων τεχνητής νοημοσύνης δημιουργεί νέες απαιτήσεις ως προς την αξιοπιστία, τη διαφάνεια, την ασφάλεια και την αξιολόγησή τους. Η λειτουργία ενός μοντέλου εξαρτάται σε μεγάλο βαθμό από τα δεδομένα στα οποία εκπαιδεύεται, από τις παραμέτρους που χρησιμοποιούνται και από το περιβάλλον μέσα στο οποίο εφαρμόζεται. Για τον λόγο αυτό, η τεχνική αξιολόγηση των δυνατοτήτων και των περιορισμών των συστημάτων τεχνητής νοημοσύνης αποτελεί βασική προϋπόθεση για την ορθή και ασφαλή χρήση τους [9], [10].



Σχήμα 2.1. Ιεραρχική σχέση μεταξύ Τεχνητής Νοημοσύνης, Μηχανικής Μάθησης και Βαθιάς Μάθησης.

Η τεχνητή νοημοσύνη, η μηχανική μάθηση και η βαθιά μάθηση αποτελούν διακριτά αλλά αλληλένδετα επίπεδα του ίδιου τεχνολογικού πεδίου. Η τεχνητή νοημοσύνη συνιστά το ευρύτερο πλαίσιο, η μηχανική μάθηση αποτελεί βασικό υποπεδίο της και η βαθιά μάθηση εξειδικεύεται στη χρήση πολυεπίπεδων νευρωνικών δικτύων. Η κατανόηση αυτής της σχέσης αποτελεί απαραίτητη βάση για την ερμηνεία των δυνατοτήτων και των περιορισμών των σύγχρονων συστημάτων ΑΙ στην κυβερνοασφάλεια. [7], [8], [12].

2.2. Βασικοί αλγόριθμοι / Αλγόριθμοι μηχανικής μάθησης και βαθιάς μάθησης

Η ανάπτυξη της τεχνητής νοημοσύνης συνδέεται άμεσα με την εξέλιξη των αλγορίθμων μηχανικής μάθησης και βαθιάς μάθησης, οι οποίοι επιτρέπουν στα υπολογιστικά συστήματα να επεξεργάζονται δεδομένα, να εντοπίζουν πρότυπα και να βελτιώνουν την απόδοσή τους μέσω διαδικασιών εκπαίδευσης. Οι αλγόριθμοι αυτοί αποτελούν τον βασικό μηχανισμό λειτουργίας των σύγχρονων συστημάτων τεχνητής νοημοσύνης και εφαρμόζονται σε ένα ευρύ φάσμα προβλημάτων, όπως η ταξινόμηση δεδομένων, η πρόβλεψη συμπεριφορών, η αναγνώριση εικόνας, η επεξεργασία φυσικής γλώσσας και η ανίχνευση ανωμαλιών [8], [16].

Η λειτουργία ενός συστήματος μηχανικής μάθησης βασίζεται στη διαδικασία εκπαίδευσης ενός μοντέλου μέσω δεδομένων εισόδου. Κατά τη διαδικασία αυτή, το σύστημα αναλύει χαρακτηριστικά των δεδομένων, προσαρμόζει εσωτερικές παραμέτρους και επιχειρεί να ελαχιστοποιήσει το σφάλμα των προβλέψεών του. Η αποτελεσματικότητα του μοντέλου εξαρτάται από πολλούς παράγοντες, όπως η ποιότητα των δεδομένων, η αρχιτεκτονική του αλγορίθμου, η διαδικασία βελτιστοποίησης και οι υπολογιστικοί πόροι που διατίθενται για την εκπαίδευση [8], [13], [18].

2.2.1. Επιβλεπόμενη Μάθηση

Η επιβλεπόμενη μάθηση αποτελεί μία από τις βασικότερες κατηγορίες μηχανικής μάθησης. Στην προσέγγιση αυτή, το μοντέλο εκπαιδεύεται χρησιμοποιώντας δεδομένα που συνοδεύονται από γνωστές ετικέτες ή προκαθορισμένα αποτελέσματα. Στόχος του αλγορίθμου είναι να μάθει τη σχέση μεταξύ των δεδομένων εισόδου και των αντίστοιχων εξόδων, ώστε να μπορεί στη συνέχεια να πραγματοποιεί προβλέψεις ή ταξινομήσεις σε νέα δεδομένα [8], [15].

Η επιβλεπόμενη μάθηση χρησιμοποιείται ευρέως σε προβλήματα ταξινόμησης και παλινδρόμησης. Στα προβλήματα ταξινόμησης, το μοντέλο καλείται να κατατάξει δεδομένα σε συγκεκριμένες κατηγορίες, ενώ στα προβλήματα παλινδρόμησης επιχειρείται η πρόβλεψη συνεχών αριθμητικών τιμών. Η διαδικασία εκπαίδευσης βασίζεται στη συνεχή σύγκριση των προβλέψεων του μοντέλου με τα πραγματικά δεδομένα, μέσω μηχανισμών υπολογισμού σφάλματος και προσαρμογής των παραμέτρων του συστήματος [15], [16].

Η αποτελεσματικότητα της επιβλεπόμενης μάθησης εξαρτάται σε σημαντικό βαθμό από την ποιότητα και την αντιπροσωπευτικότητα των δεδομένων εκπαίδευσης. Η ύπαρξη ελλιπών δεδομένων μπορεί να επηρεάσει αρνητικά την ικανότητα του μοντέλου και να οδηγήσει σε λανθασμένες προβλέψεις [16].

2.2.2. Μη Επιβλεπόμενη Μάθηση

Η μη επιβλεπόμενη μάθηση διαφοροποιείται από την επιβλεπόμενη, καθώς το μοντέλο εκπαιδεύεται χωρίς τη χρήση προκαθορισμένων ετικετών ή γνωστών αποτελεσμάτων. Στην περίπτωση αυτή, ο αλγόριθμος επιχειρεί να ανακαλύψει δομές, σχέσεις και πρότυπα που υπάρχουν

στα δεδομένα, χωρίς να διαθέτει προηγούμενη γνώση σχετικά με την κατηγοριοποίησή τους [8], [15].

Μία από τις σημαντικότερες εφαρμογές της μη επιβλεπόμενης μάθησης είναι η ομαδοποίηση δεδομένων, κατά την οποία το σύστημα εντοπίζει ομάδες στοιχείων με παρόμοια χαρακτηριστικά. Επιπλέον, χρησιμοποιείται σε διαδικασίες ανίχνευσης ανωμαλιών, ανάλυσης συμπεριφοράς και εξαγωγής συσχετίσεων μεταξύ διαφορετικών μεταβλητών [15], [18].

Η μη επιβλεπόμενη μάθηση παρουσιάζει ιδιαίτερο ενδιαφέρον σε περιπτώσεις όπου μεγάλοι όγκοι δεδομένων δεν έχουν προηγουμένως ταξινομηθεί ή επισημανθεί. Ωστόσο, η ερμηνεία των αποτελεσμάτων είναι συχνά περισσότερο σύνθετη, καθώς το μοντέλο δεν καθοδηγείται από σαφώς ορισμένα επιθυμητά αποτελέσματα κατά τη διάρκεια της εκπαίδευσης [13], [17].

2.2.3. Νευρωνικά Δίκτυα

Τα τεχνητά νευρωνικά δίκτυα αποτελούν μία από τις σημαντικότερες τεχνολογικές βάσεις της σύγχρονης τεχνητής νοημοσύνης και της βαθιάς μάθησης. Η αρχιτεκτονική βασίζεται στη διασύνδεση μεγάλου αριθμού νευρώνων, οι οποίοι επεξεργάζονται και μεταδίδουν πληροφορίες [7], [13].

Ένα νευρωνικό δίκτυο αποτελείται συνήθως από επίπεδα εισόδου, κρυφά επίπεδα επεξεργασίας και επίπεδα εξόδου. Κάθε επίπεδο περιλαμβάνει κόμβους που δέχονται δεδομένα, εφαρμόζουν υπολογιστικές λειτουργίες και μεταβιβάζουν τα αποτελέσματα στα επόμενα επίπεδα. Μέσω της διαδικασίας εκπαίδευσης, το δίκτυο προσαρμόζει τα βάρη σύνδεσης μεταξύ των νευρώνων, ώστε να βελτιστοποιεί την απόδοσή του [7].

Η εκπαίδευση των νευρωνικών δικτύων πραγματοποιείται συνήθως μέσω αλγορίθμων οπισθοδιάδοσης σφάλματος (backpropagation), οι οποίοι επιτρέπουν τον υπολογισμό και τη σταδιακή ελαχιστοποίηση του σφάλματος πρόβλεψης. Η διαδικασία αυτή συνδυάζεται με τεχνικές βελτιστοποίησης, όπως η μέθοδος gradient descent, με στόχο την αποδοτικότερη προσαρμογή των παραμέτρων του μοντέλου [7].

Η ανάπτυξη βαθιών νευρωνικών δικτύων με πολλαπλά επίπεδα επεξεργασίας οδήγησε στη δημιουργία της βαθιάς μάθησης, η οποία επιτρέπει την εξαγωγή σύνθετων αναπαραστάσεων

δεδομένων και την επίτευξη υψηλών επιδόσεων σε απαιτητικές εργασίες. Στο πλαίσιο αυτό έχουν αναπτυχθεί εξειδικευμένες αρχιτεκτονικές, όπως τα συνελκτικά νευρωνικά δίκτυα (Convolutional Neural Networks – CNNs), τα αναδρομικά νευρωνικά δίκτυα (Recurrent Neural Networks – RNNs) και οι αρχιτεκτονικές μετασχηματιστών (transformers), οι οποίες χρησιμοποιούνται σε εφαρμογές αναγνώρισης εικόνας, επεξεργασίας φυσικής γλώσσας και ανάλυσης ακολουθιών δεδομένων [7].

Η αποτελεσματικότητα των νευρωνικών δικτύων συνδέεται άμεσα με τη διαθεσιμότητα μεγάλων συνόλων δεδομένων και αυξημένης υπολογιστικής ισχύος. Η εκπαίδευση πολύπλοκων μοντέλων απαιτεί σημαντικούς υπολογιστικούς πόρους, καθώς και εξειδικευμένες διαδικασίες διαχείρισης δεδομένων και βελτιστοποίησης. Παράλληλα, η αυξανόμενη πολυπλοκότητα των αρχιτεκτονικών αυτών δημιουργεί προκλήσεις που σχετίζονται με τη διαφάνεια, την αξιοπιστία και την ερμηνευσιμότητα των αποτελεσμάτων [16], [20].

Οι αλγόριθμοι μηχανικής μάθησης και βαθιάς μάθησης αποτελούν τον βασικό τεχνολογικό μηχανισμό των σύγχρονων συστημάτων ΑΙ. Η επιβλεπόμενη και η μη επιβλεπόμενη μάθηση, καθώς και τα νευρωνικά δίκτυα, επιτρέπουν την επεξεργασία δεδομένων, την αναγνώριση προτύπων και την υποστήριξη σύνθετων αποφάσεων. Η κατανόηση της λειτουργίας τους είναι αναγκαία για την αξιολόγηση τόσο των αμυντικών εφαρμογών τους όσο και των κινδύνων κακόβουλης αξιοποίησής τους. [8], [7], [16].

2.3. Βασικές έννοιες κυβερνοεπιθέσεων / Ευπάθειες συστημάτων και κυβερνοεπιθέσεις

Η συνεχής ανάπτυξη των ψηφιακών τεχνολογιών και η αυξανόμενη διασύνδεση πληροφοριακών συστημάτων έχουν ενισχύσει σημαντικά τη σημασία της κυβερνοασφάλειας στο σύγχρονο τεχνολογικό περιβάλλον. Παράλληλα, η εξάρτηση οργανισμών, επιχειρήσεων και κρατικών φορέων από ψηφιακές υποδομές έχει αυξήσει την έκθεσή τους σε κυβερνοαπειλές, οι οποίες εξελίσσονται διαρκώς ως προς την πολυπλοκότητα, την κλίμακα και τις τεχνικές υλοποίησής τους [21], [27].

Οι κυβερνοεπιθέσεις αποτελούν κακόβουλες ενέργειες που στοχεύουν στην παραβίαση της εμπιστευτικότητας, της ακεραιότητας ή της διαθεσιμότητας πληροφοριών και πληροφοριακών συστημάτων. Οι επιθέσεις αυτές μπορούν να πραγματοποιηθούν με διαφορετικές τεχνικές και να στοχεύουν τόσο τεχνικές ευπάθειες όσο και ανθρώπινους παράγοντες. Σύμφωνα με σύγχρονες αναφορές κυβερνοασφάλειας, οι κυβερνοεπιθέσεις συνδέονται συχνά με οικονομικά κίνητρα, κλοπή δεδομένων, εκβιασμό, κατασκοπεία ή διατάραξη λειτουργίας κρίσιμων υπηρεσιών [22], [27], [30].

Κεντρική έννοια στο πεδίο της κυβερνοασφάλειας αποτελεί η έννοια της ευπάθειας. Ως ευπάθεια ορίζεται οποιαδήποτε αδυναμία ή ελάττωμα σε λογισμικό, υλικό, δίκτυα, διαδικασίες ή ανθρώπινες πρακτικές, το οποίο μπορεί να αξιοποιηθεί από έναν επιτιθέμενο για την απόκτηση μη εξουσιοδοτημένης πρόσβασης ή για την εκτέλεση κακόβουλων ενεργειών. Οι ευπάθειες μπορεί να οφείλονται σε σφάλματα σχεδιασμού, ελλιπή ενημέρωση συστημάτων, λανθασμένες ρυθμίσεις ασφαλείας ή ανεπαρκή επίπεδα προστασίας [21], [25].

Η αξιοποίηση ευπαθειών είναι μία από τις συνηθέστερες τεχνικές αρχικής διείσδυσης σε πληροφοριακά συστήματα. Σύμφωνα με σύγχρονες αναφορές περιστατικών κυβερνοασφάλειας, μεγάλος αριθμός παραβιάσεων ξεκινά μέσω εκμετάλλευσης γνωστών αδυναμιών λογισμικού, ανεπαρκούς διαχείρισης διαπιστευτηρίων πρόσβασης ή επιθέσεων social engineering [23], [25], [31].

2.3.1. Malware

Το κακόβουλο λογισμικό (malware) αποτελεί μία από τις σημαντικότερες κατηγορίες κυβερνοαπειλών. Ο όρος χρησιμοποιείται για να περιγράψει λογισμικό το οποίο σχεδιάζεται με σκοπό την πραγματοποίηση μη εξουσιοδοτημένων ή επιβλαβών ενεργειών σε ένα πληροφοριακό σύστημα. Οι λειτουργίες του κακόβουλου λογισμικού μπορεί να περιλαμβάνουν την υποκλοπή δεδομένων, την παρακολούθηση χρηστών, την καταστροφή αρχείων, τη διακοπή λειτουργίας συστημάτων ή την παροχή απομακρυσμένης πρόσβασης σε επιτιθέμενους [21], [24].

Στις συνηθέστερες μορφές malware περιλαμβάνονται οι ιοί, τα worms, τα trojans, τα spyware, τα info stealers και το ransomware. Ιδιαίτερη αύξηση παρουσιάζουν τα τελευταία χρόνια οι επιθέσεις ransomware, κατά τις οποίες τα δεδομένα του θύματος κρυπτογραφούνται και απαιτείται

καταβολή χρηματικού ποσού για την αποκατάσταση της πρόσβασης. Οι επιθέσεις αυτές έχουν προκαλέσει σημαντικές οικονομικές και λειτουργικές επιπτώσεις σε επιχειρήσεις και οργανισμούς παγκοσμίως [21], [22], [29].

Η διάδοση κακόβουλου λογισμικού πραγματοποιείται μέσω διαφορετικών μηχανισμών, όπως κακόβουλα συνημμένα αρχεία ηλεκτρονικού ταχυδρομείου, παραβιασμένες ιστοσελίδες, εκμετάλλευση ευπαθειών λογισμικού ή κακόβουλες λήψεις αρχείων. Η συνεχής εξέλιξη των τεχνικών απόκρυψης και διανομής malware καθιστά ολοένα δυσκολότερη την έγκαιρη ανίχνευσή του [24], [31].

2.3.2. Phishing

Οι επιθέσεις phishing αποτελούν μορφή social engineering, κατά την οποία ο επιτιθέμενος επιχειρεί να εξαπατήσει το θύμα ώστε να αποκαλύψει ευαίσθητες πληροφορίες, όπως κωδικούς πρόσβασης, οικονομικά στοιχεία ή προσωπικά δεδομένα. Οι επιθέσεις αυτές πραγματοποιούνται συνήθως μέσω ηλεκτρονικού ταχυδρομείου, μηνυμάτων ή παραποιημένων ιστοσελίδων που προσομοιώνουν αξιόπιστους οργανισμούς ή υπηρεσίες [21], [23].

Το phishing συγκαταλέγεται μεταξύ των πιο διαδεδομένων μορφών κυβερνοεπιθέσεων, καθώς βασίζεται κυρίως στην εκμετάλλευση ανθρώπινων αδυναμιών και όχι αποκλειστικά τεχνικών ευπαθειών. Σύμφωνα με σύγχρονες εκθέσεις κυβερνοασφάλειας, σημαντικό ποσοστό παραβιάσεων ξεκινά από επιτυχημένες επιθέσεις phishing, οι οποίες οδηγούν σε υποκλοπή διαπιστευτηρίων πρόσβασης ή εγκατάσταση κακόβουλου λογισμικού [23], [24], [28].

2.3.3. Denial-of-Service Attacks

Οι επιθέσεις άρνησης υπηρεσίας (Denial-of-Service – DoS) αποσκοπούν στη διακοπή ή υποβάθμιση της λειτουργίας ενός πληροφοριακού συστήματος, δικτύου ή διαδικτυακής υπηρεσίας, μέσω υπερφόρτωσης των διαθέσιμων πόρων του. Όταν οι επιθέσεις αυτές πραγματοποιούνται από μεγάλο αριθμό συστημάτων ταυτόχρονα, αναφέρονται ως κατανεμημένες επιθέσεις άρνησης υπηρεσίας (Distributed Denial-of-Service – DDoS) [21], [26].

Οι επιθέσεις αυτές επιδιώκουν να καταστήσουν μία υπηρεσία μη διαθέσιμη στους νόμιμους χρήστες, προκαλώντας σημαντικές λειτουργικές και οικονομικές επιπτώσεις. Οι στόχοι μπορεί να περιλαμβάνουν δημόσιους οργανισμούς, επιχειρήσεις, χρηματοπιστωτικά ιδρύματα ή κρίσιμες ψηφιακές υποδομές [26], [27].

Η πραγματοποίηση επιθέσεων DDoS βασίζεται συχνά στη χρήση δικτύων παραβιασμένων συσκευών, γνωστών ως botnets, τα οποία αποστέλλουν ταυτόχρονα μεγάλο όγκο αιτημάτων προς τον στόχο. Η αυξημένη διαθεσιμότητα συνδεδεμένων συσκευών και υπηρεσιών νέφους έχει ενισχύσει τη δυνατότητα εκτέλεσης επιθέσεων μεγάλης κλίμακας [21], [31].

2.3.4. Σύγχρονες Τάσεις Κυβερνοεπιθέσεων

Το σύγχρονο περιβάλλον κυβερνοαπειλών χαρακτηρίζεται από αυτοματοποίηση και οργανωμένη εγκληματική δραστηριότητα. Σύμφωνα με διεθνείς αναφορές κυβερνοασφάλειας, παρατηρείται ανάπτυξη μοντέλων crime-as-a-service, μέσω των οποίων κακόβουλα εργαλεία και υπηρεσίες διατίθενται σε τρίτους έναντι οικονομικού ανταλλάγματος [22], [30].

Παράλληλα, οι επιθέσεις γίνονται περισσότερο στοχευμένες, αξιοποιώντας συνδυαστικά τεχνικές social engineering, εκμετάλλευσης ευπαθειών, υποκλοπής διαπιστευτηρίων και κακόβουλου λογισμικού. Η εξέλιξη αυτή ενισχύεται από τη χρήση αυτοματοποιημένων εργαλείων και προηγμένων τεχνολογιών, οι οποίες επιτρέπουν ταχύτερη ανάλυση στόχων και μεγαλύτερη αποτελεσματικότητα των επιθέσεων [21], [25], [31].

2.4. Σύγχρονες τεχνικές κυβερνοεπιθέσεων

Το σύγχρονο περιβάλλον κυβερνοασφάλειας χαρακτηρίζεται από τη συνεχή εξέλιξη των επιθετικών τεχνικών, την αυξανόμενη πολυπλοκότητα των κυβερνοεπιθέσεων και την ενσωμάτωση προηγμένων τεχνολογιών αυτοματοποίησης και τεχνητής νοημοσύνης στις επιθετικές διαδικασίες. Οι σύγχρονες κυβερνοεπιθέσεις δεν βασίζονται πλέον αποκλειστικά σε μεμονωμένες τεχνικές παραβίασης, αλλά σε συνδυαστικές και πολυεπίπεδες στρατηγικές [32], [33], [37].

Η εξέλιξη αυτή συνδέεται άμεσα με την επαγγελματοποίηση του κυβερνοεγκλήματος και την ανάπτυξη οργανωμένων επιθετικών οικοσυστημάτων, μέσα στα οποία διακινούνται κακόβουλα εργαλεία, υπηρεσίες και έτοιμες επιθετικές υποδομές. Σύμφωνα με σύγχρονες εκθέσεις κυβερνοασφάλειας, οι επιτιθέμενοι αξιοποιούν ολοένα και περισσότερο αυτοματοποιημένα εργαλεία αναγνώρισης στόχων, πλατφόρμες malware-as-a-service και τεχνικές μαζικής εκτέλεσης επιθέσεων, γεγονός που αυξάνει σημαντικά την ταχύτητα και την κλίμακα των επιθετικών ενεργειών [33], [27].

2.4.1. Επιθέσεις Εκμετάλλευσης Ευπαθειών

Η εκμετάλλευση ευπαθειών αποτελεί μία από τις κυριότερες τεχνικές αρχικής πρόσβασης σε πληροφοριακά συστήματα. Οι επιτιθέμενοι εντοπίζουν αδυναμίες σε λογισμικό, λειτουργικά συστήματα, εφαρμογές ή δικτυακές υπηρεσίες και τις αξιοποιούν για την απόκτηση μη εξουσιοδοτημένης πρόσβασης [31], [25].

Στο σύγχρονο περιβάλλον απειλών παρατηρείται σημαντική μείωση του χρονικού διαστήματος μεταξύ της δημοσιοποίησης μιας ευπάθειας και της κακόβουλης αξιοποίησής της. Αυτοματοποιημένα εργαλεία σάρωσης και εκμετάλλευσης επιτρέπουν στους επιτιθέμενους να εντοπίζουν ευάλωτα συστήματα σε μεγάλη κλίμακα και να πραγματοποιούν επιθέσεις με αυξημένη ταχύτητα [33], [31].

Επιπλέον, ιδιαίτερη αύξηση παρουσιάζουν οι επιθέσεις που στοχεύουν περιβάλλοντα υπολογιστικού νέφους και υπηρεσίες απομακρυσμένης πρόσβασης. Η αυξημένη χρήση cloud υποδομών από οργανισμούς έχει διευρύνει την επιφάνεια επίθεσης, δημιουργώντας νέες δυνατότητες για εκμετάλλευση λανθασμένων ρυθμίσεων ασφαλείας, αδύναμων μηχανισμών αυθεντικοποίησης και ανεπαρκώς προστατευμένων λογαριασμών πρόσβασης [32], [31].

2.4.2. Επιθέσεις Ταυτότητας και Υποκλοπή Διαπιστευτηρίων

Οι επιθέσεις που βασίζονται στην κλοπή ή κατάχρηση ταυτοτήτων χρηστών αποτελούν πλέον έναν από τους σημαντικότερους μηχανισμούς διείσδυσης σε πληροφοριακά συστήματα. Οι επιτιθέμενοι επιδιώκουν την απόκτηση διαπιστευτηρίων πρόσβασης μέσω phishing, κακόβουλου λογισμικού, credential stuffing ή παραβίασης βάσεων δεδομένων [32], [31].

Η αυξανόμενη εξάρτηση οργανισμών από ψηφιακές υπηρεσίες και cloud περιβάλλοντα έχει ενισχύσει τη σημασία των επιθέσεων ταυτότητας, καθώς η απόκτηση έγκυρων διαπιστευτηρίων επιτρέπει στους επιτιθέμενους να παρακάμπτουν συμβατικούς μηχανισμούς ασφαλείας και να αποκτούν πρόσβαση σε κρίσιμα πληροφοριακά συστήματα [31], [36].

Παράλληλα, οι επιτιθέμενοι αξιοποιούν τεχνικές social engineering και ψυχολογικής χειραγώγησης, με στόχο την εξαπάτηση χρηστών και τη συλλογή ευαίσθητων πληροφοριών. Οι τεχνικές αυτές εμφανίζονται πλέον περισσότερο εξελιγμένες και προσωποποιημένες, αξιοποιώντας πληροφορίες που συλλέγονται από κοινωνικά δίκτυα, διαδικτυακές υπηρεσίες και δημόσια διαθέσιμα δεδομένα [32], [37].

2.4.3. Σύγχρονες Τεχνικές Phishing και Social Engineering

Οι επιθέσεις phishing έχουν εξελιχθεί σημαντικά τα τελευταία χρόνια και αποτελούν πλέον ιδιαίτερα σύνθετες μορφές social engineering. Δεν περιορίζονται σε γενικά και εύκολα αναγνωρίσιμα μηνύματα εξαπάτησης, αλλά αξιοποιούν προσωποποιημένο περιεχόμενο, τεχνικές μίμησης νόμιμων οργανισμών και αυτοματοποιημένη δημιουργία επιθετικών μηνυμάτων [32], [34].

Η αξιοποίηση τεχνητής νοημοσύνης και αυτοματοποιημένων συστημάτων παραγωγής περιεχομένου επιτρέπει τη δημιουργία περισσότερο πειστικών ηλεκτρονικών μηνυμάτων, ψεύτικων ιστοσελίδων και τεχνικών εξαπάτησης. Παράλληλα, αυξάνεται η χρήση επιθέσεων μέσω φωνητικής επικοινωνίας (vishing), γραπτών μηνυμάτων (smishing) και deepfake περιεχομένου, το οποίο χρησιμοποιείται για την παραποίηση φωνής ή εικόνας με στόχο την ενίσχυση της αξιοπιστίας της επίθεσης [35], [38].

Οι επιθέσεις social engineering θεωρούνται ιδιαίτερα αποτελεσματικές, διότι στοχεύουν κυρίως στον ανθρώπινο παράγοντα και όχι αποκλειστικά σε τεχνικές αδυναμίες. Η δυνατότητα εξατομίκευσης των επιθέσεων και η χρήση δεδομένων από διαφορετικές πηγές καθιστούν δυσκολότερη την αναγνώριση των κακόβουλων ενεργειών από τους τελικούς χρήστες [34], [37].

2.4.4. Επιθέσεις Ransomware και Κακόβουλου Λογισμικού

Το ransomware αποτελεί μία από τις πιο διαδεδομένες και οικονομικά επιζήμιες μορφές σύγχρονων κυβερνοεπιθέσεων. Οι επιθέσεις αυτές βασίζονται στην κρυπτογράφηση δεδομένων ή στην παρεμπόδιση πρόσβασης σε πληροφοριακά συστήματα, με σκοπό την απαίτηση χρηματικού ανταλλάγματος για την αποκατάσταση της λειτουργίας [32], [35].

Οι σύγχρονες επιχειρήσεις ransomware εμφανίζουν αυξημένο βαθμό οργάνωσης και λειτουργούν συχνά με επιχειρησιακά χαρακτηριστικά. Παρατηρείται ανάπτυξη μοντέλων ransomware-as-a-service, μέσω των οποίων κακόβουλοι παράγοντες παρέχουν εργαλεία ransomware σε τρίτους χρήστες, διευρύνοντας σημαντικά την κλίμακα των επιθέσεων [33], [27].

2.4.5. Επιθέσεις σε Υποδομές Νέφους και Εφοδιαστική Αλυσίδα

Η αυξανόμενη εξάρτηση οργανισμών από cloud υποδομές και εξωτερικούς παρόχους υπηρεσιών έχει οδηγήσει σε σημαντική αύξηση επιθέσεων που στοχεύουν περιβάλλοντα νέφους και αλυσίδες εφοδιασμού λογισμικού. Οι επιθέσεις αυτές επιδιώκουν να αξιοποιήσουν αδυναμίες σε τρίτους παρόχους, ενημερώσεις λογισμικού ή μηχανισμούς διασύνδεσης μεταξύ διαφορετικών οργανισμών [32], [36].

Οι επιθέσεις εφοδιαστικής αλυσίδας παρουσιάζουν ιδιαίτερη επικινδυνότητα, καθώς επιτρέπουν στους επιτιθέμενους να αποκτούν πρόσβαση σε πολλαπλούς οργανισμούς μέσω ενός μόνο σημείου παραβίασης. Παράλληλα, η αυξημένη χρήση cloud υπηρεσιών δημιουργεί νέες προκλήσεις ασφάλειας που σχετίζονται με τη διαχείριση ταυτοτήτων, τα δικαιώματα πρόσβασης και τη διαμόρφωση των υποδομών [31], [36].

Το σύγχρονο τοπίο κυβερνοαπειλών χαρακτηρίζεται από αυξημένη πολυπλοκότητα, αυτοματοποίηση και επιχειρησιακή οργάνωση. Η αξιοποίηση τεχνολογιών τεχνητής νοημοσύνης, cloud υποδομών και προηγμένων τεχνικών social engineering το μεταβάλλει ουσιαστικά δημιουργώντας νέες προκλήσεις για την ασφάλεια πληροφοριακών συστημάτων και την αποτελεσματική κυβερνοάμυνα.

3. Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΩΣ ΕΡΓΑΛΕΙΟ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΩΝ

3.1. Αυτοματοποίηση Επιθέσεων

Η αυτοματοποίηση επιθέσεων αποτελεί μία από τις σημαντικότερες μεταβολές που έχει επιφέρει η ενσωμάτωση τεχνολογιών τεχνητής νοημοσύνης στο πεδίο της κυβερνοασφάλειας. Η τεχνητή νοημοσύνη δεν περιορίζεται πλέον στη χρήση της ως υποστηρικτικού εργαλείου ανάλυσης δεδομένων ή αυτοματοποίησης απλών διαδικασιών, αλλά χρησιμοποιείται ολοένα και περισσότερο για την οργάνωση, επιτάχυνση και βελτιστοποίηση επιθετικών κυβερνοεπιχειρήσεων. Η εξέλιξη αυτή μεταβάλλει ουσιαστικά τον τρόπο με τον οποίο σχεδιάζονται και υλοποιούνται οι σύγχρονες κυβερνοεπιθέσεις, καθώς μεγάλο μέρος των επιθετικών διαδικασιών μπορεί πλέον να εκτελείται με περιορισμένη ανθρώπινη παρέμβαση [39], [42], [45].

Παραδοσιακά, η πραγματοποίηση σύνθετων κυβερνοεπιθέσεων απαιτούσε υψηλό επίπεδο τεχνικής εξειδίκευσης, σημαντικό χρόνο προετοιμασίας και χειροκίνητη εκτέλεση πολλών επιμέρους σταδίων της επίθεσης. Η διαδικασία συλλογής πληροφοριών, η ανάλυση πιθανών στόχων, η δημιουργία κακόβουλου λογισμικού, η ανάπτυξη τεχνικών social engineering και η εκμετάλλευση ευπαθειών προϋπέθεταν συνήθως εξειδικευμένες γνώσεις και εμπειρία στον χώρο της κυβερνοασφάλειας. Η αξιοποίηση όμως συστημάτων τεχνητής νοημοσύνης και μεγάλων γλωσσικών μοντέλων έχει μειώσει σημαντικά τις απαιτήσεις αυτές, επιτρέποντας την αυτοματοποίηση κρίσιμων επιθετικών λειτουργιών [40], [46].

Σύμφωνα με σύγχρονες αναφορές κυβερνοασφάλειας, τα εργαλεία τεχνητής νοημοσύνης μπορούν να χρησιμοποιηθούν για την αυτοματοποίηση πολλαπλών σταδίων του επιθετικού κύκλου ζωής. Οι επιτιθέμενοι είναι πλέον σε θέση να αξιοποιούν αλγορίθμους τεχνητής νοημοσύνης για αναγνώριση πιθανών στόχων, αυτοματοποιημένη συλλογή πληροφοριών από ανοιχτές πηγές, ανάλυση υποδομών, εντοπισμό ευπαθειών και δημιουργία προσωποποιημένων επιθετικών ενεργειών [39], [42].

Ιδιαίτερα σημαντική θεωρείται η χρήση μεγάλων γλωσσικών μοντέλων (Large Language Models – LLMs), τα οποία διαθέτουν δυνατότητες παραγωγής και ανάλυσης φυσικής γλώσσας σε μεγάλο βαθμό αυτοματοποίησης. Τα μοντέλα αυτά μπορούν να παράγουν πειστικά phishing emails, να

δημιουργούν κακόβουλα scripts, να προσαρμόζουν περιεχόμενο ανάλογα με το προφίλ του στόχου και να παρέχουν υποστήριξη σε διαδικασίες επιθετικού σχεδιασμού [41], [46], [47].

Η δυνατότητα αυτή επηρεάζει άμεσα το κόστος και τον χρόνο πραγματοποίησης κυβερνοεπιθέσεων. Σύμφωνα με έρευνες που εξετάζουν τη χρήση AI-assisted offensive operations, διαδικασίες που απαιτούσαν προηγουμένως εκτεταμένη ανθρώπινη εργασία μπορούν πλέον να ολοκληρώνονται ταχύτερα και με μικρότερη επιχειρησιακή επιβάρυνση. Η αυτοματοποίηση επιτρέπει στους επιτιθέμενους να πραγματοποιούν επιθέσεις σε μεγαλύτερη κλίμακα, να αυξάνουν τον αριθμό των στόχων και να προσαρμόζουν ταχύτερα τις τεχνικές τους ανάλογα με το περιβάλλον επίθεσης [42], [45], [46].

Παράλληλα, η τεχνητή νοημοσύνη συμβάλλει ουσιαστικά στη μείωση του απαιτούμενου επιπέδου τεχνικών γνώσεων. Η ανάπτυξη εργαλείων AI-as-a-service και αυτοματοποιημένων επιθετικών πλατφορμών καθιστά δυνατή την πρόσβαση σε προηγμένες επιθετικές δυνατότητες ακόμη και από χρήστες χωρίς εξειδικευμένη κατάρτιση. Η εξέλιξη αυτή θεωρείται ιδιαίτερα κρίσιμη, καθώς αυξάνει σημαντικά τον αριθμό επιτιθέμενων και διευρύνει την πρόσβαση στο κυβερνοέγκλημα [40], [43], [46].

Σύμφωνα με αναφορές διεθνών οργανισμών, η αυτοματοποίηση κυβερνοεπιθέσεων συνδέεται άμεσα με την ανάπτυξη οργανωμένων εγκληματικών οικοσυστημάτων, μέσα στα οποία κακόβουλα εργαλεία, έτοιμες επιθετικές υποδομές και υπηρεσίες τεχνητής νοημοσύνης παρέχονται εμπορικά σε τρίτους χρήστες. Η μετάβαση αυτή οδηγεί στη δημιουργία μοντέλων cybercrime-as-a-service, στα οποία η τεχνητή νοημοσύνη λειτουργεί ως εργαλείο βελτιστοποίησης και κλιμάκωσης εγκληματικών δραστηριοτήτων [43], [44].

Η αυτοματοποίηση επιθέσεων δεν περιορίζεται μόνο στην αρχική φάση διείσδυσης. Σύγχρονες επιθετικές τεχνικές αξιοποιούν τεχνητή νοημοσύνη και σε μεταγενέστερα στάδια, όπως η διατήρηση πρόσβασης σε παραβιασμένα περιβάλλοντα, η ανάλυση κλεμμένων δεδομένων, η αναγνώριση κρίσιμων πληροφοριών και η αυτοματοποιημένη λήψη αποφάσεων κατά την εξέλιξη της επίθεσης [39], [48].

Επιπλέον, η χρήση LLM-based agents δημιουργεί νέα δεδομένα ως προς την ημιαυτόνομη ή αυτόνομη λειτουργία επιθετικών συστημάτων. Σύμφωνα με πρόσφατες ερευνητικές προσεγγίσεις,

αναπτύσσονται συστήματα τα οποία μπορούν να εκτελούν αλληλουχίες επιθετικών ενεργειών με περιορισμένη ανθρώπινη καθοδήγηση, συνδυάζοντας συλλογή πληροφοριών, ανάλυση αποτελεσμάτων και προσαρμογή στρατηγικής [42], [46].

Παράλληλα, αυξάνεται σημαντικά η ταχύτητα διάδοσης και εκτέλεσης επιθέσεων. Οι επιτιθέμενοι μπορούν να αξιοποιούν αυτοματοποιημένους μηχανισμούς σάρωσης δικτύων, αναγνώρισης ευπαθειών και δημιουργίας επιθετικών σεναρίων σε πραγματικό χρόνο. Ως αποτέλεσμα, μειώνεται το χρονικό περιθώριο αντίδρασης των αμυντικών μηχανισμών και αυξάνεται η δυσκολία έγκαιρης ανίχνευσης των επιθέσεων [39], [45].

Ιδιαίτερη ανησυχία προκαλεί επίσης η δυνατότητα χρήσης τεχνητής νοημοσύνης για την παραγωγή μεγάλου όγκου κακόβουλου περιεχομένου με χαμηλό κόστος. Η δημιουργία κακόβουλων ενεργειών μπορεί πλέον να πραγματοποιείται σε βιομηχανική κλίμακα, αυξάνοντας τη μαζικότητα και την αποτελεσματικότητα των επιθέσεων social engineering [39], [44], [47].

3.2. Μείωση του κόστους και των τεχνικών απαιτήσεων για τους επιτιθέμενους

Η αξιοποίηση της τεχνητής νοημοσύνης στις κυβερνοεπιθέσεις δεν μεταβάλλει μόνο την ταχύτητα και την κλίμακα των επιθετικών ενεργειών, αλλά επηρεάζει ουσιαστικά και το επίπεδο τεχνικής εξειδίκευσης που απαιτείται από τον επιτιθέμενο. Η μεταβολή αυτή αποτελεί κρίσιμο χαρακτηριστικό του σύγχρονου κυβερνοεγκλήματος, διότι μειώνει τα εμπόδια εισόδου σε επιθετικές δραστηριότητες και καθιστά δυνατή την εκτέλεση πιο σύνθετων ενεργειών από άτομα ή ομάδες με περιορισμένες τεχνικές γνώσεις [40], [27], [43].

Παραδοσιακά, η πραγματοποίηση σύνθετων κυβερνοεπιθέσεων απαιτούσε εξειδικευμένη γνώση σε τομείς όπως ο προγραμματισμός, η ανάλυση δικτύων, η εκμετάλλευση ευπαθειών, η ανάπτυξη κακόβουλου λογισμικού και η απόκρυψη επιθετικών ιχνών. Η τεχνητή νοημοσύνη, ιδίως μέσω των μεγάλων γλωσσικών μοντέλων, μεταβάλλει αυτή τη συνθήκη, καθώς μπορεί να λειτουργήσει ως ενδιάμεσο τεχνικό εργαλείο που μεταφράζει γενικές οδηγίες σε πιο συγκεκριμένες τεχνικές ενέργειες [40], [47], [46].

Η μείωση των τεχνικών απαιτήσεων δεν σημαίνει ότι ο επιτιθέμενος καθίσταται πλήρως ανεξάρτητος από την τεχνική γνώση. Αντιθέτως, σημαίνει ότι η τεχνητή νοημοσύνη μπορεί να περιορίσει το επίπεδο εξειδίκευσης που απαιτείται για την εκκίνηση, την οργάνωση και την υποστήριξη μιας επίθεσης. Ένα άτομο με βασική κατανόηση του στόχου μπορεί πλέον να αξιοποιήσει συστήματα τεχνητής νοημοσύνης για τη σύνταξη phishing μηνυμάτων, τη δημιουργία scripts, τη διαμόρφωση επιθετικών σεναρίων, την ανάλυση πληροφοριών και την παραγωγή τεχνικών οδηγιών [40], [39], [41].

Ιδιαίτερη σημασία έχει η λειτουργία των μεγάλων γλωσσικών μοντέλων ως εργαλείων τεχνικής υποστήριξης. Τα μοντέλα αυτά μπορούν να υποστηρίξουν διαδικασίες οι οποίες παλαιότερα απαιτούσαν αναζήτηση σε εξειδικευμένα τεχνικά φόρουμ, εμπειρία σε προγραμματισμό ή γνώση συγκεκριμένων εργαλείων κυβερνοασφάλειας [47], [46].

Η εξέλιξη αυτή συνδέεται άμεσα με το φαινόμενο της ενίσχυσης ικανοτήτων, δηλαδή της δυνατότητας που παρέχει η τεχνητή νοημοσύνη σε λιγότερο ικανούς ή λιγότερο εξειδικευμένους δρώντες να προσεγγίσουν επίπεδα απόδοσης που παλαιότερα απαιτούσαν μεγαλύτερη εμπειρία. Η τεχνητή νοημοσύνη δεν δημιουργεί απλώς νέες τεχνικές επιθέσεων, αλλά καθιστά υφιστάμενες τεχνικές περισσότερο προσβάσιμες, ταχύτερες και ευκολότερες στην εφαρμογή [49], [46], [50].

Η μείωση κόστους συνδέεται επίσης με την κλιμάκωση. Όταν η παραγωγή επιθετικού περιεχομένου ή η τεχνική υποστήριξη μιας επίθεσης αυτοματοποιείται, το οριακό κόστος επανάληψης μειώνεται σημαντικά. Ένα επιθετικό πρότυπο μπορεί να προσαρμοστεί σε διαφορετικούς στόχους, γλώσσες, οργανισμούς ή τεχνικά περιβάλλοντα με μικρότερη προσπάθεια σε σχέση με τις παραδοσιακές μεθόδους. Αυτό καθιστά τις επιθέσεις και φθηνότερες, αλλά και μαζικές [27], [43], [46].

Η ίδια λογική παρατηρείται και στο οικοσύστημα του κυβερνοεγκλήματος ως υπηρεσίας. Οι εκθέσεις για το σύγχρονο κυβερνοέγκλημα καταγράφουν την ύπαρξη υποδομών, εργαλείων και υπηρεσιών που παρέχονται σε τρίτους, επιτρέποντας σε άτομα χωρίς πλήρη τεχνική επάρκεια να αποκτούν πρόσβαση σε έτοιμες επιθετικές δυνατότητες. Η τεχνητή νοημοσύνη ενισχύει αυτή την τάση, διότι μπορεί να ενσωματωθεί σε εργαλεία phishing, δημιουργίας malware, αυτοματοποίησης επικοινωνίας και τεχνικής υποστήριξης επιθετικών ενεργειών [22], [43], [39].

Η έννοια του crime-as-a-service είναι ιδιαίτερος σημαντική για την κατανόηση της μείωσης των τεχνικών απαιτήσεων. Σε τέτοια μοντέλα, ο επιτιθέμενος δεν χρειάζεται να κατασκευάσει ο ίδιος όλα τα επιμέρους στοιχεία της επίθεσης. Μπορεί να αγοράσει ή να μισθώσει υπηρεσίες, εργαλεία, πρόσβαση, υποδομές ή τεχνική υποστήριξη. Όταν σε αυτό το οικοσύστημα προστίθενται δυνατότητες τεχνητής νοημοσύνης, το τεχνικό βάρος μεταφέρεται ακόμη περισσότερο από τον τελικό χρήστη του εργαλείου προς τον πάροχο της υπηρεσίας ή προς το αυτοματοποιημένο σύστημα [22], [43].

Τα κακόβουλα εργαλεία τεχνητής νοημοσύνης που αναφέρονται σε αναλύσεις της underground αγοράς δείχνουν ότι η τεχνητή νοημοσύνη μπορεί να αξιοποιηθεί για εργασίες όπως η παραγωγή phishing περιεχομένου, η υποστήριξη δημιουργίας κακόβουλων κώδικα, η έρευνα ευπαθειών και η αυτοματοποίηση επιθετικών ενεργειών. Αυτό δεν σημαίνει ότι κάθε τέτοιο εργαλείο είναι τεχνικά προηγμένο στον ίδιο βαθμό, αλλά δείχνει την τάση εμπορευματοποίησης της τεχνητής νοημοσύνης ως υποστηρικτικού μηχανισμού κυβερνοεγκλήματος [39].

Η μείωση των τεχνικών απαιτήσεων έχει ιδιαίτερη σημασία και για τις επιθέσεις social engineering. Η παραγωγή πειστικού και γλωσσικά και συντακτικά ορθού περιεχομένου αποτελούσε μέχρι πρόσφατα σημαντικό περιορισμό για πολλούς επιτιθέμενους, ιδίως όταν στόχευαν θύματα σε ξένες γλώσσες ή σε εξειδικευμένα επαγγελματικά περιβάλλοντα. Τα συστήματα τεχνητής νοημοσύνης μπορούν να περιορίσουν αυτό το εμπόδιο, παράγοντας κείμενα που εμφανίζονται πιο φυσικά, πιο προσωποποιημένα και πιο συμβατά με το ύφος του οργανισμού ή του προσώπου που μιμούνται [40], [27], [41].

Επιπλέον, ο επιτιθέμενος δεν χρειάζεται πάντοτε να κατανοεί σε βάθος κάθε τεχνική λεπτομέρεια, εφόσον μπορεί να ζητήσει από ένα μοντέλο να εξηγήσει, να μετατρέψει, να διορθώσει ή να προσαρμόσει τεχνικό περιεχόμενο. Αυτό δημιουργεί ένα περιβάλλον στο οποίο η γνώση δεν αντικαθίσταται πλήρως, αλλά διαμεσολαβείται από συστήματα που επιταχύνουν τη μάθηση και μειώνουν την “απόσταση” ανάμεσα στην πρόθεση και την τεχνική υλοποίηση [47], [46].

Η επίδραση αυτή είναι ιδιαίτερα εμφανής στην περίπτωση των LLM-based agents. Οι πράκτορες που βασίζονται σε μεγάλα γλωσσικά μοντέλα μπορούν να οργανώνουν αλληλουχίες ενεργειών, να εκτελούν επιμέρους βήματα, να αξιολογούν ενδιάμεσα αποτελέσματα και να προσαρμόζουν την επόμενη ενέργεια με βάση τον στόχο. Σε επιθετικό πλαίσιο, τέτοιες δυνατότητες μπορούν να

μειώσουν τον χρόνο που απαιτείται, αφού μέρος του σχεδιασμού και της εκτέλεσης μπορεί να μεταφερθεί στο ίδιο το σύστημα [46].

Ωστόσο, η μείωση των τεχνικών απαιτήσεων δεν πρέπει να ερμηνεύεται ως πλήρης εξάλειψη των περιορισμών. Οι επιθέσεις υψηλής πολυπλοκότητας εξακολουθούν να απαιτούν τεχνική κρίση, επιχειρησιακή εμπειρία, επιλογή κατάλληλων εργαλείων και ικανότητα αξιολόγησης αποτελεσμάτων. Η τεχνητή νοημοσύνη ενισχύει μόνο ορισμένες δυνατότητες των επιτιθέμενων, αλλά παράλληλα μπορεί να ενισχύσει και τις αμυντικές λειτουργίες [49], [50].

Η τεχνητή νοημοσύνη δεν αντικαθιστά πλήρως τους επιτιθέμενους, αλλά επιτρέπει σε λιγότερο εξειδικευμένους να εκτελούν ευκολότερα διεργασίες υψηλών απαιτήσεων. Μέσω αυτοματοποιημένης παραγωγής περιεχομένου, τεχνικής υποστήριξης, έτοιμων υπηρεσιών και εργαλείων ως υπηρεσία, ορισμένες δυνατότητες καθίστανται περισσότερο προσβάσιμες σε λιγότερο έμπειρους δρώντες. Παράλληλα, οι επιθέσεις υψηλής πολυπλοκότητας εξακολουθούν να απαιτούν τεχνική κρίση, επιχειρησιακή εμπειρία και ικανότητα αξιολόγησης αποτελεσμάτων. [40], [27], [22], [43], [46].

3.3. Κλιμάκωση επιθέσεων (mass attacks)

Η ενσωμάτωση τεχνολογιών τεχνητής νοημοσύνης στο πεδίο των κυβερνοεπιθέσεων έχει μεταβάλει ουσιαστικά όχι μόνο τον τρόπο εκτέλεσης επιθετικών ενεργειών, αλλά και τη δυνατότητα κλιμάκωσής τους. Η σημαντικότερη αλλαγή που επιφέρει η τεχνητή νοημοσύνη αφορά τη μετάβαση από περιορισμένες, περισσότερο χειροκίνητες και χρονοβόρες επιθέσεις σε επιθετικές διαδικασίες μεγάλης κλίμακας, οι οποίες μπορούν να εκτελούνται ταχύτατα, μαζικά και με υψηλό βαθμό προσαρμοστικότητας [32], [40], [43].

Στο παραδοσιακό μοντέλο κυβερνοεπιθέσεων, η κλιμάκωση μιας επιχείρησης συναντούσε σημαντικούς περιορισμούς. Η ανάγκη ανθρώπινης συμμετοχής σε στάδια όπως η συλλογή πληροφοριών, η επιλογή στόχων, η προσαρμογή phishing μηνυμάτων, η ανάλυση αποτελεσμάτων και η διαχείριση επιθετικών υποδομών περιόριζε την ταχύτητα και τον όγκο των επιθέσεων που

μπορούσαν να εκτελεστούν ταυτόχρονα. Επιπλέον, η αύξηση της κλίμακας μιας επίθεσης απαιτούσε αναλογική αύξηση πόρων, χρόνου και ανθρώπινου δυναμικού [40], [21], [42].

Η τεχνητή νοημοσύνη μεταβάλλει ριζικά αυτή τη συνθήκη. Οι επιτιθέμενοι μπορούν πλέον να πραγματοποιούν επιθέσεις μεγάλης κλίμακας με περιορισμένη ανθρώπινη παρέμβαση. Η ΑΙ δεν λειτουργεί ως βοηθητικό εργαλείο, αλλά λειτουργεί ως ενεργός επιχειρησιακός μηχανισμός, ικανός να υποστηρίξει ταυτόχρονα πολλαπλά στάδια μιας επιθετικής διαδικασίας [40], [42], [50].

Ένα από τα σημαντικότερα χαρακτηριστικά αυτής της μεταβολής είναι η δυνατότητα ταυτόχρονης στόχευσης μεγάλου αριθμού θυμάτων. Σύμφωνα με σύγχρονες αναφορές κυβερνοασφάλειας, η τεχνητή νοημοσύνη επιτρέπει την παραγωγή και διαχείριση χιλιάδων ή ακόμη και εκατομμυρίων εξατομικευμένων επιθετικών αλληλεπιδράσεων σε πολύ μικρό χρονικό διάστημα. Η δημιουργία phishing emails, ψεύτικων ιστοσελίδων, impersonation domains και μηνυμάτων social engineering μπορεί πλέον να πραγματοποιείται ακαριαία [32], [43], [21].

Ιδιαίτερα χαρακτηριστική είναι η δυνατότητα δημιουργίας μεγάλου αριθμού domains και επιθετικών υποδομών μέσα σε ελάχιστο χρόνο. Σύμφωνα με τεχνικές αναλύσεις, εργαλεία τεχνητής νοημοσύνης μπορούν να υποστηρίξουν τη γρήγορη παραγωγή χιλιάδων domains που μιμούνται νόμιμους οργανισμούς ή υπηρεσίες, διευκολύνοντας την οργάνωση εκτεταμένων phishing και scam campaigns [32]. Η εξέλιξη αυτή αυξάνει σημαντικά τη δυσκολία ανίχνευσης των επιθέσεων, καθώς οι αμυντικοί μηχανισμοί καλούνται να αντιμετωπίσουν τεράστιο όγκο δυναμικά μεταβαλλόμενων επιθετικών στοιχείων.

Παράλληλα, η τεχνητή νοημοσύνη ενισχύει τη δυνατότητα εξατομίκευσης των μαζικών επιθέσεων. Στο παρελθόν, οι μαζικές phishing επιθέσεις βασίζονταν κυρίως σε γενικά και επαναλαμβανόμενα μηνύματα, γεγονός που περιορίζει την αποτελεσματικότητά τους. Η χρήση όμως γενετικών μοντέλων τεχνητής νοημοσύνης επιτρέπει τη δημιουργία περιεχομένου προσαρμοσμένου στα χαρακτηριστικά κάθε στόχου, αυξάνοντας την αξιοπιστία και την πειστικότητα των επιθέσεων [51], [52], [53].

Η δυνατότητα αυτή έχει ιδιαίτερη σημασία, διότι συνδυάζει δύο χαρακτηριστικά που παλαιότερα θεωρούνταν δύσκολο να συνυπάρξουν, τη μαζικότητα και την προσωποποίηση. Παραδοσιακά,

όσο αυξάνεται η εξατομίκευση μιας επίθεσης, τόσο αυξάνεται και το απαιτούμενο κόστος και ο χρόνος προετοιμασίας. [32], [51], [53].

Η κλιμάκωση επιθέσεων συνδέεται επίσης με την ανάπτυξη adaptive malware και δυναμικών επιθετικών τεχνικών. Σύγχρονες ερευνητικές προσεγγίσεις επισημαίνουν ότι τα συστήματα τεχνητής νοημοσύνης μπορούν να υποστηρίξουν κακόβουλο λογισμικό το οποίο προσαρμόζει τη συμπεριφορά του ανάλογα με το περιβάλλον στόχου, τις αμυντικές αντιδράσεις ή τα χαρακτηριστικά του συστήματος που προσβάλλεται [52], [54]. Η προσαρμοστικότητα αυτή αυξάνει τη δυναμική των επιθέσεων και επιτρέπει τη συνέχιση της επιθετικής δραστηριότητας ακόμη και όταν εφαρμόζονται μηχανισμοί άμυνας.

Σημαντικό ρόλο στην κλιμάκωση διαδραματίζουν και οι cloud υποδομές. Η διαθεσιμότητα υπολογιστικών πόρων μέσω cloud υπηρεσιών επιτρέπει στους επιτιθέμενους να αξιοποιούν υψηλή υπολογιστική ισχύ χωρίς την ανάγκη ανάπτυξης δικών τους φυσικών υποδομών. Η τεχνητή νοημοσύνη, σε συνδυασμό με cloud environments, επιτρέπει τη μαζική αυτοματοποίηση επιθετικών διαδικασιών, τη γρήγορη διανομή κακόβουλου περιεχομένου και τη δυναμική επέκταση επιθετικών επιχειρήσεων σε πολλαπλούς στόχους και γεωγραφικές περιοχές [21], [33], [42].

Παράλληλα, η τεχνητή νοημοσύνη ενισχύει τη βιομηχανοποίηση του κυβερνοεγκλήματος. Η έννοια της industrialisation of cybercrime περιγράφει τη μετάβαση από αποσπασματικές επιθέσεις μικρής κλίμακας σε οργανωμένες, επαναλαμβανόμενες και υψηλά αυτοματοποιημένες εγκληματικές διαδικασίες [21], [33]. Η ΑΙ επιτρέπει στις εγκληματικές ομάδες να αυξάνουν τον όγκο των επιθέσεων χωρίς αντίστοιχη αύξηση ανθρώπινων πόρων, δημιουργώντας μοντέλα λειτουργίας που θυμίζουν περισσότερο οργανωμένες ψηφιακές επιχειρήσεις.

Η μεταβολή αυτή επηρεάζει άμεσα το επιχειρησιακό κόστος των επιθέσεων. Η αυτοματοποίηση παραγωγής phishing περιεχομένου, η χρήση AI-enabled bots, η ταχεία δημιουργία επιθετικών υποδομών και η αυτοματοποιημένη συλλογή πληροφοριών μειώνουν σημαντικά το κόστος ανά επίθεση. Ως αποτέλεσμα, καθίσταται ιδιαιτέρως επικερδής η πραγματοποίηση μεγάλου αριθμού επιθέσεων ακόμη και όταν το ποσοστό επιτυχίας κάθε μεμονωμένης επίθεσης είναι σχετικά χαμηλό [43], [33], [51].

Η τεχνητή νοημοσύνη αυξάνει επίσης την επιχειρησιακή αποτελεσματικότητα των μαζικών επιθέσεων μέσω συνεχούς προσαρμογής και βελτιστοποίησης. Οι επιθέσεις πλέον μπορούν να τροποποιούν δυναμικά το περιεχόμενο, τη χρονική στιγμή εκτέλεσης, τη μορφή επικοινωνίας ή τη στρατηγική προσέγγισης του στόχου. Η δυνατότητα αυτή αυξάνει την πιθανότητα επιτυχίας και δυσχεραίνει την ανίχνευση από παραδοσιακά αμυντικά συστήματα [40], [52], [54].

Επιπλέον, μέσω της τεχνητής νοημοσύνης μειώνεται ο χρόνος μεταξύ εντοπισμού μιας ευπάθειας και εκμετάλλευσής της. Η ταχεία εκμετάλλευση ευπαθειών αποτελεί βασικό χαρακτηριστικό του σύγχρονου τοπίου απειλών, καθώς οι επιτιθέμενοι μπορούν να ενσωματώνουν γρήγορα νέες τεχνικές σε ήδη υπάρχουσες επιθετικές υποδομές [21], [33]. Αυτό αυξάνει την πίεση προς οργανισμούς και κρατικές δομές, οι οποίοι διαθέτουν μικρότερο χρονικό περιθώριο αντίδρασης.

Η κλιμάκωση επιθέσεων μέσω τεχνητής νοημοσύνης δημιουργεί ένα νέο περιβάλλον ασύμμετρης απειλής. Μικρότερες ομάδες ή ακόμη και μεμονωμένοι δράστες μπορούν να αποκτούν δυνατότητες που παλαιότερα απαιτούσαν σημαντικούς οικονομικούς πόρους, μεγάλη τεχνική εξειδίκευση ή οργανωμένες εγκληματικές δομές. Η ΑΙ λειτουργεί ως πολλαπλασιαστής ισχύος, επιτρέποντας την παραγωγή δυσανάλογα μεγάλου αντίκτυπου σε σχέση με τους πραγματικούς πόρους του επιτιθέμενου [40], [44], [42].

Η βασική διαφορά δεν εντοπίζεται μόνο στην ταχύτητα ή στον όγκο, αλλά κυρίως στον συνδυασμό μαζικότητας, αυτοματοποίησης, προσαρμοστικότητας και εξατομίκευσης. Η τεχνητή νοημοσύνη επιτρέπει τη μετάβαση από μαζικές αλλά γενικές επιθέσεις σε μαζικές αλλά ταυτόχρονα δυναμικά προσαρμοζόμενες επιθετικές επιχειρήσεις [32], [51], [53].

Η τεχνητή νοημοσύνη μεταβάλλει ριζικά την έννοια της κλιμάκωσης στο πεδίο των κυβερνοεπιθέσεων. Μέσω αυτοματοποίησης, εξατομίκευσης, cloud υποδομών, adaptive malware και AI-enabled επιχειρησιακών μηχανισμών, οι επιθέσεις αποκτούν μεγαλύτερη ταχύτητα, μεγαλύτερη γεωγραφική και επιχειρησιακή εμβέλεια, μειωμένο κόστος και αυξημένη αποτελεσματικότητα [32], [40], [21], [52].

3.4. Artificial Intelligence στη συλλογή πληροφοριών (reconnaissance)

Η διαδικασία συλλογής πληροφοριών και αναγνώρισης στόχων αποτελεί ένα από τα σημαντικότερα στάδια κάθε κυβερνοεπίθεσης, καθώς μέσω αυτής ο επιτιθέμενος αποκτά εικόνα για τις τεχνικές, οργανωτικές και ανθρώπινες αδυναμίες ενός πιθανού στόχου. Παραδοσιακά, το στάδιο του reconnaissance απαιτούσε σημαντικό χρόνο, εξειδικευμένες τεχνικές γνώσεις και εκτεταμένη ανθρώπινη παρέμβαση, καθώς η αναζήτηση πληροφοριών πραγματοποιούνταν μέσω χειροκίνητης ανάλυσης ιστοσελίδων, δικτυακών υπηρεσιών, δημόσιων βάσεων δεδομένων και κοινωνικών δικτύων. Η ανάπτυξη όμως τεχνολογιών τεχνητής νοημοσύνης μεταβάλλει ριζικά αυτή τη διαδικασία, επιτρέποντας την αυτοματοποίηση της συλλογής, επεξεργασίας και αξιοποίησης πληροφοριών σε πρωτοφανή κλίμακα [55], [46], [40].

Η τεχνητή νοημοσύνη ενισχύει το reconnaissance κυρίως μέσω της δυνατότητας επεξεργασίας μεγάλου όγκου δεδομένων από πολλαπλές και ετερογενείς πηγές. Σύγχρονα συστήματα AI μπορούν να συλλέγουν πληροφορίες από δημόσιες ιστοσελίδες, κοινωνικά δίκτυα, forums, repositories κώδικα, DNS εγγραφές, metadata αρχείων, τεχνικές βάσεις δεδομένων και άλλες ανοικτές πηγές πληροφόρησης, πραγματοποιώντας διαδικασίες Open Source Intelligence (OSINT) με πολύ μεγαλύτερη ταχύτητα και ακρίβεια σε σχέση με τις παραδοσιακές μεθόδους [55], [57], [58].

Στο μοντέλο reconnaissance, η διαδικασία συσχέτισης πληροφοριών ήταν πολύ απαιτητικό στάδιο. Ο επιτιθέμενος έπρεπε να εντοπίσει και να συνδυάσει χειροκίνητα δεδομένα που προέρχονταν από διαφορετικές πηγές, ώστε να δημιουργήσει ένα λειτουργικό προφίλ του στόχου. Η τεχνητή νοημοσύνη μεταβάλλει ουσιαστικά αυτή τη συνθήκη, καθώς οι αλγόριθμοι μηχανικής μάθησης και επεξεργασίας φυσικής γλώσσας μπορούν να αναγνωρίζουν πρότυπα, σχέσεις και συσχετίσεις ανάμεσα σε φαινομενικά ασύνδετα δεδομένα [55], [57], [52].

Η δυνατότητα αυτή αποκτά ιδιαίτερη σημασία στην περίπτωση οργανισμών με εκτεταμένη ψηφιακή παρουσία. Σύγχρονες τεχνικές AI-driven reconnaissance μπορούν να εντοπίζουν αυτόματα ψηφιακά assets, subdomains, exposed services, δημόσια διαθέσιμες υποδομές και πιθανές τεχνικές αδυναμίες ενός οργανισμού. Η αυτοματοποίηση της διαδικασίας asset discovery επιτρέπει στους επιτιθέμενους να δημιουργούν γρήγορα μία πλήρη εικόνα του στόχου και των ευπαθειών του [55], [59].

Η στρατηγική σημασία της AI-driven reconnaissance έγκειται κυρίως στη δυνατότητα μετατροπής μεγάλου όγκου αδόμητων δεδομένων σε αξιοποιήσιμη επιχειρησιακή πληροφορία. Η τεχνητή νοημοσύνη εκτός από συλλογή δεδομένων, πραγματοποιεί ανάλυση περιεχομένου, κατηγοριοποίηση πληροφοριών, εξαγωγή οντοτήτων και αναγνώριση μοτίβων συμπεριφοράς [57], [58]. Με τον τρόπο αυτό, οι επιτιθέμενοι αποκτούν βαθύτερη κατανόηση των αδυναμιών ενός οργανισμού πριν ακόμη πραγματοποιηθεί η επίθεση.

Ιδιαίτερη σημασία παρουσιάζει η αξιοποίηση τεχνικών Natural Language Processing (NLP) στο πλαίσιο της Cyber Threat Intelligence (CTI). Οι τεχνικές αυτές επιτρέπουν την ανάλυση μεγάλου όγκου πληροφοριών που προέρχονται από διαφορετικές μορφές κειμενικού περιεχομένου, όπως αναφορές ασφαλείας, forums, social media posts, blogs και τεχνικές δημοσιεύσεις [57], [58]. Μέσω αυτής της διαδικασίας, εντοπίζονται τάσεις, αναγνωρίζονται επαναλαμβανόμενα πρότυπα και αντλούνται χρήσιμες πληροφορίες για πιθανές επιθετικές στρατηγικές ή αδύναμα σημεία ενός στόχου.

Η δυνατότητα ανάλυσης κοινωνικών δικτύων αποτελεί επίσης κρίσιμο στοιχείο της σύγχρονης reconnaissance. Οι επιτιθέμενοι μπορούν να αξιοποιούν αλγορίθμους AI για την αυτόματη συλλογή και ανάλυση πληροφοριών σχετικά με εργαζομένους, οργανωτικές δομές, επαγγελματικές σχέσεις και ψηφιακές συνήθειες χρηστών. Μέσω αυτής της διαδικασίας δημιουργούνται λεπτομερή target profiles, τα οποία μπορούν να αξιοποιηθούν σε πιο στοχευμένες επιθέσεις social engineering [41], [40], [52].

Η εξέλιξη αυτή διαφοροποιεί ουσιαστικά τις σύγχρονες επιθέσεις spear-phishing από τις παραδοσιακές μαζικές phishing campaigns. Στο παρελθόν, οι επιθέσεις βασίζονταν συχνά σε γενικά και μη προσωποποιημένα μηνύματα, γεγονός που περιόριζε την αποτελεσματικότητά τους. Η AI επιτρέπει πλέον τη δημιουργία προσωποποιημένων μηνυμάτων που λαμβάνουν υπόψη τις πληροφορίες που έχουν συλλεχθεί κατά το reconnaissance [41], [52]. Η αυξημένη αυτή εξατομίκευση ενισχύει σημαντικά την πιθανότητα επιτυχίας της επίθεσης.

Παράλληλα, η τεχνητή νοημοσύνη συμβάλλει στην ταχύτερη αναγνώριση πιθανών ευπαθειών. AI-assisted εργαλεία reconnaissance μπορούν να πραγματοποιούν αυτοματοποιημένη ανάλυση δικτυακών υπηρεσιών, scanning υποδομών, subdomain enumeration και vulnerability discovery με περιορισμένη ανθρώπινη παρέμβαση [56], [59]. Η δυνατότητα αυτή επιτρέπει στους

επιτιθέμενους να εντοπίζουν ταχύτερα αδύναμα σημεία και να μειώνουν το χρονικό διάστημα μεταξύ reconnaissance και exploitation.

Επιπλέον, η τεχνητή νοημοσύνη ενισχύει τη δυνατότητα πρόβλεψης πιθανών αδύναμων σημείων ενός οργανισμού. Μέσω της ανάλυσης ιστορικών δεδομένων, τεχνικών προτύπων, γνωστών vulnerabilities και συμπεριφορών χρηστών, οι αλγόριθμοι AI μπορούν να υποστηρίξουν την εκτίμηση πιθανών επιφανειών επίθεσης και να κατευθύνουν αποτελεσματικότερα τη στρατηγική του επιτιθέμενου [56], [42].

Η reconnaissance αποκτά έτσι αυξημένη στρατηγική σημασία στο πλαίσιο των σύγχρονων κυβερνοεπιθέσεων. Η επιτυχία μιας επίθεσης εξαρτάται πλέον σε μεγάλο βαθμό από την ποιότητα, την ακρίβεια και την ταχύτητα συλλογής πληροφοριών πριν από την έναρξη της επιθετικής ενέργειας. Η AI επιτρέπει στους επιτιθέμενους να μετατρέπουν το reconnaissance σε μία διαρκή, δυναμική και υψηλά αυτοματοποιημένη διαδικασία, η οποία λειτουργεί ως θεμέλιο για μεταγενέστερα στάδια της επίθεσης [46], [40], [42].

Η μεταβολή αυτή επηρεάζει άμεσα και την ισορροπία μεταξύ επιτιθέμενων και αμυνόμενων. Οι οργανισμοί καλούνται πλέον να προστατεύσουν όχι μόνο τα εσωτερικά τους συστήματα, αλλά και το σύνολο της δημόσια διαθέσιμης ψηφιακής πληροφορίας που μπορεί να αξιοποιηθεί στο πλαίσιο AI-driven reconnaissance. Η ύπαρξη μεγάλου όγκου ανοικτών πληροφοριών, σε συνδυασμό με τη δυνατότητα αυτοματοποιημένης επεξεργασίας τους, αυξάνει σημαντικά την επιφάνεια απειλής [55], [58], [52].

Η τεχνητή νοημοσύνη μεταβάλλει ριζικά το reconnaissance, επιτρέποντας την αυτοματοποιημένη συλλογή, ανάλυση και αξιοποίηση πληροφοριών σε ταχύτητα και κλίμακα που δεν ήταν προηγουμένως εφικτές. Μέσω τεχνικών OSINT, NLP, Cyber Threat Intelligence, automated asset discovery και AI-assisted vulnerability discovery, οι επιτιθέμενοι αποκτούν αυξημένες δυνατότητες αναγνώρισης στόχων, δημιουργίας λεπτομερών target profiles και σχεδιασμού πιο στοχευμένων και αποτελεσματικών επιθέσεων. Το reconnaissance μετατρέπεται έτσι από υποστηρικτικό στάδιο προετοιμασίας σε κρίσιμο στρατηγικό πυλώνα των σύγχρονων AI-driven κυβερνοεπιθέσεων [55], [46], [57], [40].

3.5. Παραγωγή κακόβουλου λογισμικού

Η παραγωγή κακόβουλου λογισμικού αποτελεί έναν από τους σημαντικότερους τομείς στους οποίους η τεχνητή νοημοσύνη επηρεάζει τη σύγχρονη κυβερνοασφάλεια. Παραδοσιακά, η ανάπτυξη malware απαιτούσε υψηλό επίπεδο τεχνικής εξειδίκευσης, βαθιά γνώση προγραμματισμού, κατανόηση λειτουργικών συστημάτων και εμπειρία σε τεχνικές εκμετάλλευσης ευπαθειών και αποφυγής ανίχνευσης. Η διαδικασία δημιουργίας κακόβουλου λογισμικού ήταν συνήθως χρονοβόρα και απαιτούσε σημαντική ανθρώπινη παρέμβαση, ιδιαίτερα κατά τα στάδια συγγραφής κώδικα, obfuscation, testing και προσαρμογής του malware σε διαφορετικά περιβάλλοντα στόχων. Η ενσωμάτωση όμως τεχνολογιών τεχνητής νοημοσύνης και ειδικότερα μεγάλων γλωσσικών μοντέλων μεταβάλλει ουσιαστικά αυτή τη διαδικασία, επιτρέποντας την ταχύτερη, περισσότερο αυτοματοποιημένη και δυναμική ανάπτυξη κακόβουλου λογισμικού [40], [52], [39].

Η τεχνητή νοημοσύνη ενισχύει τη διαδικασία παραγωγής malware κυρίως μέσω της δυνατότητας αυτοματοποιημένης δημιουργίας και τροποποίησης κώδικα. Σύγχρονα μοντέλα generative AI και large language models μπορούν να παράγουν λειτουργικά τμήματα προγραμματιστικού κώδικα, να προτείνουν εναλλακτικές υλοποιήσεις, να βελτιστοποιούν υπάρχοντα scripts και να υποστηρίζουν τη δημιουργία εργαλείων που χρησιμοποιούνται σε κυβερνοεπιθέσεις [40], [54], [47]. Η εξέλιξη αυτή δεν συνεπάγεται απαραίτητα πλήρως αυτόνομη παραγωγή εξελιγμένου malware, ωστόσο μειώνει σημαντικά τον χρόνο και τη δυσκολία ανάπτυξης επιθετικών εργαλείων.

Η επίδραση αυτή είναι ιδιαίτερα σημαντική σε περιπτώσεις επιτιθέμενων με περιορισμένη τεχνική κατάρτιση. Παραδοσιακά, η δημιουργία κακόβουλου λογισμικού αποτελούσε διαδικασία που απαιτούσε προηγμένες γνώσεις reverse engineering, exploitation και low-level programming. Με τη χρήση AI-assisted εργαλείων, μεγάλο μέρος αυτής της τεχνικής πολυπλοκότητας μπορεί να απλοποιηθεί μέσω αυτοματοποιημένης παραγωγής κώδικα, debugging assistance και προσαρμογής υπαρχόντων payloads [40], [39], [41]. Ως αποτέλεσμα, παρατηρείται μείωση του τεχνικού φραγμού εισόδου στο πεδίο της ανάπτυξης malware.

Παράλληλα, η τεχνητή νοημοσύνη επιτρέπει την επιτάχυνση της διαδικασίας δημιουργίας malware variants. Παραδοσιακά, η ανάπτυξη πολλαπλών εκδόσεων ενός κακόβουλου λογισμικού

απαιτούσε χειροκίνητες αλλαγές στον κώδικα, ώστε να αποφεύγεται η αναγνώριση από antivirus και intrusion detection systems. Η generative AI επιτρέπει πλέον τη δυναμική παραγωγή παραλλαγών κακόβουλου λογισμικού, μέσω αυτοματοποιημένων αλλαγών στη δομή, στη σύνταξη και στη συμπεριφορά του κώδικα [60], [52], [62]. Η δυνατότητα αυτή οδηγεί στη δημιουργία polymorphic και adaptive malware, τα οποία μπορούν να μεταβάλλουν συνεχώς χαρακτηριστικά τους, δυσχεραίνοντας σημαντικά την ανίχνευσή τους.

Η έννοια του polymorphic malware αποκτά ιδιαίτερη σημασία στο σύγχρονο περιβάλλον κυβερνοασφάλειας. Τα polymorphic malware χρησιμοποιούν τεχνικές μετασχηματισμού του κώδικα ώστε κάθε νέα εκτέλεση ή διανομή να εμφανίζει διαφορετικά χαρακτηριστικά, χωρίς να μεταβάλλεται η βασική λειτουργικότητα του κακόβουλου λογισμικού. Η τεχνητή νοημοσύνη ενισχύει σημαντικά αυτή τη δυνατότητα, καθώς μπορεί να παράγει δυναμικά νέες μορφές κώδικα, να αναδιαμορφώνει payloads και να τροποποιεί μοτίβα συμπεριφοράς αυτόματα [52], [62]. Ως αποτέλεσμα, τα συμβατικά signature-based συστήματα ανίχνευσης καθίστανται λιγότερο αποτελεσματικά.

Επιπλέον, η AI συμβάλλει στην ενίσχυση των δυνατοτήτων evasion των σύγχρονων malware. Παραδοσιακά, οι τεχνικές αποφυγής ανίχνευσης περιλάμβαναν obfuscation, packing και encryption του κώδικα. Η ενσωμάτωση AI επιτρέπει πλέον περισσότερο προσαρμοστικές μορφές evasion, κατά τις οποίες το κακόβουλο λογισμικό μπορεί να τροποποιεί δυναμικά τη συμπεριφορά του ανάλογα με το περιβάλλον εκτέλεσης ή τους μηχανισμούς άμυνας που εντοπίζει [40], [60], [52]. Η δυνατότητα αυτή μετατρέπει το malware από στατικό εργαλείο σε περισσότερο προσαρμοστικό επιθετικό σύστημα.

Η ανάπτυξη adaptive malware αποτελεί μία από τις σημαντικότερες μεταβολές που επιφέρει η AI στο πεδίο των κυβερνοεπιθέσεων. Σε αντίθεση με το παραδοσιακό malware, το οποίο εκτελούσε προκαθορισμένες ενέργειες, τα AI-assisted malware μπορούν να προσαρμόζουν ορισμένες λειτουργίες τους βάσει δεδομένων που συλλέγουν από το περιβάλλον στόχο [60], [52], [42]. Η προσαρμοστικότητα αυτή μπορεί να αφορά την επιλογή persistence mechanisms, τη χρονική στιγμή ενεργοποίησης payloads, την επιλογή τεχνικών lateral movement ή τη συμπεριφορά απέναντι σε αμυντικούς μηχανισμούς.

Παράλληλα, τα μεγάλα γλωσσικά μοντέλα ενισχύουν τη δυνατότητα modular malware development. Σύμφωνα με σύγχρονες ερευνητικές προσεγγίσεις, η παραγωγή κακόβουλου λογισμικού μπορεί να διαμοιραστεί σε επιμέρους υπο-εργασίες, οι οποίες ανατίθενται σε διαφορετικά στάδια AI-assisted code generation [61]. Με τον τρόπο αυτό, η διαδικασία δημιουργίας malware γίνεται περισσότερο οργανωμένη, αυτοματοποιημένη και επεκτάσιμη. Η κατανομή της ανάπτυξης σε μικρότερα τμήματα δυσχεραίνει επίσης τον εντοπισμό κακόβουλης δραστηριότητας κατά τη διάρκεια παραγωγής του κώδικα.

Η AI επηρεάζει επίσης τη διαδικασία vulnerability exploitation που συνδέεται με την ανάπτυξη malware. Μέσω τεχνικών machine learning και automated vulnerability analysis, τα σύγχρονα συστήματα μπορούν να εντοπίζουν αδύναμα σημεία λογισμικού και να προτείνουν τρόπους αξιοποίησής τους [54], [47], [42]. Η δυνατότητα αυτή επιταχύνει τον κύκλο ανάπτυξης exploit-based malware και μειώνει το χρονικό διάστημα μεταξύ ανακάλυψης ευπάθειας και δημιουργίας λειτουργικού κακόβουλου εργαλείου.

Ιδιαίτερη σημασία παρουσιάζει και η διάδοση malicious AI tools στο underground οικοσύστημα κυβερνοεγκλήματος. Σύγχρονες αναφορές threat intelligence δείχνουν ότι εργαλεία AI-assisted malware creation, automated scripting και vulnerability research διατίθενται πλέον μέσω παράνομων forums και criminal-as-a-service υποδομών [39]. Η εξέλιξη αυτή ενισχύει τη βιομηχανοποίηση του κυβερνοεγκλήματος, καθώς κακόβουλοι χρήστες μπορούν να αποκτήσουν πρόσβαση σε προηγμένες επιθετικές δυνατότητες χωρίς να διαθέτουν υψηλό επίπεδο τεχνικής γνώσης.

Πραγματικά περιστατικά misuse τεχνητής νοημοσύνης δείχνουν ότι LLMs έχουν χρησιμοποιηθεί για υποστήριξη scripting, debugging, automation και δημιουργία επιθετικών εργαλείων στο πλαίσιο κυβερνοεπιχειρήσεων [41]. Παρότι τα περισσότερα προηγμένα AI συστήματα ενσωματώνουν μηχανισμούς ασφαλείας, η συνεχής εξέλιξη τεχνικών bypass και prompt manipulation δημιουργεί ανησυχίες σχετικά με τη δυνατότητα κακόβουλης αξιοποίησής τους [40], [41], [50].

Παράλληλα, η χρήση AI στην παραγωγή malware επηρεάζει σημαντικά την επιχειρησιακή αποτελεσματικότητα των επιθέσεων. Η δυνατότητα ταχείας ανάπτυξης, προσαρμογής και επαναχρησιμοποίησης κακόβουλου λογισμικού επιτρέπει στους επιτιθέμενους να αυξάνουν τον

ρυθμό επιθέσεων, να μειώνουν το κόστος ανάπτυξης νέων εργαλείων και να προσαρμόζονται ταχύτερα στις μεταβολές του αμυντικού περιβάλλοντος [40], [39], [42].

Η εξέλιξη αυτή ενισχύει και τη δυναμική των οργανωμένων εγκληματικών δικτύων. Η παραγωγή malware μετατρέπεται σταδιακά από δραστηριότητα υψηλής τεχνικής εξειδίκευσης σε περισσότερο αυτοματοποιημένη διαδικασία που μπορεί να υποστηριχθεί από AI-enabled platforms και cybercrime ecosystems [39], [42]. Το γεγονός αυτό διευρύνει τον αριθμό πιθανών επιτιθέμενων και αυξάνει τη συνολική κλίμακα των κυβερνοαπειλών.

Η Τεχνητή Νοημοσύνη μεταβάλλει ουσιαστικά τον τρόπο ανάπτυξης κακόβουλου λογισμικού, επιτρέποντας μεγαλύτερη αυτοματοποίηση, αυξημένη προσαρμοστικότητα και ταχύτερη παραγωγή malware variants. Μέσω AI-assisted code generation, polymorphic malware, adaptive evasion techniques και automated vulnerability exploitation, οι επιτιθέμενοι αποκτούν δυνατότητες που αυξάνουν σημαντικά την αποτελεσματικότητα και τη δυναμική των κυβερνοεπιθέσεων. Η μετάβαση από το παραδοσιακό στατικό malware σε περισσότερο ευφυή και προσαρμοστικά κακόβουλα συστήματα αποτελεί μία από τις σημαντικότερες μεταβολές που επιφέρει η τεχνητή νοημοσύνη στο σύγχρονο περιβάλλον κυβερνοασφάλειας [40], [60], [52], [62], [42].

3.6. Social engineering με AI

Το Social engineering αποτελεί μία από τις πιο κρίσιμες μορφές κυβερνοεπίθεσης, καθώς δεν στοχεύει πρωτίστως την τεχνική αδυναμία ενός πληροφοριακού συστήματος, αλλά την ανθρώπινη κρίση, την εμπιστοσύνη και τη συμπεριφορά του χρήστη. Σε αντίθεση με επιθέσεις που βασίζονται αποκλειστικά στην εκμετάλλευση τεχνικών ευπαθειών, οι επιθέσεις social engineering επιδιώκουν να πείσουν το θύμα να εκτελέσει μία ενέργεια προς όφελος του επιτιθέμενου, όπως η αποκάλυψη διαπιστευτηρίων, η λήψη κακόβουλου αρχείου, η έγκριση οικονομικής συναλλαγής ή η παράκαμψη εσωτερικών διαδικασιών ασφαλείας. Η ενσωμάτωση της τεχνητής νοημοσύνης στο πεδίο αυτό μεταβάλλει ουσιαστικά τη φύση των επιθέσεων, διότι ενισχύει την πειστικότητα, την εξατομίκευση, την ταχύτητα παραγωγής και τη δυσκολία αναγνώρισής τους [63], [53], [21].

Παραδοσιακά, επιθέσεις όπως το phishing, το spear-phishing και το Business Email Compromise βασίζονταν συχνά σε μαζικά ή ημι-τυποποιημένα μηνύματα, τα οποία παρουσίαζαν εμφανείς αδυναμίες. Γλωσσικά λάθη, ασυνήθιστη σύνταξη, γενικόλογες προσφωνήσεις, ακατάλληλο ύφος και μη πειστική μίμηση οργανισμών αποτελούσαν συχνά ενδείξεις απάτης. Αν και τέτοιες επιθέσεις εξακολουθούσαν να είναι αποτελεσματικές, η ποιότητά τους περιοριζόταν από τις γλωσσικές ικανότητες, τον διαθέσιμο χρόνο και την επιχειρησιακή εμπειρία του επιτιθέμενου [53], [64], [67].

Η generative AI μεταβάλλει αυτή τη συνθήκη, καθώς επιτρέπει τη δημιουργία κειμένων με υψηλή γλωσσική αρτιότητα, συνοχή και προσαρμογή στο επικοινωνιακό πλαίσιο του θύματος. Τα μεγάλα γλωσσικά μοντέλα μπορούν να παράγουν μηνύματα που εμφανίζονται φυσικά, επαγγελματικά και συμβατά με το ύφος ενός οργανισμού, ενός στελέχους ή ενός συνεργάτη. Έτσι, ένα phishing μήνυμα μπορεί να μην παρουσιάζει πλέον τα παραδοσιακά εμφανή σημάδια απάτης, γεγονός που δυσχεραίνει σημαντικά την αναγνώρισή του από τον τελικό χρήστη [63], [53], [64].

Ιδιαίτερη σημασία έχει η δυνατότητα εξατομίκευσης. Η τεχνητή νοημοσύνη μπορεί να αξιοποιήσει πληροφορίες από κοινωνικά δίκτυα, δημόσιες ιστοσελίδες, επαγγελματικά προφίλ, διαρροές δεδομένων και ανοικτές πηγές, προκειμένου να δημιουργήσει στοχευμένα μηνύματα προσαρμοσμένα στο επαγγελματικό περιβάλλον, στις σχέσεις, στη γλώσσα και στις συνήθειες του θύματος. Με τον τρόπο αυτό, η επίθεση δεν εμφανίζεται ως γενικό και απρόσωπο μήνυμα, αλλά ως επικοινωνία που φαίνεται να σχετίζεται με πραγματικές δραστηριότητες, πρόσωπα ή ανάγκες του παραλήπτη [63], [53], [68].

Η μεταβολή αυτή είναι ιδιαίτερα εμφανής στις επιθέσεις spear-phishing. Η χρήση generative AI επιτρέπει τη δημιουργία μηνυμάτων που μιμούνται το ύφος γραφής ενός συναδέλφου, ενός προϊσταμένου ή ενός αξιόπιστου οργανισμού, αυξάνοντας σημαντικά την πιθανότητα ανταπόκρισης του θύματος. Στο πλαίσιο αυτό, η επιτυχία της επίθεσης δεν βασίζεται μόνο στη μαζική αποστολή μηνυμάτων, αλλά στην ψυχολογική ακρίβεια με την οποία το περιεχόμενο προσαρμόζεται στο προφίλ του αποδέκτη [53], [64], [68].

Αντίστοιχα, οι επιθέσεις Business Email Compromise αποκτούν νέα δυναμική μέσω της τεχνητής νοημοσύνης. Σε τέτοιες επιθέσεις, ο επιτιθέμενος επιχειρεί να υποδυθεί στέλεχος, συνεργάτη ή αξιόπιστο πρόσωπο, με σκοπό να προκαλέσει οικονομική συναλλαγή, αποστολή ευαίσθητων

δεδομένων ή αλλαγή διαδικασιών πληρωμής. Η τεχνητή νοημοσύνη ενισχύει τις επιθέσεις αυτές, καθώς μπορεί να παράγει μηνύματα που αντανακλούν το ύφος, την ορολογία και το επικοινωνιακό μοτίβο ενός πραγματικού προσώπου, καθιστώντας την πλαστοπροσωπία περισσότερο πειστική [32], [64], [67].

Παράλληλα, η χρήση τεχνητής νοημοσύνης δεν περιορίζεται στο γραπτό περιεχόμενο. Η ανάπτυξη deepfake τεχνολογιών, συνθετικής φωνής και ψεύτικων οπτικοακουστικών μέσων έχει επεκτείνει το social engineering σε πιο σύνθετες μορφές πλαστοπροσωπίας. Οι επιτιθέμενοι μπορούν να αξιοποιούν τεχνολογίες voice cloning, audio spoofing και synthetic media για να μιμηθούν τη φωνή ή την εικόνα στελεχών, συναδέλφων ή αξιόπιστων εκπροσώπων οργανισμών [32], [21], [65], [66].

Η χρήση συνθετικής φωνής είναι ιδιαίτερα κρίσιμη, διότι αξιοποιεί ένα ισχυρό στοιχείο ανθρώπινης εμπιστοσύνης, την αναγνώριση της φωνής. Σε περιβάλλοντα όπου οι εργαζόμενοι έχουν συνηθίσει να λαμβάνουν οδηγίες μέσω τηλεφωνικών ή διαδικτυακών επικοινωνιών, μία πειστική φωνητική απομίμηση μπορεί να ενισχύσει σημαντικά την αποτελεσματικότητα μιας απάτης. Με τον τρόπο αυτό, το social engineering μεταβαίνει από την απλή εξαπάτηση μέσω κειμένου σε πολυτροπικές μορφές χειραγώγησης, όπου συνδυάζονται κείμενο, ήχος, εικόνα και βίντεο [32], [65], [66].

Η τεχνητή νοημοσύνη επιτρέπει επίσης τη μαζική παραγωγή προσωποποιημένων επιθέσεων. Παραδοσιακά, υπήρχε σαφής διάκριση ανάμεσα στις μαζικές phishing επιθέσεις και στις στοχευμένες επιθέσεις υψηλής εξατομίκευσης. Οι πρώτες ήταν κλιμακούμενες αλλά λιγότερο πειστικές, ενώ οι δεύτερες ήταν πιο αποτελεσματικές αλλά απαιτούσαν σημαντικό χρόνο και ανθρώπινη προσπάθεια. Η generative AI περιορίζει αυτή τη διάκριση, επιτρέποντας την ταυτόχρονη παραγωγή μεγάλου αριθμού προσωποποιημένων μηνυμάτων [63], [27], [67], [44].

Η εξέλιξη αυτή έχει ιδιαίτερη σημασία για την κλιμάκωση των επιθέσεων. Με τη χρήση AI, ένας επιτιθέμενος μπορεί να δημιουργήσει διαφορετικές εκδοχές ενός μηνύματος ανάλογα με τη γλώσσα, τον ρόλο, την επαγγελματική ιδιότητα, την οργανωτική θέση ή τα ενδιαφέροντα του θύματος. Έτσι, η επίθεση αποκτά χαρακτηριστικά μαζικότητας χωρίς να χάνει την αίσθηση προσωπικής στόχευσης. Το αποτέλεσμα είναι η αύξηση τόσο του όγκου όσο και της ποιότητας των social engineering επιθέσεων [27], [67], [43].

Το social engineering με χρήση τεχνητής νοημοσύνης συνδέεται άμεσα και με τη μείωση του επιχειρησιακού κόστους των επιθέσεων. Η παραγωγή πειστικών μηνυμάτων, η μετάφραση σε διαφορετικές γλώσσες, η προσαρμογή του ύφους και η δημιουργία πολλαπλών παραλλαγών μπορούν να πραγματοποιούνται ταχύτερα και με μικρότερη ανθρώπινη παρέμβαση. Κατά συνέπεια, ομάδες με περιορισμένους πόρους μπορούν να εκτελούν επιθέσεις υψηλότερης ποιότητας και μεγαλύτερης κλίμακας [63], [27], [44].

Επιπλέον, η τεχνητή νοημοσύνη ενισχύει την πειστικότητα των επιθέσεων μέσω της δυνατότητας ανάλυσης του επικοινωνιακού πλαισίου. Ένα ΑΙ σύστημα μπορεί να προσαρμόσει ένα μήνυμα ώστε να φαίνεται επείγον, τυπικό, φιλικό ή επαγγελματικό, ανάλογα με τον ρόλο του προσώπου που υποδύεται ο επιτιθέμενος και το προφίλ του θύματος. Η προσαρμογή αυτή αυξάνει τη δυνατότητα ψυχολογικής επιρροής, καθώς το μήνυμα δεν λειτουργεί απλώς ως δόλωμα, αλλά ως στοχευμένη επικοινωνιακή παρέμβαση [63], [53], [68].

Η χρήση deepfakes και συνθετικών μέσων ενισχύει ακόμη περισσότερο τη δυσκολία ανίχνευσης. Ενώ ένα απλό phishing email μπορεί να αξιολογηθεί με βάση γλωσσικά και τεχνικά χαρακτηριστικά, ένα πειστικό ηχητικό ή βιντεοσκοπημένο μήνυμα που φαίνεται να προέρχεται από πραγματικό πρόσωπο δημιουργεί πιο σύνθετες συνθήκες αξιολόγησης. Το θύμα καλείται να αμφισβητήσει όχι μόνο ένα κείμενο, αλλά την ίδια την αντιληπτή αυθεντικότητα της φωνής ή της εικόνας [32], [21], [65], [66].

Η εξέλιξη αυτή έχει σοβαρές συνέπειες για οργανισμούς και επιχειρήσεις. Οι διαδικασίες εσωτερικής επικοινωνίας, έγκρισης πληρωμών, αλλαγής κωδικών, επαναφοράς λογαριασμών και διαχείρισης ευαίσθητων δεδομένων μπορούν να αποτελέσουν στόχο AI-enabled social engineering. Αναφορές κυβερνοασφάλειας επισημαίνουν ότι οι επιτιθέμενοι αξιοποιούν τεχνικές πλαστοπροσωπίας και χειραγώγησης για να παρακάμπτουν μηχανισμούς ελέγχου, να αποκτούν πρόσβαση σε λογαριασμούς ή να προκαλούν οικονομικές απώλειες [32], [67], [68].

Παράλληλα, η τεχνητή νοημοσύνη καθιστά δυσκολότερη τη διάκριση μεταξύ αυθεντικής και κακόβουλης επικοινωνίας. Όσο αυξάνεται η ποιότητα του παραγόμενου περιεχομένου, τόσο λιγότερο αποτελεσματικές γίνονται οι παραδοσιακές συμβουλές αναγνώρισης phishing, όπως ο εντοπισμός ορθογραφικών λαθών ή ασυνήθιστης σύνταξης. Η εκπαίδευση των χρηστών παραμένει αναγκαία, αλλά πρέπει να προσαρμοστεί σε ένα περιβάλλον όπου τα κακόβουλα

μήνυματα μπορεί να είναι γλωσσικά άρτια, αισθητικά πειστικά και επικοινωνιακά αληθοφανή [53], [64], [67].

Από στρατηγική άποψη, το social engineering με χρήση τεχνητής νοημοσύνης δημιουργεί ένα περιβάλλον ασύμμετρης απειλής. Ο επιτιθέμενος μπορεί να χρησιμοποιήσει εργαλεία χαμηλού κόστους για να παράγει μεγάλο αριθμό πειστικών επιθέσεων, ενώ ο αμυνόμενος οργανισμός πρέπει να προστατεύσει πολλαπλά σημεία ανθρώπινης αλληλεπίδρασης. Η ασυμμετρία αυτή ενισχύεται από το γεγονός ότι ακόμη και ένα μόνο επιτυχές περιστατικό εξαπάτησης μπορεί να οδηγήσει σε παραβίαση λογαριασμών, οικονομική απώλεια ή διαρροή δεδομένων [27], [44], [43].

Η ανάλυση του φαινομένου δείχνει ότι η βασική μεταβολή δεν είναι απλώς η αυτοματοποίηση του social engineering. Η ουσιαστική αλλαγή είναι η μετάβαση σε μια πιο εξελιγμένη μορφή ψυχολογικής χειραγώγησης, όπου η τεχνική ισχύς της τεχνητής νοημοσύνης συνδυάζεται με την εκμετάλλευση της ανθρώπινης εμπιστοσύνης. Η τεχνητή νοημοσύνη επιτρέπει στους επιτιθέμενους να κατανοούν καλύτερα το περιβάλλον του θύματος, να προσαρμόζουν το μήνυμα, να μιμούνται αξιόπιστες ταυτότητες και να δημιουργούν περιεχόμενο που ενεργοποιεί συγκεκριμένες συμπεριφορικές αντιδράσεις [63], [53], [68].

Η ΑΙ επαναπροσδιορίζει το social engineering. Οι επιθέσεις γίνονται πιο πειστικές, περισσότερο προσωποποιημένες, γλωσσικά άρτιες, πολυτροπικές και δυσκολότερα αναγνωρίσιμες. Μέσω generative AI, deepfakes, synthetic voice, AI-powered phishing και αυτοματοποιημένης παραγωγής περιεχομένου, οι επιτιθέμενοι αποκτούν τη δυνατότητα να συνδυάζουν μαζικότητα και ακρίβεια στόχευσης. Η εξέλιξη αυτή καθιστά το social engineering έναν από τους πιο κρίσιμους τομείς στους οποίους η τεχνητή νοημοσύνη μεταβάλλει τη δυναμική των κυβερνοεπιθέσεων, καθώς η ανθρώπινη εμπιστοσύνη μετατρέπεται σε ακόμη πιο ευάλωτο σημείο του ψηφιακού οικοσυστήματος [63], [32], [21], [67], [68].

4. ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΩΝ ΒΑΣΕΙ ΠΟΡΩΝ ΚΑΙ ΠΟΛΥΠΛΟΚΟΤΗΤΑΣ

4.1. Επιθέσεις χαμηλού κόστους vs επιθέσεις υψηλής πολυπλοκότητας (Low-cost vs high-complexity attacks)

4.1.1. Επιθέσεις χαμηλού κόστους (Low-cost attacks)

Η ανάλυση των κυβερνοεπιθέσεων με βάση το κόστος και την πολυπλοκότητά τους επιτρέπει την ουσιαστικότερη κατανόηση της σύγχρονης δυναμικής του κυβερνοεγκλήματος. Οι κυβερνοεπιθέσεις δεν αποτελούν ενιαία και ομοιογενή κατηγορία ενεργειών, αλλά εντάσσονται σε ένα ευρύτερο φάσμα, το οποίο εκτείνεται από απλές, χαμηλού κόστους και εύκολα επαναλήψιμες επιθέσεις έως σύνθετες, πολλών σταδίων και στρατηγικά στοχευμένες επιχειρήσεις. Η διάκριση αυτή είναι ιδιαίτερα σημαντική, καθώς αναδεικνύει ότι η επικινδυνότητα μιας επίθεσης δεν εξαρτάται αποκλειστικά από την τεχνική της πολυπλοκότητα, αλλά και από την κλίμακα, τη συχνότητα, την προσβασιμότητα των εργαλείων και τον αντίκτυπο της [69], [70], [50].

Για την ταξινόμηση μιας κυβερνοεπίθεσης μπορούν να χρησιμοποιηθούν πολλαπλά κριτήρια. Τα σημαντικότερα από αυτά είναι το οικονομικό κόστος υλοποίησης, το επίπεδο τεχνικής εξειδίκευσης που απαιτείται, ο χρόνος προετοιμασίας, οι υπολογιστικοί πόροι, η ανάγκη για εξειδικευμένη υποδομή, ο βαθμός αυτοματοποίησης, η δυνατότητα επανάληψης και ο πιθανός αντίκτυπος στον στόχο. Με βάση αυτά τα κριτήρια, μια επίθεση χαμηλού κόστους δεν είναι απαραίτητα ασήμαντη, όπως και μια επίθεση υψηλής πολυπλοκότητας δεν είναι πάντοτε μαζική. Αντιθέτως, η σχέση μεταξύ κόστους, πολυπλοκότητας και αποτελεσματικότητας είναι δυναμική και επηρεάζεται σημαντικά από την τεχνητή νοημοσύνη [69], [74], [76].

Οι επιθέσεις χαμηλού κόστους χαρακτηρίζονται συνήθως από περιορισμένη ανάγκη τεχνικής εξειδίκευσης, χρήση έτοιμων εργαλείων, χαμηλό χρόνο προετοιμασίας και δυνατότητα μαζικής επανάληψης. Σε αυτή την κατηγορία μπορούν να ενταχθούν επιθέσεις phishing, credential stuffing, βασικές μορφές κακόβουλου λογισμικού, αυτοματοποιημένα spam campaigns, χρήση phishing kits, DDoS-as-a-service υπηρεσίες και επιθέσεις που βασίζονται σε ήδη διαθέσιμες υποδομές κυβερνοεγκλήματος [70], [71], [72].

Η ανάπτυξη του cybercrime-as-a-service έχει ενισχύσει σημαντικά αυτή την κατηγορία επιθέσεων. Μέσω τέτοιων μοντέλων, ο επιτιθέμενος δεν χρειάζεται να αναπτύξει ο ίδιος το σύνολο της τεχνικής υποδομής ή των εργαλείων που απαιτούνται. Αντίθετα, μπορεί να αποκτήσει πρόσβαση σε έτοιμες υπηρεσίες, όπως phishing kits, malware deployment tools, credential marketplaces ή DDoS υποδομές. Με τον τρόπο αυτό, το τεχνικό βάρος μεταφέρεται από τον τελικό χρήστη της επίθεσης στον πάροχο της παράνομης υπηρεσίας, μειώνοντας σημαντικά τη δυσκολία εισόδου [70], [72].

Ιδιαίτερα χαρακτηριστικό παράδειγμα χαμηλού κόστους επίθεσης είναι το phishing. Η χρήση έτοιμων phishing kits επιτρέπει την ταχεία δημιουργία παραπλανητικών ιστοσελίδων, φορμών συλλογής διαπιστευτηρίων και μαζικών μηνυμάτων εξαπάτησης, χωρίς να απαιτείται υψηλό επίπεδο προγραμματιστικής γνώσης. Η οικονομία των phishing kits δείχνει ότι η εμπορευματοποίηση εργαλείων επίθεσης επιτρέπει σε μεγάλο αριθμό επιτιθέμενων να πραγματοποιούν επιθέσεις με ευκολία και μικρό κόστος [71].

Η τεχνητή νοημοσύνη ενισχύει ακόμη περισσότερο τις επιθέσεις χαμηλού κόστους, διότι μειώνει τον χρόνο προετοιμασίας, βελτιώνει την ποιότητα των μηνυμάτων και επιτρέπει την ταχύτερη προσαρμογή τους σε διαφορετικά θύματα. Για παράδειγμα, ένα phishing campaign που στο παρελθόν απαιτούσε χειροκίνητη σύνταξη μηνυμάτων, μετάφραση, διόρθωση και προσαρμογή στο εκάστοτε θύμα μπορεί πλέον να υποστηριχθεί από AI εργαλεία. Έτσι, η επίθεση παραμένει χαμηλού κόστους, αλλά αποκτά υψηλότερη πειστικότητα και αυτό συνεπάγεται μεγαλύτερη πιθανότητα επιτυχίας [69], [50], [74].

Συνεπώς, μια κυβερνοεπίθεση μπορεί να είναι χαμηλού κόστους αλλά υψηλού αντίκτυπου. Η χαμηλή δαπάνη υλοποίησης δεν συνεπάγεται χαμηλή επικινδυνότητα. [71], [73].

4.1.2. Επιθέσεις υψηλής πολυπλοκότητας (High-complexity attacks)

Από την άλλη πλευρά, οι επιθέσεις υψηλής πολυπλοκότητας χαρακτηρίζονται από αυξημένες απαιτήσεις σε τεχνική γνώση, χρόνο προετοιμασίας, αναλυτικό σχεδιασμό και εξειδικευμένα εργαλεία. Σε αυτή την κατηγορία εντάσσονται στοχευμένες εισβολές, πολυσταδιακές επιθέσεις, προηγμένο κακόβουλο λογισμικό, τεχνικές αποφυγής ανίχνευσης, lateral movement, persistence,

χρήση άγνωστων ή πρόσφατα αποκαλυφθέντων ευπαθειών και επιχειρήσεις που προσομοιάζουν σε Advanced Persistent Threats [75], [35], [76].

Οι επιθέσεις υψηλής πολυπλοκότητας αποσκοπούν συνήθως στην επίτευξη ενός συγκεκριμένου στρατηγικού αποτελέσματος. Ο επιτιθέμενος επενδύει περισσότερο χρόνο στην αναγνώριση του στόχου, στη μελέτη της υποδομής, στην απόκτηση αρχικής πρόσβασης, στη διατήρηση παρουσίας στο σύστημα και στην αποφυγή ανίχνευσης. Η επιτυχία τέτοιων επιθέσεων εξαρτάται από τη δυνατότητα συνδυασμού πολλών τεχνικών σε μία αλληλένδετη επιχειρησιακή αλυσίδα [75], [35].

Σε αντίθεση με τις χαμηλού κόστους επιθέσεις, οι επιθέσεις υψηλής πολυπλοκότητας απαιτούν συνήθως οργανωμένη υποστήριξη. Η ανάπτυξη custom malware, η επιλογή τεχνικών evasion, η αξιοποίηση ευπαθειών και η πολυσταδιακή διείσδυση σε οργανωμένα περιβάλλοντα προϋποθέτουν εξειδικευμένες δεξιότητες και συχνά καταμερισμό εργασιών. Σε ορισμένες περιπτώσεις, διαφορετικά στάδια της επίθεσης μπορούν να ανατίθενται σε διαφορετικούς δρώντες ή ομάδες, γεγονός που αυξάνει την επιχειρησιακή πολυπλοκότητα αλλά και την αποτελεσματικότητα της επίθεσης [75], [35].

Η τεχνητή νοημοσύνη μεταβάλλει τα όρια μεταξύ των δύο αυτών κατηγοριών. Η σημαντικότερη αλλαγή δεν είναι ότι εξαφανίζεται η διαφορά μεταξύ απλών και σύνθετων επιθέσεων, αλλά ότι ορισμένες τεχνικές που παλαιότερα απαιτούσαν υψηλό επίπεδο γνώσεων καθίστανται περισσότερο προσβάσιμες. Η ΑΙ μπορεί να υποστηρίξει τη συγγραφή κώδικα, την ανάλυση ευπαθειών, την παραγωγή επιθετικού περιεχομένου, την αυτοματοποίηση αναγνώρισης στόχων και τη βελτίωση της επιχειρησιακής λήψης αποφάσεων [69], [50], [74].

Με τον τρόπο αυτό, η τεχνητή νοημοσύνη λειτουργεί ως παράγοντας σύγκλισης. Από τη μία πλευρά, ενισχύει τους λιγότερο έμπειρους επιτιθέμενους, μειώνοντας το τεχνικό εμπόδιο εισόδου και επιτρέποντάς τους να εκτελούν ενέργειες που προηγουμένως απαιτούσαν μεγαλύτερη εμπειρία. Από την άλλη πλευρά, ενισχύει και τους προηγμένους επιτιθέμενους, αυξάνοντας την ταχύτητα, την ακρίβεια, την προσαρμοστικότητα και την επιχειρησιακή αποτελεσματικότητα των επιθέσεών τους [69], [50], [76].

Η αυτοματοποίηση της κυβερνοεπίθεσης έχει άμεση σχέση με το κόστος και την κλίμακα. Όσο περισσότερα επιμέρους στάδια αυτοματοποιούνται, τόσο μειώνεται το κόστος επανάληψης της

επίθεσης. Αυτό σημαίνει ότι ένας επιτιθέμενος μπορεί να δοκιμάσει μεγαλύτερο αριθμό στόχων, να παράγει περισσότερες παραλλαγές επίθεσης ή να διατηρεί ταυτόχρονα περισσότερες επιχειρησιακές γραμμές δράσης. Η ΑΙ, επομένως, δεν μειώνει μόνο την πολυπλοκότητα για τον χρήστη, αλλά αυξάνει και τη δυνατότητα επιχειρησιακής κλιμάκωσης [69], [72], [74].

Η διάκριση χαμηλού κόστους και υψηλής πολυπλοκότητας δεν είναι απόλυτη, υπάρχουν και ενδιάμεσες βαθμίδες. Μια επίθεση μπορεί να ξεκινά ως χαμηλού κόστους phishing campaign, αλλά να οδηγήσει σε ιδιαίτερα σύνθετη παραβίαση εάν οι επιτιθέμενοι αποκτήσουν διαπιστευτήρια πρόσβασης σε κρίσιμα συστήματα. Αντίστοιχα, μια προηγμένη επίθεση μπορεί να αξιοποιεί χαμηλού κόστους εργαλεία σε ορισμένα στάδια, όπως για την αρχική πρόσβαση ή τη μαζική συλλογή πληροφοριών [73], [75].

Αυτή η υβριδική φύση των σύγχρονων επιθέσεων καθιστά ανεπαρκή την απλή διάκριση μεταξύ “εύκολων” και “δύσκολων” επιθέσεων. Στην πράξη, οι επιτιθέμενοι συχνά συνδυάζουν χαμηλού κόστους τεχνικές με πιο σύνθετες επιχειρησιακές μεθόδους. Ένα phishing μήνυμα μπορεί να λειτουργήσει ως αρχικό σημείο εισόδου, ενώ στη συνέχεια η επίθεση μπορεί να εξελιχθεί σε lateral movement, privilege escalation, εγκατάσταση malware ή εξαγωγή δεδομένων. Επομένως, το χαμηλό κόστος του αρχικού σταδίου δεν μειώνει απαραίτητα τη σοβαρότητα της συνολικής επίθεσης [73], [75], [35].

Η τεχνητή νοημοσύνη ενισχύει αυτή την υβριδικότητα, καθώς επιτρέπει τη γρηγορότερη μετάβαση από απλές σε πιο σύνθετες φάσεις επίθεσης. Ένας επιτιθέμενος μπορεί να χρησιμοποιήσει τεχνητή νοημοσύνη για τη δημιουργία πειστικού αρχικού δολώματος, στη συνέχεια για την ανάλυση των αποτελεσμάτων, μετά για την υποστήριξη τεχνικών ενεργειών και τελικά για τη βελτιστοποίηση της επιθετικής στρατηγικής. Με αυτόν τον τρόπο, η τεχνητή νοημοσύνη δεν ανήκει αποκλειστικά ούτε στις χαμηλού κόστους ούτε στις υψηλής πολυπλοκότητας επιθέσεις, αλλά λειτουργεί σε όλο το φάσμα [69], [74], [76].

Η ταξινόμηση των κυβερνοεπιθέσεων με βάση το κόστος και την πολυπλοκότητα αναδεικνύει μια κρίσιμη μεταβολή στο σύγχρονο περιβάλλον απειλών. Οι επιθέσεις χαμηλού κόστους γίνονται πιο πειστικές, πιο μαζικές και δυνητικά πιο επιζήμιες, ενώ οι επιθέσεις υψηλής πολυπλοκότητας γίνονται ταχύτερες, περισσότερο προσαρμοστικές και επιχειρησιακά αποδοτικότερες. Η τεχνητή νοημοσύνη μειώνει τα εμπόδια εισόδου για λιγότερο εξειδικευμένους επιτιθέμενους, ενισχύοντας

ταυτόχρονα τις δυνατότητες των πιο εξειδικευμένων χρηστών. Έτσι, η σχέση μεταξύ κόστους, πολυπλοκότητας και αντικτύπου καθίσταται πιο ρευστή, γεγονός που καθιστά αναγκαία μια πιο σύνθετη και δυναμική προσέγγιση στην αξιολόγηση των κυβερνοαπειλών [69], [70], [50], [76].

4.2. Ρόλος υπολογιστικών πόρων και πώς η διαθεσιμότητα πόρων επηρεάζει τη δυναμική του επιτιθέμενου

Η πραγματοποίηση μιας κυβερνοεπίθεσης δεν εξαρτάται αποκλειστικά από την τεχνική ιδέα ή από τη γνώση του επιτιθέμενου, αλλά και από το σύνολο των πόρων που έχει στη διάθεσή του. Οι πόροι αυτοί δεν περιορίζονται στην υπολογιστική ισχύ ή στο υλικό, αλλά περιλαμβάνουν την πρόσβαση σε δεδομένα, cloud υποδομές, αυτοματοποιημένα εργαλεία, οικονομικούς μηχανισμούς, ανθρώπινο δυναμικό, τεχνική εξειδίκευση, έτοιμες εγκληματικές υπηρεσίες και συστήματα τεχνητής νοημοσύνης. Κατά συνέπεια, η δυναμική του επιτιθέμενου διαμορφώνεται από τη σχέση ανάμεσα στους διαθέσιμους πόρους και στις επιθετικές δυνατότητες που αυτοί επιτρέπουν [70], [83].

Στο παραδοσιακό περιβάλλον κυβερνοεπιθέσεων, η διαθεσιμότητα πόρων λειτουργούσε ως σημαντικό εμπόδιο εισόδου. Οι προηγμένες επιθέσεις απαιτούσαν εξειδικευμένο προσωπικό, ιδιόκτητες υποδομές, χρόνο προετοιμασίας, τεχνική εμπειρία και συχνά οργανωμένη επιχειρησιακή υποστήριξη. Η ανάπτυξη σύνθετων επιθετικών εργαλείων, η διαχείριση botnets, η εκτέλεση εκτεταμένων DDoS επιθέσεων, η διείσδυση σε cloud περιβάλλοντα ή η πραγματοποίηση πολυσταδιακών επιθέσεων απαιτούσαν πόρους που δεν ήταν διαθέσιμοι σε όλους τους επιτιθέμενους [77], [78], [84].

Η ανάπτυξη όμως υπηρεσιών cloud, αυτοματοποιημένων εργαλείων και μοντέλων cybercrime-as-a-service έχει μεταβάλει ουσιαστικά αυτή την ισορροπία. Ο επιτιθέμενος δεν χρειάζεται πλέον να διαθέτει πάντοτε δική του υποδομή ή υψηλή τεχνική εξειδίκευση, καθώς μπορεί να αποκτήσει πρόσβαση σε έτοιμες επιθετικές δυνατότητες μέσω παράνομων αγορών. Με τον τρόπο αυτό, οι πόροι παύουν να αποτελούν αποκλειστικά ιδιόκτητο κεφάλαιο του επιτιθέμενου και μετατρέπονται σε διαθέσιμη υπηρεσία [70], [82].

Η τεχνητή νοημοσύνη ενισχύει ακόμη περισσότερο αυτή τη μεταβολή. Η πρόσβαση σε AI εργαλεία, pretrained models και agentic workflows επιτρέπει την αυτοματοποίηση επιμέρους σταδίων της επιθετικής διαδικασίας, όπως η αναγνώριση στόχων, η ανάλυση δεδομένων, η παραγωγή κώδικα, η αξιολόγηση αποτελεσμάτων και η επιλογή επόμενων ενεργειών. Έτσι, η AI λειτουργεί και ως μηχανισμός μετατροπής περιορισμένων πόρων σε αυξημένη επιχειρησιακή ικανότητα [79], [80], [83].

Ιδιαίτερα σημαντικός είναι ο ρόλος των cloud υποδομών. Το cloud computing επιτρέπει την πρόσβαση σε υπολογιστική ισχύ, αποθηκευτικό χώρο, αυτοματοποιημένες υπηρεσίες και δικτυακές δυνατότητες χωρίς την ανάγκη ανάπτυξης φυσικής υποδομής από τον ίδιο τον επιτιθέμενο. Η δυνατότητα αυτή μπορεί να αξιοποιηθεί είτε άμεσα, μέσω κατάχρησης cloud λογαριασμών και υπηρεσιών, είτε έμμεσα, μέσω εργαλείων και επιθετικών πλατφορμών που βασίζονται σε cloud-native πόρους [81], [82].

Η διαθεσιμότητα τέτοιων πόρων επηρεάζει άμεσα την κλίμακα των επιθέσεων. Όταν ένας επιτιθέμενος έχει πρόσβαση σε μεγάλη υπολογιστική και δικτυακή ισχύ, μπορεί να αυξήσει τον αριθμό των στόχων, την ταχύτητα εκτέλεσης και τον όγκο της επιθετικής δραστηριότητας. Αυτό είναι ιδιαίτερα εμφανές σε επιθέσεις DDoS, όπου η συγκεντρωμένη ισχύς botnets, μολυσμένων συσκευών και δικτυακών υποδομών μετατρέπεται σε επιχειρησιακό πλεονέκτημα. Σε τέτοιες περιπτώσεις, η ισχύς της επίθεσης δεν εξαρτάται μόνο από την τεχνική τελειότητα του επιτιθέμενου, αλλά από την ποσότητα και την οργάνωση των διαθέσιμων υπολογιστικών και δικτυακών πόρων [77], [78], [51].

Παράλληλα, οι botnets αποτελούν χαρακτηριστικό παράδειγμα μετατροπής κατανεμημένων πόρων σε επιθετική ισχύ. Μεμονωμένες συσκευές μπορεί να έχουν περιορισμένη υπολογιστική ή δικτυακή αξία, όμως η μαζική συγκέντρωσή τους δημιουργεί δυνατότητες μεγάλης κλίμακας. Η συσσωρευμένη ισχύς τέτοιων δικτύων μπορεί να αξιοποιηθεί για DDoS επιθέσεις, credential stuffing, spam campaigns ή άλλες μορφές αυτοματοποιημένης επιθετικής δραστηριότητας [77], [51].

Η τεχνητή νοημοσύνη μεταβάλλει επίσης τη σχέση μεταξύ τεχνικής εξειδίκευσης και διαθέσιμων πόρων. Στο παρελθόν, η αξιοποίηση σύνθετων πόρων απαιτούσε σημαντική ανθρώπινη ικανότητα. Σήμερα, AI agents και multi-agent workflows μπορούν να υποστηρίξουν την εκτέλεση

τεχνικών εργασιών με μικρότερη ανθρώπινη παρέμβαση. Με αυτόν τον τρόπο, η ύπαρξη ΑΙ εργαλείων μειώνει την ανάγκη για συνεχή ανθρώπινο έλεγχο και επιτρέπει σε ανειδίκευτους σχεδόν επιτιθέμενους να αξιοποιούν πιο σύνθετες υποδομές [79], [80].

Η επίδραση αυτή είναι κρίσιμη, διότι μεταβάλλει τη σχέση πόρων και ικανοτήτων. Ένας επιτιθέμενος με περιορισμένες τεχνικές γνώσεις αλλά πρόσβαση σε αυτοματοποιημένα εργαλεία, cloud υπηρεσίες και ΑΙ συστήματα μπορεί να αποκτήσει δυνατότητες που παλαιότερα θα απαιτούσαν μεγαλύτερη ομάδα ή υψηλότερη τεχνική κατάρτιση. Αντίστοιχα, ένας προηγμένος επιτιθέμενος μπορεί να αξιοποιήσει τους ίδιους πόρους για να αυξήσει την ταχύτητα, την ακρίβεια και την προσαρμοστικότητα της επίθεσης [70], [80], [83].

Η διαθεσιμότητα πόρων επηρεάζει και το κόστος της επίθεσης. Όταν εργαλεία, υποδομές και υπηρεσίες μπορούν να μισθωθούν ή να αγοραστούν, το κόστος μετατρέπεται από κόστος ανάπτυξης σε κόστος πρόσβασης. Αυτό σημαίνει ότι ο επιτιθέμενος δεν χρειάζεται να επενδύσει σε μακροχρόνια δημιουργία τεχνικής υποδομής, αλλά μπορεί να αξιοποιήσει έτοιμες υπηρεσίες για συγκεκριμένο χρονικό διάστημα ή συγκεκριμένη επιθετική ανάγκη. Η λογική αυτή ενισχύει τη μετάβαση από την ιδιοκτησία πόρων στην κατανάλωση επιθετικών δυνατοτήτων ως υπηρεσίας [70], [82].

Το cybercrime-as-a-service αποτελεί βασικό μηχανισμό αναδιαμόρφωσης της δυναμικής του επιτιθέμενου. Μέσω τέτοιων υπηρεσιών, επιτιθέμενοι με περιορισμένη τεχνική εμπειρία μπορούν να αποκτήσουν πρόσβαση σε εργαλεία, υποδομές, botnets, malware, phishing kits ή υπηρεσίες τεχνικής υποστήριξης. Η διαθεσιμότητα αυτών των υπηρεσιών μειώνει τα εμπόδια εισόδου και επιτρέπει την εκτέλεση επιθέσεων από ευρύτερο φάσμα επιτιθέμενων [70].

Η ίδια λογική επεκτείνεται και στο AI-as-a-Service. Η πρόσβαση σε ισχυρά μοντέλα τεχνητής νοημοσύνης, εργαλεία αυτοματοποίησης και agentic συστήματα επιτρέπει την υποστήριξη επιθετικών διαδικασιών χωρίς ο επιτιθέμενος να διαθέτει ο ίδιος τις υποδομές εκπαίδευσης ή ανάπτυξης προηγμένων μοντέλων. Με τον τρόπο αυτό, η τεχνητή νοημοσύνη λειτουργεί ως εξωτερικευμένος πόρος, ο οποίος μπορεί να ενισχύσει επιθέσεις σε διαφορετικά στάδια του κύκλου ζωής τους [79], [83].

Η μεταβολή αυτή επηρεάζει άμεσα την επιχειρησιακή κλιμάκωση. Όταν ένας επιτιθέμενος μπορεί να αποκτήσει υπολογιστική ισχύ, αυτοματοποιημένα εργαλεία και ΑΙ δυνατότητες χωρίς αντίστοιχη αύξηση ανθρώπινου δυναμικού, μπορεί να επεκτείνει τις επιθέσεις του σε περισσότερους στόχους, να εκτελεί ταχύτερα επαναλαμβανόμενες εργασίες και να προσαρμόζει αποτελεσματικότερα τις τεχνικές του [80], [81], [84].

Η διαθεσιμότητα δεδομένων αποτελεί επίσης κρίσιμο πόρο. Οι επιθέσεις που βασίζονται σε τεχνητή νοημοσύνη εξαρτώνται συχνά από τη δυνατότητα συλλογής, επεξεργασίας και αξιοποίησης μεγάλων όγκων πληροφοριών. Δεδομένα που αφορούν χρήστες, οργανισμούς, cloud λογαριασμούς, credentials, συστήματα και συμπεριφορικά πρότυπα μπορούν να αυξήσουν την ακρίβεια στόχευσης και να ενισχύσουν την αποτελεσματικότητα της επίθεσης. Επομένως, ο πόρος δεν είναι μόνο η υπολογιστική ισχύς, αλλά και η πληροφορία που μπορεί να μετατραπεί σε επιθετικό πλεονέκτημα [82], [83].

Ιδιαίτερη σημασία έχει και η κατάχρηση cloud credentials και non-human identities. Σε cloud-native περιβάλλοντα, η πρόσβαση σε service accounts ή αυτοματοποιημένες ταυτότητες μπορεί να δώσει στον επιτιθέμενο δυνατότητες που υπερβαίνουν κατά πολύ την αρχική του θέση. Η πρόσβαση σε τέτοιους πόρους επιτρέπει lateral movement, κατάχρηση υπηρεσιών, πρόσβαση σε δεδομένα και αξιοποίηση υπολογιστικών δυνατοτήτων του ίδιου του οργανισμού-θύματος [82].

Η επίδραση των πόρων στη δυναμική του επιτιθέμενου μπορεί να αποτυπωθεί ως εξής σε μορφή πίνακα:

Πίνακας 4.1. Σχέση διαθέσιμων πόρων και δυνατοτήτων των επιτιθέμενων

Διαθέσιμο επίπεδο πόρων	Ενδεικτικές δυνατότητες επιτιθέμενου
-------------------------	--------------------------------------

Περιορισμένοι πόροι	Βασικό phishing, απλά scripts, χρήση έτοιμων εργαλείων και περιορισμένη αυτοματοποίηση.
Μέτριοι πόροι	Αυτοματοποιημένη αναγνώριση στόχων, credential attacks, χρήση malware kits, αξιοποίηση cloud υπηρεσιών και βασικών AI εργαλείων.
Υψηλοί πόροι	Μεγάλης κλίμακας επιχειρήσεις, botnet-driven DDoS, multi-agent AI workflows, cloud intrusions, adaptive operations και σύνθετες πολυσταδιακές επιθέσεις.

Ο πίνακας αυτός δείχνει ότι οι πόροι δεν αυξάνουν απλώς την ένταση μιας επίθεσης, αλλά μεταβάλλουν την ίδια τη φύση της. Όσο αυξάνεται η διαθεσιμότητα υπολογιστικών, οικονομικών και τεχνολογικών πόρων, τόσο αυξάνονται η κλίμακα, η ταχύτητα, η ακρίβεια και η προσαρμοστικότητα των επιθέσεων. Αντίστοιχα, η τεχνητή νοημοσύνη και το cloud computing συμπιέζουν τα όρια μεταξύ περιορισμένων και υψηλών πόρων, επειδή παρέχουν πρόσβαση σε δυνατότητες που παλαιότερα απαιτούσαν ιδιόκτητη υποδομή ή εξειδικευμένο προσωπικό [70], [79], [83].

Η σχέση αυτή έχει σημαντικές επιπτώσεις για την κυβερνοασφάλεια. Όσο οι επιτιθέμενοι αποκτούν ευκολότερη πρόσβαση σε υπολογιστικούς πόρους, cloud υποδομές και AI εργαλεία, τόσο αυξάνεται ο αριθμός των πιθανών επιθέσεων και μειώνεται ο χρόνος αντίδρασης των αμυντικών μηχανισμών. Η άμυνα δεν καλείται πλέον να αντιμετωπίσει μόνο τεχνικά προηγμένους δρώντες, αλλά και ένα ευρύτερο οικοσύστημα επιτιθέμενων που μπορούν να αξιοποιούν εξωτερικούς πόρους και αυτοματοποιημένες υπηρεσίες [78], [82], [84].

Η διαθεσιμότητα πόρων αποτελεί καθοριστικό παράγοντα στη διαμόρφωση της δυναμικής του επιτιθέμενου. Η τεχνητή νοημοσύνη, το cloud computing, τα botnets, οι υπηρεσίες cybercrime-as-a-service και τα AI-enabled workflows μειώνουν τα παραδοσιακά εμπόδια εισόδου και αυξάνουν την επιχειρησιακή ισχύ ακόμη και λιγότερο εξειδικευμένων δρώντων. Η βασική μεταβολή δεν είναι μόνο ότι οι επιθέσεις γίνονται πιο τεχνικά προηγμένες, αλλά ότι οι πόροι που απαιτούνται

για την υλοποίησή τους γίνονται πιο προσβάσιμοι, πιο ευέλικτοι και πιο εύκολα αξιοποιήσιμοι. Με αυτόν τον τρόπο, η ΑΙ και οι σύγχρονες υποδομές μεταβάλλουν τη σχέση μεταξύ πόρων και επιθετικών δυνατοτήτων, δημιουργώντας ένα περιβάλλον όπου η κλίμακα, η ταχύτητα και η αποτελεσματικότητα των κυβερνοεπιθέσεων ενισχύονται σημαντικά [70], [80], [81], [83].

4.3. Συγκριτική ανάλυση πριν και μετά την τεχνητή νοημοσύνη

Η ενσωμάτωση τεχνολογιών Τεχνητής Νοημοσύνης στο πεδίο της κυβερνοασφάλειας δεν επηρέασε μόνο μεμονωμένες τεχνικές επιθέσεων, αλλά μετασχημάτισε συνολικά τη δομή, τη δυναμική και την επιχειρησιακή λογική του κυβερνοχώρου. Η βασική μεταβολή δεν αφορά αποκλειστικά την αύξηση της αποτελεσματικότητας των επιθέσεων, αλλά την αλλαγή του τρόπου με τον οποίο σχεδιάζονται, εκτελούνται, προσαρμόζονται και κλιμακώνονται. Η ΑΙ επηρεάζει πλέον όλα τα στάδια του cyber kill chain, από τη reconnaissance και την εκμετάλλευση ευπαθειών έως τη δημιουργία κακόβουλου λογισμικού, την κοινωνική μηχανική και τη διατήρηση πρόσβασης σε πληροφοριακά συστήματα [49], [92].

Πριν από την ευρεία διάδοση εργαλείων Τεχνητής Νοημοσύνης, οι περισσότερες προηγμένες κυβερνοεπιθέσεις απαιτούσαν υψηλό επίπεδο τεχνικής εξειδίκευσης, σημαντικό χρόνο προετοιμασίας και εκτεταμένη ανθρώπινη παρέμβαση. Η reconnaissance πραγματοποιούνταν κυρίως χειροκίνητα, η ανάλυση στόχων βασιζόταν σε περιορισμένες διαδικασίες OSINT, ενώ η δημιουργία κακόβουλου λογισμικού απαιτούσε εξειδικευμένη γνώση προγραμματισμού και μηχανισμών αποφυγής ανίχνευσης. Παράλληλα, οι επιθέσεις phishing συχνά βασίζονταν σε γενικά, μαζικά και εύκολα αναγνωρίσιμα μηνύματα, τα οποία παρουσίαζαν γλωσσικές αδυναμίες και περιορισμένη εξατομίκευση [49], [53].

Επιπλέον, το επιχειρησιακό κόστος αποτελούσε σημαντικό περιοριστικό παράγοντα. Η ανάπτυξη σύνθετων επιθετικών εργαλείων, η εκτέλεση μεγάλης κλίμακας επιχειρήσεων και η διατήρηση μακροχρόνιων επιθετικών επιχειρήσεων απαιτούσαν οργανωμένες ομάδες, υψηλή τεχνική κατάρτιση και σημαντική υποδομή. Κατά συνέπεια, προηγμένες επιθέσεις αποτελούσαν κυρίως χαρακτηριστικό κρατικά υποστηριζόμενων ομάδων ή highly organized cybercriminal networks [85], [86].

Η εμφάνιση όμως generative AI, agentic systems και LLM-based offensive workflows μεταβάλλει ριζικά αυτή την πραγματικότητα. Οι σύγχρονες AI τεχνολογίες επιτρέπουν αυτοματοποίηση διαδικασιών που προηγουμένως απαιτούσαν συνεχή ανθρώπινη συμμετοχή, ενώ παράλληλα μειώνουν σημαντικά τα τεχνικά εμπόδια εισόδου στο κυβερνοέγκλημα. Η μεταβολή αυτή δεν είναι απλώς ποσοτική, αλλά ποιοτική, καθώς αλλάζει τη σχέση μεταξύ πόρων, τεχνικής γνώσης και επιθετικής ικανότητας [49], [69].

Στο στάδιο της reconnaissance, η AI επιτρέπει τη γρήγορη επεξεργασία μεγάλου όγκου πληροφοριών, τη συσχέτιση ετερογενών δεδομένων και την αυτοματοποιημένη αναγνώριση πιθανών αδύναμων σημείων. Ενώ στο παραδοσιακό περιβάλλον η διαδικασία αυτή ήταν χρονοβόρα και περιορισμένη από ανθρώπινες δυνατότητες, τα AI-driven reconnaissance systems μπορούν να λειτουργούν συνεχώς, να διατηρούν context και να εκτελούν πιο σύνθετα operational workflows [85], [88].

Αντίστοιχα, στο πεδίο της εκμετάλλευσης ευπαθειών, η AI μεταβάλλει το επίπεδο πολυπλοκότητας που μπορούν να διαχειριστούν οι επιτιθέμενοι. Σύγχρονες έρευνες δείχνουν ότι ομάδες LLM agents μπορούν να συμβάλλουν στον εντοπισμό και στην αξιοποίηση σύνθετων ευπαθειών, ακόμη και σε περιβάλλοντα που παραδοσιακά απαιτούσαν υψηλή τεχνική εμπειρία. Αυτό υποδηλώνει ότι η AI δεν λειτουργεί μόνο ως εργαλείο υποστήριξης, αλλά ως παράγοντας ενίσχυσης επιθετικών δυνατοτήτων [87].

Η μεταβολή είναι ιδιαίτερα εμφανής και στο phishing. Πριν από την AI, οι επιθέσεις phishing χαρακτηρίζονταν κυρίως από μαζικά και γενικά μηνύματα, τα οποία ήταν σχετικά εύκολο να εντοπιστούν λόγω γλωσσικών λαθών, περιορισμένης προσωποποίησης ή ασυνήθιστης δομής. Με την ανάπτυξη generative AI, οι επιτιθέμενοι μπορούν πλέον να δημιουργούν προσωποποιημένα, γλωσσικά άρτια και πειστικά μηνύματα, προσαρμοσμένα στο προφίλ κάθε στόχου. Η δυνατότητα μαζικής προσωποποίησης αποτελεί μία από τις σημαντικότερες αλλαγές που επιφέρει η τεχνητή νοημοσύνη στο social engineering [53], [90], [91].

Η ίδια μετάβαση παρατηρείται και στο κακόβουλο λογισμικό. Παραδοσιακά, πολλά malware samples ήταν περισσότερο στατικά και συχνά ανιχνεύονταν μέσω signature-based μηχανισμών. Η ενσωμάτωση τεχνικών AI και generative models οδηγεί στην ανάπτυξη περισσότερο

προσαρμοστικών και polymorphic μορφών malware, οι οποίες μπορούν να μεταβάλλουν χαρακτηριστικά τους, να προσαρμόζονται σε αμυντικούς μηχανισμούς και να αυξάνουν την ικανότητα αποφυγής ανίχνευσης [62].

Η ΑΙ επηρεάζει επίσης τη χρονική διάσταση των επιθέσεων. Η αυτοματοποίηση reconnaissance, exploitation, phishing generation και malware adaptation μειώνει σημαντικά τον χρόνο μεταξύ σχεδιασμού και εκτέλεσης μιας επίθεσης. Η συμπίεση αυτή των επιθετικών διαδικασιών (attack compression) επιτρέπει στους επιτιθέμενους να εκτελούν περισσότερες ενέργειες σε μικρότερο χρονικό διάστημα και να αντιδρούν ταχύτερα σε αλλαγές του περιβάλλοντος στόχου [69], [89].

Παράλληλα, η χρήση ΑΙ αυξάνει σημαντικά την κλιμάκωση των επιθέσεων. Πριν από την ΑΙ, η δυνατότητα διαχείρισης μεγάλου αριθμού στόχων περιοριζόταν από ανθρώπινους και οργανωτικούς παράγοντες. Σήμερα, AI-enabled systems μπορούν να υποστηρίξουν μαζική παραγωγή phishing περιεχομένου, ταυτόχρονη ανάλυση πολλαπλών στόχων και αυτοματοποιημένη διαχείριση επιθετικών ενεργειών μεγάλης κλίμακας [69], [90], [92].

Σημαντική είναι και η μεταβολή της σχέσης μεταξύ κόστους και επιθετικής ικανότητας. Στο παραδοσιακό μοντέλο, οι πιο αποτελεσματικές επιθέσεις απαιτούσαν μεγαλύτερους πόρους, μεγαλύτερη ομάδα και υψηλή τεχνική εξειδίκευση. Η ΑΙ μειώνει μέρος αυτού του κόστους, επιτρέποντας σε λιγότερο έμπειρους δρώντες να αποκτούν πρόσβαση σε δυνατότητες που προηγουμένως ανήκαν κυρίως σε advanced threat actors [49], [85], [86].

Η μεταβολή αυτή επηρεάζει και την ισορροπία επίθεσης–άμυνας. Σύμφωνα με αρκετές αναλύσεις, η ΑΙ μπορεί να ενισχύσει περισσότερο την επιθετική πλευρά σε ορισμένα στάδια του cyber conflict, κυρίως λόγω της ταχύτητας, της αυτοματοποίησης και της δυνατότητας επιχειρησιακής κλιμάκωσης που παρέχει στους επιτιθέμενους. Παρότι η τεχνητή νοημοσύνη χρησιμοποιείται και αμυντικά, η αύξηση της ταχύτητας και της μαζικότητας των επιθέσεων δημιουργεί σημαντική πίεση στους αμυντικούς μηχανισμούς [49], [85], [86].

Ιδιαίτερα σημαντική είναι η μετάβαση από AI-assisted σε περισσότερο agentic και autonomous επιθετικά μοντέλα. Ενώ οι πρώτες γενιές εργαλείων λειτουργούσαν κυρίως υποστηρικτικά προς τον άνθρωπο, τα νεότερα agentic systems μπορούν να διαχειρίζονται context, να εκτελούν διαδοχικές ενέργειες και να προσαρμόζουν τη συμπεριφορά τους ανάλογα με το περιβάλλον

στόχου. Η εξέλιξη αυτή ενισχύει την επιχειρησιακή αυτονομία των επιθέσεων και μεταβάλλει τη συνολική φύση των offensive cyber operations [69], [88], [89].

Η συγκριτική ανάλυση πριν και μετά την τεχνητή νοημοσύνη δείχνει ότι η τεχνητή νοημοσύνη δεν άλλαξε απλώς την αποτελεσματικότητα των κυβερνοεπιθέσεων, αλλά μετασχημάτισε τη δομή και την οικονομία του κυβερνοεγκλήματος. Η δυνατότητα αυτοματοποίησης, προσαρμογής και μαζικής κλιμάκωσης επιθέσεων, σε συνδυασμό με τη μείωση των τεχνικών και οικονομικών εμποδίων εισόδου, δημιουργεί ένα νέο περιβάλλον κυβερνοαπειλών, στο οποίο η σχέση μεταξύ επιτιθέμενου, πόρων και επιχειρησιακής ισχύος επαναπροσδιορίζεται ουσιαστικά.

Πίνακας 4.2. Συγκριτική αποτύπωση παραδοσιακών και AI-enabled κυβερνοεπιθέσεων

Παράμετρος	Παραδοσιακές επιθέσεις	Επιθέσεις με αξιοποίηση AI
------------	------------------------	----------------------------

Συλλογή πληροφοριών	Κατά κύριο λόγο χειροκίνητη, με περιορισμένη ταχύτητα επεξεργασίας.	Αυτοματοποιημένη επεξεργασία ανοικτών πηγών, συσχέτιση δεδομένων και ταχύτερη χαρτογράφηση στόχων.
Παραγωγή περιεχομένου	Συχνά τυποποιημένο και περιορισμένα προσωποποιημένο περιεχόμενο.	Δυνατότητα γλωσσικά άρτιου, προσαρμοσμένου και πολυγλωσσικού περιεχομένου.
Τεχνικές απαιτήσεις	Αυξημένη ανάγκη για εξειδίκευση σε επιμέρους στάδια.	Μείωση του αρχικού εμποδίου εισόδου για ορισμένες εργασίες, χωρίς κατάργηση της ανάγκης τεχνικής κρίσης στις σύνθετες επιθέσεις.
Κλίμακα και επανάληψη	Η αύξηση της κλίμακας απαιτεί αντίστοιχους ανθρώπινους και τεχνικούς πόρους.	Διευκολύνεται η επανάληψη, η εξατομίκευση και η διαχείριση μεγαλύτερου όγκου επιθετικών ενεργειών.
Προσαρμοστικότητα	Συχνά βασίζεται σε στατικά scripts ή προκαθορισμένες ροές.	Μπορεί να υποστηρίζεται δυναμική προσαρμογή περιεχομένου ή επιλογής ενεργειών βάσει δεδομένων.
Αμυντική πρόκληση	Εμφαση στην ανίχνευση γνωστών τεχνικών και στατικών δεικτών.	Ανάγκη για συμπεριφορική ανάλυση, ταχεία απόκριση και διαρκή αναπροσαρμογή αμυντικών μηχανισμών.

4.4. Αυτοματοποιημένες επιθέσεις με τεχνητή νοημοσύνη και ο ρόλος του Cloud Computing (Cloud & AI-as-a-Service)

Η ανάλυση των προηγούμενων υποκεφαλαίων ανέδειξε ότι η τεχνητή νοημοσύνη επηρεάζει το κόστος, την πολυπλοκότητα, την κλίμακα και την προσβασιμότητα των κυβερνοεπιθέσεων. Ωστόσο, η επιχειρησιακή αξιοποίηση της AI σε μεγάλη κλίμακα δεν εξαρτάται μόνο από τους ίδιους τους αλγόριθμους, αλλά και από τις υποδομές που επιτρέπουν την πρακτική εφαρμογή τους. Στο πλαίσιο αυτό, το cloud computing, οι υπηρεσίες AI-as-a-Service και τα οικοσυστήματα cybercrime-as-a-service αποτελούν κρίσιμους παράγοντες που καθιστούν την τεχνητή νοημοσύνη λειτουργικά αξιοποιήσιμη στο πεδίο των αυτοματοποιημένων κυβερνοεπιθέσεων [69], [70], [45].

Οι αυτοματοποιημένες επιθέσεις με τεχνητή νοημοσύνη δεν πρέπει να νοούνται απλώς ως επιθέσεις στις οποίες ένα εργαλείο εκτελεί επαναλαμβανόμενες ενέργειες. Η ουσιαστική μεταβολή αφορά τη δυνατότητα συνδυασμού αυτοματοποίησης, ανάλυσης δεδομένων, λήψης αποφάσεων και προσαρμογής σε πραγματικό χρόνο. Η τεχνητή νοημοσύνη μπορεί να υποστηρίξει επιθετικές διαδικασίες όπως reconnaissance, victim profiling, παραγωγή phishing περιεχομένου, τεχνική υποστήριξη κακόβουλου λογισμικού, ανάλυση ευπαθειών και προσαρμογή επιθετικών σεναρίων [69], [40], [39].

Η επιχειρησιακή σημασία του cloud computing έγκειται στο ότι παρέχει δυναμικά επεκτάσιμους υπολογιστικούς και δικτυακούς πόρους. Σε παραδοσιακά περιβάλλοντα, η εκτέλεση μεγάλης κλίμακας επιθέσεων απαιτούσε ιδιόκτητες υποδομές, υψηλό αρχικό κόστος και τεχνική δυνατότητα διαχείρισης των πόρων αυτών. Αντίθετα, οι cloud υποδομές επιτρέπουν πρόσβαση σε απομακρυσμένη υπολογιστική ισχύ, αποθηκευτικούς πόρους, υπηρεσίες αυτοματοποίησης και ευέλικτη κλιμάκωση χωρίς αντίστοιχη επένδυση σε φυσικό εξοπλισμό [32], [81], [82].

Η διαθεσιμότητα cloud πόρων μεταβάλλει άμεσα την κλίμακα των επιθέσεων. Ένας επιτιθέμενος μπορεί να αξιοποιήσει απομακρυσμένες υποδομές για τη διαχείριση μεγαλύτερου όγκου δεδομένων, τη μαζική αποστολή επιθετικού περιεχομένου, τη φιλοξενία προσωρινών υποδομών

εξαπάτησης ή την εκτέλεση αυτοματοποιημένων διαδικασιών αναγνώρισης στόχων. Με τον τρόπο αυτό, το cloud δεν αποτελεί απλώς τεχνολογικό περιβάλλον λειτουργίας, αλλά παράγοντα που ενισχύει την ταχύτητα και την επεκτασιμότητα των επιθετικών ενεργειών [32], [25], [82].

Ιδιαίτερη σημασία έχει και η κατάχρηση cloud-native πόρων. Οι επιτιθέμενοι δεν χρειάζεται πάντοτε να δημιουργούν δικές τους υποδομές, καθώς μπορούν να αξιοποιήσουν παραβιασμένους λογαριασμούς, κλεμμένα credentials, μη ορθά ρυθμισμένες υπηρεσίες, non-human identities και ευάλωτα cloud environments. Σε τέτοιες περιπτώσεις, οι πόροι του ίδιου του θύματος μπορούν να μετατραπούν σε μέσο επέκτασης της επίθεσης [25], [82].

Η έννοια του AI-as-a-Service αποτελεί τον δεύτερο βασικό άξονα του κεφαλαίου. Μέσω υπηρεσιών που παρέχει η τεχνητή νοημοσύνη, ο επιτιθέμενος δεν χρειάζεται να αναπτύξει, να εκπαιδεύσει ή να συντηρήσει δικά του προηγμένα μοντέλα. Μπορεί να αξιοποιήσει διαθέσιμες πλατφόρμες, έτοιμα μοντέλα, APIs, αυτοματοποιημένα εργαλεία και agentic AI services για να υποστηρίξει επιμέρους στάδια μιας επίθεσης. Έτσι, η τεχνητή νοημοσύνη μετατρέπεται από εξειδικευμένη τεχνολογική δυνατότητα σε προσβάσιμη υπηρεσία [69], [39], [93].

Η μετάβαση αυτή μειώνει ουσιαστικά τα τεχνικά εμπόδια εισόδου. Παραδοσιακά, η αξιοποίηση προηγμένων αλγορίθμων απαιτούσε γνώση μηχανικής μάθησης, πρόσβαση σε datasets, υπολογιστική ισχύ και τεχνική εμπειρία. Η διάδοση υπηρεσιών AI-as-a-Service περιορίζει αυτές τις απαιτήσεις, επιτρέποντας σε λιγότερο εξειδικευμένους δρώντες να χρησιμοποιούν δυνατότητες παραγωγής περιεχομένου, ανάλυσης πληροφοριών, αυτοματοποίησης workflows και τεχνικής υποστήριξης κυβερνοεπιθετικών ενεργειών [70], [39], [93].

Η εξέλιξη αυτή συνδέεται άμεσα με το cybercrime-as-a-service. Στα σύγχρονα εγκληματικά οικοσυστήματα, εργαλεία, υποδομές και υπηρεσίες διατίθενται σε τρίτους ως έτοιμες λύσεις. Η προσθήκη AI σε τέτοια μοντέλα ενισχύει την επιχειρησιακή τους αποτελεσματικότητα, καθώς επιτρέπει αυτοματοποιημένη παραγωγή phishing περιεχομένου, υποστήριξη malware creation, vulnerability research, victim profiling και τεχνική καθοδήγηση λιγότερο έμπειρων επιτιθέμενων [70], [39], [72].

Η σχέση μεταξύ cloud, AI-as-a-Service και cybercrime-as-a-service δημιουργεί ένα νέο επιχειρησιακό μοντέλο κυβερνοεπιθέσεων. Ο επιτιθέμενος δεν χρειάζεται πλέον να κατέχει όλα

τα τεχνικά στοιχεία της επίθεσης. Μπορεί να συνδυάσει υπηρεσίες υποδομής, εργαλεία αυτοματοποίησης, μοντέλα AI και εγκληματικές υπηρεσίες τρίτων, διαμορφώνοντας μια επιθετική αλυσίδα που είναι περισσότερο ευέλικτη, επεκτάσιμη και οικονομικά αποδοτική [69], [70], [72].

Η αυτοματοποίηση που προκύπτει από αυτόν τον συνδυασμό επηρεάζει πολλαπλά στάδια του επιθετικού κύκλου. Στο στάδιο της reconnaissance, AI εργαλεία μπορούν να υποστηρίξουν τη συλλογή και ανάλυση πληροφοριών. Στο στάδιο του social engineering, generative AI μπορεί να παραγάγει προσωποποιημένα μηνύματα εξαπάτησης. Στο στάδιο της τεχνικής εκμετάλλευσης, αυτοματοποιημένα εργαλεία μπορούν να υποστηρίξουν την αναζήτηση αδύναμων σημείων. Στα μεταγενέστερα στάδια, cloud υποδομές μπορούν να διευκολύνουν τη διαχείριση επιθετικών επιχειρήσεων και τη διατήρηση απομακρυσμένων υποδομών [69], [40], [81].

Η έννοια του agentic AI ενισχύει ακόμη περισσότερο αυτή τη δυναμική. Τα agentic AI systems μπορούν να εκτελούν αλληλουχίες ενεργειών, να αξιολογούν ενδιάμεσα αποτελέσματα και να προσαρμόζουν τη συμπεριφορά τους ανάλογα με το περιβάλλον. Όταν τέτοια συστήματα συνδυάζονται με πρόσβαση σε cloud υπηρεσίες και εργαλεία απομακρυσμένης εκτέλεσης, δημιουργείται η δυνατότητα πιο αυτόνομων και επιχειρησιακά ολοκληρωμένων επιθετικών workflows [69], [81], [93].

Η χρήση autonomous AI cloud attack agents δείχνει ότι το cloud μπορεί να λειτουργήσει ταυτόχρονα ως στόχος και ως πεδίο επιχειρησιακής δράσης. Οι επιτιθέμενοι μπορούν να στοχεύσουν cloud περιβάλλοντα, να αξιοποιήσουν λανθασμένες ρυθμίσεις, να εκμεταλλευτούν credentials και να κινηθούν εντός cloud-native υποδομών. Παράλληλα, οι ίδιες οι cloud υπηρεσίες μπορούν να υποστηρίξουν την αυτοματοποίηση και την κλιμάκωση των επιθετικών διαδικασιών [81], [82].

Η επίδραση στο κόστος είναι καθοριστική. Το cloud computing μειώνει την ανάγκη για αρχική επένδυση σε υποδομές, ενώ το AI-as-a-Service μειώνει την ανάγκη ανάπτυξης ιδιόκτητων μοντέλων και εξειδικευμένης τεχνικής γνώσης. Το cybercrime-as-a-service, από την πλευρά του, μειώνει το κόστος πρόσβασης σε έτοιμες εγκληματικές δυνατότητες. Ο συνδυασμός αυτών των τριών στοιχείων μετατρέπει πολλές κυβερνοεπιθέσεις από δραστηριότητες υψηλής αρχικής επένδυσης σε διαδικασίες on-demand πρόσβασης σε πόρους και εργαλεία [70], [39], [72].

Η μεταβολή αυτή επηρεάζει και την πολυπλοκότητα. Ενώ οι επιθέσεις μπορεί τεχνικά να γίνονται πιο σύνθετες, η πολυπλοκότητα αυτή αποκρύπτεται από τον τελικό χρήστη των εργαλείων. Δηλαδή, ο επιτιθέμενος μπορεί να αξιοποιεί σύνθετες δυνατότητες χωρίς να κατανοεί πλήρως κάθε τεχνικό επίπεδο της υποδομής που χρησιμοποιεί. Αυτό δημιουργεί μια σημαντική διαφοροποίηση μεταξύ της αντικειμενικής πολυπλοκότητας της επίθεσης και της αντιλαμβανόμενης πολυπλοκότητας για τον δράστη [69], [70], [93].

Παράλληλα, η κλίμακα των επιθέσεων αυξάνεται σημαντικά. Η χρήση cloud και AI υπηρεσιών επιτρέπει την ταυτόχρονη διαχείριση πολλών στόχων, την παραγωγή μεγάλου όγκου επιθετικού περιεχομένου και την επανάληψη επιθέσεων με χαμηλότερο κόστος. Έτσι, επιθέσεις που παλαιότερα θα απαιτούσαν σημαντικό ανθρώπινο δυναμικό μπορούν πλέον να εκτελούνται σε υψηλότερο βαθμό αυτοματοποίησης [69], [40], [32].

Οι επιπτώσεις για την κυβερνοασφάλεια είναι σημαντικές. Η άμυνα δεν καλείται πλέον να αντιμετωπίσει μόνο μεμονωμένους τεχνικά καταρτισμένους επιτιθέμενους, αλλά ένα ευρύτερο οικοσύστημα όπου υποδομές, εργαλεία και υπηρεσίες μπορούν να συνδυαστούν ταχύτατα. Η ύπαρξη AI-enabled workflows και cloud-based επιθετικών δυνατοτήτων μειώνει τον χρόνο προετοιμασίας των επιθέσεων και αυξάνει την πίεση στους μηχανισμούς ανίχνευσης και απόκρισης [32], [25], [45].

Παράλληλα, τα agentic AI services δημιουργούν νέους κινδύνους λόγω της δυνατότητας πρόσβασης σε εργαλεία, εκτέλεσης ενεργειών και αλληλεπίδρασης με απομακρυσμένα περιβάλλοντα. Η επιχειρησιακή αυτονομία τέτοιων συστημάτων μπορεί να αυξήσει την ταχύτητα και την πολυπλοκότητα τόσο νόμιμων όσο και κακόβουλων διαδικασιών, γεγονός που καθιστά αναγκαία την προσεκτική αξιολόγηση των κινδύνων τους [93], [94].

Η πραγματική μεταβολή λοιπόν, προκύπτει από τη σύζευξη AI, cloud computing, AI-as-a-Service και cybercrime-as-a-service. Ο συνδυασμός αυτός μειώνει το κόστος, περιορίζει τις τεχνικές απαιτήσεις, αυξάνει την κλίμακα και επιταχύνει την αυτοματοποίηση των κυβερνοεπιθέσεων. Επομένως, η AI δεν λειτουργεί απλώς ως νέο εργαλείο επίθεσης, αλλά ως μέρος ενός ευρύτερου τεχνολογικού και οικονομικού οικοσυστήματος που αναδιαμορφώνει τη σύγχρονη δυναμική του κυβερνοεγκλήματος.

5. ΕΝΔΕΙΚΤΙΚΕΣ ΠΕΡΙΠΤΩΣΕΙΣ ΚΑΙ ΕΦΑΡΜΟΓΕΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΕ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΙΣ

5.1. AI-driven phishing campaigns

Η αξιοποίηση τεχνολογιών Τεχνητής Νοημοσύνης στις επιθέσεις phishing έχει επιφέρει ουσιαστικές αλλαγές στη μορφή, στην αποτελεσματικότητα και στην επιχειρησιακή δυναμική των social engineering επιθέσεων. Στο παραδοσιακό περιβάλλον κυβερνοαπειλών, οι επιθέσεις phishing βασίζονταν κυρίως σε γενικευμένα και μαζικά μηνύματα, τα οποία συχνά περιείχαν γλωσσικά λάθη, περιορισμένη προσωποποίηση και εμφανή χαρακτηριστικά εξαπάτησης. Η επιτυχία τους στηριζόταν κυρίως στον μεγάλο αριθμό πιθανών θυμάτων και όχι στην ποιότητα ή στην προσαρμογή του περιεχομένου στο προφίλ του στόχου. Η ανάπτυξη generative AI και προηγμένων μοντέλων επεξεργασίας φυσικής γλώσσας αλλάζει σημαντικά αυτή τη λογική, επιτρέποντας τη δημιουργία περισσότερο πειστικών, εξατομικευμένων και γλωσσικά άρτιων phishing μηνυμάτων με υψηλό βαθμό αυτοματοποίησης [53], [64].

Η τεχνητή νοημοσύνη επηρεάζει πολλά στάδια των phishing επιχειρήσεων, ξεκινώντας από τη συλλογή και ανάλυση πληροφοριών σχετικά με πιθανούς στόχους. Μέσω τεχνικών OSINT και αυτοματοποιημένης επεξεργασίας δεδομένων από κοινωνικά δίκτυα, δημόσιες βάσεις δεδομένων και επαγγελματικές πλατφόρμες, οι επιτιθέμενοι μπορούν να δημιουργούν αναλυτικά προφίλ θυμάτων και να προσαρμόζουν το περιεχόμενο της επίθεσης ανάλογα με τη θέση, τις επαγγελματικές σχέσεις ή τις επικοινωνιακές συνήθειες του στόχου. Οι Raja Jabir, John Le και Chau Nguyen επισημαίνουν ότι τα generative AI systems μπορούν να παράγουν spear-phishing emails με αυξημένη γλωσσική συνοχή, περιορισμένα εμφανή σημάδια απάτης και μεγαλύτερη ικανότητα μίμησης πραγματικής επαγγελματικής επικοινωνίας [53]. Αντίστοιχα, οι Green και Alam αναφέρουν ότι η χρήση generative AI επιτρέπει τη μαζική παραγωγή highly personalized phishing περιεχομένου με σημαντικά μειωμένες απαιτήσεις χρόνου και τεχνικής εξειδίκευσης [64].

Η σημαντικότερη μεταβολή που επιφέρει η τεχνητή νοημοσύνη στις επιθέσεις phishing αφορά τον συνδυασμό κλιμάκωσης και εξατομίκευσης. Στο παρελθόν, η προσωποποίηση απαιτούσε σημαντική ανθρώπινη εργασία και χρόνο προετοιμασίας, γεγονός που περιόριζε την κλίμακα

τέτοιων επιθέσεων. Σήμερα, η ΑΙ επιτρέπει τη δημιουργία χιλιάδων προσωποποιημένων phishing μηνυμάτων με αυτοματοποιημένο τρόπο και περιορισμένο επιχειρησιακό κόστος [53], [64]. Ως αποτέλεσμα, οι επιθέσεις phishing μεταβαίνουν από generic και σχετικά εύκολα αναγνωρίσιμες πρακτικές σε highly scalable και ταυτόχρονα highly convincing επιχειρήσεις social engineering.

Παράλληλα, η ανάπτυξη τεχνολογιών deepfake και synthetic voice επεκτείνει τις δυνατότητες των AI-driven phishing campaigns πέρα από τα παραδοσιακά emails ή γραπτά μηνύματα. Σύμφωνα με το ENISA Threat Landscape 2025 και το Microsoft Digital Defense Report 2025, η χρήση AI-generated audio και video επιτρέπει στους επιτιθέμενους να πραγματοποιούν επιθέσεις impersonation υψηλής αξιοπιστίας, μιμούμενοι στελέχη οργανισμών, συνεργάτες ή αξιόπιστα πρόσωπα με αυξημένο βαθμό ρεαλισμού [21], [32]. Οι τεχνικές αυτές επηρεάζουν ιδιαίτερα επιθέσεις τύπου Business Email Compromise (BEC), καθώς η ΑΙ αυξάνει σημαντικά την πειστικότητα της πλαστοπροσωπίας και δυσκολεύει την αναγνώριση της απάτης από τα θύματα [32], [68].

Χαρακτηριστικό παράδειγμα αποτελεί το περιστατικό deepfake CFO fraud στο Hong Kong το 2024, κατά το οποίο εργαζόμενος πολυεθνικής εταιρείας συμμετείχε σε βιντεοκλήση με πρόσωπα που θεωρούσε ότι ήταν ο οικονομικός διευθυντής και άλλα στελέχη της εταιρείας. Στην πραγματικότητα, οι συμμετέχοντες είχαν δημιουργηθεί μέσω τεχνολογιών deepfake και synthetic media, με αποτέλεσμα ο εργαζόμενος να προχωρήσει σε μεταφορά περίπου 25 εκατομμυρίων δολαρίων προς λογαριασμούς των επιτιθέμενων [95]. Το συγκεκριμένο περιστατικό αναδεικνύει τη μετάβαση από τις παραδοσιακές phishing επιθέσεις σε σύνθετες AI-enabled επιχειρήσεις social engineering, όπου η τεχνητή νοημοσύνη δεν χρησιμοποιείται μόνο για τη δημιουργία παραπλανητικού περιεχομένου, αλλά για την κατασκευή ολόκληρου πλαισίου εμπιστοσύνης μέσω ψεύτικης οπτικοακουστικής αλληλεπίδρασης [65], [21].

Επιπλέον, threat intelligence reports του 2025 καταγράφουν αυξημένη χρήση GenAI σε επιθέσεις Business Email Compromise (BEC), με στόχο τη δημιουργία realistic corporate impersonation campaigns και automated fraud workflows [32], [68]. Η δυνατότητα linguistic mimicry, δηλαδή μίμησης του τρόπου γραφής και επικοινωνίας συγκεκριμένων στελεχών ή οργανισμών, επιτρέπει στους επιτιθέμενους να δημιουργούν ιδιαίτερα πειστικά μηνύματα, τα οποία εμφανίζονται ως νόμιμη εταιρική επικοινωνία [68]. Σε αντίθεση με τις παραδοσιακές phishing επιθέσεις, όπου τα

γλωσσικά λάθη και η περιορισμένη προσαρμογή λειτουργούσαν συχνά ως ένδειξη απάτης, τα AI-generated phishing μηνύματα παρουσιάζουν υψηλότερο βαθμό γλωσσικής ορθότητας και καλύτερη προσαρμογή στο προφίλ του θύματος [53], [68].

5.2. Automated malware generation

Η ενσωμάτωση τεχνολογιών τεχνητής νοημοσύνης στην ανάπτυξη και τροποποίηση κακόβουλου λογισμικού αποτελεί μία από τις σημαντικότερες μεταβολές στο σύγχρονο περιβάλλον κυβερνοαπειλών. Παραδοσιακά, η δημιουργία malware απαιτούσε υψηλό επίπεδο γνώσεων από τους εκτελεστές. Η διαδικασία αυτή απαιτούσε επίσης, σημαντικό χρόνο ανάπτυξης και εξειδικευμένη γνώση offensive security. Η εμφάνιση generative AI και μεγάλων γλωσσικών μοντέλων (LLMs) μεταβάλλει ουσιαστικά αυτή τη δυναμική, καθώς επιτρέπει την αυτοματοποίηση μέρους της διαδικασίας ανάπτυξης κακόβουλου λογισμικού, τη γρηγορότερη παραγωγή malware variants και τη μείωση των τεχνικών απαιτήσεων για τους επιτιθέμενους [40], [52], [47].

Οι σύγχρονες AI-assisted malware διαδικασίες δεν περιορίζονται μόνο στη δημιουργία κακόβουλου κώδικα, αλλά επεκτείνονται σε πολλαπλά στάδια του malware lifecycle. Τα LLMs μπορούν να αξιοποιηθούν για code generation, vulnerability assessment, script modification, obfuscation και υποστήριξη κακόβουλων workflows, επιτρέποντας την επιτάχυνση της ανάπτυξης malware και τη μείωση του χρόνου προετοιμασίας επιθέσεων [54], [47]. Ως αποτέλεσμα, η AI δεν λειτουργεί απλώς ως εργαλείο υποστήριξης προγραμματισμού, αλλά ως παράγοντας operational acceleration που επηρεάζει άμεσα την ταχύτητα και την κλίμακα ανάπτυξης κακόβουλου λογισμικού.

Ιδιαίτερα σημαντική είναι η δυνατότητα της AI να υποστηρίζει adaptive και polymorphic malware. Στα παραδοσιακά malware οικοσυστήματα, μεγάλο μέρος της ανίχνευσης βασιζόταν σε signature-based μηχανισμούς, όπου στατικά χαρακτηριστικά του κακόβουλου λογισμικού μπορούσαν να αναγνωριστούν από antivirus και intrusion detection systems. Η ανάπτυξη AI-enabled polymorphic malware μεταβάλλει αυτή τη λογική, καθώς το malware μπορεί να τροποποιεί δυναμικά τμήματα του κώδικά του, να μεταβάλλει τη συμπεριφορά του ή να

δημιουργεί νέες παραλλαγές με στόχο την αποφυγή ανίχνευσης [96], [52]. Σύμφωνα με τον Morris, η χρήση generative AI σε polymorphic malware επιτρέπει τη δημιουργία malware variants που εμφανίζουν υψηλότερο επίπεδο προσαρμοστικότητας και δυσκολεύουν την αποτελεσματικότητα των παραδοσιακών signature-based αμυντικών μηχανισμών [39].

Παράλληλα, η AI επηρεάζει σημαντικά και το επίπεδο τεχνικής εξειδίκευσης που απαιτείται για την ανάπτυξη malware. Η δυνατότητα χρήσης pretrained models και automated code-generation διαδικασιών επιτρέπει ακόμη και σε λιγότερο έμπειρους επιτιθέμενους να αποκτούν πρόσβαση σε προηγμένες δυνατότητες ανάπτυξης κακόβουλου λογισμικού [40], [41]. Το φαινόμενο αυτό συνδέεται άμεσα με τη μείωση των barriers to entry στο κυβερνοέγκλημα, καθώς οι επιτιθέμενοι δεν χρειάζεται πλέον να αναπτύσσουν εξ ολοκλήρου δικά τους malware frameworks ή να διαθέτουν υψηλή εξειδίκευση σε offensive programming.

Χαρακτηριστικό παράδειγμα αποτελεί το περιστατικό που περιγράφεται στο Anthropic Threat Intelligence Report 2025 σχετικά με AI-assisted ransomware και “no-code malware” workflows [40]. Σύμφωνα με την αναφορά, η τεχνητή νοημοσύνη χρησιμοποιήθηκε για υποστήριξη ransomware διαδικασιών και malware operationalization από άτομα με περιορισμένες τεχνικές γνώσεις. Η χρήση generative AI επέτρεψε τη δημιουργία και τροποποίηση κακόβουλων scripts, την υποστήριξη deployment διαδικασιών και τη διευκόλυνση ransomware-as-a-service πρακτικών [40]. Το συγκεκριμένο case είναι ιδιαίτερα σημαντικό, καθώς αναδεικνύει ότι η AI δεν αυξάνει μόνο την αποτελεσματικότητα προηγμένων threat actors, αλλά επιτρέπει και σε χαμηλότερου επιπέδου επιτιθέμενους να αποκτούν πρόσβαση σε πιο σύνθετες offensive δυνατότητες.

Ένα ακόμη σημαντικό παράδειγμα αφορά τις αναφορές της Google Threat Intelligence Group σχετικά με AI-enabled malware και autonomous malware operations [96]. Οι αναφορές αυτές περιγράφουν malware frameworks που αξιοποιούν AI για δυναμική ερμηνεία του περιβάλλοντος εκτέλεσης, δημιουργία εντολών και προσαρμοστική συμπεριφορά κατά τη διάρκεια της επίθεσης. Η δυνατότητα αυτή μεταβάλλει τη φύση του παραδοσιακού malware, καθώς το κακόβουλο λογισμικό αποκτά αυξημένο επίπεδο adaptability και operational flexibility. Αντί για στατικά malware payloads, τα AI-assisted συστήματα μπορούν να τροποποιούν τη λειτουργία τους ανάλογα με το περιβάλλον στόχο ή τα αμυντικά μέτρα που εντοπίζουν [96], [52].

Επιπλέον, ιδιαίτερο ενδιαφέρον παρουσιάζουν οι ερευνητικές εργασίες σχετικά με LLM-assisted malware generation. Στη δημοσίευση “A Compiler Framework Exploiting Compositional Blindness in Aligned LLMs for Malware Generation”, οι συγγραφείς παρουσίασαν framework μέσω του οποίου μεγάλα γλωσσικά μοντέλα μπορούν να αξιοποιηθούν για τη σύνθεση και τον κατακερματισμό malware-related εργασιών σε επιμέρους υπο-βήματα [61]. Η προσέγγιση αυτή επιτρέπει την παραγωγή malware components μέσω compositional διαδικασιών, παρακάμπτοντας σε ορισμένες περιπτώσεις περιορισμούς ασφαλείας των ίδιων των μοντέλων [61]. Παράλληλα, οι συγγραφείς του δημοσιεύματος “Generative AI in cybersecurity: A comprehensive review of LLM applications and vulnerabilities” επισημαίνουν ότι τα LLMs μπορούν να επιταχύνουν τη διαδικασία ανάπτυξης malware μέσω automated code assistance και script adaptation, μειώνοντας τον χρόνο που απαιτείται για την παραγωγή ή τροποποίηση κακόβουλου λογισμικού [54].

Η σημαντικότερη επιχειρησιακή μεταβολή που προκύπτει από την AI-assisted παραγωγή malware αφορά τη μετάβαση από στατικές και περισσότερο χειροκίνητες διαδικασίες ανάπτυξης σε highly automated και adaptive malware ecosystems. Η AI επιτρέπει μεγαλύτερη ταχύτητα ανάπτυξης, αυξημένη δυνατότητα evasion και πιο αποτελεσματική προσαρμογή των επιθέσεων σε διαφορετικά περιβάλλοντα στόχων [96], [52], [39]. Επιπλέον, η συνδυαστική χρήση automation, generative AI και malware-as-a-service πρακτικών αυξάνει τη διαθεσιμότητα κακόβουλων εργαλείων και ενισχύει την industrialization του κυβερνοεγκλήματος [41].

5.3. Επιθέσεις με χρήση automation tools

Η αυτοματοποίηση αποτελεί διαχρονικά βασικό χαρακτηριστικό των κυβερνοεπιθέσεων, ωστόσο η ενσωμάτωση τεχνολογιών τεχνητής νοημοσύνης μεταβάλλει σημαντικά το επίπεδο αυτονομίας, προσαρμοστικότητας και επιχειρησιακής αποτελεσματικότητας των σύγχρονων automation tools. Παραδοσιακά, τα αυτοματοποιημένα εργαλεία κυβερνοεπιθέσεων περιορίζονταν κυρίως σε προκαθορισμένα scripts, exploit kits ή botnets που εκτελούσαν συγκεκριμένες ενέργειες βάσει στατικών κανόνων. Η σύγχρονη χρήση AI επιτρέπει τη δημιουργία περισσότερο δυναμικών συστημάτων, τα οποία μπορούν να προσαρμόζονται στο περιβάλλον στόχο, να λαμβάνουν αποφάσεις βάσει δεδομένων και να αυτοματοποιούν πολλαπλά στάδια του cyber kill chain [46], [41].

Η AI-assisted αυτοματοποίηση επηρεάζει κυρίως τη reconnaissance διαδικασία, την αναζήτηση ευπαθειών, τις credential attacks επιχειρήσεις, την αυτοματοποίηση phishing campaigns και την προσαρμογή επιθετικών workflows. Σύμφωνα με το άρθρο “A Survey on Large Language Model-based Agents in Cybersecurity.”, τα LLM-based agents μπορούν να αναλαμβάνουν επιμέρους λειτουργίες offensive cybersecurity workflows, όπως συλλογή πληροφοριών, vulnerability discovery, script generation και interaction με διαφορετικά συστήματα [46]. Η εξέλιξη αυτή διαφοροποιεί ουσιαστικά τα σύγχρονα automation tools από τα παραδοσιακά static scripts, καθώς τα AI-enabled συστήματα εμφανίζουν μεγαλύτερη ευελιξία και δυνατότητα επιχειρησιακής προσαρμογής.

Ιδιαίτερη σημασία έχει η μετάβαση από την απλή αυτοματοποίηση σε agentic AI μοντέλα. Σύμφωνα με τη μελέτη του Koch σχετικά με το Agentic AI and the Industrialization of Cyber Offense, τα σύγχρονα AI agents μπορούν να αυτοματοποιούν πολλαπλά στάδια επιθέσεων, να διατηρούν operational context και να προσαρμόζουν τη συμπεριφορά τους ανάλογα με τα δεδομένα που λαμβάνουν από το περιβάλλον στόχο [69]. Αυτό σημαίνει ότι τα automation tools δεν περιορίζονται πλέον στην εκτέλεση στατικών εντολών, αλλά αποκτούν δυνατότητες dynamic task execution, workflow orchestration και attack adaptation. Ως αποτέλεσμα, μειώνεται σημαντικά η ανάγκη συνεχούς ανθρώπινης παρέμβασης κατά τη διάρκεια μιας κυβερνοεπίθεσης.

Παράλληλα, η ανάπτυξη cybercrime-as-a-service και AI-enabled underground εργαλείων αυξάνει τη διαθεσιμότητα αυτοματοποιημένων επιθετικών δυνατοτήτων. Η KEΛΑ αναφέρει ότι σε underground communities διατίθενται AI-assisted phishing kits, reconnaissance tools, credential harvesting frameworks και malware automation εργαλεία, τα οποία επιτρέπουν ακόμη και σε μη εξειδικευμένους χρήστες να πραγματοποιούν σύνθετες επιθετικές ενέργειες [39]. Η αυτοματοποίηση αυτή μεταβάλλει την οικονομία του κυβερνοεγκλήματος, καθώς μειώνει το κόστος ανάπτυξης επιθέσεων και επιτρέπει τη μαζική επανάληψή τους.

Ένα χαρακτηριστικό real-life case αφορά τη χρήση AI-assisted automation frameworks σε phishing και credential theft επιχειρήσεις, όπως περιγράφεται στις αναφορές της OpenAI και της Anthropic. Στις συγκεκριμένες περιπτώσεις, threat actors αξιοποίησαν generative AI και automation εργαλεία για τη δημιουργία phishing περιεχομένου, τη διαχείριση επιθετικών workflows και τη μαζική προσωποποίηση επιθέσεων. Η χρήση AI επέτρεψε τη γρήγορη

προσαρμογή μηνυμάτων, τη δημιουργία μεγάλου όγκου επιθετικού περιεχομένου και τη μείωση του χρόνου προετοιμασίας επιθέσεων [41], [97]. Το σημαντικότερο στοιχείο των περιστατικών αυτών είναι ότι η ΑΙ χρησιμοποιήθηκε όχι μόνο για content generation, αλλά και για operational orchestration επιθετικών διαδικασιών.

Επιπλέον, ιδιαίτερο ενδιαφέρον παρουσιάζουν τα AI-driven penetration testing και autonomous offensive systems. Περιγράφονται multi-agent workflows μέσω των οποίων AI agents μπορούν να εκτελούν penetration-testing διαδικασίες, να διατηρούν context information και να προσαρμόζουν τις ενέργειές τους κατά τη διάρκεια πολύπλοκων workflows [79]. Αντίστοιχα, στο άρθρο “Autonomous Pentesting Using Artificial Intelligence”, οι συγγραφείς αναλύουν autonomous penetration testing συστήματα που αξιοποιούν ΑΙ για reconnaissance, vulnerability exploitation και attack chaining [89]. Παρότι οι συγκεκριμένες τεχνολογίες αναπτύσσονται και για αμυντικούς ή ερευνητικούς σκοπούς, οι δυνατότητές τους μπορούν να αξιοποιηθούν και σε κακόβουλα offensive scenarios, ενισχύοντας την αυτοματοποίηση και την επιχειρησιακή αποτελεσματικότητα των επιθέσεων.

Η χρήση automation tools επηρεάζει σημαντικά και την κλίμακα των επιθέσεων. Η δυνατότητα αυτοματοποιημένης διαχείρισης μεγάλου αριθμού στόχων, η ταυτόχρονη εκτέλεση reconnaissance ενεργειών και η αυτοματοποίηση credential attacks ή phishing campaigns αυξάνουν δραστικά τη δυνατότητα scaling κυβερνοεπιχειρήσεων [46], [41]. Παράλληλα, cloud υποδομές και AI-as-a-Service περιβάλλοντα ενισχύουν ακόμη περισσότερο αυτή τη δυναμική, επιτρέποντας την αξιοποίηση remote computing πόρων και την εκτέλεση distributed attack workflows [39], [97].

Σημαντική μεταβολή παρατηρείται επίσης στη σχέση μεταξύ τεχνικής εξειδίκευσης και επιθετικής δυνατότητας. Παραδοσιακά, η ανάπτυξη και διαχείριση automation tools απαιτούσε υψηλό επίπεδο προγραμματιστικών γνώσεων και εμπειρίας σε offensive security διαδικασίες. Σήμερα, η ενσωμάτωση AI interfaces, natural language interaction και pretrained AI models επιτρέπει τη χρήση πιο σύνθετων επιθετικών εργαλείων ακόμη και από actors με περιορισμένες τεχνικές γνώσεις [69], [97]. Η εξέλιξη αυτή μειώνει τα barriers to entry στο κυβερνοέγκλημα και ενισχύει τη democratization των επιθετικών δυνατοτήτων.

6. Η ΧΡΗΣΗ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΟ ΚΥΒΕΡΝΟΕΓΚΛΗΜΑ

6.1. AI-as-a-Service

Το AI-as-a-Service αναφέρεται στο μοντέλο παροχής υπηρεσιών τεχνητής νοημοσύνης μέσω διαδικτύου ή cloud υποδομών, χωρίς ο χρήστης να χρειάζεται να αναπτύξει, να εκπαιδεύσει ή να διαχειριστεί ο ίδιος τα αντίστοιχα μοντέλα. Με αυτόν τον τρόπο, προηγμένες δυνατότητες όπως η ανάλυση δεδομένων, η παραγωγή περιεχομένου, η αυτοματοποίηση διαδικασιών και η χρήση μεγάλων γλωσσικών μοντέλων καθίστανται άμεσα προσβάσιμες σε ευρύ φάσμα χρηστών, μειώνοντας σημαντικά τις τεχνικές και οικονομικές απαιτήσεις αξιοποίησης της τεχνητής νοημοσύνης.

Η ανάπτυξη της τεχνητής νοημοσύνης δεν χαρακτηρίζεται μόνο από τη δημιουργία ολοένα ισχυρότερων αλγορίθμων και μοντέλων, αλλά και από τη μεταβολή του τρόπου πρόσβασης σε αυτές τις τεχνολογίες. Η έννοια του Artificial Intelligence as-a-Service, ή AIaaS, αναφέρεται στο μοντέλο παροχής δυνατοτήτων τεχνητής νοημοσύνης μέσω διαδικτυακών και cloud υποδομών, χωρίς ο τελικός χρήστης να χρειάζεται να αναπτύξει, να εκπαιδεύσει ή να φιλοξενήσει ο ίδιος τα αντίστοιχα μοντέλα. Με τον τρόπο αυτό, λειτουργίες όπως η παραγωγή κειμένου, η ανάλυση δεδομένων, η αυτοματοποίηση διαδικασιών, η επεξεργασία φυσικής γλώσσας και η χρήση προεκπαιδευμένων μοντέλων καθίστανται προσβάσιμες ως υπηρεσίες [112], [113].

Η σημασία του AI-as-a-Service έγκειται στο γεγονός ότι μεταβάλλει τη σχέση μεταξύ τεχνολογικής δυνατότητας και πρόσβασης. Στο παρελθόν, η αξιοποίηση προηγμένων τεχνικών τεχνητής νοημοσύνης απαιτούσε εξειδικευμένο ανθρώπινο δυναμικό, μεγάλα σύνολα δεδομένων, ισχυρή υπολογιστική υποδομή, χρόνο εκπαίδευσης μοντέλων και σημαντικό οικονομικό κόστος. Η διάθεση AI δυνατοτήτων μέσω cloud υπηρεσιών μεταφέρει μεγάλο μέρος αυτής της πολυπλοκότητας στον πάροχο της υπηρεσίας, επιτρέποντας στον χρήστη να αξιοποιεί έτοιμες λειτουργίες μέσω πλατφορμών, APIs ή διαδικτυακών εφαρμογών [9], [112], [113].

Η ανάπτυξη του AIaaS συνδέεται άμεσα με την εξέλιξη του cloud computing. Οι cloud υποδομές παρέχουν την υπολογιστική ισχύ, την επεκτασιμότητα, την αποθήκευση και την απομακρυσμένη διαθεσιμότητα που απαιτούνται για την παροχή υπηρεσιών τεχνητής νοημοσύνης σε μεγάλο αριθμό χρηστών. Χωρίς την ύπαρξη τέτοιων υποδομών, η ευρεία πρόσβαση σε προηγμένα

μοντέλα θα παρέμενε περιορισμένη σε οργανισμούς με σημαντικούς τεχνικούς και οικονομικούς πόρους. Το cloud, επομένως, δεν αποτελεί απλώς τεχνικό υπόβαθρο του AIaaS, αλλά βασική προϋπόθεση για την επιχειρησιακή του διάδοση [112], [114].

Ένας ακόμη παράγοντας που ενίσχυσε την ανάπτυξη του AIaaS είναι η διάδοση μεγάλων προεκπαιδευμένων μοντέλων. Τα μοντέλα αυτά μπορούν να αξιοποιηθούν σε διαφορετικά πεδία εφαρμογής χωρίς ο χρήστης να ξεκινά από μηδενική βάση. Η δυνατότητα πρόσβασης σε προεκπαιδευμένα μοντέλα μειώνει σημαντικά τις απαιτήσεις τεχνικής γνώσης, καθώς ο τελικός χρήστης δεν χρειάζεται να κατανοεί σε βάθος τη διαδικασία εκπαίδευσης, βελτιστοποίησης και συντήρησης του μοντέλου. Αυτό επιτρέπει την αξιοποίηση προηγμένων AI δυνατοτήτων από ευρύτερο φάσμα χρηστών και οργανισμών [9], [6], [50].

Η μείωση των τεχνικών εμποδίων αποτελεί ένα από τα βασικότερα χαρακτηριστικά του AI-as-a-Service. Η χρήση υπηρεσιών AI δεν απαιτεί απαραίτητα εξειδίκευση στη μηχανική μάθηση, στην ανάπτυξη νευρωνικών δικτύων ή στη διαχείριση υπολογιστικών υποδομών μεγάλης κλίμακας. Ο χρήστης μπορεί να αλληλεπιδρά με την υπηρεσία μέσω πιο απλών διεπαφών, έτοιμων εργαλείων ή APIs. Με αυτόν τον τρόπο, δυνατότητες που στο παρελθόν ήταν διαθέσιμες κυρίως σε μεγάλους οργανισμούς, ερευνητικά κέντρα ή τεχνικά εξειδικευμένες ομάδες γίνονται περισσότερο προσβάσιμες [112], [113].

Παράλληλα, το AIaaS μειώνει και τα οικονομικά εμπόδια πρόσβασης. Αντί ένας οργανισμός ή ένας χρήστης να επενδύσει σε ιδιόκτητη υποδομή, εξοπλισμό υψηλής υπολογιστικής ισχύος, ομάδες ειδικών και διαδικασίες εκπαίδευσης μοντέλων, μπορεί να αποκτήσει πρόσβαση σε AI δυνατότητες με βάση μοντέλα συνδρομής, χρήσης ή παροχής υπηρεσιών. Αυτό μεταβάλλει το κόστος από υψηλή αρχική επένδυση σε πιο ευέλικτο κόστος πρόσβασης. Η μεταβολή αυτή έχει ιδιαίτερη σημασία, διότι αυξάνει τον αριθμό των φορέων που μπορούν να αξιοποιήσουν προηγμένες τεχνολογίες τεχνητής νοημοσύνης [9], [6], [113].

Η αυξημένη προσβασιμότητα στην AI διευρύνει το κοινό το οποίο έχει πρόσβαση σε προηγμένες τεχνολογικές δυνατότητες πέρα από τα όρια των μεγάλων τεχνολογικών οργανισμών και των εξειδικευμένων ερευνητικών ομάδων. Μέσω του AIaaS, περισσότεροι χρήστες μπορούν να αξιοποιήσουν εργαλεία ανάλυσης, αυτοματοποίησης και παραγωγής περιεχομένου, χωρίς να διαθέτουν αντίστοιχη τεχνική υποδομή ή βαθιά επιστημονική εξειδίκευση [9], [6], [50].

Ωστόσο, η ίδια αυτή διεύρυνση πρόσβασης δημιουργεί και νέες προκλήσεις. Όταν προηγμένες δυνατότητες τεχνητής νοημοσύνης καθίστανται εύκολα διαθέσιμες, δεν αξιοποιούνται μόνο από νόμιμους χρήστες, επιχειρήσεις, ερευνητές και οργανισμούς, αλλά μπορούν να αξιοποιηθούν και από κακόβουλους δρώντες. Η σχέση αυτή είναι ιδιαίτερα κρίσιμη για τη μελέτη του κυβερνοεγκλήματος, καθώς η πρόσβαση σε ΑΙ υπηρεσίες μπορεί να μειώσει το επίπεδο τεχνικής εξειδίκευσης που απαιτείται για την εκτέλεση ορισμένων επιθετικών ενεργειών [42], [41], [40].

Στο πλαίσιο του κυβερνοεγκλήματος, το AIaaS μεταβάλλει τη δυναμική πρόσβασης σε προηγμένες δυνατότητες. Εργαλεία παραγωγής κειμένου μπορούν να αξιοποιηθούν για τη δημιουργία πειστικότερων phishing μηνυμάτων, υπηρεσίες ανάλυσης δεδομένων μπορούν να υποστηρίξουν την επεξεργασία πληροφοριών στόχων, ενώ δυνατότητες αυτοματοποίησης μπορούν να μειώσουν τον χρόνο προετοιμασίας επιθέσεων. Το κρίσιμο στοιχείο δεν είναι ότι κάθε AIaaS υπηρεσία είναι κακόβουλη, αλλά ότι οι ίδιες δυνατότητες που προσφέρονται για νόμιμες εφαρμογές μπορούν να χρησιμοποιηθούν κακόβουλα [41], [40], [70].

Η εξέλιξη αυτή συνδέεται με το ευρύτερο μοντέλο service-based κυβερνοεγκλήματος. Όπως το Cybercrime-as-a-Service επιτρέπει σε χρήστες να αποκτούν πρόσβαση σε έτοιμες εγκληματικές υπηρεσίες, έτσι και η διάθεση ΑΙ δυνατοτήτων ως υπηρεσία μπορεί να ενισχύσει την επιχειρησιακή ικανότητα δρώντων που δεν διαθέτουν οι ίδιοι υψηλό επίπεδο τεχνικής γνώσης. Η διαφορά είναι ότι το AIaaS δεν αποτελεί εξ ορισμού παράνομο μοντέλο, αλλά νόμιμη τεχνολογική υποδομή η οποία μπορεί, υπό ορισμένες συνθήκες, να ενταχθεί σε κακόβουλα workflows [69].

Η σύνδεση του AIaaS με το κυβερνοεγκλημα δεν πρέπει να παρουσιαστεί ως απλή ταύτιση της τεχνολογίας με την κακόβουλη χρήση. Η τεχνητή νοημοσύνη ως υπηρεσία έχει σημαντικές θετικές εφαρμογές σε τομείς όπως η παραγωγικότητα, η ανάλυση δεδομένων, η κυβερνοάμυνα, η αυτοματοποίηση επιχειρησιακών διαδικασιών και η υποστήριξη λήψης αποφάσεων. Παράλληλα όμως, το ίδιο μοντέλο πρόσβασης μπορεί να αξιοποιηθεί από επιτιθέμενους για μείωση κόστους, αύξηση ταχύτητας και ενίσχυση της κλίμακας επιθετικών ενεργειών. Αυτή η αμφίπλευρη χρήση επιβεβαιώνει τον dual-use χαρακτήρα της ΑΙ στο ψηφιακό οικοσύστημα [45], [42], [50].

Η μείωση της τεχνικής πολυπλοκότητας μέσω AIaaS επηρεάζει άμεσα το ερευνητικό ερώτημα που αφορά το επίπεδο γνώσεων των επιτιθέμενων. Σε παραδοσιακά περιβάλλοντα, η ανάπτυξη προηγμένων εργαλείων απαιτούσε εξειδικευμένους προγραμματιστές, τεχνική εμπειρία και

πρόσβαση σε ειδικές υποδομές. Με το AIaaS, ο επιτιθέμενος μπορεί να αξιοποιήσει υπάρχουσες υπηρεσίες για να υποστηρίξει επιμέρους στάδια της επίθεσης, χωρίς να κατανοεί πλήρως το τεχνικό υπόβαθρο των μοντέλων. Έτσι, η AI δεν εξαλείφει την ανάγκη για τεχνική γνώση, αλλά μειώνει το επίπεδο γνώσης που απαιτείται για την αξιοποίηση ορισμένων προηγμένων δυνατοτήτων [42], [113], [41].

Αντίστοιχα, το AIaaS επηρεάζει το κόστος και την κλίμακα των κυβερνοεπιθέσεων. Όταν οι δυνατότητες AI παρέχονται ως υπηρεσία, το κόστος χρήσης τους γίνεται πιο ευέλικτο και η δυνατότητα επανάληψης ενεργειών αυξάνεται. Αυτό μπορεί να διευκολύνει τη μαζική παραγωγή περιεχομένου, την αυτοματοποίηση εργασιών και την ταχύτερη εκτέλεση επιμέρους σταδίων μιας επίθεσης. Ειδικά όταν οι υπηρεσίες αυτές συνδυάζονται με cloud υποδομές και αυτοματοποιημένα εργαλεία, η επιχειρησιακή κλιμάκωση καθίσταται πιο εύκολη [112], [114].

Η επίδραση του AIaaS στη φύση των κυβερνοεπιθέσεων είναι επομένως κυρίως επιχειρησιακή. Η βασική αλλαγή δεν βρίσκεται μόνο στο ότι οι επιθέσεις μπορούν να γίνουν πιο τεχνικά σύνθετες, αλλά στο ότι η πρόσβαση σε τεχνολογίες που υποστηρίζουν σύνθετες ενέργειες γίνεται περισσότερο διαθέσιμη. Αυτό σημαίνει ότι η δυναμική των επιθέσεων μετατοπίζεται από το ερώτημα «ποιος μπορεί να αναπτύξει την τεχνολογία» στο ερώτημα «ποιος μπορεί να αποκτήσει πρόσβαση και να τη χρησιμοποιήσει αποτελεσματικά» [9], [42], [50].

Η εξέλιξη αυτή δημιουργεί μια νέα μορφή ασυμμετρίας. Ένας οργανισμός μπορεί να χρειάζεται σημαντικούς πόρους για να προστατεύσει συστήματα, χρήστες, δεδομένα και υποδομές, ενώ ένας επιτιθέμενος μπορεί να αξιοποιεί έτοιμες υπηρεσίες για να υποστηρίξει στοχευμένες ή μαζικές ενέργειες. Το AIaaS, όταν χρησιμοποιείται καταχρηστικά, μπορεί να ενισχύσει αυτή την ασυμμετρία, διότι μειώνει το κόστος και τον χρόνο προετοιμασίας για την επιτιθέμενη πλευρά [42], [41], [69].

Παράλληλα, η αυξημένη διαθεσιμότητα AI υπηρεσιών δημιουργεί ζητήματα ελέγχου πρόσβασης, λογοδοσίας και υπεύθυνης χρήσης. Οι πάροχοι AIaaS καλούνται να ισορροπήσουν ανάμεσα στη διάθεση χρήσιμων τεχνολογιών και στην ανάγκη περιορισμού της κακόβουλης αξιοποίησής τους. Η ύπαρξη μηχανισμών ασφαλείας, πολιτικών χρήσης, συστημάτων παρακολούθησης κατάχρησης και διαδικασιών περιορισμού κακόβουλων δραστηριοτήτων αποκτά ιδιαίτερη σημασία σε ένα περιβάλλον όπου οι υπηρεσίες παρέχονται διεθνώς και σε μεγάλη κλίμακα [50], [40], [70].

Η διεθνής διάσταση του AIaaS επιτείνει τις προκλήσεις. Οι υπηρεσίες τεχνητής νοημοσύνης διατίθενται μέσω διαδικτύου και συχνά υπερβαίνουν τα εθνικά ρυθμιστικά πλαίσια. Αυτό δημιουργεί δυσκολίες στον έλεγχο, στη συμμόρφωση, στην επιβολή κανόνων και στη διερεύνηση κακόβουλης χρήσης. Η διασυνοριακή φύση των υπηρεσιών αυτών συνδέει το AIaaS όχι μόνο με τεχνικά ζητήματα, αλλά και με ζητήματα διακυβέρνησης, ευθύνης και διεθνούς συνεργασίας [45], [50].

Η σημασία του AI-as-a-Service βρίσκεται στη μετατροπή προηγμένων δυνατοτήτων AI σε υπηρεσίες ευρέως προσβάσιμες μέσω ψηφιακών πλατφορμών. Η εξέλιξη αυτή μειώνει τεχνικά και οικονομικά εμπόδια, αλλά παράλληλα αυξάνει τη σημασία της υπεύθυνης χρήσης, της διακυβέρνησης και των μηχανισμών πρόληψης κακόβουλης αξιοποίησης. Ως εκ τούτου, το AI-as-a-Service αποτελεί ταυτόχρονα παράγοντα τεχνολογικής καινοτομίας και πεδίο αυξημένης κυβερνοασφαλιστικής ευθύνης.

6.2. Dark web tools

Η ενσωμάτωση της τεχνητής νοημοσύνης στο κυβερνοέγκλημα δεν περιορίζεται στη μεμονωμένη χρήση νόμιμων εργαλείων από κακόβουλους χρήστες. Ένα ιδιαίτερα κρίσιμο φαινόμενο είναι η εμφάνιση, διαφήμιση και εμπορευματοποίηση AI εργαλείων σε underground οικοσυστήματα, όπως dark web forums, κλειστές κοινότητες, ιδιωτικά κανάλια επικοινωνίας, Telegram συνομιλίες, cybercrime marketplaces και search engines που λειτουργούν ως σημεία συγκέντρωσης παράνομων υπηρεσιών. Στο πλαίσιο αυτό, η τεχνητή νοημοσύνη δεν λειτουργεί μόνο ως τεχνικό μέσο επίθεσης, αλλά και ως προϊόν που μπορεί να διακινηθεί, να προωθηθεί και να ενταχθεί σε ευρύτερα μοντέλα κυβερνοεγκληματικής δραστηριότητας [115], [39], [103].

Η μετάβαση από το νόμιμο AI-as-a-Service προς τα κακόβουλα AI εργαλεία αποτελεί ουσιαστικά μορφή καταχρηστικής επέκτασης της ίδιας τεχνολογικής λογικής. Όταν δυνατότητες όπως η παραγωγή κειμένου, η δημιουργία εικόνας ή φωνής, η αυτοματοποίηση διαδικασιών και η ανάλυση δεδομένων καθίστανται ευρύτερα προσβάσιμες, μπορούν να αξιοποιηθούν όχι μόνο από νόμιμους χρήστες, αλλά και από χρήστες που έχουν ως σκοπό να προκαλέσουν ζημιά. Στα underground οικοσυστήματα, οι δυνατότητες αυτές συχνά παρουσιάζονται ως εργαλεία για

παραγωγή phishing περιεχομένου, δημιουργία scam σεναρίων, υποστήριξη κακόβουλου κώδικα, deepfake πλαστοπροσωπία ή αυτοματοποίηση επιθετικών workflows [39], [41], [96].

Το Dark Web και οι συναφείς υπόγειες κοινότητες δεν πρέπει να αντιμετωπίζονται απλώς ως απομονωμένοι χώροι παράνομης δραστηριότητας, αλλά ως δυναμικά οικοσυστήματα ανταλλαγής τεχνογνωσίας, διαφήμισης υπηρεσιών και εμπορευματοποίησης εργαλείων. Σε αυτά τα περιβάλλοντα, οι δυνατότητες της ΑΙ μπορούν να εμφανιστούν είτε ως αυτόνομα εργαλεία είτε ως υπηρεσίες που ενσωματώνονται σε ήδη υπάρχοντα εγκληματικά μοντέλα, όπως phishing-as-a-service, fraud-as-a-service και cybercrime-as-a-service. Έτσι, η επίθεση δεν εξαρτάται αποκλειστικά από την τεχνική επάρκεια του μεμονωμένου δράστη, αλλά από την πρόσβασή του σε εργαλεία, υπηρεσίες και κοινότητες που μειώνουν το κόστος και την πολυπλοκότητα της προετοιμασίας [69], [39].

Η παρουσία ΑΙ εργαλείων σε τέτοια οικοσυστήματα απαντά άμεσα στο ερευνητικό ερώτημα σχετικά με τη μείωση των τεχνικών απαιτήσεων των επιτιθέμενων. Ένας λιγότερο εξειδικευμένος χρήστης δεν χρειάζεται απαραίτητα να αναπτύξει από την αρχή ένα μοντέλο τεχνητής νοημοσύνης, να εκπαιδεύσει αλγόριθμους ή να δημιουργήσει δικό του κακόβουλο περιεχόμενο με πλήρως χειροκίνητο τρόπο. Αντιθέτως, μπορεί να αναζητήσει και να αποκτήσει πρόσβαση, με once in a lifetime αγορά ή μηνιαία ενοικίαση, σε έτοιμα εργαλεία που φημίζονται ως οι πλέον καταλληλότερες λύσεις για phishing, scam campaigns, deepfake απάτες ή υποστήριξη αυτοματοποιημένων επιθετικών ενεργειών [115], [39], [120]. Σε αυτό το σημείο αξίζει να σημειωθεί ότι παρατηρούμε, πως δεν υπάρχουν δοκιμές των εργαλείων ή δωρεάν εκδόσεις τους. Τις περισσότερες φορές παρατηρήθηκε μία δωρεαν επισκόπηση των δυνατοτήτων του εργαλείου, αλλά ποτέ free trial. Τέλος για την πληρωμή ζητούνται αποκλειστικά εικονικά νομίσματα.

Τα dark web ΑΙ εργαλεία μπορούν να ταξινομηθούν σε τέσσερις βασικές κατηγορίες. Η πρώτη κατηγορία αφορά τα dark ΑΙ chatbots ή dark LLMs, δηλαδή εργαλεία που παρουσιάζονται ως «χωρίς περιορισμούς» εκδοχές μεγάλων γλωσσικών μοντέλων. Η δεύτερη αφορά deepfake, face-swapping και voice cloning υπηρεσίες που αξιοποιούνται για πλαστοπροσωπία. Η τρίτη αφορά automation εργαλεία, τα οποία υπόσχονται υποστήριξη σε scripting, reconnaissance ή επιθετικά workflows. Η τέταρτη αφορά fraud και scam εργαλεία, τα οποία συνδέονται ειδικότερα με

οικονομική απάτη, phishing, social engineering και παραγωγή παραπλανητικού περιεχομένου [116], [119], [123].

Στην κατηγορία των dark AI chatbots, ένα από τα πιο γνωστά παραδείγματα είναι το WormGPT. Το εργαλείο αυτό αναφέρθηκε ως GPT-like σύστημα που το ανακαλύψαμε σε forums και συνδέθηκε κυρίως με phishing και Business Email Compromise σενάρια [124], [117]. Η σημασία του WormGPT δεν έγκειται αποκλειστικά στις τεχνικές του δυνατότητας, αλλά στο γεγονός ότι αποτέλεσε χαρακτηριστικό παράδειγμα προσπάθειας μετατροπής των δυνατοτήτων των μεγάλων γλωσσικών μοντέλων σε εμπορεύσιμο εργαλείο εγκληματικής χρήσης.

Η περίπτωση του WormGPT αναδεικνύει τον τρόπο με τον οποίο η ΑΙ μπορεί να ενισχύσει τις επιθέσεις social engineering. Ένας επιτιθέμενος που δεν διαθέτει υψηλή γλωσσική επάρκεια ή εμπειρία στη σύνταξη πειστικών επιχειρησιακών μηνυμάτων μπορεί να αξιοποιήσει ένα τέτοιο εργαλείο για να δημιουργήσει πιο φυσικό και προσαρμοσμένο σε κάθε θύμα περιεχόμενο. Με αυτόν τον τρόπο, η απειλή δεν ενισχύεται μόνο σε τεχνικό επίπεδο, αλλά και σε επικοινωνιακό και ψυχολογικό επίπεδο, καθώς η πειστικότητα αποτελεί βασικό παράγοντα επιτυχίας των phishing και BEC επιθέσεων [124], [117].

Η Kaspersky έχει αναφέρει ότι, μετά τη δημοσιότητα γύρω από το εργαλείο, εμφανίστηκαν ψεύτικες σελίδες και διαφημίσεις που προσποιούνταν ότι προσφέρουν πρόσβαση στο WormGPT, με χαρακτηριστικά απάτης και στόχο άλλους κακόβουλους ή περίεργους χρήστες [125]. Το στοιχείο αυτό είναι σημαντικό, διότι δείχνει ότι η underground αγορά δεν περιλαμβάνει μόνο πραγματικά εργαλεία, αλλά και απάτες μεταξύ κυβερνοεγκληματιών, αντιγραφές και προϊόντα υπερβολικής διαφήμισης.

Στην ίδια κατηγορία εντάσσονται και άλλα εργαλεία που αναφέρονται ως dark ή uncensored chatbots, όπως τα DarkGPT, DarkBard, EvilGPT, FraudGPT και άλλα GPT-like παράγωγα. Οι αναφορές αυτές δείχνουν ότι το φαινόμενο δεν περιορίζεται σε ένα μοναδικό εργαλείο, αλλά αφορά ευρύτερη τάση δημιουργίας ή διαφήμισης εναλλακτικών LLMs για χρήση σε κυβερνοεγκληματικά περιβάλλοντα [116], [119], [127]. Παράλληλα αναφέρεται από αρκετούς χρήστες των forums ότι αμφισβητείται η αποτελεσματικότητα ορισμένων εργαλείων, γεγονός που δείχνει ότι η αγορά περιλαμβάνει τόσο πραγματικές δυνατότητες όσο και υπερβολικούς ισχυρισμούς για να προσεγγιστούν περισσότεροι χρήστες [116].

Η δεύτερη κατηγορία αφορά deepfake και voice cloning εργαλεία ή υπηρεσίες. Σε αυτήν την περίπτωση, η τεχνητή νοημοσύνη δεν χρησιμοποιείται μόνο για παραγωγή κειμένου, αλλά για παραποίηση εικόνας, προσώπου, φωνής ή βίντεο. Τέτοιες δυνατότητες μπορούν να ενισχύσουν romance scams, investment fraud, vishing, CEO fraud και άλλες μορφές social engineering που βασίζονται στην πλαστοπροσωπία. Η σημασία τους είναι ότι μεταφέρουν την AI-enabled απάτη από το γραπτό μήνυμα σε οπτικοακουστικό επίπεδο, όπου η εμπιστοσύνη του θύματος μπορεί να χειραγωγηθεί μέσω πιο ρεαλιστικών μέσων [122], [123].

Χαρακτηριστικό παράδειγμα είναι το Haotian AI, το οποίο έχει συνδεθεί με face-swapping και scam οικοσυστήματα, ιδίως σε περιπτώσεις romance scams και επενδυτικής απάτης [122]. Η περίπτωση αυτή δείχνει ότι οι εφαρμογές τεχνητής νοημοσύνης για παράνομες δραστηριότητες δεν περιορίζονται στα κλασικά cyberattacks, αλλά επεκτείνονται σε ευρύτερες μορφές ψηφιακής εξαπάτησης. Μέσω τέτοιων εργαλείων, οι επιτιθέμενοι μπορούν να δημιουργήσουν πιο πειστικές ψευδείς ταυτότητες και να ενισχύσουν την αληθοφάνεια της επικοινωνίας με το θύμα.

Παράλληλα, η ύπαρξη deepfake-as-a-service και voice cloning υπηρεσιών σε underground περιβάλλοντα δείχνει ότι η παραγωγή συνθετικών μέσων μπορεί να εμπορευματοποιηθεί με τρόπο παρόμοιο με άλλα εγκληματικά εργαλεία [123], [115]. Συχνά δεν πρόκειται για ένα σταθερό και επώνυμο εργαλείο, αλλά για υπηρεσίες που προσφέρονται κατά περίπτωση μέσα από κλειστά κανάλια, forums ή ιδιωτικές αγορές. Η μείωση του κόστους παραγωγής τέτοιου περιεχομένου αυξάνει τη δυνατότητα χρήσης του σε μεγαλύτερη κλίμακα και σε περισσότερα σενάρια απάτης.

Η τρίτη κατηγορία αφορά εργαλεία αυτοματοποίησης. Σε αυτήν την κατηγορία εντάσσονται πλατφόρμες που δεν παρουσιάζονται απλώς ως chatbots, αλλά ως πολυλειτουργικά περιβάλλοντα που υπόσχονται υποστήριξη σε κακόβουλα workflows. Παράδειγμα αποτελεί το Xanthorox AI, το οποίο αναφέρθηκε ως self-hosted και modular malicious AI platform που προωθήθηκε σε κυβερνοεγκληματικές κοινότητες [121]. Η σημασία του Xanthorox AI βρίσκεται στο ότι παρουσιάζεται ως εργαλείο ευρύτερης αυτοματοποίησης και όχι μόνο ως μηχανισμός παραγωγής κειμένου.

Τέτοιου τύπου εργαλεία αναδεικνύουν τη μετατόπιση από την AI ως απλό βοηθό περιεχομένου προς την AI ως υποδομή υποστήριξης επιθετικών διαδικασιών. Ένα automation-oriented dark AI tool μπορεί να διαφημίζεται ως μέσο υποστήριξης scripting, phishing, social engineering,

vulnerability research ή διαχείρισης επιμέρους σταδίων μιας επίθεσης. Ακόμη και όταν οι σχετικοί ισχυρισμοί πρέπει να αξιολογούνται με επιφύλαξη, η ύπαρξη τέτοιων διαφημίσεων δείχνει ότι υπάρχει ζήτηση για AI εργαλεία που μειώνουν την ανάγκη χειροκίνητης εργασίας και τεχνικής εξειδίκευσης [121], [39].

Η τέταρτη κατηγορία αφορά fraud και scam εργαλεία. Το πιο χαρακτηριστικό παράδειγμα είναι το FraudGPT, το οποίο αναφέρθηκε ως generative AI εργαλείο που διαφημίστηκε σε dark web forums και Telegram channels για απάτες και κυβερνοεγκληματικές χρήσεις [126], [118], [39]. Το FraudGPT είναι σημαντικό παράδειγμα επειδή το ίδιο το εμπορικό του όνομα συνδέεται άμεσα με την οικονομική απάτη, το phishing, το social engineering και τη δημιουργία παραπλανητικού περιεχομένου.

Η περίπτωση του FraudGPT δείχνει ότι η AI μπορεί να ενσωματωθεί σε επιχειρησιακά μοντέλα απάτης χωρίς να απαιτείται απαραίτητα παραγωγή τεχνικά εξελιγμένου malware. Η ικανότητα παραγωγής πειστικών μηνυμάτων, scam scripts, ψεύτικων ειδοποιήσεων ή σεναρίων εξαπάτησης μπορεί να είναι από μόνη της επιχειρησιακά σημαντική. Με αυτόν τον τρόπο, η AI ενισχύει την απάτη σε επίπεδο γλώσσας, πειθούς, προσαρμογής και μαζικής παραγωγής περιεχομένου [126], [39].

6.3 Οικονομία κυβερνοεγκλήματος

Το σύγχρονο κυβερνοέγκλημα δεν αποτελεί πλέον μεμονωμένη τεχνική ενέργεια ενός και μόνο επιτιθέμενου. Αντιθέτως, έχει εξελιχθεί σε οργανωμένο οικονομικό έγκλημα, με το οποίο διακινούνται δεδομένα, πρόσβαση σε παραβιασμένα συστήματα, εργαλεία, υπηρεσίες, τεχνογνωσία και παράνομες ψηφιακές υποδομές. Η λειτουργία του οικοσυστήματος αυτού παρουσιάζει χαρακτηριστικά μιας πραγματικής αγοράς, καθώς περιλαμβάνει προμηθευτές, αγοραστές, διαμεσολαβητές, εξειδικευμένους παρόχους υπηρεσιών, συνδρομητικά μοντέλα και διαφορετικά επίπεδα τεχνικής εξειδίκευσης [22], [128], [70].

Η οικονομία του κυβερνοεγκλήματος στηρίζεται σε μεγάλο βαθμό στην εμπορευματοποίηση ψηφιακών πόρων. Κλεμμένα δεδομένα, κωδικοί πρόσβασης, τραπεζικές πληροφορίες, προσωπικά δεδομένα, cookies, εταιρικοί λογαριασμοί και πρόσβαση σε εσωτερικά δίκτυα μπορούν να αποκτήσουν οικονομική αξία μέσα σε παράνομες αγορές. Η Europol περιγράφει τα δεδομένα ως κεντρικό εμπόρευμα του κυβερνοεγκληματικού οικοσυστήματος, καθώς μπορούν να αποτελέσουν τόσο στόχο κλοπής όσο και μέσο για την πραγματοποίηση περαιτέρω εγκληματικών ενεργειών [22]. Με τον τρόπο αυτό, η παραβίαση ενός συστήματος δεν αποτελεί απαραίτητα το τελικό στάδιο της επίθεσης, αλλά μπορεί να λειτουργήσει ως αφετηρία για νέες μορφές εκμετάλλευσης.

Η οικονομική οργάνωση του κυβερνοεγκλήματος συνδέεται άμεσα με το μοντέλο cybercrime-as-a-service. Στο μοντέλο αυτό, διαφορετικές εγκληματικές δυνατότητες παρέχονται ως υπηρεσίες σε άλλους δράστες, οι οποίοι δεν χρειάζεται να διαθέτουν οι ίδιοι πλήρη τεχνική ικανότητα ή υποδομή. Η λογική αυτή περιλαμβάνει υποκατηγορίες όπως ransomware-as-a-service, phishing-as-a-service, malware-as-a-service, access brokerage και αγορές κλεμμένων διαπιστευτηρίων. Η σημασία του φαινομένου έγκειται στο ότι το κυβερνοέγκλημα αποκτά δομή καταμερισμού εργασίας, όπου κάθε actor μπορεί να εξειδικεύεται σε ένα επιμέρους στάδιο της εγκληματικής αλυσίδας [70], [21], [71].

Η ανάπτυξη του cybercrime-as-a-service μειώνει ουσιαστικά τα εμπόδια εισόδου στο κυβερνοέγκλημα. Ένας λιγότερο εξειδικευμένος δράστης δεν χρειάζεται απαραίτητα να αναπτύξει δικό του malware, να παραβιάσει ο ίδιος ένα δίκτυο ή να δημιουργήσει από την αρχή μια phishing υποδομή. Μπορεί να αγοράσει πρόσβαση, να νοικιάσει εργαλεία, να χρησιμοποιήσει έτοιμα phishing kits ή να συνεργαστεί με εξειδικευμένους παρόχους υπηρεσιών. Με αυτόν τον τρόπο, η τεχνική πολυπλοκότητα μεταφέρεται από τον τελικό χρήστη της εγκληματικής υπηρεσίας στον πάροχο που την αναπτύσσει και τη διαθέτει [70], [130], [71].

Ιδιαίτερα σημαντικό ρόλο στην οικονομία του κυβερνοεγκλήματος διαδραματίζουν οι Initial Access Brokers. Οι IABs εξειδικεύονται στην απόκτηση αρχικής πρόσβασης σε συστήματα και δίκτυα, την οποία στη συνέχεια πωλούν σε άλλους κακόβουλους δρώντες. Η ύπαρξή τους δείχνει ότι η πρόσβαση σε ένα παραβιασμένο περιβάλλον έχει μετατραπεί σε αυτόνομο προϊόν, το οποίο μπορεί να πωληθεί ανεξάρτητα από την τελική επίθεση που θα ακολουθήσει [22], [21], [130]. Με

τον τρόπο αυτό, η αρχική παραβίαση και η τελική εκμετάλλευση αποσυνδέονται επιχειρησιακά και οικονομικά.

Η αγορά πρόσβασης επιταχύνει σημαντικά την εκτέλεση επιθέσεων. Ένας ransomware operator ή ένας affiliate δεν χρειάζεται να αφιερώσει χρόνο στη φάση της αρχικής διείσδυσης, εφόσον μπορεί να αγοράσει πρόσβαση από έναν broker. Αυτό μειώνει τον χρόνο προετοιμασίας, αυξάνει την αποδοτικότητα και επιτρέπει σε διαφορετικές ομάδες να συνεργάζονται χωρίς να εκτελούν όλες τις φάσεις της επίθεσης εσωτερικά [130], [131]. Η εξέλιξη αυτή ενισχύει τη βιομηχανοποίηση του κυβερνοεγκλήματος, καθώς κάθε στάδιο μπορεί να εμπορευματοποιηθεί και να ανατεθεί σε διαφορετικό εξειδικευμένο ρόλο.

Αντίστοιχα, το ransomware-as-a-service αποτελεί χαρακτηριστικό παράδειγμα οικονομικής οργάνωσης του κυβερνοεγκλήματος. Στο μοντέλο αυτό, οι δημιουργοί ή διαχειριστές μιας ransomware υποδομής παρέχουν εργαλεία, τεχνική υποστήριξη, διαχείριση πληρωμών ή διαπραγματεύσεων και επιχειρησιακή καθοδήγηση σε affiliates, οι οποίοι αναλαμβάνουν την εκτέλεση επιθέσεων. Η ύπαρξη affiliates, leak sites, access brokers και οικονομικών διαμεσολαβητών δείχνει ότι το ransomware λειτουργεί πλέον ως καταναμημένο οικονομικό οικοσύστημα και όχι απλώς ως μεμονωμένο κακόβουλο λογισμικό [21], [129], [131].

Το phishing-as-a-service και η οικονομία των phishing kits αποτελούν επίσης κρίσιμες διαστάσεις της αγοράς κυβερνοεγκλήματος. Έτοιμα phishing kits, σελίδες-αντίγραφα νόμιμων υπηρεσιών, μηχανισμοί συλλογής διαπιστευτηρίων και υποδομές αποστολής μηνυμάτων μπορούν να διατεθούν σε χρήστες που δεν διαθέτουν επαρκή τεχνική γνώση για να τα δημιουργήσουν μόνοι τους. Η Flare περιγράφει την οικονομία των phishing kits ως μέρος των κυβερνοεγκληματικών αγορών, όπου η διάθεση έτοιμων εργαλείων αυξάνει τη δυνατότητα εκτέλεσης επιθέσεων σε μεγάλη κλίμακα [71]. Έτσι, η παραγωγή απάτης μετατρέπεται σε επαναλήψιμη και εμπορεύσιμη διαδικασία.

Η τεχνητή νοημοσύνη εντάσσεται σε αυτό το οικοσύστημα ως παράγοντας μείωσης κόστους και αύξησης αποδοτικότητας. Η AI μπορεί να χρησιμοποιηθεί για παραγωγή phishing περιεχομένου, μετάφραση και γλωσσική προσαρμογή μηνυμάτων, δημιουργία παραλλαγών κειμένου, αυτοματοποίηση απαντήσεων, ανάλυση στόχων και υποστήριξη scam campaigns. Η Europol επισημαίνει ότι τα LLMs και η generative AI μπορούν να βελτιώσουν την αποτελεσματικότητα

του social engineering, επιτρέποντας καλύτερη επικοινωνία με τα θύματα και αυτοματοποίηση εγκληματικών διαδικασιών [22]. Η UNODC, αντίστοιχα, συνδέει την ΑΙ και την αυτοματοποίηση με την αύξηση της αποδοτικότητας και της κλίμακας εγκληματικών πρακτικών [43].

Η μείωση του κόστους παραγωγής επιθετικού υλικού είναι ιδιαίτερα σημαντική για την οικονομία του κυβερνοεγκλήματος. Σε περιβάλλοντα phishing, fraud και social engineering, το κόστος δημιουργίας πειστικού περιεχομένου αποτελούσε παραδοσιακά παράγοντα περιορισμού, ιδίως όταν απαιτούνταν γλωσσική προσαρμογή, εξατομίκευση και επαναλαμβανόμενη επικοινωνία με το θύμα. Η ΑΙ μειώνει αυτό το κόστος, επιτρέποντας τη γρήγορη παραγωγή μεγάλου όγκου περιεχομένου και την προσαρμογή του σε διαφορετικές γλώσσες, πολιτισμικά περιβάλλοντα ή επαγγελματικά προφίλ [51], [39].

Η οικονομική σημασία αυτής της μεταβολής έγκειται στο ότι το χαμηλό κόστος αποτυχίας καθιστά βιώσιμη τη μαζική επανάληψη επιθέσεων. Στο κυβερνοέγκλημα, μια επίθεση δεν χρειάζεται να επιτύχει σε μεγάλο ποσοστό θυμάτων για να αποφέρει κέρδος, εφόσον το κόστος δημιουργίας και αποστολής της παραμένει χαμηλό. Η τεχνητή νοημοσύνη ενισχύει αυτή τη λογική, καθώς επιτρέπει σε κακόβουλους δρώντες να παράγουν περισσότερες παραλλαγές περιεχομένου, να δοκιμάζουν διαφορετικές προσεγγίσεις και να κλιμακώνουν καμπάνιες εξαπάτησης με περιορισμένη πρόσθετη δαπάνη [51], [27], [134].

Παράλληλα, η ΑΙ ενισχύει το μοντέλο crime-as-a-service, διότι μπορεί να ενσωματωθεί ως επιμέρους υπηρεσία σε ήδη υπάρχουσες εγκληματικές αγορές. Ένα phishing kit μπορεί να εμπλουτιστεί με ΑΙ-generated περιεχόμενο, ενώ μια fraud υπηρεσία μπορεί να αξιοποιεί ΑΙ για προσωποποιημένα μηνύματα. Με αυτόν τον τρόπο, η ΑΙ δεν δημιουργεί από μόνη της την εγκληματική αγορά, αλλά αυξάνει την αποδοτικότητα και την ελκυστικότητα των υπηρεσιών που ήδη κυκλοφορούν σε αυτή [70], [71], [39].

Το κυβερνοέγκλημα χαρακτηρίζεται επίσης από καταμερισμό ρόλων. Δεν συμμετέχουν όλοι οι δρώντες με τον ίδιο τρόπο ούτε απαιτούνται οι ίδιες δεξιότητες για κάθε στάδιο. Υπάρχουν developers κακόβουλων εργαλείων, operators ransomware, affiliates, initial access brokers, δημιουργοί phishing kits, πωλητές κλεμμένων δεδομένων, διαχειριστές botnets, money mules, crypto intermediaries και διαμεσολαβητές αγορών. Η ύπαρξη αυτών των ρόλων δείχνει ότι το

κυβερνοέγκλημα λειτουργεί ως κατανεμημένο οικονομικό σύστημα, όπου διαφορετικοί φορείς συμβάλλουν σε διαφορετικά στάδια της εγκληματικής αλυσίδας [128], [70], [130].

Η τεχνητή νοημοσύνη μπορεί να ενισχύσει πολλούς από αυτούς τους ρόλους. Οι δημιουργοί phishing kits μπορούν να χρησιμοποιούν ΑΙ για παραγωγή πιο πειστικών σελίδων ή μηνυμάτων. Οι fraud operators μπορούν να επικοινωνούν με τα θύματα αυτόματα. Οι πωλητές δεδομένων μπορούν να αξιοποιούν εργαλεία ανάλυσης για ταξινόμηση και αξιολόγηση πληροφοριών. Οι ransomware affiliates μπορούν να υποστηρίζονται από εργαλεία που επιταχύνουν reconnaissance ή victim profiling. Έτσι, η ΑΙ λειτουργεί ως οριζόντιος πολλαπλασιαστής δυνατοτήτων σε διαφορετικά επίπεδα της κυβερνοεγκληματικής οικονομίας [43], [32], [39].

Η λειτουργία των Initial Access Brokers αποκαλύπτει ιδιαίτερα καθαρά την οικονομική λογική του κυβερνοεγκλήματος. Η πρόσβαση σε ένα σύστημα αποκτά τιμή, περιγραφή, επίπεδο προνομίων, κατηγοριοποίηση ανά κλάδο ή χώρα και μπορεί να μεταπωληθεί σε άλλους δρώντες. Η Cyberint αναλύει την αγορά των IABs ως οργανωμένο πεδίο όπου η πρόσβαση σε εταιρικά συστήματα προσφέρεται σε underground forums και marketplaces, συχνά με διαφοροποίηση τιμών ανάλογα με το είδος και την αξία της πρόσβασης [130]. Το στοιχείο αυτό δείχνει ότι ακόμη και η είσοδος σε ένα δίκτυο έχει μετατραπεί σε εμπορεύσιμο αγαθό.

Η ΑΙ μπορεί να επηρεάσει και αυτό το επίπεδο της οικονομίας, διότι μπορεί να υποστηρίξει την αναγνώριση στόχων, την ανάλυση δημοσίων πληροφοριών, τη δημιουργία εξατομικευμένων μηνυμάτων αρχικής πρόσβασης και την αυτοματοποίηση επιμέρους διαδικασιών. Αν και η πρόσβαση σε ένα δίκτυο εξακολουθεί να απαιτεί τεχνικές ή κοινωνικές μεθόδους, η ΑΙ μπορεί να μειώσει τον χρόνο και την προσπάθεια που απαιτούνται για τη συλλογή πληροφοριών και την προετοιμασία της επίθεσης [43], [32], [33].

Η βιομηχανοποίηση του κυβερνοεγκλήματος ενισχύεται περαιτέρω από την ταχύτητα με την οποία οι κακόβουλοι δρώντες μπορούν να αξιοποιούν νέες ευπάθειες και εργαλεία. Η Fortinet αναφέρεται στη βιομηχανοποίηση του κυβερνοεγκλήματος και στην αυτοματοποίηση διαδικασιών reconnaissance, επισημαίνοντας ότι ο χρόνος μεταξύ γνωστοποίησης μιας ευπάθειας και εκμετάλλευσής της συμπίπτει σημαντικά [33]. Η εξέλιξη αυτή έχει οικονομική σημασία, διότι αυξάνει την αξία της ταχύτητας και της αυτοματοποίησης ως ανταγωνιστικών πλεονεκτημάτων μέσα στην αγορά κυβερνοεγκλήματος.

Η οικονομία του κυβερνοεγκλήματος επηρεάζεται επίσης από τις χρηματοοικονομικές υποδομές που επιτρέπουν την είσπραξη, μεταφορά και απόκρυψη παράνομων εσόδων. Τα κρυπτονομίσματα, οι υπηρεσίες ανάμειξης, οι darknet markets και τα crypto-enabled scams συνδέονται με την οικονομική λειτουργία του κυβερνοεγκλήματος, ιδίως σε περιπτώσεις ransomware, απάτης και παράνομων αγορών [132], [133], [134]. Η chainalysis αναλύει την οικονομική διάσταση του crypto crime, συμπεριλαμβανομένων ransomware, scams και darknet markets, δείχνοντας ότι η οικονομία του κυβερνοεγκλήματος δεν περιορίζεται στην τεχνική επίθεση, αλλά περιλαμβάνει και μηχανισμούς χρηματοδότησης και ρευστοποίησης [132].

Η ΑΙ μπορεί να ενισχύσει και την οικονομία των scams, ιδίως όταν συνδυάζεται με crypto-enabled fraud. Η δυνατότητα παραγωγής πειστικών μηνυμάτων, ψεύτικων ταυτοτήτων, προσωποποιημένων επενδυτικών προτάσεων ή μαζικών scam campaigns αυξάνει την αποδοτικότητα των απατών. Η European Payments Council συνδέει την ΑΙ με αυτοματοποίηση, κλιμάκωση και προσωποποίηση fraud και phishing επιθέσεων, ενώ η chainalysis αναδεικνύει τη σημασία των βιομηχανοποιημένων scams στο ευρύτερο crypto crime οικοσύστημα [51], [134].

Παρά την ενίσχυση που προσφέρει η ΑΙ, είναι σημαντικό να αποφευχθεί η υπερβολή. Η τεχνητή νοημοσύνη δεν δημιουργεί από μόνη της το κυβερνοέγκλημα ούτε αντικαθιστά πλήρως τις υπάρχουσες εγκληματικές υποδομές. Οι αγορές πρόσβασης, τα phishing kits, οι ransomware affiliates, οι data brokers και οι οικονομικοί διαμεσολαβητές προϋπήρχαν της ευρείας διάδοσης των σύγχρονων ΑΙ εργαλείων. Αυτό που αλλάζει είναι ότι η ΑΙ μπορεί να λειτουργήσει ως επιταχυντής και πολλαπλασιαστής δυνατοτήτων μέσα σε ήδη υπάρχοντα οικονομικά και οργανωτικά σχήματα [22], [43], [70].

Η κριτική αυτή προσέγγιση είναι απαραίτητη για την ακαδημαϊκή αποτίμηση του φαινομένου. Δεν αρκεί να υποστηριχθεί ότι η ΑΙ «κάνει το κυβερνοέγκλημα ευκολότερο». Πρέπει να αναλυθεί πού ακριβώς μειώνει το κόστος, σε ποια στάδια αυξάνει την αποδοτικότητα και ποιους ρόλους ενισχύει. Φαίνεται, η τεχνητή νοημοσύνη, να επηρεάζει κυρίως την παραγωγή περιεχομένου, την αυτοματοποίηση επαναλαμβανόμενων διαδικασιών, την εξατομίκευση social engineering, την ανάλυση στόχων και τη δυνατότητα μαζικής κλιμάκωσης επιθέσεων [43], [51], [39].

Η οικονομία του κυβερνοεγκλήματος αποτελεί κρίσιμο πλαίσιο για την κατανόηση της επίδρασης της τεχνητής νοημοσύνης στις κυβερνοεπιθέσεις. Η ΑΙ δεν λειτουργεί απομονωμένα, αλλά

εντάσσεται σε ένα ήδη ανεπτυγμένο οικοσύστημα υπηρεσιών, αγορών, εργαλείων και ρόλων. Η βασική της επίδραση είναι ότι μειώνει το κόστος, αυξάνει την αποδοτικότητα, ενισχύει την κλίμακα και διευκολύνει τη συμμετοχή περισσότερων δρώντων στην αγορά του κυβερνοεγκλήματος [22], [43], [70], [39].

6.4. Προκλήσεις προς αρχές επιβολής του νόμου

Η χρήση της τεχνητής νοημοσύνης στο κυβερνοέγκλημα δημιουργεί νέες και σύνθετες προκλήσεις για τις αρχές επιβολής του νόμου, διότι μεταβάλλει όχι μόνο τις τεχνικές δυνατότητες των επιτιθέμενων, αλλά και το ίδιο το περιβάλλον διερεύνησης, απόδοσης ευθύνης και ποινικής αντιμετώπισης των κυβερνοεγκλημάτων. Οι AI-assisted επιθέσεις μπορούν να εκτελούνται ταχύτερα, να εξαπλώνονται σε μεγαλύτερη κλίμακα και να παράγουν συνθετικό περιεχόμενο που δυσχεραίνει την αναγνώριση της αυθεντικότητας των αποδεικτικών στοιχείων. Κατά συνέπεια, οι αρχές δεν καλούνται απλώς να αντιμετωπίσουν περισσότερα περιστατικά, αλλά να προσαρμοστούν σε ένα περιβάλλον όπου η τεχνολογική πολυπλοκότητα, η ταχύτητα και η διεθνής διάσταση των επιθέσεων αυξάνονται ταυτόχρονα [30], [135], [43].

Μία από τις σημαντικότερες προκλήσεις είναι η διεθνής και διασυνοριακή φύση του κυβερνοεγκλήματος. Σε πολλές περιπτώσεις, ο δράστης, το θύμα, οι διακομιστές, οι πλατφόρμες επικοινωνίας, οι οικονομικές ροές και τα ψηφιακά αποδεικτικά στοιχεία βρίσκονται σε διαφορετικές χώρες. Αυτό δημιουργεί προβλήματα δικαιοδοσίας, καθυστέρησης, ανταλλαγής πληροφοριών και πρόσβασης σε δεδομένα. Η κυβερνοεπίθεση μπορεί να εκτελεστεί μέσα σε λίγα λεπτά, ενώ η συλλογή στοιχείων από άλλη έννομη τάξη μπορεί να απαιτεί θεσμικές διαδικασίες, αιτήματα δικαστικής συνδρομής και διεθνή συνεργασία [135], [147], [148].

Η κατάσταση περιπλέκεται περισσότερο από το γεγονός ότι τα εγκληματικά δίκτυα συχνά αξιοποιούν υποδομές και υπηρεσίες που βρίσκονται εκτός της χώρας όπου εντοπίζεται το θύμα. Η χρήση cloud υπηρεσιών, vrn, ανώνυμων επικοινωνιών, κρυπτονομισμάτων, πλατφορμών ανταλλαγής μηνυμάτων και παράνομων marketplaces επιτρέπει στους δράστες να διασπούν την εγκληματική δραστηριότητα σε πολλαπλά σημεία του ψηφιακού οικοσυστήματος. Ως

αποτέλεσμα, η διερεύνηση απαιτεί συντονισμό μεταξύ αρχών, παρόχων υπηρεσιών, χρηματοπιστωτικών φορέων και διεθνών οργανισμών [30], [137], [128].

Η απόδοση ευθύνης αποτελεί επίσης κρίσιμο πρόβλημα. Στο κυβερνοέγκλημα, η ανίχνευση του τεχνικού μέσου δεν ταυτίζεται απαραίτητα με την ταυτοποίηση του φυσικού δράστη. Μία διεύθυνση IP, ένας λογαριασμός, ένα domain ή ένα malware sample μπορεί να δείχνει ένα τμήμα της επιθετικής υποδομής, χωρίς να αποκαλύπτει με βεβαιότητα ποιος το χειρίστηκε, ποιος το χρηματοδότησε ή ποιος ωφελήθηκε από την επίθεση. Η χρήση AI μπορεί να δυσχεράνει ακόμη περισσότερο την attribution διαδικασία, καθώς επιτρέπει την παραγωγή περιεχομένου που μιμείται ανθρώπινο ύφος, οργανωτική επικοινωνία ή ταυτότητα τρίτων προσώπων [30], [43], [139].

Η τεχνητή νοημοσύνη ενισχύει ιδιαίτερα τις τεχνικές πλαστοπροσωπίας. deepfakes, συνθετικές φωνές, ψεύτικα βίντεο, AI-generated εικόνες και παραποιημένες συνομιλίες μπορούν να χρησιμοποιηθούν για εξαπάτηση θυμάτων, χειραγώγηση εργαζομένων, οικονομική απάτη ή παραπλάνηση ερευνητικών αρχών. Η ύπαρξη τέτοιου περιεχομένου δημιουργεί νέο επίπεδο αμφιβολίας, οι αρχές δεν αρκεί πλέον να αποδείξουν ότι ένα ψηφιακό αρχείο υπάρχει, αλλά πρέπει να αξιολογήσουν εάν είναι αυθεντικό, εάν έχει τροποποιηθεί και εάν μπορεί να χρησιμοποιηθεί αξιόπιστα ως αποδεικτικό στοιχείο [142], [143], [144].

Η αξιοπιστία των ψηφιακών αποδεικτικών στοιχείων αποτελεί κεντρικό ζήτημα στο περιβάλλον της generative AI. Φωτογραφίες, ηχητικά αρχεία, βίντεο και γραπτά μηνύματα μπορούν να παραχθούν ή να τροποποιηθούν με τρόπους που δεν είναι άμεσα ορατοί στον ανθρώπινο παρατηρητή. Αυτό δημιουργεί προκλήσεις για την ψηφιακή εγκληματολογία, καθώς απαιτείται τεχνική επαλήθευση της προέλευσης, της ακεραιότητας και της αυθεντικότητας του περιεχομένου. Οι αρχές καλούνται επομένως να αναπτύξουν διαδικασίες και εργαλεία που μπορούν να διακρίνουν γνήσιο από συνθετικό ή παραποιημένο περιεχόμενο [142], [143], [146].

Η ανάγκη αξιολόγησης deepfake και synthetic media συνδέεται άμεσα με την αποδεικτική αξιοπιστία. Σε μία ποινική διαδικασία, ένα ψηφιακό στοιχείο πρέπει να μπορεί να υποστηριχθεί τεχνικά και νομικά. Εάν η αυθεντικότητα ενός αρχείου αμφισβητείται, η αξία του ως αποδεικτικού μέσου μπορεί να μειωθεί. Για τον λόγο αυτό, η ανάπτυξη μηχανισμών provenance, content

credentials, deepfake detection και forensic verification καθίσταται κρίσιμη για τη διατήρηση της εμπιστοσύνης στα ψηφιακά στοιχεία [142], [144], [146].

Η ταχύτητα και η κλίμακα των AI-assisted επιθέσεων δημιουργούν ακόμα ένα τοίχο μεταξύ εγκληματικών ομάδων και αρχών επιβολής του νόμου. Οι δράστες μπορούν να παράγουν μαζικά phishing μηνύματα, να αυτοματοποιούν απάτες, να χρησιμοποιούν bots και να προσαρμόζουν περιεχόμενο σε διαφορετικά θύματα. Αντίθετα, οι αρχές λειτουργούν μέσα σε θεσμικά πλαίσια που απαιτούν νομιμότητα, τεκμηρίωση, σεβασμό δικαιωμάτων και διαδικαστικές εγγυήσεις. Η ταχύτητα της επίθεσης και η βραδύτητα της θεσμικής διερεύνησης δημιουργούν πρακτικό χάσμα ανταπόκρισης [43], [137], [139].

Το χάσμα αυτό είναι ιδιαίτερα έντονο σε απάτες μεγάλης κλίμακας, scam centres και οργανωμένα online fraud δίκτυα. Οι εγκληματικές ομάδες μπορούν να χρησιμοποιούν αυτοματοποίηση, social engineering, AI-generated περιεχόμενο και πολλαπλά κανάλια επικοινωνίας για να προσεγγίζουν μεγάλο αριθμό θυμάτων. Η διερεύνηση τέτοιων περιπτώσεων απαιτεί διασυνοριακή ανάλυση, ιχνηλάτηση, συνεργασία με πλατφόρμες και πρόσβαση σε δεδομένα που μπορεί να βρίσκονται σε διαφορετικές δικαιοδοσίες [137], [139], [128].

Ένα ακόμη κρίσιμο ζήτημα είναι το τεχνολογικό χάσμα μεταξύ κυβερνοεγκληματικών ομάδων και αρχών. Οι εγκληματικές ομάδες μπορούν να υιοθετούν γρήγορα νέα εργαλεία, να πειραματίζονται χωρίς θεσμικούς περιορισμούς και να αξιοποιούν underground υπηρεσίες ή εμπορικά διαθέσιμες τεχνολογίες. Αντίθετα, πολλές αρχές, ιδιαίτερα σε μικρότερα κράτη ή οργανισμούς με περιορισμένους πόρους, ενδέχεται να μην διαθέτουν επαρκές εξειδικευμένο προσωπικό, σύγχρονα εργαλεία ψηφιακής εγκληματολογίας, πρόσβαση σε threat intelligence ή κατάλληλη τεχνική εκπαίδευση [138], [139], [145].

Η INTERPOL επισημαίνει ότι σε ορισμένες περιοχές εξακολουθούν να υπάρχουν ελλείψεις σε ψηφιακή εγκληματολογία, τεχνικούς πόρους, εξειδικευμένη κατάρτιση και πρόσβαση σε ψηφιακά αποδεικτικά στοιχεία [138]. Το στοιχείο αυτό δείχνει ότι η αντιμετώπιση του AI κυβερνοεγκλήματος δεν είναι μόνο τεχνικό ζήτημα, αλλά και ζήτημα θεσμικής ικανότητας. Ακόμη και όταν υπάρχουν κατάλληλοι νόμοι, η αποτελεσματική επιβολή τους προϋποθέτει ανθρώπινους πόρους, υποδομές, διεθνή συνεργασία και συνεχή εκπαίδευση.

Η οργανωτική ετοιμότητα των αρχών αποτελεί επομένως κρίσιμο παράγοντα. Η αντιμετώπιση AI-driven cybercrime και deepfake fraud απαιτεί εξειδικευμένες ομάδες, τεχνική κατάρτιση, σαφείς διαδικασίες, πρόσβαση σε εργαλεία ανάλυσης και δυνατότητα συνεργασίας με ιδιωτικούς φορείς. Χωρίς αυτά τα στοιχεία, οι αρχές κινδυνεύουν να αντιδρούν εκ των υστέρων, όταν η ζημία έχει ήδη προκληθεί και τα ψηφιακά ίχνη έχουν κατακερματιστεί ή εξαφανιστεί [145], [138], [141].

Η πρόσβαση σε δεδομένα αποτελεί ιδιαίτερη πρόκληση. Πολλά στοιχεία που είναι απαραίτητα για μία έρευνα κρατούνται από ιδιωτικές εταιρείες, πλατφόρμες επικοινωνίας, παρόχους cloud υπηρεσιών, παρόχους τηλεπικοινωνιών ή χρηματοπιστωτικούς οργανισμούς. Οι αρχές χρειάζονται νόμιμη και αποτελεσματική πρόσβαση σε τέτοια δεδομένα, χωρίς όμως να παραβιάζονται τα δικαιώματα προστασίας προσωπικών δεδομένων. Η ισορροπία ανάμεσα στην αποτελεσματική διερεύνηση και στην προστασία δικαιωμάτων αποτελεί βασικό θεσμικό ζήτημα [140], [147], [148].

Η συλλογή ηλεκτρονικών αποδεικτικών στοιχείων απαιτεί επίσης ταχύτητα. Σε ψηφιακά περιβάλλοντα, τα δεδομένα μπορεί να διαγραφούν, να μετακινηθούν, να κρυπτογραφηθούν ή να αποθηκεύονται προσωρινά. Εάν οι αρχές δεν αποκτήσουν έγκαιρη πρόσβαση, η έρευνα μπορεί να καθυστερήσει σημαντικά. Για τον λόγο αυτό, οι διεθνείς πρωτοβουλίες για κοινά πλαίσια συνεργασίας, ανταλλαγής δεδομένων και ηλεκτρονικών αποδεικτικών στοιχείων έχουν ιδιαίτερη σημασία για την αντιμετώπιση κυβερνοεγκλημάτων σε διασυνοριακό περιβάλλον [147], [148].

Η ανάγκη διεθνούς συνεργασίας είναι κεντρική. Καμία αρχή και κανένα κράτος δεν μπορεί να αντιμετωπίσει μόνο του ένα οικοσύστημα κυβερνοεγκλήματος που λειτουργεί μέσω διεθνών δικτύων, πλατφορμών, αγορών και χρηματοοικονομικών ροών. Η συνεργασία μεταξύ κρατών, διεθνών οργανισμών, αστυνομικών αρχών, εισαγγελικών αρχών και ιδιωτικού τομέα είναι αναγκαία για τον εντοπισμό δραστών, την ανταλλαγή πληροφοριών, τη διακοπή εγκληματικών υποδομών και τη συλλογή αποδεικτικών στοιχείων [30], [135], [128].

Η συνεργασία δημόσιου και ιδιωτικού τομέα έχει επίσης καθοριστική σημασία. Πολλά κρίσιμα δεδομένα, τεχνικές πληροφορίες και δείκτες απειλής βρίσκονται στα χέρια εταιρειών κυβερνοασφάλειας, παρόχων cloud, τραπεζών, τηλεπικοινωνιακών παρόχων, πλατφορμών κοινωνικής δικτύωσης και υπηρεσιών ανταλλαγής μηνυμάτων. Οι αρχές χρειάζονται μηχανισμούς συνεργασίας με αυτούς τους φορείς, ώστε να μπορούν να εντοπίζουν υποδομές, να διακόπτουν

απάτες, να αναλύουν οικονομικές ροές και να υποστηρίζουν έρευνες μεγάλης κλίμακας [135], [141], [128].

Παράλληλα, η τεχνητή νοημοσύνη δεν αποτελεί μόνο πρόβλημα για τις αρχές, αλλά και πιθανό εργαλείο υποστήριξης των ερευνών. Οι αρχές μπορούν να αξιοποιήσουν ΑΙ για ανάλυση μεγάλων συνόλων δεδομένων, OSINT, NLP, ψηφιακή εγκληματολογία, αναγνώριση μοτίβων, ανάλυση οικονομικών συναλλαγών, computer vision, βιομετρικές εφαρμογές και εντοπισμό συνδέσεων μεταξύ διαφορετικών περιστατικών. Η Europol Innovation Lab αναφέρει ότι η ΑΙ μπορεί να ενισχύσει την ικανότητα των αρχών να επεξεργάζονται μεγάλες ποσότητες πληροφοριών και να εξάγουν επιχειρησιακά χρήσιμα συμπεράσματα [141].

Η χρήση ΑΙ από τις αρχές μπορεί να συμβάλλει στη μείωση του ανθρώπινου φόρτου. Οι ψηφιακές έρευνες συχνά περιλαμβάνουν τεράστιους όγκους αρχείων, logs, εικόνων, επικοινωνιών, οικονομικών συναλλαγών και δεδομένων από πολλές πλατφόρμες. Η χειροκίνητη ανάλυση όλων αυτών των στοιχείων είναι συχνά αδύνατη ή εξαιρετικά χρονοβόρα. Η ΑΙ μπορεί να χρησιμοποιηθεί για φιλτράρισμα, κατηγοριοποίηση, εντοπισμό μοτίβων και προτεραιοποίηση ευρημάτων, επιτρέποντας στους ερευνητές να επικεντρώνονται στα πιο κρίσιμα στοιχεία [141], [140].

Η ΑΙ μπορεί επίσης να υποστηρίξει την ανίχνευση deepfakes και συνθετικού περιεχομένου. Καθώς η παραγωγή ψεύτικων εικόνων, βίντεο και φωνής γίνεται ευκολότερη, οι αρχές χρειάζονται εργαλεία που μπορούν να αξιολογούν την αυθεντικότητα ψηφιακών μέσων. Η τεχνολογία deepfake detection μπορεί να συμβάλει σε fraud prevention, identity verification, ερευνητική υποστήριξη και προστασία της αποδεικτικής αξιοπιστίας [142], [143], [144].

Ωστόσο, η χρήση ΑΙ από τις αρχές πρέπει να συνοδεύεται από αυστηρούς κανόνες διαφάνειας, λογοδοσίας και προστασίας θεμελιωδών δικαιωμάτων. Η εφαρμογή ΑΙ στην αστυνόμευση μπορεί να δημιουργήσει κινδύνους μεροληψίας, υπερβολικής επιτήρησης, λανθασμένης ταυτοποίησης ή αδιαφανούς λήψης αποφάσεων. Επομένως, η αξιοποίηση τέτοιων εργαλείων δεν πρέπει να γίνεται χωρίς ανθρώπινη εποπτεία, τεχνική αξιολόγηση, και νομικό έλεγχο [141], [145], [148].

Ιδιαίτερη σημασία έχει η αποδεικτική χρήση αποτελεσμάτων που παράγονται ή υποστηρίζονται από ΑΙ. Εάν ένα εργαλείο ΑΙ χρησιμοποιείται για ανάλυση δεδομένων, αναγνώριση προσώπων,

εντοπισμό deepfake ή συσχέτιση περιστατικών, πρέπει να είναι δυνατό να εξηγηθεί πώς προέκυψε το αποτέλεσμα, ποια δεδομένα χρησιμοποιήθηκαν και ποια είναι τα όρια αξιοπιστίας του. Διαφορετικά, υπάρχει κίνδυνος τα αποτελέσματα να αμφισβητηθούν σε δικαστικό επίπεδο ή να υπονομεύσουν την αξιοπιστία της έρευνας [141], [142], [144].

Η πρόκληση για τις αρχές είναι επομένως διπλή. Από τη μία πλευρά, πρέπει να αντιμετωπίσουν εγκληματικές ομάδες που αξιοποιούν ΑΙ για ταχύτητα, κλίμακα, απόκρυψη και πλαστοπροσωπία. Από την άλλη πλευρά, πρέπει να ενσωματώσουν οι ίδιες ΑΙ εργαλεία χωρίς να παραβιάσουν δικαιώματα και χωρίς να δημιουργήσουν παράνομους μηχανισμούς επιτήρησης [30], [141], [145].

Η χρήση τεχνητής νοημοσύνης στο κυβερνοέγκλημα δημιουργεί ένα νέο επίπεδο προκλήσεων για τις αρχές επιβολής του νόμου. Η διεθνής φύση των επιθέσεων, η ανωνυμία, η δυσκολία attribution, τα deepfakes, η ανάγκη συλλογής ηλεκτρονικών αποδεικτικών στοιχείων, το τεχνολογικό χάσμα και η ταχύτητα των επιθέσεων καθιστούν την παραδοσιακή αντίδραση ανεπαρκή από μόνη της [135], [138], [139]. Η αποτελεσματική αντιμετώπιση απαιτεί τεχνική αναβάθμιση, θεσμική προσαρμογή, διεθνή συνεργασία και υπεύθυνη αξιοποίηση της ΑΙ από τις ίδιες τις αρχές [141], [147], [128].

7. Ο ΡΟΛΟΣ ΤΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΤΗΝ ΚΥΒΕΡΝΟΑΜΥΝΑ

7.1. Χρήση αλγορίθμων Τεχνητής Νοημοσύνης για την ανίχνευση επιθέσεων

Η ανάλυση των προηγούμενων κεφαλαίων ανέδειξε τον τρόπο με τον οποίο η τεχνητή νοημοσύνη μπορεί να ενισχύσει τη δυναμική των επιτιθέμενων, μειώνοντας το κόστος, αυξάνοντας την ταχύτητα και διευρύνοντας την κλίμακα των κυβερνοεπιθέσεων. Ωστόσο, η ίδια τεχνολογική εξέλιξη δεν περιορίζεται στην επιθετική χρήση. Στο πεδίο της κυβερνοάμυνας, οι αλγόριθμοι τεχνητής νοημοσύνης αξιοποιούνται για την έγκαιρη ανίχνευση επιθέσεων, την ανάλυση μεγάλου όγκου δεδομένων ασφαλείας, τη συσχέτιση συμβάντων και την αναγνώριση ύποπτων μοτίβων που ενδέχεται να υποδηλώνουν κακόβουλη δραστηριότητα [98], [99].

Η ανάγκη χρήσης τεχνητής νοημοσύνης στην ανίχνευση κυβερνοεπιθέσεων προκύπτει από τη σημαντική αύξηση του όγκου, της ταχύτητας και της πολυπλοκότητας των δεδομένων που παράγονται στα σύγχρονα πληροφοριακά περιβάλλοντα. Δικτυακή κίνηση, αρχεία καταγραφής, συμπεριφορές χρηστών, δραστηριότητες endpoints, ειδοποιήσεις ασφαλείας και δεδομένα από cloud υποδομές δημιουργούν ένα ιδιαίτερα σύνθετο περιβάλλον παρακολούθησης. Τα παραδοσιακά συστήματα ανίχνευσης, τα οποία βασίζονται κυρίως σε προκαθορισμένους κανόνες ή γνωστές υπογραφές, δυσκολεύονται να ανταποκριθούν σε άγνωστες, ταχέως μεταβαλλόμενες ή προσαρμοστικές απειλές [98], [54], [32].

Η βασική συμβολή της τεχνητής νοημοσύνης στην κυβερνοάμυνα εντοπίζεται στην ικανότητά της να αναγνωρίζει μοτίβα και αποκλίσεις μέσα σε μεγάλα και ετερογενή σύνολα δεδομένων. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν ώστε να διακρίνουν μεταξύ φυσιολογικής και ύποπτης δραστηριότητας, να ταξινομούν συμβάντα ασφαλείας και να υποστηρίζουν την ανάλυση πιθανών επιθέσεων σε πραγματικό ή σχεδόν πραγματικό χρόνο [98], [9]. Με τον τρόπο αυτό, η ΑΙ δεν λειτουργεί ως απλή επέκταση ενός παραδοσιακού antivirus, αλλά ως μηχανισμός ανάλυσης, συσχέτισης και προτεραιοποίησης απειλών.

Μία από τις βασικότερες εφαρμογές της Τεχνητής Νοημοσύνης στην ανίχνευση επιθέσεων είναι η ανίχνευση ανωμαλιών. Η ανίχνευση ανωμαλιών βασίζεται στην αναγνώριση αποκλίσεων από ένα αναμενόμενο πρότυπο συμπεριφοράς, είτε αυτό αφορά χρήστες, συστήματα, εφαρμογές, endpoints ή δικτυακή κίνηση. Σε αντίθεση με τις μεθόδους που αναζητούν αποκλειστικά γνωστά

μοτίβα κακόβουλης δραστηριότητας, οι αλγόριθμοι anomaly detection μπορούν να συμβάλλουν στον εντοπισμό άγνωστων ή μη προηγουμένως καταγεγραμμένων απειλών [98], [99], [101].

Η επιβλεπόμενη μάθηση χρησιμοποιείται κυρίως σε περιπτώσεις όπου υπάρχουν διαθέσιμα σύνολα δεδομένων με γνωστές ετικέτες, δηλαδή παραδείγματα νόμιμης και κακόβουλης δραστηριότητας. Μέσω αυτής της προσέγγισης, τα μοντέλα μπορούν να εκπαιδευτούν ώστε να αναγνωρίζουν συγκεκριμένες κατηγορίες επιθέσεων, όπως malware, phishing ή γνωστά μοτίβα κακόβουλης διαδικτυακής συμπεριφοράς. Η αποτελεσματικότητα της επιβλεπόμενης μάθησης εξαρτάται σε μεγάλο βαθμό από την ποιότητα, την αντιπροσωπευτικότητα και την επικαιρότητα των δεδομένων εκπαίδευσης [98], [54], [47].

Αντίθετα, η μη επιβλεπόμενη μάθηση είναι ιδιαίτερα χρήσιμη όταν δεν υπάρχουν επαρκώς επισημασμένα δεδομένα ή όταν ο στόχος είναι η αναγνώριση άγνωστων προτύπων. Σε τέτοιες περιπτώσεις, το μοντέλο δεν εκπαιδεύεται με προκαθορισμένες ετικέτες, αλλά επιχειρεί να εντοπίσει δομές, ομάδες και αποκλίσεις μέσα στα δεδομένα. Η προσέγγιση αυτή είναι σημαντική για την ανίχνευση αγνώστων απειλών, εσωτερικών ανωμαλιών, ασυνήθιστης συμπεριφοράς χρηστών και πιθανών ενδείξεων αρχικής παραβίασης [98], [99].

Η βαθιά μάθηση μπορεί να αξιοποιηθεί σε περιβάλλοντα όπου τα δεδομένα είναι ιδιαίτερα σύνθετα ή μεγάλης κλίμακας. Νευρωνικά δίκτυα και άλλες deep learning αρχιτεκτονικές μπορούν να υποστηρίξουν την ανάλυση ακολουθιών γεγονότων, τη συσχέτιση πολλών παραμέτρων και την επεξεργασία δεδομένων από διαφορετικές πηγές. Η προσέγγιση αυτή είναι ιδιαίτερα χρήσιμη σε περιπτώσεις όπου η κακόβουλη δραστηριότητα δεν εμφανίζεται ως μεμονωμένο γεγονός, αλλά ως αλληλουχία ενεργειών που αποκτούν σημασία μόνο όταν αναλυθούν συνδυαστικά [98], [54], [47].

Σημαντικό πεδίο εφαρμογής αποτελεί η ανίχνευση κακόβουλου λογισμικού. Τα AI-based συστήματα μπορούν να αξιοποιούν στατικά και δυναμικά χαρακτηριστικά αρχείων, συμπεριφορικά δεδομένα, μοτίβα εκτέλεσης και ενδείξεις ύποπτης δραστηριότητας για την αναγνώριση πιθανών malware απειλών. Σε αντίθεση με τις αποκλειστικά signature-based μεθόδους, οι οποίες βασίζονται σε ήδη γνωστά χαρακτηριστικά κακόβουλου λογισμικού, τα μοντέλα μηχανικής μάθησης μπορούν να συμβάλουν στην ανίχνευση παραλλαγών malware ή συμπεριφορών που δεν αντιστοιχούν πλήρως σε γνωστές υπογραφές [54], [47], [101].

Αντίστοιχα, τεχνητή νοημοσύνη αξιοποιείται στην ανίχνευση phishing επιθέσεων. Η ανάλυση γλώσσας, δεδομένων, συνδέσμων, domain patterns και χαρακτηριστικών επικοινωνίας μπορεί να υποστηρίξει τον εντοπισμό παραπλανητικών μηνυμάτων. Ιδιαίτερη σημασία αποκτά η χρήση μεγάλων γλωσσικών μοντέλων και τεχνικών επεξεργασίας φυσικής γλώσσας, καθώς οι σύγχρονες phishing επιθέσεις γίνονται περισσότερο γλωσσικά άρτιες και προσωποποιημένες. Η τεχνητή νοημοσύνη μπορεί να συμβάλει στην αξιολόγηση του περιεχομένου, της δομής και του επικοινωνιακού πλαισίου ενός μηνύματος, ενισχύοντας την ικανότητα εντοπισμού ύποπτων στοιχείων [54], [47].

Η τεχνητή νοημοσύνη αξιοποιείται επίσης στην Cyber Threat Intelligence. Η συλλογή, επεξεργασία και αξιολόγηση πληροφοριών απειλών αποτελεί κρίσιμη διαδικασία για την πρόληψη και ανίχνευση κυβερνοεπιθέσεων. Οι αλγόριθμοι AI και οι τεχνικές Natural Language Processing μπορούν να υποστηρίξουν την εξαγωγή δεικτών παραβίασης, την ανάλυση τεχνικών αναφορών, τη συσχέτιση πληροφοριών από διαφορετικές πηγές και την αναγνώριση αναδυόμενων απειλών.

Η σημασία της AI-assisted threat intelligence έγκειται στη δυνατότητα μετατροπής μεγάλου όγκου αδόμητων πληροφοριών σε αξιοποιήσιμη γνώση για την άμυνα. Αναφορές ευπαθειών, τεχνικά blogs, dark web πληροφορίες, indicators of compromise και δεδομένα από περιστατικά ασφαλείας μπορούν να αναλυθούν ταχύτερα μέσω αυτοματοποιημένων μεθόδων. Με τον τρόπο αυτό, η τεχνητή νοημοσύνη μπορεί να υποστηρίξει την έγκαιρη προειδοποίηση και την προτεραιοποίηση απειλών, επιτρέποντας στους οργανισμούς να ανταποκρίνονται πιο αποτελεσματικά [58], [57].

Ιδιαίτερα σημαντική είναι η ενσωμάτωση της τεχνητής νοημοσύνης σε περιβάλλοντα SIEM, SOAR και SOC. Τα συστήματα SIEM συγκεντρώνουν και συσχετίζουν μεγάλο αριθμό συμβάντων ασφαλείας, ενώ τα SOAR περιβάλλοντα υποστηρίζουν την αυτοματοποίηση διαδικασιών απόκρισης. Η τεχνητή νοημοσύνη μπορεί να ενισχύσει αυτά τα συστήματα μέσω αυτοματοποιημένης ταξινόμησης ειδοποιήσεων, alert triage, συσχέτισης περιστατικών, μείωσης θορύβου και υποστήριξης των αναλυτών ασφαλείας στη λήψη αποφάσεων [32], [102].

Η συμβολή της AI σε SOC environments είναι ιδιαίτερα σημαντική, διότι οι ομάδες ασφαλείας συχνά αντιμετωπίζουν μεγάλο όγκο ειδοποιήσεων, από τις οποίες μόνο ένα μέρος αντιστοιχεί σε πραγματικά κρίσιμα περιστατικά. Η χρήση AI μπορεί να υποστηρίξει την ιεράρχηση των

συμβάντων με βάση τη σοβαρότητα, τη συσχέτιση με προηγούμενη δραστηριότητα και την πιθανότητα κακόβουλης πρόθεσης. Έτσι, μειώνεται ο χρόνος που απαιτείται για την αρχική αξιολόγηση ενός περιστατικού και βελτιώνεται η επιχειρησιακή αποδοτικότητα των αμυντικών μηχανισμών [32], [102].

Η τεχνητή νοημοσύνη μπορεί επίσης να χρησιμοποιηθεί σε endpoint detection και behavioral analytics. Μέσω της ανάλυσης τηλεμετρίας από endpoints, οι αλγόριθμοι μπορούν να εντοπίζουν ύποπτες ενέργειες, ασυνήθιστη χρήση εφαρμογών, μη τυπικές ακολουθίες ενεργειών ή πιθανές ενδείξεις lateral movement. Η προσέγγιση αυτή είναι σημαντική, διότι πολλές σύγχρονες επιθέσεις δεν γίνονται άμεσα αντιληπτές μέσω μεμονωμένων γεγονότων, αλλά μέσω συνδυασμού συμπεριφορικών αποκλίσεων [24], [101].

Σε cloud περιβάλλοντα, η τεχνητή νοημοσύνη μπορεί να ενισχύσει την παρακολούθηση και ανίχνευση επιθέσεων μέσω behavioral analytics, ανάλυσης πρόσβασης, εντοπισμού ύποπτης χρήσης credentials και συσχέτισης γεγονότων σε διαφορετικές υπηρεσίες. Καθώς οι οργανισμοί μεταφέρουν κρίσιμες λειτουργίες σε cloud υποδομές, η δυνατότητα ανίχνευσης ασυνήθιστων μοτίβων πρόσβασης και δραστηριότητας καθίσταται κεντρική για την κυβερνοάμυνα [100], [24].

Ωστόσο, η αξιοποίηση της τεχνητής νοημοσύνης στην ανίχνευση επιθέσεων δεν πρέπει να αντιμετωπίζεται ως πανάκεια. Ένας από τους βασικούς περιορισμούς αφορά τα ψευδώς θετικά αποτελέσματα. Ένα AI-based σύστημα μπορεί να χαρακτηρίσει νόμιμη δραστηριότητα ως ύποπτη, επιβαρύνοντας τους αναλυτές ασφαλείας και μειώνοντας την επιχειρησιακή αποτελεσματικότητα. Αντίστοιχα, ψευδώς αρνητικά αποτελέσματα μπορούν να επιτρέψουν σε κακόβουλη δραστηριότητα να παραμείνει μη ανιχνευμένη [50], [8].

Ένας δεύτερος περιορισμός αφορά την ποιότητα των δεδομένων εκπαίδευσης. Τα μοντέλα μηχανικής μάθησης εξαρτώνται σε μεγάλο βαθμό από τα δεδομένα με τα οποία εκπαιδεύονται. Εάν τα δεδομένα είναι ελλιπή, μεροληπτικά, παρωχημένα ή μη αντιπροσωπευτικά του πραγματικού περιβάλλοντος απειλών, τότε το μοντέλο μπορεί να παρουσιάζει περιορισμένη ικανότητα γενίκευσης. Αυτό είναι ιδιαίτερα κρίσιμο στην κυβερνοασφάλεια, όπου οι επιθέσεις εξελίσσονται διαρκώς και οι επιτιθέμενοι τροποποιούν τις τεχνικές τους [98], [50], [8].

Η ερμηνευσιμότητα αποτελεί επίσης σημαντικό ζήτημα. Ορισμένα προηγμένα μοντέλα, ιδίως στο πεδίο της βαθιάς μάθησης, μπορεί να παράγουν αποτελέσματα χωρίς να είναι πάντα σαφές ποιοι παράγοντες οδήγησαν στην εκάστοτε απόφαση. Σε περιβάλλοντα κυβερνοάμυνας, όπου οι αποφάσεις πρέπει να είναι ελέγξιμες και επιχειρησιακά αιτιολογημένες, η δυσκολία ερμηνείας μπορεί να περιορίσει την εμπιστοσύνη των αναλυτών στα AI-based συστήματα [50], [9].

Ιδιαίτερη προσοχή απαιτείται και στο ζήτημα των adversarial attacks εναντίον των ίδιων των AI μοντέλων. Οι επιτιθέμενοι μπορούν να επιχειρήσουν να παραπλανήσουν ένα μοντέλο μέσω τεχνικών evasion, poisoning ή άλλων adversarial μεθόδων, με στόχο να μειώσουν την αποτελεσματικότητα της ανίχνευσης. Το NIST επισημαίνει ότι η ασφάλεια των μοντέλων μηχανικής μάθησης αποτελεί κρίσιμο ζήτημα, καθώς τα ίδια τα AI συστήματα μπορούν να αποτελέσουν στόχο επιθετικών ενεργειών [8].

Η AI μπορεί να ενισχύσει σημαντικά την ανίχνευση κυβερνοεπιθέσεων μέσω της ανάλυσης μεγάλων όγκων δεδομένων, της αναγνώρισης ανωμαλιών και της ταχύτερης αξιολόγησης ύποπτων δραστηριοτήτων. Ωστόσο, η αποτελεσματικότητά της εξαρτάται από την ποιότητα των δεδομένων, τη σωστή διαμόρφωση των μοντέλων και την ενσωμάτωσή της σε ώριμες διαδικασίες κυβερνοάμυνας. Η μεγαλύτερη αξία της προκύπτει όταν λειτουργεί συμπληρωματικά προς την ανθρώπινη κρίση, την πληροφορία απειλών και τους υφιστάμενους μηχανισμούς απόκρισης.

7.2. Intrusion Detection Systems (IDS)

Τα Intrusion Detection Systems (IDS) αποτελούν έναν από τους βασικότερους μηχανισμούς ανίχνευσης κυβερνοεπιθέσεων στο σύγχρονο περιβάλλον κυβερνοάμυνας. Ο κύριος ρόλος τους είναι η παρακολούθηση δραστηριοτήτων σε δίκτυα, συστήματα και τελικές συσκευές, με σκοπό τον εντοπισμό ενδείξεων μη εξουσιοδοτημένης πρόσβασης, ύποπτης συμπεριφοράς ή κακόβουλης δραστηριότητας. Σε αντίθεση με τα συστήματα που επικεντρώνονται αποκλειστικά στην αποτροπή ή στην απόκριση, τα IDS λειτουργούν κυρίως ως μηχανισμοί έγκαιρης προειδοποίησης, καθώς παράγουν ειδοποιήσεις ασφαλείας οι οποίες υποστηρίζουν την περαιτέρω διερεύνηση και αντιμετώπιση περιστατικών [103], [32], [107].

Η σημασία των IDS έχει αυξηθεί ιδιαίτερα λόγω της πολυπλοκότητας των σύγχρονων κυβερνοεπιθέσεων. Οι επιθέσεις δεν εμφανίζονται πάντοτε ως άμεσα αναγνωρίσιμα περιστατικά, αλλά συχνά εκδηλώνονται μέσα από μικρές αποκλίσεις στη συμπεριφορά χρηστών, συστημάτων ή δικτυακής κίνησης. Για τον λόγο αυτό, τα IDS αποτελούν κρίσιμο επίπεδο παρακολούθησης, καθώς επιτρέπουν στους οργανισμούς να εντοπίζουν ύποπτες ενέργειες πριν αυτές εξελιχθούν σε πλήρη παραβίαση ή προκαλέσουν εκτεταμένη ζημιά [32], [31].

Τα IDS διακρίνονται συνήθως σε βασικές κατηγορίες ανάλογα με το σημείο παρακολούθησης και το είδος των δεδομένων που αναλύουν. Τα Network-based Intrusion Detection Systems (NIDS) παρακολουθούν τη δικτυακή κίνηση και αναλύουν πακέτα δεδομένων, επικοινωνίες, πρωτόκολλα και συνδέσεις, με στόχο τον εντοπισμό ύποπτων μοτίβων στο επίπεδο του δικτύου. Αντίθετα, τα Host-based Intrusion Detection Systems (HIDS) εγκαθίστανται σε συγκεκριμένα συστήματα ή endpoints και παρακολουθούν στοιχεία όπως αρχεία καταγραφής, αλλαγές σε κρίσιμα αρχεία, διεργασίες, ενέργειες χρηστών και συμπεριφορά του λειτουργικού συστήματος [103], [32], [107].

Πέρα από τις δύο αυτές βασικές κατηγορίες, χρησιμοποιούνται και υβριδικές προσεγγίσεις, οι οποίες συνδυάζουν δεδομένα από δίκτυα, hosts, endpoints και cloud περιβάλλοντα. Τα hybrid IDS επιτρέπουν πιο ολοκληρωμένη εικόνα της δραστηριότητας ενός οργανισμού, καθώς συνδυάζουν τη δικτυακή ορατότητα των NIDS με τη λεπτομερέστερη παρακολούθηση που παρέχουν τα HIDS σε επίπεδο συστήματος. Η υβριδική αυτή προσέγγιση είναι ιδιαίτερα σημαντική σε σύγχρονα περιβάλλοντα όπου οι επιθέσεις μπορούν να ξεκινούν από το δίκτυο, να συνεχίζονται σε endpoints και να επεκτείνονται σε cloud υπηρεσίες [32], [104], [107].

Παραδοσιακά, τα IDS βασίζονταν σε signature-based και rule-based μεθόδους ανίχνευσης. Η signature-based ανίχνευση λειτουργεί με βάση γνωστά μοτίβα κακόβουλης δραστηριότητας, όπως συγκεκριμένες ακολουθίες πακέτων, χαρακτηριστικά αρχείων ή γνωστές τεχνικές επίθεσης. Το βασικό της πλεονέκτημα είναι η υψηλή ακρίβεια απέναντι σε γνωστές απειλές. Ωστόσο, η αποτελεσματικότητά της περιορίζεται σημαντικά όταν οι επιθέσεις είναι νέες, τροποποιημένες ή σχεδιασμένες ώστε να αποφεύγουν γνωστές υπογραφές [103], [32].

Αντίστοιχα, η rule-based ανίχνευση βασίζεται σε προκαθορισμένους κανόνες, οι οποίοι ενεργοποιούνται όταν εντοπιστεί συγκεκριμένη συμπεριφορά ή συγκεκριμένη ακολουθία γεγονότων. Οι κανόνες αυτοί μπορούν να είναι ιδιαίτερα χρήσιμοι για την αναγνώριση γνωστών

επιθετικών τεχνικών, αλλά παρουσιάζουν περιορισμούς απέναντι σε πιο δυναμικές, πιο πολύπλοκες ή άγνωστες επιθέσεις. Όσο οι επιτιθέμενοι τροποποιούν τη συμπεριφορά τους, χρησιμοποιούν polymorphic malware ή αξιοποιούν τεχνικές απόκρυψης, τόσο δυσκολότερο γίνεται για ένα καθαρά rule-based IDS να εντοπίσει αποτελεσματικά την απειλή [52], [103], [32].

Η ενσωμάτωση αλγορίθμων τεχνητής νοημοσύνης και μηχανικής μάθησης στα IDS αποσκοπεί ακριβώς στην αντιμετώπιση αυτών των περιορισμών. Τα AI-enhanced IDS δεν βασίζονται αποκλειστικά σε γνωστές υπογραφές ή στατικούς κανόνες, αλλά μπορούν να αναλύουν μεγάλα σύνολα δεδομένων, να αναγνωρίζουν μοτίβα, να εντοπίζουν ανωμαλίες και να προσαρμόζονται σε μεταβαλλόμενα περιβάλλοντα απειλών. Η δυνατότητα αυτή είναι ιδιαίτερα σημαντική σε ένα περιβάλλον όπου οι κυβερνοεπιθέσεις εξελίσσονται γρήγορα και συχνά δεν αντιστοιχούν σε προηγούμενως γνωστά μοτίβα [54], [47], [52].

Η επιβλεπόμενη μάθηση αποτελεί μία από τις βασικές προσεγγίσεις που μπορούν να αξιοποιηθούν στα IDS. Σε αυτή την περίπτωση, τα μοντέλα εκπαιδεύονται με δεδομένα τα οποία έχουν ήδη ταξινομηθεί ως νόμιμη ή κακόβουλη δραστηριότητα. Με βάση αυτά τα δεδομένα, το σύστημα μαθαίνει να διακρίνει ανάμεσα σε διαφορετικές κατηγορίες συμπεριφοράς και να ταξινομεί νέα περιστατικά. Η επιβλεπόμενη μάθηση είναι ιδιαίτερα χρήσιμη όταν υπάρχουν αξιόπιστα και αντιπροσωπευτικά datasets, ωστόσο η απόδοσή της εξαρτάται σε μεγάλο βαθμό από την ποιότητα και την επικαιρότητα των δεδομένων εκπαίδευσης [54], [47], [105].

Η μη επιβλεπόμενη μάθηση είναι εξίσου σημαντική, ιδίως για την ανίχνευση ανωμαλιών και αγνώστων απειλών. Σε αυτή την περίπτωση, το σύστημα δεν χρειάζεται απαραίτητα προκαθορισμένες ετικέτες, αλλά επιχειρεί να εντοπίσει αποκλίσεις από το φυσιολογικό πρότυπο συμπεριφοράς. Η προσέγγιση αυτή είναι ιδιαίτερα χρήσιμη στα IDS, διότι πολλές επιθέσεις, ειδικά zero-day ή προσαρμοστικές επιθέσεις, δεν διαθέτουν γνωστές υπογραφές κατά τη στιγμή της εκδήλωσής τους [54], [47], [31].

Η βαθιά μάθηση μπορεί επίσης να ενισχύσει σημαντικά τα IDS, ιδιαίτερα όταν απαιτείται ανάλυση πολύπλοκων και πολυδιάστατων δεδομένων. Τα deep learning μοντέλα μπορούν να επεξεργάζονται μεγάλους όγκους δικτυακής κίνησης, ακολουθίες γεγονότων, συμπεριφορικά μοτίβα και δεδομένα από endpoints, εντοπίζοντας συσχετίσεις που είναι δύσκολο να

αναγνωριστούν με παραδοσιακές μεθόδους. Με αυτόν τον τρόπο, τα IDS αποκτούν μεγαλύτερη ικανότητα αναγνώρισης σύνθετων επιθετικών μοτίβων [54], [47], [105].

Ένα από τα σημαντικότερα πλεονεκτήματα των AI-enhanced IDS είναι η δυνατότητα ανίχνευσης ανωμαλιών. Αντί να αναζητούν μόνο γνωστές επιθέσεις, τα συστήματα αυτά μπορούν να εντοπίζουν συμπεριφορές που αποκλίνουν από το κανονικό πρότυπο λειτουργίας. Για παράδειγμα, ασυνήθιστη δικτυακή κίνηση, μη τυπική πρόσβαση σε δεδομένα, απρόσμενες συνδέσεις ή ασυνήθιστη δραστηριότητα endpoint μπορούν να αξιολογηθούν ως πιθανές ενδείξεις εισβολής. Η προσέγγιση αυτή αυξάνει την πιθανότητα έγκαιρου εντοπισμού επιθέσεων που δεν έχουν ακόμη καταγραφεί σε βάσεις γνωστών απειλών [54], [31], [105].

Ιδιαίτερη σημασία έχει και η δυνατότητα εντοπισμού zero-day επιθέσεων. Οι επιθέσεις αυτές αξιοποιούν ευπάθειες που δεν είναι ακόμη γνωστές ή δεν έχουν διορθωθεί, γεγονός που καθιστά ανεπαρκείς τις παραδοσιακές signature-based μεθόδους. Τα AI-based IDS μπορούν να συμβάλουν στον εντοπισμό τέτοιων επιθέσεων μέσω συμπεριφορικής ανάλυσης και αναγνώρισης μη φυσιολογικών μοτίβων, ακόμη και όταν δεν υπάρχει διαθέσιμη γνωστή υπογραφή για την απειλή [54], [47], [31].

Η χρήση τεχνητής νοημοσύνης στα IDS ενισχύει επίσης την ικανότητα επεξεργασίας μεγάλου όγκου δεδομένων. Οι σύγχρονοι οργανισμοί παράγουν τεράστιες ποσότητες logs, alerts, network flows, endpoint telemetry και cloud security events. Η χειροκίνητη ανάλυση αυτών των δεδομένων είναι πρακτικά αδύνατη σε πραγματικό χρόνο. Τα AI-enhanced IDS μπορούν να φιλτράρουν, να συσχετίζουν και να αξιολογούν τέτοια δεδομένα, επιτρέποντας στους αναλυτές ασφαλείας να επικεντρώνονται σε περιστατικά υψηλότερης προτεραιότητας [32], [102], [104], [106].

Τα IDS δεν λειτουργούν απομονωμένα από τη συνολική αρχιτεκτονική κυβερνοάμυνας ενός οργανισμού. Στη σύγχρονη πρακτική ενσωματώνονται σε Security Operations Centers, SIEM πλατφόρμες, XDR οικοσυστήματα και cloud security environments. Η ενσωμάτωση αυτή επιτρέπει τη συσχέτιση δεδομένων από διαφορετικές πηγές, όπως δίκτυα, endpoints, cloud υπηρεσίες και συστήματα ταυτότητας. Έτσι, το IDS δεν παράγει απλώς μεμονωμένες ειδοποιήσεις, αλλά συμβάλλει σε μια ευρύτερη εικόνα κινδύνου και υποστηρίζει την έγκαιρη ανάλυση περιστατικών [32], [102], [104], [106].

Η σύνδεση IDS με XDR και SIEM περιβάλλοντα ενισχύει τη δυνατότητα εντοπισμού πολυσταδιακών επιθέσεων. Πολλές σύγχρονες κυβερνοεπιθέσεις δεν εκδηλώνονται ως ένα μεμονωμένο γεγονός, αλλά ως αλληλουχία ενεργειών, όπως αρχική πρόσβαση, credential misuse, lateral movement και εξαγωγή δεδομένων. Ένα AI-enhanced IDS μπορεί να συμβάλει στην αναγνώριση τέτοιων ακολουθιών, ιδίως όταν τα επιμέρους γεγονότα φαίνονται μεμονωμένα λιγότερο ύποπτα, αλλά αποκτούν σημασία όταν αναλυθούν συνδυαστικά [32], [31], [102].

Παρά τα πλεονεκτήματά τους, τα AI-based IDS αντιμετωπίζουν σημαντικούς περιορισμούς. Ένας από τους πιο κρίσιμους είναι τα ψευδώς θετικά αποτελέσματα. Ένα IDS μπορεί να χαρακτηρίσει νόμιμη δραστηριότητα ως ύποπτη, δημιουργώντας υπερβολικό αριθμό ειδοποιήσεων και αυξάνοντας τον φόρτο των αναλυτών ασφαλείας. Αντίστοιχα, τα ψευδώς αρνητικά αποτελέσματα είναι εξίσου επικίνδυνα, καθώς μπορεί να επιτρέψουν σε πραγματική κακόβουλη δραστηριότητα να παραμείνει μη ανιχνευμένη [54], [105], [106].

Ένας ακόμη περιορισμός αφορά την ανάγκη για ποιοτικά, αντιπροσωπευτικά και επικαιροποιημένα δεδομένα εκπαίδευσης. Εάν ένα μοντέλο εκπαιδευτεί σε παρωχημένα ή μη αντιπροσωπευτικά δεδομένα, ενδέχεται να αποτύχει να αναγνωρίσει νέες επιθετικές τεχνικές ή να εμφανίσει χαμηλή ικανότητα γενίκευσης. Το πρόβλημα αυτό είναι ιδιαίτερα σημαντικό στην κυβερνοασφάλεια, όπου το περιβάλλον απειλών μεταβάλλεται συνεχώς και οι επιτιθέμενοι προσαρμόζουν τις τεχνικές τους [54], [50], [9].

Η ερμηνευσιμότητα αποτελεί επίσης κρίσιμη πρόκληση. Ορισμένα μοντέλα, ιδιαίτερα τα βαθιά νευρωνικά δίκτυα, μπορούν να παράγουν αποτελέσματα υψηλής ακρίβειας, χωρίς όμως να είναι πάντοτε σαφές γιατί μια συγκεκριμένη δραστηριότητα χαρακτηρίστηκε ως κακόβουλη. Σε περιβάλλοντα SOC, η έλλειψη ερμηνευσιμότητας μπορεί να δυσχεράνει την αξιολόγηση των ειδοποιήσεων και να περιορίσει την εμπιστοσύνη των αναλυτών στα αποτελέσματα του συστήματος [50], [9], [105].

Ιδιαίτερη προσοχή απαιτείται και στο Adversarial Machine Learning. Τα IDS που βασίζονται σε ΑΙ μπορούν να αποτελέσουν στόχο επιθέσεων οι οποίες επιχειρούν να παραπλανήσουν το μοντέλο. Οι επιτιθέμενοι ενδέχεται να τροποποιήσουν χαρακτηριστικά της δραστηριότητάς τους ώστε να αποφύγουν την ανίχνευση ή να επιχειρήσουν να επηρεάσουν τα δεδομένα εκπαίδευσης ενός μοντέλου. Το NIST επισημαίνει ότι οι τεχνικές evasion, poisoning και άλλες μορφές

adversarial manipulation αποτελούν κρίσιμες απειλές για την αξιοπιστία των AI-based συστημάτων [8], [50].

Επομένως, η αξιοποίηση AI στα IDS δεν πρέπει να αντιμετωπίζεται ως πλήρης αντικατάσταση των παραδοσιακών μεθόδων ανίχνευσης ή της ανθρώπινης εποπτείας. Αντίθετα, η μεγαλύτερη αξία της εντοπίζεται στον συνδυασμό τεχνικών. Τα signature-based και rule-based συστήματα παραμένουν χρήσιμα για γνωστές απειλές, ενώ τα AI-based μοντέλα ενισχύουν την ανίχνευση ανωμαλιών, άγνωστων επιθέσεων και σύνθετων συμπεριφορικών μοτίβων. Η αποτελεσματική κυβερνοάμυνα απαιτεί συνδυασμό κανόνων, υπογραφών, AI αλγορίθμων, ανθρώπινης κρίσης και συνεχούς αξιολόγησης [54], [8], [103], [105].

Τα συστήματα ανίχνευσης εισβολών αποτελούν βασικό στοιχείο της σύγχρονης κυβερνοάμυνας, ενώ η ενσωμάτωση AI ενισχύει την ικανότητά τους να επεξεργάζονται σύνθετα δεδομένα και να αναγνωρίζουν γνωστές ή αναδυόμενες απειλές. Η αποτελεσματική λειτουργία τους εξαρτάται, ωστόσο, από αξιόπιστα δεδομένα, σωστή εκπαίδευση των μοντέλων, ερμηνευσιμότητα των αποτελεσμάτων και ανθρώπινη εποπτεία. Η AI δεν αντικαθιστά τα IDS, αλλά διευρύνει τις δυνατότητές τους σε δυναμικά περιβάλλοντα κυβερνοαπειλών.

7.3. Αυτοματοποιημένες στρατηγικές απόκρισης σε κυβερνοεπιθέσεις

Η ανίχνευση μιας κυβερνοεπίθεσης αποτελεί κρίσιμο στάδιο της κυβερνοάμυνας, ωστόσο δεν επαρκεί από μόνη της για την αποτελεσματική προστασία ενός οργανισμού. Μετά τον εντοπισμό μιας απειλής, απαιτείται ταχεία, συντονισμένη και τεκμηριωμένη απόκριση, ώστε να περιοριστεί η έκταση της παραβίασης, να μειωθούν οι πιθανές απώλειες και να αποκατασταθεί η επιχειρησιακή συνέχεια. Στο σύγχρονο περιβάλλον απειλών, όπου οι επιθέσεις εξελίσσονται με αυξημένη ταχύτητα και πολυπλοκότητα, η χρονική απόσταση ανάμεσα στην ανίχνευση και στην αντιμετώπιση ενός περιστατικού αποτελεί καθοριστικό παράγοντα για το τελικό μέγεθος της ζημίας [32], [108].

Οι αυτοματοποιημένες στρατηγικές απόκρισης αναφέρονται στη χρήση τεχνολογιών που επιτρέπουν την εκτέλεση ενεργειών ασφάλειας χωρίς άμεση ανθρώπινη παρέμβαση σε κάθε

επιμέρους βήμα ή με περιορισμένη συμμετοχή των αναλυτών ασφαλείας. Οι ενέργειες αυτές μπορεί να περιλαμβάνουν την απομόνωση ενός ύποπτου endpoint, τον αποκλεισμό κακόβουλων διευθύνσεων IP, την καραντίνα ύποπτων αρχείων ή μηνυμάτων ηλεκτρονικού ταχυδρομείου, την προσωρινή απενεργοποίηση λογαριασμών και την ενεργοποίηση διαδικασιών ειδοποίησης της ομάδας ασφαλείας. Η βασική λογική της αυτοματοποιημένης απόκρισης είναι ότι ο οργανισμός δεν αρκεί να γνωρίζει πως δέχεται επίθεση, αλλά πρέπει να μπορεί να αντιδρά άμεσα και με συνέπεια [102], [109], [110].

Η τεχνητή νοημοσύνη ενισχύει ουσιαστικά τη διαδικασία απόκρισης, επειδή μπορεί να αξιολογεί περιστατικά, να ιεραρχεί ειδοποιήσεις, να συσχετίζει δεδομένα από διαφορετικές πηγές και να προτείνει ή να ενεργοποιεί ενέργειες αντιμετώπισης. Με τον τρόπο αυτό, η ΑΙ δεν λειτουργεί μόνο ως μηχανισμός ανίχνευσης, αλλά και ως μηχανισμός υποστήριξης αποφάσεων στο πλαίσιο της κυβερνοάμυνας. Η λειτουργία αυτή είναι ιδιαίτερα σημαντική σε περιβάλλοντα Security Operations Center, όπου οι αναλυτές καλούνται να διαχειριστούν μεγάλο όγκο alerts και να αποφασίσουν γρήγορα ποια περιστατικά απαιτούν άμεση παρέμβαση [54], [47], [106].

Η ανάγκη για ταχεία απόκριση γίνεται ακόμη πιο εμφανής σε επιθέσεις όπως ransomware, credential theft, lateral movement και cloud account compromise. Σε τέτοιες περιπτώσεις, ακόμη και μικρή καθυστέρηση μπορεί να επιτρέψει στον επιτιθέμενο να επεκτείνει την πρόσβασή του, να κλιμακώσει τα δικαιώματά του, να μετακινηθεί σε άλλα συστήματα ή να εξαγάγει κρίσιμα δεδομένα. Η αυτοματοποίηση επιτρέπει τη λήψη αρχικών μέτρων περιορισμού πριν ολοκληρωθεί η πλήρης ανθρώπινη διερεύνηση, μειώνοντας τον κίνδυνο περαιτέρω εξάπλωσης της επίθεσης [32], [31].

Κεντρικό ρόλο στις αυτοματοποιημένες στρατηγικές απόκρισης διαδραματίζουν τα συστήματα Security Orchestration, Automation and Response. Τα SOAR συστήματα επιτρέπουν την ενοποίηση διαφορετικών εργαλείων ασφαλείας, τη συγκέντρωση πληροφοριών, την ενεργοποίηση προκαθορισμένων workflows και την αυτοματοποίηση επαναλαμβανόμενων διαδικασιών αντιμετώπισης περιστατικών. Μέσω αυτών, ένα περιστατικό που εντοπίζεται από σύστημα ανίχνευσης μπορεί να εμπλουτιστεί με πρόσθετα δεδομένα, να αξιολογηθεί ως προς τη σοβαρότητά του και να οδηγήσει σε συγκεκριμένες ενέργειες περιορισμού [109], [110].

Η σημασία των SOAR συστημάτων δεν έγκειται μόνο στην αυτοματοποίηση μεμονωμένων ενεργειών, αλλά κυρίως στην ενορχήστρωση της συνολικής διαδικασίας απόκρισης. Σε έναν μεγάλο οργανισμό, τα δεδομένα ασφαλείας προέρχονται από πολλαπλές πηγές, όπως endpoints, firewalls, cloud υπηρεσίες, συστήματα ταυτότητας, IDS, SIEM και εργαλεία threat intelligence. Η αυτοματοποίηση επιτρέπει τη σύνδεση αυτών των πηγών σε ένα ενιαίο επιχειρησιακό πλαίσιο, ώστε η απόκριση να μην είναι αποσπασματική, αλλά οργανωμένη και επαναλήψιμη [102], [104], [110].

Ένα παράδειγμα αυτοματοποιημένης απόκρισης αφορά την πιθανή ανίχνευση ransomware. Εάν ένα σύστημα εντοπίσει ασυνήθιστη μαζική τροποποίηση αρχείων, ύποπτη κρυπτογραφική δραστηριότητα ή συμπεριφορά που αποκλίνει από τη φυσιολογική λειτουργία ενός endpoint, μπορεί να ενεργοποιηθεί αυτόματη απομόνωση του συστήματος από το δίκτυο. Η ενέργεια αυτή περιορίζει την πιθανότητα εξάπλωσης της επίθεσης σε άλλα συστήματα και παρέχει στους αναλυτές χρόνο για περαιτέρω διερεύνηση [32], [31], [110].

Αντίστοιχα, σε περίπτωση ύποπτης δραστηριότητας λογαριασμού, η αυτοματοποιημένη απόκριση μπορεί να περιλαμβάνει προσωρινό περιορισμό πρόσβασης, απαίτηση πρόσθετης επαλήθευσης ή ειδοποίηση της ομάδας ασφαλείας. Η προσέγγιση αυτή είναι ιδιαίτερα σημαντική, καθώς η κατάχρηση διαπιστευτηρίων αποτελεί συχνό μηχανισμό αρχικής πρόσβασης σε πληροφοριακά συστήματα. Η ΑΙ μπορεί να συμβάλει στην αξιολόγηση του κινδύνου μιας σύνδεσης, λαμβάνοντας υπόψη στοιχεία όπως η τοποθεσία, η συσκευή, η ώρα πρόσβασης και η απόκλιση από τη συνηθισμένη συμπεριφορά του χρήστη [32], [104], [31].

Στην περίπτωση κακόβουλων συνδέσεων ή ύποπτης δικτυακής δραστηριότητας, η αυτοματοποιημένη απόκριση μπορεί να οδηγήσει σε αποκλεισμό IP διευθύνσεων, ενημέρωση κανόνων firewall ή ενεργοποίηση πολιτικών περιορισμού πρόσβασης. Παρόμοια, σε περιπτώσεις ύποπτων emails, το σύστημα μπορεί να θέσει το μήνυμα σε καραντίνα, να απομακρύνει αντίγραφα από γραμματοκιβώτια χρηστών και να ενημερώσει το προσωπικό ασφαλείας. Έτσι, η απόκριση δεν περιορίζεται στην αναγνώριση της απειλής, αλλά επεκτείνεται στη μείωση της έκθεσης των χρηστών και των συστημάτων [54], [47], [110].

Η αυτοματοποίηση μπορεί επίσης να εφαρμοστεί σε περιστατικά malware detection. Εάν εντοπιστεί ύποπτη διεργασία ή κακόβουλο αρχείο, ένα σύστημα απόκρισης μπορεί να τερματίσει

τη διεργασία, να απομονώσει το αρχείο, να συλλέξει forensic δεδομένα και να αποστείλει ειδοποίηση στους αναλυτές. Η προσέγγιση αυτή περιορίζει τον χρόνο κατά τον οποίο το κακόβουλο λογισμικό παραμένει ενεργό και μειώνει την πιθανότητα εξάπλωσης ή περαιτέρω εκτέλεσης κακόβουλων ενεργειών [54], [47], [102].

Ένα από τα σημαντικότερα πλεονεκτήματα των αυτοματοποιημένων στρατηγικών απόκρισης είναι η μείωση του χρόνου αντίδρασης. Σε αντίθεση με τις αποκλειστικά χειροκίνητες διαδικασίες, όπου κάθε βήμα απαιτεί ανθρώπινη αξιολόγηση και εκτέλεση, η αυτοματοποίηση επιτρέπει την άμεση εφαρμογή βασικών μέτρων περιορισμού. Αυτό είναι ιδιαίτερα κρίσιμο σε επιθέσεις μεγάλης ταχύτητας, όπου η καθυστέρηση μπορεί να οδηγήσει σε εξάπλωση της παραβίασης και αύξηση του κόστους αποκατάστασης [108], [106].

Παράλληλα, η αυτοματοποίηση μειώνει τον ανθρώπινο φόρτο των ομάδων ασφαλείας. Τα σύγχρονα SOC's συχνά αντιμετωπίζουν μεγάλο αριθμό alerts, εκ των οποίων μόνο ένα μέρος αντιστοιχεί σε κρίσιμα περιστατικά. Η ΑΙ μπορεί να υποστηρίξει την αρχική αξιολόγηση, την ιεράρχηση και την ομαδοποίηση των ειδοποιήσεων, ώστε οι αναλυτές να επικεντρώνονται σε περιστατικά υψηλότερης σοβαρότητας. Με τον τρόπο αυτό, η αυτοματοποίηση δεν καταργεί τον ανθρώπινο ρόλο, αλλά τον μετατοπίζει από επαναλαμβανόμενες ενέργειες προς πιο σύνθετη ανάλυση και λήψη αποφάσεων [102], [106].

Επιπλέον, οι αυτοματοποιημένες στρατηγικές απόκρισης ενισχύουν την κλιμάκωση των αμυντικών δυνατοτήτων. Σε μεγάλους οργανισμούς, η χειροκίνητη αντιμετώπιση κάθε περιστατικού είναι συχνά ανεπαρκής, ιδίως όταν οι επιθέσεις εκδηλώνονται ταυτόχρονα σε πολλαπλά endpoints, cloud υπηρεσίες ή λογαριασμούς χρηστών. Η αυτοματοποίηση επιτρέπει την ταυτόχρονη εφαρμογή μέτρων σε πολλά σημεία του οργανισμού, προσδίδοντας στην άμυνα μεγαλύτερη επιχειρησιακή ανθεκτικότητα [104], [110], [106].

Η χρήση ΑΙ και αυτοματοποίησης συμβάλλει επίσης στη μετάβαση από μια αντιδραστική μορφή κυβερνοάμυνας σε ένα περισσότερο προσαρμοστικό μοντέλο ασφάλειας. Παραδοσιακά, η απόκριση βασιζόταν κυρίως σε ανθρώπινη αξιολόγηση μετά την εκδήλωση ενός περιστατικού. Στα σύγχρονα περιβάλλοντα, η τεχνητή νοημοσύνη μπορεί να αξιοποιεί δεδομένα από προηγούμενα περιστατικά, threat intelligence και συμπεριφορικά μοτίβα, προκειμένου να προτείνει ταχύτερες και πιο στοχευμένες ενέργειες αντιμετώπισης [54], [102], [103].

Ωστόσο, η αυτοματοποιημένη απόκριση δεν πρέπει να αντιμετωπίζεται ως πλήρως αυτόνομη και αλάνθαστη λύση. Ένας βασικός περιορισμός είναι ο κίνδυνος λανθασμένων αποφάσεων. Ένα false positive μπορεί να οδηγήσει σε απομόνωση νόμιμου συστήματος, αποκλεισμό χρήστη ή διακοπή υπηρεσίας, προκαλώντας επιχειρησιακή αναστάτωση. Αντίστοιχα, ένα false negative μπορεί να επιτρέψει σε πραγματική απειλή να παραμείνει ενεργή χωρίς κατάλληλη αντίδραση [103], [8].

Η υπερβολική εξάρτηση από αυτοματοποιημένους μηχανισμούς αποτελεί επίσης σημαντική πρόκληση. Εάν ένας οργανισμός βασιστεί αποκλειστικά σε αυτοματοποιημένες αποφάσεις, χωρίς επαρκή ανθρώπινη εποπτεία, μπορεί να απολέσει τη δυνατότητα κριτικής αξιολόγησης σύνθετων ή ασαφών περιστατικών. Η ανθρώπινη κρίση παραμένει αναγκαία, ιδιαίτερα σε περιπτώσεις όπου οι συνέπειες μιας απόφασης είναι σημαντικές ή όπου τα δεδομένα δεν επιτρέπουν ασφαλές συμπέρασμα [103], [50], [9].

Η ερμηνευσιμότητα των AI-based αποφάσεων αποτελεί επιπλέον κρίσιμο ζήτημα. Σε περιβάλλοντα κυβερνοάμυνας, δεν αρκεί ένα σύστημα να προτείνει ή να εκτελεί μια ενέργεια· πρέπει να είναι δυνατό να εξηγηθεί γιατί η συγκεκριμένη ενέργεια θεωρήθηκε αναγκαία. Η έλλειψη διαφάνειας μπορεί να περιορίσει την εμπιστοσύνη των αναλυτών και να δυσχεράνει την αξιολόγηση των αυτοματοποιημένων ενεργειών [50], [9].

Ιδιαίτερη προσοχή απαιτείται και στο adversarial machine learning. Οι επιτιθέμενοι μπορούν να επιχειρήσουν να παραπλανήσουν τα AI-based συστήματα ανίχνευσης και απόκρισης, τροποποιώντας τη συμπεριφορά τους ώστε να αποφύγουν την ενεργοποίηση μέτρων ή να προκαλέσουν λανθασμένες αντιδράσεις. Το NIST επισημαίνει ότι τεχνικές όπως evasion και poisoning μπορούν να επηρεάσουν την αξιοπιστία των μοντέλων μηχανικής μάθησης, γεγονός που καθιστά απαραίτητη τη συνεχή αξιολόγηση και ανθεκτικότητα των αμυντικών συστημάτων [8].

Συνεπώς, η αποτελεσματική χρήση αυτοματοποιημένων στρατηγικών απόκρισης απαιτεί συνδυασμό τεχνολογικής αυτοματοποίησης και ανθρώπινης εποπτείας. Οι πιο κρίσιμες ενέργειες, όπως η απομόνωση κρίσιμων servers, η απενεργοποίηση λογαριασμών υψηλών δικαιωμάτων ή η διακοπή υπηρεσιών, πρέπει να υπόκεινται σε σαφείς κανόνες, μηχανισμούς επαλήθευσης και δυνατότητα ανθρώπινης παρέμβασης. Με αυτόν τον τρόπο, η αυτοματοποίηση λειτουργεί ως

ενισχυτικός μηχανισμός και όχι ως ανεξέλεγκτη αντικατάσταση της ανθρώπινης κρίσης [110], [103], [9].

Οι αυτοματοποιημένες στρατηγικές απόκρισης μπορούν να μειώσουν σημαντικά τον χρόνο αξιολόγησης και περιορισμού ενός περιστατικού ασφαλείας. Η εφαρμογή τους πρέπει να είναι προσαρμοσμένη στο επίπεδο κινδύνου κάθε ενέργειας, ώστε οι χαμηλού κινδύνου ενέργειες να μπορούν να εκτελούνται άμεσα, ενώ οι κρίσιμες παρεμβάσεις να υπόκεινται σε έλεγχο και έγκριση από αρμόδιο προσωπικό. Με αυτόν τον τρόπο, η αυτοματοποίηση ενισχύει την επιχειρησιακή ετοιμότητα χωρίς να υποκαθιστά ανεξέλεγκτα την ανθρώπινη κρίση.

7.4. Attack AI vs Defence AI

Η αντιπαράθεση μεταξύ επιθετικών και αμυντικών εφαρμογών τεχνητής νοημοσύνης αποτελεί ένα από τα πιο κρίσιμα ζητήματα της σύγχρονης κυβερνοασφάλειας. Η AI δεν μπορεί να χαρακτηριστεί μονοσήμαντα ως τεχνολογία επίθεσης ή άμυνας, καθώς η ίδια τεχνολογική βάση μπορεί να αξιοποιηθεί τόσο για την ενίσχυση κυβερνοεπιθέσεων όσο και για την ενδυνάμωση των αμυντικών μηχανισμών. Πρόκειται, επομένως, για τεχνολογία που εξυπηρετεί δύο σκοπούς, η αξία και η επικινδυνότητά της οποίας εξαρτώνται από το πλαίσιο αξιοποίησης, τον σκοπό του χρήστη και το επίπεδο ελέγχου που εφαρμόζεται κατά τη χρήση της [54], [47], [52].

Η έννοια της διπλής χρήσης είναι καθοριστική για την κατανόηση της σχέσης μεταξύ Attack AI και Defence AI. Οι ίδιες δυνατότητες που επιτρέπουν την παραγωγή κειμένου, την ανάλυση μεγάλου όγκου δεδομένων, την αναγνώριση μοτίβων και την αυτοματοποίηση ενεργειών μπορούν να χρησιμοποιηθούν από έναν επιτιθέμενο για phishing, reconnaissance, ανάπτυξη κακόβουλου περιεχομένου ή προσαρμογή επιθετικών τεχνικών. Ταυτόχρονα, οι ίδιες δυνατότητες μπορούν να αξιοποιηθούν από έναν αμυνόμενο για ανίχνευση επιθέσεων, threat intelligence, ανάλυση περιστατικών, υποστήριξη αποφάσεων και αυτοματοποιημένη απόκριση [54], [47], [111].

Από την πλευρά των επιτιθέμενων, η τεχνητή νοημοσύνη λειτουργεί ως μηχανισμός επιτάχυνσης, αυτοματοποίησης και κλιμάκωσης. Η χρήση μοντέλων τεχνητής νοημοσύνης μπορεί να μειώσει τον χρόνο που απαιτείται για τη συλλογή πληροφοριών, τη δημιουργία phishing περιεχομένου,

την παραγωγή scripts, την προσαρμογή κακόβουλων εργαλείων και την επιχειρησιακή οργάνωση μιας επίθεσης [96], [41], [39]. Με τον τρόπο αυτό, η ΑΙ αυξάνει την επιχειρησιακή αποδοτικότητα των επιτιθέμενων και μειώνει την ανάγκη συνεχούς ανθρώπινης παρέμβασης σε επαναλαμβανόμενα ή τεχνικά απαιτητικά στάδια.

Ιδιαίτερη σημασία έχει η επίδραση της ΑΙ στη μείωση των τεχνικών εμποδίων εισόδου. Δυνατότητες που παλαιότερα απαιτούσαν υψηλή τεχνική εξειδίκευση μπορούν πλέον να υποστηριχθούν από εργαλεία ΑΙ, τα οποία παρέχουν βοήθεια σε επίπεδο κώδικα, ανάλυσης, προσαρμογής περιεχομένου και επιχειρησιακού σχεδιασμού. Αυτό δεν σημαίνει ότι η ΑΙ καθιστά κάθε χρήστη ικανό να εκτελέσει προηγμένες κυβερνοεπιθέσεις, αλλά ότι μειώνει σημαντικά την απόσταση μεταξύ λιγότερο έμπειρων και περισσότερο εξειδικευμένων επιτιθέμενων [52], [41], [39].

Η χρήση ΑΙ από επιτιθέμενους ενισχύει επίσης την κλίμακα των επιθέσεων. Η μαζική δημιουργία προσωποποιημένων phishing μηνυμάτων, η αυτοματοποίηση της αναγνώρισης στόχων, η ταχύτερη ανάλυση τεχνικών πληροφοριών και η δυνατότητα προσαρμογής σε διαφορετικά περιβάλλοντα αυξάνουν την ικανότητα των επιτιθέμενων να επιχειρούν εναντίον μεγάλου αριθμού στόχων με μειωμένο κόστος [96], [41], [69]. Η εξέλιξη αυτή μεταβάλλει τη δυναμική του κυβερνοεγκλήματος, καθώς επιτρέπει σε μικρότερες ομάδες ή ακόμη και μεμονωμένους δράστες να αποκτούν δυσανάλογα υψηλή επιχειρησιακή επίδραση.

Η εμφάνιση agentic ΑΙ ενισχύει περαιτέρω αυτή τη μεταβολή. Τα agentic συστήματα μπορούν να εκτελούν αλληλουχίες ενεργειών, να διατηρούν επιχειρησιακό πλαίσιο, να αξιολογούν ενδιάμεσα αποτελέσματα και να προσαρμόζουν τη συμπεριφορά τους ανάλογα με τα δεδομένα που λαμβάνουν. Στο επιθετικό πεδίο, αυτό οδηγεί σε μεγαλύτερη αυτοματοποίηση του κύκλου επίθεσης και σε μια μορφή βιομηχανοποίησης των κυβερνοεπιθέσεων, όπου η ανθρώπινη συμμετοχή μετατοπίζεται από τη χειροκίνητη εκτέλεση στην επίβλεψη και καθοδήγηση των εργαλείων [97], [81], [69].

Από την πλευρά των αμυνόμενων, η τεχνητή νοημοσύνη προσφέρει επίσης σημαντικά πλεονεκτήματα. Τα ΑΙ-based αμυντικά συστήματα μπορούν να αναλύουν μεγάλο όγκο δεδομένων ασφαλείας, να εντοπίζουν ανωμαλίες, να συσχετίζουν συμβάντα από διαφορετικές πηγές και να υποστηρίζουν ταχύτερη λήψη αποφάσεων. Συστήματα IDS, SIEM, XDR, SOC και automated

response μπορούν να ενισχυθούν μέσω αλγορίθμων μηχανικής μάθησης, μοντέλων ανάλυσης συμπεριφοράς και τεχνικών threat intelligence [47], [32], [111].

Η βασική αμυντική αξία της ΑΙ έγκειται στην ικανότητά της να εντοπίζει μοτίβα που δεν είναι άμεσα ορατά με παραδοσιακές μεθόδους. Σε ένα περιβάλλον όπου οι επιθέσεις γίνονται πιο γρήγορες, πολυσταδιακές και προσαρμοστικές, η άμυνα χρειάζεται μηχανισμούς που μπορούν να λειτουργούν σε πραγματικό ή σχεδόν πραγματικό χρόνο. Η ΑΙ μπορεί να βοηθήσει στην ανίχνευση ανωμαλιών, στην ιεράρχηση ειδοποιήσεων, στην ανάλυση endpoint δραστηριότητας και στην υποστήριξη αμυντικών workflows [32], [111], [8].

Ωστόσο, η Defence AI δεν είναι ανεξάρτητη από περιορισμούς. Η αποτελεσματικότητά της εξαρτάται από την ποιότητα των δεδομένων, την αντιπροσωπευτικότητα των datasets, την ερμηνευσιμότητα των μοντέλων και την ανθεκτικότητα απέναντι σε adversarial attacks. Τα αμυντικά συστήματα που βασίζονται σε ΑΙ μπορούν να στοχοποιηθούν από επιτιθέμενους μέσω τεχνικών evasion, poisoning ή άλλων μορφών adversarial manipulation, με στόχο την αποφυγή ανίχνευσης ή την πρόκληση λανθασμένων αποφάσεων [50], [8].

Η σύγκριση μεταξύ Attack AI και Defence AI δείχνει ότι η τεχνητή νοημοσύνη ενισχύει και τις δύο πλευρές, αλλά με διαφορετικό τρόπο. Για τους επιτιθέμενους, η ΑΙ προσφέρει κυρίως αυτοματοποίηση, κλιμάκωση, προσωποποίηση και μείωση του κόστους. Για τους αμυνόμενους, προσφέρει κυρίως ανάλυση μεγάλης κλίμακας, ανίχνευση ανωμαλιών, ταχύτερη αξιολόγηση περιστατικών και υποστήριξη απόκρισης [54], [52], [32].

Πίνακας 7.1. Διαφορές για χρήσεις τεχνητής νοημοσύνης

ΠΑΡΑΜΕΤΡΟΣ	ATTACK AI	DEFENCE AI
Αυτοματοποίηση	Αυτοματοποιεί reconnaissance, phishing, scripting και επιθετικά workflows	Αυτοματοποιεί ανίχνευση, alert triage, συσχέτιση συμβάντων και απόκριση

Κλιμάκωση	Επιτρέπει μαζική παραγωγή επιθετικού περιεχομένου και ταυτόχρονη στόχευση πολλών θυμάτων	Επιτρέπει ανάλυση μεγάλου όγκου logs, alerts, endpoints και cloud events
Κόστος	Μειώνει το κόστος σχεδιασμού και εκτέλεσης επιθέσεων	Μειώνει τον ανθρώπινο φόρτο και τον χρόνο αξιολόγησης περιστατικών
Τεχνικές Απαιτήσεις	Μειώνει τα εμπόδια εισόδου για λιγότερο έμπειρους επιτιθέμενους	Υποστηρίζει αναλυτές ασφαλείας σε σύνθετα περιβάλλοντα
Προσαρμοστικότητα	Επιτρέπει δυναμική προσαρμογή επιθετικών ενεργειών	Επιτρέπει δυναμική ανάλυση και προσαρμογή αμυντικών μηχανισμών
Εξάρτηση από δεδομένα	Αξιοποιεί OSINT, κλεμμένα δεδομένα, logs και τεχνικές πληροφορίες	Εξαρτάται από ποιοτικά δεδομένα ασφαλείας και threat intelligence
Κίνδυνοι	Κακόβουλη χρήση, αυτοματοποίηση επιθέσεων, agentic exploitation	False positives, adversarial attacks, υπερβολική εξάρτηση

Η συγκριτική αυτή ανάλυση αναδεικνύει ότι η ΑΙ δεν δημιουργεί απλώς ισχυρότερα εργαλεία επίθεσης ή άμυνας, αλλά μεταβάλλει τους όρους του ανταγωνισμού μεταξύ των δύο πλευρών. Η επίθεση αποκτά μεγαλύτερη ταχύτητα και κλίμακα, ενώ η άμυνα αποκτά μεγαλύτερη ικανότητα ανάλυσης και αυτοματοποίησης. Το κρίσιμο ζήτημα είναι ότι οι δύο αυτές εξελίξεις δεν συμβαίνουν ανεξάρτητα, αλλά αλληλοτροφοδοτούνται [42], [50].

Η σχέση επίθεσης και άμυνας στον κυβερνοχώρο χαρακτηρίζεται από εγγενή ασυμμετρία. Ο επιτιθέμενος συχνά χρειάζεται να επιτύχει μία φορά για να προκαλέσει παραβίαση, ενώ ο αμυνόμενος πρέπει να προστατεύει συνεχώς ένα μεγάλο και δυναμικό σύνολο συστημάτων,

χρηστών, εφαρμογών και δεδομένων. Η τεχνητή νοημοσύνη δεν εξαλείφει αυτή την ασυμμετρία. Αντίθετα, σε ορισμένες περιπτώσεις μπορεί να την ενισχύσει, επειδή επιτρέπει στους επιτιθέμενους να δοκιμάζουν περισσότερες τεχνικές, να προσαρμόζουν ταχύτερα τις μεθόδους τους και να εκτελούν επιθέσεις με χαμηλότερο κόστος [42], [50], [69].

Από την άλλη πλευρά, η Defence AI μπορεί να περιορίσει μέρος αυτής της ασυμμετρίας, εφόσον ενσωματωθεί αποτελεσματικά σε πολυεπίπεδες αρχιτεκτονικές κυβερνοάμυνας. Η δυνατότητα συνεχούς παρακολούθησης, ανάλυσης συμπεριφοράς, αυτοματοποιημένης συσχέτισης συμβάντων και ταχύτερης απόκρισης επιτρέπει στους οργανισμούς να αντιδρούν πιο γρήγορα και πιο στοχευμένα. Ωστόσο, η αποτελεσματικότητα αυτή δεν είναι δεδομένη. Εξαρτάται από την ποιότητα εφαρμογής, το επίπεδο ωριμότητας του οργανισμού, την ανθρώπινη εποπτεία και την ικανότητα συνεχούς προσαρμογής των αμυντικών μηχανισμών [32], [111], [8].

Η σύγκρουση Attack AI και Defence AI οδηγεί σε μια μορφή συνεχούς τεχνολογικής κούρσας. Καθώς οι επιτιθέμενοι αξιοποιούν AI για να δημιουργούν πιο πειστικές, ταχύτερες και προσαρμοστικές επιθέσεις, οι αμυνόμενοι αναπτύσσουν AI-based μηχανισμούς για να ανιχνεύουν, να αναλύουν και να αντιμετωπίζουν αυτές τις απειλές. Με τη σειρά τους, οι βελτιωμένοι αμυντικοί μηχανισμοί ωθούν τους επιτιθέμενους να αναπτύξουν πιο εξελιγμένες τεχνικές αποφυγής, παραπλάνησης ή παράκαμψης [42], [97], [81], [69].

Η έννοια του AI arms race δεν πρέπει να εκληφθεί ως απλή μεταφορά, αλλά ως πραγματική περιγραφή μιας δυναμικής αλληλεξάρτησης. Κάθε βελτίωση στην επιθετική χρήση της AI δημιουργεί ανάγκη για αντίστοιχη αμυντική προσαρμογή. Αντίστοιχα, κάθε βελτίωση στην αμυντική AI μπορεί να οδηγήσει σε νέα επιθετικά μοτίβα, σχεδιασμένα ώστε να αποφεύγουν ή να παραπλανούν τα μοντέλα ανίχνευσης [42], [8], [69].

Το ερώτημα εάν η AI ευνοεί περισσότερο τον επιτιθέμενο ή τον αμυνόμενο δεν μπορεί να απαντηθεί απόλυτα. Η επίδραση της AI εξαρτάται από το επίπεδο τεχνολογικής ωριμότητας, τους διαθέσιμους πόρους, την πρόσβαση σε δεδομένα, την ποιότητα των μοντέλων και την ικανότητα επιχειρησιακής ενσωμάτωσης. Σε οργανισμούς με ώριμες δομές κυβερνοασφάλειας, η AI μπορεί να ενισχύσει σημαντικά την άμυνα. Αντίθετα, σε οργανισμούς με περιορισμένες αμυντικές δυνατότητες, η χρήση AI από επιτιθέμενους μπορεί να αυξήσει δυσανάλογα την πίεση προς τα συστήματα προστασίας [42], [50], [32].

Παρά την αμφίπλευρη ενίσχυση, αρκετές εφαρμογές ΑΙ φαίνεται να προσφέρουν άμεσα και χαμηλού κόστους οφέλη στους επιτιθέμενους, κυρίως ως προς την αυτοματοποίηση, την κλιμάκωση, την προσωποποίηση και τη μείωση των τεχνικών απαιτήσεων. Αυτό δημιουργεί αυξημένη πίεση στους αμυνόμενους να υιοθετήσουν αντίστοιχες ή ανώτερες τεχνολογικές δυνατότητες, ώστε να διατηρήσουν αποτελεσματικό επίπεδο προστασίας [52], [96], [39], [69].

Στο επίπεδο της στρατηγικής κυβερνοασφάλειας, η Defence AI δεν πρέπει να αντιμετωπίζεται ως προαιρετική τεχνολογική αναβάθμιση, αλλά ως αναγκαία προσαρμογή στο νέο περιβάλλον απειλών. Εφόσον οι επιτιθέμενοι χρησιμοποιούν την τεχνητή νοημοσύνη για να αυξήσουν την ταχύτητα, την κλίμακα και την πολυπλοκότητα των επιθέσεων τους, οι αμυνόμενοι δεν μπορούν να βασίζονται αποκλειστικά σε παραδοσιακές μεθόδους ανίχνευσης και χειροκίνητης απόκρισης. Η αμυντική ενσωμάτωση την τεχνητής νοημοσύνης καθίσταται αναγκαία για τη διατήρηση της επιχειρησιακής ισορροπίας [32], [111], [8].

Παράλληλα, η ανάπτυξη Defence AI πρέπει να συνοδεύεται από υπεύθυνη διακυβέρνηση, διαφάνεια, αξιολόγηση κινδύνων και ανθρώπινη εποπτεία. Η χρήση ΑΙ στην κυβερνοάμυνα δεν είναι απαλλαγμένη από κινδύνους, καθώς λανθασμένες αποφάσεις, ανεπαρκής ερμηνευσιμότητα ή adversarial manipulation μπορούν να οδηγήσουν σε επιχειρησιακές αστοχίες. Επομένως, η αποτελεσματική άμυνα δεν εξαρτάται μόνο από την ύπαρξη ΑΙ εργαλείων, αλλά και από τον τρόπο με τον οποίο αυτά ενσωματώνονται σε θεσμικά, τεχνικά και οργανωτικά πλαίσια ελέγχου [50], [9], [8].

Συνολικά, η αντιπαράθεση Attack AI και Defence AI δείχνει ότι η Τεχνητή Νοημοσύνη μετασχηματίζει ταυτόχρονα και τις δύο πλευρές της κυβερνοασφάλειας. Η ΑΙ αυξάνει την ταχύτητα, την κλίμακα και την προσαρμοστικότητα των επιθέσεων, αλλά παράλληλα ενισχύει την ανίχνευση, την ανάλυση και την απόκριση των αμυνόμενων. Η ισορροπία μεταξύ επίθεσης και άμυνας δεν μετατοπίζεται οριστικά προς μία πλευρά, αλλά γίνεται πιο δυναμική, πιο ανταγωνιστική και περισσότερο εξαρτημένη από την ικανότητα κάθε πλευράς να αξιοποιεί αποτελεσματικά την τεχνολογία [54], [42], [50], [32].

Η σχέση μεταξύ Attack AI και Defence AI διαμορφώνει ένα δυναμικό περιβάλλον τεχνολογικού ανταγωνισμού. Η ΑΙ μπορεί να αυξήσει την ταχύτητα, την προσαρμοστικότητα και την κλίμακα επιθετικών ενεργειών, ενώ ταυτόχρονα ενισχύει την ανίχνευση, την ανάλυση και την απόκριση

των αμυνόμενων. Η ισορροπία δεν μετατοπίζεται οριστικά προς μία πλευρά, αλλά εξαρτάται από την ποιότητα των δεδομένων, την τεχνολογική ωριμότητα και τη στρατηγική αξιοποίηση των διαθέσιμων δυνατοτήτων.

8. ΝΟΜΙΚΕΣ ΚΑΙ ΗΘΙΚΕΣ ΔΙΑΣΤΑΣΕΙΣ ΤΗΣ ΧΡΗΣΗΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΣΕ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΙΣ

8.1 Ηθικά διλήμματα της χρήσης Τεχνητής Νοημοσύνης σε κυβερνοεπιθέσεις

Η χρήση της τεχνητής νοημοσύνης σε κυβερνοεπιθέσεις δημιουργεί ένα σύνθετο πεδίο ηθικού προβληματισμού, το οποίο δεν μπορεί να περιοριστεί στη γενική διαπίστωση ότι η κακόβουλη χρήση της τεχνολογίας είναι ηθικά προβληματική. Το ουσιώδες ζήτημα είναι ότι η ΑΙ μεταβάλλει τους όρους με τους οποίους σχεδιάζονται, εκτελούνται και αξιολογούνται οι κυβερνοεπιθέσεις. Η δυνατότητα αυτοματοποίησης, η αύξηση της κλίμακας, η μείωση της ανθρώπινης εμπλοκής, η χρήση συνθετικού περιεχομένου και η δυσκολία απόδοσης ευθύνης δημιουργούν νέα ηθικά διλήμματα, τα οποία συνδέονται άμεσα με τη φύση της ΑΙ αυτής καθεαυτής [149], [50], [52].

Η τεχνητή νοημοσύνη δεν είναι από τη φύση της ούτε αμυντική ούτε επιθετική τεχνολογία. Η ηθική της αξιολόγηση εξαρτάται από το σκοπό χρήσης, το πλαίσιο εφαρμογής, τον βαθμό ανθρώπινου ελέγχου και τις συνέπειες που προκαλεί σε τρίτους. Οι ίδιες τεχνικές που μπορούν να χρησιμοποιηθούν για την ανίχνευση απειλών, την ανάλυση κακόβουλου λογισμικού και την ενίσχυση της κυβερνοάμυνας μπορούν επίσης να αξιοποιηθούν για phishing, social engineering, αυτοματοποιημένη αναγνώριση στόχων, παραγωγή κακόβουλου περιεχομένου ή δημιουργία deepfakes [149], [52], [54]. Αυτός ο διπλός χαρακτήρας καθιστά την ΑΙ ιδιαίτερα δύσκολη από ηθικής πλευράς, διότι η ίδια τεχνολογική δυνατότητα μπορεί να έχει θετική ή επιβλαβή λειτουργία ανάλογα με το ποιος τη χρησιμοποιεί και για ποιο σκοπό.

Το πρώτο βασικό ηθικό δίλημμα αφορά την ευθύνη. Όταν μια κυβερνοεπίθεση υποστηρίζεται από ΑΙ, η ευθύνη δεν εξαφανίζεται, αλλά γίνεται πιο σύνθετη. Σε μια παραδοσιακή επίθεση, η ανθρώπινη πρόθεση και η ανθρώπινη εκτέλεση είναι πιο άμεσα αναγνωρίσιμες. Αντίθετα, σε ένα ΑΙ-assisted περιβάλλον μπορεί να εμπλέκονται ο δημιουργός του μοντέλου, ο πάροχος της υπηρεσίας, ο χρήστης που το αξιοποιεί κακόβουλα, οι διαχειριστές της πλατφόρμας και οι οργανισμοί που δεν έθεσαν επαρκείς περιορισμούς. Η ηθική ευθύνη, επομένως, δεν μπορεί να αποδοθεί μονοσήμαντα σε ένα τεχνικό σύστημα, αλλά πρέπει να εξετάζεται σε σχέση με τις ανθρώπινες αποφάσεις που επέτρεψαν, διευκόλυναν ή δεν απέτρεψαν την κακόβουλη χρήση [45], [50], [154].

Η αυτοματοποίηση εντείνει το πρόβλημα της ευθύνης. Όσο περισσότερα στάδια μιας επίθεσης εκτελούνται από ΑΙ συστήματα, τόσο πιο δύσκολο γίνεται να προσδιοριστεί ποιο τμήμα της βλάβης οφείλεται στην ανθρώπινη εντολή, ποιο στην αυτόνομη λειτουργία του εργαλείου και ποιο σε ανεπαρκείς δικλίδες ασφαλείας. Αναφορές πραγματικών περιστατικών κατάχρησης ΑΙ δείχνουν ότι τα μοντέλα μπορούν να χρησιμοποιηθούν για υποστήριξη phishing, scripting, επιχειρησιακής οργάνωσης και άλλων επιμέρους σταδίων κακόβουλης δραστηριότητας [41], [40]. Η ύπαρξη τέτοιων περιστατικών δεν σημαίνει ότι η ΑΙ λειτουργεί ανεξάρτητα από τον άνθρωπο, αλλά δείχνει ότι η ανθρώπινη ευθύνη διαχέεται σε περισσότερα επίπεδα σχεδιασμού, χρήσης και ελέγχου.

Συναφές ηθικό ζήτημα είναι η μείωση της ανθρώπινης εμπλοκής και η δημιουργία ηθικής απόστασης. Όταν ένας επιτιθέμενος χρησιμοποιεί ΑΙ εργαλεία για να παράγει μηνύματα, να επιλέγει στόχους ή να αυτοματοποιεί επαναλαμβανόμενες ενέργειες, η βλάβη μπορεί να εμφανίζεται ως αποτέλεσμα μιας τεχνικής διαδικασίας και όχι ως άμεση ανθρώπινη πράξη. Αυτή η απόσταση ενδέχεται να μειώσει το ψυχολογικό εμπόδιο της επίθεσης, διότι ο δράστης δεν έρχεται αντιμέτωπος με κάθε επιμέρους συνέπεια της ενέργειάς του. Η αυτοματοποίηση, όμως, δεν μειώνει την ηθική βαρύτητα της πράξης. Αντίθετα, μπορεί να την αυξήσει, επειδή επιτρέπει την πρόκληση βλάβης σε μεγαλύτερη κλίμακα και με μικρότερη ανθρώπινη προσπάθεια [42], [50], [97].

Η κλίμακα της βλάβης αποτελεί ξεχωριστό ηθικό ζήτημα. Η ΑΙ μπορεί να επιτρέψει μαζικό phishing, εξατομικευμένη παραπλάνηση, παραγωγή μεγάλου όγκου κακόβουλου περιεχομένου και αυτοματοποιημένη στόχευση πολλών θυμάτων. Το ηθικό πρόβλημα δεν περιορίζεται στο ότι μια επίθεση μπορεί να γίνει πιο αποτελεσματική. Το βαθύτερο ζήτημα είναι ότι η βλάβη μπορεί να επεκταθεί από μεμονωμένα άτομα σε οργανισμούς, δημόσιες υπηρεσίες, οικονομικά συστήματα και ευρύτερες κοινωνικές ομάδες. Η ΑΙ, ως πολλαπλασιαστής κλίμακας, αυξάνει την πιθανότητα οι επιθέσεις να προκαλέσουν πολλές συνέπειες που να υπερβαίνουν την αρχική πρόθεση του επιτιθέμενου [45], [42], [52].

Ιδιαίτερα κρίσιμη είναι η ηθική διάσταση των επιθέσεων social engineering. Η χρήση generative ΑΙ επιτρέπει τη δημιουργία πιο φυσικών, πειστικών και προσωποποιημένων μηνυμάτων, μειώνοντας τα γλωσσικά ή τεχνικά σημάδια που θα μπορούσαν να προειδοποιήσουν το θύμα. Στο

πλαίσιο αυτό, η επίθεση δεν στρέφεται μόνο εναντίον ενός πληροφοριακού συστήματος, αλλά εναντίον της ανθρώπινης κρίσης και εμπιστοσύνης. Η ΑΙ μπορεί να αξιοποιήσει διαθέσιμες πληροφορίες για το θύμα, να προσαρμόσει το περιεχόμενο στο προφίλ του και να δημιουργήσει επικοινωνία που φαίνεται αξιόπιστη [54], [47], [150]. Το ηθικό πρόβλημα εδώ είναι η χειραγώγηση της ικανότητας του ατόμου να λάβει ελεύθερη και ενημερωμένη απόφαση.

Η χρήση deepfakes και συνθετικού οπτικοακουστικού περιεχομένου ενισχύει ακόμη περισσότερο αυτή τη μορφή χειραγώγησης. Μέσω τεχνικών παραγωγής ψεύτικης φωνής, εικόνας ή βίντεο, ο επιτιθέμενος μπορεί να μιμηθεί στελέχη εταιρειών, δημόσιες αρχές, συγγενείς ή άλλες αξιόπιστες ταυτότητες. Η παραπλάνηση δεν βασίζεται πλέον μόνο στο περιεχόμενο ενός μηνύματος, αλλά στην ίδια την ψευδή αναπαράσταση της παρουσίας ενός προσώπου. Αυτό υπονομεύει τη δυνατότητα διάκρισης ανάμεσα σε αυθεντική και κατασκευασμένη επικοινωνία και πλήττει την εμπιστοσύνη στις ψηφιακές αλληλεπιδράσεις [150], [151], [142].

Τα deepfakes δημιουργούν και ηθικό ζήτημα γύρω από την αξιοπιστία των αποδεικτικών στοιχείων. Σε ένα περιβάλλον όπου εικόνες, φωνές και βίντεο μπορούν να παραχθούν ή να τροποποιηθούν τεχνητά, η ύπαρξη ενός ψηφιακού αρχείου δεν αρκεί για να θεωρηθεί αυτόματα αυθεντικό. Η δυνατότητα κατασκευής πειστικού ψευδούς υλικού μπορεί να χρησιμοποιηθεί όχι μόνο για εξαπάτηση θυμάτων, αλλά και για παραπλάνηση ερευνών, δυσφήμιση, εκβιασμό ή αμφισβήτηση πραγματικών αποδεικτικών στοιχείων. Η ανάγκη τεχνικής επαλήθευσης της αυθεντικότητας αποκτά επομένως και ηθική διάσταση, επειδή συνδέεται με την προστασία της αλήθειας, της δικαιοσύνης και της εμπιστοσύνης στους θεσμούς [142], [144].

Ένα ακόμη σημαντικό δίλημμα αφορά την ιδιωτικότητα και την κατάχρηση δεδομένων. Πολλές AI-assisted επιθέσεις βασίζονται στη συλλογή, συσχέτιση και ανάλυση δεδομένων από κοινωνικά δίκτυα, δημόσια προφίλ, προηγούμενες διαρροές, επαγγελματικές πλατφόρμες ή άλλες ανοιχτές πηγές. Το ηθικό πρόβλημα δεν αφορά μόνο την τελική επίθεση, αλλά και την εκμετάλλευση προσωπικών δεδομένων για τη δημιουργία στοχευμένων επιθετικών ενεργειών. Όταν τα δεδομένα ενός ατόμου χρησιμοποιούνται για να κατασκευαστεί εναντίον του μια πιο πειστική απάτη, η προσβολή της ιδιωτικότητας μετατρέπεται σε εργαλείο χειραγώγησης [45], [50], [155].

Η κατάχρηση δεδομένων συνδέεται και με το πρόβλημα της πληροφοριακής ανισότητας. Ο επιτιθέμενος μπορεί να αξιοποιήσει ΑΙ για να συγκεντρώσει και να ερμηνεύσει μεγάλο όγκο

πληροφοριών για ένα θύμα, ενώ το ίδιο το θύμα δεν γνωρίζει ποια δεδομένα έχουν συλλεχθεί, πώς έχουν συσχετιστεί και με ποιον τρόπο χρησιμοποιούνται εναντίον του. Η ανισορροπία αυτή περιορίζει την πραγματική αυτονομία του ατόμου, καθώς η απόφασή του επηρεάζεται από περιεχόμενο που έχει κατασκευαστεί με βάση πληροφορίες τις οποίες δεν γνωρίζει ότι έχουν αξιοποιηθεί [150], [155].

Η ηθική ανάλυση δεν πρέπει να περιορίζεται στην επιθετική χρήση της ΑΙ. Ακόμη και η αμυντική αξιοποίηση της τεχνητής νοημοσύνης δημιουργεί σημαντικά διλήμματα. Τα AI-based συστήματα κυβερνοάμυνας μπορούν να παρακολουθούν δικτυακή δραστηριότητα, να αναλύουν συμπεριφορά χρηστών, να αξιολογούν κινδύνους και να ενεργοποιούν αυτοματοποιημένες ενέργειες. Αν και οι λειτουργίες αυτές μπορούν να ενισχύσουν την ασφάλεια, ενδέχεται να επηρεάσουν την ιδιωτικότητα, την εργασιακή αυτονομία, τη διαφάνεια και την αναλογικότητα της επιτήρησης [141], [153], [154].

Ένα αμυντικό σύστημα ΑΙ μπορεί, για παράδειγμα, να χαρακτηρίσει εσφαλμένα έναν χρήστη ως ύποπτο, να περιορίσει την πρόσβασή του, να μπλοκάρει δραστηριότητες ή να ενεργοποιήσει διαδικασίες ελέγχου χωρίς επαρκή ανθρώπινη αξιολόγηση. Τα false positives δεν αποτελούν μόνο τεχνικό πρόβλημα, αλλά και ηθικό ζήτημα, επειδή μπορούν να οδηγήσουν σε αδικαιολόγητους περιορισμούς ή λανθασμένη στοχοποίηση. Η ανάγκη ανθρώπινης εποπτείας είναι επομένως κρίσιμη, ιδίως όταν οι αποφάσεις ενός συστήματος επηρεάζουν δικαιώματα, πρόσβαση ή φήμη προσώπων και οργανισμών [141], [152], [153].

Η ανθρώπινη εποπτεία δεν πρέπει να αντιμετωπίζεται τυπικά, ως απλή παρουσία ανθρώπου στο τελικό στάδιο της διαδικασίας. Για να έχει ουσιαστικό περιεχόμενο, πρέπει ο άνθρωπος να μπορεί να κατανοήσει τη βάση της απόφασης, να αμφισβητήσει το αποτέλεσμα, να παρέμβει όταν υπάρχει αβεβαιότητα και να αναλάβει ευθύνη για την τελική ενέργεια. Διαφορετικά, δημιουργείται ο κίνδυνος της ψευδούς εποπτείας, όπου ο άνθρωπος απλώς επικυρώνει αυτόματα αποφάσεις που δεν έχει ουσιαστικά ελέγξει [152], [153].

Σημαντικό ηθικό ζήτημα αποτελεί και η ασφάλεια των ίδιων των ΑΙ συστημάτων. Τα συστήματα που χρησιμοποιούνται για άμυνα μπορούν να γίνουν στόχος adversarial attacks, όπως evasion, poisoning ή άλλες μορφές χειραγώγησης εισόδων και δεδομένων. Εάν ένας επιτιθέμενος καταφέρει να παραπλανήσει ένα αμυντικό μοντέλο, μπορεί να αποφύγει την ανίχνευση ή να

προκαλέσει λανθασμένες αποφάσεις. Η ανθεκτικότητα των ΑΙ συστημάτων δεν είναι επομένως μόνο τεχνική απαίτηση, αλλά και ηθική υποχρέωση, διότι η χρήση μη ασφαλών συστημάτων σε κρίσιμα περιβάλλοντα μπορεί να δημιουργήσει νέους κινδύνους για τρίτους [8], [154], [155].

Η ανάγκη για διαφάνεια και λογοδοσία είναι κεντρική σε κάθε μορφή ΑΙ χρήσης στην κυβερνοασφάλεια. Όταν ένα σύστημα προτείνει, υποστηρίζει ή εκτελεί ενέργειες που επηρεάζουν ανθρώπους, οργανισμούς ή θεσμούς, πρέπει να υπάρχουν μηχανισμοί εξήγησης, ελέγχου και κατανομής ευθύνης. Η αδιαφάνεια μπορεί να οδηγήσει σε αδυναμία αμφισβήτησης μιας απόφασης, σε μεταφορά ευθύνης στο τεχνικό σύστημα και σε υποβάθμιση του ανθρώπινου ρόλου. Για τον λόγο αυτό, αρχές όπως η explainability, η accountability, η privacy, η fairness και η robustness αποτελούν βασικά στοιχεία υπεύθυνης διακυβέρνησης τεχνητής νοημοσύνης [45], [154], [155].

Στο πλαίσιο των κυβερνοεπιθέσεων, η ηθική διακυβέρνηση πρέπει να καλύπτει τόσο το στάδιο ανάπτυξης όσο και το στάδιο διάθεσης και χρήσης των ΑΙ συστημάτων. Οι δημιουργοί και πάροχοι προηγμένων μοντέλων οφείλουν να λαμβάνουν υπόψη πιθανές μορφές κατάχρησης, να εφαρμόζουν μηχανισμούς περιορισμού, να παρακολουθούν κακόβουλα πρότυπα χρήσης και να ενισχύουν την ασφάλεια των συστημάτων τους. Αναφορές για διακοπή κακόβουλων χρήσεων ΑΙ δείχνουν ότι η υπεύθυνη λειτουργία των παρόχων δεν μπορεί να περιορίζεται στην αρχική σχεδίαση, αλλά απαιτεί συνεχή αξιολόγηση και προσαρμογή [41], [40], [97].

Παράλληλα, η υπεύθυνη χρήση ΑΙ στην κυβερνοασφάλεια απαιτεί συνεργασία τεχνικών, νομικών, ηθικών και οργανωτικών φορέων. Τα ηθικά διλήμματα δεν μπορούν να λυθούν μόνο με καλύτερους αλγόριθμους, διότι αφορούν ανθρώπινη βλάβη, θεσμική εμπιστοσύνη, ευθύνη, ιδιωτικότητα και ασφάλεια. Η αντιμετώπιση των AI-enabled κυβερνοεπιθέσεων χρειάζεται πλαίσια στα οποία η τεχνική καινοτομία συνοδεύεται από κανόνες χρήσης, διαδικασίες ελέγχου, μηχανισμούς λογοδοσίας και αξιολόγηση συνεπειών [45], [42], [50].

Η χρήση τεχνητής νοημοσύνης σε κυβερνοεπιθέσεις αναδεικνύει ένα νέο επίπεδο ηθικής πολυπλοκότητας. Η ΑΙ μπορεί να αυτοματοποιήσει επιβλαβείς ενέργειες, να αυξήσει την κλίμακα της βλάβης, να μειώσει την ανθρώπινη εμπλοκή, να δυσχεράνει την απόδοση ευθύνης, να ενισχύσει την παραπλάνηση και να αξιοποιήσει προσωπικά δεδομένα για στοχευμένη χειραγώγηση [149], [52], [150]. Ταυτόχρονα, η αμυντική χρήση της ΑΙ δημιουργεί δικά της ηθικά

ερωτήματα σχετικά με την εποπτεία, τη διαφάνεια, την ιδιωτικότητα, την αξιοπιστία και την ασφάλεια των συστημάτων [141], [8], [153].

Η ηθική διάσταση των AI-enabled κυβερνοεπιθέσεων δεν περιορίζεται στην τεχνική αποτελεσματικότητα των συστημάτων, αλλά αφορά την ευθύνη, την ανθρώπινη εποπτεία, την προστασία της ιδιωτικότητας και την πρόληψη της κατάχρησης. Η AI μπορεί να ενισχύσει την κλίμακα της βλάβης, να δυσχεράνει την απόδοση ευθύνης και να αυξήσει την αποτελεσματικότητα της παραπλάνησης. Για τον λόγο αυτό, η υπεύθυνη χρήση της στην κυβερνοασφάλεια απαιτεί σαφείς αρχές διαφάνειας, λογοδοσίας, ασφάλειας σχεδιασμού και ουσιαστικού ανθρώπινου ελέγχου.

8.2 Νομοθετικό πλαίσιο για την Τεχνητή Νοημοσύνη και την κυβερνοασφάλεια

Η αντιμετώπιση των κυβερνοεπιθέσεων που αξιοποιούν τεχνητή νοημοσύνη δεν μπορεί να βασιστεί αποκλειστικά σε τεχνικά μέτρα άμυνας. Η χρήση AI σε επιθετικά περιβάλλοντα δημιουργεί ζητήματα προστασίας προσωπικών δεδομένων, ασφάλειας συστημάτων, διαφάνειας, ανθρώπινης εποπτείας και οργανωτικής ευθύνης. Για τον λόγο αυτό, το νομικό πλαίσιο αποκτά ιδιαίτερη σημασία, καθώς επιχειρεί να θέσει κανόνες πρόληψης, συμμόρφωσης και διαχείρισης κινδύνου πριν από την εκδήλωση μιας κυβερνοεπίθεσης. Στο ευρωπαϊκό περιβάλλον, η ρύθμιση των κινδύνων που συνδέονται με την τεχνητή νοημοσύνη και την κυβερνοασφάλεια διαμορφώνεται μέσα από ένα πολυεπίπεδο σύστημα κανόνων, στο οποίο ο GDPR προστατεύει τα προσωπικά δεδομένα, ο AI Act ρυθμίζει τη χρήση συστημάτων AI με βάση τον κίνδυνο, η NIS2 ενισχύει την κυβερνοανθεκτικότητα οργανισμών και κρίσιμων τομέων, ενώ ο Cyber Resilience Act εισάγει απαιτήσεις ασφάλειας για προϊόντα με ψηφιακά στοιχεία [5], [157], [160], [162].

Η ανάγκη για νομική ρύθμιση γίνεται πιο έντονη λόγω της φύσης των AI-enabled κυβερνοεπιθέσεων. Όπως αναλύθηκε στα προηγούμενα κεφάλαια, η τεχνητή νοημοσύνη μπορεί να αξιοποιηθεί για αυτοματοποίηση επιθέσεων, εξατομίκευση phishing, social engineering, παραγωγή συνθετικού περιεχομένου, συλλογή πληροφοριών και βελτιστοποίηση κακόβουλων ενεργειών. Οι πρακτικές αυτές δεν απειλούν μόνο την ασφάλεια των πληροφοριακών συστημάτων, αλλά και τα δικαιώματα των χρηστών ή όσων αναφέρονται στα πληροφοριακά

συστήματα, την εμπιστοσύνη στις ψηφιακές υπηρεσίες και τη συνολική λειτουργία οργανισμών. Επομένως, η νομοθετική προσέγγιση δεν περιορίζεται στην καταστολή της παράνομης συμπεριφοράς μετά την επίθεση, αλλά αποσκοπεί και στη δημιουργία υποχρεώσεων πρόληψης, διακυβέρνησης και ασφαλούς σχεδιασμού [21], [103].

Κεντρική θέση στο ευρωπαϊκό πλαίσιο κατέχει ο Γενικός Κανονισμός για την Προστασία Δεδομένων, γνωστός ως GDPR. Ο GDPR αποτελεί τον βασικό κανόνα προστασίας προσωπικών δεδομένων στην Ευρωπαϊκή Ένωση και εφαρμόζεται ανεξάρτητα από την τεχνολογία που χρησιμοποιείται για την επεξεργασία των δεδομένων. Η σημασία του GDPR είναι ιδιαίτερη, επειδή πολλές AI-assisted κυβερνοεπιθέσεις βασίζονται στη συλλογή, ανάλυση και εκμετάλλευση προσωπικών δεδομένων. Τέτοια δεδομένα μπορούν να προέρχονται από διαρροές, δημόσια προφίλ, κοινωνικά δίκτυα, παράνομες αγορές διαπιστευτηρίων ή προηγούμενα περιστατικά παραβίασης [157], [158].

Ο GDPR συνδέεται άμεσα με τις AI-enabled επιθέσεις σε τρία επίπεδα. Πρώτον, θέτει τις προϋποθέσεις νόμιμης επεξεργασίας προσωπικών δεδομένων, στοιχείο κρίσιμο σε περιβάλλοντα όπου δεδομένα μπορεί να χρησιμοποιηθούν για profiling, στοχευμένο phishing ή social engineering. Δεύτερον, επιβάλλει υποχρεώσεις ασφάλειας, οι οποίες συνδέονται με την ανάγκη προστασίας των δεδομένων από μη εξουσιοδοτημένη πρόσβαση, διαρροή ή κατάχρηση. Τρίτον, κατοχυρώνει δικαιώματα των υποκειμένων των δεδομένων (άτομα των οποίων τα προσωπικά δεδομένα χρησιμοποιούνται), τα οποία αποκτούν ιδιαίτερη σημασία όταν η επεξεργασία γίνεται με αλγοριθμικά ή αυτοματοποιημένα μέσα [157], [158], [159].

Η διάσταση του profiling είναι ιδιαίτερα κρίσιμη. Οι AI-assisted επιθέσεις μπορούν να αξιοποιούν δεδομένα για τη δημιουργία στοχευμένων μηνυμάτων, την προσαρμογή περιεχομένου στο προφίλ του θύματος ή την εκμετάλλευση επαγγελματικών και προσωπικών πληροφοριών. Οι κατευθυντήριες γραμμές για την αυτοματοποιημένη λήψη αποφάσεων και το profiling αναδεικνύουν τη σημασία της διαφάνειας, της ενημέρωσης και της προστασίας του ατόμου όταν χρησιμοποιούνται αλγοριθμικές μέθοδοι επεξεργασίας [159]. Αν και ο GDPR δεν είναι κανονισμός ειδικά για την κυβερνοασφάλεια ή την AI, λειτουργεί ως θεμελιώδες νομικό πλαίσιο προστασίας απέναντι στην κατάχρηση δεδομένων που μπορεί να τροφοδοτήσει AI-enabled επιθέσεις.

Παράλληλα, ο GDPR συνδέεται με την κυβερνοασφάλεια μέσω της υποχρέωσης προστασίας των προσωπικών δεδομένων. Η ασφάλεια της επεξεργασίας δεν αφορά μόνο τεχνικά μέτρα, αλλά και οργανωτικές διαδικασίες, αξιολόγηση κινδύνου και δυνατότητα ανταπόκρισης σε περιστατικά παραβίασης. Σε περιβάλλον όπου οι επιτιθέμενοι μπορούν να αξιοποιήσουν ΑΙ για ταχύτερη ανάλυση διαρροών, στοχευμένη εξαπάτηση ή εκμετάλλευση κλεμμένων διαπιστευτηρίων, η προστασία των δεδομένων αποκτά άμεση σχέση με την πρόληψη κυβερνοεπιθέσεων [157], [21].

Ο δεύτερος βασικός πυλώνας του ευρωπαϊκού ρυθμιστικού πλαισίου είναι ο AI Act. Ο Κανονισμός (ΕΕ) 2024/1689 αποτελεί οριζόντιο νομικό πλαίσιο για την τεχνητή νοημοσύνη και εισάγει προσέγγιση βασισμένη στον κίνδυνο. Η λογική αυτή σημαίνει ότι δεν αντιμετωπίζονται όλα τα συστήματα ΑΙ με τον ίδιο τρόπο. Αντίθετα, οι υποχρεώσεις διαφοροποιούνται ανάλογα με το επίπεδο κινδύνου που δημιουργεί η χρήση τους για την ασφάλεια, τα θεμελιώδη δικαιώματα και την κοινωνία [5], [156].

Η risk-based προσέγγιση του AI Act είναι ιδιαίτερα σημαντική για τη σύνδεση ΑΙ και κυβερνοασφάλειας. Ο κανονισμός περιλαμβάνει κατηγορίες κινδύνου, όπως μη αποδεκτό κίνδυνο, υψηλό κίνδυνο, περιορισμένο κίνδυνο και ελάχιστο ή μηδενικό κίνδυνο [156]. Για τα συστήματα υψηλού κινδύνου, προβλέπονται αυξημένες απαιτήσεις, όπως διαχείριση κινδύνου, ποιότητα δεδομένων, τεχνική τεκμηρίωση, διαφάνεια, καταγραφή λειτουργίας, ανθρώπινη εποπτεία και ακρίβεια [5]. Με αυτόν τον τρόπο, ο AI Act δεν λειτουργεί ως νόμος ειδικά για τις κυβερνοεπιθέσεις, αλλά ως πλαίσιο που επιδιώκει να διασφαλίσει ότι τα συστήματα ΑΙ αναπτύσσονται και χρησιμοποιούνται με τρόπο ελεγχόμενο, ασφαλή και υπεύθυνο.

Ο AI Act είναι κρίσιμος επειδή αντιμετωπίζει την τεχνητή νοημοσύνη όχι μόνο ως τεχνολογική καινοτομία, αλλά και ως πηγή κινδύνου. Η χρήση ΑΙ σε κυβερνοεπιθέσεις αναδεικνύει την ανάγκη για συστήματα που δεν είναι μόνο αποτελεσματικά, αλλά και ελέγξιμα, τεκμηριωμένα και ασφαλή έναντι κατάχρησης. Οι υποχρεώσεις τεχνικής τεκμηρίωσης, διαφάνειας και ανθρώπινης εποπτείας συνδέονται με την ανάγκη να αποτρέπεται η ανεξέλεγκτη ή αδιαφανής χρήση συστημάτων που μπορεί να επηρεάσουν άτομα, οργανισμούς και κρίσιμες υποδομές [5], [165], [166].

Ο AI Act αποκτά ιδιαίτερη σημασία και για την αμυντική χρήση της ΑΙ. Όταν οργανισμοί χρησιμοποιούν ΑΙ συστήματα για κυβερνοάμυνα, ανίχνευση απειλών ή αυτοματοποιημένη απόκριση, προκύπτουν ζητήματα ακρίβειας, αξιοπιστίας, ανθρώπινης εποπτείας και δυνατότητας

ελέγχου. Ένα αμυντικό AI σύστημα που λαμβάνει λανθασμένες αποφάσεις, παράγει false positives ή λειτουργεί χωρίς επαρκή διαφάνεια μπορεί να δημιουργήσει επιχειρησιακές και νομικές συνέπειες. Συνεπώς, ο AI Act συμβάλλει στη διαμόρφωση ενός πλαισίου όπου η χρήση AI στην κυβερνοασφάλεια πρέπει να συνοδεύεται από μηχανισμούς ελέγχου και υπευθυνότητας [5], [156], [165].

Τρίτος πυλώνας είναι η NIS2 Directive, η οποία αφορά την ασφάλεια δικτύων και πληροφοριακών συστημάτων. Η NIS2 θεσπίζει ένα ενιαίο νομικό πλαίσιο για την επίτευξη υψηλού επιπέδου κυβερνοασφάλειας στην Ευρωπαϊκή Ένωση, καλύπτοντας κρίσιμους τομείς και επιβάλλοντας υποχρεώσεις σε essential και important entities [160], [161]. Η σημασία της για τις AI-enabled κυβερνοεπιθέσεις έγκειται στο ότι οι επιθέσεις αυτές δεν στοχεύουν μόνο μεμονωμένα άτομα, αλλά μπορούν να επηρεάσουν οργανισμούς, δημόσιες υπηρεσίες, παρόχους κρίσιμων υποδομών και αλυσίδες εφοδιασμού.

Η NIS2 δίνει έμφαση στη διαχείριση κινδύνου, στην αναφορά περιστατικών, στην κυβερνοανθεκτικότητα, στη συνεργασία και στην ευθύνη των οργανισμών. Σε περιβάλλον όπου η AI μπορεί να αυξήσει την ταχύτητα και την κλίμακα των επιθέσεων, η ύπαρξη διαδικασιών risk management και incident reporting καθίσταται ουσιώδης. Οι οργανισμοί δεν αρκεί να αντιδρούν αποσπασματικά σε περιστατικά, αλλά οφείλουν να διαθέτουν συστηματικές πολιτικές πρόληψης, ανίχνευσης, αντιμετώπισης και αποκατάστασης [160], [161].

Η NIS2 συνδέεται άμεσα με το ερευνητικό ερώτημα σχετικά με την ισορροπία μεταξύ επίθεσης και άμυνας. Όταν οι επιτιθέμενοι αξιοποιούν AI για αυτοματοποίηση, αναγνώριση στόχων και ταχύτερη εκμετάλλευση ευπαθειών, οι οργανισμοί χρειάζονται υψηλότερο επίπεδο ετοιμότητας. Η οδηγία επιδιώκει να ενισχύσει αυτή την ετοιμότητα μέσω ελάχιστων απαιτήσεων κυβερνοασφάλειας, εθνικών στρατηγικών, συνεργασίας μεταξύ κρατών-μελών και μηχανισμών επιβολής [160], [161]. Έτσι, η NIS2 δεν ρυθμίζει την AI άμεσα, αλλά ενισχύει το οργανωτικό και θεσμικό υπόβαθρο που απαιτείται για την αντιμετώπιση AI-enabled απειλών.

Τέταρτος πυλώνας είναι το Cyber Resilience Act. Ο Κανονισμός (ΕΕ) 2024/2847 εισάγει οριζόντιες απαιτήσεις κυβερνοασφάλειας για προϊόντα με ψηφιακά στοιχεία, συμπεριλαμβανομένου λογισμικού και hardware [162]. Η σημασία του είναι ότι μεταφέρει την κυβερνοασφάλεια στο στάδιο του σχεδιασμού, της ανάπτυξης, της παραγωγής και της

συντήρησης των προϊόντων. Η λογική αυτή συνδέεται με την αρχή security-by-design, σύμφωνα με την οποία η ασφάλεια δεν πρέπει να προστίθεται εκ των υστέρων, αλλά να αποτελεί βασικό στοιχείο του ψηφιακού προϊόντος από την αρχή [162], [163], [164].

Η σύνδεση του Cyber Resilience Act με τις AI-enabled κυβερνοεπιθέσεις είναι ουσιαστική. Οι επιτιθέμενοι που αξιοποιούν AI μπορούν να αναζητούν ευπάθειες ταχύτερα, να αυτοματοποιούν reconnaissance ή να προσαρμόζουν επιθετικές μεθόδους σε διαφορετικά ψηφιακά προϊόντα. Όσο περισσότερες ευπάθειες υπάρχουν σε λογισμικό, συσκευές και ψηφιακά συστήματα, τόσο μεγαλύτερη είναι η επιφάνεια επίθεσης. Ο CRA επιχειρεί να περιορίσει αυτόν τον κίνδυνο επιβάλλοντας απαιτήσεις ασφάλειας, ενημερώσεων και υπεύθυνης διαχείρισης ευπαθειών σε προϊόντα με ψηφιακά στοιχεία [162], [163], [164].

Το Cyber Resilience Act βασίζεται στην πρόληψη, ενώ ο GDPR προστατεύει τα προσωπικά δεδομένα, ο AI Act ρυθμίζει την ανάπτυξη και χρήση AI συστημάτων και η NIS2 ενισχύει την οργανωτική κυβερνοασφάλεια, ο CRA στοχεύει στη μείωση των ευπαθειών στο ίδιο το ψηφιακό προϊόν. Με αυτόν τον τρόπο, το ευρωπαϊκό πλαίσιο δεν περιορίζεται στη ρύθμιση της συμπεριφοράς των οργανισμών, αλλά επεκτείνεται στην ασφάλεια των τεχνολογικών αντικειμένων που χρησιμοποιούνται στην ψηφιακή οικονομία [162], [164].

Είναι πλέον κατανοητό ότι η Ευρωπαϊκή Ένωση αντιμετωπίζει την AI και την κυβερνοασφάλεια μέσα από συμπληρωματικές ρυθμιστικές λογικές. Ο GDPR προστατεύει τα δεδομένα που συχνά χρησιμοποιούνται ως πρώτη ύλη για profiling, phishing και social engineering. Ο AI Act εισάγει κανόνες για ασφαλή, αξιόπιστη και υπεύθυνη χρήση της τεχνητής νοημοσύνης. Η NIS2 ενισχύει την κυβερνοανθεκτικότητα οργανισμών και κρίσιμων τομέων. Ο Cyber Resilience Act επιδιώκει ασφαλέστερα προϊόντα με ψηφιακά στοιχεία, ώστε να περιοριστούν οι ευπάθειες που μπορούν να αξιοποιηθούν σε κυβερνοεπιθέσεις [5], [157], [160], [162].

Η σημασία αυτής της πολυεπίπεδης προσέγγισης είναι ότι οι AI-enabled κυβερνοεπιθέσεις δεν μπορούν να αντιμετωπιστούν με ένα μόνο νομικό εργαλείο. Η κατάχρηση προσωπικών δεδομένων απαιτεί κανόνες προστασίας δεδομένων. Η χρήση και ανάπτυξη AI συστημάτων απαιτεί ρυθμιστικό πλαίσιο για ασφάλεια, διαφάνεια και εποπτεία. Η προστασία οργανισμών απαιτεί διαχείριση κινδύνων και αναφορά περιστατικών. Η μείωση τεχνικών ευπαθειών απαιτεί ασφαλή

ψηφιακά προϊόντα. Επομένως, η νομική αντιμετώπιση των κινδύνων της τεχνητής νοημοσύνης στην κυβερνοασφάλεια είναι εξαιρετικά σύνθετη [156], [161], [163], [156].

Παρά την πληρότητα του ευρωπαϊκού πλαισίου, υπάρχουν και πρακτικές προκλήσεις. Η τεχνολογική εξέλιξη της τεχνητής νοημοσύνης είναι ταχύτερη από τον ρυθμό με τον οποίο συνήθως προσαρμόζονται οι θεσμοί. Επιπλέον, οι AI-enabled επιθέσεις συχνά εκδηλώνονται διασυνοριακά, αξιοποιώντας πλατφόρμες, υποδομές και δεδομένα που δεν βρίσκονται πάντα εντός της ίδιας δικαιοδοσίας. Αυτό σημαίνει ότι η αποτελεσματικότητα των κανονισμών εξαρτάται όχι μόνο από το περιεχόμενό τους, αλλά και από την εφαρμογή τους, τη συνεργασία μεταξύ αρχών, τη συμμόρφωση των οργανισμών και τη δυνατότητα επιβολής σε διεθνές περιβάλλον [21], [103], [161].

Επιπλέον, το νομικό πλαίσιο δεν μπορεί να υποκαταστήσει την τεχνική και οργανωτική ετοιμότητα. Η ύπαρξη κανονισμών δεν αρκεί, εάν οι οργανισμοί δεν εφαρμόζουν ουσιαστικά μέτρα ασφάλειας, δεν εκπαιδεύουν προσωπικό, δεν αξιολογούν τους κινδύνους των συστημάτων τεχνητής νοημοσύνης και δεν διαθέτουν συγκεκριμένα πρωτόκολλα διαχείρισης περιστατικών. Το δίκαιο λειτουργεί ως πλαίσιο υποχρεώσεων και ευθύνης, αλλά η πραγματική κυβερνοανθεκτικότητα απαιτεί συνεχή τεχνική εφαρμογή, οργανωτική προσαρμογή και ενημέρωση απέναντι στο εξελισσόμενο threat landscape [21], [160], [164].

Η χρήση AI αυξάνει την ανάγκη για προστασία δεδομένων, διαφάνεια και αυξημένη ασφάλεια συστημάτων. Ο GDPR, ο AI Act, η NIS2 και ο Cyber Resilience Act δεν ρυθμίζουν το ίδιο αντικείμενο, αλλά συγκροτούν ένα ευρωπαϊκό ρυθμιστικό πλέγμα που επιδιώκει να περιορίσει τους κινδύνους τόσο στο επίπεδο των δεδομένων όσο και στο επίπεδο των συστημάτων, των οργανισμών και των ψηφιακών προϊόντων [5], [157], [160], [162].

Πίνακας 8.1. Βασικά ευρωπαϊκά κανονιστικά πλαίσια με συνάφεια προς την AI και την κυβερνοασφάλεια

Πλαίσιο	Νομική μορφή	Κύριο πεδίο	Συνάφεια με την εργασία
---------	--------------	-------------	-------------------------

GDPR - Κανονισμός (ΕΕ) 2016/679	Κανονισμός	Προστασία δεδομένων προσωπικού χαρακτήρα και ασφάλεια επεξεργασίας.	Συνδέεται με την προστασία προσωπικών δεδομένων, την ασφάλεια επεξεργασίας και τους κινδύνους από διαρροές ή κακόβουλη αξιοποίηση δεδομένων.
AI Act - Κανονισμός (ΕΕ) 2024/1689	Κανονισμός	Εναρμονισμένοι κανόνες για συστήματα ΑΙ, με προσέγγιση βασισμένη στον κίνδυνο.	Σχετίζεται με απαιτήσεις διακυβέρνησης, διαφάνειας, ασφάλειας και ανθρώπινης εποπτείας, ιδίως για κατηγορίες συστημάτων αυξημένου κινδύνου.
NIS2 - Οδηγία (ΕΕ) 2022/2555	Οδηγία	Μέτρα για υψηλό κοινό επίπεδο κυβερνοασφάλειας στην Ένωση.	Συνδέεται με διαχείριση κινδύνου, μέτρα κυβερνοασφάλειας και υποχρεώσεις αναφοράς περιστατικών για τους φορείς που καλύπτονται από την οδηγία.
Cyber Resilience Act - Κανονισμός (ΕΕ) 2024/2847	Κανονισμός	Οριζόντιες απαιτήσεις κυβερνοασφάλειας για προϊόντα με ψηφιακά στοιχεία.	Συνδέεται με την ασφάλεια από τον σχεδιασμό, τη διαχείριση ευπαθειών και την ενίσχυση της ανθεκτικότητας προϊόντων.

Τελικό συμπέρασμα αποτελεί ότι η αντιμετώπιση των AI-enabled κυβερνοεπιθέσεων απαιτεί συνδυασμό τεχνικής άμυνας και νομικής διακυβέρνησης. Η Ευρωπαϊκή Ένωση επιχειρεί να απαντήσει στους κινδύνους της τεχνητής νοημοσύνης και της κυβερνοασφάλειας μέσα από ένα πολυεπίπεδο πλαίσιο, στο οποίο η προστασία προσωπικών δεδομένων, η αξιόπιστη AI, η κυβερνοανθεκτικότητα οργανισμών και η ασφάλεια ψηφιακών προϊόντων λειτουργούν συμπληρωματικά. Η ρύθμιση δεν εξαλείφει τον κίνδυνο, αλλά δημιουργεί κανόνες, ευθύνες και μηχανισμούς πρόληψης που είναι αναγκαίοι για την ασφαλέστερη λειτουργία του ψηφιακού περιβάλλοντος στο σήμερα [156], [161], [163], [166].

8.3. Διακρατικές διαφορές και παγκόσμιες στρατηγικές πρόληψης

Η αντιμετώπιση των κυβερνοεπιθέσεων που αξιοποιούν τεχνητή νοημοσύνη δεν μπορεί να περιοριστεί στα όρια ενός κράτους ή ακόμη και μιας περιφερειακής έννομης τάξης, όπως για παράδειγμα η Ευρωπαϊκή Ένωση. Οι AI-enabled κυβερνοεπιθέσεις έχουν κατεξοχήν διασυνοριακό χαρακτήρα, καθώς ο επιτιθέμενος, το θύμα, οι ψηφιακές υποδομές, οι διακομιστές, τα εργαλεία, οι πλατφόρμες επικοινωνίας και τα αποδεικτικά στοιχεία μπορεί να βρίσκονται σε διαφορετικά κράτη και να υπάγονται σε διαφορετικές δικαιοδοσίες. Η διεθνής διάσταση του φαινομένου καθιστά την πρόληψη και την αντιμετώπισή του όχι μόνο τεχνικό ζήτημα, αλλά και νομικό, θεσμικό, γεωπολιτικό και επιχειρησιακό πρόβλημα [170], [22], [30].

Η τεχνητή νοημοσύνη εντείνει αυτή τη διασυνοριακή πολυπλοκότητα, επειδή αυξάνει την ταχύτητα, την κλίμακα και την προσαρμοστικότητα των επιθέσεων. Ένας δράστης μπορεί να χρησιμοποιεί AI εργαλεία που έχουν αναπτυχθεί σε μία χώρα, cloud υποδομές που φιλοξενούνται σε άλλη, παράνομα marketplaces που λειτουργούν διεθνώς και θύματα που βρίσκονται σε πολλές διαφορετικές έννομες τάξεις. Η κατανομή αυτή δυσχεραίνει την απόδοση ευθύνης, την άμεση διακοπή της επίθεσης και τη συλλογή ηλεκτρονικών αποδεικτικών στοιχείων [22], [21], [148].

Η διακρατική διάσταση των AI-enabled κυβερνοεπιθέσεων συνδέεται άμεσα με τις διαφορές μεταξύ εθνικών ρυθμιστικών προσεγγίσεων. Η Ευρωπαϊκή Ένωση ακολουθεί μια περισσότερο δεσμευτική και risk-based προσέγγιση, όπως αναλύθηκε στο προηγούμενο υποκεφάλαιο, μέσω του AI Act, του GDPR, της NIS2 και του Cyber Resilience Act. Αντίθετα, άλλες χώρες μπορεί να

στηρίζονται περισσότερο σε μη δεσμευτικά πλαίσια, τομεακές ρυθμίσεις, κατευθυντήριες αρχές ή πρότυπα διαχείρισης κινδύνου. Η διαφοροποίηση αυτή δημιουργεί δυσκολίες, διότι οι κυβερνοεπιθέσεις δεν ακολουθούν τα σύνορα των νομικών συστημάτων [173], [4], [177].

Στο επίπεδο της τεχνητής νοημοσύνης, το συμβούλιο της Ευρώπης έχει συμβάλλει στη διαμόρφωση διεθνούς νομικού πλαισίου μέσω της Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. Η σύμβαση αυτή αποτελεί σημαντικό βήμα, διότι επιχειρεί να συνδέσει την ανάπτυξη και χρήση ΑΙ συστημάτων με την προστασία των ανθρωπίνων δικαιωμάτων, της δημοκρατίας και του κράτους δικαίου [167], [168], [169].

Η σύμβαση του συμβουλίου της Ευρώπης δεν αφορά αποκλειστικά το κυβερνοέγκλημα, αλλά προσφέρει ευρύτερο θεσμικό υπόβαθρο για υπεύθυνη χρήση της τεχνητής νοημοσύνης. Η έμφαση στον κύκλο ζωής των ΑΙ συστημάτων, στη διαχείριση κινδύνων και στη συμβατότητα με θεμελιώδεις αξίες συνδέεται με την ανάγκη να αποτρέπεται η ανεξέλεγκτη ή κακόβουλη χρήση της ΑΙ. Στο πλαίσιο των κυβερνοεπιθέσεων, η σύμβαση συμβάλλει στη διαμόρφωση κοινής διεθνούς γλώσσας γύρω από την ευθύνη, την ασφάλεια και τη θεσμική εποπτεία της ΑΙ [167], [168].

Παράλληλα, οι OECD AI Principles αποτελούν σημαντικό παράδειγμα διεθνών αρχών ήπιας ρύθμισης. Οι αρχές αυτές προωθούν την αξιόπιστη ΑΙ, τον σεβασμό των ανθρωπίνων δικαιωμάτων, τη δημοκρατία, τη διαφάνεια, τη λογοδοσία και τη διαχείριση κινδύνων [173], [174]. Παρότι δεν έχουν την ίδια δεσμευτική ισχύ με έναν κανονισμό ή μία σύμβαση, λειτουργούν ως κοινό πλαίσιο αναφοράς για κράτη, οργανισμούς και επιχειρήσεις. Η αξία τους βρίσκεται στη δυνατότητα ευθυγράμμισης διαφορετικών εθνικών προσεγγίσεων γύρω από βασικές αρχές υπεύθυνης ανάπτυξης και χρήσης ΑΙ.

Η προσέγγιση των Ηνωμένων Πολιτειών, όπως αποτυπώνεται στο NIST AI Risk Management Framework, δίνει έμφαση στη διαχείριση κινδύνων μέσω ενός μη δεσμευτικού αλλά επιχειρησιακά χρήσιμου πλαισίου. Το AI RMF οργανώνει τη διαχείριση κινδύνων γύρω από λειτουργίες όπως govern, map, measure και manage, παρέχοντας σε οργανισμούς μεθοδολογία για την αναγνώριση, αξιολόγηση και μείωση κινδύνων που συνδέονται με ΑΙ συστήματα [4]. Η

σημασία του για τις AI-enabled κυβερνοεπιθέσεις είναι ότι προσφέρει τρόπο συστηματικής διαχείρισης κινδύνων, χωρίς να στηρίζεται αποκλειστικά σε δεσμευτικούς νομικούς κανόνες.

Η σύγκριση μεταξύ ευρωπαϊκής και αμερικανικής προσέγγισης δείχνει ότι η διεθνής ρύθμιση της AI δεν είναι ενιαία. Η ΕΕ δίνει μεγαλύτερη έμφαση σε δεσμευτικούς κανόνες και risk-based κατηγοριοποίηση, ενώ το NIST AI RMF λειτουργεί ως πλαίσιο καθοδήγησης και διαχείρισης κινδύνου. Παρά τις διαφορές τους, και οι δύο προσεγγίσεις αναγνωρίζουν ότι η AI απαιτεί μηχανισμούς αξιοπιστίας, ασφάλειας και διαφάνειας [173], [4], [177]. Η ύπαρξη τέτοιων κοινών αρχών είναι κρίσιμη για τη διεθνή συνεργασία, επειδή οι κυβερνοεπιθέσεις συχνά επηρεάζουν ταυτόχρονα οργανισμούς που υπάγονται σε διαφορετικά ρυθμιστικά καθεστώτα.

Σε επίπεδο παγκόσμιου συντονισμού, ιδιαίτερη σημασία έχει και το G7 Hiroshima AI Process. Η πρωτοβουλία αυτή επιδιώκει να διαμορφώσει κοινές αρχές και κώδικα συμπεριφοράς για προηγμένα AI συστήματα, ενισχύοντας τη διεθνή συζήτηση γύρω από την ασφάλεια, τη διαφάνεια και την υπεύθυνη ανάπτυξη AI συστημάτων [175], [176], [177]. Η σημασία του Hiroshima Process βρίσκεται στο ότι επιχειρεί να γεφυρώσει διαφορετικές ρυθμιστικές προσεγγίσεις και να δημιουργήσει κοινή βάση για κράτη και εταιρείες που αναπτύσσουν ή χρησιμοποιούν προηγμένα συστήματα AI.

Η διεθνής συνεργασία δεν αφορά μόνο το AI governance, αλλά και την ποινική αντιμετώπιση του κυβερνοεγκλήματος. Στο πεδίο αυτό, η σύμβαση του ΟΗΕ κατά του κυβερνοεγκλήματος αποτελεί κεντρική εξέλιξη. Η σύμβαση υιοθετήθηκε από τη γενική συνέλευση του ΟΗΕ στις 24 Δεκεμβρίου 2024 και τέθηκε για υπογραφή στις 25 Οκτωβρίου 2025 στο Ανόι του Βιετνάμ [170], [171]. Η σημασία της έγκειται στην προσπάθεια δημιουργίας παγκόσμιου πλαισίου για την αντιμετώπιση εγκλημάτων που τελούνται μέσω πληροφοριακών συστημάτων, καθώς και για τη διεθνή συνεργασία και τα ηλεκτρονικά αποδεικτικά στοιχεία.

Η σύμβαση του ΟΗΕ είναι ιδιαίτερα σημαντική για τις AI-enabled κυβερνοεπιθέσεις, διότι οι επιθέσεις αυτές συχνά απαιτούν διασυνοριακή έρευνα. Η συλλογή logs, δεδομένων λογαριασμών, πληροφοριών υποδομής, οικονομικών στοιχείων και στοιχείων επικοινωνίας απαιτεί συνεργασία πολλών κρατών και ιδιωτικών φορέων. Η ύπαρξη διεθνούς σύμβασης δεν εξαλείφει τις πρακτικές δυσκολίες, αλλά μπορεί να δημιουργήσει κοινά πρωτόκολλα άμυνας και θεσμικά κανάλια συνεργασίας [170], [172], [147].

Πριν από τη νέα Σύμβαση του ΟΗΕ, σημαντικό ρόλο είχε ήδη η Budapest Convention on Cybercrime του συμβουλίου της Ευρώπης, η οποία αποτελεί διεθνές πλαίσιο συνεργασίας για την αντιμετώπιση του κυβερνοεγκλήματος [178]. Η σύμβαση αυτή και τα πρόσθετα πρωτόκολλά της έχουν ιδιαίτερη σημασία για τη διασυνοριακή συλλογή ψηφιακών αποδεικτικών στοιχείων, καθώς ενισχύουν την πρακτική συνεργασία μεταξύ αρχών και, σε ορισμένες περιπτώσεις, μεταξύ κρατών και ιδιωτικών φορέων [179], [180].

Το ζήτημα των ηλεκτρονικών αποδεικτικών στοιχείων είναι ένα από τα πιο κρίσιμα στην αντιμετώπιση AI-enabled κυβερνοεπιθέσεων. Τα στοιχεία μπορεί να βρίσκονται σε cloud υποδομές, πλατφόρμες επικοινωνίας, παρόχους υπηρεσιών, χρηματοπιστωτικά συστήματα ή συσκευές σε διαφορετικές χώρες. Η καθυστέρηση στη συλλογή τους μπορεί να οδηγήσει σε απώλεια κρίσιμων δεδομένων, ενώ οι διαφορές στα νομικά συστήματα μπορεί να δυσχεράνουν την αποδοχή ή αξιοποίησή τους σε ποινικές διαδικασίες [179], [180], [148]. Επομένως, η διεθνής συνεργασία δεν είναι απλώς πολιτική επιλογή, αλλά πρακτική προϋπόθεση αποτελεσματικής διερεύνησης.

Σημαντική πρόκληση αποτελεί και η απόδοση ευθύνης. Στις AI-enabled κυβερνοεπιθέσεις, οι δράστες μπορούν να χρησιμοποιούν botnets, VPNs, cloud υπηρεσίες, dark web αγορές και εργαλεία παραγωγής περιεχομένου. Η χρήση αυτών των μέσων δυσκολεύει τον εντοπισμό του πραγματικού δράστη και τη σύνδεσή του με συγκεκριμένη ενέργεια. Η διεθνής συνεργασία είναι αναγκαία, διότι κανένα κράτος δεν έχει πλήρη εικόνα όλων των τεχνικών, οικονομικών και κοινωνικών στοιχείων μιας διασυνοριακής επίθεσης [22], [30], [139].

Οι δυνατότητες αντιμετώπισης δεν είναι ίδιες σε όλες τις χώρες. Ορισμένα κράτη διαθέτουν ανεπτυγμένες δομές κυβερνοασφάλειας, εξειδικευμένες αρχές και ισχυρή συνεργασία με τον ιδιωτικό τομέα, ενώ άλλα αντιμετωπίζουν περιορισμούς σε τεχνικούς πόρους, ψηφιακή εγκληματολογία, εκπαίδευση και πρόσβαση σε threat intelligence. Οι περιφερειακές αναφορές της INTERPOL αναδεικνύουν τη σημασία της ενίσχυσης ικανοτήτων και της διεθνούς συνεργασίας, ιδίως σε περιοχές όπου η τεχνολογική ετοιμότητα παραμένει υπανάπτυκτη [138], [139]. Η ανισότητα αυτή μεταξύ κρατών μπορεί να αξιοποιηθεί από εγκληματικές ομάδες, οι οποίες επιλέγουν υποδομές ή διαδρομές που δυσχεραίνουν την επιβολή του νόμου.

Οι παγκόσμιες στρατηγικές πρόληψης πρέπει να στηρίζονται στην ανταλλαγή πληροφοριών. Οι AI-enabled απειλές εξελίσσονται γρήγορα, επομένως η έγκαιρη ανταλλαγή threat intelligence μεταξύ κρατών, οργανισμών, παρόχων τεχνολογίας και ομάδων incident response είναι κρίσιμη. Η ENISA αναδεικνύει τη σημασία της ανταλλαγής πληροφοριών, των βέλτιστων πρακτικών και των στρατηγικών απόκρισης σε περιστατικά, ενώ το FIRST λειτουργεί ως διεθνές δίκτυο συνεργασίας ομάδων αντιμετώπισης περιστατικών ασφάλειας [99], [181]. Η πρόληψη δεν εξαρτάται μόνο από την ύπαρξη κανόνων, αλλά και από την ταχύτητα με την οποία διαμοιράζονται οι ενδείξεις απειλής.

Η ανταλλαγή πληροφοριών αποκτά ιδιαίτερη αξία όταν οι επιθέσεις είναι μαζικές ή επαναλαμβανόμενες. Μία AI-assisted phishing καμπάνια ή ένα αυτοματοποιημένο exploit campaign μπορεί να επηρεάσει οργανισμούς σε πολλές χώρες. Εάν οι πληροφορίες για την επίθεση παραμείνουν απομονωμένες σε ένα κράτος ή έναν οργανισμό, οι υπόλοιποι στόχοι παραμένουν εκτεθειμένοι. Αντίθετα, η κοινοποίηση τεχνικών δεικτών, μεθόδων επίθεσης και μοτίβων συμπεριφοράς επιτρέπει ταχύτερη πρόληψη, ανίχνευση και απόκριση [21], [99], [181].

Ο δημόσιος και ιδιωτικός συντονισμός είναι επίσης απαραίτητος. Μεγάλο μέρος των κρίσιμων πληροφοριών βρίσκεται σε ιδιωτικούς φορείς, όπως πάροχοι cloud, εταιρείες κυβερνοασφάλειας, τηλεπικοινωνιακοί πάροχοι, πλατφόρμες κοινωνικής δικτύωσης, χρηματοπιστωτικά ιδρύματα και εταιρείες ανάπτυξης AI. Οι αρχές χρειάζονται συνεργασία με αυτούς τους φορείς για να εντοπίζουν υποδομές, να συλλέγουν στοιχεία, να διακόπτουν κακόβουλες δραστηριότητες και να αναλύουν διασυννοριακές απειλές [22], [148]. Χωρίς τέτοιο συντονισμό, η νομική αντίδραση συχνά υπολείπεται της ταχύτητας της τεχνικής εξέλιξης.

Παρά τη σημασία της διεθνούς συνεργασίας, υπάρχουν σοβαρές προκλήσεις συντονισμού. Τα κράτη έχουν διαφορετικά νομικά συστήματα, διαφορετικές αντιλήψεις για την ιδιωτικότητα, διαφορετικά επίπεδα τεχνολογικής ετοιμότητας και διαφορετικές πολιτικές προτεραιότητες. Επιπλέον, η συνεργασία για ηλεκτρονικά αποδεικτικά στοιχεία πρέπει να ισορροπεί ανάμεσα στην αποτελεσματικότητα της έρευνας και στην προστασία θεμελιωδών δικαιωμάτων. Εάν η διεθνής συνεργασία δεν συνοδεύεται από εγγυήσεις νομιμότητας, αναλογικότητας και λογοδοσίας, μπορεί να δημιουργήσει κινδύνους υπερβολικής επιτήρησης ή κατάχρησης δεδομένων [167], [170], [148].

Η πρόκληση αυτή είναι εμφανής και στο πεδίο της AI governance. Η παγκόσμια ανάγκη για κοινά πρότυπα δεν σημαίνει ότι όλα τα κράτη συμφωνούν στον ίδιο βαθμό ρύθμισης ή στον ίδιο τρόπο εφαρμογής. Ορισμένες έννομες τάξεις δίνουν προτεραιότητα στην καινοτομία, άλλες στην προστασία δικαιωμάτων, άλλες στην εθνική ασφάλεια και άλλες στην οικονομική ανταγωνιστικότητα. Η διαμόρφωση διεθνών αρχών, όπως οι OECD AI Principles, το NIST AI RMF, η σύμβαση του συμβουλίου της Ευρώπης και το Hiroshima AI Process, συμβάλλουν στη δημιουργία κοινής γλώσσας, αλλά δεν εξαλείφουν πλήρως τις διαφορές [173], [4], [175], [177].

Οι παγκόσμιες στρατηγικές πρόληψης πρέπει επομένως να είναι πολυεπίπεδες. Πρώτον, απαιτούνται κοινές αρχές υπεύθυνης AI, ώστε τα συστήματα να αναπτύσσονται ταυτόχρονα. Δεύτερον, χρειάζονται τεχνικά πρότυπα και πρακτικές διαχείρισης κινδύνου, ώστε οργανισμοί και κράτη να μπορούν να αξιολογούν και να περιορίζουν τις επιπτώσεις της AI σε κυβερνοαπειλές. Τρίτον, απαιτείται επιχειρησιακή συνεργασία για ανταλλαγή threat intelligence, incident response και συλλογή αποδεικτικών στοιχείων [168], [4], [99], [181].

Επιπλέον, η πρόληψη απαιτεί ενίσχυση ικανοτήτων σε παγκόσμιο επίπεδο. Εάν ορισμένα κράτη δεν διαθέτουν επαρκείς τεχνικές υποδομές, εξειδικευμένο προσωπικό ή μηχανισμούς διερεύνησης κυβερνοεγκλημάτων, δημιουργούνται αδύναμοι κρίκοι στο διεθνές σύστημα ασφάλειας. Οι εγκληματικές ομάδες μπορούν να εκμεταλλευτούν αυτές τις ανισότητες, χρησιμοποιώντας χώρες με περιορισμένες δυνατότητες επιβολής ως σημεία υποδομής ή απόκρυψης. Η εκπαίδευση, η τεχνική βοήθεια και η ενίσχυση της ψηφιακής εγκληματολογίας αποτελούν βασικές προϋποθέσεις παγκόσμιας πρόληψης [138], [139], [172].

Η διακρατική πρόληψη πρέπει επίσης να λαμβάνει υπόψη την προστασία ανθρωπίνων δικαιωμάτων. Η αντιμετώπιση AI-enabled κυβερνοεπιθέσεων δεν πρέπει να οδηγήσει σε ανεξέλεγκτη παρακολούθηση, υπερβολική πρόσβαση σε δεδομένα ή περιορισμό νόμιμης τεχνολογικής έρευνας. Τα διεθνή πλαίσια για την AI και το κυβερνοέγκλημα οφείλουν να εξισορροπούν την ανάγκη ασφάλειας με την προστασία της ιδιωτικότητας και της ελευθερίας έκφρασης [167], [168], [170]. Η ισορροπία αυτή είναι κρίσιμη, διότι η απάντηση στην κατάχρηση της AI δεν πρέπει να δημιουργεί νέους θεσμικούς κινδύνους.

Συνολικά, οι AI-enabled κυβερνοεπιθέσεις καθιστούν σαφές ότι η κυβερνοασφάλεια δεν μπορεί να αντιμετωπιστεί ως αποκλειστικά εθνικό ζήτημα. Η παγκόσμια αντιμετώπιση απαιτεί σύγκλιση

αρχών, θεσμική συνεργασία, κοινές διαδικασίες, τεχνικά πρότυπα και διαρκή επικοινωνία μεταξύ κρατών, διεθνών οργανισμών, αρχών επιβολής του νόμου και ιδιωτικού τομέα. Η διεθνής ρύθμιση της ΑΙ και η διεθνής συνεργασία κατά του κυβερνοεγκλήματος δεν αποτελούν δύο ξεχωριστά πεδία, αλλά συμπληρωματικούς μηχανισμούς για την αντιμετώπιση ενός κοινού προβλήματος [167], [170], [173], [178].

Η πρόληψη των AI-enabled κυβερνοεπιθέσεων απαιτεί παγκόσμια στρατηγική και όχι αποσπασματικές εθνικές αντιδράσεις. Η ύπαρξη διεθνών αρχών για αξιόπιστη ΑΙ, νομικών συμβάσεων για το κυβερνοέγκλημα, μηχανισμών συλλογής ηλεκτρονικών αποδεικτικών στοιχείων, δικτύων ανταλλαγής πληροφοριών και συνεργασίας δημόσιου και ιδιωτικού τομέα αποτελεί αναγκαία προϋπόθεση για την αποτελεσματική άμυνα. Ωστόσο, η συνεργασία αυτή πρέπει να συνοδεύεται από σεβασμό στα ανθρώπινα δικαιώματα, διαφάνεια και λογοδοσία, ώστε η αντιμετώπιση της τεχνολογικής απειλής να μη μετατραπεί σε νέα πηγή θεσμικού κινδύνου.

9. ΟΙΚΟΝΟΜΙΚΕΣ ΚΑΙ ΚΟΙΝΩΝΙΚΕΣ ΕΠΙΠΤΩΣΕΙΣ ΤΩΝ ΚΥΒΕΡΝΟΕΠΙΘΕΣΕΩΝ ΜΕ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

9.1. Οικονομικές απώλειες από AI-driven κυβερνοεπιθέσεις

Οι οικονομικές απώλειες από κυβερνοεπιθέσεις αποτελούν μία από τις σημαντικότερες συνέπειες της σύγχρονης ψηφιακής απειλής. Στην περίπτωση των AI-driven κυβερνοεπιθέσεων, το οικονομικό ζήτημα αποκτά ακόμη μεγαλύτερη σημασία, καθώς η τεχνητή νοημοσύνη μπορεί να αυξήσει την ταχύτητα, την κλίμακα, την πειστικότητα και την αποδοτικότητα των επιθετικών ενεργειών. Η οικονομική ζημία δεν περιορίζεται στην άμεση απώλεια χρημάτων ή δεδομένων, αλλά περιλαμβάνει κόστος αποκατάστασης, διακοπή λειτουργίας, νομικές υποχρεώσεις, απώλεια εμπιστοσύνης, φθορά φήμης, αύξηση ασφαλιστρών και ανάγκη πρόσθετων επενδύσεων σε μηχανισμούς κυβερνοασφάλειας [108], [27], [188].

Η ανάλυση των οικονομικών απωλειών πρέπει να γίνει με προσοχή, διότι δεν είναι επιστημονικά ορθό να αποδοθούν όλες οι σύγχρονες οικονομικές απώλειες του κυβερνοεγκλήματος αποκλειστικά στην τεχνητή νοημοσύνη. Οι περισσότερες κυβερνοεπιθέσεις εξακολουθούν να συνδυάζουν παραδοσιακές τεχνικές, ανθρώπινη παρέμβαση, social engineering, εκμετάλλευση ευπαθειών και οργανωμένες εγκληματικές υποδομές. Ωστόσο, η τεχνητή νοημοσύνη λειτουργεί ως επιταχυντής και οικονομικός πολλαπλασιαστής κινδύνου, διότι μειώνει το κόστος προετοιμασίας για τον επιτιθέμενο και αυξάνει το πιθανό κόστος αντιμετώπισης για το θύμα [32], [43], [189].

Το οικονομικό πλαίσιο μέσα στο οποίο εξετάζονται οι AI-driven κυβερνοεπιθέσεις είναι ήδη ιδιαίτερα βαρύ. Η IBM αναφέρει ότι το παγκόσμιο μέσο κόστος μιας παραβίασης δεδομένων το 2025 ανήλθε περίπου στα 4,44 εκατομμύρια δολάρια, γεγονός που αποτυπώνει το υψηλό οικονομικό βάρος των περιστατικών ασφαλείας για οργανισμούς και επιχειρήσεις [108]. Αντίστοιχα, το FBI IC3 κατέγραψε για το 2025 περισσότερες από ένα εκατομμύριο καταγγελίες και συνολικές αναφερόμενες απώλειες που ξεπέρασαν τα 20 δισεκατομμύρια δολάρια, ενώ οι καταγγελίες με αναφορά σε AI-related στοιχεία συνδέθηκαν με προσαρμοσμένες απώλειες άνω των 893 εκατομμυρίων δολαρίων [182]. Τα στοιχεία αυτά δεν αποδεικνύουν ότι κάθε ζημία είναι αποτέλεσμα τεχνητής νοημοσύνης, αλλά δείχνουν το οικονομικό περιβάλλον μέσα στο οποίο αυτή, ενισχύει ήδη υπάρχουσες απειλές.

Οι άμεσες οικονομικές απώλειες περιλαμβάνουν την κλοπή χρημάτων, την απώλεια δεδομένων, τις πληρωμές λύτρων σε ransomware, τις απάτες πληρωμών, τις επιθέσεις Business Email Compromise και τις οικονομικές απώλειες από credential theft. Σε αυτές τις περιπτώσεις, η ζημία είναι συχνά μετρήσιμη, καθώς συνδέεται με συγκεκριμένες χρηματικές απώλειες, μεταφορές κεφαλαίων, απαιτήσεις λύτρων ή κόστος ανάκτησης συστημάτων. Το IC3 καταγράφει ιδιαίτερα υψηλές απώλειες από κατηγορίες όπως investment fraud, Business Email Compromise, tech support scams, personal data breaches και phishing/spoofing, γεγονός που δείχνει ότι η οικονομική διάσταση του κυβερνοεγκλήματος είναι πολυμορφική και δεν περιορίζεται σε έναν μόνο τύπο επίθεσης [182].

Οι έμμεσες οικονομικές απώλειες είναι δυσκολότερο να υπολογιστούν, αλλά μπορεί να είναι και περισσότερο σημαντικές. Περιλαμβάνουν το κόστος τεχνικής αποκατάστασης, τη διακοπή λειτουργίας, την απώλεια παραγωγικότητας, την ανάγκη πρόσληψης εξωτερικών συμβούλων, τις νομικές δαπάνες, τις αποζημιώσεις, τα πρόστιμα, την απώλεια πελατών και τη μακροχρόνια φθορά της φήμης ενός οργανισμού. Η ENISA επισημαίνει ότι οι οικονομικές και λειτουργικές επιπτώσεις και οι επιπτώσεις στη φήμη ενός οργανισμού, είναι δύσκολο να εκτιμηθούν πλήρως, καθώς δεν αποτυπώνονται πάντα σε άμεσες χρηματικές καταβολές [188]. Αυτό σημαίνει ότι το πραγματικό κόστος μιας κυβερνοεπίθεσης μπορεί να υπερβαίνει σημαντικά το εμφανές αρχικό ποσό.

Η τεχνητή νοημοσύνη επηρεάζει την οικονομία των επιθέσεων κυρίως επειδή μειώνει το κόστος παραγωγής επιθετικού υλικού. Η δημιουργία πειστικών phishing μηνυμάτων, η μετάφραση περιεχομένου, η γλωσσική προσαρμογή, η παραγωγή πολλαπλών παραλλαγών και η εξατομίκευση επικοινωνίας μπορούν να γίνουν ταχύτερα και φθηνότερα με τη χρήση generative AI. Η Microsoft αναφέρει ότι τα AI-automated phishing emails εμφάνισαν σημαντικά υψηλότερο click-through rate από τις τυπικές προσπάθειες phishing, ενώ η αυτοματοποίηση μπορεί να αυξήσει σημαντικά την οικονομική αποδοτικότητα τέτοιων επιθέσεων μέσω της κλιμάκωσης στοχευμένων εκστρατειών [32]. Το στοιχείο αυτό είναι κρίσιμο, διότι δείχνει ότι η AI δεν αυξάνει μόνο την τεχνική δυνατότητα του επιτιθέμενου, αλλά και την οικονομική απόδοση της επίθεσης.

Οι επιθέσεις Business Email Compromise έχουν σημαντική οικονομική διάσταση. Η AI μπορεί να ενισχύσει αυτές τις επιθέσεις μέσω παραγωγής επίσημου ύφους, μίμησης οργανωτικής επικοινωνίας και χρήσης voice cloning για πιο πειστικές εντολές πληρωμής. Το FBI IC3 αναφέρει

απώλειες από BEC scams με AI involvement, γεγονός που δείχνει ότι η AI έχει αρχίσει να συνδέεται με οικονομικά μετρήσιμες μορφές εταιρικής εξαπάτησης [182].

Το deepfake fraud αποτελεί μία ακόμη μορφή οικονομικού κινδύνου. Η Pindrop συνδέει τη generative AI με την αύξηση φωνητικής απάτης και αναδεικνύει τον κίνδυνο οικονομικών απωλειών από synthetic voice attacks [65]. Αντίστοιχα, η Resemble AI καταγράφει περιστατικά deepfake fraud και reported fraud losses, υπογραμμίζοντας ότι η οικονομική ζημία από συνθετικό περιεχόμενο δεν είναι πλέον θεωρητικό σενάριο, αλλά αναδυόμενη πρακτική απειλή [184].

Η σημασία των deepfakes είναι ότι μεταβάλλουν τη σχέση μεταξύ εμπιστοσύνης και οικονομικής συναλλαγής. Σε πολλά επιχειρησιακά περιβάλλοντα, οι συναλλαγές βασίζονται σε φωνητικές εντολές, τηλεδιασκέψεις, emails ή γρήγορες εγκρίσεις από στελέχη. Όταν η AI επιτρέπει την πειστική μίμηση αυτών των μορφών επικοινωνίας, αυξάνεται ο κίνδυνος λανθασμένων πληρωμών ή παραπλανητικών εγκρίσεων. Η οικονομική ζημία προκύπτει όχι μόνο από την τεχνική παραβίαση ενός συστήματος, αλλά και από τη χειραγώγηση των μηχανισμών εμπιστοσύνης που στηρίζουν την επιχειρηματική λειτουργία [65], [184], [51].

Το ransomware παραμένει μία από τις πιο σημαντικές πηγές οικονομικού κόστους. Η Verizon αναφέρει ότι το ransomware εμφανίστηκε σε σημαντικό ποσοστό των breaches που εξέτασε το 2025, ενώ το median ransomware payment ήταν περίπου 115.000 δολάρια [23]. Παράλληλα, η chainalysis αναφέρει ότι τα on-chain ransomware payments το 2025 ανήλθαν περίπου στα 820 εκατομμύρια δολάρια, παρά τη μείωση σε σχέση με το προηγούμενο έτος, ενώ τα claimed attacks αυξήθηκαν [133]. Η διάκριση αυτή είναι σημαντική καθώς γίνεται κατανοητό πως οι πληρωμές λύτρων μπορεί να μειώνονται, αλλά ο αριθμός των επιθέσεων και το κόστος αποκατάστασης μπορούν να παραμένουν υψηλά.

Η οικονομική ζημία του ransomware δεν ταυτίζεται με το ποσό των λύτρων. Ακόμη και όταν ένας οργανισμός δεν πληρώνει, μπορεί να υποστεί διακοπή λειτουργίας, απώλεια δεδομένων, ανάγκη επαναφοράς συστημάτων, νομικές επιπτώσεις και φθορά στη φήμη του. Η Allianz αναφέρει ότι το ransomware αντιπροσώπευσε σημαντικό μέρος της αξίας μεγάλων cyber claims, ενώ η συμμετοχή της κλοπής δεδομένων στις μεγάλες αξιώσεις αυξήθηκε [187]. Επομένως, το πραγματικό κόστος ενός ransomware περιστατικού πρέπει να αξιολογείται ως συνδυασμός λύτρων, διακοπής εργασιών, τεχνικής αποκατάστασης και δευτερογενών ζημιών.

Η τεχνητή νοημοσύνη μπορεί να επιδεινώσει το οικονομικό βάρος του ransomware με έμμεσο τρόπο. Δεν είναι απαραίτητο να δημιουργεί αυτόνομα ransomware από την αρχή για να αυξήσει τον κίνδυνο. Μπορεί να υποστηρίξει reconnaissance, victim profiling, social engineering, παραγωγή παραλλαγών phishing, ανάλυση αδύναμων σημείων και ταχύτερη πραγματοποίηση επιθετικών ενεργειών. Σε συνδυασμό με ransomware-as-a-service, initial access brokers και underground υπηρεσίες, η ΑΙ μπορεί να μειώσει τον χρόνο και το κόστος προετοιμασίας των επιτιθέμενων [43], [33], [129].

Το cybercrime-as-a-service ενισχύει την οικονομική διάσταση των AI-driven απειλών. Όταν οι επιτιθέμενοι μπορούν να αγοράζουν πρόσβαση, να νοικιάζουν εργαλεία, να χρησιμοποιούν έτοιμα phishing kits ή να αξιοποιούν υπηρεσίες αυτοματοποίησης, το κόστος εισόδου στο κυβερνοέγκλημα μειώνεται. Η Group-IB περιγράφει το ransomware ως σύνθετη underground βιομηχανία με ransomware-as-a-service, dedicated leak sites και initial access brokers [129]. Η Europol, αντίστοιχα, αναδεικνύει την οικονομική εκμετάλλευση κλεμμένων δεδομένων και την οργανωμένη λειτουργία του κυβερνοεγκλήματος [22]. Σε αυτό το πλαίσιο, η ΑΙ λειτουργεί ως πρόσθετος παράγοντας αποδοτικότητας μέσα σε ήδη οργανωμένες παράνομες αγορές.

Η διαφορά μεταξύ κόστους για τον επιτιθέμενο και κόστους για το θύμα είναι κεντρική. Για τον δράστη, η ΑΙ μπορεί να μειώσει το κόστος παραγωγής περιεχομένου, να περιορίσει την ανάγκη εξειδικευμένης γνώσης, να επιταχύνει την αναγνώριση στόχων και να επιτρέψει μαζική επανάληψη. Για το θύμα, όμως, η ίδια τεχνολογία μπορεί να αυξήσει το κόστος ανίχνευσης, απόκρισης, εκπαίδευσης προσωπικού, τεχνικής θωράκισης και αποκατάστασης. Η ασυμμετρία αυτή σημαίνει ότι μια φθηνότερη επίθεση για τον δράστη μπορεί να οδηγήσει σε δυσανάλογα υψηλό οικονομικό βάρος για τον οργανισμό που πλήττεται [108], [32], [43].

Οι απάτες πληρωμών και οι οικονομικές απάτες αποτελούν ακόμη ένα πεδίο όπου η ΑΙ μπορεί να αυξήσει τον κίνδυνο. Η European Payments Council συνδέει την ΑΙ με την αυτοματοποίηση και εξατομίκευση fraud και phishing επιθέσεων, ενώ αναδεικνύει κινδύνους που σχετίζονται με payment fraud, credential attacks και social engineering [51]. Σε επίπεδο καταναλωτών, οι απώλειες από investment scams, payment redirection, romance scams και phishing δείχνουν ότι η οικονομική ζημία δεν αφορά μόνο μεγάλους οργανισμούς, αλλά και φυσικά πρόσωπα [185].

Η ΑΙ μπορεί επίσης να επηρεάσει ασφαλιστικές και χρηματοπιστωτικές αγορές. Η δυνατότητα παραγωγής συνθετικών εικόνων, βίντεο ή ψευδών αποδεικτικών στοιχείων μπορεί να χρησιμοποιηθεί για insurance fraud, όπως δείχνει η σχετική ανάλυση για vehicle insurance fraud fueled by generative AI [191]. Αντίστοιχα, σε χρηματοπιστωτικές συναλλαγές, AI-enhanced phishing, deepfakes και AI-driven malware μπορούν να αυξήσουν τον κίνδυνο απάτης, να μειώσουν τα εμπόδια για τους δράστες και να πιέσουν τις επιχειρήσεις να επενδύσουν σε ισχυρότερους μηχανισμούς ταυτοποίησης, παρακολούθησης και ανίχνευσης [190].

Η οικονομική επίπτωση των AI-driven κυβερνοεπιθέσεων πρέπει να εξεταστεί και σε επίπεδο συστημικού κινδύνου. Το World Economic Forum συνδέει το cyber-enabled fraud με την οικονομική ανθεκτικότητα και αναδεικνύει ότι η κυβερνοαπάτη, η ΑΙ και η γεωπολιτική αστάθεια μπορούν να οδηγήσουν σε ευρύτερο επιχειρησιακό και οικονομικό κίνδυνο [186]. Αυτό σημαίνει ότι οι κυβερνοεπιθέσεις δεν επηρεάζουν μόνο τον άμεσο οργανισμό θύμα, αλλά μπορούν να δημιουργήσουν αλυσιδωτές συνέπειες σε προμηθευτές, πελάτες, αγορές και κρίσιμες υπηρεσίες.

Η ανάγκη επενδύσεων σε κυβερνοανθεκτικότητα αποτελεί επίσης οικονομική συνέπεια. Καθώς οι επιθέσεις γίνονται ταχύτερες και πιο πειστικές, οι οργανισμοί υποχρεούνται να αυξήσουν τις δαπάνες τους σε ανίχνευση, απόκριση, εκπαίδευση προσωπικού, identity security, backup strategies, incident response plans και AI-based defensive tools. Η IBM επισημαίνει ότι η εκτεταμένη χρήση ΑΙ και security automation μπορεί να μειώσει το κόστος παραβίασης σε σύγκριση με οργανισμούς που δεν χρησιμοποιούν τέτοιες λύσεις [108]. Άρα η ΑΙ δεν είναι μόνο πηγή οικονομικού κινδύνου, αλλά και πιθανός μηχανισμός μείωσης κόστους όταν χρησιμοποιείται αμυντικά και υπεύθυνα.

Η “διπρόσωπη” αυτή οικονομική λειτουργία της ΑΙ είναι σημαντική. Από τη μία πλευρά, η ΑΙ μειώνει το κόστος και αυξάνει την αποτελεσματικότητα των επιθέσεων. Από την άλλη, μπορεί να συμβάλει στη μείωση του κόστους παραβίασης μέσω ταχύτερης ανίχνευσης, αυτοματοποιημένης απόκρισης και καλύτερης ανάλυσης απειλών. Η οικονομική ισορροπία, επομένως, δεν εξαρτάται μόνο από το αν οι επιτιθέμενοι χρησιμοποιούν ΑΙ, αλλά και από το αν οι οργανισμοί επενδύουν σε αντίστοιχες αμυντικές δυνατότητες [108], [27].

Σε παγκόσμιο επίπεδο, οι οικονομικές ζημιές του κυβερνοεγκλήματος είναι δύσκολο να εκτιμηθούν με ακρίβεια. Η ακαδημαϊκή μελέτη των Lukošiusė, Halstead και Righetti εκτιμά

παγκόσμιες ζημιές κυβερνοεγκλήματος περίπου 500 δισεκατομμυρίων δολαρίων ετησίως και εξετάζει πώς μια AI-driven αύξηση θα μπορούσε να προσθέσει σημαντικό επιπλέον κόστος [189]. Η αξία αυτής της προσέγγισης είναι ότι αντιμετωπίζει την AI όχι ως μεμονωμένη τεχνική μεταβολή, αλλά ως παράγοντα που μπορεί να αυξήσει το συνολικό οικονομικό αποτύπωμα του κυβερνοεγκλήματος.

Τελικά οι AI-driven κυβερνοεπιθέσεις λειτουργούν ως οικονομικός πολλαπλασιαστής κινδύνου. Η τεχνητή νοημοσύνη δεν δημιουργεί από μόνη της όλες τις οικονομικές απώλειες του κυβερνοεγκλήματος, αλλά ενισχύει μηχανισμούς που ήδη παράγουν σημαντική ζημία όπως phishing, ransomware, BEC, fraud, data breaches και cybercrime-as-a-service. Με τη μείωση του κόστους για τους επιτιθέμενους και την αύξηση του κόστους άμυνας και αποκατάστασης για τα θύματα, η AI μετατρέπει την κυβερνοασφάλεια σε κρίσιμο ζήτημα οικονομικής ανθεκτικότητας για οργανισμούς, επιχειρήσεις, δημόσιες υποδομές και κοινωνίες.

9.2. Κοινωνικές επιπτώσεις (σε εταιρείες, κυβερνήσεις και άτομα)

Οι κυβερνοεπιθέσεις που ενισχύονται από τεχνητή νοημοσύνη δεν προκαλούν μόνο οικονομικές απώλειες ή τεχνικές δυσλειτουργίες. Οι επιπτώσεις τους επεκτείνονται στην εμπιστοσύνη, στη λειτουργικότητα οργανισμών, στην αξιοπιστία δημόσιων θεσμών και στην ασφάλεια των πολιτών. Η AI αυξάνει την ταχύτητα, την πειστικότητα και την κλίμακα επιθέσεων όπως phishing, social engineering, deepfake fraud, identity theft και ransomware, μετατρέποντας την κυβερνοασφάλεια και σε κοινωνικό ζήτημα [188], [32], [27].

Η έκταση του φαινομένου αποτυπώνεται στο ENISA Threat Landscape 2025, όπου αναλύθηκαν 4.875 επιλεγμένα περιστατικά για την περίοδο από Ιούλιο του 2024 έως Ιούνιο του 2025 [188]. Το στοιχείο αυτό δείχνει ότι οι κυβερνοαπειλές επηρεάζουν διαφορετικούς τομείς της οικονομικής και κοινωνικής ζωής, από επιχειρήσεις και δημόσιους οργανισμούς έως φυσικά πρόσωπα. Συνεπώς, οι κοινωνικές επιπτώσεις των AI-driven επιθέσεων δεν πρέπει να εξετάζονται ως δευτερεύουσες συνέπειες, αλλά ως βασικό μέρος της συνολικής βλάβης που προκαλείται.

Σε επίπεδο επιχειρήσεων, οι AI-driven επιθέσεις μπορούν να πλήξουν τη φήμη, την εμπιστοσύνη των πελατών και την καθημερινή λειτουργία του οργανισμού. Μια επιτυχής επίθεση phishing, ransomware ή παραβίαση δεδομένων μπορεί να ερμηνευθεί από πελάτες και συνεργάτες ως ένδειξη αδυναμίας προστασίας πληροφοριών και υπηρεσιών. Η απώλεια εμπιστοσύνης μπορεί να οδηγήσει σε απομάκρυνση πελατών, δυσκολία συνεργασιών και υποβάθμιση της ανταγωνιστικής θέσης της επιχείρησης [186], [27], [28].

Παράλληλα, οι επιθέσεις αυτές επιβαρύνουν την εσωτερική λειτουργία των οργανισμών. Το προσωπικό πρέπει να εκπαιδευτεί εκ νέου, οι διαδικασίες επικοινωνίας γίνονται πιο αυστηρές και οι εργαζόμενοι καλούνται να αξιολογούν συνεχώς αν ένα μήνυμα, μια φωνητική εντολή ή ένα αίτημα είναι αυθεντικό. Το UK Government Cyber Security Breaches Survey 2025 καταγράφει το phishing ως τη συχνότερη και πιο διαταρακτική μορφή επίθεσης για επηρεασμένες επιχειρήσεις και charities, γεγονός που δείχνει ότι ακόμη και οι φαινομενικά απλές επιθέσεις μπορούν να έχουν σοβαρό οργανωτικό αντίκτυπο [28].

Η τεχνητή νοημοσύνη εντείνει αυτόν τον κίνδυνο, επειδή καθιστά τις επιθέσεις social engineering πιο πειστικές. AI-generated emails, πλαστά προφίλ, deepfake φωνές και αυτοματοποιημένες συνομιλίες μειώνουν τα εμφανή σημάδια απάτης και αυξάνουν την πιθανότητα ανθρώπινου λάθους. Η Microsoft συνδέει τη χρήση AI με πιο εξελιγμένα phishing σενάρια, deepfakes και πολυσταδιακές επιθετικές αλυσίδες, οι οποίες δυσκολεύουν εργαζομένους και οργανισμούς να διακρίνουν την αυθεντική από την πλαστή επικοινωνία [32].

Στο επίπεδο των κυβερνήσεων και των δημόσιων οργανισμών, οι κοινωνικές επιπτώσεις είναι ακόμη ευρύτερες. Επιθέσεις σε δημόσιες υπηρεσίες, κρίσιμες υποδομές, οργανισμούς υγείας, ενέργειας, μεταφορών ή διοικητικά συστήματα μπορούν να επηρεάσουν άμεσα την καθημερινότητα των πολιτών. Σε τέτοιες περιπτώσεις, η κυβερνοεπίθεση δεν πλήττει μόνο έναν οργανισμό, αλλά την ικανότητα του κράτους να παρέχει αξιόπιστες και συνεχείς υπηρεσίες [188], [192], [199].

Η χρήση deepfakes και συνθετικού περιεχομένου μπορεί επίσης να επηρεάσει την πολιτική και θεσμική εμπιστοσύνη. Η Europol αναφέρεται σε AI-powered voice cloning, live video deepfakes και synthetic media ως εργαλεία που μπορούν να χρησιμοποιηθούν σε fraud, extortion και identity theft [30]. Όταν οι πολίτες δεν μπορούν εύκολα να διακρίνουν αν μια φωνή, ένα βίντεο ή μια

δημόσια ανακοίνωση είναι πραγματική, υπονομεύεται η εμπιστοσύνη στο ψηφιακό περιβάλλον και στους θεσμούς.

Η απειλή αυτή γίνεται ιδιαίτερα σοβαρή σε περιόδους κρίσης, όπου η παραποιημένη πληροφορία μπορεί να προκαλέσει σύγχυση, πανικό ή απώλεια εμπιστοσύνης σε αξιόπιστους οργανισμούς. Οι AI-driven κυβερνοεπιθέσεις μπορούν έτσι να συνδυαστούν με παραπληροφόρηση, κοινωνική χειραγωγή και επιθέσεις σε κρίσιμες υπηρεσίες, δημιουργώντας επιπτώσεις που υπερβαίνουν το τεχνικό περιστατικό [186], [30], [198].

Για τα άτομα, οι συνέπειες είναι συχνά άμεσες και προσωπικές. Οι πολίτες μπορεί να υποστούν οικονομική εξαπάτηση, κλοπή ταυτότητας, παραβίαση ιδιωτικότητας, έκθεση προσωπικών δεδομένων ή χειραγωγή μέσω προσωποποιημένων μηνυμάτων. Η Europol συνδέει το οργανωμένο κυβερνοέγκλημα με online fraud, social engineering, fake social media profiles και χρήση GenAI για παραπλάνηση [22].

Η οικονομική εξαπάτηση μέσω social media, ψεύτικων προφίλ και προσωποποιημένων scams έχει επίσης σημαντικό κοινωνικό βάρος. Εκθέσεις για scams δείχνουν ότι investment scams, romance scams, payment redirection και phishing δεν προκαλούν μόνο οικονομικές απώλειες, αλλά και απώλεια εμπιστοσύνης στις ψηφιακές συναλλαγές και στις διαδικτυακές σχέσεις [185], [195].

Η κλοπή ταυτότητας αποτελεί χαρακτηριστικό παράδειγμα επίπτωσης που ξεπερνά την οικονομική διάσταση. Τα θύματα μπορεί να αντιμετωπίσουν δυσκολίες αποκατάστασης λογαριασμών, απώλεια ευκαιριών, επαναλαμβανόμενες επιθέσεις και σημαντική ψυχολογική επιβάρυνση. Το Identity Theft Resource Center αναδεικνύει ότι τα identity crimes έχουν τόσο οικονομικό όσο και συναισθηματικό αντίκτυπο, ιδίως όταν συνδέονται με AI-enabled μορφές εξαπάτησης [194].

Ιδιαίτερη σημασία έχουν οι επιθέσεις που αξιοποιούν deepfakes και συνθετικές φωνές. Η Pindrop αναλύει την αύξηση των κινδύνων από voice cloning και AI-driven fraud, ενώ η Resemble AI καταγράφει περιστατικά deepfake fraud, impersonation και άλλες μορφές κοινωνικής βλάβης που προκαλούνται από AI [65], [184]. Οι τεχνικές αυτές δεν παραβιάζουν μόνο συστήματα, αλλά εκμεταλλεύονται την ανθρώπινη εμπιστοσύνη προς φωνές, πρόσωπα και οικείες μορφές επικοινωνίας.

Οι υπερ-προσωποποιημένες επιθέσεις αποτελούν έναν από τους πιο ανησυχητικούς κοινωνικούς κινδύνους. Η GenAI μπορεί να αξιοποιήσει δημόσια δεδομένα από κοινωνικά δίκτυα για να δημιουργήσει context-aware spear-phishing μηνύματα, τα οποία φαίνονται πιο προσωπικά, πιο σχετικά και λιγότερο ύποπτα για το θύμα [90]. Η δυσκολία αναγνώρισης τέτοιου περιεχομένου επιβεβαιώνεται και από μελέτες που δείχνουν ότι σημαντικό ποσοστό χρηστών δεν μπορεί να αναγνωρίσει AI-generated audio ή video ως ψεύτικο [150], [196].

Οι ευάλωτες ομάδες επηρεάζονται δυσανάλογα. Ηλικιωμένοι, άτομα με περιορισμένη ψηφιακή εξοικείωση ή πολίτες που βρίσκονται σε συναισθηματική πίεση μπορεί να είναι πιο εκτεθειμένοι σε deepfake scams και πλαστοπροσωπία συγγενικών ή αξιόπιστων προσώπων. Η σχετική μελέτη για deepfake scams targeting older adults δείχνει ότι τέτοιες επιθέσεις έχουν έντονη κοινωνική και ψυχολογική διάσταση [197].

Η AI μεταβάλλει τις κοινωνικές επιπτώσεις των κυβερνοεπιθέσεων, επειδή οι επιθέσεις δεν στοχεύουν πλέον μόνο τεχνικές αδυναμίες, αλλά και σχέσεις εμπιστοσύνης μεταξύ ατόμων, οργανισμών και θεσμών. Μέσω deepfakes, συνθετικής φωνής, προσωποποιημένου phishing και πλαστοπροσωπίας, η διάκριση μεταξύ αυθεντικής και κακόβουλης επικοινωνίας γίνεται δυσκολότερη. Η αντιμετώπιση της απειλής απαιτεί, συνεπώς, όχι μόνο τεχνικά μέτρα, αλλά και ενίσχυση της ψηφιακής παιδείας, των διαδικασιών επαλήθευσης και της κοινωνικής ανθεκτικότητας.

Πίνακας 9.1. Οικονομικές, οργανωτικές και κοινωνικές επιπτώσεις των AI-enabled κυβερνοεπιθέσεων.

Επίπεδο επίπτωσης	Ενδεικτικές συνέπειες	Επιχειρησιακή ή κοινωνική σημασία	Ανάγκη αντιμετώπισης
----------------------	--------------------------	--------------------------------------	----------------------

Οικονομικό	Απώλειες από απάτη, κόστος αποκατάστασης, διακοπή λειτουργίας, νομικές και ασφαλιστικές επιβαρύνσεις.	Οι επιθέσεις μπορεί να καταστούν περισσότερο αποδοτικές για τους επιτιθέμενους και ακριβότερες για τα θύματα.	Προληπτική κυβερνοασφάλεια, σχέδια επιχειρησιακής συνέχειας και ταχεία απόκριση σε περιστατικά.
Οργανωτικό	Παραβίαση λογαριασμών, διαρροή δεδομένων, διατάραξη υπηρεσιών, φθορά φήμης και αυξημένος φόρτος ομάδων ασφάλειας.	Επηρεάζονται η εμπιστοσύνη, η λειτουργική συνέχεια και η ικανότητα του οργανισμού να διαχειριστεί κρίσεις.	Διαχείριση ταυτοτήτων, εκπαίδευση προσωπικού, παρακολούθηση απειλών και ελεγχόμενη αυτοματοποίηση απόκρισης.
Κοινωνικό και θεσμικό	Παραπληροφόρηση, deepfakes, απώλεια εμπιστοσύνης σε ψηφιακές επικοινωνίες και αυξημένη ευαλωτότητα χρηστών.	Δυσκολεύεται η διάκριση μεταξύ αυθεντικής και κακόβουλης πληροφορίας, ενώ κλονίζεται η εμπιστοσύνη σε θεσμούς και οργανισμούς.	Ψηφιακή παιδεία, διαδικασίες επαλήθευσης, θεσμική συνεργασία και προστασία θεμελιωδών δικαιωμάτων.

9.3. Αλλαγές στη στρατηγική κυβερνοασφάλειας λόγω AI

Η άνοδος των AI-enabled κυβερνοεπιθέσεων οδηγεί σε σημαντική αναθεώρηση των στρατηγικών κυβερνοασφάλειας που ακολουθούν οργανισμοί, επιχειρήσεις και κρατικοί φορείς. Οι

οικονομικές απώλειες και οι κοινωνικές επιπτώσεις που αναλύθηκαν στα προηγούμενα υποκεφάλαια δείχνουν ότι η κυβερνοασφάλεια δεν μπορεί πλέον να αντιμετωπίζεται ως στενά τεχνικό ζήτημα. Η τεχνητή νοημοσύνη αυξάνει την ταχύτητα, την κλίμακα και την προσαρμοστικότητα των επιθέσεων, ενώ ταυτόχρονα επιτρέπει πιο πειστικές μορφές phishing, deepfake απάτες, αυτοματοποιημένη αναγνώριση στόχων και ταχύτερη εκμετάλλευση ευπαθειών [27], [32], [21]. Ως αποτέλεσμα, οι οργανισμοί οδηγούνται σταδιακά από μια κυρίως αντιδραστική λογική άμυνας σε πιο προληπτικές, προσαρμοστικές και AI-supported στρατηγικές κυβερνοασφάλειας.

Οι παραδοσιακές στρατηγικές κυβερνοασφάλειας βασίζονταν σε μεγάλο βαθμό σε κανόνες πρωτοκόλλου, συχνούς ελέγχους, perimeter-based προστασία, εκπαίδευση προσωπικού ανά αραιά χρονικά διαστήματα και απόκριση μετά την εκδήλωση του περιστατικού. Η προσέγγιση αυτή καθίσταται λιγότερο επαρκής απέναντι σε απειλές που μπορούν να μεταβάλλονται γρήγορα, να εξατομικεύονται και να εκτελούνται σε μεγάλη κλίμακα. Οι AI-generated επιθέσεις social engineering, το adaptive malware και η αυτοματοποιημένη συλλογή πληροφοριών απαιτούν αμυντικά μοντέλα που δεν περιμένουν απλώς την εμφάνιση της επίθεσης, αλλά παρακολουθούν συνεχώς το περιβάλλον, αναγνωρίζουν ανωμαλίες και προσαρμόζουν τις αντιδράσεις τους [206], [54], [47].

Η πρώτη βασική αλλαγή είναι η μετάβαση προς την προληπτική ανίχνευση και τη συνεχή παρακολούθηση. Οι οργανισμοί χρειάζεται να αξιοποιούν μηχανισμούς που συλλέγουν και αναλύουν δεδομένα ασφαλείας σε πραγματικό χρόνο, ώστε να εντοπίζουν αποκλίσεις από τη φυσιολογική συμπεριφορά χρηστών, συστημάτων και δικτύων. Η τεχνητή νοημοσύνη μπορεί να υποστηρίξει την ανίχνευση τέτοιων μοτίβων, την ιεράρχηση συμβάντων, την ανάλυση μεγάλου όγκου δεδομένων και την ενίσχυση των λειτουργιών SOC. Ωστόσο, η χρήση της πρέπει να συνοδεύεται από ανθρώπινη εποπτεία, αξιολόγηση ποιότητας δεδομένων και σαφείς διαδικασίες ελέγχου, ώστε να αποφεύγονται λανθασμένες αποφάσεις ή υπερβολική εξάρτηση από αυτοματοποιημένα συστήματα [201], [206], [54], [47].

Δεύτερη στρατηγική αλλαγή είναι η ενίσχυση της αυτοματοποίησης στην απόκριση. Καθώς οι επιθέσεις εκτελούνται ταχύτερα, η χειροκίνητη ανάλυση και αντίδραση συχνά δεν επαρκεί. Το security automation μπορεί να συμβάλλει στη μείωση του χρόνου εντοπισμού και αντιμετώπισης

περιστατικών, στην προτεραιοποίηση alerts και στην αυτοματοποιημένη εφαρμογή προκαθορισμένων ενεργειών, όπως απομόνωση συστημάτων, μπλοκάρισμα ύποπτης δραστηριότητας ή ενεργοποίηση διαδικασιών incident response [108], [25], [106]. Η αυτοματοποίηση, όμως, δεν πρέπει να λειτουργεί ανεξέλεγκτα. Η αποτελεσματική στρατηγική απαιτεί ισορροπία ανάμεσα στην ταχύτητα των συστημάτων ΑΙ και στην ανθρώπινη κρίση των αναλυτών ασφαλείας.

Τρίτη σημαντική κατεύθυνση είναι η υιοθέτηση Zero Trust αρχιτεκτονικών. Η παραδοσιακή λογική που θεωρούσε ότι οι χρήστες ή οι συσκευές εντός του δικτύου είναι περισσότερο αξιόπιστοι δεν ανταποκρίνεται πλέον σε περιβάλλοντα cloud, remote work, identity-based attacks και AI-supported credential abuse. Η Zero Trust προσέγγιση βασίζεται στην αρχή ότι καμία πρόσβαση δεν πρέπει να θεωρείται δεδομένη και ότι κάθε χρήστης, συσκευή, εφαρμογή και αίτημα πρέπει να ελέγχεται συνεχώς [202], [204], [207]. Έτσι, η στρατηγική ασφάλειας μετατοπίζεται από την προστασία περιμέτρου στη συνεχή επαλήθευση ταυτότητας, στην ελάχιστη αναγκαία πρόσβαση και στην παρακολούθηση συμπεριφοράς.

Η Zero Trust Architecture αποκτά ιδιαίτερη σημασία απέναντι σε AI-enabled επιθέσεις, επειδή οι επιτιθέμενοι μπορούν να αξιοποιήσουν κλεμμένα διαπιστευτήρια, deepfake πλαστοπροσωπία ή αυτοματοποιημένη αναγνώριση για να παρακάμψουν παραδοσιακούς ελέγχους. Η χρήση continuous authentication, least privilege, microsegmentation και real-time analytics περιορίζει την πιθανότητα μια μεμονωμένη παραβίαση να οδηγήσει σε εκτεταμένη πρόσβαση σε κρίσιμα συστήματα [202], [207], [208], [209]. Με αυτόν τον τρόπο, η Zero Trust στρατηγική δεν αποτρέπει απαραίτητα κάθε αρχική επίθεση, αλλά μειώνει το εύρος και την ένταση της πιθανής ζημίας.

Τέταρτη αλλαγή αφορά την κυβερνοανθεκτικότητα. Η στρατηγική κυβερνοασφάλειας δεν μπορεί να περιορίζεται στην αποτροπή κάθε επίθεσης, διότι σε ένα περιβάλλον σύνθετων και AI-enhanced απειλών η πλήρης αποφυγή περιστατικών δεν είναι ρεαλιστική. Οι οργανισμοί πρέπει να σχεδιάζουν με βάση την υπόθεση ότι κάποια επίθεση μπορεί να επιτύχει και να επενδύουν σε δυνατότητες συνέχισης λειτουργίας, ταχείας ανάκτησης, εφεδρικών συστημάτων, ασκήσεων incident response και σαφών διαδικασιών αποκατάστασης [27], [205], [192], [25]. Η κυβερνοανθεκτικότητα μετατρέπει την ασφάλεια από απλή άμυνα σε ικανότητα επιβίωσης και επαναφοράς.

Η στρατηγική εκπαίδευσης του προσωπικού πρέπει επίσης να αλλάξει. Η παραδοσιακή εκπαίδευση σε γενικά παραδείγματα phishing δεν επαρκεί απέναντι σε AI-generated μηνύματα, deepfake φωνές, πλαστά προφίλ και context-aware social engineering. Οι εργαζόμενοι χρειάζεται να εκπαιδεύονται σε πιο ρεαλιστικά σενάρια, να αναγνωρίζουν ασυνήθιστες διαδικασίες έγκρισης, να επαληθεύουν ευαίσθητα αιτήματα μέσα από δεύτερα κανάλια και να κατανοούν ότι το γλωσσικά άρτιο περιεχόμενο δεν αποτελεί πλέον ένδειξη αξιοπιστίας [205], [23]. Η ανθρώπινη επίγνωση παραμένει κρίσιμος αμυντικός παράγοντας, αλλά πρέπει να προσαρμοστεί στη νέα ποιότητα των AI-assisted επιθέσεων.

Η διακυβέρνηση δεδομένων και η διακυβέρνηση AI αποτελούν αναγκαίο μέρος της σύγχρονης στρατηγικής. Οι οργανισμοί δεν χρειάζεται μόνο να προστατεύουν τα δεδομένα τους από εξωτερικούς επιτιθέμενους, αλλά και να γνωρίζουν ποια δεδομένα χρησιμοποιούνται σε AI συστήματα, με ποιον σκοπό, με ποιους περιορισμούς και υπό ποια εποπτεία. Το NIST AI Risk Management Framework και το Cybersecurity Framework Profile for Artificial Intelligence υπογραμμίζουν τη σημασία της συστηματικής διαχείρισης κινδύνων στο σημείο τομής AI και κυβερνοασφάλειας [201], [4].

Η ανάγκη threat intelligence sharing γίνεται επίσης πιο έντονη. Οι AI-enabled επιθέσεις μπορούν να εξαπλώνονται γρήγορα και να επηρεάζουν πολλούς οργανισμούς ή κλάδους. Η ανταλλαγή πληροφοριών για δείκτες απειλής, μοτίβα επίθεσης, exploited vulnerabilities και νέες τεχνικές επιτρέπει στους οργανισμούς να αντιδρούν προληπτικά και όχι μόνο μετά την παραβίαση. Η ENISA επισημαίνει τη σημασία της ανταλλαγής πληροφοριών, της συνεργασίας και των στρατηγικών απόκρισης σε περιβάλλον όπου η AI επηρεάζει το threat landscape [21], [99]. Η συλλογική γνώση μειώνει την ασυμμετρία μεταξύ επιτιθέμενων και αμυνόμενων.

Η στρατηγική κυβερνοασφάλειας λόγω AI πρέπει να γίνει ολιστική. Δεν αρκεί να περιορίζεται στο IT department, διότι οι AI-enabled επιθέσεις επηρεάζουν νομική συμμόρφωση, προστασία δεδομένων, οικονομική συνέχεια, εταιρική φήμη, ανθρώπινο δυναμικό και επικοινωνιακή διαχείριση κρίσεων. Η διοίκηση, το νομικό τμήμα, ο υπεύθυνος προστασίας δεδομένων, το HR, η οικονομική διεύθυνση, οι ομάδες επικοινωνίας και οι εξωτερικοί πάροχοι πρέπει να συμμετέχουν στον σχεδιασμό και στην εφαρμογή της στρατηγικής [27], [45], [206]. Με τον τρόπο αυτό, η

κυβερνοασφάλεια μετατρέπεται από τεχνική λειτουργία σε ζήτημα επιχειρησιακής διακυβέρνησης.

Η οικονομική διάσταση ενισχύει αυτή τη μεταβολή. Τα υψηλά κόστη των data breaches, η διακοπή λειτουργίας, οι απαιτήσεις αποκατάστασης και οι επιπτώσεις στη φήμη καθιστούν την κυβερνοασφάλεια ζήτημα επιχειρησιακής βιωσιμότητας. Η χρήση security AI και automation μπορεί να συμβάλλει στη μείωση του χρόνου εντοπισμού και αντιμετώπισης περιστατικών, εφόσον εντάσσεται σε ώριμη στρατηγική και όχι σε αποσπασματική τεχνολογική υιοθέτηση [108]. Η AI, επομένως, δεν πρέπει να αντιμετωπίζεται μόνο ως εργαλείο άμυνας, αλλά ως παράγοντας που επηρεάζει συνολικά τη σχέση κόστους, κινδύνου και ανθεκτικότητας.

Η AI μεταβάλλει ριζικά τη στρατηγική κυβερνοασφάλειας. Οι οργανισμοί δεν μπορούν να στηρίζονται μόνο σε αντιδραστικά μέτρα, στατικούς κανόνες και αποσπασματική εκπαίδευση. Απαιτείται μετάβαση σε προληπτικά και προσαρμοστικά μοντέλα, τα οποία συνδυάζουν AI-based detection, Zero Trust, συνεχή παρακολούθηση, αυτοματοποιημένη απόκριση, κυβερνοανθεκτικότητα, threat intelligence sharing, εκπαίδευση προσωπικού και υπεύθυνη AI governance. Με αυτή την έννοια, η AI δεν αλλάζει μόνο τις κυβερνοεπιθέσεις, αλλά και τον τρόπο με τον οποίο οργανισμοί και κράτη οφείλουν να σχεδιάζουν την άμυνά τους [27], [201], [45], [4].

10. ΜΕΛΛΟΝΤΙΚΕΣ ΤΑΣΕΙΣ ΚΑΙ ΠΡΟΚΛΗΣΕΙΣ

Η εξέλιξη της τεχνητής νοημοσύνης αναμένεται να επηρεάσει καθοριστικά το μέλλον της κυβερνοασφάλειας, τόσο στο επίπεδο των επιθέσεων όσο και στο επίπεδο της άμυνας. Τα προηγούμενα κεφάλαια έδειξαν ότι η ΑΙ έχει ήδη μεταβάλει το κόστος, την ταχύτητα, την κλίμακα και την πολυπλοκότητα των κυβερνοεπιθέσεων. Το παρόν κεφάλαιο εξετάζει τις μελλοντικές τάσεις που προκύπτουν από αυτή τη μεταβολή, εστιάζοντας στην εξέλιξη των αλγορίθμων, στη χρήση της ΑΙ στην κυβερνοάμυνα και στις πιθανές ρυθμιστικές εξελίξεις που συνδέονται με την ασφάλεια πληροφοριών.

Το βασικό ζήτημα για το μέλλον δεν είναι αν η τεχνητή νοημοσύνη θα συνεχίσει να επηρεάζει τον κυβερνοχώρο, αλλά με ποιον τρόπο και σε ποιον βαθμό. Η αύξηση των δυνατοτήτων των γενικών ΑΙ συστημάτων, η ανάπτυξη agentic ΑΙ, η διάδοση LLM εφαρμογών και η ενσωμάτωση ΑΙ σε επιχειρησιακά περιβάλλοντα δημιουργούν νέες ευκαιρίες, αλλά και νέες επιφάνειες επίθεσης [83], [211]. Συνεπώς, η κυβερνοασφάλεια του μέλλοντος θα απαιτεί ταυτόχρονα τεχνική προσαρμογή, θεσμική ετοιμότητα και υπεύθυνη διακυβέρνηση της ΑΙ.

10.1. Εξέλιξη αλγορίθμων και νέες απειλές

Η εξέλιξη των αλγορίθμων τεχνητής νοημοσύνης αναμένεται να δημιουργήσει νέες μορφές κυβερνοαπειλών, καθώς τα ΑΙ συστήματα αποκτούν μεγαλύτερες δυνατότητες συλλογισμού, παραγωγής κώδικα, ανάλυσης δεδομένων και αυτοματοποίησης σύνθετων εργασιών. Το International AI Safety Report 2026 επισημαίνει ότι τα γενικά ΑΙ συστήματα χρησιμοποιούνται ήδη σε κυβερνοεπιθέσεις και ότι παρουσιάζουν ιδιαίτερη ικανότητα στην ανακάλυψη λογισμικών ευπαθειών και στη συγγραφή κακόβουλου κώδικα [83]. Η εξέλιξη αυτή δεν σημαίνει ότι η ΑΙ θα αντικαταστήσει πλήρως τον ανθρώπινο επιτιθέμενο, αλλά ότι μπορεί να αυξήσει την επιχειρησιακή του αποτελεσματικότητα.

10.1.1. Αυτόνομες επιθετικές ροές

Μία από τις σημαντικότερες μελλοντικές απειλές είναι η ανάπτυξη περισσότερο αυτόνομων επιθετικών ροών. Η agentic ΑΙ μπορεί να επιτρέψει σε συστήματα να εκτελούν διαδοχικά βήματα

μιας επιχείρησης, όπως reconnaissance, επιλογή στόχων, ανάλυση ευπαθειών, παραγωγή scripts και προσαρμογή της επίθεσης ανάλογα με τα αποτελέσματα. Η μελέτη για την βιομηχανοποίηση του κυβερνοεγκλήματος υποστηρίζει ότι η agentic AI μπορεί να συμπίσει τον χρόνο που απαιτείται για μία επίθεση και να αυξήσει την κλιμάκωση επιθετικών διαδικασιών [69]. Αυτό καθιστά πιθανή τη μετάβαση από απλή AI-assisted επίθεση σε περισσότερες οργανωμένες και αυτοματοποιημένες επιθετικές αλυσίδες.

10.1.2. Ταχύτερη εκμετάλλευση ευπαθειών

Η αύξηση των δυνατοτήτων των μοντέλων ενισχύει επίσης τον κίνδυνο ταχύτερης εκμετάλλευσης ευπαθειών. Εάν τα AI συστήματα μπορούν να αναλύουν τεχνική τεκμηρίωση, κώδικα, ενημερώσεις ασφαλείας και δημόσιες αναφορές ευπαθειών με μεγάλη ταχύτητα, τότε ο χρόνος μεταξύ αποκάλυψης μιας ευπάθειας και πιθανής εκμετάλλευσής της μπορεί να μειωθεί. Η ENISA αναδεικνύει ήδη τη γρήγορη αξιοποίηση ευπαθειών ως σημαντικό στοιχείο του threat landscape, ενώ η ενσωμάτωση AI μπορεί να επιταχύνει περαιτέρω αυτή τη διαδικασία [21].

10.1.3. Επιθέσεις εναντίον AI συστημάτων

Ιδιαίτερη κατηγορία μελλοντικών απειλών αποτελούν οι επιθέσεις εναντίον των ίδιων των AI συστημάτων. Καθώς οι οργανισμοί ενσωματώνουν LLMs, AI agents και μοντέλα μηχανικής μάθησης σε παραγωγικές διαδικασίες, τα συστήματα αυτά μετατρέπονται σε νέοι στόχοι επίθεσης. Το OWASP Top 10 for LLM Applications αναδεικνύει κινδύνους όπως prompt injection, insecure output handling, data poisoning, model denial of service και supply-chain vulnerabilities [211]. Παράλληλα, το MITRE ATLAS καταγράφει τακτικές και τεχνικές επιθέσεων που στοχεύουν AI-enabled systems [210].

10.1.4. Επιθέσεις adversarial machine learning

Οι adversarial machine learning επιθέσεις αναμένεται επίσης να παραμείνουν κρίσιμο ζήτημα. Το NIST καταγράφει τεχνικές όπως evasion, poisoning, privacy attacks και άλλες μορφές υπονόμευσης της αξιοπιστίας των μοντέλων [8]. Σε περιβάλλον κυβερνοασφάλειας, τέτοιες επιθέσεις μπορούν να οδηγήσουν σε λανθασμένη ταξινόμηση κακόβουλης δραστηριότητας, παράκαμψη συστημάτων ανίχνευσης ή υποβάθμιση της αμυντικής αποτελεσματικότητας. Αυτό

σημαίνει ότι η μελλοντική κυβερνοάμυνα θα πρέπει να προστατεύει τόσο τα παραδοσιακά πληροφοριακά συστήματα όσο και τα ΑΙ μοντέλα που χρησιμοποιούνται για την προστασία τους.

10.1.5. AI-powered social engineering επιθέσεις

Μία ακόμη μελλοντική τάση είναι η αύξηση των AI-powered social engineering επιθέσεων. Η GenAI μπορεί να δημιουργεί πειστικά μηνύματα, ψεύτικες ταυτότητες, συνθετικές φωνές και εξατομικευμένο περιεχόμενο σε μεγάλη κλίμακα. Οι πραγματικές αναφορές χρήσης GenAI από threat actors δείχνουν ότι η τεχνολογία αξιοποιείται ήδη για operational support, phishing και άλλες κακόβουλες δραστηριότητες [96], [41]. Με τη βελτίωση των μοντέλων, η διάκριση μεταξύ αυθεντικής και κατασκευασμένης επικοινωνίας ενδέχεται να γίνει ακόμη δυσκολότερη.

Η μελλοντική απειλή δεν θα προκύπτει απαραίτητα από μία εντελώς νέα μορφή κυβερνοεπίθεσης, αλλά από τη βελτίωση και κλιμάκωση ήδη γνωστών τεχνικών. Phishing, malware, credential theft, reconnaissance και vulnerability exploitation θα συνεχίσουν να υπάρχουν, αλλά θα εκτελούνται ταχύτερα, πιο προσωποποιημένα και κυρίως με μικρότερη ανάγκη ανθρώπινης παρέμβασης [83], [212]. Συνεπώς, η εξέλιξη των αλγορίθμων δεν αλλάζει μόνο το τεχνικό οπλοστάσιο των επιτιθέμενων, αλλά και την οικονομία, την ταχύτητα και την προσβασιμότητα των κυβερνοεπιθέσεων.

Τέλος, η ανάπτυξη cyber threat intelligence ειδικά για ΑΙ συστήματα θα αποκτήσει μεγαλύτερη σημασία. Καθώς οι οργανισμοί θα εξαρτώνται περισσότερο από ΑΙ μοντέλα, θα χρειάζονται μηχανισμούς παρακολούθησης AI-specific assets, supply chains, datasets, model behavior και AI-focused indicators of compromise [213]. Η μελλοντική πρόκληση δεν θα είναι μόνο η προστασία από επιθέσεις που χρησιμοποιούν ΑΙ, αλλά και η κατανόηση του πώς απειλούνται τα ίδια τα ΑΙ οικοσυστήματα.

10.2. Χρήση Τεχνητής Νοημοσύνης στην κυβερνοάμυνα

Η μελλοντική χρήση της τεχνητής νοημοσύνης στην κυβερνοάμυνα αναμένεται να αποτελέσει βασικό παράγοντα προσαρμογής απέναντι σε ταχύτερες, πιο σύνθετες και περισσότερο αυτοματοποιημένες απειλές. Καθώς οι επιθέσεις που αξιοποιούν ΑΙ μπορούν να εκτελούνται σε

μεγαλύτερη κλίμακα και με αυξημένη προσαρμοστικότητα, οι αμυντικοί μηχανισμοί θα χρειαστεί να αξιοποιούν αντίστοιχες τεχνολογίες για την ανάλυση δεδομένων, την ανίχνευση ανωμαλιών, την προτεραιοποίηση περιστατικών και την ταχύτερη απόκριση [214], [54], [47].

10.2.1. Security Operations Centers

Σε αυτό το πλαίσιο, η τεχνητή νοημοσύνη μπορεί να υποστηρίξει σημαντικά τη λειτουργία των Security Operations Centers, μειώνοντας τον όγκο των μη κρίσιμων ειδοποιήσεων και βοηθώντας τους αναλυτές να εστιάζουν στα περιστατικά που παρουσιάζουν μεγαλύτερο κίνδυνο. Η αξία της δεν βρίσκεται στην πλήρη αντικατάσταση του ανθρώπινου παράγοντα, αλλά στην ενίσχυση της ικανότητας των ομάδων ασφαλείας να επεξεργάζονται μεγάλο όγκο πληροφοριών και να λαμβάνουν ταχύτερες αποφάσεις [214], [106].

10.2.2. Threat intelligence

Παράλληλα, η AI μπορεί να συμβάλει στην καλύτερη αξιοποίηση του threat intelligence, στην ανάλυση κακόβουλου λογισμικού, στην αναγνώριση ύποπτης συμπεριφοράς και στην υποστήριξη διαδικασιών incident response. Ωστόσο, η αποτελεσματικότητά της εξαρτάται από την ποιότητα των δεδομένων, την ωριμότητα των διαδικασιών κυβερνοασφάλειας και την ύπαρξη σαφών μηχανισμών ανθρώπινης εποπτείας [201], [4], [25].

Η μελλοντική πρόκληση δεν θα είναι απλώς η υιοθέτηση περισσότερων AI εργαλείων, αλλά η υπεύθυνη ενσωμάτωσή τους σε πρωτόκολλα αμυντικής λειτουργίας. Τα αμυντικά AI συστήματα πρέπει να είναι αξιόπιστα, ελέγξιμα και να δρουν αποτρεπτικά απέναντι σε κακόβουλη χειραγώγηση, καθώς μπορούν και τα ίδια να αποτελέσουν στόχο επιθέσεων [8], [99]. Επομένως, η κυβερνοάμυνα του μέλλοντος θα βασίζεται σε έναν συνδυασμό τεχνολογικής αυτοματοποίησης, ανθρώπινης κρίσης, οργανωτικής ετοιμότητας και συνεχούς διαχείρισης κινδύνου.

Συνολικά, η τεχνητή νοημοσύνη αναμένεται να ενισχύσει σημαντικά την κυβερνοάμυνα, χωρίς όμως να αποτελεί αυτόνομη λύση. Η αποτελεσματική αξιοποίησή της θα εξαρτηθεί από το κατά πόσο οι οργανισμοί θα μπορέσουν να τη συνδυάσουν με κατάλληλη διακυβέρνηση, εκπαιδευμένο προσωπικό, σαφείς διαδικασίες απόκρισης και συνεχή αξιολόγηση των κινδύνων [214], [205].

10.3. Πιθανές εξελίξεις στο ρυθμιστικό πλαίσιο και στην ασφάλεια πληροφοριών

Οι μελλοντικές εξελίξεις στο ρυθμιστικό πλαίσιο για την τεχνητή νοημοσύνη και την κυβερνοασφάλεια αναμένεται να κινηθούν προς την κατεύθυνση της αυξημένης διακυβέρνησης, της διαχείρισης κινδύνου και της ενίσχυσης της κυβερνοανθεκτικότητας. Η Ευρωπαϊκή Ένωση έχει ήδη διαμορφώσει ένα πολυεπίπεδο πλαίσιο μέσα από τον AI Act, τη NIS2 Directive και τον Cyber Resilience Act, το οποίο συνδυάζει ρύθμιση της AI, προστασία κρίσιμων οργανισμών και ασφάλεια ψηφιακών προϊόντων [5], [160], [162].

Ο AI Act αναμένεται να αποτελέσει βασικό σημείο αναφοράς για την ασφαλή και υπεύθυνη χρήση της τεχνητής νοημοσύνης. Η σταδιακή εφαρμογή του έως το 2027 θα απαιτήσει από οργανισμούς και παρόχους AI συστημάτων να ενισχύσουν τη διαφάνεια, την τεκμηρίωση, την ανθρώπινη εποπτεία, τη διαχείριση κινδύνων και την ασφάλεια των συστημάτων τους [215], [156]. Ειδικά στο πεδίο της κυβερνοασφάλειας, η αξιοπιστία και η ανθεκτικότητα των AI συστημάτων θα αποτελέσουν κρίσιμο ζήτημα, καθώς τα ίδια τα μοντέλα μπορούν να γίνουν στόχος επιθέσεων ή χειραγώγησης [216].

Παράλληλα, η NIS2 και ο Cyber Resilience Act αναμένεται να ενισχύσουν τη μετάβαση προς πιο οργανωμένες και προληπτικές στρατηγικές ασφάλειας πληροφοριών. Η NIS2 δίνει έμφαση στη διαχείριση κινδύνων, στην αναφορά περιστατικών και στην προστασία κρίσιμων τομέων, ενώ ο Cyber Resilience Act προωθεί την ασφάλεια ψηφιακών προϊόντων ήδη από το στάδιο του σχεδιασμού [160], [162]. Η κατεύθυνση αυτή είναι ιδιαίτερα σημαντική σε περιβάλλον όπου οι AI-assisted επιτιθέμενοι μπορούν να εντοπίζουν και να εκμεταλλεύονται ευπάθειες με μεγαλύτερη ταχύτητα.

Σε διεθνές επίπεδο, η ρύθμιση της AI και του κυβερνοεγκλήματος αναμένεται να βασιστεί περισσότερο σε κοινές αρχές και συνεργασία. Η σύμβαση του συμβουλίου της Ευρώπης για την AI, η σύμβαση του ΟΗΕ κατά του κυβερνοεγκλήματος, οι OECD AI Principles και το G7 Hiroshima AI Process δείχνουν ότι η παγκόσμια συζήτηση μετατοπίζεται προς την ανάγκη κοινών κανόνων, ανταλλαγής πληροφοριών και προστασίας ανθρωπίνων δικαιωμάτων στο ψηφιακό περιβάλλον [167], [170], [173], [176].

Η βασική πρόκληση για τα επόμενα χρόνια θα είναι η ισορροπία ανάμεσα στην καινοτομία και στην ασφάλεια. Η χαλαρότητα μπορεί να αφήσει περιθώρια κατάχρησης προηγμένων ΑΙ συστημάτων, ενώ η υπερβολική αυστηρότητα μπορεί να περιορίσει την αμυντική καινοτομία. Συνεπώς, η μελλοντική ασφάλεια πληροφοριών θα απαιτεί ευέλικτα ρυθμιστικά πλαίσια, συνεχή επικαιροποίηση τεχνικών προτύπων, υπεύθυνη ανάπτυξη ΑΙ και συνεργασία μεταξύ κρατών, οργανισμών και τεχνολογικών εταιρειών [201], [4], [173].

Συνολικά, οι εξελίξεις στο ρυθμιστικό πλαίσιο και στην ασφάλεια πληροφοριών θα καθοριστούν από την ανάγκη προσαρμογής στους κινδύνους της ΑΙ. Η νομοθεσία δεν θα μπορεί να περιορίζεται σε γενικούς κανόνες συμμόρφωσης. Έτσι, αυτή η ρύθμιση θα λειτουργεί συμπληρωματικά προς την τεχνική άμυνα, δημιουργώντας ένα πιο ασφαλές και υπεύθυνο πλαίσιο χρήσης της τεχνητής νοημοσύνης στην ψηφιακή κοινωνία [5], [160], [162].

11. Συμπεράσματα και προτάσεις

11.1 Απαντήσεις στα ερευνητικά ερωτήματα

Στην παρούσα διπλωματική εργασία εξετάσαμε την επίδραση της τεχνητής νοημοσύνης στην πραγματοποίηση κυβερνοεπιθέσεων, με έμφαση στον τρόπο με τον οποίο οι αλγόριθμοι ΑΙ μεταβάλλουν τη φύση, το κόστος, την πολυπλοκότητα, την κλίμακα και την αποτελεσματικότητα των επιθέσεων.

Ως προς το πρώτο ερευνητικό ερώτημα, δηλαδή πώς η τεχνητή νοημοσύνη επηρεάζει τη φύση και τη δυναμική των κυβερνοεπιθέσεων, η εργασία έδειξε ότι η ΑΙ μετατρέπει τις επιθέσεις από χειροκίνητες ενέργειες σε πιο αυτοματοποιημένες. Η χρήση ΑΙ μπορεί να ενισχύσει στάδια όπως η συλλογή πληροφοριών, η αναγνώριση ευπαθειών, η παραγωγή επιθετικού περιεχομένου, το social engineering και η προσαρμογή της επίθεσης στις αντιδράσεις του αμυνόμενου.

Ως προς το δεύτερο ερευνητικό ερώτημα, που αφορά το κόστος, την πολυπλοκότητα, την κλίμακα και την αποτελεσματικότητα των επιθέσεων, διαπιστώθηκε ότι η ΑΙ λειτουργεί ως πολλαπλασιαστής ισχύος. Μειώνει τον χρόνο και το κόστος παραγωγής επιθετικού υλικού, αυξάνει τη δυνατότητα μαζικής εκτέλεσης επιθέσεων και βελτιώνει την πειστικότητα τεχνικών όπως το phishing, το social engineering και οι απάτες με πλαστό περιεχόμενο.

Ως προς το τρίτο ερευνητικό ερώτημα, η τεχνητή νοημοσύνη δηλαδή μειώνει σε σημαντικό βαθμό τις τεχνικές απαιτήσεις για τους επιτιθέμενους. Δείξαμε ότι η ύπαρξη έτοιμων εργαλείων, υπηρεσιών AI-as-a-Service, dark web εφαρμογών, αυτοματοποιημένων πλατφορμών και εγκληματικών οικοσυστημάτων επιτρέπει ακόμη και σε λιγότερο εξειδικευμένους δράστες να αποκτούν πρόσβαση σε δυνατότητες που παλαιότερα απαιτούσαν υψηλό επίπεδο τεχνικής γνώσης. Αυτό δεν σημαίνει ότι η τεχνική εξειδίκευση παύει να έχει σημασία, αλλά ότι μειώνεται το αρχικό εμπόδιο εισόδου στο κυβερνοέγκλημα.

Ως προς το τέταρτο ερευνητικό ερώτημα, που αφορά την ισορροπία μεταξύ επίθεσης και άμυνας, η εργασία κατέληξε ότι η ΑΙ ενισχύει και τις δύο πλευρές. Από τη μία πλευρά, οι επιτιθέμενοι αξιοποιούν ΑΙ για αυτοματοποίηση, κλιμάκωση, προσωποποίηση και ταχύτερη εκτέλεση επιθέσεων. Από την άλλη πλευρά, οι αμυνόμενοι μπορούν να χρησιμοποιήσουν ΑΙ για ανίχνευση

ανωμαλιών, threat intelligence, αυτοματοποιημένη απόκριση, ανάλυση μεγάλου όγκου δεδομένων και ενίσχυση των μηχανισμών κυβερνοάμυνας. Ωστόσο, η ισορροπία δεν εξαρτάται αποκλειστικά από την ύπαρξη της τεχνολογίας, αλλά από το ποια πλευρά τη χρησιμοποιεί πιο στρατηγικά και αποτελεσματικά.

Τέλος, ως προς το πέμπτο ερευνητικό ερώτημα, δείξαμε ότι οι επιπτώσεις των AI-enabled κυβερνοεπιθέσεων είναι οικονομικές, κοινωνικές, ηθικές, νομικές και θεσμικές. Οι οικονομικές απώλειες περιλαμβάνουν άμεσες ζημιές, κόστος αποκατάστασης, διακοπή λειτουργίας και φθορά φήμης. Οι κοινωνικές επιπτώσεις σχετίζονται με την απώλεια εμπιστοσύνης, την κλοπή ταυτότητας, τις deepfake απάτες και την πίεση προς άτομα, επιχειρήσεις και δημόσιους οργανισμούς. Παράλληλα, ανακύπτουν ηθικά ζητήματα ευθύνης, ανθρώπινης εποπτείας και κατάχρησης δεδομένων, καθώς και νομικές προκλήσεις που απαιτούν διεθνή συνεργασία, ρυθμιστική προσαρμογή και ενίσχυση της κυβερνοανθεκτικότητας.

11.2 Σύνοψη βασικών ευρημάτων της έρευνας

Το βασικό εύρημα της εργασίας είναι ότι η τεχνητή νοημοσύνη δεν δημιουργεί εξ ολοκλήρου νέες μορφές κυβερνοεπιθέσεων, αλλά ενισχύει, επιταχύνει και μετασχηματίζει ήδη υπάρχουσες τεχνικές. Εκείνο που αλλάζει είναι η ταχύτητα, η ακρίβεια, η δυνατότητα αυτοματοποίησης και η κλίμακα με την οποία μπορούν να πραγματοποιηθούν.

Μέσω έτοιμων εργαλείων, αυτοματοποιημένων υπηρεσιών και underground αγορών, ένας δράστης μπορεί να αποκτήσει πρόσβαση σε τεχνικές δυνατότητες χωρίς να διαθέτει απαραίτητα βαθιά εξειδίκευση. Αυτό οδηγεί σε αύξηση του κινδύνου, επειδή κάνει το κυβερνοέγκλημα εύκολα προσβάσιμο.

Παράλληλα, η εργασία έδειξε ότι η AI μεταβάλλει τη σχέση μεταξύ τεχνικής επίθεσης και ανθρώπινου παράγοντα. Οι επιθέσεις social engineering γίνονται πιο πειστικές, καθώς η AI μπορεί να παράγει γλωσσικά άρθριο, προσωποποιημένο και δύσκολα αναγνωρίσιμο περιεχόμενο. Η χρήση deepfakes, πλαστής φωνής και πλαστοπροσωπίας υπονομεύει την εμπιστοσύνη στις ψηφιακές επικοινωνίες και καθιστά τον άνθρωπο ακόμη πιο κρίσιμο σημείο ευπάθειας.

Ένα ακόμη βασικό εύρημα είναι ότι η κυβερνοάμυνα πρέπει να μεταβεί από την αντιδραστική αντιμετώπιση στην προληπτική και προσαρμοστική στρατηγική. Η χρήση ΑΙ από επιτιθέμενους καθιστά αναγκαία την ενίσχυση μηχανισμών ανίχνευσης, αυτοματοποιημένης απόκρισης, Zero Trust, threat intelligence και κυβερνοανθεκτικότητας. Η ΑΙ μπορεί να αποτελέσει ισχυρό εργαλείο άμυνας, αλλά μόνο όταν συνοδεύεται από ανθρώπινη εποπτεία και ορθή χρήση.

Τέλος, η εργασία ανέδειξε ότι η αντιμετώπιση των AI-enabled κυβερνοεπιθέσεων δεν είναι μόνο τεχνικό ζήτημα. Απαιτείται συνδυασμός τεχνολογικής προστασίας, νομικής ρύθμισης, διεθνούς συνεργασίας, ηθικής εποπτείας και άρτιας εκπαίδευσης χρηστών.

11.3 Προτάσεις για μελλοντική έρευνα

Μελλοντική έρευνα θα μπορούσε να εστιάσει στην αξιολόγηση της αποτελεσματικότητας των AI-assisted κυβερνοεπιθέσεων σε σύγκριση με τις παραδοσιακές μορφές επιθέσεων. Ιδιαίτερο ενδιαφέρον παρουσιάζει η μελέτη παραμέτρων όπως ο χρόνος εκτέλεσης, το κόστος, το ποσοστό επιτυχίας, η δυνατότητα ανίχνευσης και ο βαθμός αυτοματοποίησης. Παράλληλα, σημαντικό πεδίο αποτελεί η διερεύνηση των agentic AI συστημάτων και της πιθανής χρήσης τους σε κυβερνοεπιχειρήσεις.

Επιπλέον, απαιτείται περαιτέρω μελέτη της αμυντικής αξιοποίησης της τεχνητής νοημοσύνης, ιδίως σε συστήματα ανίχνευσης εισβολών, αυτοματοποιημένης απόκρισης, threat intelligence και εντοπισμού deepfakes. Η αξιολόγηση τέτοιων τεχνολογιών σε πραγματικά επιχειρησιακά περιβάλλοντα μπορεί να συμβάλει στην καλύτερη κατανόηση τόσο των δυνατοτήτων όσο και των περιορισμών τους, όπως τα false positives και η πιθανότητα χειραγώγησης των ίδιων των ΑΙ μοντέλων.

Τέλος, ιδιαίτερη σημασία έχει η μελλοντική έρευνα γύρω από τις κοινωνικές, ηθικές και νομικές συνέπειες των AI-driven κυβερνοεπιθέσεων. Η επίδραση των deepfakes, της συνθετικής φωνής, της κλοπής ταυτότητας και του προσωποποιημένου phishing στην εμπιστοσύνη των πολιτών και στη λειτουργία των οργανισμών αποτελεί κρίσιμο πεδίο μελέτης. Συνολικά, η αντιμετώπιση των

κινδύνων της ΑΙ στην κυβερνοασφάλεια απαιτεί συνδυασμό τεχνικής προστασίας, θεσμικής προσαρμογής, ηθικής ευθύνης και συνεχούς εκπαίδευσης.

12. ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] European Parliamentary Research Service. (2024). Artificial intelligence and cybersecurity. European Parliament. https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/762292/EPRS_ATA%282024%29762292_EN.pdf
- [2] ENISA. (2024). ENISA Threat Landscape 2024. European Union Agency for Cybersecurity. <https://www.enisa.europa.eu/publications/enisa-threat-landscape-2024#contentList>
- [3] Europol. (2024). Internet Organised Crime Threat Assessment (IOCTA) 2024. European Union Agency for Law Enforcement Cooperation. <https://www.europol.europa.eu/publication-events/main-reports/internet-organised-crime-threat-assessment-iocta-2024>
- [4] National Institute of Standards and Technology. (2023). Artificial Intelligence Risk Management Framework (AI RMF 1.0). U.S. Department of Commerce. <https://nvlpubs.nist.gov/nistpubs/AI/nist.AI.100-1.pdf>
- [5] European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. Official Journal of the European Union. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng>
- [6] Stanford University Human-Centered Artificial Intelligence (HAI). (2025). Artificial Intelligence Index Report 2025. Stanford University. <https://hai.stanford.edu/AI-index/2025-AI-index-report>
- [7] Tian, C., et al. (2025). A survey on deep learning fundamentals. Springer Nature. <https://link.springer.com/article/10.1007/s10462-025-11368-7>
- [8] National Institute of Standards and Technology (NIST). (2025). Adversarial Machine Learning: A Taxonomy and Terminology (AI 100-2 E2025). U.S. Department of Commerce. <https://csrc.nist.gov/pubs/AI/100/2/e2025/final>

- [9] Organisation for Economic Co-operation and Development (OECD). (2025). OECD AI Capability Indicators Technical Report. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/11/oecd-AI-capability-indicators-technical-report_d3762d1a/9cdb3dd1-en.pdf
- [10] Organisation for Economic Co-operation and Development (OECD). (2025). Mapping Relevant Data Collection Mechanisms for AI Training. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/10/mapping-relevant-data-collection-mechanisms-for-AI-training_62921889/3264cd4c-en.pdf
- [11] Klingelhöfer, D., et al. (2025). Research on artificial intelligence, machine learning and deep learning: A bibliometric and scientific overview. PMC. <https://pmc.ncbi.nlm.nih.gov/articles/PMC12147299/>
- [12] Fraunhofer Institute for Cognitive Systems (IKS). (2025). Künstliche Intelligenz und maschinelles Lernen. Fraunhofer IKS. <https://www.iks.fraunhofer.de/de/themen/kuenstliche-intelligenz.html>
- [13] Eisbein, H., et al. (2025). Grundlagen der Künstlichen Intelligenz und des Maschinellen Lernens. FIR e. V. an der RWTH Aachen. https://epub.fir.de/frontdoor/deliver/index/docId/4237/file/fir_Pickert_PrEvelOp_Schulungsunterlage_Grundlagen_ML_KI_2025.pdf
- [14] Ministère de l'Éducation Nationale / Direction du Numérique pour l'Éducation (MEN/DNE). (2025). Intelligence artificielle et éducation. République Française. https://edunumrech.hypotheses.org/files/2025/02/MEN_DNE_brochure_IA_2025_FR_web.pdf
- [15] Machine Learning and Deep Learning Paradigms: From Techniques to Practical Applications and Research Frontiers, Computers, MDPI <https://www.mdpi.com/2073-431X/14/3/93>

[16] A survey on state-of-the-art deep learning applications and challenges
<https://www.sciencedirect.com/science/article/abs/pii/S0952197625012266>

[17] Bojja (2025) – A Comparative Study of Supervised and Unsupervised Machine Learning Techniques <https://imrjr.com/wp-content/uploads/2025/04/IMRJR.2025.020411.pdf>

[18] Neto et al. (2025) – Supervised and unsupervised deep learning applied to the dynamics of complex systems <https://www.sciencedirect.com/science/article/abs/pii/S037843712500860X>

[19] Massachusetts Institute of Technology – Computer Science and Artificial Intelligence Laboratory (MIT CSAIL). (2025). Machine Learning Systems Review 2025. MIT.
<https://www.csail.mit.edu/>

[20] Elsevier. (2025). Neural Networks and Deep Learning Optimization Techniques. Artificial Intelligence Journal, Elsevier. <https://www.sciencedirect.com/journal/artificial-intelligence>

[21] European Union Agency for Cybersecurity (ENISA). (2025). ENISA Threat Landscape 2025. ENISA.
https://www.enisa.europa.eu/sites/default/files/2026-01/ENISA%20Threat%20Landscape%202025_v1.2.pdf

[22] Europol. (2025). Steal, Deal and Repeat: IOCTA 2025. European Union Agency for Law Enforcement Cooperation.
https://www.europol.europa.eu/cms/sites/default/files/documents/Steal-deal-repeat-IOCTA_2025.pdf

[23] Verizon. (2025). Data Breach Investigations Report 2025. Verizon Business.
<https://www.verizon.com/business/resources/T16f/reports/2025-dbir-data-breach-investigations-report.pdf>

[24] IBM X-Force. (2025). Threat Intelligence Index 2025. IBM Security.
<https://www.bosconsulting.lt/reports/ibm-x-force-threat-intelligence-index-2025-report.pdf>

[25] Google Cloud / Mandiant. (2025). M-Trends 2025 Report. Google Cloud.
<https://services.google.com/fh/files/misc/m-trends-2025-en.pdf>

[26] European Union Agency for Cybersecurity (ENISA). (2025). Sectorial Threat Landscape: Public Administration. ENISA. <https://www.enisa.europa.eu/sites/default/files/2025-12/ENISA%20Public%20Administration%20TL%202024%20-%20v1.1.pdf>

[27] World Economic Forum. (2025). Global Cybersecurity Outlook 2025. World Economic Forum. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2025.pdf

[28] UK Government. (2025). Cyber Security Breaches Survey 2025. UK Government. <https://www.gov.uk/government/statistics/cyber-security-breaches-survey-2025/cyber-security-breaches-survey-2025>

[29] Cybersecurity and Infrastructure Security Agency (CISA). (2025). #StopRansomware Guide. U.S. Department of Homeland Security. <https://www.cisa.gov/sites/default/files/2025-03/StopRansomware-Guide%20508.pdf>

[30] Europol. (2025). EU-SOCTA 2025: The Changing DNA of Serious and Organised Crime. European Union Agency for Law Enforcement Cooperation. <https://www.europol.europa.eu/cms/sites/default/files/documents/EU-SOCTA-2025.pdf>

[31] CrowdStrike. (2025). Global Threat Report 2025. CrowdStrike. <https://go.crowdstrike.com/rs/281-OBQ-266/images/CrowdStrikeGlobalThreatReport2025.pdf>

[32] Microsoft. (2025). Microsoft Digital Defense Report 2025. Microsoft Corporation. <https://cdn-dynmedia-1.microsoft.com/is/content/microsoftcorp/microsoft/msc/documents/presentations/CSR/Microsoft-Digital-Defense-Report-2025.pdf>

- [33] Fortinet. (2025). 2025 Global Threat Landscape Report. Fortinet. <https://www.fortinet.com/content/dam/fortinet/assets/threat-reports/threat-landscape-report-2025.pdf>
- [34] OpenText. (2025). 2025 Cybersecurity Threat Report. OpenText. <https://www.opentext.com/fr/media/report/2025-opentext-cybersecurity-threat-report-en.pdf>
- [35] Honeywell. (2025). 2025 Cyber Threat Report. Honeywell. <https://www.honeywell.com/content/dam/honeywellbt/en/documents/gated/hon-corp-honeywell-2025-cyber-threat-report.pdf>
- [36] National Cyber Security Centre New Zealand (NCSC). (2025). Cyber Threat Report 2025. Government of New Zealand. <https://www.ncsc.govt.nz/assets/insights/cyber-threat-report/NCSC-CyberReport2025-FINAL.pdf>
- [37] Abdullah, A. (2025). Evolution Cybercrime—Key Trends, Cybersecurity Threats, and Mitigation Strategies. MDPI. <https://www.mdpi.com/2813-2203/4/3/25>
- [38] Veprytska, I. (2025). Cybersecurity in the Age of Artificial Intelligence: A Literature Review. Communications in Applied Information Technologies. <https://reference-global.com/download/article/10.2478/cait-2025-0028.pdf>
- [39] KELA. (2025). 2025 AI Threat Report. KELA Cyber. <https://info.kela.com/hubfs/Reports/KELA%20Report%20-%202025%20AI%20Threat%20Report.pdf>
- [40] Anthropic. (2025). Threat Intelligence Report: August 2025. Anthropic. <https://www-cdn.anthropic.com/b2a76c6f6992465c09a6f2fce282f6c0cea8c200.pdf>
- [41] OpenAI. (2025). Disrupting Malicious Uses of AI: June 2025. OpenAI. <https://cdn.openai.com/threat-intelligence-reports/5f73af09-a3a3-4a55-992e-069237681620/disrupting-malicious-uses-of-AI-june-2025.pdf>

[42] RAND Corporation. (2025). Facing the Artificial Intelligence–Cyber Nexus. RAND. https://www.rand.org/content/dam/rand/pubs/research_reports/RAA4200/RAA4250-1/RAND_RRA4250-1.pdf

[43] United Nations Office on Drugs and Crime (UNODC). (2025). Emerging Threats: The Intersection of Criminal and Technological Innovation in the Use of Automation and AI. United Nations. https://www.unodc.org/roseap/uploads/documents/Publications/2025/UNODC_Report_Emerging_threats_-_The_intersection_of_criminal_and_technological_innovation_in_the_use_of_automation_and_AI.pdf

[44] Alan Turing Institute / Centre for Emerging Technology and Security (CETaS). (2025). AI and Serious Online Crime. Alan Turing Institute. https://cetas.turing.ac.uk/sites/default/files/2025-03/cetas_research_report_-_AI_and_serious_online_crime_0.pdf

[45] World Economic Forum. (2025). Artificial Intelligence and Cybersecurity: Balancing Risks and Rewards. World Economic Forum. https://reports.weforum.org/docs/WEF_Artificial_Intelligence_and_Cybersecurity_Balancing_Risks_and_Rewards_2025.pdf

[46] Zhang, Y., et al. (2025). A Survey on Large Language Model-based Agents in Cybersecurity. arXiv. <https://arxiv.org/html/2505.12786v2>

[47] Jaffal, M., et al. (2025). Large Language Models in Cybersecurity: A Survey of Applications, Vulnerabilities, and Defense Techniques. MDPI. <https://www.mdpi.com/2673-2688/6/9/216>

[48] ESET. (2025). ESET Threat Report H2 2025. ESET Research. <https://web-assets.esetstatic.com/wls/en/papers/threat-reports/eset-threat-report-h22025.pdf>

[49] The Impact of AI on the Cyber Offense-Defense Balance and the Character of Cyber Conflict. (2025). arXiv. <https://arxiv.org/html/2504.13371v1>

[50] International AI Safety Report. (2025). International AI Safety Report 2025. International AI Safety Institute. <https://internationalaisafetyreport.org/publication/international-AI-safety-report-2025>

[51] European Payments Council. (2025). 2025 Payments Threats and Fraud Trends Report. European Payments Council. https://www.europeanpaymentscouncil.eu/sites/default/files/kb/file/2025-12/EPC162-24%20v2.0%202025%20Payments%20Threats%20and%20Fraud%20Trends%20Report_0.pdf

[52] Ibrar, M., et al. (2025). Generative AI: A Double-Edged Sword in the Cyber Threat Landscape. Springer. <https://link.springer.com/article/10.1007/s10462-025-11285-9>

[53] Jabir, M., et al. (2025). Phishing Attacks in the Age of Generative Artificial Intelligence. MDPI. <https://www.mdpi.com/2673-2688/6/8/174>

[54] Ferrag, M. A., et al. (2025). Generative AI in Cybersecurity: A Comprehensive Review of LLM Applications and Vulnerabilities. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S2667345225000082>

[55] Babenko, O., et al. (2025). Automated OSINT Techniques for Digital Asset Discovery. MDPI Computers. <https://www.mdpi.com/2073-431X/14/10/430>

[56] CAI Team. (2025). CAI: An Open, Bug Bounty-Ready Cybersecurity AI. arXiv. <https://arxiv.org/html/2504.06017v1>

[57] Arazzi, M., et al. (2025). NLP-based Techniques for Cyber Threat Intelligence. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S1574013725000413>

[58] Santos, R., et al. (2025). A Systematic Review of Cyber Threat Intelligence. MDPI Sensors. <https://www.mdpi.com/1424-8220/25/14/4272>

[59] ACM. (2025). Weaponizing AI in Cyberattacks: A Comparative Study of AI Tools in Kali Linux. Association for Computing Machinery. <https://dl.acm.org/doi/10.1145/3726122.3726164>

[60] Google Threat Intelligence Group. (2025). Advances in Threat Actor Usage of AI Tools. Google Cloud. <https://cloud.google.com/blog/topics/threat-intelligence/threat-actor-usage-of-AI-tools>

[61] Yan, X., et al. (2025). MGC: A Compiler Framework Exploiting Compositional Blindness in Aligned LLMs for Malware Generation. arXiv. <https://arxiv.org/abs/2507.02057>

[62] Morris, J. (2025). Detecting Generative-AI-Enabled Polymorphic Malware. Old Dominion University. <https://digitalcommons.odu.edu/cgi/viewcontent.cgi?article=1134&context=covacci-undergraduateresearch>

[63] Schmitt, M. (2024/2025). Digital Deception: Generative Artificial Intelligence in Social Engineering. Springer. <https://link.springer.com/article/10.1007/s10462-024-10973-2>

[64] Green, C., & Alam, A. (2025). Utilising Generative AI To Improve Practical Spear Phishing Training. <https://eprints.glos.ac.uk/15561/1/15581%20Green%2C%20C%20and%20Alam%2C%20A%20%282025%29%20%20Utilising%20Generative%20AI%20To%20Improve%20Practical%20Spear%20Phishing%20Training.pdf>

[65] Pindrop. (2025). Voice Intelligence and Security Report. Pindrop. <https://www.pindrop.com/wp-content/uploads/2025/07/Voice-Intelligence-and-Security-Report-Report-June-2025.pdf>

- [66] Resemble AI. (2025). Q1 2025 Deepfake Incident Report. Resemble AI. <https://39879154.fs1.hubspotusercontent-na1.net/hubfs/39879154/ResembleAI-Q1-Deepfake-Threats.pdf>
- [67] TitanHQ. (2025). The State of email Security in 2025. TitanHQ. https://www.titanhq.com/documentation/TitanHQ_State_of_email_Security_in_2025_final_1.pdf
- [68] CrowdStrike. (2025). How GenAI Powers Social Engineering / Global Threat Report Spotlight. CrowdStrike. <https://www.crowdstrike.com/en-us/resources/articles/crowdstrike-2025-global-threat-report-genai-powers-social-engineering>
- [69] Koch, R. (2026). Agentic AI and the Industrialization of Cyber Offense. arXiv. <https://arxiv.org/pdf/2605.06713>
- [70] Mauko, A., et al. (2026). Cybercrime as a Service: A Scoping Review. arXiv. <https://arxiv.org/pdf/2604.00063>
- [71] Flare. (2026). The Phishing Kits Economy in Cybercrime Markets. Flare. <https://flare.io/learn/resources/phishing-kits-economy-cybercrime>
- [72] Cyble. (2025). Global Cybersecurity Report 2025. Cyble. <https://cyble.com/wp-content/uploads/2025/12/Global-Cybersecurity-Report-2025.pdf>
- [73] Ontinue. (2026). 2H 2025 Threat Intelligence Report. Ontinue. https://www.ontinue.com/wp-content/uploads/2026/03/2026_2H2025-Threat-Intelligence-Report.pdf
- [74] AI-driven Attack Research Topics. (2026). SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=6364978

- [75] Ensign. (2025). Cyber Threat Landscape Report 2025. Ensign. <https://raw.githubusercontent.com/jacobdjwilson/awesome-annual-security-reports/main/Annual%20Security%20Reports/2025/Ensign-Cyber-Threat-Landscape-Report-2025.pdf>
- [76] Rohozinski, R. (2026). Weaponising AI: The New Cyber Attack Surface. Taylor & Francis. <https://www.tandfonline.com/doi/full/10.1080/00396338.2026.2620282>
- [77] NETSCOUT. (2026). 2H 2025 DDoS Threat Intelligence Report. NETSCOUT. <https://www.netscout.com/threatreport/wp-content/uploads/2026/03/NETSCOUT-2H-2025-Threat-Report.pdf>
- [78] Cloudflare. (2026). Cloudflare Threat Report 2026. Cloudflare. <https://cf-assets.www.cloudflare.com/slt3lc6tev37/sWDBUMNVtEJB9ZFLt1dUU/8d69e92de2edfb3bf59e7d21d57e7e1a/Cloudflare-2026-threat-report.pdf>
- [79] Somma, G., et al. (2026). Environment-Grounded Multi-Agent Workflow for Autonomous Penetration Testing. arXiv. <https://arxiv.org/abs/2603.24221>
- [80] Lin, J., et al. (2025). Comparing AI Agents to Cybersecurity Professionals in Real-World Networks. arXiv. <https://arxiv.org/html/2512.09882v1>
- [81] Palo Alto Networks Unit 42. (2026). Can AI Attack the Cloud? Lessons From Building an Autonomous AI Cloud Attack Agent. Unit 42. <https://unit42.paloaltonetworks.com/autonomous-AI-cloud-attacks/>
- [82] Recorded Future. (2026). 2025 Cloud Threat Hunting and Defense Landscape. Recorded Future. <https://www.recordedfuture.com/research/2025-cloud-threat-hunting-defense-landscape>

[83] International AI Safety Report. (2026). International AI Safety Report 2026. International AI Safety Report. <https://internationalaisafetyreport.org/sites/default/files/2026-02/international-AI-safety-report-2026.pdf>

[84] Federation of American Scientists / Future of Life Institute. (2026). Artificial Intelligence, Cyber, and Global Risk. FAS / Future of Life Institute. <https://fas.org/wp-content/uploads/2026/01/January-2026-AI-Cyber-Global-Risk.pdf>

[85] CSET. Lohn, A. J. (2025). Anticipating AI's Impact on the Cyber Offense-Defense Balance. <https://cset.georgetown.edu/wp-content/uploads/CSET-Anticipating-AIs-Impact-on-the-Cyber-Offense-Defense-Balance.pdf>

[86] Center for a New American Security (CNAS). (2025). Tipping the Scales: AI, Cybersecurity, and the Offense-Defense Balance. CNAS. https://s3.us-east-1.amazonaws.com/files.cnas.org/documents/Report_Tipping-the-Scales_TECH_Sep-2025-Final.pdf

[87] Zhu, Y., et al. (2026). Teams of LLM Agents can Exploit Zero-Day Vulnerabilities. ACL Anthology. <https://aclanthology.org/2026.eacl-long.2.pdf>

[88] Vangeli, A., et al. (2026). Context Relay for Long-Running Penetration-Testing Agents. NDSS Symposium. <https://www.ndss-symposium.org/wp-content/uploads/lastx2026-42.pdf>

[89] Fernandes, R., et al. (2026). Autonomous pentesting using artificial intelligence. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S2542660526000624>

[90] Context-Aware Spear Phishing: Generative AI-Enabled Social Engineering and Mass Personalization. (2026). arXiv. <https://arxiv.org/html/2605.11268v1>

[91] UK Government. (2025). Safety and security risks of generative AI to 2025: Annex B. UK Government. <https://www.gov.uk/government/publications/frontier-AI-capabilities-and-risks-discussion-paper/safety-and-security-risks-of-generative-artificial-intelligence-to-2025-annex-b>

[92] A Systematic Review of Artificial Intelligence (AI)-Enabled Cybercrime and Self-Directed Attack Systems. (2026). ResearchGate. https://www.researchgate.net/publication/400864975_A_Systematic_Review_of_Artificial_Intelligence_AI-Enabled_Cybercrime_and_Self-Directed_Attack_Systems

[93] NCSC / NSA / CISA et al. (2026). Careful Adoption of Agentic AI Services. NCSC / NSA / CISA. https://media.defense.gov/2026/Apr/30/2003922823/-1/-1/0/CAREFUL%20ADOPTION%20OF%20AGENTIC%20AI%20SERVICES_FINAL.PDF

[94] KPMG. (2026). Cybersecurity Considerations 2026. KPMG. <https://assets.kpmg.com/content/dam/kpmgsites/xx/pdf/2026/03/cybersecurity-considerations-2026.pdf>

[95] Trend Micro. (2024). Deepfake CFO Video Calls Result in \$25MM in Damages. https://www.trendmicro.com/en_us/research/24/b/deepfake-video-calls.html

[96] Google Threat Intelligence Group. (2025). Advances in Threat Actor Usage of AI Tools. <https://cloud.google.com/blog/topics/threat-intelligence/adversarial-misuse-generative-AI>

[97] Anthropic. (2025). Threat Intelligence Report: August 2025. <https://www.anthropic.com/news/disrupting-AI-espionage>

[98] Mohamed, N. (2025). Artificial intelligence and machine learning in cybersecurity. Springer. <https://link.springer.com/article/10.1007/s10115-025-02429-y>

[99] European Union Agency for Cybersecurity (ENISA). (2025). ENISA Advisory Group Opinion Paper on Artificial Intelligence. ENISA.

<https://www.enisa.europa.eu/sites/default/files/2025-07/AG%20opinion%20paper%20on%20AI.pdf>

[100] Google Cloud Security. (2025). AI-Powered Threat Detection and Response. https://www.researchgate.net/publication/389435965_AI-Powered_Threat_Detection_and_Response_in_Multi-Cloud_Environments

[101] CrowdStrike. (2025). AI-Powered Behavioral Analysis in Cybersecurity. CrowdStrike. <https://www.crowdstrike.com/en-us/cybersecurity-101/artificial-intelligence/AI-powered-behavioral-analysis/>

[102] Palo Alto Networks. (2025). AI-Driven Security Operations. Palo Alto Networks. <https://www.paloaltonetworks.com/blog/security-operations/2025-the-year-of-the-autonomous-soc-the-year-of-xsiam/>

[103] ENISA. (2025). Artificial Intelligence and Cybersecurity. <https://www.enisa.europa.eu/topics/artificial-intelligence-and-next-gen-technologies>

[104] Google Cloud Security. (2025). AI-Powered Security Operations and Threat Detection. <https://cloud.google.com/security>

[105] IBM Security. (2025). AI and Machine Learning in Cybersecurity. <https://www.ibm.com/think/topics/AI-cybersecurity>

[106] Splunk. (2025). The State of Security Operations. https://www.splunk.com/en_us/form/state-of-security.html

[107] Cisco. (2025). AI in Security Operations and Threat Detection. <https://www.cisco.com/site/us/en/products/security/artificial-intelligence-security/index.html>

[108] IBM. (2025). Cost of a Data Breach Report 2025. <https://www.ibm.com/reports/data-breach>

[109] Gartner. (2025). Security Orchestration, Automation and Response (SOAR) Solutions. <https://www.gartner.com/reviews/market/security-orchestration-automation-and-response-solutions>

[110] IBM. (2025). SOAR: Security Orchestration, Automation and Response Explained. <https://www.ibm.com/think/topics/security-orchestration-automation-response>

[111] ENISA. (2025). Artificial Intelligence and Cybersecurity. <https://www.enisa.europa.eu/topics/artificial-intelligence-and-next-gen-technologies>

[112] Microsoft Azure. (2025). What is AI as a Service (AIaaS)? <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-AIaaS>

[113] IBM. (2025). AI as a Service (AIaaS): Benefits, Challenges and Applications. <https://www.ibm.com/think/topics/AI-as-a-service-AIaaS>

[114] Google Cloud. (2025). Artificial Intelligence and Machine Learning Solutions. <https://cloud.google.com/AI>

[115] Kaspersky. (2024). Cybercrime AI Experimentation in the Dark Web. <https://www.kaspersky.co.uk/about/press-releases/cybercrime-AI-experimentation-in-the-dark-web-new-kaspersky-study>

[116] Sophos. (2025). Cybercriminals Can't Agree on GPTs. <https://www.sophos.com/en-us/blog/cybercriminals-cant-agree-on-gpts>

[117] WIRED. (2025). Criminals Have Created Their Own ChatGPT Clones. <https://www.wired.com/story/chatgpt-scams-fraudgpt-wormgpt-crime>

[118] Dark Reading. (2025). FraudGPT: Malicious Chatbot Now for Sale on the Dark Web. <https://www.darkreading.com/threat-intelligence/fraudgpt-malicious-chatbot-for-sale-dark-web>

[119] Infosecurity Europe. (2025). The Dark Side of Generative AI: Five Malicious LLMs. <https://www.infosecurityeurope.com/en-gb/blog/threat-vectors/generative-AI-dark-web-bots.html>

[120] TechRadar. (2025). Malicious LLMs Are Letting Even Unskilled Hackers Craft Dangerous Malware. <https://www.techradar.com/pro/security/malicious-llms-are-letting-even-unskilled-hackers-to-craft-dangerous-new-malware>

[121] SiliconANGLE. (2025). Xanthorox AI Emerges as a New Malicious Threat in Cybercrime Communities. <https://siliconangle.com/2025/04/07/xanthorox-AI-emerges-new-malicious-threat-cybercrime-communities>

[122] WIRED. (2025). The Ultra-Realistic AI Face Swapping Platform Driving Romance Scams. <https://www.wired.com/story/the-ultra-realistic-AI-face-swapping-platform-driving-romance-scams>

[123] IT Voice. (2025). Kaspersky: Deepfake Services Now 400 Times Cheaper. <https://www.itvoice.in/kaspersky-deepfake-services-now-400-times-cheaper>

[124] Varonis / SlashNext. (2025). WormGPT: The Generative AI Tool Cybercriminals Are Using to Launch Business Email Compromise Attacks. <https://www.varonis.com/blog/wormgpt>

[125] Kaspersky. (2025). WormGPT-Mimicking Phishing Scams Surface on the Darknet. <https://usa.kaspersky.com/about/press-releases/wormgpt-mimicking-phishing-scams-surface-on-the-darknet>

[126] Netenrich. (2023). FraudGPT: The Villain Avatar of ChatGPT. <https://netenrich.com/blog/fraudgpt-the-villain-avatar-of-chatgpt>

[127] Barracuda Networks. (2025). DarkBard: The “Evil Twin” of Google Bard. <https://blog.barracuda.com/2025/09/04/darkbard-evil-twin-google-bard>

[128] World Economic Forum. (2025). Cybercrime Atlas: Impact Report 2025. https://reports.weforum.org/docs/WEF_Cybercrime_Atlas_Impact_Report_2025.pdf

[129] Group-IB. (2025). High-Tech Crime Trends Report 2025. <https://www.group-ib.com/landing/high-tech-crime-trends-2025/>

[130] Cyberint. (2025). Initial Access Brokers Report. <https://e.cyberint.com/hubfs/IAB%20Report%202025.pdf>

[131] Akamai. (2025). Ransomware Report 2025. <https://www.akamai.com/site/en/documents/state-of-the-internet/2025/ransomware-trends-2025.pdf>

[132] chainalysis. (2025). The 2025 Crypto Crime Report. <https://go.chainalysis.com/2025-Crypto-Crime-Report.html>

[133] chainalysis. (2026). Crypto Ransomware: 2026 Crypto Crime Report. <https://www.chainalysis.com/blog/crypto-ransomware-2026/>

[134] chainalysis. (2026). 2026 Crypto Crime Report: Scams. <https://www.chainalysis.com/blog/crypto-scams-2026/>

[135] Europol. (2025). Steal, Deal and Repeat: IOCTA 2025. <https://www.europol.europa.eu/publications-events/main-reports/iocta-report>

[136] UNODC. (2025). The Nexus Between Cybercrime and Corruption. https://www.unodc.org/roseap/uploads/documents/Publications/2025/2025.10.21_The_Nexus_Between_Cybercrime_and_Corruption.pdf

[137] UNODC. (2025). Inflection Point: Global Implications of Scam Centres, Underground Banking and Illicit Online Marketplaces in Southeast Asia. https://www.unodc.org/roseap/uploads/documents/Publications/2025/Inflection_Point_2025.pdf

[138] INTERPOL. (2025). Africa Cyberthreat Assessment Report 2025. <https://www.interpol.int/content/download/23222/file/2025%20Africa%20Cyberthreat%20Assessment%20Report.pdf>

[139] INTERPOL. (2025/2026). Asia and South Pacific Cyber Threat Assessment Report 2025/2026.

https://www.interpol.int/content/download/24327/file/CYBER_ASP%20Cyber%20Threat%20Assessment%20Report_2025_2026_v4.pdf

[140] Council of Europe. (2025). Cybercrime, e-Evidence and AI. <https://www.coe.int/en/web/cybercrime/main-session-4>

[141] Europol Innovation Lab. (2024). AI and Policing: The Benefits and Challenges of Artificial Intelligence for Law Enforcement. <https://www.europol.europa.eu/cms/sites/default/files/documents/AI-and-policing.pdf>

[142] NIST. Guan, H., Horan, J., & Zhang, A. (2025). Guardians of Forensic Evidence: Evaluating Analytic Systems Against AI-Generated Deepfakes. <https://www.nist.gov/publications/guardians-forensic-evidence-evaluating-analytic-systems-against-AI-generated-deepfakes>

[143] UK Government. (2026). Deepfake Detection Technology. <https://www.gov.uk/government/publications/deepfake-detection-technology/deepfake-detection-technology>

[144] Romero-Moreno, F. (2025). Deepfake Detection in Generative AI: A Legal Framework for Forensic and Evidentiary Reliability. ScienceDirect. <https://www.sciencedirect.com/science/article/pii/S2212473X25000355>

[145] Lin, L. S. F. (2025). Organisational Challenges in US Law Enforcement's Response to AI-Driven Cybercrime and Deepfake Fraud. Laws, MDPI. <https://www.mdpi.com/2075-471X/14/4/46>

[146] NSA / CISA / FBI / International Partners. (2025). Strengthening Multimedia Integrity in the Generative AI Era. <https://media.defense.gov/2025/Jan/29/2003634788/-1/-1/0/CSI-CONTENT-CREDENTIALS.PDF>

[147] European Commission / Council of the EU. (2025). Proposal concerning the United Nations Convention against Cybercrime. <https://data.consilium.europa.eu/doc/document/ST-11553-2025-INIT/en/pdf>

[148] European Commission / Council of the EU. (2025). Roadmap on Lawful and Effective Access to Data for Law Enforcement. <https://data.consilium.europa.eu/doc/document/ST-11556-2025-INIT/en/pdf>

[149] Chlasta, K. (2026). The Dual-Use Dilemma of Generative Artificial Intelligence in Cybersecurity: Navigating the Explosive Growth in Offensive and Defensive Applications. <https://securityanddefence.pl/The-dual-use-dilemma-of-generative-artificial-intelligence-in-cybersecurity-Navigating%2C217364%2C0%2C2.html>

[150] Kemp, M., Kalutarage, H., & Al-Kadri, M. O. (2025). AI-Powered Spearphishing Cyber Attacks: Fact or Fiction? <https://arxiv.org/abs/2502.00961>

[151] Macron, T. (2025). Generative AI and Cybersecurity: Analyzing the Dual Use of Deepfake Technology for Threats and Defensive Measures.

https://www.researchgate.net/publication/388323713_Generative_AI_and_Cybersecurity_Analyzing_the_Dual_Use_of_Deepfake_Technology_for_Threats_and_Defensive_Measures

[152] Ditz, J. C., et al. (2025). Secure Human Oversight of AI: Threat Modeling in a Socio-Technical Context. <https://arxiv.org/abs/2509.12290>

[153] Mohsin, A., et al. (2025). A Unified Framework for Human-AI Collaboration in Security Operations Centers with Trusted Autonomy. <https://arxiv.org/abs/2505.23397>

[154] TechScience. (2025). Ethical Implications of AI-Driven Ethical Hacking: A Systematic Review and Governance Framework. <https://www.techscience.com/JCS/v7n1/62970/html>

[155] Gunasekara, L., et al. (2025). A Systematic Review of Responsible Artificial Intelligence Principles and Practice. MDPI. <https://www.mdpi.com/2571-5577/8/4/97>

[156] European Commission. (2025). AI Act — Shaping Europe’s Digital Future. <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-AI>

[157] European Union. (2016). Regulation (EU) 2016/679 — General Data Protection Regulation (GDPR). EUR-Lex. <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>

[158] European Commission. (2025). Information for Individuals — Data Protection under GDPR. https://commission.europa.eu/law/law-topic/data-protection/information-individuals_en

[159] European Data Protection Board. (2018). Guidelines on Automated Individual Decision-Making and Profiling. https://www.edpb.europa.eu/our-work-tools/our-documents/guidelines/automated-decision-making-and-profiling_en

[160] European Union. (2022). Directive (EU) 2022/2555 — NIS2 Directive. EUR-Lex. <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>

[161] European Commission. (2026). NIS2 Directive: Securing Network and Information Systems. <https://digital-strategy.ec.europa.eu/en/policies/nis2-directive>

[162] European Union. (2024). Regulation (EU) 2024/2847 — Cyber Resilience Act. EUR-Lex. <https://eur-lex.europa.eu/eli/reg/2024/2847/oj/eng>

[163] European Commission. (2025). Cyber Resilience Act. <https://digital-strategy.ec.europa.eu/en/policies/cyber-resilience-act>

[164] European Commission. (2025). The Cyber Resilience Act — Summary of the Legislative Text. <https://digital-strategy.ec.europa.eu/en/policies/cra-summary>

[165] Sousa e Silva, N. (2024). The Artificial Intelligence Act: Critical Overview. arXiv. <https://arxiv.org/abs/2409.00264>

[166] Ho-Dac, M. (2024). First Analysis of the EU Artificial Intelligence Act: Towards a Global Standard for Trustworthy AI? arXiv. <https://arxiv.org/abs/2408.08318>

[167] Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. <https://www.coe.int/en/web/artificial-intelligence/the-framework-convention-on-artificial-intelligence>

[168] Council of Europe. (2024). CETS No. 225 — Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. <https://rm.coe.int/1680afae3c>

[169] Council of Europe. (2024). Council of Europe Opens First Ever Global Treaty on AI for Signature. <https://www.coe.int/en/web/portal/-/council-of-europe-opens-first-ever-global-treaty-on-ai-for-signature>

[170] UNODC. (2025). United Nations Convention against Cybercrime. <https://www.unodc.org/unodc/en/cybercrime/convention/home.html>

[171] UNODC. (2025). UN Convention against Cybercrime Opens for Signature in Hanoi, Viet Nam. <https://www.unodc.org/unodc/en/press/releases/2025/October/un-convention-against-cybercrime-opens-for-signature-in-hanoi--viet-nam.html>

[172] UNODC. (2025). Pre-ratification Compendium — United Nations Convention against Cybercrime. <https://www.unodc.org/unodc/en/cybercrime/convention/pre-ratification-compendium.html>

[173] OECD. (2024/2025). OECD AI Principles. <https://www.oecd.org/en/topics/sub-issues/AI-principles.html>

[174] OECD. (2024). Recommendation of the Council on Artificial Intelligence. <https://legalinstruments.oecd.org/en/instruments/oecd-legal-0449>

[175] European Commission. (2023/2025). G7 Leaders' Statement on the Hiroshima AI Process. <https://digital-strategy.ec.europa.eu/en/library/g7-leaders-statement-hiroshima-AI-process>

[176] OECD. (2025). OECD Launches Global Framework to Monitor Application of G7 Hiroshima AI Code of Conduct. <https://www.oecd.org/en/about/news/press-releases/2025/02/oecd-launches-global-framework-to-monitor-application-of-g7-hiroshima-AI-code-of-conduct.html>

[177] Habuka, H. (2025). Enhancements and Next Steps for the G7 Hiroshima AI Process: Toward a Common Framework to Advance Human Rights, Democracy and Rule of Law. Cambridge Forum on AI: Law and Governance. <https://www.cambridge.org/core/journals/cambridge-forum-on-AI-law-and-governance/article/enhancements-and-next-steps-for-the-g7-hiroshima-AI-process-toward-a-common-framework-to-advance-human-rights-democracy-and-rule-of-law>

[178] Council of Europe. (2025). Budapest Convention on Cybercrime. <https://www.coe.int/en/web/cybercrime/the-budapest-convention>

[179] Eurojust. (2024). Article 9 of the Second Additional Protocol to the Convention on Cybercrime: Expedited Disclosure of Stored Computer Data. <https://www.eurojust.europa.eu/publication/article-9-second-additional-protocol-convention-cybercrime-expedited-disclosure-stored-computer-data>

[180] Council of Europe / EUCrim. (2022). Second Additional Protocol to the Cybercrime Convention. <https://eucrim.eu/news/council-of-europe-second-additional-protocol-to-the-cybercrime-convention/>

[181] FIRST. (2025/2026). Forum of Incident Response and Security Teams. <https://www.first.org/>

[182] FBI Internet Crime Complaint Center. (2026). 2025 Internet Crime Report. https://www.ic3.gov/AnnualReport/Reports/2025_IC3Report.pdf

[183] chainalysis. (2026). The 2026 Crypto Crime Report. <https://www.chainalysis.com/reports/crypto-crime-2026/>

[184] Resemble AI. (2026). 2025 Deepfake Threat Report. <https://www.resemble.ai/resources/2025-deepfake-threat-report>

[185] Australian Competition and Consumer Commission / National Anti-Scam Centre. (2026). Targeting Scams Report 2025. <https://www.scamwatch.gov.au/research-and-resources/targeting-scams-report>

[186] World Economic Forum. (2026). Global Cybersecurity Outlook 2026. https://reports.weforum.org/docs/WEF_Global_Cybersecurity_Outlook_2026.pdf

[187] Allianz Commercial. (2026). Allianz Risk Barometer 2026 — Cyber Incidents. <https://commercial.allianz.com/news-and-insights/expert-risk-articles/allianz-risk-barometer-2026-cyber-incidents.html>

[188] ENISA. (2025). ENISA Threat Landscape 2025. <https://www.enisa.europa.eu/sites/default/files/2025-11/ENISA%20Threat%20Landscape%202025.pdf>

[189] Lukošiuūtė, K., Halstead, J., & Righetti, L. (2026). Global Cybercrime Damages: A Baseline for Frontier AI Risk Assessment. arXiv. <https://arxiv.org/abs/2603.20570>

[190] Elmisery, A. M., Sertovic, M., Zayin, A., & Watson, P. (2025). Cyber Threats in Financial Transactions — Addressing the Dual Challenge of AI and Quantum Computing. arXiv. <https://arxiv.org/abs/2503.15678>

[191] Hever, A., & Orr, I. (2025). A New Wave of Vehicle Insurance Fraud Fueled by Generative AI. arXiv. <https://arxiv.org/abs/2510.19957>

[192] NCSC. (2025). NCSC Annual Review 2025. <https://www.ncsc.gov.uk/files/ncsc-annual-review-2025.pdf>

[193] Canadian Centre for Cyber Security. (2024). National Cyber Threat Assessment 2025–2026. <https://www.cyber.gc.ca/en/guidance/national-cyber-threat-assessment-2025-2026>

[194] Identity Theft Resource Center. (2025). 2025 Consumer Impact Report. <https://www.idtheftcenter.org/post/2025-consumer-impact-report-financial-emotional-impacts-rise/>

[195] Federal Trade Commission. (2026). New FTC Data Show People Have Lost Billions to Social Media Scams. <https://www.ftc.gov/news-events/news/press-releases/2026/04/new-ftc-data-show-people-have-lost-billions-social-media-scams>

- [196] Bhatti, Z. H., et al. (2026). Can You Tell It's AI? Human Perception of Synthetic Voices in Vishing Scenarios. <https://arxiv.org/abs/2602.20061>
- [197] LaRubbio, K., et al. (2025). Intergenerational Support for Deepfake Scams Targeting Older Adults. <https://arxiv.org/abs/2508.11579>
- [198] Popa, C., et al. (2025). Deepfake Technology Unveiled: The Commoditization of AI and Its Impact on Digital Trust. <https://arxiv.org/abs/2506.07363>
- [199] Yigit, Y., et al. (2024). Critical Infrastructure Protection: Generative AI, Challenges, and Opportunities. <https://arxiv.org/abs/2405.04874>
- [200] CSIS. (2025). Significant Cyber Incidents. <https://www.csis.org/programs/strategic-technologies-program/significant-cyber-incidents>
- [201] NIST. (2025). Cybersecurity Framework Profile for Artificial Intelligence. <https://nvlpubs.nist.gov/nistpubs/ir/2025/NIST.IR.8596.iprd.pdf>
- [202] NIST. (2025). Implementing a Zero Trust Architecture: High-Level Document. <https://www.nist.gov/publications/implementing-zero-trust-architecture-high-level-document>
- [203] NIST. (2025). NIST Offers 19 Ways to Build Zero Trust Architectures. <https://www.nist.gov/news-events/news/2025/06/nist-offers-19-ways-build-zero-trust-architectures>
- [204] CISA. (2025). Zero Trust Maturity Model. <https://www.cisa.gov/zero-trust-maturity-model>
- [205] Accenture. (2025). State of Cybersecurity Resilience 2025. <https://www.accenture.com/content/dam/accenture/final/accenture-com/document-3/State-of-Cybersecurity-report.pdf>

- [206] Nott, C. (2025). Organizational Adaptation to Generative AI in Cybersecurity: A Systematic Review. <https://arxiv.org/abs/2506.12060>
- [207] Nasiruzzaman, M., Ali, M., Salam, I., & Miraz, M. H. (2025). The Evolution of Zero Trust Architecture from Concept to Implementation. <https://arxiv.org/abs/2504.11984>
- [208] Lund, B. D., Lee, T. H., Wang, Z., Wang, T., & Mannuru, N. R. (2025). Zero Trust Cybersecurity: Procedures and Considerations in Context. <https://arxiv.org/abs/2505.18872>
- [209] Aiello, S. (2025). Prescriptive Zero Trust: Assessing the Impact of Zero Trust on Cyber Attack Prevention. <https://arxiv.org/abs/2508.12953>
- [210] MITRE. (2025/2026). MITRE ATLAS. <https://atlas.mitre.org/>
- [211] OWASP. (2025). OWASP Top 10 for Large Language Model Applications 2025. <https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/>
- [212] Erukude, S. T., Marella, V. C., & Veluru, S. R. (2026). AI-Driven Cybersecurity Threats: A Survey of Emerging Risks and Defensive Strategies. <https://arxiv.org/abs/2601.03304>
- [213] Krawczyk, N., Szczepkowski, M., Brodzik, A., & Bocianiak, K. (2026). Cyber Threat Intelligence for Artificial Intelligence Systems. <https://arxiv.org/abs/2603.05068>
- [214] World Economic Forum. (2026). Empowering Defenders: AI for Cybersecurity. https://reports.weforum.org/docs/WEF_Empowering_Defenders_AI_for_Cybersecurity_2026.pdf
- [215] European Commission. (2025/2026). Timeline for the Implementation of the EU AI Act. <https://ai-act-service-desk.ec.europa.eu/en/ai-act/timeline/timeline-implementation-eu-ai-act>

[216] Nolte, H., Rateike, M., & Finck, M. (2025). Robustness and Cybersecurity in the EU Artificial Intelligence Act. <https://arxiv.org/abs/2502.16184>

13. ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ – ΞΕΝΟΓΛΩΣΣΟΙ ΟΡΟΙ

Όρος / συντομογραφία	Σύντομη ελληνική απόδοση
AI Act (EU)	Κανονισμός της Ευρωπαϊκής Ένωσης για την Τεχνητή Νοημοσύνη
AI as a Service (AIaaS)	Τεχνητή νοημοσύνη ως υπηρεσία μέσω υπολογιστικού νέφους
AI-assisted malware creation	Δημιουργία κακόβουλου λογισμικού με υποβοήθηση AI
AI-assisted offensive operations	Επιθετικές κυβερνοεπιχειρήσεις με υποστήριξη AI
AI-assisted threat intelligence	Ανάλυση πληροφοριών απειλών με υποβοήθηση AI
AI-assisted vulnerability discovery	Εντοπισμός ευπαθειών με υποβοήθηση AI
AI-driven phishing campaigns	Εκστρατείες phishing που σχεδιάζονται ή ενισχύονται από AI
AI-driven reconnaissance	Συλλογή πληροφοριών στόχου με χρήση AI
AI-enabled bots	Αυτοματοποιημένα bots ενισχυμένα με AI
AI-enabled tools / platforms	Εργαλεία ή πλατφόρμες που αξιοποιούν AI
AI-enhanced IDS	Συστήματα ανίχνευσης εισβολών ενισχυμένα με AI
Adaptive malware	Προσαρμοστικό κακόβουλο λογισμικό που μεταβάλλει τη συμπεριφορά του

Advanced Persistent Threats (APTs)	Στοχευμένες, μακροχρόνιες και εξελιγμένες κυβερνοεπιθέσεις
Adversarial Machine Learning	Επιθέσεις χειραγώγησης ή παραπλάνησης μοντέλων μηχανικής μάθησης
Alert triage	Αρχική αξιολόγηση και προτεραιοποίηση ειδοποιήσεων ασφαλείας
Anomaly detection	Ανίχνευση ασυνήθιστης ή ύποπτης συμπεριφοράς
Antivirus	Λογισμικό ανίχνευσης και αντιμετώπισης κακόβουλου λογισμικού
Asset discovery	Εντοπισμός ψηφιακών πόρων ενός οργανισμού ή δικτύου
Attack AI / Defence AI	Χρήση AI για επιθέσεις έναντι χρήσης AI για κυβερνοάμυνα
Attack compression	Επιτάχυνση και συντόμευση των σταδίων μιας κυβερνοεπίθεσης
Automated asset discovery	Αυτοματοποιημένος εντοπισμός ψηφιακών πόρων
Automated code generation	Αυτοματοποιημένη παραγωγή κώδικα
Automated malware generation	Αυτοματοποιημένη δημιουργία κακόβουλου λογισμικού
Automated scripting	Αυτοματοποιημένη δημιουργία ή εκτέλεση scripts
Autonomous AI cloud attack agents	Αυτόνομοι πράκτορες AI για επιθέσεις σε περιβάλλοντα cloud

Autonomous penetration testing systems	Αυτόνομα συστήματα ελέγχου διείσδυσης
Backpropagation	Οπισθοδιάδοση σφάλματος στα νευρωνικά δίκτυα
Behavioral analytics	Ανάλυση προτύπων συμπεριφοράς χρηστών ή συστημάτων
Botnets	Δίκτυα μολυσμένων συσκευών που ελέγχονται από επιτιθέμενο
Business Email Compromise (BEC)	Απάτη μέσω πλαστοπροσωπίας ή παραβίασης εταιρικής ηλεκτρονικής αλληλογραφίας
Bypass	Παράκαμψη μηχανισμών ασφαλείας ή ελέγχων προστασίας
Cloud computing	Υπολογιστική παροχή πόρων μέσω διαδικτύου
Cloud credentials	Διαπιστευτήρια πρόσβασης σε υπηρεσίες υπολογιστικού νέφους
Cloud environments	Περιβάλλοντα υπολογιστικού νέφους
Cloud-native resources	Πόροι σχεδιασμένοι για λειτουργία σε περιβάλλον cloud
Convolutional Neural Networks (CNNs)	Συνελκτικά νευρωνικά δίκτυα, κυρίως για ανάλυση εικόνων
Credential attacks	Επιθέσεις με στόχο την υποκλοπή ή κατάχρηση διαπιστευτηρίων
Credential harvesting	Συλλογή κωδικών και στοιχείων σύνδεσης χρηστών

Credential stuffing	Χρήση κλεμμένων διαπιστευτηρίων σε πολλαπλές υπηρεσίες
Credentials	Στοιχεία ταυτοποίησης και πρόσβασης χρήστη ή συστήματος
Crime-as-a-Service / Cybercrime-as-a-Service (CaaS)	Παροχή κυβερνοεγκληματικών εργαλείων ή υπηρεσιών έναντι αμοιβής
Cyber kill chain	Αλληλουχία σταδίων μιας κυβερνοεπίθεσης
Cyber Resilience Act (CRA)	Ευρωπαϊκός κανονισμός κυβερνοασφάλειας για προϊόντα με ψηφιακά στοιχεία
Cyber Threat Intelligence (CTI)	Πληροφορίες και ανάλυση σχετικά με κυβερνοαπειλές
Dark web tools	Εργαλεία και υπηρεσίες που διακινούνται στον σκοτεινό ιστό
Darknet markets	Παράνομες ηλεκτρονικές αγορές του σκοτεινού δικτύου
Deep Learning	Βαθιά μάθηση με πολυεπίπεδα νευρωνικά δίκτυα
Deepfake	Συνθετικό οπτικοακουστικό περιεχόμενο που μιμείται πραγματικό πρόσωπο
Deepfake CFO fraud	Απάτη με deepfake που μιμείται υψηλόβαθμο οικονομικό στέλεχος
Denial-of-Service (DoS)	Επίθεση άρνησης υπηρεσίας
Deployment	Ανάπτυξη και ενεργοποίηση ενός εργαλείου ή λογισμικού σε σύστημα

Distributed Denial-of-Service (DDoS)	Κατανεμημένη επίθεση άρνησης υπηρεσίας από πολλές συσκευές
DNS records	Εγγραφές του Συστήματος Ονομάτων Τομέων
Domain enumeration	Εντοπισμός και καταγραφή ονομάτων τομέων ή υποτομέων
Domain patterns	Μοτίβα ονομάτων διαδικτυακών τομέων
Endpoint telemetry	Τηλεμετρικά δεδομένα από τερματικές συσκευές
Endpoints	Τερματικές συσκευές ή σημεία πρόσβασης ενός δικτύου
ENISA Threat Landscape	Έκθεση αποτύπωσης κυβερνοαπειλών του ENISA
Evasion	Αποφυγή εντοπισμού από μηχανισμούς ασφαλείας
Exploitation	Εκμετάλλευση ευπάθειας για μη εξουσιοδοτημένη πρόσβαση
Exploit kits	Εργαλεία που αυτοματοποιούν την εκμετάλλευση ευπαθειών
Exploit-based malware	Κακόβουλο λογισμικό που αξιοποιεί ευπάθειες συστημάτων
Exposed services	Δημόσια προσβάσιμες ή ανεπαρκώς προστατευμένες υπηρεσίες
Face swapping	Αντικατάσταση προσώπου σε εικόνα ή βίντεο
Fraud-as-a-Service (FaaS)	Παροχή εργαλείων και υπηρεσιών ηλεκτρονικής απάτης έναντι αμοιβής

GDPR	Γενικός Κανονισμός για την Προστασία Δεδομένων
Generative AI	Παραγωγική τεχνητή νοημοσύνη που δημιουργεί νέο περιεχόμενο
Gradient descent	Μέθοδος βελτιστοποίησης για ελαχιστοποίηση του σφάλματος μοντέλου
Host-based Intrusion Detection Systems (HIDS)	Συστήματα ανίχνευσης εισβολών σε επίπεδο τερματικού συστήματος
Impersonation	Πλαστοπροσωπία χρήστη, οργανισμού ή στελέχους
Indicators of Compromise (IoCs)	Δείκτες που υποδηλώνουν πιθανή παραβίαση συστήματος
Industrialisation of cybercrime	Οργανωμένη και μαζική παροχή κυβερνοεγκληματικών υπηρεσιών
Infostealers	Κακόβουλα προγράμματα κλοπής προσωπικών και ευαίσθητων δεδομένων
Intrusion Detection Systems (IDS)	Συστήματα ανίχνευσης εισβολών
Lateral movement	Πλάγια μετακίνηση του επιτιθέμενου μέσα σε δίκτυο
Large Language Models (LLMs)	Μεγάλα γλωσσικά μοντέλα
LLM-based agents	Αυτόνομοι πράκτορες που βασίζονται σε μεγάλα γλωσσικά μοντέλα
Linguistic mimicry	Μίμηση του γλωσσικού ύφους ενός προσώπου ή οργανισμού

Logs	Αρχεία καταγραφής συμβάντων συστήματος ή δικτύου
Machine Learning	Μηχανική μάθηση από δεδομένα
Malware	Κακόβουλο λογισμικό
Malware-as-a-Service (MaaS)	Παροχή κακόβουλου λογισμικού ως υπηρεσία
Malware samples	Δείγματα κακόβουλου λογισμικού προς ανάλυση
Mass attacks	Μαζικές κυβερνοεπιθέσεις σε μεγάλο αριθμό στόχων
Metadata	Δεδομένα που περιγράφουν άλλα δεδομένα ή αρχεία
Modular malware development	Ανάπτυξη κακόβουλου λογισμικού από ανεξάρτητες λειτουργικές ενότητες
Multi-agent workflows	Ροές εργασίας που εκτελούνται από πολλαπλούς πράκτορες AI
Natural Language Processing (NLP)	Επεξεργασία φυσικής γλώσσας
Network-based Intrusion Detection Systems (NIDS)	Συστήματα ανίχνευσης εισβολών σε επίπεδο δικτύου
Network flows	Ροές δεδομένων και επικοινωνίας μέσα σε δίκτυο
NIS2 Directive	Ευρωπαϊκή οδηγία για υψηλό κοινό επίπεδο κυβερνοασφάλειας
NIST	Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας των ΗΠΑ

Non-human identities	Ταυτότητες εφαρμογών, υπηρεσιών ή συστημάτων χωρίς φυσικό χρήστη
Obfuscation	Απόκρυψη ή δυσνόητη διαμόρφωση κώδικα
Offensive cyber operations	Επιθετικές επιχειρήσεις στον κυβερνοχώρο
Offensive programming	Προγραμματισμός με επιθετικό σκοπό στην κυβερνοασφάλεια
Open-Source Intelligence (OSINT)	Συλλογή πληροφοριών από δημόσια διαθέσιμες πηγές
Packing	Συμπίεση ή πακετάρισμα εκτελέσιμου κώδικα για δυσκολότερη ανάλυση
Payloads	Ωφέλιμα φορτία που εκτελούν την κακόβουλη ενέργεια
Persistence mechanisms	Μηχανισμοί διατήρησης πρόσβασης σε παραβιασμένο σύστημα
Phishing	Ηλεκτρονική απάτη μέσω παραπλανητικών μηνυμάτων
Phishing-as-a-Service (PhaaS)	Παροχή έτοιμων εργαλείων phishing ως υπηρεσία
Phishing campaigns	Οργανωμένες εκστρατείες παραπλανητικών μηνυμάτων
Phishing emails	Παραπλανητικά emails με στόχο την υποκλοπή στοιχείων
Poisoning	Χειραγώγηση δεδομένων εκπαίδευσης μοντέλου μηχανικής μάθησης

Polymorphic malware	Κακόβουλο λογισμικό που μεταβάλλει συνεχώς τον κώδικά του
Pretrained models	Προεκπαιδευμένα μοντέλα τεχνητής νοημοσύνης
Prompt manipulation	Χειραγώγηση εντολών προς συστήματα τεχνητής νοημοσύνης
Ransomware	Κακόβουλο λογισμικό εκβιασμού μέσω κρυπτογράφησης δεδομένων
Ransomware-as-a-Service (RaaS)	Παροχή ransomware ως υπηρεσία σε συνεργάτες επιτιθέμενους
Reconnaissance	Συλλογή πληροφοριών και αναγνώριση του στόχου
Repositories	Αποθετήρια κώδικα ή αρχείων
Rule-based IDS	IDS που βασίζονται σε προκαθορισμένους κανόνες
Scam campaigns	Οργανωμένες εκστρατείες ηλεκτρονικής απάτης
Scripts	Σύντομα προγράμματα αυτοματοποίησης ενεργειών
Security Information and Event Management (SIEM)	Διαχείριση πληροφοριών και συμβάντων ασφαλείας
Security Operations Center (SOC)	Κέντρο επιχειρήσεων ασφάλειας
Security Orchestration, Automation and Response (SOAR)	Ενορχήστρωση, αυτοματοποίηση και απόκριση σε περιστατικά ασφάλειας

Service accounts	Λογαριασμοί υπηρεσιών που χρησιμοποιούνται από εφαρμογές ή συστήματα
Signature-based detection	Ανίχνευση απειλών με βάση γνωστές ψηφιακές υπογραφές
Smishing	Phishing μέσω γραπτών μηνυμάτων SMS
Social engineering	Κοινωνική μηχανική για χειραγώγηση χρηστών
Spear phishing	Εξατομικευμένο phishing με συγκεκριμένο στόχο
Spyware	Λογισμικό παρακολούθησης και υποκλοπής πληροφοριών
Static scripts	Προκαθορισμένα scripts χωρίς δυναμική προσαρμογή
Subdomain enumeration	Εντοπισμός και καταγραφή υποτομέων ενός domain
Subdomains	Υποτομείς ενός διαδικτυακού domain
Supervised learning	Επιβλεπόμενη μάθηση με δεδομένα που έχουν ετικέτες
Synthetic media	Συνθετικό περιεχόμενο εικόνας, ήχου ή βίντεο
Synthetic voice	Τεχνητά παραγόμενη ή μιμούμενη ανθρώπινη φωνή
Target profiles	Προφίλ πιθανών στόχων μιας επίθεσης
Threat actors	Άτομα ή ομάδες που πραγματοποιούν κυβερνοεπιθέσεις

Threat intelligence	Πληροφορίες για απειλές, επιτιθέμενους και τεχνικές επίθεσης
Trojans	Δούρειοι ίπποι που αποκρύπτουν κακόβουλη λειτουργία
Unsupervised learning	Μη επιβλεπόμενη μάθηση χωρίς προκαθορισμένες ετικέτες
Vishing	Phishing μέσω φωνητικής επικοινωνίας
Voice cloning	Κλωνοποίηση ή μίμηση ανθρώπινης φωνής
Vulnerabilities	Ευπάθειες ή αδυναμίες συστημάτων
Vulnerability discovery	Εντοπισμός ευπαθειών
Vulnerability research	Έρευνα και αξιολόγηση ευπαθειών
Worms	Αυτοαναπαραγόμενα κακόβουλα προγράμματα
XDR	Εκτεταμένη ανίχνευση και απόκριση σε απειλές
Zero Trust Architecture	Αρχιτεκτονική ασφάλειας μηδενικής εμπιστοσύνης