
# Τίτλος Διπλωματικής Εργασίας:

# Τεχνητή Νοημοσύνη και Μηχανική Μάθηση ως εργαλεία κυβερνοασφάλειας

# «Artificial intelligence and machine learning as cybersecurity tools»

**ΜΙΧΑΗΛΙΔΗΣ ΜΙΧΑΗΛ**
**Α.Μ. 235**

**Επιβλέπων Καθηγητής**

**Χειλάς Κωνσταντίνος**

**ΣΕΡΡΕΣ - ΑΠΡΙΛΙΟΣ 2024**

# Περιεχόμενα

## Πίνακας Εικόνων

## Πίνακας Πινάκων

## Περίληψη

Η Τεχνητή Νοημοσύνη και η Μηχανική Μάθηση είναι πολύτιμα εργαλεία σε πολλές εφαρμογές κυβερνοασφάλειας. Η Τεχνητή Νοημοσύνη μπορεί να βρει τόσο αμυντικές όσο και επιθετικές εφαρμογές και όσο βελτιώνεται θα αποτελέσει κύριο συστατικό στην ασφάλεια πληροφοριακών συστημάτων. Μερικά από τα προτερήματά της είναι ότι μπορεί να βρει εφαρμογή στην αυτοματοποίηση επαναλαμβανόμενων διαδικασιών, στη βελτιωμένη και έγκαιρη ανίχνευση και αντιμετώπιση απειλών, στη βελτίωση της επίγνωσης της κατάστασης εφαρμογών και δικτύων αλλά και στην λήψη αποφάσεων. Από την άλλη, εμφανίζονται πολλές προκλήσεις κατά την εφαρμογή της. Η έλλειψη διαφάνειας στην ερμηνεία των αποφάσεων (black-box), οι ανησυχία για τη δικαιοσύνη ή την πιθανή πόλωση των λύσεων που προσφέρει, η ενσωμάτωση σε υφιστάμενα συστήματα είναι μερικές από αυτές. Στα πλαίσια της διπλωματικής θα διερευνηθούν εφαρμογές, λύσεις και προβλήματα με σκοπό να σκιαγραφηθεί το τοπίο που δημιουργείται.



## Abstract

Artificial intelligence and machine learning are valuable tools in many cybersecurity applications. Artificial intelligence can find both defensive and offensive applications, and as it improves it will become a major component in the security of information systems. Some of its advantages are that it can find application in the automation of repetitive processes, in the improved and timely detection and response of threats, in the improvement of the awareness of the state of applications and networks and also in decision-making. On the other hand, many challenges appear during its implementation. The lack of transparency in the interpretation of decisions (black-box), concerns about justice or the possible polarization of the solutions it offers, integration into existing systems are some of them. In the context of the diploma, applications, solutions and problems will be investigated in order to outline the landscape that is being created.

# Κεφάλαιο 1. Εισαγωγή

Η κυβερνοασφάλεια έχει καταστεί απαραίτητη για όποιον χρησιμοποιεί την τεχνολογία σήμερα, καθώς η τεχνολογία συνεχίζει να εμπλέκεται περισσότερο σε όλες τις πτυχές της καθημερινής μας ζωής (Cavelty, 2012). Με την αυξανόμενη υιοθέτηση ψηφιακών πλατφορμών και λύσεων, η ανάγκη για μέτρα ασφάλειας στον κυβερνοχώρο για την προστασία από τις αυξανόμενες απειλές στον κυβερνοχώρο γίνεται όλο και πιο σημαντική. Η ασφάλεια στον κυβερνοχώρο στοχεύει στην προστασία και προστασία συσκευών, δικτύων και δεδομένων από μη εξουσιοδοτημένη πρόσβαση, απειλή και ζημιά από οποιαδήποτε κυβερνοεπίθεση ή απειλή (Griffiths, 2023). Ο τομέας της ασφάλειας στον κυβερνοχώρο εξελίσσεται με ταχείς ρυθμούς και με νέες απειλές και τρωτά σημεία που αναδύονται καθημερινά, οι εταιρείες και τα άτομα πρέπει να είναι σε εγρήγορση και να βρίσκονται μπροστά από τους επιτιθέμενους που επιδιώκουν να βλάψουν την ασφάλειά τους. Πολλοί οργανισμοί και κυβερνήσεις έχουν βάλει το βλέμμα τους σε πιο προηγμένες και έξυπνες λύσεις για την καταπολέμηση αυτών των αναδυόμενων εγκλημάτων στον κυβερνοχώρο, την Τεχνητή Νοημοσύνη (European Commission, 2020; Kasper, 2020).

Η Τεχνητή Νοημοσύνη τα τελευταία χρόνια έχει γίνει μια ολοένα και πιο διαδεδομένη τεχνολογία σε πολλούς κλάδους, όπως η υγειονομική περίθαλψη, τα οικονομικά, οι μεταφορές, η ψυχαγωγία και πολλά άλλα χρησιμοποιούν τη δύναμη αυτής της τεχνολογίας για να προωθήσουν τις δραστηριότητές τους (Kasper, 2020). Με την ικανότητά της να αυτοματοποιεί σύνθετες εργασίες και να κάνει προβλέψεις με βάση μεγάλους όγκους δεδομένων στα οποία εκπαιδεύτηκε, η Τεχνητή Νοημοσύνη έχει, με πολλούς τρόπους, μεταμορφώσει τον αριθμό των επιχειρήσεων που λειτουργούν και τον τρόπο με τον οποίο ζουν οι άνθρωποι την καθημερινή τους ζωή (Haenlein, 2019). Ένα από τα πιο σημαντικά είναι η ικανότητα των συστημάτων Τεχνητής Νοημοσύνης να μαθαίνουν και να βελτιώνονται με την πάροδο του χρόνου μέσω τεχνικών όπως η Μηχανική Μάθηση (Machine Learning) (Abonamah, 2021). Αυτές οι τεχνικές χρησιμοποιούνται για την ανάλυση και την ερμηνεία δεδομένων και τη λήψη προβλέψεων και αποφάσεων σχετικά με νέα δεδομένα που επεξεργάζονται, μαθαίνοντας ακόμη περισσότερα κάθε φορά. Με τον αυξανόμενο

όγκο των διαθέσιμων δεδομένων, τα συστήματα Τεχνητής Νοημοσύνης μπορούν να μπορούν να απομνημονεύουν μεγάλα και ποικίλα σύνολα δεδομένων, τα οποία με τη βοήθεια της μηχανικής μάθησης επιτρέπουν στα εργαλεία της Τεχνητής Νοημοσύνης να λαμβάνει ταχύτερες αποφάσεις για συγκεκριμένες εργασίες και ακόμη και να κάνει προβλέψεις για μελλοντικές εργασίες.

Ιδιαίτερα στον κλάδο της ασφάλειας στον κυβερνοχώρο, η Τεχνητής Νοημοσύνης θα μπορούσε να μάθει από προηγούμενες απειλές και επιθέσεις στον οργανισμό για να προβλέψει μελλοντικές απειλές, να τις εντοπίσει και να τις αποτρέψει πριν προλάβουν να βλάψουν έναν οργανισμό. Επίσης, θα μπορούσε επίσης να σαρώνει το δίκτυο συνεχώς για οποιαδήποτε μη φυσιολογική δραστηριότητα ή ελαττώματα στο σύστημα, γεγονός που θα μείωνε τις ευπάθειες που θα μπορούσαν να εκμεταλλευτούν οι φορείς απειλών (IBM, 2019). Οι οργανισμοί σε όλο τον κόσμο έχουν ενστερνιστεί πλήρως τη Τεχνητή Νοημοσύνη ως πολύτιμο εργαλείο στην επιχείρησή τους αλλά και στην θωράκιση της ασφάλειά τους, θεωρώντας την ως πλεονέκτημα για την άμυνά τους έναντι των εγκληματιών του κυβερνοχώρου που θα ήθελαν να βλάψουν την ασφάλειά τους (Grand View Research, 2022). Αν και δεν αυτό δεν ισχύει για την πλειοψηφία των εταιρειών στον ανεπτυγμένο κόσμο, πολλές εξακολουθούν να προσπαθούν να αποφασίσουν εάν θα υιοθετήσουν τη Τεχνητής Νοημοσύνης στην επιχείρησή τους, θεωρώντας ότι είναι πολύ ακριβή επένδυση ή απλώς λόγω έλλειψης εμπιστοσύνης (KPMG, 2023)

Βέβαια, παρά το γεγονός ότι οι υπολογιστές δεν διαθέτουν «κληρονομική» ευφυΐα επί του παρόντος, η ιδέα της μεταφοράς αυτής της νοημοσύνης σε ανθρωπογενείς συσκευές είναι δελεαστική. Όσον αφορά την ασφάλεια στο διαδίκτυο, η Τεχνητή Νοημοσύνη είναι ζωτικής σημασίας, καθώς ο στόχος της ασφάλειας στο Διαδίκτυο είναι η δημιουργία διασφαλίσεων για την προστασία συστημάτων τεχνολογίας πληροφοριών, συνδέσεων, λογισμικού και δεδομένων από μη εξουσιοδοτημένη συνδεσιμότητα και τροποποίηση δεδομένων (Winston, 1992). Περιέχει επίσης ένα ευρύ φάσμα εργαλείων για την προστασία λογισμικού και δεδομένων αμοιβαίας υποστήριξης από απώλεια, μη εξουσιοδοτημένες συνδέσεις και κυβερνοεπίθεση. Νέοι κίνδυνοι εμφανίζονται και εμφανίζονται γρήγορα καθώς προχωρά η καινοτομία στη γνώση και στις Τεχνολογίες Πληροφοριών και

Επικοινωνιών (ΤΠΕ). Τα τελευταία χρόνια, η Τεχνητή Νοημοσύνη, έχει κερδίσει έλξη, επηρεάζοντας κάθε κομμάτι της επιχείρησης και την επιβίωσή της.

Πλέον, η Τεχνητή Νοημοσύνη και η Μηχανική Μάθηση έχουν γίνει γρήγορα μερικές από τις πιο βασικές τεχνολογίες στον τομέα της κυβερνοασφάλειας καθώς, με τον αυξανόμενο όγκο δεδομένων και τις εξελιγμένες απειλές στον κυβερνοχώρο, χρησιμοποιούνται για την ενίσχυση της ασφάλειας των οργανισμών. Πιο συγκεκριμένα, θα μπορούσαν να βοηθήσουν στην ανάλυση μεγάλων ποσοτήτων δεδομένων και στον εντοπισμό μοτίβων που μπορεί να υποδηλώνουν την παρουσία απειλής στον κυβερνοχώρο. Αυτό επιτρέπει στους οργανισμούς να εντοπίζουν και να ανταποκρίνονται σε απειλές στον κυβερνοχώρο πιο γρήγορα και με ακρίβεια από τις παραδοσιακές μεθόδους.

# Κεφάλαιο 2 Κυβερνοασφάλεια

## 2.1 Εισαγωγή

Σε έναν κόσμο που βασιλεύει η ταχύτητα και η τελειότητα, η τεχνολογία βασίζεται κυρίως στην επιστήμη των υπολογιστών. Είτε πρόκειται για μια απλή πράξη αποστολής ενός email είτε για μια κρίσιμη πράξη μεταφοράς δισεκατομμυρίων δολαρίων, σχεδόν όλα απέχουν μόνο ένα κλικ. Ο κόσμος της επιστήμης των υπολογιστών κρατά τους ανθρώπους να ασχολούνται με δραστηριότητες όπως παιχνίδια, περιήγηση σε ιστότοπους, μέσα κοινωνικής δικτύωσης, τραπεζικές συναλλαγές, ψηφιακή ιθαγένεια κ.λπ. που καλύπτει όλους τους τομείς όπως το υλικό, το λογισμικό, το δίκτυο, τα δεδομένα κ.λπ. Επειδή τόσες πολλές δραστηριότητες βασίζονται σε υπολογιστές, προσελκύει εγκληματίες, το οποίο τελικά οδηγεί σε έγκλημα στον κυβερνοχώρο, το οποίο θα μπορούσε να είναι τόσο στοιχειώδες όσο ένα βασικό «hacking» ή τόσο περίπλοκο όσο οι επιθέσεις με λογισμικό λύτρων ή τα οικονομικά εγκλήματα στον κυβερνοχώρο. Οι συνέπειες μπορεί να ποικίλλουν από απώλεια προσωπικών ή ευαίσθητων πληροφοριών έως απώλεια τεράστιου χρηματικού ποσού.

Επομένως, η ανάγκη για διασφάλιση της κυβερνοασφάλειας είναι πρωταρχικής σημασίας. Σε αυτό το κεφάλαιο, θα ρίξουμε μια ματιά στην έννοια της

κυβερνοασφάλειας, τις αιτίες, τις συνέπειες και τις αρχές της. Η ιδέα της κυβερνοασφάλειας δεν περιορίζεται μόνο σε μικρές επιχειρήσεις και εκπαιδευτικά ιδρύματα, αλλά εξαπλώνεται επίσης σε διάφορους κλάδους και την κυβέρνηση, καθιστώντας την έναν από τους πιο σημαντικούς τομείς μελέτης. Στο παρελθόν, έχουν προταθεί ορισμένοι στόχοι για τη διαφύλαξη τέτοιων κρίσιμων υποδομών στον κυβερνοχώρο. Ορισμένα πρότυπα, οδηγίες και πρακτικές βρίσκουν τη θέση τους στα πλαίσια κυβερνοασφάλειας για να διασφαλίσουν ότι η υποδομή και η αρχιτεκτονική του κυβερνοχώρου είναι ασφαλείς. Δεδομένου ότι οι λειτουργίες είναι πολλαπλές καθώς και διορατικές, πρέπει να εκτελούνται από υπεύθυνο προσωπικό στο οποίο συνήθως ανατίθενται ρόλοι στην υποδομή του κυβερνοχώρου ανάλογα με τη φύση της εργασίας τους, όπως ο διαχειριστής ασφαλείας ή η ομάδα αντιμετώπισης περιστατικών.

Η φύση των εγκλημάτων στον κυβερνοχώρο τα τελευταία χρόνια έχει αλλάξει δραστικά λόγω της αλλαγής στα κίνητρα πίσω από τα εγκλήματα, τα εργαλεία και τις τεχνικές που εμπλέκονται και τις συνολικές συνέπειες. Η αντίθεση μεταξύ των παραδοσιακών εγκλημάτων ηλεκτρονικών υπολογιστών και των σύγχρονων εγκλημάτων ηλεκτρονικών υπολογιστών τα τελευταία χρόνια. Η γενική εξέλιξη των εγκλημάτων στον κυβερνοχώρο έχει οδηγήσει σε κινδύνους που βασίζονται στο διαδίκτυο που επηρεάζουν επιχειρήσεις, οργανισμούς κ.λπ., οι οποίοι έχουν τη δυνατότητα να βλάψουν την ευθύνη και τις περιουσίες. Ωστόσο, η ασφάλεια στον κυβερνοχώρο έχει ενισχυθεί σημαντικά με την συμβολή των αναδυόμενων τεχνολογιών, όπως η Τεχνητή Νοημοσύνη και η Μηχανική Μάθηση, οι οποίες επιτρέπουν στα συστήματα να αναγνωρίζουν αυτόματα χαρακτηριστικά, να ταξινομούν πληροφορίες, να βρίσκουν μοτίβα σε δεδομένα, να κάνουν προσδιορισμούς και προβλέψεις και να αποκαλύπτουν πληροφορίες. Επίσης, έχουν την δυνατότητα να χρησιμοποιούν αλγόριθμους για τη δημιουργία μοντέλων μηχανικής εκμάθησης που εκπαιδεύουν συνεχώς τα συστήματα ώστε να αυξάνουν την ακρίβεια.

## 2.2 Ορισμός και Έννοιες

Η κυβερνοασφάλεια μπορεί να οριστεί ως η ικανότητα άμυνας και ανάκαμψης από επιθέσεις στον κυβερνοχώρο. Σύμφωνα με το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (National Institute of Standards and Technology - NIST), η κυβερνοασφάλεια είναι η ικανότητα προστασίας ή υπεράσπισης της χρήσης του κυβερνοχώρου από επιθέσεις στον κυβερνοχώρο (Kshetri, 2010). Ολόκληρος ο κυβερνοχώρος αποτελείται από πολλά αλληλοεξαρτώμενα δίκτυα της υποδομής πληροφοριακών συστημάτων, τα οποία θα μπορούσαν να είναι το διαδίκτυο, το δίκτυο τηλεπικοινωνιών, τα συστήματα υπολογιστών, τα ενσωματωμένα συστήματα και οι ελεγκτές. Έτσι, η κυβερνοασφάλεια αφορά κρίσιμες υποδομές, ασφάλεια δικτύου, ασφάλεια cloud, ασφάλεια εφαρμογών, διαδίκτυο των πραγμάτων και αρκετούς άλλους τομείς όπου η ανάγκη διασφάλισης της ασφάλειας είναι πρωταρχικής σημασίας.

- ✓ **Υποδομή Ζωτικής Σημασίας (Critical Infrastructure):** Η ασφάλεια σε υποδομές ζωτικής σημασίας ασχολείται με φυσικά συστήματα στον κυβερνοχώρο και σε πραγματικές αναπτύξεις. Βιομηχανίες όπως η αυτοματοποίηση, η αεροπορία, η υγειονομική περίθαλψη, τα φανάρια, τα δίκτυα ηλεκτρικής ενέργειας κ.λπ. είναι επιρρεπείς σε επιθέσεις στον κυβερνοχώρο, όπως η υποκλοπή, οι επιθέσεις σε κίνδυνο, οι επιθέσεις από τον άνθρωπο στη μέση και οι επιθέσεις άρνησης υπηρεσίας (Wang, et al., 2010).

- ✓ **Ασφάλεια Δικτύου (Network Security):** Η ασφάλεια δικτύου ασχολείται με μέτρα και ανησυχίες για την προστασία των συστημάτων πληροφοριών. Προστατεύει από μη εξουσιοδοτημένες εισβολές και προστατεύει τη χρηστικότητα και την ακεραιότητα του δικτύου και των δεδομένων. Οι επιθέσεις στον κυβερνοχώρο σε δίκτυα θα μπορούσαν να είναι παθητικές, όπως η σάρωση θυρών, η υποκλοπή και η κρυπτογράφηση, και ενεργές όπως το ηλεκτρονικό "ψάρεμα" (phishing), η δημιουργία σεναρίων μεταξύ τοποθεσιών και οι επιθέσεις άρνησης υπηρεσίας (DDos).

✓ **Ασφάλεια Υπολογιστικού Νέφους (Cloud Security):** Η ασφάλεια στο υπολογιστικό νέφος (ΥΝ) λαμβάνει υπόψη διάφορες τεχνολογίες και πολιτικές που βασίζονται στον έλεγχο για την προστασία των πληροφοριών, των εφαρμογών δεδομένων και της υποδομής εντός του ΥΝ. Δεδομένου ότι το ΥΝ είναι ένας κοινόχρηστος πόρος, οι επιθέσεις στον κυβερνοχώρο στο ΥΝ μπορεί να οδηγήσουν σε παραβιάσεις δεδομένων, ευπάθειες συστήματος, κακόβουλους εμπιστευτικούς χρήστες, απώλεια δεδομένων και ευπάθειες κοινής τεχνολογίας. Ορισμένες επιθέσεις στο περιβάλλον υπολογιστικού νέφους είναι η παραβίαση λογαριασμού, το ηλεκτρονικό ψάρεμα, οι επιθέσεις άρνησης υπηρεσίας και τα παραβιασμένα διαπιστευτήρια.

✓ **Ασφάλεια Εφαρμογών (Applications Security):** Η ασφάλεια μιας εφαρμογής διασφαλίζεται με τον μετριασμό των τρωτών σημείων ασφαλείας. Δεδομένου ότι μια ανάπτυξη εφαρμογής έχει πολλά στάδια όπως σχεδιασμός, ανάπτυξη, αναβάθμιση και συντήρηση, κάθε στάδιο είναι επιρρεπές σε επιθέσεις στον κυβερνοχώρο. Συνήθεις επιθέσεις που σχετίζονται με την ασφάλεια εφαρμογών Ιστού είναι η δέσμη ενεργειών μεταξύ τοποθεσιών, η έγχυση SQL (SQL injection), οι υπερχειλίσεις buffer και οι κατανεμημένες επιθέσεις άρνησης υπηρεσίας. Σε εφαρμογές για κινητές συσκευές, λαμβάνουν χώρα επιθέσεις όπως spyware, botnets, απάτες διαφημίσεων και κλικ και μολύνσεις από κακόβουλο λογισμικό.

✓ **Ασφάλεια Internet of Things:** Το Διαδίκτυο των πραγμάτων (IoT) αποτελείται από υπολογιστικές, μηχανικές και ψηφιακές συσκευές με μοναδικά αναγνωριστικά, ικανά να μεταφέρουν δεδομένα μέσω του δικτύου χωρίς ανθρώπινη παρέμβαση. Η ασφάλεια του IoT προστατεύει αυτές τις συνδεδεμένες συσκευές και δίκτυα στο IoT. Οι επιθέσεις περιλαμβάνουν spyware και botnet.

Η τριάδα της CIA (Εμπιστευτικότητα (Confidentiality), Ακεραιότητα (Integrity), Διαθεσιμότητα (Availability) είναι το ενοποιητικό χαρακτηριστικό για την ασφάλεια στον κυβερνοχώρο που χρησιμοποιείται για την αξιολόγηση

της ασφάλειας ενός οργανισμού, χρησιμοποιώντας τους τρεις βασικούς τομείς που σχετίζονται με την ασφάλεια, δηλαδή την εμπιστευτικότητα, την ακεραιότητα και τη διαθεσιμότητα. Τα τρία χαρακτηριστικά έχουν συγκεκριμένες απαιτήσεις και λειτουργίες.

## 2.3 Στόχοι Κυβερνοασφάλειας

Η έννοια της κυβερνοασφάλειας προσπαθεί να διατηρήσει έναν ασφαλή κυβερνοχώρο έτσι ώστε να προστατεύεται η κρίσιμη υποδομή. Για την ανάκαμψη από περιστατικά και επιθέσεις στον κυβερνοχώρο, θα πρέπει να υπάρχει κατάλληλη απάντηση, επίλυση και ανάκτηση. Ένα νομικό πλαίσιο διασφαλίζει ασφαλή κυβερνοχώρο. Ακολουθούν ορισμένοι στόχοι που οδηγούν στην πρόληψη από απειλές στον κυβερνοχώρο και στην προστασία από επιθέσεις στον κυβερνοχώρο.

**1. Αποτροπή Απειλών:** Για την αποτροπή απειλών, είναι σημαντικό να αναλύονται οι επιθέσεις και να διασφαλίζεται ο σχεδιασμός, η ανάπτυξη και η λειτουργία των απαιτούμενων Πρωτοκόλλων Ελέγχου Δικτύου (NCP). Πρέπει να εντοπιστούν δείκτες απειλών και να θεσπιστούν ορισμένες κατευθυντήριες γραμμές για την αναφορά συμβάντων. Η υιοθέτηση βέλτιστων πρακτικών και ο εντοπισμός κακόβουλης τεχνολογίας σε συνδυασμό με την έρευνα μπορεί να χρησιμοποιηθούν για την αποτροπή ορισμένων απειλών.

**2. Αναγνώριση και σκλήρυνση του συστήματος:** Ένας από τους πρωταρχικούς στόχους της κυβερνοασφάλειας είναι ο εντοπισμός απειλών προκειμένου να σκληρύνει το σύστημα. Η διαδικασία διασφαλίζει την αξιολόγηση κινδύνου και την υιοθέτηση μέτρων ασφαλείας. Ο σκοπός της σκλήρυνσης του συστήματος είναι ο μετριασμός ορισμένων κινδύνων που σχετίζονται με την ασφάλεια. Μερικές φορές χρησιμοποιείται προηγμένη προσέγγιση σκλήρυνσης συστήματος, η οποία ενσωματώνει επαναδιαμόρφωση σκληρών δίσκων και εγκατάσταση μόνο συγκεκριμένων προγραμμάτων στο σύστημα.

**3. Διεξαγωγή επιχειρησιακών, αρχιτεκτονικών και τεχνικών καινοτομιών:** Η εισαγωγή δυναμικών προσεγγίσεων για τη διαχείριση κινδύνων στον κυβερνοχώρο προστατεύει την υποδομή του κυβερνοχώρου από συγκεκριμένες επιθέσεις στον κυβερνοχώρο.

**4. Προετοιμασία για απρόοπτες καταστάσεις:** Η ιδέα του σχεδιασμού έκτακτης ανάγκης είναι βασικά η ετοιμότητα για επιθέσεις στον κυβερνοχώρο. Μπορεί να περιέχει πολιτικές, βέλτιστες πρακτικές, διαδικασίες και σχέδια ανάκαμψης.

**5. Κατανομή Πληροφοριών:** Οι πληροφορίες που υποτίθεται ότι κυκλοφορούν σε ολόκληρο το σύστημα πρέπει να είναι αποτελεσματικές. Οι κυβερνοαπειλές, τα τρωτά σημεία και τα περιστατικά θα μπορούσαν να αναφέρονται με την έκδοση ειδοποιήσεων. Οι πληροφορίες ενδέχεται να διανεμηθούν με επιτυχία σε διάφορες πλατφόρμες (Reveron, 2023).

**6. Εξειδικευμένη εκπαίδευση σε θέματα ασφάλειας:** Το εργατικό δυναμικό πρέπει να είναι εξοπλισμένο με εξειδικευμένη εκπαίδευση σε θέματα ασφάλειας. Οι πληροφορίες και οι υπηρεσίες πρέπει να παρέχονται στους κοινούς ομοσπονδιακούς εταίρους, ώστε το εργατικό δυναμικό να είναι αρκετά ισχυρό κατά τη διάρκεια περιστατικών στον κυβερνοχώρο.

**7. Ενίσχυση του συστήματος στην ανοχή σφαλμάτων:** Η ανοχή σφαλμάτων ενός συστήματος μπορεί να υπολογιστεί εκτελώντας αξιολόγηση τρωτότητας. Τα συστήματα υψηλής διασφάλισης ενδέχεται να αντέχουν σε επιθέσεις στον κυβερνοχώρο.

**8. Μείωση τρωτών σημείων:** Αρκετές πρακτικές ασφαλείας βοηθούν στη μείωση των τρωτών σημείων. Η ενημέρωση κώδικα, η χρήση τείχους προστασίας και η χρήση ισχυρών κωδικών πρόσβασης μπορούν να αποτρέψουν την κακόβουλη πρόσβαση στα συστήματα.

**9. Βελτίωση Χρηστικότητας:** Ο όρος ευχρηστία ορίζεται ως ο βαθμός στον οποίο κάτι είναι εύκολο στη χρήση. Οι απαιτήσεις χρηστικότητας ενδέχεται να ενσωματωθούν στα συστήματα μαζί με αξιόπιστη τεχνολογία.

**10. Έλεγχος Ταυτότητας:** Η επαλήθευση της ταυτότητας ενός χρήστη ή μιας διαδικασίας είναι μια σημαντική διαδικασία για την ασφάλεια στον κυβερνοχώρο. Ανάλογα με τη συσκευή, μπορεί να εφαρμοστεί έλεγχος ταυτότητας ενός παράγοντα ή πολλαπλών παραγόντων. Ο έλεγχος ταυτότητας υποστηρίζει αυτά που έχουμε, αυτά που είμαστε και όσα γνωρίζουμε.

**11. Αυτοματοποίησης Διαδικασίας Ταυτοποίησης**: Ο αυτοματισμός οδηγεί σε αποτελεσματικότητα, καλύτερη πρόβλεψη συμπεριφοράς και ταχύτερη εκτέλεση. Η κατάλληλη εφαρμογή του αυτοματισμού οδηγεί στην πρόληψη επιθέσεων στον κυβερνοχώρο. Ο αυτοματισμός μπορεί να συσχετίσει δεδομένα, να προωθήσει την πρόληψη ταχύτερα από την εξάπλωση επιθέσεων και να εντοπίσει μολύνσεις δικτύου.

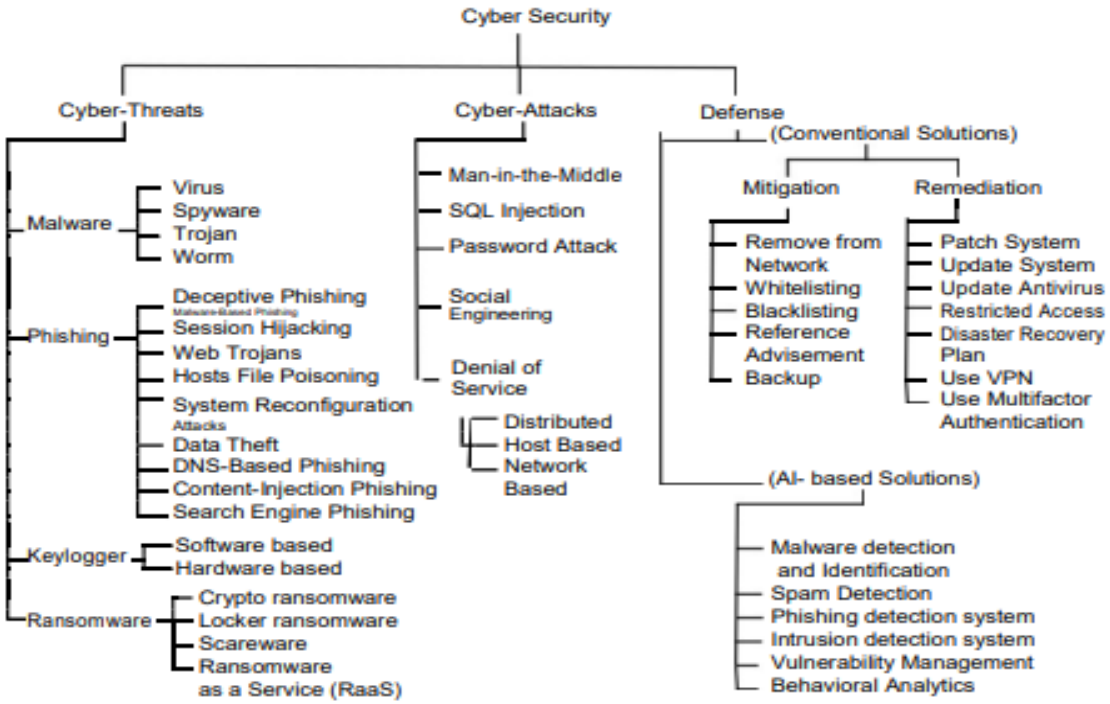
**12. Εγγύηση Διαλειτουργικότητας Συσκευών:** Διαλειτουργικότητα είναι η ικανότητα των συστημάτων να συντονίζονται προκειμένου να συνεργάζονται ή μεταξύ των οργανισμών. Η διασφάλιση της διαλειτουργικότητας οδηγεί στην αποτελεσματική διανομή των πληροφοριών στον οργανισμό.

**13. Επισήμανση και Αντιμετώπιση Δυσμενών Γεγονότων:** Είναι σημαντικό να επισημαίνονται δυσμενή γεγονότα στον κυβερνοχώρο, ώστε να βρεθεί λύση προκειμένου να αποφευχθεί ο βανδαλισμός των συστημάτων. Πληροφορίες σχετικά με την αιτία, την έκταση και τον αντίκτυπο των δυσμενών συμβάντων ενδέχεται να παρατίθενται για μελλοντική χρήση.

**14. Καθορισμός Αποτελεσματικών Μέτρων Ασφαλείας:** Με την εισαγωγή μέτρων ασφαλείας, μπορεί κανείς να εντοπίσει επιθέσεις στον κυβερνοχώρο, να τις αποτρέψει και να τις διορθώσει. Μερικά μέτρα ασφαλείας είναι η τμηματοποίηση δικτύου και η χρήση τείχους προστασίας, η ασφαλής απομακρυσμένη πρόσβαση, οι έλεγχοι πρόσβασης, η προστασία με κωδικό πρόσβασης, η εξασφάλιση προγραμμάτων εκπαίδευσης και ο καθορισμός πολιτικών (Paula & Cruz , 2023).

Η κυβερνοασφάλεια είναι η πρακτική της προστασίας κρίσιμων συστημάτων και ευαίσθητων πληροφοριών από ψηφιακές επιθέσεις. Υπάρχουν πολλοί τρόποι για την προστασία των δεδομένων και της οργανωτικής υποδομής, όπως η ανίχνευση εισβολών, η προστασία από κακόβουλο

λογισμικό, η αυστηρή τήρηση ορθών πρακτικών ασφαλείας και πολλά άλλα. Μια απειλή για την ασφάλεια στον κυβερνοχώρο μπορεί να είναι μια κυβερνοεπίθεση που χρησιμοποιεί κακόβουλο λογισμικό ή ransomware για να αποκτήσει πρόσβαση σε δεδομένα, να διακόψει τις ψηφιακές λειτουργίες ή να βλάψει πληροφορίες. Υπάρχουν κάθε είδους απειλές στον κυβερνοχώρο, συμπεριλαμβανομένων των εταιρικών κατασκόπων, των κακόβουλων χρηστών και των τρομοκρατών [28]. Στην εικόνα 1, παρουσιάζεται η ταξινόμηση της κυβερνοασφάλειας. Αν και όλοι έχουν διαφορετικούς λόγους για να επιτεθούν, όλοι θα πρέπει να αντιμετωπίζονται με εξαιρετική προσοχή καθώς αποτελούν κίνδυνο για τα δεδομένα ενός οργανισμού και τα προσωπικά δεδομένα. Η άνοδος του διαδικτύου έχει φέρει μια νέα εποχή ανησυχιών για την ασφάλεια στον κυβερνοχώρο. Εκτός από την απειλή των ηλεκτρονικών εγκληματιών (hackers) και των ξένων κυβερνήσεων, νέες προκλήσεις συνδέονται με την προστασία πληροφοριών από εσωτερικές απειλές, όπως παραβιάσεις δεδομένων και κλοπή εμπιστευτικών πληροφοριών (Obotivere & Nwaezeigwe, 2020). Η ασφάλεια στον κυβερνοχώρο είναι επίσης μια ουσιαστική διατομεακή ανησυχία για ευαίσθητες υποδομές, κρίσιμα περιουσιακά στοιχεία και ευαίσθητες πληροφορίες. Αυτός είναι ο λόγος για τον οποίο σημειώθηκε αξιοσημείωτη άνοδος των επαγγελματιών της κυβερνοασφάλειας στο σύνολό της και γιατί γίνεται ολοένα και πιο σημαντικό να διασφαλιστεί ότι οι μηχανισμοί άμυνας έναντι των επιθέσεων στον κυβερνοχώρο είναι ολοκληρωμένοι και ισχυροί .




Εικόνα 1. Ταξινόμηση της κυβερνοασφάλειας. Πηγή: *(Chakraborty, et al., 2022)*

## 2.4 Υποδομή κυβερνοασφάλειας και Αρχιτεκτονική Διαδικτύου

Δεδομένου ότι οι επιθέσεις στον κυβερνοχώρο γίνονται όλο και πιο περίπλοκες, υπάρχει ανάγκη εισαγωγής τυποποιημένων πρακτικών για τη διασφάλιση της ασφάλειας. Το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (National Institute of Standards and Technology - NIST) ενσωματώνει ορισμένες πολιτικές, πρότυπα, κατευθυντήριες γραμμές και βέλτιστες πρακτικές για την αντιμετώπιση ζητημάτων ασφάλειας στον κυβερνοχώρο (National Institute of Standards and Technology, 2018). Αυτό το πλαίσιο χωρίζεται σε πυρήνα πλαισίου, επίπεδα υλοποίησης και προφίλ.

**1. Πυρήνας Πλαισίου (Framework Core):** Ο πυρήνας του πλαισίου αποτελείται από ορισμένα σχήματα που οδηγούν σε συγκεκριμένα αποτελέσματα. Μπορεί να έχει τη μορφή συναρτήσεων, κατηγοριών, υποκατηγοριών και ενημερωτικών αναφορών. Πιο συγκεκριμένα:

- ✓ **Λειτουργίες (Functions):** Για την ασφάλεια των συστημάτων και την απόκριση σε επιθέσεις, οι βασικές λειτουργίες είναι ο εντοπισμός, η αναγνώριση, η προστασία, η απόκριση και η ανάκτηση, για τις οποίες θα συζητήσουμε αργότερα σε αυτήν την ενότητα.

- ✓ **Κατηγορίες (Categories):** Διαφορετικές συναρτήσεις έχουν αντίστοιχες κατηγορίες για τον προσδιορισμό διαφορετικών λειτουργιών και δραστηριοτήτων. Για πχ. Για την προστασία, μπορεί κανείς να κάνει χρήση ελέγχου πρόσβασης, ενημερώσεων λογισμικού και προγραμμάτων κατά του κακόβουλου λογισμικού.

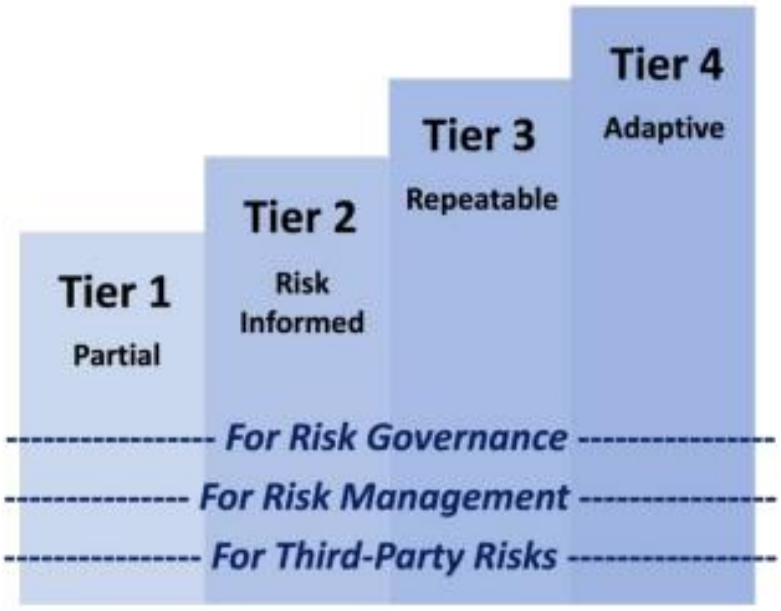
- ✓ **Υποκατηγορίες (Subcategories):** Οι κατηγορίες με συγκεκριμένους στόχους ονομάζονται υποκατηγορίες. Για παράδειγμα, η διαδικασία ενημέρωσης λογισμικού θα μπορούσε να έχει συγκεκριμένες λειτουργίες όπως σωστή διαμόρφωση ή μη αυτόματη ενημέρωση συσκευών.

- ✓ **Ενημερωτικές αναφορές (Informative References):** Οι ενημερωτικές αναφορές περιλαμβάνουν πολιτικές, πρότυπα και οδηγίες. Για παράδειγμα, ορισμένα βήματα που απαιτούν μη αυτόματη ενημέρωση του συστήματος.

**2. Επίπεδα Υλοποίησης (Implementation Tiers):** Τα ακόλουθα είναι τα τέσσερα επίπεδα υλοποίησης:

Η επιλογή των επιπέδων πλαισίου (Tiers) βοηθά να διαμορφωθεί ο συνολικός τόνος για τον τρόπο διαχείρισης των κινδύνων κυβερνοασφάλειας εντός του οργανισμού και να καθοριστεί η προσπάθεια που απαιτείται για την επίτευξη ενός επιλεγμένου επιπέδου. Οι οργανισμοί μπορούν να επιλέξουν να χρησιμοποιήσουν τα Επίπεδα για να ενημερώσουν τα τρέχοντα προφίλ και τα προφίλ στόχων τους. Τα εν λόγω επίπεδα χαρακτηρίζουν την αυστηρότητα των αποτελεσμάτων διακυβέρνησης και διαχείρισης των κινδύνων στον κυβερνοχώρο ενός οργανισμού και παρέχουν το πλαίσιο για το πώς ένας οργανισμός βλέπει τους κινδύνους κυβερνοασφάλειας και τις διαδικασίες που

εφαρμόζονται για τη διαχείριση αυτών των κινδύνων (National Institute of Standards and Technology, 2023).




Εικόνα 2. Επίπεδα Πλαισίου Κυβερνοασφάλειας .Πηγή: *(National Institute of Standards and Technology, 2023)*

- ✓ Το επίπεδο 1 (Tier 1) ή μερική εφαρμογή χειρίζεται τους οργανωτικούς κινδύνους με ασυνέπεια λόγω της ad-hoc υποδομής ασφάλειας στον κυβερνοχώρο.
- ✓ Το επίπεδο 2 (Tier 2) ασχολείται με κινδύνους, σχέδια και πόρους για την προστασία της υποδομής στον κυβερνοχώρο σε βαθύτερο επίπεδο από τη μερική εφαρμογή.
- ✓ Το επίπεδο 3 (Tier 3) ή η επαναλαμβανόμενη υλοποίηση μπορεί επανειλημμένα να τείνει σε κρίσεις στον κυβερνοχώρο. Οι πολιτικές μπορούν να εφαρμοστούν στο ίδιο επίπεδο και η ευαισθητοποίηση για την ασφάλεια στον κυβερνοχώρο μπορεί να ελαχιστοποιήσει τους κινδύνους που σχετίζονται με τον κυβερνοχώρο.

- ✓ Το Tier 4 είναι υπεύθυνο για τον εντοπισμό απειλών και την πρόβλεψη ζητημάτων σε σχέση με την υποδομή ασφαλείας.

**3. Προφίλ (Profiles):** Το Πλαίσιο Κυβερνοασφάλειας έχει ορισμένους συγκεκριμένους στόχους. Τα προφίλ συνοψίζουν την κατάσταση της κυβερνοασφάλειας ενός οργανισμού. Τα πολλαπλά προφίλ σε ένα πλαίσιο κυβερνοασφάλειας διασφαλίζουν τον εντοπισμό πολλών αδύναμων σημείων που αποτελούν μέρος της εφαρμογής της κυβερνοασφάλειας. Μπορούν επίσης να υποστηρίξουν τη σύνδεση μεταξύ λειτουργιών, κατηγοριών και υποκατηγοριών με πόρους και την ανοχή κινδύνου των οργανισμών (National Institute of Standards and Technology, 2023).


Εικόνα 3. Προφίλ Πλαισίου Κυβερνοασφάλεια.Πηγή: *(National Institute of Standards and Technology, 2023)*

## 2.5 Εργαλεία και Τεχνικές Ασφάλειας στο Κυβερνοχώρο

### 2.5.1 Τείχη Προστασίας (Firewalls)

Υπάρχει μια ποικιλία διαφορετικών τύπων τείχους προστασίας, αλλά οι 3 πιο συνηθισμένοι είναι:

✓ **Φίλτρο Πακέτων:** Αυτός είναι ο αρχικός και πιο βασικός τύπος τείχους προστασίας που αναπτύσσουν οι επαγγελματίες της ασφάλειας στον

κυβερνοχώρο. Επιθεωρεί πακέτα που μεταφέρονται μεταξύ υπολογιστών και επιτρέπει ή αρνείται την πρόσβαση βάσει μιας λίστας ελέγχου πρόσβασης. Αυτή η λίστα λέει στο τείχος προστασίας ποια πακέτα πρέπει να διερευνηθούν και ποιες πληροφορίες θα οδηγήσουν σε απόρριψη ή διαγραφή αρχείου. Αυτά τα τείχη προστασίας είναι παλαιότερα και δεν μπορούν να ασφαλίσουν πλήρως ένα δίκτυο από μόνα τους, αλλά εξακολουθούν να είναι χρήσιμα για το φιλτράρισμα κυβερνοεπιθέσεων χαμηλής προσπάθειας.

✓ **Παρακολούθηση Σύνδεσης:** Τα τείχη προστασίας παρακολούθησης σύνδεσης, γνωστά και ως τείχη προστασίας δεύτερης γενιάς, εκτελούν εργασίες με τρόπο παρόμοιο με τα φίλτρα πακέτων πρώτης γενιάς. Εκτελούν παρόμοιο τύπο επιθεώρησης πακέτων, αλλά καταγράφουν επίσης τον αριθμό θύρας που χρησιμοποιεί κάθε διεύθυνση IP για την αποστολή και τη λήψη πληροφοριών. Αυτό επιτρέπει την εξέταση της ανταλλαγής δεδομένων επιπλέον του περιεχομένου του πακέτου.

✓ **Επίπεδο 7 (Εφαρμογής):** Τα τείχη προστασίας εφαρμογών είναι σημαντικά πιο ισχυρά από τα τείχη προστασίας παρακολούθησης σύνδεσης ή φίλτρων πακέτων. Είναι σε θέση να κατανοήσουν διάφορες εφαρμογές όπως το πρωτόκολλο μεταφοράς αρχείων (FTP), το πρωτόκολλο μεταφοράς υπερκειμένου (HTTP) και το σύστημα ονομάτων τομέα (DNS). Αυτό τους επιτρέπει να αναγνωρίζουν μη τυπικές θύρες ή ανεπιθύμητες εφαρμογές. Αυτά είναι επίσης χρήσιμα στο διαδίκτυο χάρη στην ικανότητά τους να πραγματοποιούν φιλτράρισμα ιστού.

### 2.5.2 Λογισμικό Anti-Malware

Το Anti-malware είναι ένας τύπος εργαλείου ασφάλειας στον κυβερνοχώρο που βασίζεται σε λογισμικό που εμποδίζει το κακόβουλο λογισμικό (κακόβουλο λογισμικό) να μολύνει έναν υπολογιστή και αφαιρεί υπάρχον κακόβουλο λογισμικό από συσκευές και συστήματα. Υπάρχουν 3 συνήθεις τύποι λογισμικού κατά του κακόβουλου λογισμικού, ο καθένας με τη δική του μέθοδο για τον εντοπισμό και την αφαίρεση κακόβουλου λογισμικού:

✓ **Ανίχνευση βάσει Συμπεριφοράς (Behavior-based Detection)**: Αυτός είναι ένας ισχυρός τύπος λογισμικού που εφαρμόζει τεχνολογία όπως αλγόριθμους μηχανικής εκμάθησης για τον εντοπισμό κακόβουλου λογισμικού μέσω μιας ενεργής προσέγγισης. Αντί να εξετάζει την εμφάνιση του κακόβουλου λογισμικού, εστιάζει στον τρόπο συμπεριφοράς του προκειμένου να το εξαλείψει πιο γρήγορα.

✓ **Sandboxing:** Το Sandboxing είναι μια δυνατότητα που τοποθετεί επικίνδυνο λογισμικό σε μια απομονωμένη τοποθεσία. Μπορεί να φιλτράρει τα αρχεία προτού προκαλέσουν ζημιά στο σύστημα γενικότερα. Μόλις απομονωθεί, το anti-malware μπορεί να διαγράψει το επικίνδυνο λογισμικό.

✓ **Ανίχνευση βάσει Υπογραφών (Signature-based Detection):** Η ανίχνευση βάσει υπογραφών είναι πιο χρήσιμη για την εξάλειψη κοινών κακόβουλων προγραμμάτων, όπως adware και keyloggers. Χρησιμοποιεί ανίχνευση υπογραφών για να εντοπίσει κοινό κακόβουλο λογισμικό και να το διαγράψει. Μόλις εξαλείψει ένα κομμάτι κακόβουλου λογισμικού, θα αφαιρέσει αυτόματα όλους τους τύπους κακόβουλου λογισμικού που φέρουν την ίδια υπογραφή.

### 2.5.3 Λογισμικό Εξουδετέρωσης Ιών

Το λογισμικό προστασίας από ιούς είναι ένα άλλο από τα εργαλεία για την ασφάλεια στον κυβερνοχώρο που είναι πιθανό να γνωρίζουν πολλοί χρήστες υπολογιστών. Συνιστάται γενικά ο καθένας να εγκαταστήσει κάποιο είδος λογισμικού προστασίας από ιούς στις συσκευές του για να μην το μολύνει επικίνδυνο λογισμικό.

Επί του παρόντος, το πιο ισχυρό λογισμικό προστασίας από ιούς ονομάζεται «λογισμικό επόμενης γενιάς». Χρησιμοποιείται από το 2014 και είναι γνωστό από μια στροφή προς την ανίχνευση χωρίς υπογραφή. Αυτός ο τύπος λογισμικού προστασίας από ιούς μπορεί να ενσωματώσει στον

προγραμματισμό του την ανίχνευση συμπεριφοράς και η έκρηξη αρχείων που βασίζεται σε σύννεφο.

Οι επαγγελματίες της ασφάλειας στον κυβερνοχώρο πρέπει να ενημερώνονται για τις τελευταίες εξελίξεις στο λογισμικό προστασίας από ιούς για να προστατεύουν τις εταιρείες για τις οποίες εργάζονται. Επειδή οι ιοί εξελίσσονται συνεχώς, είναι σημαντικό οι εταιρείες να γνωρίζουν την πιο αποτελεσματική, αιχμής τεχνολογία προστασίας από ιούς και να κάνουν αναβαθμίσεις στο υπάρχον λογισμικό όταν αυτό είναι διαθέσιμο.

### 2.5.4 Δοκιμή Διείσδυσης (Penetration Testing)

Η δοκιμή διείσδυσης είναι μια τεχνική ασφάλειας στον κυβερνοχώρο που προσομοιώνει μια κυβερνοεπίθεση σε ένα σύστημα. Αυτό μπορεί επίσης να είναι γνωστό ως δοκιμή στυλό ή ηθική παραβίαση. Το τεστ έχει σχεδιαστεί για να εντοπίζει τις αδυναμίες ενός συστήματος και να προσδιορίζει την πιθανότητα παραβίασης. Βοηθά επίσης τους επαγγελματίες της ασφάλειας στον κυβερνοχώρο να προσδιορίσουν ποια μέρη του συστήματος είναι ισχυρότερα και δεν απαιτούν επί του παρόντος βελτίωση.

Για να εκτελέσει μια δοκιμή διείσδυσης, ο κακόβουλος χρήστης θα περάσει συνήθως από 6 διαφορετικές φάσεις:

- ✓ **Αναγνώριση:** Ο επαγγελματίας της ασφάλειας στον κυβερνοχώρο συλλέγει δεδομένα για το σύστημα για να του επιτεθεί καλύτερα. Αυτές οι δοκιμές εκτελούνται συνήθως από κάποιον που δεν είναι καλά εξοικειωμένος με το σύστημα για την καλύτερη προσομοίωση ενός ρεαλιστικού σεναρίου παραβίασης.

- ✓ **Σάρωση:** Ο εισβολέας αναπτύσσει εργαλεία που σαρώνουν το δίκτυο και ανοίγουν τις θύρες, αυξάνοντας περαιτέρω το ποσό που γνωρίζει για το δίκτυο.

✓ **Απόκτηση Πρόσβασης:** Ο κακόβουλος χρήστης χρησιμοποιεί τα δεδομένα που συλλέγονται από τις προηγούμενες 2 φάσεις για να εισχωρήσει στο δίκτυο. Αυτό μπορεί να γίνει χειροκίνητα ή με λογισμικό.

✓ **Συντήρηση Πρόσβασης:** Μόλις εισχωρήσουν στο δίκτυο, ο ελεγκτής διείσδυσης πρέπει να προσπαθήσει να διατηρήσει την παρουσία του εντός του δικτύου για να κλέψει όσο το δυνατόν περισσότερα δεδομένα.

✓ **Αφαίρεση Αποδεικτικών Στοιχείων:** Αφού συγκεντρώσει τα δεδομένα και κάνει τη διαφυγή τους, ο δοκιμαστής καλύπτει τα ίχνη του για να διασφαλίσει ότι δεν μπορούν να ενοχοποιηθούν για την επίθεση. Αυτό γίνεται με την αφαίρεση στοιχείων σχετικά με τα δεδομένα που συγκεντρώθηκαν και την εξάλειψη των γεγονότων καταγραφής για τη διατήρηση της ανωνυμίας.

✓ **Διοχέτευση Κίνησης (Pivoting):** Στη δοκιμή διείσδυσης, η διοχέτευση κίνησης είναι η πράξη της χρήσης ενός παραβιασμένου συστήματος για εξάπλωση μεταξύ διαφορετικών συστημάτων ηλεκτρονικών υπολογιστών μόλις εισέλθει στο δίκτυο. Αυτή η διαδικασία επαναλαμβάνει τα βήματα 2 έως 5 για τη λήψη πρόσθετων δεδομένων.

Μόλις ολοκληρωθεί, ο κακόβουλος χρήστης συντάσσει μια αναφορά για το πώς μπόρεσαν να εισβάλουν στο σύστημα. Ο διαχειριστής του δικτύου ή οι επαγγελματίες της ασφάλειας στον κυβερνοχώρο στην εταιρεία που κατέχει το δίκτυο θα χρησιμοποιήσουν στη συνέχεια αυτές τις πληροφορίες για να ενισχύσουν την άμυνα του δικτύου. Οι ελεγκτές διείσδυσης χρησιμοποιούν συνήθως εργαλεία ασφάλειας στον κυβερνοχώρο όπως το Kali Linux, μια διανομή Linux ανοιχτού κώδικα, καθώς και τα Metasploit, Intruder και Core Impact (Diogenes & Ozkaya, 2018).

### 2.5.5 Έλεγχος Κωδικών Πρόσβασης και Ανιχνευτές Πακέτων

Οι επαγγελματίες της ασφάλειας στον κυβερνοχώρο χρησιμοποιούν εξειδικευμένα εργαλεία για την αξιολόγηση των κωδικών πρόσβασης και την παρακολούθηση δικτύων. Γνωρίζουν ότι οι αδύναμοι κωδικοί πρόσβασης

μπορούν να θέσουν σε κίνδυνο ένα ολόκληρο δίκτυο και τα κρίσιμα δεδομένα που διαχειρίζεται. Έτσι, χρησιμοποιώντας τεχνικές ελέγχου κωδικών πρόσβασης, οι διαχειριστές συστήματος και οι αναλυτές μπορούν να παρακολουθούν τους κωδικούς πρόσβασης και να προσδιορίζουν τη δύναμή τους έναντι απόπειρες εισβολής.

Αρχικά, το «John the Ripper» είναι ένα εργαλείο που χρησιμοποιείται για τη δοκιμή της ισχύος των κωδικών πρόσβασης γρήγορα και αποτελεσματικά, για να ελαχιστοποιηθεί η πιθανότητα ένας αδύναμος κωδικός πρόσβασης να θέσει ένα δίκτυο σε κίνδυνο.

Ένα άλλο εργαλείο αποτελεί το «Hashcat», το οποίο βοηθά στη διάσπαση του κωδικού πρόσβασης που χρησιμοποιείται από ελεγκτές διείσδυσης και διαχειριστές συστήματος. Ο κατακερματισμός κωδικού πρόσβασης είναι μια μέθοδος προστασίας των κωδικών πρόσβασης με τη μετατροπή τους σε μια σειρά τυχαίων χαρακτήρων, γνωστών ως κατακερματισμός (αυτή η διαδικασία διαφέρει από την κρυπτογράφηση, η οποία χρησιμοποιείται για την απόκρυψη πληροφοριών). Το λογισμικό ουσιαστικά μαντεύει έναν κωδικό πρόσβασης, τον κατακερματίζει και συγκρίνει τον κατακερματισμό με αυτόν που προσπαθεί να σπάσει.

Επίσης, ο ανιχνευτής πακέτων, γνωστός και ως αναλυτής πακέτων, αναλυτής πρωτοκόλλου ή αναλυτής δικτύου, είναι ένα εργαλείο υλικού ή λογισμικού που χρησιμοποιείται για την παρακολούθηση της κυκλοφορίας του δικτύου. Ένα από τα σημαντικότερα είναι το Wireshark, το οποίο αποτελεί εργαλείο ασφάλειας στον κυβερνοχώρο που βασίζεται σε κονσόλα (παλαιότερα γνωστό ως Ethereal) που χρησιμοποιείται για τη μελέτη πρωτοκόλλων δικτύου και την ανάλυση της ασφάλειας του δικτύου σε πραγματικό χρόνο. Ακόμη, το Tcpdump αποτελεί ένα πρόγραμμα ανίχνευσης πακέτων δεδομένων δικτύου που χρησιμοποιείται από επαγγελματίες της ασφάλειας στον κυβερνοχώρο για την παρακολούθηση και την καταγραφή της κυκλοφορίας TCP (Transmission Control Protocol) και IP (Internet Protocol) που διέρχεται από ένα δίκτυο υπολογιστών.

### 2.5.6 Παρακολούθηση Ασφάλειας Δικτύου

Μέσω της χρήσης λογισμικού παρακολούθησης δικτύου, οι διαχειριστές μπορούν να προσδιορίσουν εάν ένα δίκτυο λειτουργεί βέλτιστα και να εντοπίσουν προληπτικά τις ελλείψεις. Η παρακολούθηση δικτύου παρέχει μια σαφή εικόνα όλων των συνδεδεμένων συσκευών σε ένα δίκτυο, επιτρέποντας στους διαχειριστές συστήματος να βλέπουν πώς μετακινούνται τα δεδομένα μεταξύ τους και να διορθώνουν γρήγορα τυχόν ελαττώματα που θα μπορούσαν να υπονομεύσουν την απόδοση του δικτύου ή να οδηγήσουν σε διακοπές λειτουργίας.

Οι τύποι πρωτοκόλλων παρακολούθησης δικτύου περιλαμβάνουν:

- ✓ **Simple Network Management Protocol (SNMP):** Το Simple Network Management Protocol χρησιμοποιεί ένα σύστημα κλήσης και απόκρισης για να ελέγξει την κατάσταση συσκευών όπως διακόπτες και εκτυπωτές και μπορεί να χρησιμοποιηθεί για την παρακολούθηση της κατάστασης και της διαμόρφωσης του συστήματος.

- ✓ **Internet Control Message Protocol (ICMP):** Οι δρομολογητές, οι διακομιστές και άλλες συσκευές δικτύου χρησιμοποιούν το Πρωτόκολλο Μηνυμάτων Ελέγχου Διαδικτύου για την αποστολή πληροφοριών λειτουργιών IP και τη δημιουργία μηνυμάτων όταν οι συσκευές αποτυγχάνουν.

- ✓ **Πρωτόκολλο Ανακαλύψεων Cisco (Cisco Discover):** Αυτό το πρωτόκολλο διευκολύνει τη διαχείριση των συσκευών Cisco ανακαλύπτοντάς τις, προσδιορίζοντας τον τρόπο διαμόρφωσης τους και επιτρέποντας στα συστήματα που χρησιμοποιούν διαφορετικά πρωτόκολλα επιπέδου δικτύου να μαθαίνουν το ένα το άλλο.

- ✓ **ThousandEyes Synthetics**: Ένα σύστημα συνθετικής παρακολούθησης με γνώση του Διαδικτύου που εντοπίζει προβλήματα απόδοσης σύγχρονων δικτυωμένων εφαρμογών.

### 2.5.7 Σαρωτές Ευπάθειας

Οι σαρωτές ευπάθειας βοηθούν τους οργανισμούς να προσδιορίσουν ποιες απειλές για την ασφάλεια στον κυβερνοχώρο ενδέχεται να αντιμετωπίζουν ως αποτέλεσμα των τρωτών σημείων που εντοπίζονται στην υποδομή πληροφορικής τους. Οι οργανισμοί χρησιμοποιούν συχνά πολλαπλούς σαρωτές ευπάθειας για να διασφαλίσουν ότι λαμβάνουν μια σαφή αξιολόγηση των απειλών. Ένα δείγμα αυτών των εργαλείων ασφάλειας στον κυβερνοχώρο περιλαμβάνει:

✓ **Acunetix:** Αυτός ο σαρωτής ευπάθειας ιστού διαθέτει προηγμένη τεχνολογία ανίχνευσης που του επιτρέπει να αποκαλύπτει τρωτά σημεία για αναζήτηση κάθε τύπου ιστοσελίδας, ακόμη και σε σελίδες που προστατεύονται με κωδικό πρόσβασης.

✓ **Nessus:** Το Nessus, το οποίο έχει ληφθεί περισσότερες από 2 εκατομμύρια φορές παγκοσμίως, παρέχει ενδελεχή κάλυψη και σαρώνει για περισσότερες από 59.000 κοινές ευπάθειες και εκθέσεις (CVE).

✓ **Burp Suite:** Με πολλαπλές δυνατότητες σάρωσης, ενοποίησης και αναφοράς, το Burp Suite είναι ένας σαρωτής ευπάθειας που ενσωματώνεται με συστήματα παρακολούθησης σφαλμάτων όπως το Jira και ενημερώνεται συχνά.

✓ **GFI Languard:** Ένας σαρωτής ευπάθειας για εφαρμογές δικτύου και web που μπορεί να αναπτύξει αυτόματα ενημερώσεις κώδικα σε λειτουργικά συστήματα, προγράμματα περιήγησης ιστού και εφαρμογές τρίτων.

✓ **Tripwire IP360:** Ένα επεκτάσιμο εργαλείο σάρωσης ευπάθειας που μπορεί να σαρώσει το συνολικό περιβάλλον ενός οργανισμού, συμπεριλαμβανομένων των στοιχείων που δεν είχαν εντοπιστεί στο παρελθόν.

### 2.5.8 Ανίχνευση Εισβολής Δικτύου

Για να βελτιώσουν την προστασία από κακόβουλη επισκεψιμότητα IP στα δίκτυά τους, οι οργανισμοί χρησιμοποιούν συχνά συστήματα ανίχνευσης εισβολής και προστασίας (IDPS) για να προστατεύονται από απειλές που ενδέχεται να διεισδύσουν στα τείχη προστασίας τους. Τα συστήματα ανίχνευσης εισβολής (IDS) χρησιμοποιούν λογισμικό για την αυτοματοποίηση της διαδικασίας ανίχνευσης και τα συστήματα προστασίας από εισβολή (IPS) χρησιμοποιούν λογισμικό για τον εντοπισμό και την προσπάθεια αποτροπής πιθανών παραβιάσεων δεδομένων. Μόλις εντοπιστεί ένα κακόβουλο μοτίβο ή παραβίαση, το IDS ειδοποιεί τους διαχειριστές του συστήματος ώστε να προβούν στις κατάλληλες ενέργειες. Το IPS αναλύει την κίνηση IP και αποκλείει την κακόβουλη κυκλοφορία, αποτρέποντας έτσι μια επίθεση.

Σύμφωνα με το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (NIST), υπάρχουν 4 ταξινομήσεις τεχνολογιών IDPS:

- ✓ **Βασισμένο σε δίκτυο (Network-based):** Αυτές οι τεχνολογίες IDPS παρακολουθούν την κυκλοφορία δικτύου για συγκεκριμένα τμήματα ή συσκευές δικτύου και αναλύουν τη δραστηριότητα του πρωτοκόλλου δικτύου και εφαρμογών για τον εντοπισμό ύποπτων δραστηριοτήτων.

- ✓ **Ασύρματο (Wireless):** Οι ασύρματες τεχνολογίες IDPS παρακολουθούν και αναλύουν την κίνηση σε ασύρματα δίκτυα για να εντοπίσουν ύποπτη δραστηριότητα που περιλαμβάνει πρωτόκολλα ασύρματης δικτύωσης.\

- ✓ **Ανάλυση συμπεριφοράς δικτύου (Network Behavior Analysis (NBA):** Το NBA εξετάζει την κυκλοφορία δικτύου για να εντοπίσει απειλές που δημιουργούν ασυνήθιστες ροές κυκλοφορίας, όπως επιθέσεις κατανεμημένης άρνησης υπηρεσίας (DDoS) ή ορισμένες μορφές κακόβουλου λογισμικού.

- ✓ **Βασισμένο σε κεντρικό υπολογιστή (Host-based):** Οι τεχνολογίες IDPS που βασίζονται σε κεντρικούς υπολογιστές παρακολουθούν τα

χαρακτηριστικά ενός μεμονωμένου κεντρικού υπολογιστή (π.χ. υπολογιστή ή διακομιστή) και τα συμβάντα που συμβαίνουν σε αυτόν τον κεντρικό υπολογιστή για ύποπτη δραστηριότητα.

### 2.5.9 Εργαλεία Κρυπτογράφησης

Διαδραματίζοντας ουσιαστικό ρόλο στη διαφύλαξη των δεδομένων που αποθηκεύονται ή μεταδίδονται, η κρυπτογράφηση είναι μια διαδικασία που ανακατεύει αναγνώσιμο κείμενο, ώστε να μπορεί να διαβαστεί μόνο από το άτομο που έχει το κλειδί αποκρυπτογράφησης. Τεράστιες ποσότητες προσωπικών πληροφοριών – τραπεζικοί λογαριασμοί, προφίλ πιστωτικών καρτών, αρχεία υγείας και άλλα – διαχειρίζονται διαδικτυακά και αποθηκεύονται στο cloud ή σε διακομιστές που είναι συνδεδεμένοι στο διαδίκτυο.

Η κρυπτογράφηση μετατρέπει το αναγνώσιμο κείμενο σε μια μη αναγνώσιμη μορφή που ονομάζεται cypher text. Όταν ο προοριζόμενος παραλήπτης ανοίγει το μήνυμα, οι πληροφορίες αποκρυπτογραφούνται ή μετατρέπονται ξανά στην αναγνώσιμη μορφή τους. Για να συμβεί αυτό, ο αποστολέας και ο παραλήπτης πρέπει και οι δύο να χρησιμοποιήσουν ένα κλειδί κρυπτογράφησης, το οποίο είναι μια συλλογή αλγορίθμων που κάνουν την κρυπτογράφηση και την αποκωδικοποίηση.

Παραδείγματα αλγορίθμων κρυπτογράφησης που χρησιμοποιούνται σήμερα περιλαμβάνουν:

- ✓ **Triple DES**: Ενισχύοντας το αρχικό DES (Data Encryption Standard), το οποίο ιδρύθηκε το 1977 και θεωρείται πλέον πολύ αδύναμο για την προστασία ευαίσθητων δεδομένων, το Triple DES εκτελεί κρυπτογράφηση 3 φορές – κρυπτογράφηση, αποκρυπτογράφηση και κρυπτογράφηση ξανά.

- ✓ **RSA (Rivest–Shamir–Adleman):** Παίρνοντας το όνομά της από τα αρχικά των 3 εφευρετών της επιστήμονες υπολογιστών (Rivest, Shamir και Adleman), η RSA χρησιμοποιεί έναν ισχυρό και ευρέως χρησιμοποιούμενο

αλγόριθμο για κρυπτογράφηση. Είναι δημοφιλές λόγω του μήκους του κλειδιού του και χρησιμοποιείται συνήθως για ασφαλή μετάδοση δεδομένων.

✓ **Προηγμένο Πρότυπο Κρυπτογράφησης (Advanced Encryption Standard (AES):** Το AES που χρησιμοποιείται παγκοσμίως, είναι το κυβερνητικό πρότυπο των ΗΠΑ από το 2002.

✓ **TwoFish:** Αυτό το δωρεάν λογισμικό κρυπτογράφησης χρησιμοποιείται σε υλικό και λογισμικό. Θεωρείται ότι είναι ένας από τους ταχύτερους αλγόριθμους κρυπτογράφησης.

# Κεφάλαιο 3 Τεχνητή Νοημοσύνη και Κυβερνοασφάλεια

## 3.1 Εισαγωγή

Η ταχεία πρόοδος της τεχνολογίας έχει μεταμορφώσει τη ζωή μας με πολλούς τρόπους, προσφέροντας ευκολία, αποτελεσματικότητα και αμέτρητες ευκαιρίες. Ωστόσο, αυτά τα οφέλη έχουν ένα τίμημα: ένα διαρκώς εξελισσόμενο ψηφιακό τοπίο που παρουσιάζει νέες προκλήσεις και απειλές. Καθώς ο κόσμος γίνεται πιο συνδεδεμένος και εξαρτάται από την τεχνολογία, η ασφάλεια στον κυβερνοχώρο γίνεται όλο και πιο σημαντική.

Η Τεχνητή Νοημοσύνη είναι ένας ζωτικός τομέας της υπολογιστικής βιολογίας που ασχολείται με τη γενική νοημοσύνη, τη διδασκαλία και την ικανότητα προσαρμογής σε μηχανήματα. Η έρευνα για τη Τεχνητή Νοημοσύνη στοχεύει στην ανάπτυξη μηχανημάτων που μπορούν να απλοποιήσουν εργασίες που απαιτούν ευφυΐα. Η ρύθμιση, η κατασκευή σχεδίων, ο προγραμματισμός, η γραφή, η έκφραση και η αναγνώριση προσώπου είναι μερικά μόνο παραδείγματα. Ως αποτέλεσμα, έχει εξελιχθεί σε έναν τομέα επιστήμης αφιερωμένου στην εξεύρεση λύσεων σε αυτά τα προβλήματα. Εκτός από την κατασκευή πολλών κοινών οικιακών πακέτων λογισμικού, συμβατικών παιχνιδιών και άλλες κονσόλες παιχνιδιών, τα

συστήματα Τεχνητής Νοημοσύνης χρησιμοποιούνται ευρέως στην οικονομία, την υγειονομική περίθαλψη, την τεχνολογία και τον στρατό.

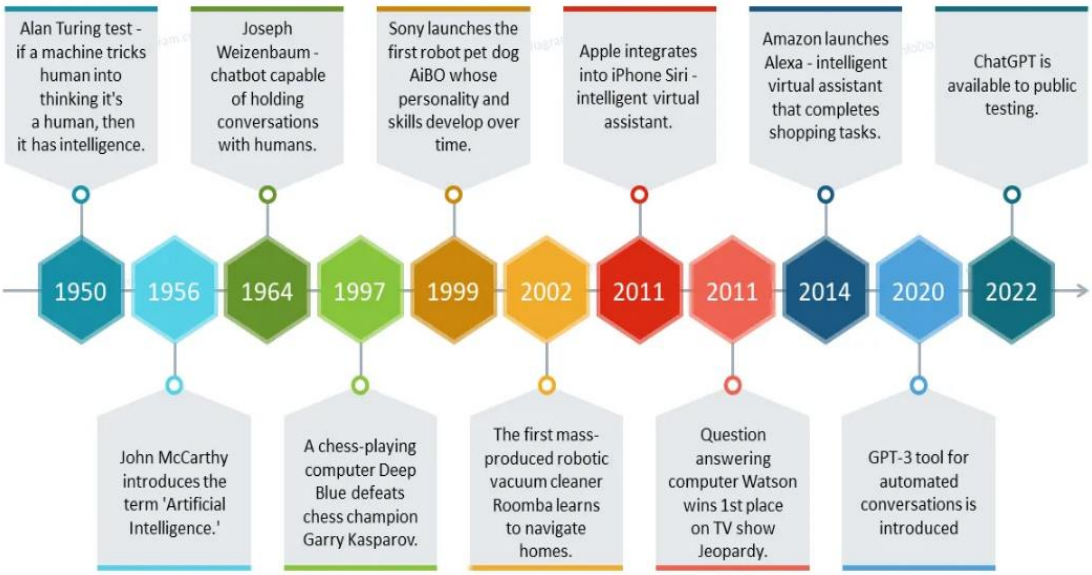
Το τοπίο της κυβερνοασφάλειας βρίσκεται σε συνεχή κατάσταση ροής, καθώς η τεχνολογία συνεχίζει να εξελίσσεται και οι φορείς απειλών προσαρμόζουν τις τακτικές τους για να εκμεταλλευτούν νέα τρωτά σημεία. Η υιοθέτηση της Τεχνητής Νοημοσύνης έχει αλλάξει σημαντικά αυτό το τοπίο, εισάγοντας τόσο νέες ευκαιρίες όσο και προκλήσεις που έχουν εκτεταμένες επιπτώσεις τόσο για τους επαγγελματίες της κυβερνοασφάλειας όσο και για τους χρήστες. Η Τεχνητή Νοημοσύνη, με την ικανότητά της να επεξεργάζεται τεράστιες ποσότητες δεδομένων και να λαμβάνει αποφάσεις με ταχύτητα άνευ προηγουμένου, έχει τη δυνατότητα να φέρει επανάσταση στον τρόπο με τον οποίο προστατεύουμε τα δίκτυά και τα ψηφιακά μας στοιχεία. Για παράδειγμα, τα εργαλεία που τροφοδοτούνται με τεχνικές Τεχνητής Νοημοσύνης μπορούν να αναλύουν μοτίβα κυκλοφορίας δικτύου για να εντοπίσουν ανωμαλίες και πιθανές εισβολές, επιτρέποντας στις ομάδες ασφαλείας να ανταποκρίνονται πιο γρήγορα και με ακρίβεια σε πιθανές απειλές.

Ωστόσο, οι ίδιες δυνατότητες που καθιστούν τη Τεχνητή Νοημοσύνη ισχυρό σύμμαχο στην καταπολέμηση του εγκλήματος στον κυβερνοχώρο μπορούν, αντίθετα να αξιοποιηθούν από κακόβουλους χρήστες για την ανάπτυξη πιο εξελιγμένων και στοχευμένων επιθέσεων. Οι κακόβουλους χρήστες χρησιμοποιούν τώρα τη Τεχνητή Νοημοσύνη για να αυτοματοποιήσουν τις επιθέσεις τους, καθιστώντας τις πιο γρήγορες, πιο προσαρμόσιμες και πιο δύσκολο να εντοπιστούν. Αυτό περιλαμβάνει καμπάνιες phishing με τη βοήθεια της Τεχνητής Νοημοσύνης, οι οποίες δημιουργήσουν μηνύματα ηλεκτρονικού ταχυδρομείου εξαιρετικά πιστού περιεχομένου, εργαλεία ανακάλυψης ευπάθειας με γνώμονα τη Τεχνητής Νοημοσύνης που μπορούν να εντοπίσουν και να εκμεταλλευτούν τα αδύνατα σημεία των συστημάτων, τα οποία μπορούν να παρακάμψουν τα μέτρα ασφαλείας προστασίας των δικτύων υπολογιστών.

## 3.2 Ιστορία της Τεχνητής Νοημοσύνης

Η πραγματική γέννηση και η προέλευση της Τεχνητής Νοημοσύνης μπορεί να είναι δύσκολο να εντοπιστούν, αλλά τα πρώτα παραδείγματα των βασικών αρχών της Τεχνητής Νοημοσύνης μπορούν να αναχθούν στη δεκαετία του 1940. Κατά τη διάρκεια αυτής της δεκαετίας, ο συγγραφέας Isaac Asimov έγραψε μυθιστορήματα για τα ρομπότ που μπορούσαν να μιμηθούν την ανθρώπινη συμπεριφορά και τη λήψη αποφάσεων (Haenlein & Kaplan, 2019). Ένα άλλο, πιο πρακτικό παράδειγμα κατά την ίδια δεκαετία, έλαβε χώρα στην Αγγλία, όπου ο Άγγλος μαθηματικός Άλαν Τούρινγκ δημιούργησε το "The Bombe". Αυτό το μηχάνημα, που θεωρείται επίσης ως ο πρώτος υπολογιστής, μπόρεσε να σπάσει και να αποκρυπτογραφήσει τον γερμανικό κώδικα «Enigma» κατά τη διάρκεια του Β' Παγκοσμίου Πολέμου. Ο Τούρινγκ δημοσίευσε αργότερα ένα άρθρο που περιγράφει τις μεθόδους δοκιμής των υπολογιστών και την ευφυΐα τους που εξακολουθεί να στέκεται ως εφαλτήριο για τη Τεχνητή Νοημοσύνη όπως τη γνωρίζουμε σήμερα (Haenlein & Kaplan, 2019).

Στην συνέχεια, ο Τούρινγκ έθεσε το ζήτημα της πιθανής νοημοσύνης μιας μηχανής για πρώτη φορά στο διάσημο άρθρο του 1950 «Υπολογιστικές Μηχανές και Νοημοσύνη (Computing Machinery and Intelligence)» και περιέγραψε ένα «παιχνίδι μίμησης (Game of Imitation)», όπου ένας άνθρωπος θα πρέπει να είναι σε θέση να διακρίνει μια συνομιλία μεταξύ άντρα και μηχανής (Turing, 1950). Όσο αμφιλεγόμενο κι αν είναι αυτό το άρθρο (αυτή η "Turing test" δεν φαίνεται να πληροί τις προϋποθέσεις για πολλούς ειδικούς), θα αναφέρεται συχνά ως η πηγή της αμφισβήτησης του ορίου μεταξύ του ανθρώπου και της μηχανής.




Εικόνα 4. Ιστορική Αναδρομή της Τεχνητής Νοημοσύνης. Πηγή: *(infoDiagram LTD, 2021)*

Μεταξύ 1964 και 1966 δημιουργήθηκε το πρώτο πρόγραμμα υπολογιστή ικανό να επεξεργάζεται φυσική γλώσσα. Αυτό το πρόγραμμα που ονομάζεται Eliza θα μπορούσε να μιμηθεί την ανθρώπινη γλώσσα και να προσομοιώσει έναν διάλογο με έναν πραγματικό άνθρωπο. Τα πρωτόγονα προγράμματα υπολογιστών επίλυσης προβλημάτων επινοήθηκαν επίσης κατά την ίδια εποχή και μπορούσαν να λύσουν αυτόματα ορισμένα παιχνίδια όπως το Towers of Hanoi (Bieszczad & Kuchar, 2020). Οι επιτυχίες στον τομέα οδήγησαν στη χρηματοδότηση της Τεχνητής Νοημοσύνης και της έρευνας σε αυτήν κατά τη δεκαετία του 1970, ενώ οι συνεχώς αυξανόμενες κατανομές περιουσιακών στοιχείων στην Τεχνητή Νοημοσύνη συνάντησαν κάποια απώθηση. Βέβαια, κάποιοι ήταν ενάντια στην περαιτέρω ανάπτυξη της Τεχνητής Νοημοσύνης καθώς υποστήριξαν ότι οι μηχανές δεν θα μπορούσαν ποτέ να επιτύχουν την πολυπλοκότητα της ανθρώπινης συμπεριφοράς (Anderson & Corbett, 1993).

Το μεγαλύτερο πρόβλημα που καθυστερούσε την πρόοδο της Τεχνητής Νοημοσύνης στην αρχή ήταν ο τρόπος με τον οποίο εφαρμόστηκε η πτυχή της ανθρώπινης συμπεριφοράς. Η αρχική προσέγγιση ήταν να δημιουργηθεί μια ιεραρχία, σαν στοίβα, για τη λήψη αποφάσεων με τη μορφή πολλαπλών δηλώσεων if-then. Αυτά τα συστήματα που βασίζονται σε κανόνες, για παράδειγμα τα έμπειρα συστήματα, λειτουργούν καλά σε ένα περιορισμένο

περιβάλλον όπως το σκάκι και άλλα παιχνίδια, και τη δεκαετία του 1990, μια μηχανή ήταν σε θέση να νικήσει τον Παγκόσμιο Πρωταθλητή του σκακιού με ευκολία. Το επόμενο βήμα για τη Τεχνητή Νοημοσύνη ήταν να έχει την ικανότητα να επεξεργάζεται εξωτερικά δεδομένα, να μαθαίνει από αυτά και να προσαρμόζεται στο μεταβαλλόμενο περιβάλλον στο οποίο λειτουργεί. Η προηγούμενη πρόταση παρέχει κατά προσέγγιση τον ορισμό της Τεχνητής Νοημοσύνης που χρησιμοποιούμε πλέον σήμερα (Norvig & Russell, 2021).

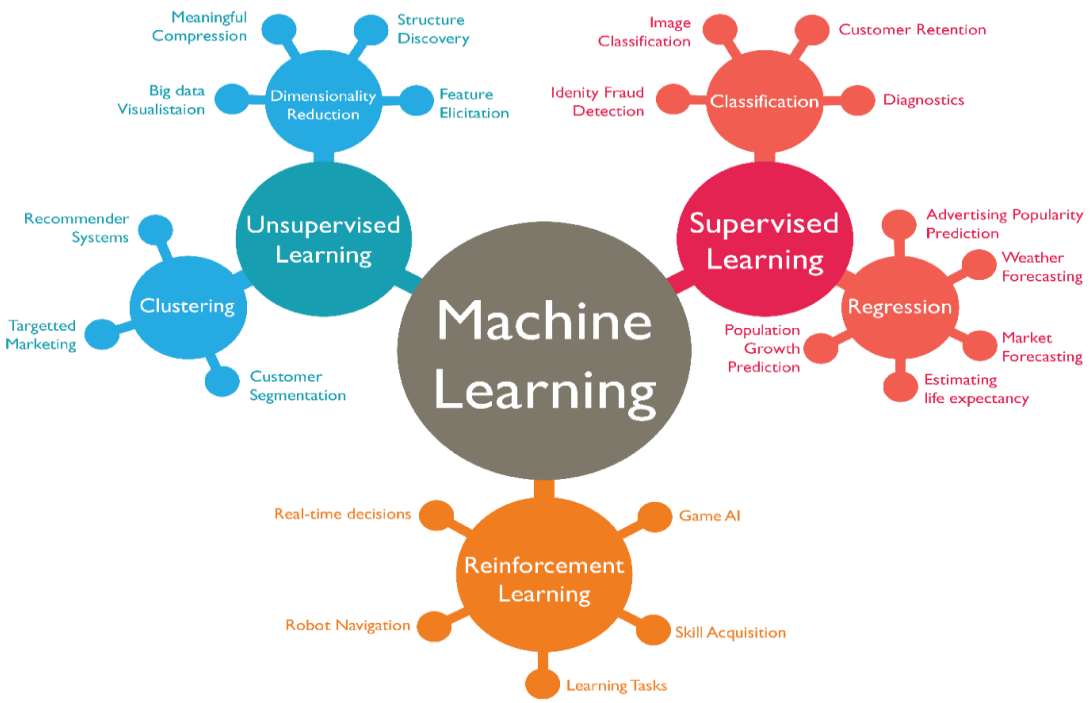
Από την αρχή της νέας χιλιετίας, η Τεχνητή Νοημοσύνη έχει δει κάποιες μεγάλες προόδους και οι εφευρέσεις των νευρωνικών δικτύων και η βαθιά μάθηση έχουν επιταχύνει περαιτέρω την εξέλιξη των προγραμμάτων Τεχνητής Νοημοσύνης. Αυτές οι δύο έννοιες θα εξηγηθούν αργότερα σε αυτή τη διατριβή καθώς είναι βασικές στις λύσεις Τεχνητής Νοημοσύνης που παρουσιάζονται σε αυτή τη μελέτη. Το 2015, ένας ακρογωνιαίος λίθος της τεχνολογίας Τεχνητής Νοημοσύνης δημιουργήθηκε όταν η Google ανέπτυξε το AlphaGo, το οποίο είναι ένα πρόγραμμα υπολογιστή που έχει σχεδιαστεί για να παίζει το παιχνίδι Go (Pumperla & Ferguson, 2019). Το AlphaGo χρησιμοποίησε την υπολογιστική δύναμη της μηχανικής μάθησης και των νευρωνικών δικτύων για να νικήσει τους καλύτερους παίκτες Go στον κόσμο και εξακολουθεί να θεωρείται ως ένας από τους πιο εξελιγμένους αλγόριθμους Τεχνητής Νοημοσύνης. Εφευρέσεις όπως η αναγνώριση ομιλίας και προσώπου, οι αλγόριθμοι διαφημίσεων και τα έξυπνα ηχεία αποτελούν πολύ πρόσφατες εξελίξεις στον τομέα της Τεχνητής Νοημοσύνης και το πλήρες δυναμικό των εργαλείων δεν έχει ακόμη διερευνηθεί (Russell & Norvig, 2020).

## 3.3 Η εξέλιξη των Τεχνολογιών Τεχνητής Νοημοσύνης

Καθώς οι τεχνολογίες Τεχνητής Νοημοσύνης συνεχίζουν να εξελίσσονται και να ωριμάζουν, οι εφαρμογές τους σε διάφορους τομείς, συμπεριλαμβανομένης της ασφάλειας στον κυβερνοχώρο, έχουν γίνει πιο ισχυρές και εξελιγμένες. Η εξέλιξη των τεχνολογιών Τεχνητής Νοημοσύνης έχει συμβάλει τόσο σε θετικές όσο και σε αρνητικές επιπτώσεις στην ασφάλεια στον κυβερνοχώρο.

Μερικές από τις βασικές εξελίξεις στις τεχνολογίες Τεχνητής Νοημοσύνης που έχουν επηρεάσει το τοπίο της κυβερνοασφάλειας περιλαμβάνουν:

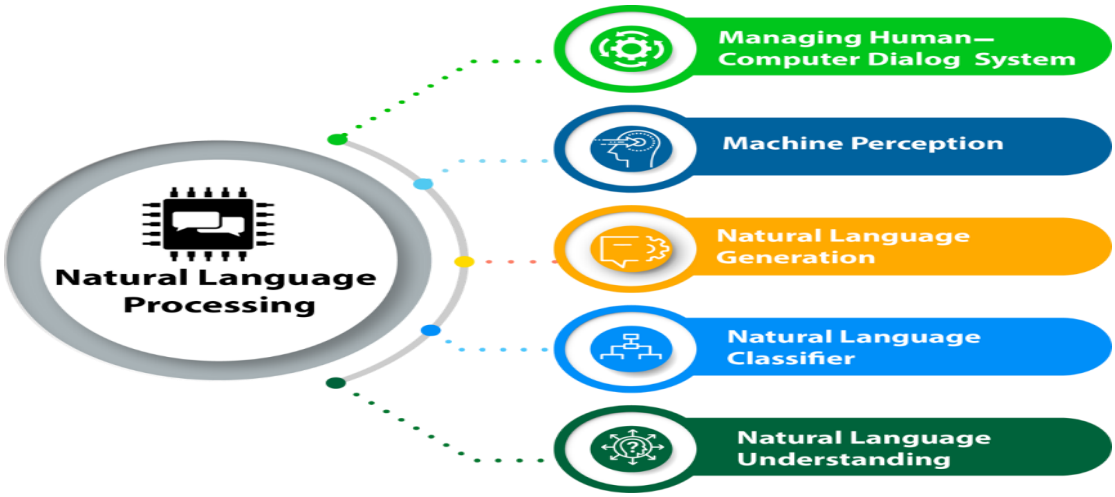
✓ **Μηχανική Μάθηση (Machine Learning) και Βαθιά Μάθηση (Deep Learning):** Οι αλγόριθμοι ML — και το πιο προηγμένο υποσύνολο τους, η βαθιά μάθηση — έχουν κάνει σημαντικά βήματα τα τελευταία χρόνια. Αυτές οι εξελίξεις επέτρεψαν την ανάπτυξη ισχυρών εργαλείων ασφαλείας με γνώμονα τη Τεχνητή Νοημοσύνη, ικανών να εντοπίζουν και να ανταποκρίνονται σε απειλές πιο αποτελεσματικά. Από την άλλη πλευρά, οι αντίπαλοι έχουν επίσης χρησιμοποιήσει αυτές τις τεχνικές για να δημιουργήσουν πιο εξελιγμένες και προσαρμοστικές επιθέσεις (Kubat, 2018).




Εικόνα 5. Τύποι Μηχανικής Μάθησης / Βαθιάς Μάθησης. Πηγή: *(Kubat, 2018)*

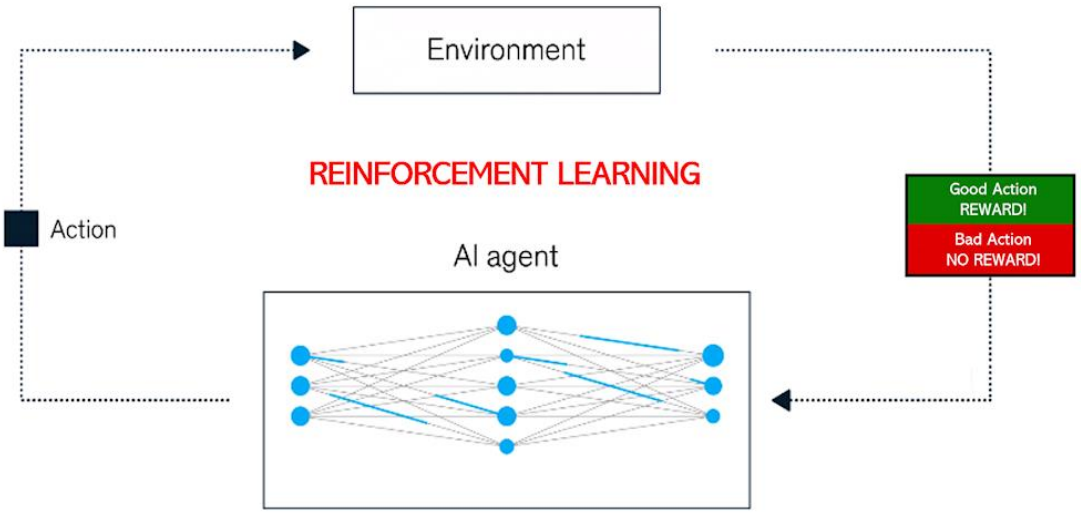
✓ **Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP):** Οι τεχνικές NLP έχουν βελτιωθεί δραματικά, επιτρέποντας στα συστήματα Τεχνητής Νοημοσύνης να κατανοούν και να επεξεργάζονται καλύτερα την ανθρώπινη γλώσσα. Αυτό οδήγησε στην ανάπτυξη προηγμένων

επιθέσεων κοινωνικής μηχανικής που αξιοποιούν περιεχόμενο που δημιουργείται από αλγορίθμους Τεχνητής Νοημοσύνης, όπως μηνύματα ηλεκτρονικού ψαρέματος και ψεύτικα βίντεο, καθιστώντας όλο και πιο δύσκολο για τους χρήστες να διακρίνουν μεταξύ γνήσιου και κακόβουλου περιεχομένου (Tunstall , et al., 2022).




Εικόνα 6. Χαρακτηριστικά Natural Language Processing - NLP. Πηγή: (Coursesteach, 2023)

✓ **Ενισχυτική Μάθηση (Reinforcement Learning - RL):** Το RL είναι ένας τομέας της Τεχνητής Νοημοσύνης που εστιάζει σε μοντέλα εκπαίδευσης για τη λήψη βέλτιστων αποφάσεων με βάση τη δοκιμή και το σφάλμα. Το RL έχει χρησιμοποιηθεί για τη δημιουργία εργαλείων κυβερνοασφάλειας με γνώμονα τη Τεχνητή Νοημοσύνη που μπορούν να προσαρμοστούν και να μάθουν από το περιβάλλον τους, βελτιώνοντας την αποτελεσματικότητά τους με την πάροδο του χρόνου. Ωστόσο, οι αντίπαλοι μπορούν επίσης να χρησιμοποιήσουν το RL για να αναπτύξουν στρατηγικές επίθεσης που μπορούν να παρακάμψουν τα παραδοσιακά μέτρα ασφαλείας και να προσαρμοστούν στην άμυνα που υπάρχει (DeAngelis, 2021).




Εικόνα 7. Ενισχυτική Μάθηση (Reinforcement Learning). Πηγή: *(DeAngelis, 2021)*

✓ **Παραγωγικό Αντιπαραθετικό Δίκτυο (Generative adversarial networks - GANs):** Τα GAN είναι ένας τύπος αρχιτεκτονικής βαθιάς μάθησης στην οποία δύο νευρωνικά δίκτυα, ένας γεννήτρια και ένας διαχωριστής, εκπαιδεύονται σε ανταγωνισμό μεταξύ τους. Τα GAN έχουν χρησιμοποιηθεί για τη δημιουργία ρεαλιστικών συνθετικών δεδομένων, όπως εικόνες, ήχος και κείμενο. Ενώ τα GAN έχουν πολυάριθμες νόμιμες εφαρμογές, μπορούν επίσης να χρησιμοποιηθούν από εγκληματίες του κυβερνοχώρου για να δημιουργήσουν ψεύτικο περιεχόμενο, να μιμηθούν νόμιμους χρήστες ή να δημιουργήσουν ρεαλιστικά μηνύματα ηλεκτρονικού ψαρέματος.

✓ **Αυτόνομοι και Ευφυείς Πράκτορες (Autonomous and Intelligent Agents:):** Οι αυτόνομοι και ευφυείς πράκτορες που βασίζονται στη Τεχνητή Νοημοσύνη έχουν τη δυνατότητα να φέρουν επανάσταση στον τρόπο με τον οποίο οι οργανισμοί διαχειρίζονται και ανταποκρίνονται σε περιστατικά ασφάλειας στον κυβερνοχώρο. Αυτοί οι πράκτορες μπορούν να αυτοματοποιήσουν χρονοβόρες εργασίες, όπως το κυνήγι απειλών και την απόκριση συμβάντων, επιτρέποντας στις ομάδες ασφαλείας να επικεντρωθούν σε πιο στρατηγικές πρωτοβουλίες. Ωστόσο, οι εγκληματίες του κυβερνοχώρου μπορούν επίσης να αναπτύξουν κακόβουλους αυτόνομους πράκτορες που

μπορούν να εντοπίσουν και να εκμεταλλευτούν αυτόνομα τις ευπάθειες, καθιστώντας τις επιθέσεις πιο γρήγορες και πιο δύσκολο να εντοπιστούν.

## 3.4 Ο ρόλος της Τεχνητής Νοημοσύνης στην ασφάλεια στον κυβερνοχώρο

Η Τεχνητή Νοημοσύνη έχει αποδειχθεί ότι είναι ένα κρίσιμο πλεονέκτημα για την αντιμετώπιση ανησυχιών για την ασφάλεια στον κυβερνοχώρο, προσφέροντας την ανάπτυξη ευφυών πρακτόρων για την αποτελεσματική αντιμετώπιση συγκεκριμένων προκλήσεων ασφαλείας. Ένας Ευφυής Πράκτορας, είτε με τη μορφή υλικού είτε λογισμικού, έχει σχεδιαστεί για να βελτιστοποιεί την πιθανότητα επίτευξης ενός καθορισμένου στόχου μέσω της ικανότητάς του να παρατηρεί, να μαθαίνει και να λαμβάνει τεκμηριωμένες αποφάσεις. Αυτοί οι Έξυπνοι Πράκτορες μπορούν να ανιχνεύσουν τρωτά σημεία σε πολύπλοκες δομές κώδικα, να εντοπίσουν παρατυπίες στα μοτίβα σύνδεσης των χρηστών και ακόμη και να αναγνωρίσουν αναδυόμενους τύπους κακόβουλου λογισμικού που αποφεύγουν τις συμβατικές μεθόδους ανίχνευσης.

Οι Ευφυείς Πράκτορες επεξεργάζονται επίσης τεράστιες ποσότητες δεδομένων για να μάθουν και να κατανοήσουν μοτίβα. Όταν αναπτύσσονται σε αμυντικά συστήματα, αυτοί οι πράκτορες εφαρμόζουν τις γνώσεις τους αναλύοντας εισερχόμενα δεδομένα, συμπεριλαμβανομένων πληροφοριών που δεν είχαν προηγουμένως εμφανιστεί. Η Τεχνητή Νοημοσύνη στην ασφάλεια στον κυβερνοχώρο βοηθά τους επαγγελματίες ασφάλειας αναγνωρίζοντας πολύπλοκα μοτίβα δεδομένων, παρέχοντας επίσης συστάσεις που μπορούν να εφαρμοστούν και επιτρέποντας τον αυτόνομο μετριασμό. Ενισχύει την ανίχνευση απειλών, υποστηρίζει τη λήψη αποφάσεων και επιταχύνει την απόκριση σε περιστατικά.

Η Τεχνητή Νοημοσύνη χρησιμοποιεί τρεις θεμελιώδεις μηχανισμούς για την αντιμετώπιση πολύπλοκων προβλημάτων ασφάλειας:

- ✓ **Πληροφορίες μοτίβων (Pattern Insights):** Η Τεχνητή Νοημοσύνη υπερέχει στην αναγνώριση και την ταξινόμηση μοτίβων δεδομένων που μπορεί να είναι δύσκολο για τους ανθρώπους να αναλύσουν. Παρουσιάζει

αυτά τα πρότυπα σε επαγγελματίες ασφαλείας για περαιτέρω εξέταση και ανάλυση.

- ✓ **Συστάσεις με δυνατότητα δράσης (Actionable Recommendations):** Οι ευφυείς πράκτορες προσφέρουν συστάσεις που να μπορούν να ενεργήσουν με βάση τα προσδιορισμένα πρότυπα, παρέχοντας στους επαγγελματίες ασφαλείας καθοδήγηση σχετικά με τα κατάλληλα μέτρα.

- ✓ **Αυτόνομος Μετριασμός (Autonomous Mitigation):** Ορισμένοι Ευφυείς Πράκτορες μπορούν να αναλάβουν άμεση δράση εκ μέρους επαγγελματιών ασφάλειας για την αντιμετώπιση και τη διόρθωση ζητημάτων ασφαλείας.

Ενώ ένας οργανισμός μπορεί να έχει ήδη ειδικευμένους επαγγελματίες ασφαλείας, προηγμένα εργαλεία και καθιερωμένες διαδικασίες, οι Ευφυείς Πράκτορες στοχεύουν να ενισχύσουν και να αυξήσουν αυτούς τους υπάρχοντες πόρους, ενισχύοντας τις συνολικές αμυντικές ικανότητες. Το αρχικό βήμα στην άμυνα είναι συχνά ο εντοπισμός τρωτών σημείων ή σφαλμάτων που θα μπορούσαν να εκμεταλλευτούν οι επιτιθέμενοι. Με τη βοήθεια της Τεχνητής Νοημοσύνης, η σάρωση του πηγαίου κώδικα γίνεται πιο ακριβής και δίνοντας τη δυνατότητα στους μηχανικούς να αποκαλύψουν σφάλματα ασφαλείας πριν από την ανάπτυξη εφαρμογών στο περιβάλλον παραγωγής.

Ο ρόλος της Τεχνητής Νοημοσύνης στην ασφάλεια στον κυβερνοχώρο εκτείνεται πέρα από τις παραδοσιακές μεθόδους, φέρνοντας επανάσταση στον τρόπο με τον οποίο οι οργανισμοί προστατεύουν τα συστήματα και τα δεδομένα τους. Αξιοποιώντας τη δύναμη της Τεχνητής Νοημοσύνης και της κυβερνοασφάλειας, οι επαγγελματίες ασφάλειας αποκτούν πρόσβαση σε βελτιωμένη ανίχνευση, προληπτικό μετριασμό απειλών και έξυπνο αυτοματισμό, επιτρέποντάς τους να παραμείνουν ένα βήμα μπροστά από τις απειλές στον κυβερνοχώρο σε ένα συνεχώς εξελισσόμενο τοπίο.

Η εφαρμογή της Τεχνητής Νοημοσύνης στην ασφάλεια στον κυβερνοχώρο έχει τεράστιες δυνατότητες για την αντιμετώπιση των

περίπλοκων προκλήσεων που αντιμετωπίζουμε σήμερα. Καθώς το τοπίο απειλών αρχίζει να μεγαλώνει και οι συσκευές γίνονται όλο και πιο διαδεδομένες, η Τεχνητή Νοημοσύνη και η μηχανική μάθηση μπορούν να διαδραματίσουν ζωτικό ρόλο στην καταπολέμηση των επιθέσεων στον κυβερνοχώρο, αυτοματοποιώντας την ανίχνευση και την απόκριση απειλών, ξεπερνώντας τις παραδοσιακές προσεγγίσεις που βασίζονται σε λογισμικό.

### 3.4.1 Η Τεχνητή Νοημοσύνη ως εργαλείο για κυβερνοεπιθέσεις

Ένας από τους βασικούς παράγοντες που επέτρεψαν την ύπαρξη του Διαδικτύου είναι η αποκεντρωμένη φύση του. Το Διαδίκτυο δεν ανήκει σε καμία οντότητα, γεγονός που καθιστά δύσκολο για οποιαδήποτε οντότητα να το τερματίσει ή να το ελέγξει. Αυτή η μοναδική πτυχή του Διαδικτύου οδήγησε εν μέρει στην επιτυχία του και επέτρεψε να γίνουν εφικτές νέες τεχνολογίες όπως η Τεχνητή Νοημοσύνη. Ωστόσο, καθώς η Τεχνητή Νοημοσύνη γίνεται πιο διαδεδομένη, το Διαδίκτυο μπορεί γρήγορα να γίνει ένα πολύ διαφορετικό μέρος. Για παράδειγμα, εάν η Τεχνητή Νοημοσύνη μπορεί να ελέγξει τη ροή πληροφοριών στο διαδίκτυο, θα μπορούσε να χρησιμοποιηθεί για να χειραγωγήσει την κοινή γνώμη (π.χ. να δώσει στους ανθρώπους ψευδείς πληροφορίες που οδηγούν σε νοοτροπία αγέλης) ή ακόμη και να προκαλέσει πόλεμο. Πιθανώς ένας από τους πιο διάσημους παράγοντες που οδήγησαν στο να γίνει εφικτό η Τεχνητή Νοημοσύνη ήταν η μοναδικότητα (Singularity). Η ιδιομορφία είναι μια κερδοσκοπική έννοια στην οποία η τεχνολογική ανάπτυξη γίνεται τόσο γρήγορη και πλήρης που διασχίζει ένα σημείο χωρίς επιστροφή, πυροδοτώντας αδιάκοπες τεχνολογικές αλλαγές. Το αποτέλεσμα είναι μια «μετα-ανθρώπινη» εποχή στην οποία οι ευφυείς μηχανές ξεπερνούν την ανθρώπινη νοημοσύνη.

Επίσης, η Τεχνητή Νοημοσύνη μπορεί να χρησιμοποιηθεί για τη δημιουργία κακόβουλου λογισμικού που μπορεί να αποφύγει τον εντοπισμό από λογισμικό προστασίας από ιούς. Μπορεί επίσης να χρησιμοποιηθεί για τη δημιουργία πλαστών προφίλ μέσων κοινωνικής δικτύωσης και τη διάδοση παραπληροφόρησης σε πλατφόρμες μέσων κοινωνικής δικτύωσης. Η Τεχνητή

Νοημοσύνη χρησιμοποιείται από τις κοινότητες του στρατού και των πληροφοριών για τον εντοπισμό συγκεκριμένων αντικειμένων σε μια φωτογραφία ή ένα βίντεο. Η πιθανότητα κατάχρησης της Τεχνητής Νοημοσύνης συμβαδίζει με τη δυνατότητά της να λαμβάνει αυτόνομες αποφάσεις, όπως πόσοι άνθρωποι θα πεθάνουν με βάση ένα προβλεπόμενο ποσοστό εγκληματικότητας. Η Τεχνητή Νοημοσύνη χρησιμοποιείται για την πρόβλεψη συντριβών στο χρηματιστήριο, μια μελέτη του 2019 έδειξε ότι πάνω από το 92% των συναλλαγών Forex γινόταν από αλγορίθμους Τεχνητής Νοημοσύνης και όχι από ανθρώπους. Πάνω από το 60% των συναλλαγών άνω των 10 εκατομμυρίων δολαρίων εκτελούνται επί του παρόντος με τη χρήση αλγορίθμων και ο αριθμός αυτός αναμένεται να αυξηθεί σημαντικά τα επόμενα τέσσερα χρόνια (Kissell, 2021).

## 3.5 Ασφάλεια της Τεχνητής Νοημοσύνης

Οι πρόσφατες εξελίξεις στην Τεχνητή Νοημοσύνη είναι μετασχηματιστικές και ήδη υπερβαίνουν τις επιδόσεις σε ανθρώπινο επίπεδο σε εργασίες όπως η αναγνώριση εικόνας, η επεξεργασία φυσικής γλώσσας και η ανάλυση δεδομένων. Οι οικονομικοί παράγοντες θα οδηγήσουν στην υιοθέτηση νέων εφαρμογών Τεχνητής Νοημοσύνης που διαταράσσουν σχεδόν κάθε πτυχή της επιχείρησης, τόσο καλή όσο και κακή. Τα συστήματα Τεχνητής Νοημοσύνης μπορούν να χειραγωγηθούν, να παρακαμφθούν και να παραπλανηθούν με αποτέλεσμα βαθιές επιπτώσεις στην ασφάλεια για εφαρμογές όπως εργαλεία παρακολούθησης δικτύου, οικονομικά συστήματα ή αυτόνομα οχήματα. Ως εκ τούτου, οι ασφαλείς και ανθεκτικές τεχνικές και οι βέλτιστες πρακτικές είναι ζωτικής σημασίας.

### 3.5.1 Προδιαγραφή και Επαλήθευση Συστημάτων Τεχνητής Νοημοσύνης

Τα ολοκληρωμένα συστήματα Τεχνητής Νοημοσύνης περιλαμβάνουν τέσσερα στοιχεία: αντίληψη, μάθηση, αποφάσεις και ενέργειες. Αυτά τα συστήματα λειτουργούν σε πολύπλοκα περιβάλλοντα που απαιτούν κάθε στοιχείο να

αλληλεπιδρά και να είναι αλληλεξάρτηση (π.χ. σφάλματα στην αντίληψη μπορεί να προκαλέσουν λανθασμένη απόφαση). Επιπλέον, υπάρχουν μοναδικά τρωτά σημεία σε καθένα από τα στοιχεία (π.χ. η αντίληψη είναι επιρρεπής σε εκπαιδευτικές επιθέσεις, ενώ οι αποφάσεις είναι επιρρεπείς σε κλασικές εκμεταλλεύσεις στον κυβερνοχώρο). Τέλος, η έννοια της ορθότητας δεν είναι ένα καθαρά λογικό ζήτημα και η αβεβαιότητα απαιτεί όρια για κάθε στοιχείο για την προστασία του συστήματος από κακή συμπεριφορά.

Έτσι, υπάρχει επιτακτική ανάγκη για επίσημες μεθόδους για την επαλήθευση των στοιχείων της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης, τόσο ανεξάρτητα όσο και από κοινού, καθώς σχετίζονται με τη λογική ορθότητα, τη θεωρία αποφάσεων και την ανάλυση κινδύνου. Απαιτούνται νέες τεχνικές που προσδιορίζουν τι αναμένεται να κάνει ένα σύστημα και πώς πρέπει να ανταποκρίνεται στην επίθεση. Στα παραδοσιακά συστήματα, οι ιδιότητες που ταιριάζουν με την προδιαγραφή είναι εφικτές για κάθε εξάρτημα. Επειδή τα συστήματα Τεχνητής Νοημοσύνης είναι τόσο πολύπλοκα, η εφαρμογή και η διαμόρφωσή τους είναι δύσκολο να αξιολογηθούν. Απαιτείται έρευνα σε αρχιτεκτονικές δομές και τεχνικές ανάλυσης που επιτρέπουν την επαλήθευση αυτών των στοιχείων και αποτελεί μέρος μιας ευρύτερης προσπάθειας για την ανάπτυξη διαχειρίσιμων προτύπων, βέλτιστων πρακτικών, εργαλείων και μεθόδων για τον συλλογισμό σχετικά με τη συμπεριφορά ενός συστήματος.

### 3.5.2 Αξιόπιστη Λήψη Αποφάσεων με Τεχνητή Νοημοσύνη

Καθώς τα συστήματα Τεχνητής Νοημοσύνης αναπτύσσονται σε περιβάλλοντα υψηλής αξίας, το ζήτημα της διασφάλισης ότι η διαδικασία λήψης αποφάσεων είναι αξιόπιστη, ιδιαίτερα σε αντίθετα σενάρια, είναι πρωταρχικής σημασίας. Ενώ υπάρχουν πολυάριθμες απεικονίσεις τρωτών σημείων της Μηχανικής Μάθησης, οι τεχνικές που βασίζονται στην επιστήμη για την πρόβλεψη της αξιοπιστίας είναι αόριστες. Απαιτείται έρευνα για την ανάπτυξη μεθόδων και αρχών για ένα ευρύ φάσμα συστημάτων Τεχνητής Νοημοσύνης, συμπεριλαμβανομένης της Μηχανικής Μάθησης, του σχεδιασμού, του

συλλογισμού και της αναπαράστασης γνώσης. Οι τομείς που πρέπει να αντιμετωπιστούν για τη λήψη αξιόπιστων αποφάσεων περιλαμβάνουν τον καθορισμό μετρήσεων απόδοσης, την ανάπτυξη τεχνικών, τη δυνατότητα επεξήγησης και υπευθυνότητας των συστημάτων Τεχνητής Νοημοσύνης, τη βελτίωση της εκπαίδευσης και του συλλογισμού σε συγκεκριμένο τομέα και τη διαχείριση δεδομένων εκπαίδευσης.

Η έρευνα του μοντέλου απειλών πρέπει να εντοπίσει μετρήσιμες ιδιότητες που καθορίζουν την αξιοπιστία, έτσι ώστε ένας υπερασπιστής να μπορεί να ενσωματώσει την ευρωστία, το απόρρητο και τη δικαιοσύνη στους αλγόριθμους λήψης αποφάσεων. Δεδομένου ενός συγκεκριμένου μοντέλου απειλής, το σύστημα θα πρέπει να αιτιολογήσει την αντίπαλη παρέμβαση και να καθορίσει τις απαραίτητες προϋποθέσεις για την επίτευξη αυτών των ιδιοτήτων αξιοπιστίας. Οι δυνατότητες περιλαμβάνουν την προσαρμογή ορισμών από την κρυπτογραφία ή την ασφάλεια υπολογιστών, την ενοποίηση ιδιοτήτων σε ένα ενιαίο πλαίσιο συλλογιστικής και την αντιμετώπισή τους ως παραλλαγές μιας ενιαίας έννοιας σταθερότητας στα συστήματα Μηχανικής Μάθησης και Ευφυΐας τόσο για τη λήψη αποφάσεων όσο και για τα μοντέλα ασφάλειας ευρύτερα.

Η νέα κατανόηση του πόσο ευάλωτα στοιχεία Τεχνητής Νοημοσύνης εγείρει ανησυχίες σχετικά με την ασφάλεια ολόκληρου του αγωγού επεξεργασίας δεδομένων στον οποίο χρησιμοποιούνται. Τα στοιχεία Τεχνητής Νοημοσύνης αψηφούν τη συμβατική ανάλυση λογισμικού και μπορούν να εισάγουν νέα διανύσματα επίθεσης σε περιβάλλοντα όπου λειτουργούν οι αλγόριθμοι Τεχνητής Νοημοσύνης, υλοποιήσεις πλαισίων και εφαρμογών Τεχνητής Νοημοσύνης, μοντέλα Μηχανικής Μάθησης και δεδομένα εκπαίδευσης. Λόγω των κρυφών εξαρτήσεων στον αγωγό, μπορούν να πραγματοποιηθούν πολλαπλές εφαρμογές. Απαιτείται έρευνα για την ανάπτυξη θεωρίας, αρχών μηχανικής και βέλτιστων πρακτικών κατά τη χρήση της Τεχνητής Νοημοσύνης ως συστατικού ενός συστήματος. Αυτό θα πρέπει να περιλαμβάνει μοντελοποίηση απειλών, εργαλεία ασφαλείας, ευπάθειες τομέα και διασφάλιση της ομαδοποίησης ανθρώπινης μηχανής. Αυτά τα μοντέλα πρέπει να επιτρέπουν επαναληπτικές αφαιρέσεις επιθέσεων και βελτιώσεων, να

σχεδιάζονται σύμφωνα με έναν ειδικό της Τεχνητής Νοημοσύνης και να λαμβάνουν υπόψη τη διαθεσιμότητα και την ακεραιότητα των δεδομένων, τους ελέγχους πρόσβασης, την ενορχήστρωση και λειτουργία δικτύου, την επίλυση ανταγωνιστικών ενδιαφερόντων, το απόρρητο και ένα δυναμικό περιβάλλον πολιτικής.

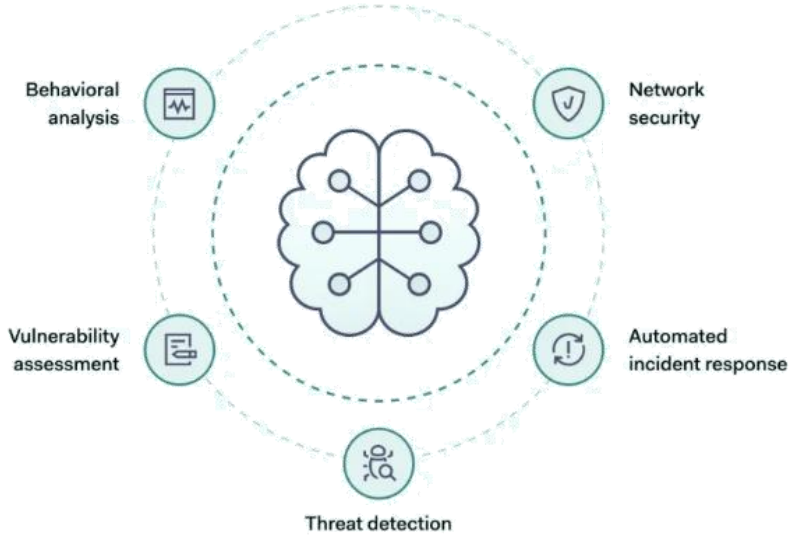
Για να καταστούν πιο αξιόπιστα τα συστήματα Τεχνητής Νοημοσύνης, οι αρχές μηχανικής θα πρέπει να βασίζονται στην επιστήμη, την κοινοτική εμπειρία και την έρευνα λειτουργικότητας που περιλαμβάνει πλεονασμό, εποπτικά και άλλα πλαίσια. Η κατανόηση των συνθηκών, των απειλών, των τομέων και των περιορισμών είναι απαραίτητοι αλλά επικουρικοί στόχοι.

Καθώς οι τεχνολογίες Τεχνητής Νοημοσύνης γίνονται πανταχού παρούσες, άνθρωποι και μηχανές θα συνεργάζονται απρόσκοπτα για να βελτιώσουν την αποτελεσματικότητα και την ακρίβεια κρίσιμων εργασιών (π.χ. βοηθώντας τους γιατρούς να διαγνώσουν ασθένειες ή τους δασκάλους που προσαρμόζονται στις ανάγκες των μεμονωμένων μαθητών). Η πρόκληση είναι ότι το μηχάνημα ή η λειτουργικότητα του ανθρώπου μπορεί να αυξηθεί ή να υποβαθμιστεί από πολλούς παράγοντες. Απαιτείται περαιτέρω έρευνα για να βοηθήσει τόσο τη μηχανή όσο και τον άνθρωπο να συνεργαστούν, να παρακολουθήσουν και να αξιολογήσουν ο ένας την απόδοση και την αξιοπιστία του άλλου. Τι γίνεται αν ένας άνθρωπος δεν μπορεί να ανταποκριθεί αρκετά γρήγορα σε μια κρίσιμη, ευαίσθητη στο χρόνο, ανθρώπινη εφαρμογή; Τι γίνεται αν τα αποτελέσματα του μηχανήματος και του ανθρώπου διαφωνούν; Θεωρία, τεχνικές και μετρήσεις απαιτούνται για την υποστήριξη σύνθετων αποφάσεων, σε πραγματικό χρόνο, όπου οι πληροφορίες είναι ασαφείς ή υποκειμενικές και όταν μια καθυστερημένη απάντηση θα μπορούσε να έχει σοβαρές συνέπειες.

# Κεφάλαιο 4 Αλγόριθμοι Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης στην Κυβερνοασφάλεια

## 4.1 Αλγόριθμοι και Εργαλεία της Τεχνητής Νοημοσύνης

Η Τεχνητή Νοημοσύνη έχει σημειώσει αξιοσημείωτη πρόοδο τα τελευταία χρόνια και έχει αποδείξει την αξία της σε διάφορους τομείς, συμπεριλαμβανομένης της ασφάλειας στον κυβερνοχώρο. Με την αύξηση των απειλών στον κυβερνοχώρο και την αυξανόμενη πολυπλοκότητα των κυβερνοεπιθέσεων, η Τεχνητή Νοημοσύνη έχει γίνει κεντρικό εργαλείο για την προστασία από το έγκλημα στον κυβερνοχώρο. Τα ολοκληρωμένα συστήματα Τεχνητής Νοημοσύνης έχουν τη δυνατότητα να εκπαιδεύονται για την αυτόματη αναγνώριση απειλών στον κυβερνοχώρο, την ειδοποίηση των χρηστών και την προστασία ευαίσθητων πληροφοριών των επιχειρήσεων (NordLayer, 2023).




Εικόνα 8. Αλγόριθμοι και Εργαλεία της Τεχνητής Νοημοσύνης. Πηγή: (NordLayer, 2023)

Η Τεχνητή Νοημοσύνη συνδυάζει μεγάλα σύνολα δεδομένων και τα χρησιμοποιεί με διαισθητικούς αλγόριθμους επεξεργασίας. Καθώς το εύρος των δικτύων και των συστημάτων διευρύνεται, η Τεχνητή Νοημοσύνη στην

ασφάλεια στον κυβερνοχώρο βοηθά στην αυτοματοποίηση των λειτουργιών, επεξεργάζοντας μεγάλες ποσότητες δεδομένων πολύ πιο γρήγορα από ό,τι θα μπορούσε ποτέ ένας άνθρωπος. Για αυτόν τον λόγο, τα περισσότερα εργαλεία κυβερνοασφάλειας ενσωματώνουν βαθιά μάθηση και άλλες δυνατότητες που προορίζονται για εργασία με μεγάλα δεδομένα. Οι κύριοι τρόποι με τους οποίους η Τεχνητή Νοημοσύνη χρησιμοποιείται στην ασφάλεια στον κυβερνοχώρο είναι οι κάτωθι:

- ✓ **Ανίχνευση Απειλών.** Η Τεχνητή Νοημοσύνη μπορεί να λειτουργήσει ως φίλτρο για την ανάλυση αρχείων και κώδικα λογισμικού για τον εντοπισμό πιθανών απειλών κακόβουλου λογισμικού, αποφεύγοντας τα ψευδώς θετικά. Οι αλγόριθμοι Μηχανικής Μάθησης μπορούν να εκπαιδευτούν για την ανίχνευση απειλών ώστε να αναγνωρίζουν μοτίβα και χαρακτηριστικά γνωστού κακόβουλου λογισμικού και να επισημαίνουν οποιονδήποτε νέο κώδικα ταιριάζει με αυτά τα μοτίβα.

- ✓ **Ασφάλεια Δικτύου.** Οι αλγόριθμοι Τεχνητής Νοημοσύνης μπορούν να αναλύσουν δεδομένα κίνησης δικτύου για να ανιχνεύσουν μοτίβα και ανωμαλίες που υποδεικνύουν απόπειρα εισβολής ή επίθεσης. Επίσης μπορεί να επισημάνει τυχόν αποκλίσεις από αυτήν τη γραμμή βάσης ως πιθανές απειλές, μαθαίνοντας πώς μοιάζουν τα κανονικά μοτίβα κυκλοφορίας δικτύου.

- ✓ **Ανάλυση Συμπεριφοράς.** Η Τεχνητή Νοημοσύνη μπορεί να χρησιμοποιηθεί για την ανάλυση της συμπεριφοράς των χρηστών και τον εντοπισμό κακόβουλων ενεργειών που μπορεί να υποδηλώνουν μη εξουσιοδοτημένη πρόσβαση ή κακόβουλη δραστηριότητα χρησιμοποιώντας μηχανική εκμάθηση. Αυτό επιτρέπει την πιο αποτελεσματική παρακολούθηση της δραστηριότητας των χρηστών και τον εντοπισμό πιθανών απειλών.

- ✓ **Αυτοματοποιημένη Απόκριση Συμβάντων**. Τα συστήματα που βασίζονται σε Τεχνητή Νοημοσύνη μπορούν να χρησιμοποιηθούν για αυτόματη απόκριση σε απειλές που έχουν εντοπιστεί, όπως τερματισμός συνδέσεων, καραντίνα μολυσμένων μηχανημάτων και απενεργοποίηση

λογαριασμών χρηστών. Τα προηγμένα μοντέλα μηχανικής εκμάθησης συμβάλλουν στον περιορισμό των προσπαθειών «hacking» και στην ελαχιστοποίηση πιθανών ζημιών.

✓ **Αξιολόγηση Ευπάθειας.** Οι αλγόριθμοι Τεχνητή Νοημοσύνη μπορούν να εντοπίσουν πιθανές ευπάθειες σε συστήματα και δίκτυα. Αυτό επιτρέπει τη λήψη προληπτικών μέτρων για τον μετριασμό των πιθανών απειλών προτού μπορέσουν να χρησιμοποιηθούν.

### 4.1.1 Ανίχνευση Απειλών και Ανάλυση Συμπεριφοράς

Η ανίχνευση απειλών και η ανάλυση συμπεριφοράς είναι απαραίτητες για τον εντοπισμό και την απόκριση σε κυβερνοεπιθέσεις σε πραγματικό χρόνο. Με την εφαρμογή της Τεχνητής Νοημοσύνης σε αυτούς τους τομείς, η ασφάλεια στον κυβερνοχώρο έχει σημειώσει σημαντική βελτίωση στην αποτελεσματικότητα και την ακρίβεια ανίχνευσης.

Ο όγκος των δεδομένων που επεξεργάζονται οι οργανισμοί (δημόσιοι ή ιδιωτικοί) σε καθημερινή βάση είναι τεράστιος. Η μη αυτόματη ανίχνευση απειλών σε τέτοιους τόμους είναι σχεδόν αδύνατη. Οι σύγχρονες επιθέσεις στον κυβερνοχώρο χρησιμοποιούν συχνά μυστικές τακτικές, όπως η lateral movement και η low-profile persistence, γεγονός που καθιστά δύσκολο τον εντοπισμό τους με παραδοσιακές μεθόδους.

Έτσι, αντί να βασίζεται μόνο σε γνωστές υπογραφές κακόβουλου λογισμικού, η Τεχνητή Νοημοσύνη εστιάζει σε μοτίβα μη φυσιολογικής συμπεριφοράς. Αυτό καθιστά δυνατό τον εντοπισμό προηγουμένως άγνωστων απειλών ή παραλλαγών κακόβουλου λογισμικού που έχουν ελαφρώς τροποποιηθεί. Αναλύοντας τη συμπεριφορά των χρηστών και του συστήματος, η Τεχνητή Νοημοσύνη μπορεί να εντοπίσει ασυνήθιστη δραστηριότητα, όπως η πρόσβαση σε αρχεία σε περίεργες ώρες ή η ασυνήθιστη μεταφορά μεγάλων ποσοτήτων δεδομένων.

Η ανίχνευση απειλών συμπεριφοράς έχει γνωρίσει ταχεία αύξηση στη δημοτικότητα και την υιοθέτηση, και έχουν αναδυθεί ορισμένα εργαλεία και

συστήματα, τόσο εμπορικά όσο και ανοιχτού κώδικα, που ειδικεύονται σε αυτήν την προσέγγιση. Μερικά από τα πιο δημοφιλή εργαλεία παρατίθενται παρακάτω:

✓ **Darktrace**: Το Darktrace χρησιμοποιεί Μηχανική Μάθηση και αλγόριθμους Τεχνητής Νοημοσύνης για τον εντοπισμό, την απόκριση και τον μετριασμό των απειλών σε πραγματικό χρόνο με βάση μοτίβα ασυνήθιστης συμπεριφοράς. Το εργαλείο είναι γνωστό για το "Enterprise Immune System", το οποίο «μαθαίνει» και καθιερώνει αυτό που μπορεί να γίνει κατανοητό ως μια κατάσταση "business as usual" στο δίκτυο και στη συνέχεια εντοπίζει αποκλίσεις από αυτόν τον κανόνα.

✓ **Vectra:** Το Vectra προσφέρει ανίχνευση απειλών σε πραγματικό χρόνο χρησιμοποιώντας Τεχνητή Νοημοσύνη. Επικεντρώνεται στον εντοπισμό κακόβουλης συμπεριφοράς εντός της κυκλοφορίας του δικτύου και παρέχει μια λεπτομερή εικόνα της συνεχιζόμενης αλυσίδας επιθέσεων, επιτρέποντας στις ομάδες ασφαλείας να ανταποκρίνονται γρήγορα.

✓ **CrowdStrike Falcon**: Η CrowdStrike είναι γνωστή για τις λύσεις προστασίας τελικού σημείου (endpoint). Η πλατφόρμα Falcon χρησιμοποιεί τεχνικές που βασίζονται στη συμπεριφορά για να ανιχνεύσει και να αποτρέψει απειλές που ενδέχεται να παραλείψουν άλλα συστήματα που βασίζονται σε υπογραφές.

✓ **Cylance:** Το CylancePROTECT είναι μια λύση προστασίας τελικού σημείου που χρησιμοποιεί μοντέλα Τεχνητής Νοημοσύνης για τον εντοπισμό και τον αποκλεισμό κακόβουλου λογισμικού με βάση τα χαρακτηριστικά και τις συμπεριφορές του και όχι με γνωστές υπογραφές.

✓ **Gurucul:** Παρέχει λύσεις ανάλυσης συμπεριφοράς χρηστών και οντοτήτων (User and Entity Behavioural Analytics - UEBA) που χρησιμοποιούν αλγόριθμους Μηχανικής Μάθησης για τον εντοπισμό εσωτερικών απειλών, απάτης και μη εξουσιοδοτημένης πρόσβασης.

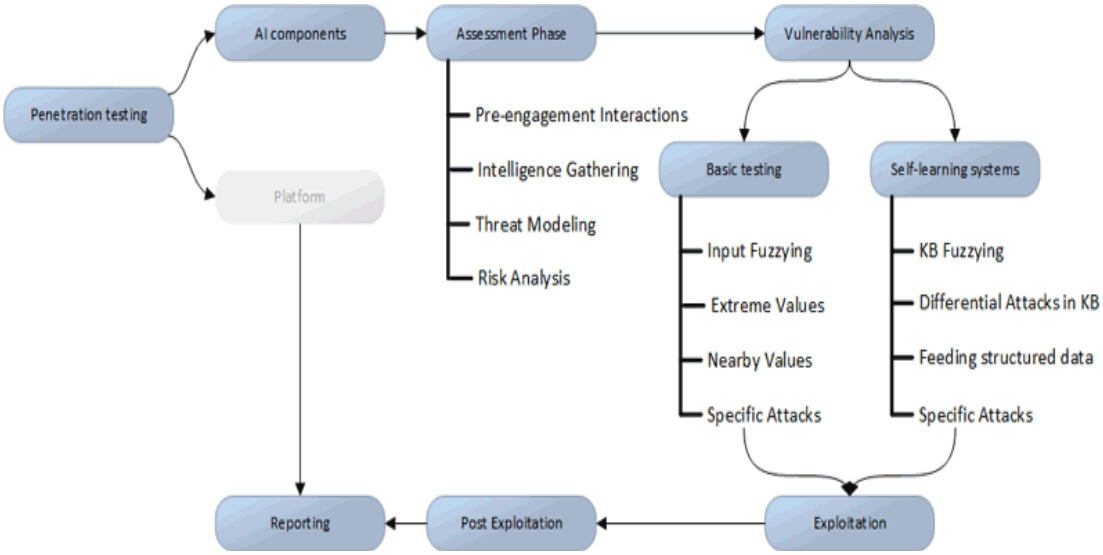
✓ **Wazuh:** Πρόκειται για μια πλατφόρμα ανοιχτού κώδικα για ανίχνευση απειλών, διαχείριση ευπάθειας και παρακολούθηση ακεραιότητας. Χρησιμοποιεί κανόνες και αποκωδικοποιητές για να αναλύει συμβάντα ασφαλείας και να ανιχνεύει ασυνήθιστη συμπεριφορά.

✓ **Snort:** Αν και πιο γνωστό ως σύστημα ανίχνευσης και πρόληψης εισβολής (Intrusion Detection and Prevention System - IDPS), το Snort έχει εξελιχθεί για να ενσωματώνει ικανότητες που βασίζονται στη συμπεριφορά. Η κοινότητα Snort αναπτύσσει και μοιράζεται νέους κανόνες που μπορούν να ανιχνεύσουν ανώμαλη συμπεριφορά.

### 4.1.2 Σάρωση Ευπάθειας και Αυτοματοποιημένη Διενέργεια Δοκιμών

Όπως είναι γνωστό, η ανάλυση τρωτών σημείων είναι μια συστηματική διαδικασία αξιολόγησης, εντοπισμού και ταξινόμησης τρωτών σημείων ασφαλείας στα πληροφοριακά συστήματα. Αυτά τα τρωτά σημεία μπορεί να προκληθούν από σφάλματα λογισμικού, ανεπαρκείς διαμορφώσεις, αστοχίες υλικού ή ακόμα και κακές πρακτικές διαχείρισης ασφάλειας.

Αυτή η διαδικασία σάρωσης συνήθως περιλαμβάνει ταυτοποίηση (εργαλεία σάρωση συστημάτων, δικτύων και εφαρμογών για γνωστά τρωτά σημεία), ταξινόμηση (αφού εντοπιστούν, τα τρωτά σημεία ταξινομούνται ανάλογα με τη σοβαρότητα και τον κίνδυνο), αποκατάσταση (προτείνονται λύσεις για τον μετριασμό ή την επιδιόρθωση των ανιχνευόμενων τρωτών σημείων) και επαλήθευση (μετά την αποκατάσταση, πραγματοποιείται περαιτέρω επαλήθευση για να επιβεβαιωθεί ότι τα τρωτά σημεία έχουν αντιμετωπιστεί επαρκώς).




Εικόνα 9. Προσέγγιση υψηλού επιπέδου για τη δοκιμή διείσδυσης συστημάτων Τεχνητής Νοημοσύνης. Πηγή: *(Weidman, 2014)*

Η δοκιμή διείσδυσης (Penetration Testing), κοινώς γνωστή ως pentesting, είναι μια προσομοιωμένη επίθεση σε ένα σύστημα με στόχο την ανακάλυψη τρωτών σημείων προτού το κάνουν οι πραγματικοί εισβολείς. Σε αντίθεση με την ανάλυση ευπάθειας, η οποία συνήθως χρησιμοποιεί αυτοματοποιημένες σαρώσεις για τον εντοπισμό γνωστών τρωτών σημείων, η διείσδυση συχνά περιλαμβάνει ειδικούς που προσπαθούν ενεργά να εκμεταλλευτούν τις ευπάθειες και να διεισδύσουν σε συστήματα, προσομοιώνοντας τις τακτικές, τις τεχνικές και τις διαδικασίες (Tactics, Techniques and Procedures - TTP) υπαρκτών αντιπάλων (Weidman, 2014).

Η διαδικασία γενικά περιλαμβάνει αναγνώριση (συλλογή πληροφοριών για τον στόχο), σάρωση (εντοπισμός πιθανών σημείων εισόδου), διείσδυση (εκμετάλλευση τρωτών σημείων), συντήρηση πρόσβασης (προσομοίωση κινήσεων του εισβολέα μετά την απόκτηση πρόσβασης) και ανάλυση (που περιέχει την αναφορά ευρημάτων και συστάσεις για οχύρωση το σύστημα). Βέβαια, η Τεχνητή Νοημοσύνη έχει επίσης ενσωματωθεί στην ανάλυση ευπάθειας και στις δοκιμές διείσδυσης, με τις ακόλουθες διαδικασίες (Cordero & Pascual, 2023):

| | |
|---|---|
| **Βελτιωμένος Αυτοματισμός** | Με τη Τεχνητή Νοημοσύνη, τα εργαλεία μπορούν να σαρώσουν δίκτυα και συστήματα πιο γρήγορα και με μεγαλύτερη ακρίβεια, εντοπίζοντας τρωτά σημεία που μπορεί να χάνουν τα παραδοσιακά εργαλεία. |
| **Συνεχής Μάθηση** | Τα εργαλεία που βασίζονται στη Τεχνητή Νοημοσύνη μπορούν να μάθουν από κάθε σάρωση, προσαρμόζονται σε νέα τρωτά σημεία και τεχνικές επίθεσης. |
| **Προηγμένη Προσομοίωση** | Στο pentesting, η Τεχνητή Νοημοσύνη μπορεί να προσομοιώσει πιο περίπλοκη συμπεριφορά εισβολέα, δοκιμάζοντας συστήματα έναντι αναδυόμενων και προηγμένων απειλών. |
| **Προτεραιοποίηση Κινδύνων** | Η Τεχνητή Νοημοσύνη μπορεί να βοηθήσει στην ιεράρχηση των τρωτών σημείων με βάση το πλαίσιο και τα ιστορικά δεδομένα, επιτρέποντας στις ομάδες ασφαλείας να επικεντρωθούν στις πιο επικείμενες ή επιζήμιες απειλές. |
| **Ένταξη / Συσχέτιση** | Οι λύσεις που βασίζονται στη Τεχνητή Νοημοσύνη μπορούν να συσχετίσουν δεδομένα από πολλές πηγές, προσφέροντας μια πιο ολιστική άποψη της στάσης ασφαλείας ενός οργανισμού. |

Πίνακας 1. Διαδικασίες Ανάλυσης Τρωτότητας και Δοκιμών Διείσδυσης. Πηγή: (Cordero & Pascual, 2023)

Βέβαια, εργαλεία όπως το Tenable.io, η Qualys Cloud Platform ή το Checkmarx χρησιμοποιούν ήδη δυνατότητες Τεχνητής Νοημοσύνης για να βελτιώσουν τη σάρωση και την ανάλυσή τους. Επιπλέον, πλατφόρμες που δοκιμάζουν, όπως η Cobalt, ενσωματώνουν Τεχνητή Νοημοσύνη για να αυτοματοποιήσουν και να βελτιώσουν μέρη της διαδικασίας. Η ενσωμάτωση της Τεχνητής Νοημοσύνης σε αυτούς τους τομείς είναι πολλά υποσχόμενη, αλλά, προς το παρόν, ο συνδυασμός ανθρώπινων ειδικών με αυτά τα προηγμένα εργαλεία παρέχει την πιο ισχυρή και ολοκληρωμένη προσέγγιση για την ασφάλεια στον κυβερνοχώρο.

## 4.2 Ταξινόμηση Αλγορίθμων Μηχανικής Μάθησης στην ασφάλεια του κυβερνοχώρου

Η υιοθέτηση και η διεισδυτικότητα των μηχανισμών της Μηχανικής Μάθησης ολοένα και αυξάνεται ενώ, παράλληλα, βελτιώνονται οι υπάρχουσες μέθοδοι και η ικανότητά τους να κατανοούν και να αντιμετωπίζουν πραγματικά ζητήματα στον τομέα της κυβερνοασφάλειας. Αυτά τα επιτεύγματα οδήγησαν στην υιοθέτηση της μηχανικής μάθησης σε διάφορους τομείς, όπως η ασφάλεια των υπολογιστών, η ιατρική ανάλυση, τα παιχνίδια και το μάρκετινγκ μέσων κοινωνικής δικτύωσης (Jordan & Mitchell, 2015). Σε ορισμένα σενάρια, οι τεχνικές μηχανικής μάθησης αντιπροσωπεύουν την καλύτερη επιλογή έναντι των παραδοσιακών αλγορίθμων που βασίζονται σε κανόνες και ακόμη και των ανθρώπινων χειριστών (LeCun, et al., 2015). Αυτή η τάση επηρεάζει επίσης τον τομέα της ασφάλειας στον κυβερνοχώρο όπου ορισμένα συστήματα ανίχνευσης αναβαθμίζονται με στοιχεία Μηχανικής Μάθησης (Buczak & Guven, 2015). Αν και η δημιουργία ενός πλήρως αυτοματοποιημένου συστήματος άμυνας στον κυβερνοχώρο είναι ακόμη ένας μακρινός στόχος, οι χειριστές πρώτου επιπέδου στα Επιχειρησιακά Κέντρα Δικτύων και Ασφάλειας (Network Operations Center (NOC) και Security Operations Center SOC) μπορούν να επωφεληθούν από εργαλεία ανίχνευσης και ανάλυσης που βασίζονται στη μηχανική μάθηση. Αυτή η ενότητα κεφάλαιο απευθύνεται ειδικά σε χειριστές ασφάλειας και στοχεύει να αξιολογήσει την τρέχουσα ωριμότητα αυτών των λύσεων, να εντοπίσει τους κύριους περιορισμούς τους και να επισημάνει κάποια περιθώρια βελτίωσης.

Η ΜΜ περιλαμβάνει μια μεγάλη ποικιλία παραδειγμάτων σε συνεχή εξέλιξη, παρουσιάζοντας αδύναμα όρια και διασταυρούμενες σχέσεις. Επιπλέον, διαφορετικές απόψεις και εφαρμογές μπορεί να οδηγήσουν σε διαφορετικές ταξινομήσεις. Ως εκ τούτου, δεν είναι ευρέως αποδεκτή μια πλήρως αποδεκτή ταξινόμηση από τη βιβλιογραφία, αλλά, μια πρωτότυπη ταξινόμηση ικανή να καταγράψει τις διαφορές μεταξύ των μυριάδων τεχνικών που εφαρμόζονται στην ανίχνευση του κυβερνοχώρου, όπως φαίνεται στο Σχήμα 8. Αυτή η ταξινόμηση είναι ειδικά προσανατολισμένη στους χειριστές ασφαλείας και αποφεύγει τον φιλόδοξο στόχο της παρουσίασης της τελικής ταξινόμησης που

μπορεί να ικανοποιήσει όλους τους ειδικούς Τεχνητής Νοημοσύνης και τις περιπτώσεις εφαρμογών. Η πρώτη διάκριση που αποδεικνύεται στην Σχήμα 10 είναι μεταξύ των παραδοσιακών αλγορίθμων ΜΜ, οι οποίοι σήμερα μπορούν να αναφέρονται ως Shallow Learning (SL), σε αντίθεση με την πιο πρόσφατη Deep Learning (DL). Η Shallow Learning απαιτεί έναν ειδικό τομέα (δηλαδή, έναν μηχανικό χαρακτηριστικών) που μπορεί να εκτελέσει την κρίσιμη εργασία του προσδιορισμού των σχετικών χαρακτηριστικών δεδομένων πριν από την εκτέλεση του αλγόριθμου SL. Η Deep Learning βασίζεται σε μια πολυεπίπεδη αναπαράσταση των δεδομένων εισόδου και μπορεί να εκτελέσει την επιλογή χαρακτηριστικών, με αυτόνομο τρόπο, μέσω μιας εκμάθησης αναπαράστασης που ορίζεται από τη διαδικασία.

Οι προσεγγίσεις SL και DL μπορούν περαιτέρω να χαρακτηριστούν με τη διάκριση μεταξύ εποπτευόμενων και μη εποπτευόμενων αλγορίθμων. Οι προηγούμενες τεχνικές απαιτούν μια εκπαιδευτική διαδικασία με ένα μεγάλο και αντιπροσωπευτικό σύνολο δεδομένων που έχουν προηγουμένως ταξινομηθεί από έναν άνθρωπο ή με άλλα μέσα. Οι τελευταίες προσεγγίσεις δεν απαιτούν προ-επισημασμένο σύνολο δεδομένων εκπαίδευσης. Σε αυτή την ενότητα, εξετάζουμε και συγκρίνουμε τις πιο δημοφιλείς κατηγορίες αλγορίθμων ML, οι οποίοι εμφανίζονται ως φύλλα του δέντρου ταξινόμησης στην εικόνα 8. Παρατηρούμε ότι κάθε κατηγορία μπορεί να περιλαμβάνει δεκάδες διαφορετικές τεχνικές.

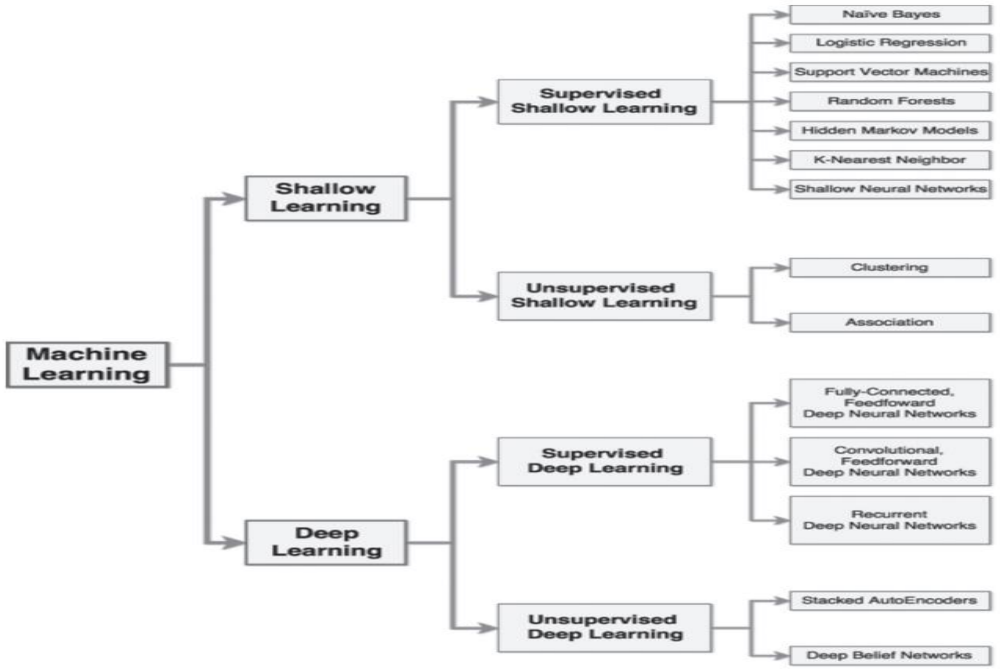
### 4.2.1 Shallow Learning

**Εποπτευόμενοι Αλγόριθμοι SL**

- ✓ **Naïve Bayes (NB).** Αυτοί οι αλγόριθμοι είναι πιθανολογικοί ταξινομητές που κάνουν την a-priori υπόθεση ότι τα χαρακτηριστικά του συνόλου δεδομένων εισόδου είναι ανεξάρτητα μεταξύ τους. Είναι επεκτάσιμα και δεν απαιτούν τεράστια σύνολα δεδομένων εκπαίδευσης για να παράγουν αξιόλογα αποτελέσματα.

✓ **Logistic Regression (LR)**. Αυτοί είναι κατηγορικοί ταξινομητές που υιοθετούν ένα διακριτικό μοντέλο. Όπως οι αλγόριθμοι NB, οι μέθοδοι LR κάνουν την εκ των προτέρων υπόθεση ανεξαρτησίας των χαρακτηριστικών εισόδου. Η απόδοσή τους εξαρτάται σε μεγάλο βαθμό από το μέγεθος των δεδομένων εκπαίδευσης.

✓ **Υποστήριξη Vector Machines (SVM).** Αυτοί είναι μη πιθανολογικοί ταξινομητές που χαρτογραφούν δείγματα δεδομένων σε έναν χώρο χαρακτηριστικών με στόχο τη μεγιστοποίηση της απόστασης μεταξύ κάθε κατηγορίας δειγμάτων. Δεν κάνουν καμία υπόθεση για τα χαρακτηριστικά εισόδου, αλλά έχουν κακή απόδοση σε ταξινομήσεις πολλαπλών τάξεων. Ως εκ τούτου, θα πρέπει να χρησιμοποιούνται ως δυαδικοί ταξινομητές. Η περιορισμένη επεκτασιμότητα τους μπορεί να οδηγήσει σε μεγάλους χρόνους επεξεργασίας.




Εικόνα 10. Ταξινόμηση Αλγορίθμων ΜΜ για εφαρμογές κυβερνοασφάλειας. Πηγή: *(Apruzzese, et al., 2018)*

✓ **Random Forest (RF).** Ένα τυχαίο «δάσος» είναι ένα σύνολο δέντρων απόφασης και εξετάζει την έξοδο κάθε δέντρου πριν δώσει μια ενοποιημένη τελική απόκριση. Κάθε απόφαση (δέντρο) είναι ένας ταξινομητής υπό όρους:

το δέντρο επισκέπτεται από την κορυφή και, σε κάθε κόμβο, μια δεδομένη συνθήκη ελέγχεται σε σχέση με ένα ή περισσότερα χαρακτηριστικά των αναλυόμενων δεδομένων. Αυτές οι μέθοδοι είναι αποτελεσματικές για μεγάλα σύνολα δεδομένων και υπερέχουν σε προβλήματα πολλαπλών κλάσεων, αλλά τα βαθύτερα δέντρα μπορεί να οδηγήσουν σε υπερβολική προσαρμογή (Şen , 2023).

- ✓ **Hidden Markov Models (HMM).** Αυτά μοντελοποιούν το σύστημα ως ένα σύνολο καταστάσεων που παράγουν εκροές με διαφορετικές πιθανότητες. Ο στόχος είναι να προσδιοριστεί η ακολουθία των καταστάσεων που παρήγαγαν τα παρατηρούμενα αποτελέσματα. Τα HMM είναι αποτελεσματικά για την κατανόηση της χρονικής συμπεριφοράς των παρατηρήσεων και για τον υπολογισμό της πιθανότητας μιας δεδομένης ακολουθίας γεγονότων. Παρόλο που το HMM μπορεί να εκπαιδευτεί σε σύνολα δεδομένων με ετικέτα ή χωρίς ετικέτα, στην ασφάλεια στον κυβερνοχώρο έχουν χρησιμοποιηθεί ως επί το πλείστον με επισημασμένα σύνολα δεδομένων.

- ✓ **K-Nearest Neighbour (KNN).** Τα KNN χρησιμοποιούνται για ταξινόμηση και μπορούν να χρησιμοποιηθούν για προβλήματα πολλαπλών τάξεων. Βέβαια, τόσο η εκπαίδευσή τους όσο και η φάση δοκιμής είναι υπολογιστικά απαιτητικές για την ταξινόμηση κάθε δείγματος δοκιμής.

- ✓ **Shallow Neural Network (SNN).** Αυτοί οι αλγόριθμοι βασίζονται σε νευρωνικά δίκτυα, τα οποία αποτελούνται από ένα σύνολο στοιχείων επεξεργασίας (δηλαδή νευρώνες) οργανωμένα σε δύο ή περισσότερα επίπεδα επικοινωνίας. Το SNN περιλαμβάνει όλους εκείνους τους τύπους νευρωνικών δικτύων με περιορισμένο αριθμό νευρώνων και επιπέδων. Παρά την ύπαρξη SNN χωρίς επίβλεψη, στην ασφάλεια στον κυβερνοχώρο έχουν χρησιμοποιηθεί κυρίως για εργασίες ταξινόμησης (Bishop & Bishop, 2023).

**Μη εποπτευόμενοι Αλγόριθμοι SL**

✓ **Clustering.** Οι αλγόριθμοι Clustering ομαδοποιούν σημεία δεδομένων που παρουσιάζουν παρόμοια χαρακτηριστικά. Οι πολύ γνωστές προσεγγίσεις περιλαμβάνουν το k-means και την ιεραρχική ομαδοποίηση. Οι μέθοδοι ομαδοποίησης έχουν περιορισμένη επεκτασιμότητα, αλλά αντιπροσωπεύουν μια ευέλικτη λύση που χρησιμοποιείται συνήθως ως προκαταρκτική φάση πριν από την υιοθέτηση ενός εποπτευόμενου αλγόριθμου ή για σκοπούς ανίχνευσης ανωμαλιών (Hsiao & Chang, 2008).

✓ **Association.** Στοχεύουν στον εντοπισμό άγνωστων μοτίβων μεταξύ των δεδομένων, καθιστώντας τα κατάλληλα για σκοπούς πρόβλεψης. Ωστόσο, τείνουν να παράγουν υπερβολικό αποτέλεσμα όχι απαραίτητα έγκυρων κανόνων, επομένως πρέπει να συνδυάζονται με ακριβείς επιθεωρήσεις από έναν άνθρωπο ειδικό (Abdelhamid, et al., 2014).

### 4.3.2 Deep Learning

Όλοι οι αλγόριθμοι DL βασίζονται στα Deep Neural Networks (DNN), τα οποία είναι μεγάλα νευρωνικά δίκτυα οργανωμένα σε πολλά επίπεδα ικανά για αυτόνομη εκμάθηση αναπαράστασης.

**Εποπτευόμενοι Αλγόριθμοι DL**

✓ **Πλήρως συνδεδεμένα Feedforward Deep Neural Networks (FNN).** Είναι μια παραλλαγή του DNN όπου κάθε νευρώνας συνδέεται με όλους τους νευρώνες στο προηγούμενο στρώμα. Το FNN δεν κάνει καμία υπόθεση για τα δεδομένα εισόδου και παρέχει μια ευέλικτη και γενικής χρήσης λύση για ταξινόμηση, σε βάρος του υψηλού υπολογιστικού κόστους (Dahl, et al., 2013).

✓ **Convolutional Feedforward Deep Neural Networks (CNN).** Είναι μια παραλλαγή του DNN όπου κάθε νευρώνας λαμβάνει την είσοδο του μόνο

από ένα υποσύνολο νευρώνων του προηγούμενου στρώματος. Αυτό το χαρακτηριστικό καθιστά το CNN αποτελεσματικό στην ανάλυση χωρικών δεδομένων, αλλά η απόδοσή του μειώνεται όταν εφαρμόζεται σε μη χωρικά δεδομένα. Το CNN έχει χαμηλότερο υπολογιστικό κόστος από το FNN (Hill & Bellekens, 2017).

✓ **Recurrent Deep Neural Networks (RNN).** Μια παραλλαγή του DNN του οποίου οι νευρώνες μπορούν να στείλουν την έξοδο τους και σε προηγούμενα επίπεδα. Αυτός ο σχεδιασμός τους κάνει πιο δύσκολο να εκπαιδεύσουν από το FNN. Υπερέχουν ως γεννήτριες ακολουθιών, ειδικά η πρόσφατη παραλλαγή τους, η μακροπρόθεσμη μνήμη (Pascanu, et al., 2015).

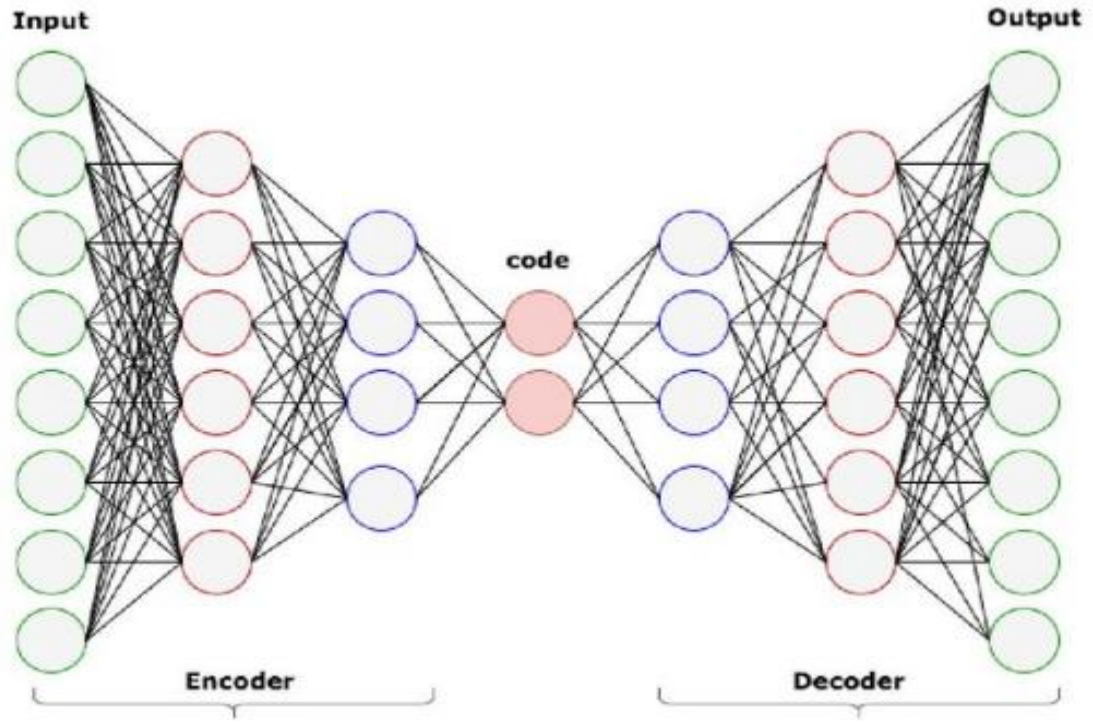
**Μη εποπτευόμενοι Αλγόριθμοι DL**

✓ **Deep Belief Networks (DBN).** Μοντελοποιούνται μέσω μιας σύνθεσης περιορισμένων μηχανών Boltzmann (Restricted Boltzmann Machines - RBM), μιας κατηγορίας νευρωνικών δικτύων χωρίς στρώμα εξόδου. Τα DBN είναι ένας τύπος αλγόριθμου βαθιάς μάθησης που αντιμετωπίζει τα προβλήματα που σχετίζονται με τα κλασικά νευρωνικά δίκτυα. Το κάνουν αυτό χρησιμοποιώντας στρώματα στοχαστικών λανθάνουσας μεταβλητής, που συνθέτουν το δίκτυο. Αυτές οι δυαδικές λανθάνουσες μεταβλητές, ή οι ανιχνευτές χαρακτηριστικών και οι κρυφές μονάδες, είναι δυαδικές μεταβλητές και είναι γνωστές ως στοχαστικές επειδή μπορούν να λάβουν οποιαδήποτε τιμή εντός ενός συγκεκριμένου εύρους με κάποια πιθανότητα (Li, et al., 2015).

✓ **Stacked Autoencoders (SAE).** Ένας SAE αποτελεί ένα νευρωνικό δίκτυο που αποτελείται από πολλαπλά επίπεδα αυτόματων κωδικοποιητών, όπου κάθε επίπεδο εκπαιδεύεται στην έξοδο του προηγούμενου. Αυτή η «στοίβαξη» των αυτόματων κωδικοποιητών επιτρέπει στο δίκτυο να μάθει πιο σύνθετες αναπαραστάσεις των δεδομένων εισόδου (Hardy, et al., 2016).

Ένας αυτόματος κωδικοποιητής αποτελείται από δύο κύρια μέρη: έναν κωδικοποιητή και έναν αποκωδικοποιητή. Ο κωδικοποιητής μειώνει τη

διάσταση των δεδομένων εισόδου (κωδικοποίηση) και ο αποκωδικοποιητής ανακατασκευάζει τα αρχικά δεδομένα από αυτή τη μειωμένη αναπαράσταση (αποκωδικοποίηση). Ο στόχος ενός αυτόματου κωδικοποιητή είναι να ελαχιστοποιήσει τη διαφορά μεταξύ της αρχικής εισόδου και της ανακατασκευασμένης εξόδου, ένα μέτρο γνωστό ως σφάλμα ανακατασκευής (Paper, 2021).




Εικόνα 11. Δομή στοιβαγμένων αυτόματων κωδικοποιητών. Πηγή: *(Paper, 2021)*

### 4.3 Εφαρμογές Αλγορίθμων Μηχανικής Μάθησης στην ασφάλεια του κυβερνοχώρου

Η ανίχνευση εισβολής (Intrusion Detection) στοχεύει στην ανακάλυψη παράνομων δραστηριοτήτων εντός ενός υπολογιστή ή ενός δικτύου μέσω των Συστημάτων Ανίχνευσης Εισβολής (Intrusion Detection Systems - IDS). Τα IDS δικτύου έχουν αναπτυχθεί ευρέως σε σύγχρονα εταιρικά δίκτυα. Αυτά τα συστήματα βασίζονταν παραδοσιακά σε πρότυπα γνωστών επιθέσεων, αλλά οι σύγχρονες αναπτύξεις περιλαμβάνουν άλλες προσεγγίσεις για τον εντοπισμό ανωμαλιών, τον εντοπισμό απειλών και την ταξινόμηση με βάση τη μηχανική

μάθηση. Εντός της ευρύτερης περιοχής ανίχνευσης εισβολής, δύο συγκεκριμένα προβλήματα σχετίζονται με την ανάλυσή μας: η ανίχνευση των botnets και των αλγορίθμων δημιουργίας τομέα Domain Generation Algorithms - DGA). Ένα botnet είναι ένα δίκτυο μολυσμένων μηχανημάτων που ελέγχονται από εισβολείς και χρησιμοποιούνται κατάχρηση για τη διεξαγωγή πολλαπλών παράνομων δραστηριοτήτων. Η ανίχνευση botnet στοχεύει στον εντοπισμό επικοινωνιών μεταξύ μολυσμένων μηχανημάτων εντός του παρακολουθούμενου δικτύου και των εξωτερικών διακομιστών εντολών και ελέγχου. Παρά τις πολλές ερευνητικές προτάσεις και τα εμπορικά εργαλεία που αντιμετωπίζουν αυτήν την απειλή, εξακολουθούν να υπάρχουν αρκετά botnet.

Οι DGA δημιουργούν αυτόματα ονόματα τομέα και συχνά χρησιμοποιούνται από ένα μολυσμένο μηχάνημα για επικοινωνία με εξωτερικούς διακομιστές δημιουργώντας περιοδικά νέα ονόματα κεντρικών υπολογιστών. Αντιπροσωπεύουν μια πραγματική απειλή για τους οργανισμούς επειδή, μέσω της DGA που βασίζεται σε τεχνικές επεξεργασίας γλώσσας, είναι δυνατό να αποφευχθούν οι άμυνες που βασίζονται σε στατικές μαύρες λίστες ονομάτων τομέα. Εξετάζουμε τεχνικές ανίχνευσης DGA που βασίζονται σε τεχνικές ΜΜ (Antonakakis, et al., 2012).

Η ανάλυση κακόβουλου λογισμικού είναι ένα εξαιρετικά σημαντικό πρόβλημα, επειδή το σύγχρονο κακόβουλο λογισμικό μπορεί να δημιουργήσει αυτόματα νέες παραλλαγές με τα ίδια κακόβουλα αποτελέσματα, αλλά να εμφανίζονται ως εντελώς διαφορετικά εκτελέσιμα αρχεία. Αυτά τα πολυμορφικά και μεταμορφικά χαρακτηριστικά νικούν τις παραδοσιακές προσεγγίσεις αναγνώρισης κακόβουλου λογισμικού που βασίζονται σε κανόνες. Οι τεχνικές ΜΜ μπορούν να χρησιμοποιηθούν για την ανάλυση παραλλαγών κακόβουλου λογισμικού και την απόδοση τους στη σωστή οικογένεια κακόβουλου λογισμικού.

Η ανίχνευση ανεπιθύμητων μηνυμάτων και phishing περιλαμβάνει ένα μεγάλο σύνολο τεχνικών που στοχεύουν στη μείωση της σπατάλης χρόνου και των πιθανών κινδύνων που προκαλούνται από ανεπιθύμητα μηνύματα

ηλεκτρονικού ταχυδρομείου. Σήμερα, τα ανεπιθύμητα μηνύματα ηλεκτρονικού ταχυδρομείου, δηλαδή το ηλεκτρονικό ψάρεμα, αντιπροσωπεύουν τον προτιμώμενο τρόπο μέσω του οποίου ένας εισβολέας δημιουργεί ένα πρώτο βήμα σε ένα εταιρικό δίκτυο. Τα μηνύματα ηλεκτρονικού ψαρέματος περιλαμβάνουν κακόβουλο λογισμικό ή συνδέσμους προς παραβιασμένους ιστότοπους. Ο εντοπισμός ανεπιθύμητων μηνυμάτων και ηλεκτρονικού ψαρέματος γίνεται όλο και πιο δύσκολος λόγω των προηγμένων στρατηγικών αποφυγής που χρησιμοποιούνται από τους εισβολείς για να παρακάμψουν τα παραδοσιακά φίλτρα. Οι προσεγγίσεις μηχανικής μάθησης μπορούν να βελτιώσουν τη διαδικασία εντοπισμού ανεπιθύμητων μηνυμάτων.

| | | **Ανίχνευση Εισβολής** | | | **Ανάλυση Κακόβουλου Λογισμικού** | **Ανίχνευση Ανεπιθύμητων Μηνυμάτων** |
|---|---|---|---|---|---|---|
| | | **Δίκτυο** | **Botnet** | **DGA** | | |
| **Deep Learning** | **Εποπτευόμενοι Αλγόριθμοι** | RNN | RNN | | FNN<br>CNN<br>RNN | |
| | **Αλγόριθμοι χωρίς επίβλεψη** | DBN<br>SAE | | | DBN<br>SAE | DBN<br>SAE |
| **Shallow Learning** | **Εποπτευόμενοι Αλγόριθμοι** | RF<br>NB<br>SVM<br>LR<br>HMM<br>KNN<br>SNN | RF<br>NB<br>SVM<br>LR<br>KNN<br>SNN | RF<br>HMM | RF<br>NB<br>SVM<br>LR<br>HMM<br>KNN<br>SNN | RF<br>NB<br>SVM<br>LR<br>KNN<br>SNN |

| | | | | | | |
|---|---|---|---|---|---|---|
| | **Αλγόριθμοι χωρίς επίβλεψη** | Clustering Association | Clustering | Clustering | Clustering Association | Clustering Association |

Πίνακας 2. Εφαρμογή της Μηχανικής Μάθησης σε προβλήματα κυβερνοασφάλειας. Πηγή: *(Apruzzese, et al., 2018)*

Στο πίνακα 2 φαίνονται οι κυριότεροι αλγόριθμοι ΜΜ που έχουν προταθεί για την αντιμετώπιση των προβλημάτων ασφάλειας στον κυβερνοχώρο που εντοπίστηκαν προηγουμένως. Σε αυτόν τον πίνακα, οι σειρές αναφέρουν την οικογένεια αλγορίθμων που παρουσιάζονται σε αυτήν την ενότητα, ενώ οι στήλες υποδηλώνουν ζητήματα στον κυβερνοχώρο. Κάθε κελί υποδεικνύει ποιοι αλγόριθμοι ΜΜ χρησιμοποιούνται για κάθε πρόβλημα. Τα κενά κελιά υποδηλώνουν ότι δεν υπάρχει ακόμη πρόταση για αυτήν την κατηγορία προβλημάτων. Από αυτόν τον πίνακα, προκύπτει ότι οι αλγόριθμοι SL εφαρμόζονται σε όλα τα εξεταζόμενα προβλήματα. Οι εποπτευόμενοι αλγόριθμοι DL βρίσκουν ευρεία εφαρμογή στην ανάλυση κακόβουλου λογισμικού, λιγότερο στην ανίχνευση εισβολών ενώ η ανίχνευση ανεπιθύμητων μηνυμάτων βασίζεται μόνο σε αλγόριθμους DL χωρίς επίβλεψη. Παρά τη σχέση του με την επεξεργασία φυσικής γλώσσας (LeCun, et al., 2015), δεν εφαρμόζεται αλγόριθμος DL στην ανίχνευση DGA. Όπως αναμενόταν, ο συνολικός αριθμός αλγορίθμων που βασίζονται στο DL είναι σημαντικά μικρότερος από αυτούς που βασίζονται στο SL. Πράγματι, οι προτάσεις DL που βασίζονται σε τεράστια νευρωνικά δίκτυα είναι πιο πρόσφατες από τις προσεγγίσεις SL. Αυτό το κενό ανοίγει πολλές ερευνητικές ευκαιρίες**.**

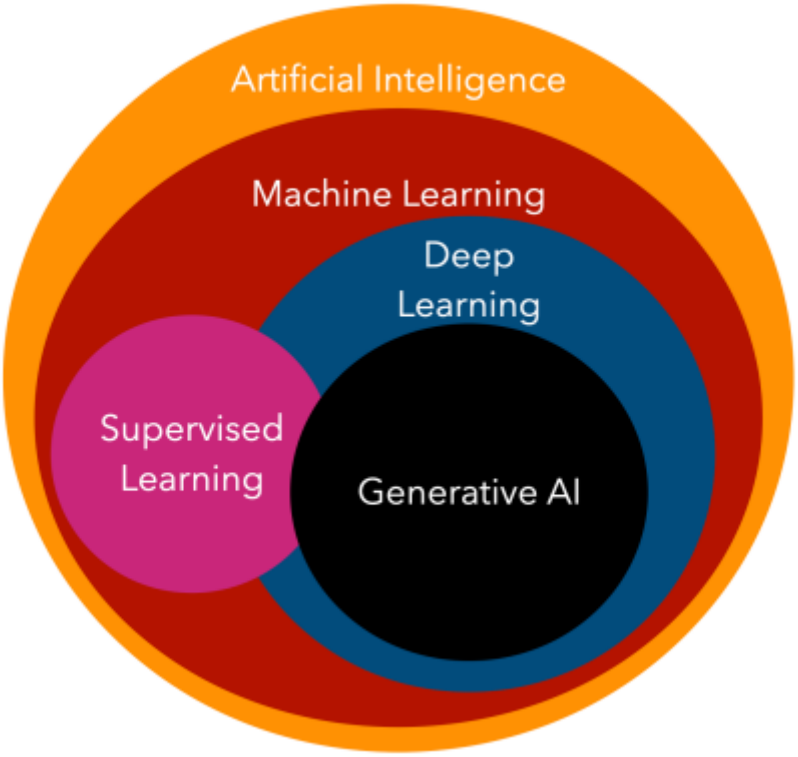
Τέλος, επισημαίνεται μια σημαντική διαφορά μεταξύ εποπτευόμενων και μη εποπτευόμενων προσεγγίσεων: οι προηγούμενοι αλγόριθμοι χρησιμοποιούνται για σκοπούς ταξινόμησης και μπορούν να εφαρμόσουν πλήρεις ανιχνευτές. Οι τελευταίες τεχνικές εκτελούν βοηθητικές δραστηριότητες (Rieck, et al., 2011)[35]. Οι αλγόριθμοι SL χωρίς επίβλεψη χρησιμοποιούνται συχνά για την ομαδοποίηση δεδομένων με παρόμοια χαρακτηριστικά ανεξάρτητα από προκαθορισμένα κριτήρια ταξινόμησης και

υπερέχουν στον εντοπισμό χρήσιμων χαρακτηριστικών κάθε φορά που τα προς ανάλυση δεδομένα παρουσιάζουν μεγάλη διάσταση (Hardy, et al., 2016).

## 4.4 Παραγόμενη Τεχνητή Νοημοσύνη (GenAI)

Τα εργαλεία παραγόμενη Τεχνητής Νοημοσύνης (GenAI) είναι μια αναδυόμενη κατηγορία αλγορίθμων Τεχνητής Νοημοσύνης νέας εποχής, ικανοί να παράγουν νέο περιεχόμενο — σε ποικίλες μορφές όπως κείμενο, ήχο, βίντεο, εικόνες και κώδικα — με βάση τις προτροπές των χρηστών. Οι πρόσφατες εξελίξεις στη ΜΜ, τα τεράστια σύνολα δεδομένων και οι σημαντικές αυξήσεις στην υπολογιστική ισχύ έχουν ωθήσει τέτοια εργαλεία σε επιδόσεις σε ανθρώπινο επίπεδο σε ακαδημαϊκά και επαγγελματικά σημεία αναφοράς (benchmarks).

Αυτή η ταχεία πρόοδος οδήγησε πολλούς να πιστέψουν ότι η μεταμόρφωση αυτών των τεχνολογιών από επιδείξεις ερευνητικού επιπέδου σε προσιτά και εύχρηστα προϊόντα και υπηρεσίες παραγωγικής ποιότητας έχει τη δυνατότητα να επιβαρύνει τις επιχειρηματικές διαδικασίες και λειτουργίες, ενώ επιτρέπει εντελώς νέα παραδοτέα μέχρι τώρα. είναι ανέφικτο από οικονομικούς ή τεχνολογικούς παράγοντες. Το ChatGPT της OpenAI, μια διαδικτυακή εφαρμογή συνομιλίας που βασίζεται σε ένα παραγωγικό (πολυτροπικό) γλωσσικό μοντέλο, χρειάστηκε περίπου πέντε ημέρες για να φτάσει το ένα εκατομμύριο χρήστες. Από την πλευρά των επιχειρήσεων, ο Economist αναφέρει ότι ο αριθμός των θέσεων εργασίας που αναφέρουν δεξιότητες που σχετίζονται με την Τεχνητή Νοημοσύνη τετραπλασιάστηκε από το 2022 έως το 2023. Αυτός ο ενθουσιασμός δεν έχει περάσει απαρατήρητος από τους επενδυτές. Σύμφωνα με πληροφορίες, οι νεοφυείς επιχειρήσεις Τεχνητής Νοημοσύνης συγκέντρωσαν 600% περισσότερα κεφάλαια το 2022 από ό,τι το 2020 (Dhamani & Engler, 2023).

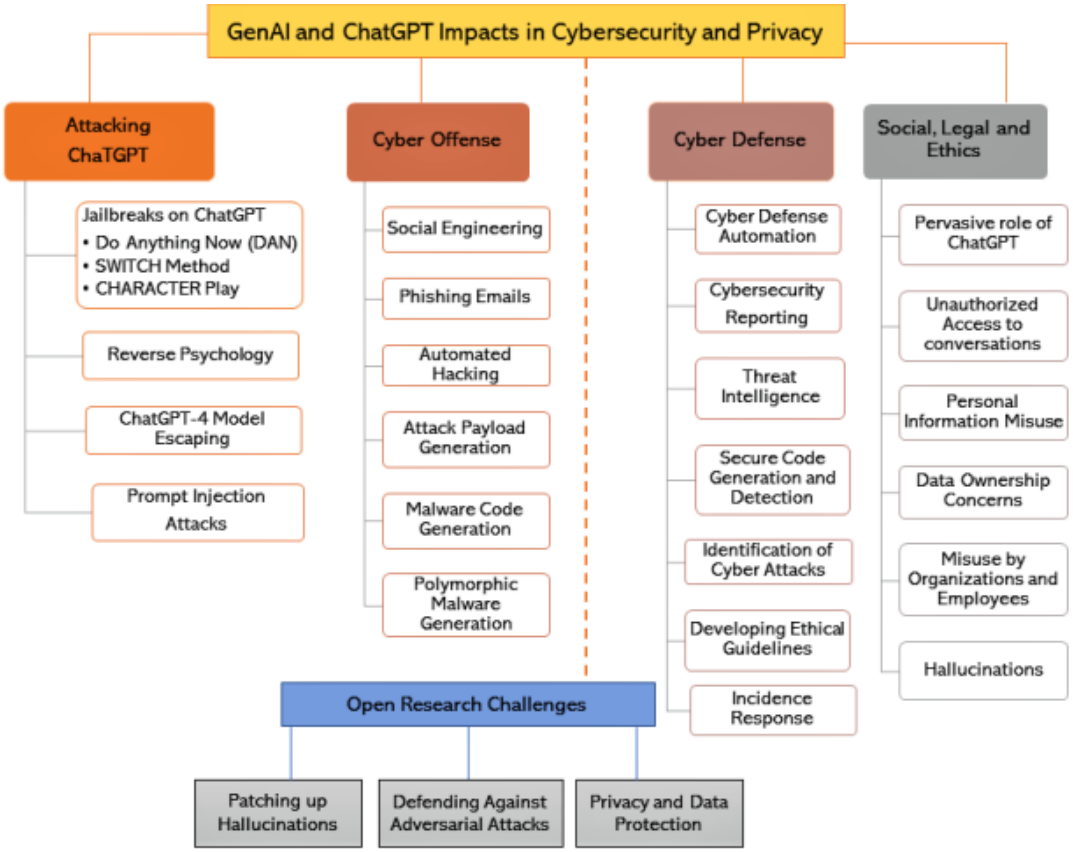



Εικόνα 12. Ταξινόμηση κλάδων που σχετίζονται με το GenAI. Πηγή: *(Singh, 2021)*

Η δύναμη γενίκευσης της Τεχνητής Νοημοσύνης ήταν επιτυχής στην αντικατάσταση των παραδοσιακών προσεγγίσεων που βασίζονται σε κανόνες με πιο έξυπνη τεχνολογία. Ωστόσο, το εξελισσόμενο ψηφιακό τοπίο δεν αναβαθμίζει μόνο την τεχνολογία, αλλά και αναβαθμίζει την πολυπλοκότητα των φορέων απειλής στον κυβερνοχώρο. Παραδοσιακά, ο κυβερνοχώρος αντιμετώπιζε σχετικά απλές προσπάθειες εισβολής, αλλά σε πολύ μεγάλο όγκο. Ωστόσο, η εισαγωγή επιθέσεων με τη βοήθεια Τεχνητής Νοημοσύνης από παραβάτες στον κυβερνοχώρο έχει ξεκινήσει μια εντελώς νέα εποχή, εξαπολύοντας γνωστούς και άγνωστους μετασχηματισμούς στους φορείς κυβερνοεπιθέσεων. Οι τεχνολογίες Τεχνητή Νοημοσύνη και Μηχανικής Μάθησης έχει αναβαθμίσει την αποτελεσματικότητα των επιθέσεων στον κυβερνοχώρο κάνοντας τους παραβάτες του κυβερνοχώρου πιο ισχυρούς από ποτέ. Προφανώς, με αρκετές πρόσφατες περιπτώσεις να γίνονται αντιληπτές, η GenAI έχει κερδίσει μεγάλο ενδιαφέρον από την κοινότητα της

κυβερνοασφάλειας, τόσο για την άμυνα στον κυβερνοχώρο όσο και για την επίθεση.

Τα εξελισσόμενα εργαλεία GenAI αποτελούν «δίκοπο μαχαίρι» στην ασφάλεια στον κυβερνοχώρο, ωφελώντας τόσο τους αμυνόμενους όσο και τους επιτιθέμενους. Τα εργαλεία GenAI όπως το ChatGPT μπορούν να χρησιμοποιηθούν από υπερασπιστές στον κυβερνοχώρο για να προστατεύσουν το σύστημα από κακόβουλους εισβολείς. Αυτά τα εργαλεία αξιοποιούν τις πληροφορίες από τα Large Language Models (LLM) που έχουν εκπαιδευτεί σχετικά με τον τεράστιο όγκο δεδομένων ευφυΐας απειλών στον κυβερνοχώρο που περιλαμβάνουν τρωτά σημεία, μοτίβα επιθέσεων και ενδείξεις επίθεσης. Οι υπερασπιστές του κυβερνοχώρου μπορούν να χρησιμοποιήσουν αυτό το μεγάλο άθροισμα πληροφοριών για να βελτιώσουν την ικανότητά τους για τη νοημοσύνη των απειλών εξάγοντας πληροφορίες και εντοπίζοντας αναδυόμενες απειλές. Τα εργαλεία GenAI μπορούν επίσης να χρησιμοποιηθούν για την ανάλυση του μεγάλου όγκου αρχείων καταγραφής, εξόδου συστήματος ή δεδομένων κίνησης δικτύου σε περίπτωση εμφάνισης στον κυβερνοχώρο. Αυτό επιτρέπει στους υπερασπιστές να επιταχύνουν και να αυτοματοποιήσουν τη διαδικασία απόκρισης περιστατικού. Τα μοντέλα που βασίζονται στο GenAI είναι επίσης χρήσιμα στη δημιουργία μιας ανθρώπινης συμπεριφοράς με επίγνωση της ασφάλειας, εκπαιδεύοντας τους ανθρώπους για αυξανόμενες εξελιγμένες επιθέσεις. Τα εργαλεία GenAI μπορούν επίσης να βοηθήσουν σε ασφαλείς πρακτικές κωδικοποίησης, τόσο με τη δημιουργία των ασφαλών κωδικών όσο και με την παραγωγή δοκιμών για την επιβεβαίωση της ασφάλειας του γραπτού κώδικα. Επιπλέον, τα μοντέλα LLM είναι επίσης χρήσιμα για την ανάπτυξη καλύτερων δεοντολογικών κατευθυντήριων γραμμών για την ενίσχυση της άμυνας στον κυβερνοχώρο σε ένα σύστημα (Gupta, et al., 2023).




Εικόνα 13. Roadmap της GenAI και του ChatGPT στην Κυβερνοασφάλεια και το Απόρρητο. Πηγή: *(Gupta, et al., 2023)*

Από την άλλη πλευρά, η χρήση της GenAI κατά της κυβερνοασφάλειας και οι κίνδυνοι κακής χρήσης του δεν μπορούν να υπονομευθούν. Οι παραβάτες του κυβερνοχώρου μπορούν να χρησιμοποιήσουν τη GenAI για να πραγματοποιήσουν επιθέσεις στον κυβερνοχώρο είτε εξάγοντας απευθείας τις πληροφορίες είτε παρακάμπτοντας τις ηθικές πολιτικές του OpenAI. Οι εισβολείς χρησιμοποιούν τη παραγόμενη δύναμη των εργαλείων GenAI για να δημιουργήσουν μια «πειστική» επίθεση κοινωνικής μηχανικής, επίθεση phishing και διάφορα είδη κακόβουλων αποσπασμάτων κώδικα που μπορούν να μεταγλωττιστούν σε ένα εκτελέσιμο αρχείο κακόβουλου λογισμικού. Αν και η ηθική πολιτική του OpenAI περιορίζει τα LLM, όπως το ChatGPT, να παρέχουν απευθείας κακόβουλες πληροφορίες στους επιτιθέμενους, υπάρχουν τρόποι να παρακαμφθούν οι περιορισμούς που επιβάλλονται σε αυτά τα μοντέλα χρησιμοποιώντας jailbreaking και άλλες τεχνικές, όπως συζητείται σε αυτήν την ενότητα. Επιπλέον, τα εργαλεία GenAI βοηθούν περαιτέρω τους

εισβολείς στον κυβερνοχώρο λόγω έλλειψης πλαισίου, άγνωστων προκαταλήψεων, τρωτών σημείων ασφαλείας και υπερβολικής εξάρτησης από αυτές τις μετασχηματιστικές τεχνολογίες (Επαυξημένη πραγματικότητα, Blockchain, Κβαντική Υπολογιστική κ.α.).

Τα σημαντικότερα οφέλη της παραγόμενης Τεχνητής Νοημοσύνης στην ασφάλεια του κυβερνοχώρου φαίνονται στον παρακάτω πίνακα:

| **Οφέλη** | **Περιγραφή** |
|---|---|
| **Παραγωγή Συνθετικών Δεδομένων** | Η Generative AI μπορεί να χρησιμοποιηθεί για τη δημιουργία συνθετικών συνόλων δεδομένων που προσομοιώνουν την κυκλοφορία δικτύου ή τη συμπεριφορά των χρηστών, χωρίς να διακυβεύονται τα πραγματικά δεδομένα. Αυτά τα δεδομένα μπορούν να χρησιμοποιηθούν για την εκπαίδευση συστημάτων ανίχνευσης εισβολών, χωρίς να παραβιάζεται το απόρρητο των χρηστών. |
| **Προσομοίωση Επιθέσεων** | Μέσω των παραγωγικών δικτύων αντιπάλου (Generative Adversarial Networks - GANs) , είναι δυνατό να προσομοιωθεί ο τρόπος με τον οποίο θα ενεργούσε ένας εισβολέας, επιτρέποντας στους οργανισμούς να δοκιμάσουν την ευρωστία των συστημάτων τους και να κάνουν βελτιώσεις πριν συμβούν πραγματικά περιστατικά. |
| **Δημιουργία Σεναρίων Δοκιμών** | Η Generative AI μπορεί να βοηθήσει στη δημιουργία ρεαλιστικών σεναρίων δοκιμών διείσδυσης, βελτιώνοντας τις παραδοσιακές πρακτικές που συχνά βασίζονται σε προκαθορισμένα και λιγότερο δυναμικά σενάρια. |
| **Ενίσχυση Μάθησης** | Η Generative AI, ειδικά τα GAN, μπορεί να είναι χρήσιμα στην ενισχυτική μάθηση, όπου ένας πράκτορας και ένας αντίπαλος συνεργάζονται. Αυτή η τεχνική μπορεί να χρησιμοποιηθεί για να διδάξει τα συστήματα κυβερνοασφάλειας πώς να βελτιώσουν τον εντοπισμό και |

## 4.5 Προκλήσεις και περιορισμοί της Τεχνητής Νοημοσύνης στην κυβερνοασφάλεια

Η Τεχνητή Νοημοσύνη επηρεάζει άμεσα τον τομέα της κυβερνοασφάλειας, προσφέροντας καινοτόμες λύσεις για τον εντοπισμό και την πρόληψη απειλών, την ανάλυση συμπεριφοράς και την αυτοματοποιημένη απόκριση συμβάντων. Ωστόσο, όπως κάθε αναδυόμενη τεχνολογία, η Τεχνητή Νοημοσύνη δεν είναι χωρίς προκλήσεις και περιορισμούς. Παρά τις μετασχηματιστικές της δυνατότητες, οι προσδοκίες για την Τεχνητή Νοημοσύνη πρέπει να εξισορροπούνται με μια σαφή κατανόηση των περιορισμών της.

Αυτές οι προκλήσεις δεν περιλαμβάνουν μόνο τεχνικές πτυχές, όπως η ποιότητα της εκπαίδευσης δεδομένων ή η ερμηνεία των αποτελεσμάτων, αλλά και ηθικά διλήμματα και ανησυχίες σχετικά με το απόρρητο. Επιπλέον, καθώς οι κακόβουλοι χρήστες του κυβερνοχώρου προσαρμόζονται και εξελίσσονται, αναδύονται νέα εμπόδια στα συστήματα που βασίζονται σε Τεχνητή Νοημοσύνη, από αντίπαλες επιθέσεις μέχρι χειραγώγηση μοντέλων.

Σε αυτή την ενότητα, θα διερευνήσουμε λεπτομερώς τις προκλήσεις που ενυπάρχουν στη χρήση της Τεχνητής Νοημοσύνης στην ασφάλεια στον κυβερνοχώρο, τους τρέχοντες περιορισμούς αυτής της τεχνολογίας και τους τομείς όπου, παρά την πρόοδο, η ανθρώπινη παρέμβαση και η κρίση παραμένουν αναντικατάστατες. Με αυτόν τον τρόπο, επιδιώκουμε να παρέχουμε μια ισορροπημένη και ρεαλιστική προοπτική που επιτρέπει στους οργανισμούς να μεγιστοποιούν τα οφέλη της Τεχνητής Νοημοσύνης, παραμένοντας σε εγρήγορση για τους πιθανούς περιορισμούς της.

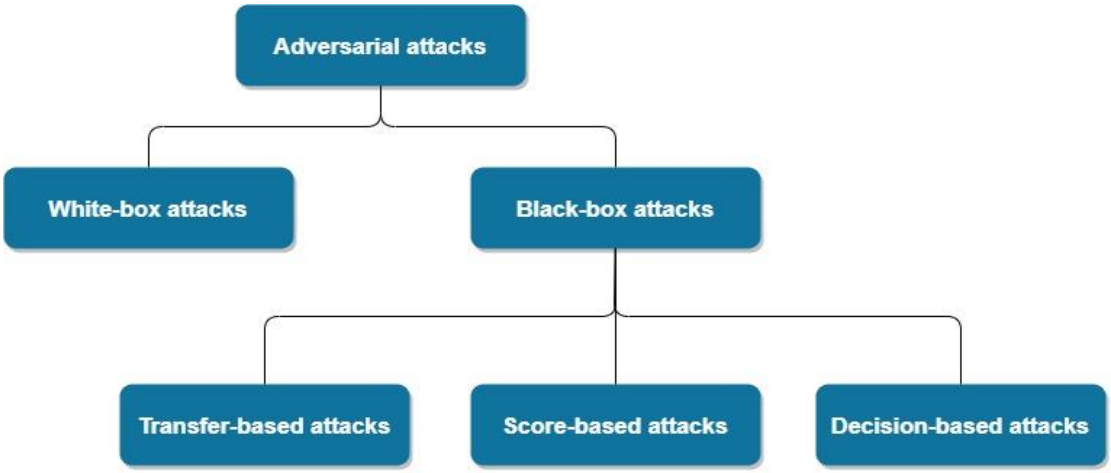
### 4.5.1 Αντιπαραθετικές Επιθέσεις εναντίον Μοντέλων Τεχνητής Νοημοσύνης

Οι αντιπαραθετικές επιθέσεις κατά μοντέλων Τεχνητής Νοημοσύνης έχουν αναδειχθεί ως κρίσιμη ανησυχία στον τομέα της κυβερνοασφάλειας. Όπως τονίσαμε προηγουμένως, αυτές οι επιθέσεις έχουν σχεδιαστεί για να

εξαπατήσουν ή να «μπερδέψουν» τα μοντέλα Μηχανικής Μάθησης, τα οποία θα μπορούσαν να οδηγήσουν σε λανθασμένες ή κακόβουλες αποφάσεις από τέτοια συστήματα. Ουσιαστικά, μια αντιπαραθετική επίθεση περιλαμβάνει την εισαγωγή μικρών διαταραχών στα δεδομένα εισόδου, σχεδιασμένων να είναι σχεδόν ανεπαίσθητες από τον άνθρωπο, αλλά που μπορούν να οδηγήσουν το μοντέλο σε λανθασμένες προβλέψεις. Αυτές οι διαταραχές υπολογίζονται προσεκτικά για να μεγιστοποιηθεί το σφάλμα πρόβλεψης του μοντέλου.

Οι αντιπαραθετικές επιθέσεις μπορεί να είναι δύο τύπων:

| | |
|---|---|
| **Επιθέσεις Λευκού Κουτιού** | Σε αυτό το σενάριο, ο εισβολέας έχει πλήρη γνώση του μοντέλου, συμπεριλαμβανομένης της αρχιτεκτονικής και των παραμέτρων του. Αυτό του επιτρέπει να σχεδιάζει διαταραχές που είναι ιδιαίτερα αποτελεσματικές έναντι του συγκεκριμένου μοντέλου. |
| **Επιθέσεις Μαύρου Κουτιού** | Σε αυτήν την περίπτωση, ο εισβολέας δεν έχει άμεση πρόσβαση στο μοντέλο και τις παραμέτρους του, αλλά μπορεί να έχει πρόσβαση στις προβλέψεις του. Αν και αυτό το σενάριο είναι πιο προκλητικό για τον εισβολέα, είναι ακόμα δυνατό να δημιουργηθούν αποτελεσματικές αντιπαραθετικές διαταραχές (Bezirganyan & Sergoyan, 2022). |




Εικόνα 14.Τύποι Αντιπαραθετικών Επιθέσεων. Πηγή: (Bezirganyan & Sergoyan, 2022)

Για παράδειγμα, στην περίπτωση εντοπισμού κακόβουλου λογισμικού, εάν ένα σύστημα Τεχνητής Νοημοσύνης χρησιμοποιείται για τον εντοπισμό κακόβουλου λογισμικού, ένας εισβολέας θα μπορούσε να σχεδιάσει κακόβουλο λογισμικό που, μόλις τροποποιηθεί, δεν ανιχνεύεται από το μοντέλο. Στην περίπτωση συστημάτων ελέγχου ταυτότητας, εάν ένα σύστημα που βασίζεται στη Τεχνητή Νοημοσύνη χειρίζεται έλεγχο ταυτότητας, π.χ. μέσω της αναγνώρισης προσώπου, μια επίθεση αντιπάλου θα μπορούσε να επιτρέψει μη εξουσιοδοτημένη πρόσβαση σε έναν εισβολέα. Τέλος, στην περίπτωση ανάλυσης της κυκλοφορίας δικτύου, οι εισβολείς μπορούν να χειραγωγήσουν συγκεκριμένα χαρακτηριστικά της κίνησης του δικτύου για να αποφύγουν τον εντοπισμό από ένα σύστημα που βασίζεται στη Τεχνητή Νοημοσύνη.

Ως απάντηση σε αυτό, μπορούν να αναπτυχθούν διάφορα αντίμετρα, μεταξύ των οποίων:

✓ **Adversarial Training**: Αυτή η τεχνική περιλαμβάνει εκπαίδευση του μοντέλου με αντίθετα παραδείγματα, τα οποία μπορούν να αυξήσουν την ευρωστία του έναντι τέτοιων επιθέσεων.

✓ **Disturbance Detection:** Ορισμένες μέθοδοι στοχεύουν στον άμεσα εντοπισμό των αντίθετων διαταραχών αντί να επιχειρούν να κάνουν ακριβείς προβλέψεις παρουσία τους.

✓ **Regularisation and Defence Techniques:** Αυτές είναι τεχνικές που έχουν σχεδιαστεί για να κάνουν τα μοντέλα εγγενώς πιο ανθεκτικά σε αντίπαλες επιθέσεις προσαρμόζοντας τη συμπεριφορά τους κατά τη διάρκεια της εκπαίδευσης.

Οι αντιπαραθετικές επιθέσεις εναντίον μοντέλων Τεχνητής Νοημοσύνης είναι μια εκδήλωση μιας θεμελιώδης αλήθειας (fundamental truth) στην ασφάλεια στον κυβερνοχώρο: οποιοδήποτε σύστημα, όσο προηγμένο κι αν είναι, έχει τρωτά σημεία. Ο στόχος θα ήταν να παραμείνουμε ένα βήμα μπροστά από τους επιτιθέμενους, να προσαρμόζονται συνεχώς και να εξελίσσονται ως απάντηση σε νέες απειλές. Τόσο οι εισβολείς όσο και οι αμυνόμενοι χρησιμοποιούν προηγμένα εργαλεία και πολλά από αυτά τα εργαλεία ενσωματώνουν δυνατότητες Τεχνητής Νοημοσύνη. Μερικά από τα πιο δημοφιλή, τόσο για επίθεση όσο και για άμυνα, φαίνονται παρακάτω.

| | |
|---|---|
| **OFFENSIVEt Εργαλεία** | **DeepExploit:t** είναιt έναt αυτοματοποιημένοt εργαλείοt pentestingt πουt χρησιμοποιείt βαθιάt εκμάθηση.t Είναιt σετ θέσηt ναt μάθειt απότ ταt αποτελέσματαt προηγούμενωνt δοκιμώνt διείσδυσηςt καιt ναt προσαρμόσειt τιςt τεχνικέςt τουt ανάλογα. **https://github.com/13o-bbr-bbq/machine_learning_security/tree/master/DeepExploit** |
| | Snallygaster: έναt εργαλείοt πουt αναζητάt εκτεθειμέναt αρχείαt σεt διακομιστέςt ιστού,t χρησιμοποιώνταςt τεχνικέςt Τεχνικήςt Νοημοσύνηςt γιαt τονt εντοπισμόt πιθανώνt φορέωνt επίθεσης. **https://github.com/hannob/snallygaster** |
| | GPT-2: αν και δεν σχεδιάστηκε αρχικά ως εργαλείο επίθεσης, αυτή η τεχνολογία φυσικής γλώσσας που αναπτύχθηκε από το OpenAI μπορεί να χρησιμοποιηθεί για τη δημιουργία πλαστού περιεχομένου, όπως μηνύματα ηλεκτρονικού ψαρέματος. |

| | |
|---|---|
| | **https://github.com/openai/gpt-2** |
| **DEFENSIVE Εργαλεία** | **TensorFlow Privacy**: Βιβλιοθήκη που βοηθά τους προγραμματιστές να εκπαιδεύουν μοντέλα μηχανικής εκμάθησης με διαφορικό απόρρητο, το οποίο μπορεί να βοηθήσει στην προστασία των δεδομένων εκπαίδευσης. **https://github.com/tensorflow/privacy** |
| | Adversarial Robustness Toolbox (ART) της IBM: Βιβλιοθήκη Python που παρέχει εργαλεία για τη βελτίωση της ευρωστίας των μοντέλων Μηχανικής Μάθησης και την εμβάθυνσή τους έναντι των αντίθετων επιθέσεων. **(https://github.com/Trusted-AI/adversarial-robustness-toolbox)** |
| | **DeepArmor:** είναι μια λύση κυβερνοασφάλειας που χρησιμοποιεί τεχνικές βαθιάς εκμάθησης για τον εντοπισμό και την πρόληψη κακόβουλου λογισμικού σε πραγματικό χρόνο. https://www.sparkcognition.com/deeparmor-endpoint-security/ |
| | **CylancePROTECT:** Λογισμικό τελικού σημείου (endpoint) που χρησιμοποιεί μοντέλα Τεχνητής Νοημοσύνης για την πρόβλεψη και την πρόληψη κακόβουλου λογισμικού και προηγμένων σεναρίων. https://www.cylance.com/cylanceprotect |

Πίνακας 3. Εργαλεία Αντιπαραθετικών Επιθέσεων. Πηγή: (Cordero & Pascual, 2023)

## Κεφάλαιο 5. Τεχνητή Νοημοσύνη και Μηχανική Μάθηση ως εργαλεία κυβερνοασφάλειας

Όπως είναι γνωστό, η ασφάλεια στον κυβερνοχώρο είναι ένας ατελείωτος αγώνας μεταξύ επιτιθέμενων και υπερασπιστών. Ενώ οι επιτιθέμενοι αναζητούν νέα τρωτά σημεία και τρόπους για να παραβιάσουν συστήματα, οι υπερασπιστές επιδιώκουν να προβλέψουν (πρόληψη), να εντοπίσουν (ανίχνευση) και να ανταποκριθούν σε αυτές τις επιθέσεις (απόκριση). Η Τεχνητή Νοημοσύνη, με την ικανότητά της να επεξεργάζεται μεγάλες ποσότητες δεδομένων με εξαιρετική ταχύτητα και να μαθαίνει από αυτά, προσφέρει σημαντικές λύσεις σε (τρέχουσες και αναδυόμενες) προκλήσεις στον κυβερνοχώρο. Παραδοσιακά, μερικές από τις βασικές εφαρμογές της Τεχνητής Νοημοσύνης στην ασφάλεια στον κυβερνοχώρο φαίνονται στον παρακάτω πίνακα (Cordero & Pascual, 2023):

| Μηχανισμός | Περιγραφή |
|---|---|
| **Ανίχνευση και Αντιμετώπιση Απειλών** | Τα συστήματα που βασίζονται σε Τεχνητή Νοημοσύνη μπορούν να αναλύσουν μοτίβα στην κίνηση δικτύου ή στη συμπεριφορά των χρηστών για να εντοπίσουν ανωμαλίες ή ύποπτη δραστηριότητα. Μόλις εντοπιστεί, η Τεχνητή Νοημοσύνη μπορεί να δράσει γρήγορα, συχνά πιο γρήγορα από μια ανθρώπινη ομάδα, για να μετριάσει ή να εξουδετερώσει την απειλή. |
| **Προγνωστική Ανάλυση** | Η Τεχνητή Νοημοσύνη μπορεί να χρησιμοποιήσει ιστορικά δεδομένα για να προβλέψει μελλοντικές απειλές ή τρωτά σημεία, επιτρέποντας στους οργανισμούς να προετοιμαστούν προληπτικά και να προστατευτούν. |
| **Έλεγχος και Διαχείριση Ταυτότητας** | Η Τεχνητή Νοημοσύνη μπορεί να χρησιμοποιήσει προηγμένα βιομετρικά στοιχεία, τη συμπεριφορά των χρηστών και άλλους παράγοντες για τον έλεγχο |

| | |
|---|---|
| | ταυτότητας ατόμων με υψηλή ακρίβεια, μειώνοντας τον κίνδυνο μη εξουσιοδοτημένης πρόσβασης. |
| **Προστασία από το Ηλεκτρονικό Ψάρεμα (phishing)** | Αναλύοντας το περιεχόμενο, τις εικόνες και τα μοτίβα κειμένων ή εγγράφων (π.χ. μηνύματα ηλεκτρονικού ταχυδρομείου), η Τεχνητή Νοημοσύνη μπορεί να εντοπίσει απόπειρες phishing με υψηλή ακρίβεια, προστατεύοντας τους χρήστες από πιθανές απάτες. |
| **Βελτιστοποίηση Ρυθμίσεων Ασφαλείας** | Η Τεχνητή Νοημοσύνη μπορεί να αξιολογήσει διαμορφώσεις και πολιτικές ασφαλείας για να εντοπίσει πιθανές αδυναμίες και να προτείνει βελτιώσεις |

Πίνακας 4. Βασικά Εργαλεία της τεχνητής νοημοσύνης στην ασφάλεια στον κυβερνοχώρο. Πηγή: *(Cordero & Pascual, 2023)*

## 5.1 Ανίχνευση Απειλών και Ανάλυση Συμπεριφοράς

Η ανίχνευση απειλών και η ανάλυση συμπεριφοράς είναι απαραίτητες για τον εντοπισμό και την απόκριση σε κυβερνοεπιθέσεων σε πραγματικό χρόνο. Με την εφαρμογή της τεχνητής νοημοσύνης σε αυτούς τους τομείς, η ασφάλεια στον κυβερνοχώρο έχει σημειώσει σημαντική βελτίωση στην αποτελεσματικότητα και την ακρίβεια ανίχνευσης. Ο όγκος των δεδομένων που επεξεργάζονται οι οργανισμοί (δημόσιοι ή ιδιωτικοί) σε καθημερινή βάση είναι τεράστιος. Η μη αυτόματη ανίχνευση απειλών σε τέτοιους τόμους είναι σχεδόν αδύνατη. Οι σύγχρονες επιθέσεις στον κυβερνοχώρο χρησιμοποιούν συχνά μυστικές τακτικές, όπως οι επιθέσεις διείσδυσης (Lateral Movement Attacks), γεγονός που καθιστά δύσκολο τον εντοπισμό τους με παραδοσιακές μεθόδους.

Έτσι, αντί να βασίζεται μόνο σε γνωστές υπογραφές κακόβουλου λογισμικού, η Τεχνητή Νοημοσύνη εστιάζει σε μοτίβα ανώμαλης συμπεριφοράς. Αυτό καθιστά δυνατό τον εντοπισμό προηγουμένως άγνωστων απειλών ή παραλλαγών κακόβουλου λογισμικού που έχουν ελαφρώς τροποποιηθεί. Αναλύοντας τη συμπεριφορά των χρηστών και του συστήματος,

η Τεχνητή Νοημοσύνη μπορεί να εντοπίσει ασυνήθιστη δραστηριότητα, όπως η πρόσβαση σε αρχεία σε περίεργες ώρες ή η ασυνήθιστη μεταφορά μεγάλων ποσοτήτων δεδομένων. Η ανίχνευση απειλών συμπεριφοράς έχει γνωρίσει ταχεία αύξηση στη δημοτικότητα και την υιοθέτηση, και έχει εμφανιστεί μια σειρά από εργαλεία και συστήματα, τόσο εμπορικά όσο και ανοιχτού κώδικα, που ειδικεύονται σε αυτήν την προσέγγιση. Μερικά από τα πιο δημοφιλή εργαλεία παρατίθενται παρακάτω:

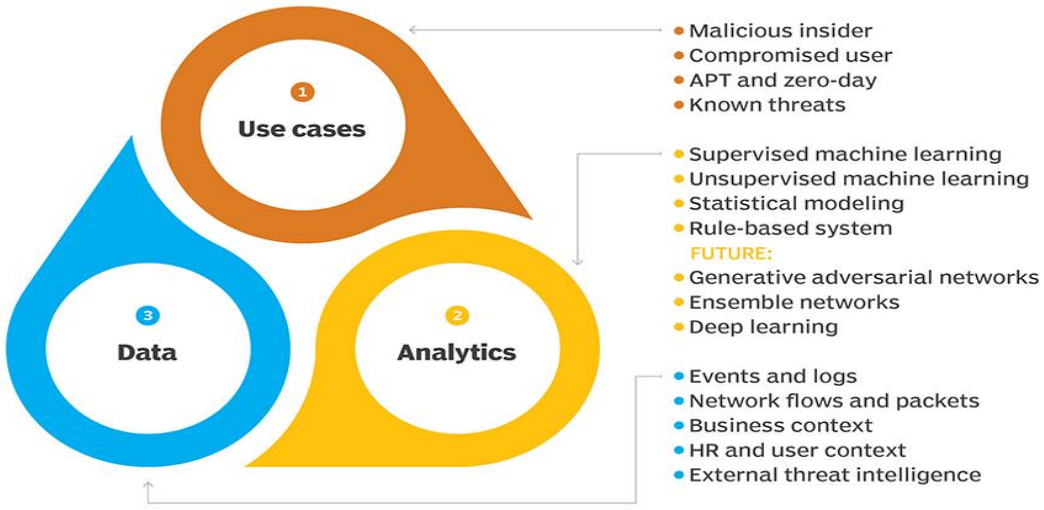
- ✓ **Darktrace:** Το Darktrace χρησιμοποιεί Μηχανική Μάθηση και αλγόριθμους Τεχνητής Νοημοσύνης για τον εντοπισμό, την απόκριση και τον μετριασμό των απειλών σε πραγματικό χρόνο με βάση μοτίβα ανώμαλης συμπεριφοράς. Το εργαλείο είναι γνωστό για το "Enterprise Immune System", το οποίο μαθαίνει και καθιερώνει αυτό που μπορεί να γίνει κατανοητό ως μια κατάσταση "business as usual" στο δίκτυο και στη συνέχεια εντοπίζει αποκλίσεις από αυτόν τον κανόνα.

- ✓ **Vectra**: Το Vectra προσφέρει ανίχνευση απειλών σε πραγματικό χρόνο χρησιμοποιώντας τεχνικές Τεχνητής Νοημοσύνης. Επικεντρώνεται στον εντοπισμό κακόβουλης συμπεριφοράς εντός της κυκλοφορίας του δικτύου και παρέχει μια λεπτομερή εικόνα της συνεχιζόμενης αλυσίδας επιθέσεων, επιτρέποντας στις ομάδες ασφαλείας να ανταποκρίνονται γρήγορα.

- ✓ **CrowdStrike Falcon**: Το CrowdStrike είναι γνωστό για τις λύσεις προστασίας τελικού σημείου (endpoint protection). Η πλατφόρμα Falcon χρησιμοποιεί τεχνικές που βασίζονται στη συμπεριφορά για να ανιχνεύσει και να αποτρέψει απειλές που ενδέχεται να παραλείψουν άλλα συστήματα που βασίζονται σε υπογραφές.

- ✓ **Cylance**: Το CylancePROTECT είναι μια λύση προστασίας τελικού σημείου που χρησιμοποιεί μοντέλα τεχνητής νοημοσύνης για τον εντοπισμό και τον αποκλεισμό κακόβουλου λογισμικού με βάση τα χαρακτηριστικά και τις συμπεριφορές του και όχι με γνωστές υπογραφές.

✓ **Gurucul:** Παρέχει λύσεις ανάλυσης συμπεριφοράς χρηστών και οντοτήτων (User and Entity Behavioural Analytics (UEBA)) που χρησιμοποιούν αλγόριθμους Μηχανικής Μάθησης για τον εντοπισμό εσωτερικών απειλών, απάτης και μη εξουσιοδοτημένης πρόσβασης. Οι τεχνολογίες UBA αναλύουν αρχεία καταγραφής ιστορικών δεδομένων -- συμπεριλαμβανομένων των αρχείων καταγραφής δικτύου και ελέγχου ταυτότητας που συλλέγονται και αποθηκεύονται σε συστήματα διαχείρισης αρχείων καταγραφής και πληροφοριών ασφαλείας και διαχείρισης συμβάντων (Security Information and Event Management - SIEM) -- για τον εντοπισμό μοτίβων επισκεψιμότητας που προκαλούνται από τη συμπεριφορά των χρηστών, κανονικών και κακόβουλων. Τα συστήματα UBA και UEBA προορίζονται κυρίως να παρέχουν στις ομάδες κυβερνοασφάλειας χρήσιμες πληροφορίες όταν τα συστήματα εντοπίζουν ασυνήθιστη συμπεριφορά. Στην παρακάτω εικόνα φαίνονται τα 3 βασικά δομικά στοιχεία των UEBA, τα οποία εξαρτώνται από περιπτώσεις χρήσης, αναλύσεις και δεδομένα για την ανάθεση βαθμολογιών κινδύνου σε συγκεκριμένες συμπεριφορές (Loshin, 2022).




Εικόνα 15. Οι 3 πυλώνες της UEBA. Πηγή: *(Loshin, 2022)*

✓ **Wazuh:** Πρόκειται για μια πλατφόρμα ανοιχτού κώδικα για ανίχνευση απειλών, διαχείριση ευπάθειας και παρακολούθηση ακεραιότητας. Χρησιμοποιεί κανόνες και αποκωδικοποιητές για να αναλύει συμβάντα ασφαλείας και να ανιχνεύει ανώμαλη συμπεριφορά.

✓ **Snort:** Αν και είναι περισσότερο γνωστό ως σύστημα ανίχνευσης και πρόληψης εισβολής (Intrusion Detection and Prevention System - IDPS), το Snort έχει εξελιχθεί για να ενσωματώνει ικανότητες που βασίζονται στη συμπεριφορά. Η κοινότητα Snort αναπτύσσει και μοιράζεται νέους κανόνες που μπορούν να ανιχνεύσουν ανώμαλη συμπεριφορά.

✓ **Stack ELK (Elasticsearch, Logstash, Kibana):** Αν και το ίδιο το ELK δεν είναι εργαλείο ανίχνευσης συμπεριφοράς, μπορεί να διαμορφωθεί με συγκεκριμένα πρόσθετα και κανόνες για την εκτέλεση ανάλυσης συμπεριφοράς αρχείων καταγραφής και συμβάντων.

Τα συστήματα Τεχνητής Νοημοσύνης που λειτουργούν στο πλαίσιο του μοντέλου Machine Learning for Behavioral Analysis εκπαιδεύονται χρησιμοποιώντας μεγάλα σύνολα δεδομένων τόσο νόμιμης όσο και κακόβουλης συμπεριφοράς. Μέσω της εποπτευόμενης μάθησης, η Τεχνητή Νοημοσύνη μπορεί να μάθει να ταξινομεί και να ανιχνεύει κάθε ασυνήθιστη δραστηριότητα. Έτσι, με την πάροδο του χρόνου και καθώς επεξεργάζονται περισσότερα δεδομένα, αυτά τα συστήματα μπορούν να βελτιώσουν την ακρίβειά τους μέσω της μάθησης χωρίς επίβλεψη και της ενισχυτικής μάθησης. Πολλά σύγχρονα εργαλεία κυβερνοασφάλειας έχουν ενσωματώσει τη Μηχανική Μάθηση στις δυνατότητές τους για τη βελτίωση του εντοπισμού και της απόκρισης απειλών. Αυτά τα εργαλεία χρησιμοποιούν τεχνικές Μηχανικής Μάθησης για να «εκπαιδευτούν» και να προσαρμοστούν σε νέες απειλές μελετώντας πρότυπα και συμπεριφορές στα δεδομένα. Εκτός από τα εργαλεία που αναφέρονται παραπάνω, μερικές από τις πιο δημοφιλείς λύσεις παρατίθενται παρακάτω:

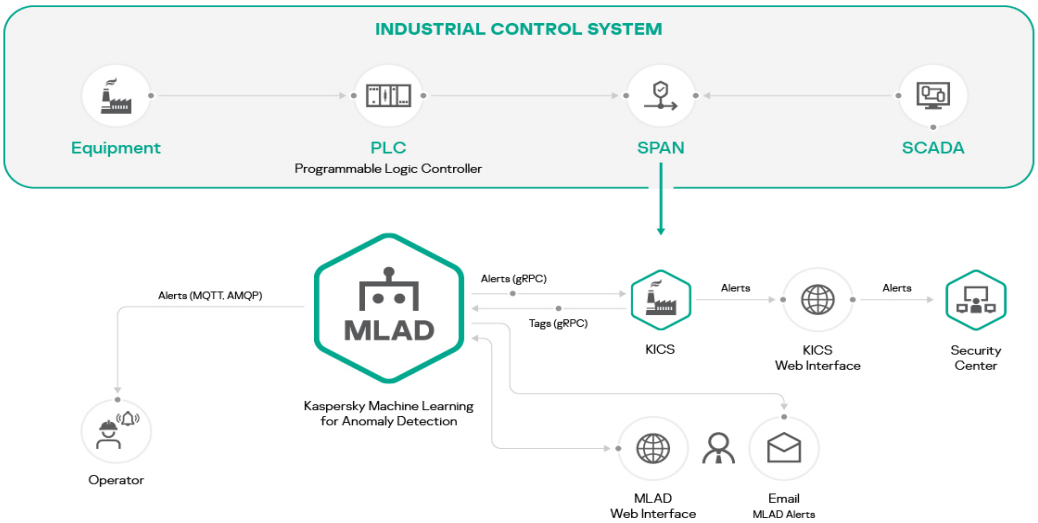
✓ **Endgame:** Αυτή η πλατφόρμα χρησιμοποιεί τεχνικές Μηχανικής Μάθησης για προστασία τελικού σημείου, ανίχνευση απειλών και απόκριση. Η

ικανότητά της Μηχανικής Μάθησης εστιάζει στον εντοπισμό τεχνικών και τακτικών επίθεσης χωρίς να βασίζεται αποκλειστικά σε υπογραφές.

✓ **PatternEx:** είναι μια λύση ανάλυσης συμπεριφοράς χρήστη και οντοτήτων (UEBA) που χρησιμοποιεί τεχνικές Μηχανικής Μάθησης. Αναλύει μεγάλους όγκους δεδομένων για να εντοπίσει μοτίβα που υποδηλώνουν κακόβουλη δραστηριότητα.

✓ **SentinelOne:** Αποτελεί μια λύση προστασίας τελικού σημείου (endpoint protection) που χρησιμοποιεί τεχνικές Μηχανικής Μάθησης για τον εντοπισμό, την ταξινόμηση και την απόκριση σε κακόβουλη και ασυνήθιστη συμπεριφορά.

✓ **Kaspersky Machine Learning for Anomaly Detection (MLAD):** σχεδιασμένο για βιομηχανικά συστήματα, το MLAD της Kaspersky χρησιμοποιεί τεχνικές Μηχανικής Μάθησης για τον εντοπισμό αποκλίσεων στη λειτουργία βιομηχανικών μηχανών (Grzemba, 2019).




Εικόνα 16. Kaspersky Machine Learning for Anomaly Detection. Πηγή: *(Grzemba, 2019)*

✓ **Splunk**: Ενώ το Splunk είναι κυρίως ένα εργαλείο ανάλυσης δεδομένων και SIEM, έχει δυνατότητες που επιτρέπουν στους χρήστες να εφαρμόζουν μοντέλα μηχανικής εκμάθησης για τον εντοπισμό μοτίβων και

ανωμαλιών σε μεγάλους όγκους δεδομένων.

Τα νευρωνικά δίκτυα, ειδικά τα νευρωνικά δίκτυα βαθιάς μάθησης, έχουν αποδειχθεί αποτελεσματικά στην ανίχνευση προτύπων σε μεγάλα σύνολα δεδομένων και μπορούν να χρησιμοποιηθούν για τον εντοπισμό κακόβουλου λογισμικού σε αρχεία με βάση τα χαρακτηριστικά τους, τον εντοπισμό επιθέσεων DDoS με βάση μοτίβα κυκλοφορίας ή τον εντοπισμό προσπαθειών phishing μέσω κειμένου και περιεχομένου ανάλυση. Εκτός από τις εμπορικές εταιρείες, οι οποίες επίσης εργάζονται σε νευρωνικά δίκτυα, μερικά από τα πιο γνωστά εργαλεία που χρησιμοποιούν αυτές τις τεχνικές παρατίθενται παρακάτω:

- ✓ **Deep Instinct:** αυτό το εργαλείο χρησιμοποιεί νευρωνικά δίκτυα βαθιάς εκμάθησης για την πρόληψη και τον εντοπισμό κακόβουλου λογισμικού σε πραγματικό χρόνο, προσφέροντας λύσεις τόσο για τερματικά σημεία όσο και για κινητές συσκευές.
- ✓ **SparkCognition:** αυτό το εργαλείο προσφέρει τον μηχανισμό DeepArmor, μια λύση που χρησιμοποιεί νευρωνικά δίκτυα για την παροχή προστασίας από απειλές σε πραγματικό χρόνο.

- ✓ **NVIDIA:** Αν και δεν είναι ένα εργαλείο ασφαλείας από μόνο του, το εργαλείο NVIDIA προσφέρει πλατφόρμες και βιβλιοθήκες όπως το CUDA και το cuDNN που επιταχύνουν τον υπολογισμό των νευρωνικών δικτύων.

Ωστόσο, υπάρχουν πολλά εργαλεία που μπόρεσαν να ενσωματώσουν μηχανισμούς Τεχνητής Νοημοσύνης στις τεχνολογίες τους, βασισμένοι σε πιο παραδοσιακές έννοιες, προκειμένου να τους κάνουν πιο αποτελεσματικούς.

### 5.1.1 Συστήματα Ανίχνευσης και Πρόληψης Εισβολών (IDPS)

Χρησιμοποιώντας τεχνικές Τεχνητής Νοημοσύνης, τα συστήματα IDPS μπορούν να ανιχνεύουν και να αποκλείσουν κακόβουλη κυκλοφορία σε πραγματικό χρόνο με μεγαλύτερη ακρίβεια. Ένα σύστημα ανίχνευσης και

πρόληψης εισβολής (IDPS) είναι απαραίτητο για τον εντοπισμό και την απόκριση σε κακόβουλη δραστηριότητα σε ένα δίκτυο ή σύστημα. Η ενσωμάτωση της Τεχνητής Νοημοσύνης σε αυτά τα συστήματα έχει βελτιώσει σημαντικά την ικανότητά τους να εντοπίζουν και να αντιδρούν σε απειλές σε πραγματικό χρόνο.

Μερικά εργαλεία είναι τα εξής:

- ✓ **Darktrace** (https://www.darktrace.com): Όπως αναφέρθηκε παραπάνω, το Darktrace είναι γνωστό για την προσέγγισή του που βασίζεται στην Τεχνητή Νοημοσύνη στον εντοπισμό και την πρόληψη απειλών. Η τεχνολογία του Enterprise Immune System χρησιμοποιεί Μηχανική Μάθηση για να ανιχνεύσει ασυνήθιστη συμπεριφορά σε πραγματικό χρόνο.

- ✓ **Vectra** (https://www.vectra.ai): Το Vectra Cognito χρησιμοποιεί Τεχνητή Νοημοσύνη για να εντοπίζει αυτόματα και να δίνει προτεραιότητα σε ασυνήθιστη συμπεριφορά σε πραγματικό χρόνο για να ανακαλύψει ενεργές επιθέσεις και εσωτερικές απειλές.
- ✓ **Cisco Stealthwatch** (https://www.cisco.com): αν και δεν είναι IDPS με την παραδοσιακή έννοια, το Stealthwatch χρησιμοποιεί μηχανική εκμάθηση για να ανιχνεύσει ασυνήθιστη συμπεριφορά στο δίκτυο και ενσωματώνεται με άλλες λύσεις της Cisco για να παρέχει δυνατότητες πρόληψης.


Εικόνα 17. Διάγραμμα υψηλού επιπέδου της αρχιτεκτονικής StealthWatch. Πηγή: *(McNamara, 2016)*

✓ **Lastline** (https://www.lastline.com): προσφέρει λύσεις που χρησιμοποιούν τεχνικές Τεχνητής Νοημοσύνης, όπως η Μηχανική Μάθηση, για τον εντοπισμό και την απόκριση σε προηγμένες, αποφυγές και απειλές zero-day[1].

✓ **Awake Security** (https://www.awakesecurity.com): η πλατφόρμα της χρησιμοποιεί Τεχνητή Νοημοσύνη για να αναλύει την κυκλοφορία του δικτύου και να ανιχνεύει απειλές. Μπορεί να εντοπίσει κακόβουλη και επικίνδυνη συμπεριφορά χωρίς να βασίζεται σε υπογραφές ή προηγούμενη γνώση.

✓ **Fortinet** (https://www.fortinet.com): Ενώ το Fortinet προσφέρει μια ποικιλία λύσεων ασφαλείας, το FortiGate με ενσωματωμένη λειτουργία IDPS έχει επίσης συμπεριλάβει Τεχνητή Νοημοσύνη για τη βελτίωση της ανίχνευσης απειλών.

### 5.1.2 Εργαλεία Εγκληματολογικής Ανάλυσης (Forensic Analysis)

[1] Μια απειλή ή επίθεση μηδενικής ημέρας είναι μια άγνωστη ευπάθεια στο λογισμικό ή το υλικό του υπολογιστή ή της κινητής συσκευής σας.

Η Τεχνητή Νοημοσύνη μπορεί να επιταχύνει τις έρευνες μετά από ένα συμβάν ασφαλείας, εντοπίζοντας και χαρτογραφώντας την πορεία ενός εισβολέα. Η ψηφιακή εγκληματολογία, ειδικά όταν εφαρμόζεται σε συμβάντα ασφαλείας, μπορεί να δημιουργήσει μεγάλο όγκο δεδομένων προς διερεύνηση. Η Τεχνητή Νοημοσύνη και, ειδικότερα, η Μηχανική Μάθηση διαδραματίζει σημαντικό ρόλο σε αυτόν τον τομέα, βοηθώντας στον εντοπισμό προτύπων, στην πραγματοποίηση ταχύτερης ανάλυσης και στην απόκτηση ακριβέστερων πληροφοριών.

Μερικά από τα πιο δημοφιλή εργαλεία που ενσωματώνουν την Τεχνητή Νοημοσύνη στις δυνατότητές τους εγκληματολογικής ανάλυσης είναι τα εξής:

- ✓ **Autopsy** (https://www.sleuthkit.org/autopsy/): αν και είναι κατά κύριο λόγο ένα εργαλείο ψηφιακής εγκληματολογικής ανάλυσης, διαθέτει ενότητες και πρόσθετα που μπορούν να αξιοποιήσουν τις δυνατότητες που βασίζονται σε Τεχνητή Νοημοσύνη για την ανάλυση δεδομένων και την αναζήτηση συγκεκριμένων μοτίβων.

- ✓ **Cellebrite** (https://www.cellebrite.com): Γνωστή για τις εγκληματολογικές λύσεις κινητών συσκευών της, η Cellebrite χρησιμοποιεί τεχνικές Τεχνητής Νοημοσύνης για να βοηθήσει στον εντοπισμό και την κατηγοριοποίηση των σχετικών δεδομένων σε κινητές συσκευές.

- ✓ **Brainspace** (https://www.brainspace.com): πρόκειται για μια πλατφόρμα αναλυτικών στοιχείων και οπτικοποίησης που χρησιμοποιεί Μηχανική Μάθηση για να βοηθήσει σε έρευνες, αναθεωρήσεις εγγράφων και ανάλυση δεδομένων. Χρησιμοποιείται σε νομικές έρευνες αλλά μπορεί επίσης να εφαρμοστεί στην ψηφιακή εγκληματολογία.

- ✓ **Cyber Triage** (https://www.cybertriage.com): Σε συνεργασία με το Autopsy, αυτό το εργαλείο χρησιμοποιεί τεχνικές Τεχνητής Νοημοσύνης για ταχεία αξιολόγηση των παραβιασμένων συστημάτων, αναζητώντας στοιχεία κακόβουλης δραστηριότητας.

✓ **Endgame** (μέρος του Elastic, https://www.elastic.co): η πλατφόρμα του παρέχει δυνατότητες αντιμετώπισης περιστατικών και απειλών και χρησιμοποιεί τεχνικές Μηχανικής Μάθησης για την ανάλυση δεδομένων και τον εντοπισμό κακόβουλης δραστηριότητας.

✓ **ReversingLabs** (https://www.reversinglabs.com): παρέχει λύσεις για την ανάλυση κακόβουλων αρχείων και τεχνουργημάτων με δυνατότητες που βασίζονται σε Τεχνητή Νοημοσύνη για τον εντοπισμό, την ταξινόμηση και τον διαχωρισμό των απειλών.

Αυτά τα εργαλεία, σε συνδυασμό με την ανθρώπινη τεχνογνωσία, μπορούν να παρέχουν ταχύτερη και ακριβέστερη ιατροδικαστική ανάλυση, η οποία είναι ζωτικής σημασίας για την αντιμετώπιση περιστατικών και τις έρευνες.

### 5.1.3 Αυτοματοποιημένα Συστήματα Απόκρισης (Automated Response Systems)

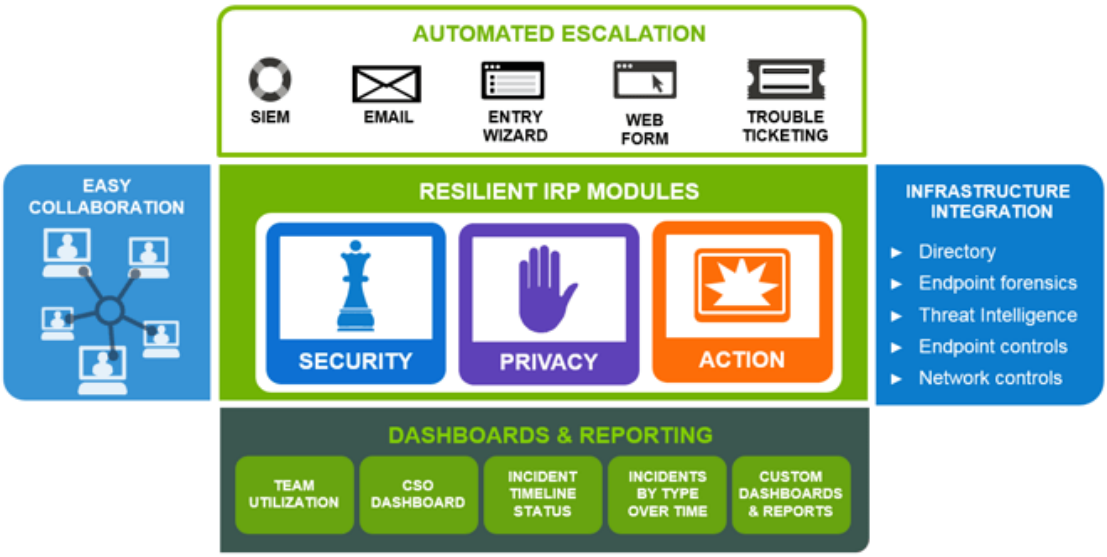
Μόλις εντοπιστεί μια απειλή, οι μηχανισμοί της Τεχνητής Νοημοσύνης μπορεί να ξεκινήσουν προκαθορισμένες ενέργειες για τον περιορισμό ή τον μετριασμό της επίθεσης, όπως η απομόνωση ενός παραβιασμένου συστήματος ή ο αποκλεισμός μιας ύποπτης διεύθυνσης IP. Η αυτοματοποιημένη απόκριση, συχνά σε συνδυασμό με την ανίχνευση απειλών, είναι ένα κρίσιμο συστατικό της σύγχρονης ασφάλειας. Με τη χρήση Τεχνητής Νοημοσύνης, αυτά τα συστήματα μπορούν να λαμβάνουν αποφάσεις σε πραγματικό χρόνο για τον περιορισμό, τον μετριασμό ή την εξουδετέρωση απειλών χωρίς άμεση ανθρώπινη παρέμβαση. Βέβαια, μερικά από τα πιο δημοφιλή εργαλεία και λύσεις που ενσωματώνουν την Τεχνητή Νοημοσύνη για την παροχή δυνατοτήτων αυτόματης απόκρισης φαίνονται παρακάτω:

✓ **Darktrace Antigena** (https://www.darktrace.com): Το Antigena είναι μια επέκταση του συστήματος ανίχνευσης που βασίζεται σε Τεχνητή Νοημοσύνη του Darktrace, το οποίο έχει τη δυνατότητα να αναλαμβάνει αυτόματες ενέργειες ως απάντηση σε απειλές που έχουν εντοπιστεί, όπως το μπλοκάρισμα συνδέσεων ή η απομόνωση (καραντίνα) συσκευών.

✓ **Palo Alto Networks - Cortex XDR** (https://www.paloaltonetworks.com): Αυτή η πλατφόρμα εντοπίζει απειλές και αυτοματοποιεί την απόκριση. Χρησιμοποιεί τεχνικές μηχανικής εκμάθησης για τον εντοπισμό απειλών και μπορεί να εκτελέσει ενέργειες όπως ο αποκλεισμός κακόβουλων διαδικασιών ή η αυτόματη ενημέρωση κανόνων τείχους προστασίας.

✓ **FireEye Helix** (https://www.fireeye.com): είναι μια πλατφόρμα ασφαλείας που χρησιμοποιεί τεχνικές Τεχνητής Νοημοσύνης για τον εντοπισμό απειλών και την αυτοματοποίηση των απαντήσεων. Μπορεί να ενσωματωθεί με μια ποικιλία εργαλείων και συστημάτων για την εκτέλεση ενεργειών απόκρισης.

✓ **IBM Resilient** (https://www.ibm.com): είναι μια πλατφόρμα απόκρισης συμβάντων που, σε συνδυασμό με το Watson, το σύστημα Τεχνητής Νοημοσύνης της IBM, μπορεί να παρέχει συστάσεις και να αυτοματοποιεί ενέργειες ως απάντηση σε συμβάντα ασφαλείας (IBM, 2022).




Εικόνα 18. IBM Resilient Incident Response. Πηγή: (IBM, 2022)

✓ **Fortinet FortiResponder** (https://www.fortinet.com): είναι μια λύση αντιμετώπισης περιστατικών που ενσωματώνεται με άλλα προϊόντα Fortinet

για να παρέχει αυτοματοποιημένες, βασισμένες σε κανόνες δυνατότητες. Ενώ η απόκριση βασίζεται κυρίως σε καθορισμένους κανόνες, ο εντοπισμός και οι πληροφορίες μπορούν να τροφοδοτηθούν από τεχνικές Τεχνητής Νοημοσύνης.

Βέβαια, η αυτοματοποίηση πρέπει να χρησιμοποιείται με προσοχή. Η εσφαλμένη διαμόρφωση ή η έλλειψη κατάλληλης επίβλεψης μπορεί να οδηγήσει σε ανεπιθύμητες αποκρίσεις που επηρεάζουν αρνητικά τις λειτουργίες. Η Τεχνητή Νοημοσύνη και ο αυτοματισμός θα πρέπει να θεωρούνται εργαλεία που συμπληρώνουν, αλλά δεν αντικαθιστούν, τους ειδικούς σε θέματα ανθρώπινης ασφάλειας.

## 5.2 Ενορχήστρωση, Αυτοματοποίηση και Απόκριση Ασφάλειας

Η ενορχήστρωση, αυτοματοποίηση και απόκριση ασφαλείας (SOAR) αναφέρεται στην ικανότητα ενός συστήματος ασφαλείας να εντοπίζει αυτόματα και να ανταποκρίνεται σε μια απειλή ή ευπάθεια χωρίς ανθρώπινη παρέμβαση, συχνά συντονίζοντας πολλαπλά συστήματα και εργαλεία στη διαδικασία. Επομένως, αυτό το μοντέλο απαιτεί τρία στοιχεία: ενορχήστρωση, κατανοητή ως συντονισμός και ολοκληρωμένη διαχείριση εργαλείων και συστημάτων ασφαλείας, που τους επιτρέπει να συνεργάζονται αρμονικά, αυτοματοποίηση, κατανοητή ως ικανότητα εκτέλεσης συγκεκριμένων εργασιών χωρίς ανθρώπινη παρέμβαση και απόκριση, αναφερόμενη σε ενέργειες που γίνονται σε απόκριση σε ένα συμβάν ασφαλείας, το οποίο μπορεί να είναι αυτόματο (π.χ. αποκλεισμός IP) ή μπορεί να απαιτεί ανθρώπινη παρέμβαση (π.χ. διερεύνηση πιθανής εισβολής).

Η χρήση εργαλείων που έχουν διαμορφωθεί σύμφωνα με το μοντέλο SOAR είναι ιδιαίτερα χρήσιμη καθώς, δεδομένης της ταχύτητας διάδοσης της απειλής, η δυνατότητα αυτόματης απόκρισης σε απειλές μπορεί να είναι καθοριστική για την ελαχιστοποίηση της ζημιάς. Επιπλέον, ο αυτοματισμός επιτρέπει στις ομάδες ασφαλείας να εστιάζουν σε εργασίες υψηλότερης αξίας ή πιο εξελιγμένες απειλές, αφήνοντας επαναλαμβανόμενες ή συνήθεις εργασίες σε αυτοματοποιημένες λύσεις, αφαιρώντας τον ανθρώπινο παράγοντα από το χειρισμό των τελευταίων και, κατά συνέπεια, μειώνοντας τον κίνδυνο

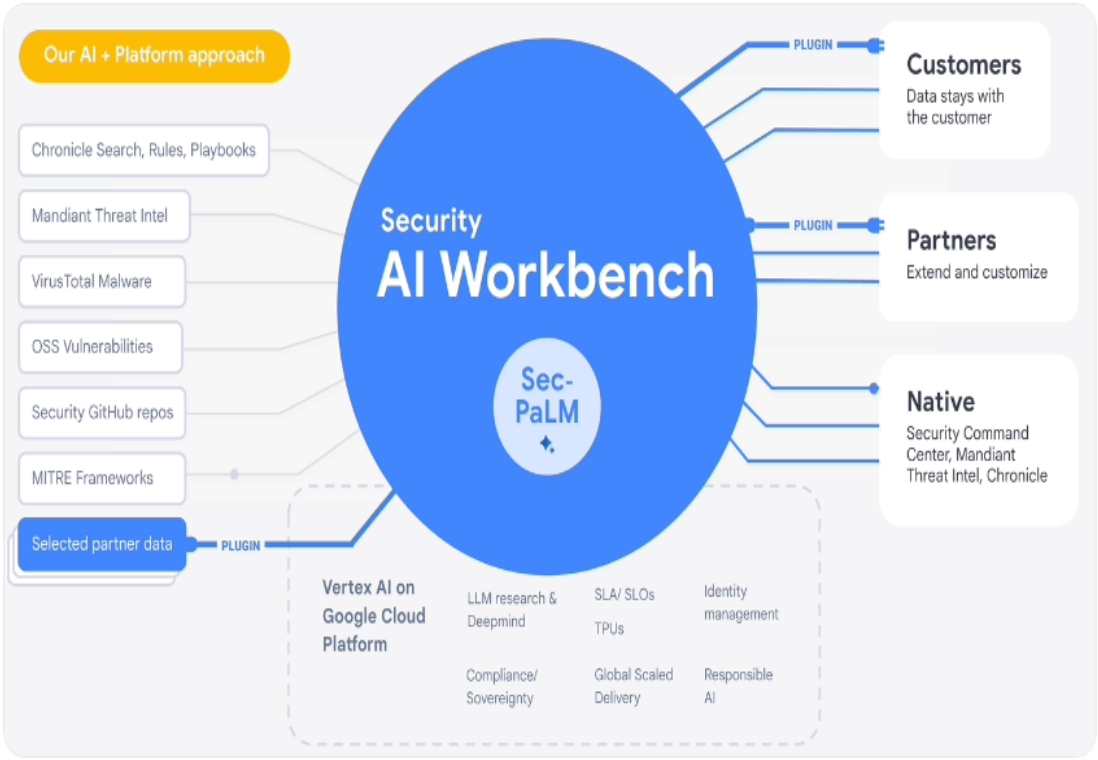
σφαλμάτων ή ασυνεπειών. Τέλος, τα συστήματα SOAR μπορούν να χειριστούν μεγάλο όγκο ειδοποιήσεων και συμβάντων, πολλά από τα οποία θα ήταν συντριπτικά για μια ανθρώπινη ομάδα.

## 5.3 Εργαλεία GenAI στην Κυβερνοασφάλεια

Η παραγωγική Τεχνητή Νοημοσύνη έχει γίνει ένα πολύτιμο εργαλείο σε διάφορους τομείς, από τη δημιουργία τέχνης έως τη σύνθεση δεδομένων. Στο πλαίσιο της κυβερνοασφάλειας, η παραγωγική Τεχνητή Νοημοσύνη μπορεί να είναι ταυτόχρονα λύση και πιθανή απειλή. Τα σημαντικότερα διαθέσιμα εργαλεία είναι τα εξής:

### 5.3.1 Google Cloud Security AI Workbench

Το Google Cloud Security AI Workbench έχει σχεδιαστεί για να υποστηρίζει προηγμένες απειλές και ευφυΐα ασφαλείας, ανίχνευση κακόβουλου λογισμικού, ανάλυση συμπεριφοράς και διαχείριση ευπάθειας. Επίσης, τροφοδοτεί νέες προσφορές που μπορούν πλέον να αντιμετωπίσουν μοναδικά κορυφαίες προκλήσεις ασφαλείας όπως υπερφόρτωση απειλών και επίπονα εργαλεία. Διαθέτει επίσης ενσωματώσεις plug-in συνεργατών για να προσφέρει στους πελάτες την ευφυΐα απειλών, τη ροή εργασιών και άλλες κρίσιμες λειτουργίες ασφάλειας, με την Accenture να είναι ο πρώτος «συνεργάτης» που χρησιμοποιεί το Security AI Workbench (Potti, 2022).




Εικόνα 19. Google Cloud Security AI Workbench. Πηγή: (Potti, 2022)

Η πλατφόρμα θα επιτρέπει επίσης στους πελάτες να διαθέσουν τα προσωπικά τους δεδομένα στην πλατφόρμα κατά τη στιγμή του συμπεράσματος, διασφαλίζοντας την τήρηση όλων των δεσμεύσεων περί απορρήτου δεδομένων προς τους πελάτες. Επειδή το Security AI Workbench είναι χτισμένο στην υποδομή Vertex AI του Google Cloud, οι πελάτες ελέγχουν τα δεδομένα τους με δυνατότητες εταιρικού επιπέδου, όπως απομόνωση δεδομένων, προστασία δεδομένων, κυριαρχία και υποστήριξη συμμόρφωσης.

### 5.3.2 Microsoft Security Copilot

Το Microsoft Security Copilot είναι μια από τις πιο στοχευμένες λύσεις ασφαλείας στο οπλοστάσιο των προϊόντων τεχνητής νοημοσύνης της Microsoft. Λειτουργεί για τη βελτιστοποίηση της απόκρισης συμβάντων, της ανίχνευσης απειλών και της αναφοράς ασφάλειας για τους χρήστες και ενσωματώνει πληροφορίες και πληροφορίες από εργαλεία όπως το Microsoft Sentinel, το Microsoft Defender και το Microsoft Intune. Το Security Copilot

παρέχει μια φυσική γλώσσα, υποβοηθητική εμπειρία Copilot που βοηθά στην υποστήριξη των επαγγελματιών ασφαλείας σε σενάρια από άκρο σε άκρο, όπως απόκριση περιστατικού, κυνήγι απειλών, συλλογή πληροφοριών και διαχείριση στάσης του σώματος.

Η λύση αξιοποιεί την πλήρη ισχύ της αρχιτεκτονικής OpenAI για να δημιουργήσει μια απάντηση σε μια προτροπή χρήστη χρησιμοποιώντας πρόσθετα ειδικά για την ασφάλεια, συμπεριλαμβανομένων πληροφοριών για συγκεκριμένους οργανισμούς, έγκυρων πηγών και πληροφοριών παγκόσμιας απειλής. Χρησιμοποιώντας πρόσθετα ως πηγές σημείων δεδομένων, οι επαγγελματίες ασφάλειας έχουν ευρύτερη ορατότητα στις απειλές και αποκτούν περισσότερο πλαίσιο και έχουν την ευκαιρία να επεκτείνουν τις λειτουργίες της λύσης. Σχεδιασμένο με γνώμονα την ενοποίηση, το Security Copilot ενσωματώνεται απρόσκοπτα με προϊόντα του χαρτοφυλακίου Microsoft Security, όπως το Microsoft 365 Defender, το Microsoft Sentinel, το Microsoft Intune, καθώς και με άλλες υπηρεσίες τρίτων, όπως το ServiceNow.

### 5.3.3 CrowdStrike Charlotte AI

Αυτό το εργαλείο CrowdStrike επιτρέπει στους χρήστες να διαχειρίζονται την ασφάλεια στον κυβερνοχώρο μέσω φυσικής γλώσσας στην πλατφόρμα Falcon. Όπως πολλά από αυτά τα αναδυόμενα εργαλεία τεχνητής νοημοσύνης στον κυβερνοχώρο, το Charlotte AI έχει σχεδιαστεί για να συμπληρώνει τις υπάρχουσες ομάδες ασφαλείας και να μειώνει τον αντίκτυπο των κενών δεξιοτήτων. Το Charlotte AI χρησιμοποιείται γενικά για την υποστήριξη των προσπαθειών ανίχνευσης απειλών και αποκατάστασης.

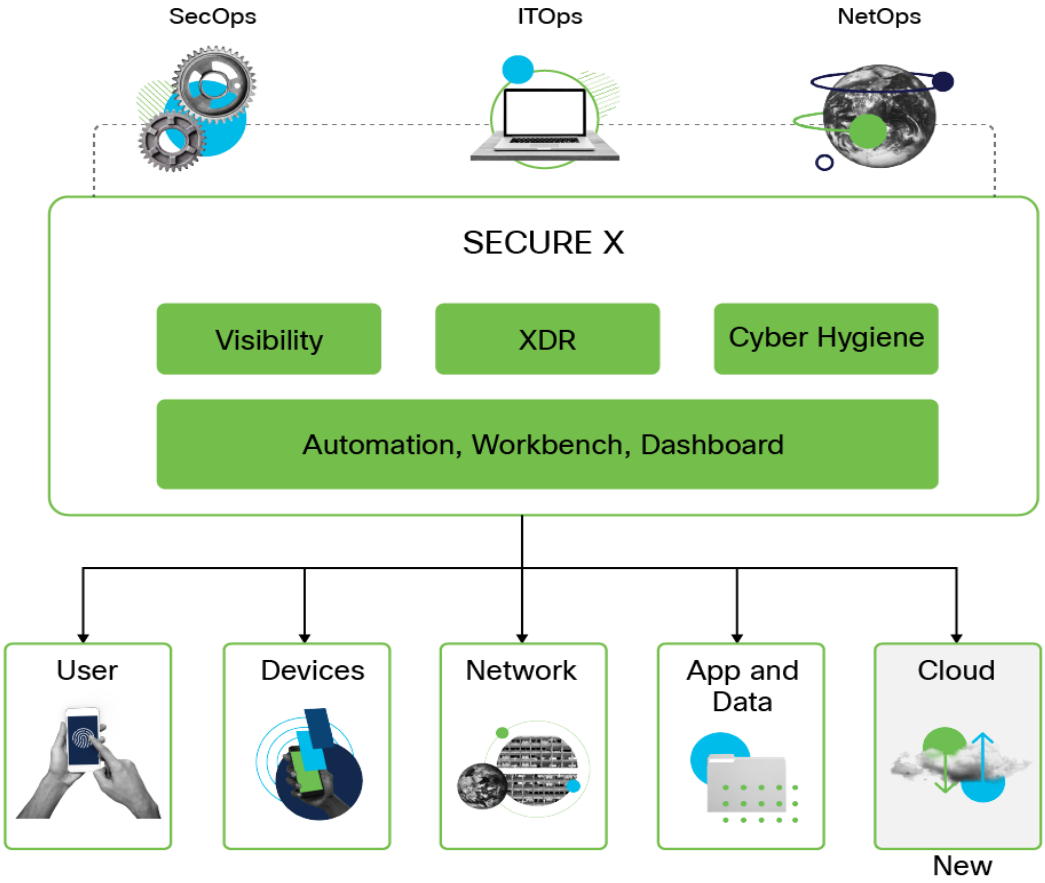
Τέλος, η Charlotte AI λειτουργεί ως ο απόλυτος πολλαπλασιαστής δύναμης για τους ειδικούς σε θέματα ασφάλειας. Εξουσιοδοτεί έμπειρους επαγγελματίες να αυτοματοποιούν επαναλαμβανόμενες εργασίες, συμπεριλαμβανομένης της συλλογής δεδομένων, της εξαγωγής και της βασικής αναζήτησης και ανίχνευσης απειλών. Επιπλέον, απλοποιεί την εκτέλεση πιο προηγμένων ενεργειών ασφαλείας. Το Charlotte AI ανοίγει επίσης τον δρόμο για ταχεία χρήση εκτεταμένης ανίχνευσης και απόκρισης (Extended Detection and Response - XDR) σε όλη την επιχείρηση, που καλύπτουν κάθε επιφάνεια

επίθεσης και προϊόν τρίτου μέρους, απευθείας από την πλατφόρμα CrowdStrike Falcon.

### 5.3.4 Cisco Security Cloud

Η Cisco κάνει μια προσπάθεια να προωθήσει την Τεχνητή Νοημοσύνη βαθύτερα στην πλατφόρμα ασφάλειας cloud της, λανσάροντας μια νέα δυνατότητα, το AI Assistant for Security, έναν βοηθό μέσω τεχνητής νοημοσύνης μεταξύ τομέων που έχει σχεδιαστεί για να βοηθά οργανισμούς όλων των μεγεθών να βελτιώσουν την άμυνά τους ενάντια στο διαρκώς αυξανόμενο αριθμός απειλών.

Με το Encrypted Visibility Engine που λειτουργεί με AI για όλα τα μοντέλα τείχους προστασίας, η Cisco στοχεύει να αντιμετωπίσει μια πρόκληση που πιστεύει ότι εμποδίζει την ανίχνευση κακόβουλου λογισμικού. Δεδομένου ότι το μεγαλύτερο μέρος της κίνησης των κέντρων δεδομένων είναι κρυπτογραφημένη, η αδυναμία επιθεώρησης της κρυπτογραφημένης κίνησης αποτελεί βασική ανησυχία για την ασφάλεια. Επίσης, η Cisco προσθέτει δυνατότητες παραγωγής τεχνητής νοημοσύνης στο Security Cloud και στα χαρτοφυλάκια συνεργασίας και ασφάλειας. Οι νέες δυνατότητες έχουν σχεδιαστεί για να διευκολύνουν τη διαχείριση πολιτικής και την απόκριση σε απειλές.




Εικόνα 20. **Cisco Security Cloud. Πηγή:** (Cisco, 2021)

### 5.3.5 Airgap Networks Threat GPT

Βασισμένη στο Generative Pre-trained Transformer 3 (GPT-3) και στις βάσεις δεδομένων γραφημάτων, το ThreatGPT είναι μια λύση Airgap Networks που βοηθά τις επιχειρήσεις να αναλύουν πιο αποτελεσματικά και ολιστικά τις απειλές ασφαλείας σε περιβάλλοντα Λειτουργικής Τεχνολογίας (Operational Technology - OT) και παλαιού τύπου συστήματα.

Το ThreatGPT της Airgap Networks είναι ένα νέο εργαλείο που στοχεύει στη βελτίωση της ασφάλειας στον κυβερνοχώρο για περιβάλλοντα ΛΤ. Τα συστήματα ΛΤ, αναπόσπαστα σε τομείς όπως η κατασκευή και η ενέργεια, περιλαμβάνουν παρακολούθηση και έλεγχο φυσικών συσκευών. Η πολυπλοκότητα και τα απαρχαιωμένα συστήματα που χρησιμοποιούνται συχνά στα περιβάλλοντα ΛΤ τα καθιστούν ευάλωτα σε απειλές στον κυβερνοχώρο.

Η αντιμετώπιση αυτών των προκλήσεων απαιτεί λύσεις που συνδυάζουν την παραδοσιακή ασφάλεια με προηγμένη Τεχνητή Νοημοσύνη και Μηχανική Μάθηση. Το ThreatGPT σκοπεύει να βελτιώσει την ασφάλεια των συστημάτων ΛΤ, γεφυρώνοντας αυτό το χάσμα. Ως νέο εργαλείο, υποδηλώνει τις δυνατότητες των εργαλείων με δυνατότητα τεχνητής νοημοσύνης που είναι προσαρμοσμένα στην ασφάλεια των συστημάτων ΛΤ έναντι των αναδυόμενων απειλών. Το ThreatGPT και παρόμοια εργαλεία μπορεί να αποδειχθούν χρήσιμα για την ενίσχυση της ασφάλειας των συστημάτων ΛΤ εν μέσω διαρκώς εξελισσόμενων κινδύνων.


Εικόνα 21. Airgap Networks Threat GPT. Πηγή: (Airgap, 2021)

Η δυναμική του ThreatGPT έγκειται στην ικανότητά του να αξιοποιεί τις τεχνολογίες Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης για να θωρακίζει περιβάλλοντα ΛΤ. Αυτά τα περιβάλλοντα, βασικά στοιχεία της σύγχρονης υποδομής, έχουν συχνά «παραμεληθεί» στα συμβατικά μοντέλα κυβερνοασφάλειας. Ο σκοπός του ThreatGPT είναι να αναλύει δεδομένα που σχετίζονται με την ασφάλεια εντός συστημάτων ΛΤ, παρακολουθώντας διάφορους παράγοντες όπως η κυκλοφορία δικτύου, οι συμπεριφορές του συστήματος, οι δραστηριότητες των χρηστών και άλλα σήματα που μπορεί να υποδηλώνουν επικείμενο κίνδυνο ασφάλειας.

### 5.3.6 SentinelOne

Το SentinelOne εστιάζει στο να ενεργεί πιο γρήγορα και πιο έξυπνα μέσω της πρόληψης με Τεχνητή Νοημοσύνη και της αυτόνομης ανίχνευσης και απόκρισης. Με την πλατφόρμα Singularity XDR, οι οργανισμοί αποκτούν πρόσβαση σε δεδομένα υποστήριξης σε ολόκληρο τον οργανισμό μέσω μιας ενιαίας λύσης, παρέχοντας μια συνεκτική εικόνα του δικτύου και των στοιχείων τους, προσθέτοντας ένα αυτόνομο επίπεδο ασφαλείας σε πραγματικό χρόνο σε όλα τα εταιρικά στοιχεία.

Ο οργανισμός πρόσφατα αναβάθμισε (και περιόρισε) την πλατφόρμα καταγραφής απειλών με δυνατότητες δημιουργίας τεχνητής νοημοσύνης. Έχει σχεδιαστεί για να κλιμακώνει τις λειτουργίες ασφαλείας και την ανίχνευση απειλών, βασιζόμενος σε ολοκληρωμένα νευρωνικά δίκτυα και εκτεταμένη μοντελοποίηση γλώσσας για να παρέχει καλύτερες και πιο κοντά σε πραγματικό χρόνο πληροφορίες σχετικά με πιθανές απειλές και λύσεις.

### 5.3.7 Synthesis Humans

Το SentinelOne είναι μια πολυεπίπεδη πλατφόρμα ασφάλειας στον κυβερνοχώρο με Τεχνητή Νοημοσύνη και ανάλυση συμπεριφοράς που ανταποκρίνεται σε απειλές με ταχύτητα μηχανής και σας προστατεύει άμεσα από προηγμένες και επίμονες επιθέσεις. Όλα τα δεδομένα συσχετίζονται στη λειτουργία Storyline™, η οποία σας δίνει πλήρη ορατότητα οποιωνδήποτε προσπαθειών παραβίασης και επιτρέπει στους αναλυτές να διερευνήσουν σε βάθος τι συνέβη. Όλα αυτά σε συνδυασμό με την αυτοματοποιημένη αποκατάσταση, τη δυνατότητα επαναφοράς με ένα κλικ και την προστασία σε περισσότερα λειτουργικά συστήματα από οποιαδήποτε άλλη πλατφόρμα ασφαλείας. Αυτό διασφαλίζει ότι οι υπάλληλοί σας προστατεύονται κάθε χιλιοστό του δευτερολέπτου κάθε μέρα.

Το Synthesis Humans είναι ένα από τα πολλά εργαλεία παραγωγής που προσφέρει η Synthesis AI. Αυτή η λύση έχει σχεδιαστεί για να

εκπαιδεύει βιομετρικά συστήματα ελέγχου πρόσβασης με πιο ευέλικτο τρόπο. Σε συνδυασμό με τα σενάρια σύνθεσης, αυτό το εργαλείο μπορεί να χρησιμοποιηθεί για την υποστήριξη της ασφάλειας εγκαταστάσεων καθώς και για την ασφάλεια στον κυβερνοχώρο.


Εικόνα 22. AI Synthesis Humans. Πηγή: (Joseph, 2023)

### 5.3.8 SecurityScorecard

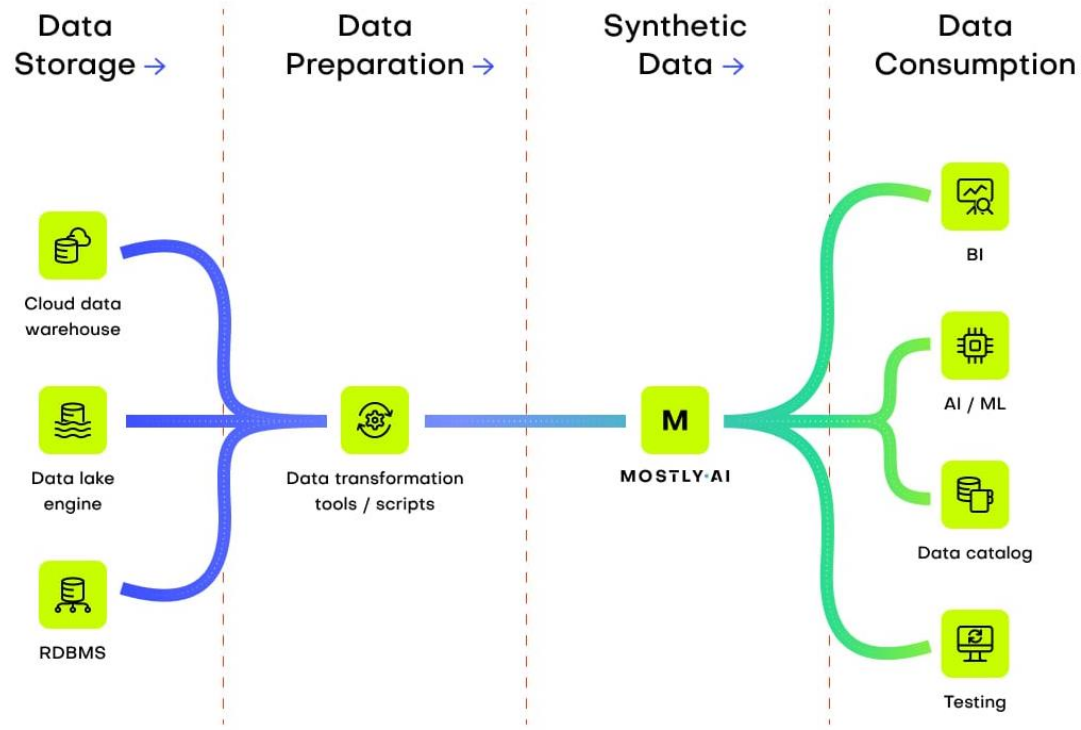
Η SecurityScorecard κυκλοφόρησε μια πλατφόρμα αξιολόγησης ασφαλείας που βασίζεται εν μέρει στο GPT-4 του OpenAI. Με αυτήν τη λύση, οι ομάδες ασφαλείας μπορούν να κάνουν ανοιχτές ερωτήσεις σε απλή γλώσσα σχετικά με την ασφάλεια του δικτύου τους και τρίτων προμηθευτών και να λαμβάνουν προληπτικές απαντήσεις και καθοδήγηση διαχείρισης κινδύνου.

Η πλατφόρμα SecurityScorecard εισάγει συνεχώς βελτιώσεις που βελτιστοποιούν τις αξιολογήσεις ασφαλείας, έτσι ώστε οι χρήστες να μπορούν να έχουν την ακριβέστερη κατανόηση του κινδύνου για να λαμβάνουν πιο έξυπνες και πιο ενημερωμένες επιχειρηματικές αποφάσεις. Επίσης, δύναται να αναλύσει μεγάλο όγκο δεδομένων από περισσότερες από 1,5 εκατομμύριο εταιρείες παγκοσμίως και αξιοποιήσει τη μηχανική εκμάθηση για να προσδιοριστούν τα βάρη των παραγόντων κινδύνου. Αυτό παρέχει στους

χρήστες μας μοναδικές πληροφορίες σχετικά με ζωτικής σημασίας συμβάντα και τάσεις κυβερνοασφάλειας σε κλίμακα και σε ένα ευρύ φάσμα εταιρειών.

### 5.3.9 MOSTLY AI

Το MOSTLY AI είναι ένα συνθετικό εργαλείο δημιουργίας δεδομένων ειδικά σχεδιασμένο για τη δημιουργία ανώνυμων δεδομένων που πληρούν διάφορες απαιτήσεις ασφάλειας και συμμόρφωσης. Λόγω της έντονης εστίασής του στην ασφάλεια και τη συμμόρφωση, χρησιμοποιείται συχνά σε ρυθμιζόμενους τομείς όπως η τραπεζική και η κοινωνική ασφάλιση (Mostly, 2022).




Εικόνα 23. Δημιουργία συνθετικών δεδομένων για ανωνυμοποίηση δεδομένων, αύξηση δεδομένων, καταλογισμό και επανεξισορρόπηση. Πηγή: (Mostly, 2022)

## Συμπεράσματα

Συμπερασματικά, ο τομέας της κυβερνοασφάλειας έχει επηρεαστεί σημαντικά από την Τεχνητή Νοημοσύνη. Είναι ένα ουσιαστικό εργαλείο για την προστασία των πληροφοριών λόγω της ικανότητάς του να εντοπίζει, να αξιολογεί και να αποφεύγει τους κινδύνους στον κυβερνοχώρο.

Η Τεχνητή Νοημοσύνη μπορεί να σαρώσει τεράστιους όγκους δεδομένων, να εντοπίσει ανωμαλίες και να προβλέψει τις αδυναμίες, βοηθώντας τις εταιρείες και τους ανθρώπους να αμυνθούν πιο γρήγορα και αποτελεσματικά από επιθέσεις. Οι επιτυχημένες εφαρμογές της Τεχνητής Νοημοσύνης στην κυβερνοασφάλεια δείχνουν ήδη την υπόσχεση και τα πλεονεκτήματα της τεχνολογίας. Ωστόσο, ο αντίκτυπος της στα εργαλεία κυβερνοασφάλειας είναι βαθύς και πολύπλευρος. Από τη βελτίωση της ανίχνευσης απειλών έως την αυτοματοποίηση των διαδικασιών ασφαλείας, η Τεχνητή Νοημοσύνη έχει γίνει ένας απαραίτητος σύμμαχος στη συνεχιζόμενη μάχη κατά των απειλών στον κυβερνοχώρο. Καθώς η τεχνολογία προχωρά, η συνεργασία μεταξύ της ανθρώπινης τεχνογνωσίας και των δυνατοτήτων Τεχνητής Νοημοσύνης θα είναι ζωτικής σημασίας για τη διασφάλιση ενός ασφαλούς ψηφιακού περιβάλλοντος.

Τα τελευταία χρόνια, τα εργαλεία της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης είναι πλέον καταλυτικής σημασίας στην ασφάλεια πληροφοριών και στην κυβερνοασφάλειας εν γένει, καθώς είναι σε θέση να αναλύουν γρήγορα εκατομμύρια συμβάντα και να εντοπίζουν πολλούς διαφορετικούς τύπους απειλών – από κακόβουλο λογισμικό που εκμεταλλεύεται zero-day τρωτά σημεία έως τον εντοπισμό ασυνήθιστης συμπεριφοράς που μπορεί να οδηγήσει σε ηλεκτρονικό ψάρεμα επίθεση ή λήψη κακόβουλου κώδικα. Αυτές οι τεχνολογίες μαθαίνουν με την πάροδο του χρόνου, αντλώντας από το παρελθόν για να εντοπίσουν νέους τύπους επιθέσεων. Τα ιστορικά συμπεριφοράς δημιουργούν προφίλ σε χρήστες, περιουσιακά στοιχεία και δίκτυα, επιτρέποντας στη Τεχνητή Νοημοσύνη να

εντοπίζει και να ανταποκρίνεται σε αποκλίσεις από τους καθιερωμένους κανόνες.

Στο μέλλον, οι αλγόριθμοι Τεχνητής Νοημοσύνης και Μηχανικής Μάθησης θα οδηγήσουν το μέλλον των δοκιμών ασφάλειας στον κυβερνοχώρο, οι οργανισμοί θα έχουν τα εργαλεία που χρειάζονται για να παραμείνουν μπροστά από την καμπύλη και να διατηρήσουν τα δεδομένα τους ασφαλή.

# Βιβλιογραφία


Abdelhamid, N., Ayesh, A. & Thabtah, F., 2014. *Phishing detection based associative classification data mining.* s.l.:Expert Systems with Applications.

Abonamah, A., 2021. *On the Commoditization of Artificial Intelligence.* 12 επιμ. s.l.:Frontiers in Psychology.

Airgap, 2021. *Airgap.* [Ηλεκτρονικό] Available at: https://airgap.io/discovery-and-visibility/ [Πρόσβαση 12 Ιανουαρίος 2024].

Anderson, J. & Corbett, A., 1993. *Tutoring of Cognitive Skill.* s.l.:Psychology Press.

Antonakakis, M. και συν., 2012. *From throw-away traffic to bots: detecting the rise of DGA-based malware.* s.l.:USENIX Security Symposium.

Apruzzese, G. και συν., 2018. *On the Effectiveness of Machine and Deep Learning for Cyber Security.* s.l.:10th International Conference on Cyber Conflict.

Bezirganyan, G. & Sergoyan, H., 2022. *A Brief Comparison Between White Box, Targeted Adversarial Attacks in Deep Neural Networks.* s.l.:Mathematical Problems of Computer Science.

Bieszczad, A. & Kuchar, S., 2020. *Neurosolver Learning to Solve Towers of Hanoi Puzzles.* s.l.:Computer Science.

Bishop, C. & Bishop, H., 2023. *Deep Learning: Foundations and Concepts.* s.l.:Springer.

Buczak, A. & Guven, E., 2015. *A survey of data mining and machine learning methods for cyber security intrusion detection.* s.l.:IEEE Communications Surveys & Tutorials.

Cavelty, D., 2012. *Cyber-security.* p. 33 επιμ. s.l.:Oxford University Press.

Chakraborty, A., Biswas, A. & Khan, A., 2022. *Artificial Intelligence for Cybersecurity: Threats Attacks and Mitigation.* s.l.:Computer Science.

Cisco, 2021. *Secure Cloud Insights At-a-Glance.* s.l.:Cisco.

Cordero, C. & Pascual, C., 2023. *Approach to Artificial Intelligence and Cybersecurity.* s.l.:CCN-CERT.

Coursesteach, 2023. *Medium.* [Ηλεκτρονικό] Available at: https://medium.com/@Coursesteach/natural-language-processing-part-1-5727b4efc8b4 [Πρόσβαση 11 Νοέμβριος 2023].

Dahl, G., Stokes, W., Deng, L. & Yu, D., 2013. *Large-scale malware classification using random projections and neural networks.* s.l.:IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

DeAngelis, S., 2021. *Is Reinforcement Learning the Future of Artificial Intelligence?.* [Ηλεκτρονικό] Available at: https://enterrasolutions.com/is-reinforcement-learning-the-future-of-artificial-intelligence/ [Πρόσβαση 10 Νοέμβριος 2023].

Dhamani, N. & Engler, M., 2023. *Introduction to Generative AI.* s.l.:Manning.

Diogenes, Y. & Ozkaya, E., 2018. *Cybersecurity - Attack and Defense Strategies.* s.l.:Packt Publishing Limited.

European Commission, 2020. *New EU Cybersecurity Strategy and new rules to make physical and digital critical entities more resilient.* s.l.:European Commission.

Grand View Research, 2022. *Artificial Intelligence In Cybersecurity Market Size Report.* s.l.:Grand View Research.

Griffiths, C., 2023. *The Latest 2023 Cyber Crime Statistics.* s.l.:AAG IT Services.

Grzemba, A., 2019. *Decentralized Anomaly Detection with unused Computing Power in Avionic and Automotive Applications.* s.l.:Kaspersky Industrial Cybersecurity Conference.

Gupta, M. και συν., 2023. *From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy.* s.l.:arXiv.

Haenlein, M., 2019. *A Brief History of Artificial Intelligence: On the Past Present, and Future of Artificial Intelligence.* vol. 61, nr 4 επιμ. s.l.:Sage Journals Home.

Haenlein, M. & Kaplan, A., 2019. *A brief history of artificial intelligence: On the past, present and future of artificial intelligence.* 61(4), pp.5-14 επιμ. s.l.:California Management Review.

Hardy, W. και συν., 2016. *DL4MD: A Deep Learning Framework for Intelligent Malware Detection.* s.l.:International Conference on Data Mining (DMIN).

Hill, G. & Bellekens, X., 2017. *Deep Learning Based Cryptographic Primitive Classification.* s.l.:arXiv preprint.

Hsiao, W. & Chang, T., 2008. *An incremental cluster-based approach to spam filtering.* s.l.:Expert Systems with Applications.

IBM, 2019. *Building trust in AI.* s.l.:IBM.

IBM, 2022. *IBM Resilient SOAR Platform.* s.l.:IBM.

infoDiagram LTD, 2021. *Infodiagram.* [Ηλεκτρονικό] Available at: https://www.infodiagram.com/slides/ai-development-timeline/ [Πρόσβαση 15 Νοέμβριος 2023].

Jordan, M. & Mitchell, T., 2015. *Machine learning: Trends, perspectives, and prospects.* s.l.:Science.

Joseph, T., 2023. *Synthesis AI cybersecurity.* s.l.:Synthesis.

Kasper, A., 2020. *EU Cybersecurity Governance -Stakeholders and Normative Intentions towards Integration.* pp. 166-185 επιμ. s.l.:Institute for European Studies.

Kissell, L., 2021. Algorithmic Trading. Στο: *Algorithmic Trading Methods.* s.l.:ScienceDirect, p. 23–56.

KPMG, 2023. *Trust in Artifical Intelligence - A global study.* s.l.:KPMG.

Kshetri, N., 2010. *The global cybercrime industry: economic, institutional and strategic perspectives.* s.l.:Springer Science & Business Media.

Kubat, M., 2018. *Introduction to Machine Learning.* s.l.:Springer International Publishing AG.

LeCun, Y., Bengio, Y. & Hinton, G., 2015. *Deep learning.* s.l.:Nature.

Li, Y., Ma, R. & Jiao, R., 2015. *A hybrid malicious code detection method based on deep learning.* s.l.:International Journal of Security and Its Applications.

Loshin, P., 2022. *User Behavior Analytics (UBA).* [Ηλεκτρονικό] Available at: https://www.techtarget.com/searchsecurity/definition/user-behavior-analytics-UBA [Πρόσβαση 16 Δεκέμβριος 2023].

McNamara, K., 2016. *StealthWatch Introduction.* s.l.:Networking.

Mostly, 2022. s.l.:Mostly.

National Institute of Standards and Technology, 2018. *Framework for Improving Critical Infrastructure Cybersecurity.* s.l.:National Institute of Standards and Technology.

National Institute of Standards and Technology, 2023. *Public Draft: The NIST Cybersecurity Framework 2.0.* s.l.:National Institute of Standards and Technology.

NordLayer, 2023. *NordLayer.* [Ηλεκτρονικό] Available at: https://nordlayer.com/blog/using-ai-in-cybersecurity/ [Πρόσβαση 23 Μάϊος 2024].

Norvig, P. & Russell, S., 2021. *Artificial Intelligence: A Modern Approach.* s.l.:Pearson.

Obotivere, B. & Nwaezeigwe, A., 2020. *Cyber security threats on the internet and possible solutions.* s.l.:IJARCCE.

Paper, D., 2021. *Stacked Autoencoders. In: State-of-the-Art Deep Learning Models in TensorFlow.* s.l.:Berkeley.

Pascanu, R. και συν., 2015. *Malware classification with recurrent networks.* s.l.:IEEE International Conference on Acoustics, Speech and Signal Processing.

Paula, D. & Cruz , M., 2023. *Cybersecurity Essentials Made Easy: A No-Nonsense Guide to Cyber Security For Beginners.* s.l.:Independently.

Potti, S., 2022. *Cloud.google.com.* [Ηλεκτρονικό] Available at: https://cloud.google.com/blog/products/identity-security/rsa-google-cloud-security-ai-workbench-generative-ai [Πρόσβαση 10 Ιανουάριος 2024].

Pumperla , M. & Ferguson, K., 2019. *Deep Learning and the Game of Go.* s.l.:Manning.

Reveron, D., 2023. *Security in the Cyber Age.* s.l.:Cambridge University Press.

Rieck, K., Trinius, P., Willems, C. & Holz, T., 2011. *Automatic Analysis of Malware Behavior Using Machine Learning.* s.l.:Journal of Computer Security.

Roy, R., 2020. *Towardsdatascience.* [Ηλεκτρονικό] Available at: https://towardsdatascience.com/understanding-the-difference-between-ai-ml-and-dl-cceb63252a6c [Πρόσβαση 11 Νοέμβριος 2023].

Russell, S. & Norvig, P., 2020. *Artificial Intelligence: A Modern Approach.* 4η επιμ. s.l.:Pearson.

Şen , Z., 2023. *Shallow and Deep Learning Principles: Scientific, Philosophical, and Logical Perspectives.* s.l.:Springer.

Singh, K., 2021. *Principles of Generative AI - A Technical Introduction.* s.l.:Tepper School of Business.

Tunstall , L., von Werra, L. & Wolf, T., 2022. *Natural Language Processing with Transformers.* s.l.:O'Reilly Media.

Turing, A., 1950. *Computing Machinery and Intelligence.* Vol. 59, No. 236 επιμ. s.l.:Oxford University Press.

Wang, K. και συν., 2010. *Security issues and challenges for cyber physical system. In Green Computing and Communications.* s.l.:IEEE.

Weidman, G., 2014. *Penetration Testing.* s.l.:No Starch Press.

Winston, H., 1992. *Artificial intelligence.* s.l.:Addison-Wesley Longman Publishing Co., Inc..