

2024

$\beta \ddot{y} \cdot \mathcal{A} \pm \hat{A} \frac{1}{4} \zeta^3 \textcircled{R} \quad 0 \pm 1 \quad \tilde{A} \cdot \frac{1}{4} \pm \tilde{A}^- \pm \ddot{A} \cdot \hat{A} \quad \ddot{A}$
 $\beta \ddot{y} \frac{1}{2} \zeta \cdot \frac{1}{4} \zeta \tilde{A} \hat{I} \frac{1}{2} \cdot \hat{A} \quad \tilde{A} \ddot{A}^1 \hat{A} \quad '1 \pm '1^0 \pm \tilde{A}^- \mu \hat{A}$
 $\beta \ddot{y} \mu \frac{1}{2} \ddot{A} \zeta \hat{A}^1 \tilde{A} \frac{1}{4} \zeta \hat{I} \quad 0 \pm 1 \quad '1 \mu \hat{A} \mu \hat{I} \frac{1}{2} \cdot \tilde{A} \cdot \ddot{A}$

$\beta \ddot{y} \pm \hat{A} \pm \tilde{A} \ddot{A} \pm \ddot{A}^- \cdot \hat{A}, \quad ' \frac{1}{2} \ddot{A} \hat{I} \frac{1}{2}^1 \zeta \hat{A}$

$\beta \ddot{y} \in \mu \ddot{A} \pm \hat{A} \ddot{A} \hat{A} \zeta^1 \pm 0 \hat{I} \quad \hat{A} \hat{I}^3 \hat{A} \pm \frac{1}{4} \frac{1}{4} \pm \tilde{A} \ddot{A} \cdot \frac{1}{2} \cdot \bullet^{30} \gg \cdot \frac{1}{4} \pm \ddot{A} \zeta \gg \zeta^{310} \textcircled{R} \quad \gg \zeta^{31} \tilde{A} \ddot{A}^{10} \textcircled{R}^0 \pm 1 \quad \ddot{A} \pm \S \hat{A} \cdot \frac{1}{4} \pm$
 $\beta \ddot{y} \in \zeta \zeta \gg \textcircled{R} \quad \ddot{Y}^{10} \zeta \frac{1}{2} \zeta \frac{1}{4}^{10} \hat{I} \frac{1}{2} \cdot \hat{A}^1 \tilde{A} \ddot{A} \cdot \frac{1}{4} \hat{I} \frac{1}{2} \quad 0 \pm 1 \quad "1 \zeta^{-0} \cdot \tilde{A} \cdot \hat{A}, \quad \pm \frac{1}{2} \mu \hat{A}^1 \tilde{A} \ddot{A} \textcircled{R} \frac{1}{4}^1 \zeta \cdot \mu \neg \hat{A} \zeta \gg^1 \hat{A}$

<http://hdl.handle.net/11728/12821>

Downloaded from HEPHAESTUS Repository, Neapolis University institutional repository

**Μεταπτυχιακό Πρόγραμμα Σπουδών στην Εγκληματολογική
Λογιστική**

**Τίτλος: Εφαρμογή και σημασία της τεχνητής νοημοσύνης στις
διαδικασίες εντοπισμού και διερεύνηση της απάτης**

Ονοματεπώνυμο:

Αντώνιος Παραστατίδης



UNIVERSITY OF
WESTERN MACEDONIA

Ευχαριστίες

Με την ολοκλήρωση της μεταπτυχιακής διπλωματικής μου εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλλαν στην εκπόνησή της.

Ευχαριστώ θερμά τον επιβλέπων καθηγητή μου, Κύριο Κωνσταντινίδη Αναστάσιο, για την επιστημονική του καθοδήγηση, τις υποδείξεις του, την επιμονή του, το αμείωτο ενδιαφέρον του και την συμπαράσταση του όλα αυτά τα χρόνια τόσο στο προπτυχιακό μου όσο και στο μεταπτυχιακό μου.

Ευχαριστώ θερμά όλους τους καθηγητές του μεταπτυχιακού που με μεγάλη χαρά μοιράστηκαν τις γνώσεις τους μαζί μας.

Ευχαριστώ βαθιά την Σύζυγό μου Πασχαλίνα Στουγιάννη και τις κόρες μου Μάρθα και Μαρία Παραστατίδου για την παρακίνηση, την συμπαράσταση αλλά και την κατανόηση που έδειξαν καθ'όλη την διάρκεια των μεταπτυχιακών μου σπουδών.

Τέλος θέλω να ευχαριστήσω πολύ τους γονείς μου, Κωνσταντίνο Παραστατίδη και Μάρθα Ελευθεριάδου όπως επίσης και την θεία μου Όλγα Παραστατίδου, οι οποίοι υπήρξαν πάντα ένα ανεκτίμητο στήριγμα για μένα και στους οποίους οφείλω όλη την διαδρομή των σπουδών μου εως σήμερα.

Κ.Π. Καβάφη, «Το Πρώτο Σκαλί»

Εἰς τὸν Θεόκριτο παραπονιούνταν
μια μέρα ὁ νέος ποιητὴς Εὐμένης·
«Τώρα δυὸ χρόνια πέρασαν που γράφω
κ' ἓνα εἰδύλλιο ἕκαμα μονάχα.
Τὸ μόνον ἄρτιον μου ἔργον εἶναι.
Ἀλλοίμονον, εἶν' ὑψηλὴ τὸ βλέπω,
πολὺ ὑψηλὴ τῆς Ποιήσεως ἡ σκάλα·
κι ἀπ' τὸ σκαλί τὸ πρῶτο ἐδῶ που εἶμαι,
ποτέ δεν θ' ἀνεβῶ ὁ δυστυχισμένος.»
Εἶπ' ὁ Θεόκριτος· «Αὐτὰ τὰ λόγια
ἀνάρμοστα καὶ βλασφημίες εἶναι.
Κι ἀν εἶσαι στὸ σκαλί τὸ πρῶτο, πρέπει
νά'σαι υπερήφανος κ' εὐτυχισμένος.
Εἰδῶ που ἐφθάσες, λίγο δεν εἶναι·
τόσο που ἕκαμες, μεγάλη δόξα.
Κι αὐτὸ ἀκόμη τὸ σκαλί τὸ πρῶτο
πολὺ ἀπὸ τὸν κοινὸ τὸν κόσμον ἀπέχει.
Εἰς τὸ σκαλί γιὰ νὰ πατήσεις τούτο
πρέπει με τὸ δικαίωμα σου νάσαι
πολίτης εἰς τῶν ιδεῶν τὴν πόλιν.
Καὶ δύσκολο στὴν πόλιν ἐκείνην εἶναι
καὶ σπάνιο νὰ σε πολιτογραφήσουν.
Στὴν ἀγορὰ τῆς βρίσκεις Νομοθέτας
που δεν γελᾶ κανένας τυχοδιώκτης.
Εἰδῶ που ἐφθάσες, λίγο δεν εἶναι·
τόσο που ἕκαμες, μεγάλη δόξα.»

[πηγή: Κ.Π. Καβάφης, *Τα Ποιήματα* (1897-1918), τόμ. Α', επιμ. Γ.Π. Σαββίδης, Ἴκαρος, Ἀθήνα 1995 (4η ἐκδ.), σ. 105]

Πνευματικά δικαιώματα

Copyright © **Αντώνιος Παραστατίδης 2024**

Με επιφύλαξη παντός δικαιώματος. Allrightsreserved.

Η έγκριση της διατριβής από το Πανεπιστημίου Νεάπολις δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Πανεπιστημίου.

Όνοματεπώνυμο Φοιτητή/Φοιτήτριας: Αντώνιος Παραστατίδης

Τίτλος Μεταπτυχιακής Διατριβής: Εφαρμογή και σημασία της τεχνητής νοημοσύνης στις διαδικασίες εντοπισμού και διερεύνηση της απάτης

Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση εξ αποστάσεως μεταπτυχιακού τίτλου στο Πανεπιστήμιο Νεάπολις και εγκρίθηκε στις 08.09.2023 από τα μέλη της Εξεταστικής Επιτροπής.

Εξεταστική Επιτροπή:

Πρώτος επιβλέπων: Κωνσταντίνος Αναστάσιος, Επίκουρος Καθηγητής

Μέλος Εξεταστικής Επιτροπής: Νικόλαος Σαριαννίδης, Καθηγητής

Μέλος Εξεταστικής Επιτροπής: Ανδρονίκη Καταραχιά, Καθηγήτρια

Η ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

Εγώ ο Παραστατίδης Αντώνιος γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο **«Εφαρμογή και σημασία της τεχνητής νοημοσύνης στις διαδικασίες εντοπισμού και διερεύνηση της απάτης»**, αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει, έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές παραπομπές και αναφορές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Ο Δηλών

Αντώνιος Παραστατίδης

Περίληψη

Η παρούσα εργασία διερευνά την εφαρμογή και τη σημασία της τεχνητής νοημοσύνης (TN) στο πλαίσιο των διαδικασιών ανίχνευσης και διερεύνησης της απάτης. Οι παραδοσιακές μέθοδοι ανίχνευσης απάτης, που χαρακτηρίζονται από χειροκίνητες διαδικασίες, συστήματα βασισμένα σε κανόνες και στατιστικές μεθόδους, εξετάζονται διεξοδικά για τους περιορισμούς τους, γεγονός που καθιστά αναγκαία μια βαθύτερη διερεύνηση των δυνατοτήτων της TN. Οι θεμελιώδεις έννοιες της TN, συμπεριλαμβανομένης της μηχανικής μάθησης, της βαθιάς μάθησης και της επεξεργασίας φυσικής γλώσσας, διασαφηνίζονται για να παρέχουν ένα ολοκληρωμένο υπόβαθρο για τη μετέπειτα εξέταση.

Ειδικές τεχνικές TN για την ανίχνευση απάτης, που περιλαμβάνουν αλγορίθμους μηχανικής μάθησης, μάθηση χωρίς επίβλεψη για την ανίχνευση ανωμαλιών, εφαρμογές βαθιάς μάθησης, νευρωνικά δίκτυα και αυτοκωδικοποιητές, διερευνώνται λεπτομερώς. Εξετάζονται πραγματικές εφαρμογές της TN στην ανίχνευση απάτης για να φωτιστούν οι τρέχουσες πρακτικές του κλάδου. Η εργασία αξιολογεί τα οφέλη και

τους κινδύνους που συνδέονται με την ΤΝ στην ανίχνευση απάτης, δίνοντας έμφαση στις δυνατότητες ανίχνευσης σε πραγματικό χρόνο και συνεχούς βελτίωσης, ενώ αναγνωρίζει προκλήσεις όπως η κοινωνική απάτη, η αλγοριθμική προκατάληψη και τα ψευδώς θετικά αποτελέσματα.

Τέλος, αναλύονται οι δεοντολογικοί προβληματισμοί και οι προκλήσεις κατά την εφαρμογή της ΤΝ για την ανίχνευση απάτης, με έμφαση στις ανησυχίες για την προστασία της ιδιωτικής ζωής και των δεδομένων. Οι προσπάθειες για την αντιμετώπιση αυτών των ηθικών προκλήσεων, συμπεριλαμβανομένης της προώθησης μιας επαγγελματικής νοοτροπίας εντός της κοινότητας της ΤΝ, συζητούνται κριτικά. Η εργασία εξετάζει επίσης τη μελλοντική πορεία της ηθικής της ΤΝ, προτρέποντας για συνεχή έλεγχο, ρυθμιστικά πλαίσια και διαφάνεια.

Λέξεις κλειδιά: Τεχνητή νοημοσύνη, ανίχνευση απάτης, διαδικασίες διερεύνησης, εφαρμογή της τεχνητής νοημοσύνης, σημασία στην πρόληψη της απάτης

Abstract

This paper explores the application and importance of artificial intelligence (AI) in the context of fraud detection and investigation processes. Traditional fraud detection methods, characterised by manual processes, rule-based systems and statistical methods, are thoroughly examined for their limitations, necessitating a deeper exploration of the potential of AI. The fundamental concepts of AI, including machine learning, deep learning and natural language processing, are clarified to provide a comprehensive background for subsequent consideration.

Specific AI techniques for fraud detection, including machine learning algorithms, unsupervised learning for anomaly detection, deep learning applications, neural networks and auto-encoders, are explored in detail. Real-world applications of AI in fraud detection are examined to shed light on current practices in the field. The paper evaluates the benefits and risks associated with AI in fraud detection, emphasizing the

potential for real-time detection and continuous improvement, while acknowledging challenges such as social fraud, algorithmic bias, and false positives.

Finally, ethical considerations and challenges in applying AI for fraud detection are discussed, with a focus on privacy and data protection concerns. Efforts to address these ethical challenges, including promoting a professional mindset within the AI community, are critically discussed. The paper also examines the future path of AI ethics, urging for continued scrutiny, regulatory frameworks and transparency.

Key words: Artificial intelligence, fraud detection, investigation procedures, application of artificial intelligence, importance in fraud prevention

Περιεχόμενα

Περίληψη.....	7
Abstract	9
Εισαγωγή	13
Κεφάλαιο 1: Παραδοσιακές μέθοδοι ανίχνευσης της απάτης	20
1.1 Χειροκίνητες διαδικασίες και οι περιορισμοί τους	20
1.2 Συστήματα βασισμένα σε κανόνες και οι αδυναμίες τους	22
1.3 Στατιστικές μέθοδοι και η αποτελεσματικότητά τους	24
Κεφάλαιο 2: Εισαγωγή στην τεχνητή νοημοσύνη (TN).....	27
2.1 Ορισμός και επεξήγηση της TN	27
2.2 Τύποι TN	29
2.2.1 Μηχανική μάθηση	29
2.2.2 Βαθιά μάθηση	31
2.2.3 Επεξεργασία φυσικής γλώσσας	32
Κεφάλαιο 3: Τεχνικές τεχνητής νοημοσύνης για την ανίχνευση απάτης	33
3.1 Αλγόριθμοι μηχανικής μάθησης στην ανίχνευση απάτης	33
3.2 Μάθηση χωρίς επίβλεψη για την ανίχνευση ανωμαλιών	34
3.3 Βαθιά μάθηση για την ανίχνευση απάτης	35
3.4 Κοινά σύνολα δεδομένων ανίχνευσης απάτης	38
3.5 Νευρωνικά δίκτυα και οι εφαρμογές τους.....	39
3.6 Αυτοκωδικοποιητές για την ανίχνευση ανωμαλιών	41
Κεφάλαιο 4: Πραγματικές εφαρμογές της TN στην ανίχνευση απάτης	45

Κεφάλαιο 5: Οφέλη και ρίσκα της ΤΝ στην ανίχνευση απάτης	48
5.1 Οφέλη	48
5.2 Ρίσκα	49
Κεφάλαιο 6: Δεοντολογικές εκτιμήσεις και προκλήσεις.....	51
6.1 Προβληματισμοί σχετικά με την προστασία της ιδιωτικής ζωής και των δεδομένων..	51
6.2 Προσπάθειες για την αντιμετώπιση των ηθικών προκλήσεων της Τεχνητής Νοημοσύνης	53
6.3 Προώθηση επαγγελματικής νοοτροπίας	54
6.4 Το μέλλον της ηθικής της Τεχνητής Νοημοσύνης	55
Συμπεράσματα	57

Εισαγωγή

Το πρώτο δίλημμα στην ανίχνευση της απάτης είναι ο προσδιορισμός του τι ακριβώς συνιστά απάτη. Το Webster's Ninth New Collegiate Dictionary (1990, 490) ορίζει την απάτη ως "εσκεμμένη διαστρέβλωση της αλήθειας με σκοπό να παρακινήσει κάποιον άλλο να αποχωριστεί κάτι αξίας ή να παραιτηθεί από ένα νόμιμο δικαίωμα" ή ως "πράξη εξαπάτησης ή παραποίησης". Το New World Dictionary του Webster αναφέρει ότι "η απάτη είναι ένας γενικός όρος και περιλαμβάνει όλα τα πολυποίκιλα μέσα που μπορεί να επινοήσει η ανθρώπινη εφευρετικότητα και στα οποία καταφεύγει ένα άτομο για να αποκτήσει πλεονέκτημα έναντι ενός άλλου με ψευδείς παραστάσεις. Κανένας συγκεκριμένος και αναλλοίωτος κανόνας δεν μπορεί να τεθεί ως γενική πρόταση για τον ορισμό της απάτης, καθώς περιλαμβάνει τον αιφνιδιασμό, το τέχνασμα, την πονηριά και τους αθέμιτους τρόπους με τους οποίους εξαπατάται

κάποιος άλλος. Τα μόνα όρια που την ορίζουν είναι εκείνα που περιορίζουν την ανθρώπινη δολιότητα" (Albrecht, 2003, 6). Ο Albrecht, (2003, 6) συνεχίζει λέγοντας ότι "η απάτη είναι μια εξαπάτηση που περιλαμβάνει τα ακόλουθα στοιχεία: μια παράσταση σχετικά με ένα ουσιώδες σημείο η οποία είναι ψευδής, και μάλιστα σκόπιμα ή απερίσκεπτα- η οποία γίνεται πιστευτή και ακολουθείται από το θύμα, προς ζημία του θύματος". "Σύμφωνα με το νομικό λεξικό Black's Law Dictionary, η απάτη είναι μια υπολογισμένη, παραπλανητική πράξη που έχει ως αποτέλεσμα τη ζημία κάποιου άλλου... Όσοι ασχολούνται με την επιβολή του νόμου βλέπουν την απάτη αυστηρά από την άποψη του κατά πόσον παραβιάζονται νόμοι. Οι ψυχολόγοι αναφέρουν την απάτη σύμφωνα με τις προθέσεις και τα κίνητρα των ατόμων- δεν τους ενδιαφέρουν τόσο τα αντικειμενικά κριτήρια για τον ορισμό των δόλιων ενεργειών" (Kolman, 1999, 88). Ο Wells, (2005, 8) δηλώνει ότι κάθε έγκλημα που χρησιμοποιεί την εξαπάτηση με σκοπό το κέρδος συνιστά απάτη. "Σύμφωνα με το κοινό δίκαιο υπάρχουν τέσσερα γενικά στοιχεία που πρέπει να συντρέχουν για να υπάρχει απάτη: (1) μια ουσιώδης ψευδής δήλωση, (2) γνώση ότι η δήλωση ήταν ψευδής όταν εκφωνήθηκε, (3) εμπιστοσύνη του θύματος στην εν λόγω ψευδή δήλωση, (4) ζημία που προκύπτει από την εμπιστοσύνη του θύματος στην ψευδή δήλωση" (Wells 2005, 8).

Οκτώ από τους δέκα ορισμούς κάνουν αναφορά στη νομική έννοια της απάτης ή της διάπραξης παράνομων πράξεων. Η νομική πτυχή της απάτης δεν προσδιορίζεται, αλλά απλώς αναφέρεται στον ορισμό της απάτης. Η νομική έννοια της απάτης απαιτεί την ερμηνεία του νόμου από ένα μέρος με νομική εμπειρογνομosύνη για να αποφασίσει εάν έχει συμβεί απάτη. Τόσο το AICPA όσο και το ISACA δηλώνουν ρητά ότι ο ελεγκτής δεν είναι υπεύθυνος για τον τελικό νομικό προσδιορισμό της απάτης/παράνομων πράξεων. Σύμφωνα με το SAS No. 99. ο ελεγκτής ασχολείται μόνο με "πράξεις που οδηγούν σε ουσιώδη παραποίηση των οικονομικών καταστάσεων". "Ο ελεγκτής IS θα ασχολείται συνήθως με ύποπτες, παρά με αποδειγμένες, απάτες ή άλλες παράνομες πράξεις" (ISACA, 2004. 20). Τέσσερις από τους δέκα οργανισμούς κατηγοριοποιούν την απάτη σε ουσιώδεις ανακρίβειες που προκύπτουν από τη χρηματοοικονομική πληροφόρηση και σε εκείνες που προκύπτουν από υπεξαίρεση περιουσιακών στοιχείων. Το SAS No. 99 αναφέρει ότι "οι ανακρίβειες που προκύπτουν από δόλια χρηματοοικονομική αναφορά είναι σκόπιμες ανακρίβειες ή παραλείψεις ποσών ή γνωστοποιήσεων στις

χρηματοοικονομικές καταστάσεις που αποσκοπούν στην εξαπάτηση των χρηστών των χρηματοοικονομικών καταστάσεων". Αυτό μπορεί να περιλαμβάνει παραποίηση των λογιστικών αρχείων και/ή των δικαιολογητικών εγγράφων, παράλειψη σημαντικών πληροφοριών ή εσφαλμένη εφαρμογή λογιστικών αρχών (SASNo. 99). Οι ανακρίβειες από υπεξαίρεση περιουσιακών στοιχείων προκύπτουν όταν η κλοπή περιουσιακών στοιχείων προκαλεί ουσιώδεις ανακρίβειες στις οικονομικές καταστάσεις (SASNo. 99).

Η απάτη έχει γίνει ένα διάχυτο ζήτημα στο σημερινό πολύπλοκο επιχειρηματικό περιβάλλον και οι οργανισμοί πρέπει να θέτουν ως προτεραιότητα την εφαρμογή ισχυρών διαδικασιών ανίχνευσης και διερεύνησης της απάτης. Στο πλαίσιο των επιχειρήσεων, η απάτη μπορεί να λάβει διάφορες μορφές, όπως οικονομικές ανακρίβειες, υπεξαίρεση, δωροδοκία και άλλες παραπλανητικές πρακτικές που θέτουν σε κίνδυνο την ακεραιότητα των οργανισμών. Κατά συνέπεια, η εφαρμογή αποτελεσματικών διαδικασιών ανίχνευσης και διερεύνησης της απάτης είναι ζωτικής σημασίας για τη διασφάλιση της οικονομικής υγείας και της φήμης των επιχειρήσεων (Huangetal., 2017).

Ένας από τους πρωταρχικούς λόγους για την ιεράρχηση των διαδικασιών ανίχνευσης και διερεύνησης της απάτης είναι ο οικονομικός αντίκτυπος των δόλιων δραστηριοτήτων. Οι δόλιες δραστηριότητες μπορούν να οδηγήσουν σε σημαντικές οικονομικές απώλειες για τους οργανισμούς, επηρεάζοντας το τελικό τους αποτέλεσμα και θέτοντας σε κίνδυνο τη μακροπρόθεσμη βιωσιμότητά τους. Σύμφωνα με την Association of Certified Fraud Examiners (ACFE), ο τυπικός οργανισμός χάνει περίπου το 5% των ετήσιων εσόδων του λόγω απάτης ("Report to the Nations", 2020). Με την καθιέρωση και τη διατήρηση ισχυρών μηχανισμών ανίχνευσης απάτης, οι οργανισμοί μπορούν να μετριάσουν τους οικονομικούς κινδύνους και να προστατεύσουν τα περιουσιακά τους στοιχεία από τις επιζήμιες συνέπειες των δόλιων δραστηριοτήτων.

Επιπλέον, η βλάβη της φήμης που προκαλείται από την απάτη μπορεί να είναι σοβαρή και διαρκής. Οι δόλιες δραστηριότητες διαβρώνουν την εμπιστοσύνη των ενδιαφερομένων μερών, συμπεριλαμβανομένων των πελατών, των επενδυτών και των εργαζομένων. Μόλις η φήμη μιας εταιρείας αμαυρωθεί από ισχυρισμούς περί απάτης, μπορεί να χρειαστούν χρόνια για να αποκατασταθεί η εμπιστοσύνη και να ανακτηθεί

η αξιοπιστία της στην αγορά. Μια κατεστραμμένη φήμη μπορεί να οδηγήσει σε απώλεια πελατών, μειωμένη εμπιστοσύνη των μετόχων και προκλήσεις στην προσέλκυση και διατήρηση κορυφαίων ταλέντων. Η εφαρμογή αποτελεσματικών διαδικασιών ανίχνευσης και διερεύνησης της απάτης είναι απαραίτητη για τον άμεσο εντοπισμό και την αντιμετώπιση των δόλιων δραστηριοτήτων, την ελαχιστοποίηση της ζημίας της φήμης και την επίδειξη της δέσμευσης για ηθικές επιχειρηματικές πρακτικές (Coram, Ferguson, & Moroney, 2006).

Οι διαδικασίες ανίχνευσης και διερεύνησης της απάτης είναι επίσης καθοριστικές για τη διασφάλιση της νομικής συμμόρφωσης. Οι ρυθμιστικοί φορείς και οι κυβερνητικές υπηρεσίες έχουν θεσπίσει αυστηρούς νόμους και κανονισμούς για την καταπολέμηση της απάτης και των οικονομικών εγκλημάτων. Οι οργανισμοί που δεν εφαρμόζουν επαρκή μέτρα πρόληψης της απάτης ενδέχεται να αντιμετωπίσουν νομικές συνέπειες, συμπεριλαμβανομένων προστίμων, ποινών και νομικών διαδικασιών. Με την προληπτική εφαρμογή ισχυρών διαδικασιών ανίχνευσης και διερεύνησης απάτης, οι οργανισμοί μπορούν να αποδείξουν τη συμμόρφωση με τους σχετικούς νόμους και κανονισμούς, μειώνοντας τον κίνδυνο νομικών ενεργειών και συναφών οικονομικών κυρώσεων (Clements, & Knudstrup, 2016).

Επιπλέον, οι διαδικασίες ανίχνευσης και διερεύνησης της απάτης συμβάλλουν στη διαμόρφωση κουλτούρας λογοδοσίας και διαφάνειας σε έναν οργανισμό. Όταν οι εργαζόμενοι γνωρίζουν την ύπαρξη μέτρων καταπολέμησης της απάτης και τις συνέπειες των δόλιων δραστηριοτήτων, είναι πιο πιθανό να τηρούν τα πρότυπα δεοντολογίας και να αναφέρουν ύποπτη συμπεριφορά. Η καθιέρωση μιας κουλτούρας που εκτιμά την ακεραιότητα και τη δεοντολογική συμπεριφορά είναι απαραίτητη για την πρόληψη της απάτης και την προώθηση ενός θετικού εργασιακού περιβάλλοντος (Huang et al., 2017).

Εν κατακλείδι, η σημασία των διαδικασιών ανίχνευσης και διερεύνησης της απάτης δεν μπορεί να υπερεκτιμηθεί στο σημερινό επιχειρηματικό τοπίο. Οι οργανισμοί πρέπει να αναγνωρίζουν τους οικονομικούς, φημιστικούς και νομικούς κινδύνους που συνδέονται με την απάτη και να λαμβάνουν προληπτικά μέτρα για τη διασφάλιση των συμφερόντων τους. Με την εφαρμογή αποτελεσματικών μηχανισμών ανίχνευσης απάτης και διαδικασιών διερεύνησης, οι επιχειρήσεις μπορούν να προστατεύσουν τα περιουσιακά τους στοιχεία, να διατηρήσουν τη φήμη τους, να διασφαλίσουν τη

νομική συμμόρφωση και να προωθήσουν μια κουλτούρα ακεραιότητας και υπευθυνότητας. Καθώς οι οργανισμοί συνεχίζουν να περιηγούνται στις πολυπλοκότητες του επιχειρηματικού κόσμου, η ιεράρχηση της ανίχνευσης και της διερεύνησης της απάτης παραμένει θεμελιώδης πτυχή της διαχείρισης κινδύνων και της εταιρικής διακυβέρνησης.

Σε μια εποχή που χαρακτηρίζεται από ραγδαίες τεχνολογικές εξελίξεις, το τοπίο της πρόληψης της απάτης έχει υποστεί σημαντικό μετασχηματισμό. Καθώς οι δόλιες δραστηριότητες γίνονται ολοένα και πιο εξελιγμένες, ο ρόλος της τεχνολογίας στην ενίσχυση της πρόληψης της απάτης έχει καταστεί υψίστης σημασίας. Τα τεχνολογικά εργαλεία και οι καινοτομίες προσφέρουν στους οργανισμούς τα μέσα για τον προληπτικό εντοπισμό, τον μετριασμό και την πρόληψη των δόλιων δραστηριοτήτων, παρέχοντας μια κρίσιμη γραμμή άμυνας κατά των οικονομικών απωλειών και της ζημίας της φήμης (Bhasin, 2016).

Μία από τις βασικές συνεισφορές της τεχνολογίας στην πρόληψη της απάτης είναι η ανάπτυξη και η εφαρμογή προηγμένων αναλυτικών δεδομένων. Η ανάλυση τεράστιων συνόλων δεδομένων σε πραγματικό χρόνο επιτρέπει στους οργανισμούς να εντοπίζουν μοτίβα, ανωμαλίες και πιθανούς δείκτες δόλιας συμπεριφοράς. Οι αλγόριθμοι μηχανικής μάθησης, ένα υποσύνολο της τεχνητής νοημοσύνης, επιτρέπουν στα συστήματα να εξελίσσονται συνεχώς και να προσαρμόζονται σε νέα συστήματα απάτης. Αυτές οι τεχνολογίες δίνουν τη δυνατότητα στους οργανισμούς να προχωρήσουν πέρα από τα παραδοσιακά συστήματα που βασίζονται σε κανόνες και να εντοπίζουν λεπτά, προηγουμένως μη ανιχνεύσιμα μοτίβα που σχετίζονται με δόλιες δραστηριότητες (Hassanetal., 2023).

Επιπλέον, η τεχνητή νοημοσύνη (AI) και η μηχανική μάθηση διαδραματίζουν καθοριστικό ρόλο στην προγνωστική μοντελοποίηση. Αναλύοντας ιστορικά δεδομένα, οι τεχνολογίες αυτές μπορούν να προβλέψουν πιθανούς κινδύνους και μοτίβα απάτης, επιτρέποντας στους οργανισμούς να εφαρμόζουν προληπτικά μέτρα πριν από την εμφάνιση δόλιων δραστηριοτήτων. Η προγνωστική μοντελοποίηση όχι μόνο ενισχύει την αποτελεσματικότητα της πρόληψης της απάτης, αλλά και ελαχιστοποιεί τα ψευδώς θετικά αποτελέσματα, επιτρέποντας στους οργανισμούς να εστιάσουν τους πόρους τους στις πιο κρίσιμες απειλές (Kumar, 2022).

Η τεχνολογία blockchain είναι μια άλλη επαναστατική εξέλιξη που έχει τη δυνατότητα να μεταμορφώσει την πρόληψη της απάτης. Η αποκεντρωμένη και ανθεκτική στην παραποίηση φύση της αλυσίδας μπλοκ την καθιστά ιδανικό εργαλείο για τη διασφάλιση της ακεραιότητας των συναλλαγών και των δεδομένων. Στις χρηματοοικονομικές συναλλαγές, για παράδειγμα, το blockchain μπορεί να παρέχει ένα διαφανές και αμετάβλητο βιβλίο, μειώνοντας τον κίνδυνο δόλιων δραστηριοτήτων, όπως η παραποίηση αρχείων ή η διπλή δαπάνη. Η διαφάνεια και η ιχνηλασιμότητα που προσφέρει το blockchain συμβάλλουν σε ένα πιο ασφαλές και ανθεκτικό στην απάτη περιβάλλον (Kumar, 2022).

Ο βιομετρικός έλεγχος ταυτότητας είναι μια ακόμη τεχνολογική πρόοδος που ενισχύει τις προσπάθειες πρόληψης της απάτης. Οι παραδοσιακές μέθοδοι ελέγχου ταυτότητας, όπως οι κωδικοί πρόσβασης και οι κωδικοί PIN, είναι ευάλωτες σε παραβιάσεις με διάφορα μέσα, όπως το phishing και η πειρατεία. Οι βιομετρικές μέθοδοι ελέγχου ταυτότητας, όπως η αναγνώριση δακτυλικών αποτυπωμάτων, η αναγνώριση προσώπου και η σάρωση ίριδας, παρέχουν έναν πιο ασφαλή και αξιόπιστο τρόπο επαλήθευσης της ταυτότητας των ατόμων. Με την εφαρμογή βιομετρικού ελέγχου ταυτότητας, οι οργανισμοί μπορούν να μειώσουν σημαντικά τον κίνδυνο μη εξουσιοδοτημένης πρόσβασης και κλοπής ταυτότητας (Aziz, &Andriansyah, 2023).

Επιπλέον, ο ρόλος της τεχνολογίας στην ενίσχυση της πρόληψης της απάτης επεκτείνεται και στον τομέα της ασφάλειας στον κυβερνοχώρο. Καθώς οι απατεώνες εξελίσσουν συνεχώς τις τακτικές τους, οι οργανισμοί πρέπει να ενισχύουν τις άμυνες τους στον κυβερνοχώρο. Τα προηγμένα εργαλεία κυβερνοασφάλειας, συμπεριλαμβανομένων των συστημάτων ανίχνευσης εισβολών, των τειχών προστασίας και των πρωτοκόλλων κρυπτογράφησης, συμβάλλουν στην προστασία των ευαίσθητων πληροφοριών και στην αποτροπή μη εξουσιοδοτημένης πρόσβασης. Η συνεχής παρακολούθηση των δραστηριοτήτων του δικτύου και η τροφοδοσία πληροφοριών για απειλές σε πραγματικό χρόνο συμβάλλουν στον έγκαιρο εντοπισμό πιθανών παραβιάσεων της ασφάλειας, επιτρέποντας στους οργανισμούς να αντιδράσουν άμεσα και να αποτρέψουν την απάτη (Hassanetal., 2023).

Συμπερασματικά, ο ρόλος της τεχνολογίας στην ενίσχυση της πρόληψης της απάτης είναι απαραίτητος στο σύγχρονο επιχειρηματικό τοπίο. Η ενσωμάτωση προηγμένων

αναλύσεων δεδομένων, τεχνητής νοημοσύνης, blockchain, βιομετρικού ελέγχου ταυτότητας και μέτρων κυβερνοασφάλειας παρέχει στους οργανισμούς μια ολοκληρωμένη εργαλειοθήκη για τον εντοπισμό και την πρόληψη δόλιων δραστηριοτήτων. Καθώς η τεχνολογία συνεχίζει να εξελίσσεται, το ίδιο θα συμβεί και με τις δυνατότητες των μηχανισμών πρόληψης της απάτης. Οι οργανισμοί που αγκαλιάζουν και αξιοποιούν αυτές τις τεχνολογικές καινοτομίες είναι σε καλύτερη θέση να διαφυλάξουν τα περιουσιακά τους στοιχεία, να προστατεύσουν τη φήμη τους και να διατηρήσουν την εμπιστοσύνη των ενδιαφερομένων σε ένα περιβάλλον όπου η απειλή της απάτης είναι διαρκώς παρούσα.

Κεφάλαιο 1: Παραδοσιακές μέθοδοι ανίχνευσης της απάτης

1.1 Χειροκίνητες διαδικασίες και οι περιορισμοί τους

Στο διαρκώς εξελισσόμενο τοπίο των χρηματοπιστωτικών συναλλαγών, η ανίχνευση της απάτης έχει καταστεί κρίσιμη πτυχή της διατήρησης της ακεραιότητας και της ασφάλειας των διαφόρων βιομηχανιών. Οι παραδοσιακές μέθοδοι ανίχνευσης της απάτης βασίζονται συχνά σε χειροκίνητες διαδικασίες, όπου η ανθρώπινη παρέμβαση είναι το κλειδί για τον εντοπισμό παρατυπιών και πιθανών περιπτώσεων εξαπάτησης. Ενώ αυτές οι μέθοδοι υπήρξαν θεμελιώδεις για την καταπολέμηση της απάτης, έχουν εγγενείς περιορισμούς που εμποδίζουν την αποτελεσματικότητά τους στη σύγχρονη, ταχέως εξελισσόμενη ψηφιακή εποχή (Phuaetal., 2010).

Οι χειροκίνητες διαδικασίες για την ανίχνευση απάτης περιλαμβάνουν συνήθως τον ανθρώπινο έλεγχο των οικονομικών εγγράφων, των συναλλαγών και των μοτίβων. Αυτή η προσέγγιση βασίζεται στην εμπειρία, τη διαίσθηση και το οξύ μάτι για ανωμαλίες. Μια συνήθης χειροκίνητη μέθοδος είναι η σχολαστική εξέταση των οικονομικών καταστάσεων και των αρχείων συναλλαγών. Οι ειδικοί αναλύουν τα δεδομένα για ασυνέπειες, ασυνήθιστα μοτίβα ή ύποπτες δραστηριότητες που μπορεί να υποδηλώνουν δόλια συμπεριφορά (Bolton, &Hand, 2002).

Ένας από τους κύριους περιορισμούς των χειροκίνητων διαδικασιών είναι ο χρονοβόρος χαρακτήρας τους. Οι άνθρωποι μπορούν να επεξεργαστούν μόνο έναν περιορισμένο όγκο πληροφοριών μέσα σε ένα δεδομένο χρονικό διάστημα. Καθώς αυξάνεται ο όγκος και η πολυπλοκότητα των χρηματοοικονομικών συναλλαγών, η

ικανότητα των χειροκίνητων μεθόδων να συμβαδίζουν μειώνεται. Οι απατεώνες εκμεταλλεύονται αυτόν τον περιορισμό με τη διεξαγωγή εξελιγμένων συστημάτων που μπορούν να παραμείνουν απαρατήρητα για μεγάλα χρονικά διαστήματα (Phuaetal., 2010).

Επιπλέον, η εξάρτηση από την ανθρώπινη διαίσθηση εισάγει το ενδεχόμενο υποκειμενικών προκαταλήψεων. Οι αναλυτές απάτης μπορεί να ερμηνεύουν διαφορετικά το ίδιο σύνολο δεδομένων, με αποτέλεσμα να υπάρχουν ασυνέπειες στον εντοπισμό των δόλιων δραστηριοτήτων. Επιπλέον, οι χειροκίνητες διαδικασίες είναι επιρρεπείς στην εποπτεία και την κόπωση, καθώς η επαναλαμβανόμενη φύση των σχετικών εργασιών μπορεί να οδηγήσει σε κενά προσοχής και συγκέντρωσης (Fawcett, &Provost, 1997).

Ένα άλλο σημαντικό μειονέκτημα της χειροκίνητης ανίχνευσης απάτης είναι ο αντιδραστικός της χαρακτήρας. Οι αναλυτές συχνά εντοπίζουν την απάτη αφού έχει συμβεί, βασιζόμενοι σε ιστορικά δεδομένα για να αναγνωρίσουν μοτίβα και ανωμαλίες. Αυτή η αντιδραστική προσέγγιση καθιστά δύσκολη την πρόληψη της απάτης σε πραγματικό χρόνο, επιτρέποντας στους εγκληματίες να εκμεταλλεύονται τα τρωτά σημεία πριν αυτά ανακαλυφθούν. Σε έναν κόσμο όπου οι απειλές στον κυβερνοχώρο και η οικονομική απάτη εξελίσσονται με ταχείς ρυθμούς, αυτή η καθυστέρηση μπορεί να έχει σοβαρές συνέπειες. Η επεκτασιμότητα των χειροκίνητων διαδικασιών αποτελεί επίσης κρίσιμη ανησυχία. Καθώς οι επιχειρήσεις αναπτύσσονται και ο όγκος των συναλλαγών αυξάνεται, η απαίτηση για αποτελεσματικούς και επεκτάσιμους μηχανισμούς ανίχνευσης απάτης καθίσταται επιτακτική. Οι χειροκίνητες μέθοδοι δυσκολεύονται να προσαρμοστούν στη δυναμική φύση των σύγχρονων χρηματοοικονομικών συστημάτων, γεγονός που τις καθιστά ακατάλληλες για οργανισμούς που διαχειρίζονται καθημερινά μεγάλο όγκο συναλλαγών (Phuaetal., 2010).

Ο χειροκίνητος χαρακτήρας των παραδοσιακών μεθόδων ανίχνευσης απάτης περιορίζει επίσης την ικανότητά τους να ενσωματώνονται απρόσκοπτα με προηγμένες τεχνολογίες. Η μηχανική μάθηση, η τεχνητή νοημοσύνη και η ανάλυση δεδομένων έχουν αναδειχθεί ως ισχυρά εργαλεία στην ανίχνευση απάτης, προσφέροντας τη δυνατότητα ανάλυσης τεράστιων συνόλων δεδομένων σε πραγματικό χρόνο και εντοπισμού μοτίβων που μπορεί να διαφεύγουν της

ανθρώπινης παρατήρησης. Οι χειροκίνητες διαδικασίες, αντίθετα, δεν διαθέτουν την ευελιξία και την ακρίβεια που παρέχουν αυτές οι τεχνολογίες (West, & Bhattacharya, 2016).

Εν κατακλείδι, ενώ οι παραδοσιακές μέθοδοι εντοπισμού απάτης μέσω χειροκίνητων διαδικασιών ήταν χρήσιμες στο παρελθόν, αντιμετωπίζουν σημαντικούς περιορισμούς στο σύγχρονο χρηματοπιστωτικό τοπίο. Ο αργός και αντιδραστικός χαρακτήρας, η ευαισθησία σε προκαταλήψεις και σφάλματα και η αδυναμία αποτελεσματικής κλιμάκωσης είναι μεταξύ των κύριων μειονεκτημάτων. Καθώς η τεχνολογία συνεχίζει να εξελίσσεται, υπάρχει αυξανόμενη ανάγκη για τους οργανισμούς να ενσωματώσουν αυτοματοποιημένες και έξυπνες λύσεις ανίχνευσης απάτης για να παραμείνουν μπροστά από το διαρκώς εξελισσόμενο τοπίο απειλών.

1.2 Συστήματα βασισμένα σε κανόνες και οι αδυναμίες τους

Τα συστήματα που βασίζονται σε κανόνες χρησιμοποιούνται εδώ και καιρό ως μέθοδος για την ανίχνευση και την πρόληψη της απάτης σε διάφορους κλάδους. Τα συστήματα αυτά βασίζονται σε προκαθορισμένα σύνολα κανόνων και λογικής για τον εντοπισμό δυνητικά δόλιων δραστηριοτήτων βάσει συγκεκριμένων κριτηρίων. Ενώ προσφέρουν πλεονεκτήματα όσον αφορά τη διαφάνεια και την απλότητα, τα συστήματα που βασίζονται σε κανόνες παρουσιάζουν επίσης αξιοσημείωτες αδυναμίες που μπορούν να περιορίσουν την αποτελεσματικότητά τους στην αντιμετώπιση της πολυπλοκότητας των σύγχρονων συστημάτων απάτης (Hajek, 2019).

Μία από τις θεμελιώδεις αδυναμίες των συστημάτων που βασίζονται σε κανόνες είναι ο στατικός και άκαμπος χαρακτήρας τους. Τα συστήματα αυτά λειτουργούν με βάση προκαθορισμένους κανόνες που συχνά δημιουργούνται από ανθρώπινους εμπειρογνώμονες με βάση ιστορικά δεδομένα και γνωστά μοτίβα απάτης. Κατά συνέπεια, μπορεί να δυσκολεύονται να προσαρμοστούν στις αναδυόμενες και εξελισσόμενες τακτικές απάτης που αποκλίνουν από τα καθιερωμένα πρότυπα. Οι απατεώνες αναπτύσσουν συνεχώς νέες στρατηγικές για να παρακάμπτουν την ανίχνευση βάσει κανόνων, καθιστώντας τα συστήματα αυτά ευάλωτα σε νέες και εξελιγμένες επιθέσεις (Kim, & Vasarhelyi, 2012).

Τα ψευδώς θετικά και τα ψευδώς αρνητικά αποτελέσματα είναι κοινές προκλήσεις που σχετίζονται με τα συστήματα που βασίζονται σε κανόνες. Στην προσπάθειά τους να ρίξουν ένα ευρύ δίκτυο και να εντοπίσουν πιθανή απάτη, τα συστήματα αυτά μπορεί να δημιουργήσουν σημαντικό αριθμό ψευδώς θετικών αποτελεσμάτων - περιπτώσεις όπου νόμιμες συναλλαγές χαρακτηρίζονται ως ύποπτες. Αυτό μπορεί να οδηγήσει σε αυξημένες χειροκίνητες αναθεωρήσεις, λειτουργική αναποτελεσματικότητα και δυσαρέσκεια των πελατών. Αντίθετα, τα συστήματα που βασίζονται σε κανόνες μπορεί να παραβλέπουν τις λεπτές και διαφοροποιημένες δόλιες δραστηριότητες, με αποτέλεσμα να προκύπτουν ψευδώς αρνητικά αποτελέσματα και να μην εντοπίζονται παράνομες συναλλαγές(Hajek, 2019).

Η επεκτασιμότητα των συστημάτων που βασίζονται σε κανόνες είναι ένας άλλος τομέας προβληματισμού. Καθώς αυξάνεται ο όγκος των συναλλαγών, η διαχείριση και η ενημέρωση ενός τεράστιου συνόλου κανόνων καθίσταται δυσχερής εργασία. Οι οργανισμοί ενδέχεται να δυσκολεύονται να συμβαδίσουν με την εξελισσόμενη φύση της απάτης, με αποτέλεσμα να καθυστερεί η ενημέρωση των κανόνων ή να παραβλέπονται κρίσιμες αλλαγές στα πρότυπα απάτης. Αυτή η έλλειψη επεκτασιμότητας μπορεί να εμποδίσει την ικανότητα του συστήματος να προσαρμόζεται στο δυναμικό και ταχέως μεταβαλλόμενο τοπίο των χρηματοπιστωτικών συναλλαγών(Kim, & Vasarhelyi, 2012).

Επιπλέον, τα συστήματα που βασίζονται σε κανόνες μπορεί να πάσχουν από έλλειψη κατανόησης του πλαισίου. Συνήθως αξιολογούν τις συναλλαγές με βάση μεμονωμένους κανόνες χωρίς να λαμβάνουν υπόψη το ευρύτερο πλαίσιο ή τις σχέσεις μεταξύ των σημείων δεδομένων. Αυτός ο περιορισμός καθιστά δύσκολο τον εντοπισμό λεπτών και πολύπλοκων συστημάτων απάτης που περιλαμβάνουν πολλαπλά αλληλένδετα στοιχεία. Οι προηγμένες τακτικές απάτης, όπως η απάτη με συνθετική ταυτότητα ή η συμπαιγνία, μπορεί να περάσουν απαρατήρητες από τα συστήματα που βασίζονται σε κανόνες και λειτουργούν με απομονωμένο τρόπο(Hajek, 2019).

Η συντήρηση των κανόνων αποτελεί συνεχή πρόκληση για τους οργανισμούς που χρησιμοποιούν συστήματα βασισμένα σε κανόνες. Η ανάγκη τακτικής ενημέρωσης και τελειοποίησης των κανόνων απαιτεί σημαντική ανθρώπινη προσπάθεια και εμπειρογνωμοσύνη. Αυτή η διαδικασία μπορεί να είναι εντατική σε πόρους και

χρονοβόρα, με αποτέλεσμα να υπάρχουν πιθανώς καθυστερήσεις στην ανταπόκριση σε αναδυόμενες απειλές. Επιπλέον, ο χειροκίνητος χαρακτήρας της δημιουργίας και της συντήρησης κανόνων εισάγει την πιθανότητα σφαλμάτων, παραλείψεων και προκαταλήψεων, υπονομεύοντας τη συνολική αποτελεσματικότητα του συστήματος (Kim, & Vasarhelyi, 2012).

Συμπερασματικά, ενώ τα συστήματα που βασίζονται σε κανόνες αποτελούν πολύτιμα εργαλεία στο οπλοστάσιο κατά της απάτης, οι αδυναμίες τους θέτουν σημαντικές προκλήσεις στο σημερινό τοπίο. Ο στατικός χαρακτήρας των προκαθορισμένων κανόνων, η ευαισθησία σε ψευδώς θετικά και αρνητικά αποτελέσματα, τα ζητήματα κλιμάκωσης, η έλλειψη κατανόησης του πλαισίου και ο συνεχής φόρτος συντήρησης αποτελούν κρίσιμα μειονεκτήματα. Καθώς οι οργανισμοί προσπαθούν να μείνουν μπροστά από τους ολοένα και πιο εξελιγμένους απατεώνες, αναγνωρίζεται ολοένα και περισσότερο η ανάγκη για πιο προσαρμοστικά και ευφυή συστήματα ανίχνευσης απάτης που αξιοποιούν προηγμένες τεχνολογίες όπως η μηχανική μάθηση και η τεχνητή νοημοσύνη.

1.3 Στατιστικές μέθοδοι και η αποτελεσματικότητά τους

Οι στατιστικές μέθοδοι διαδραματίζουν κρίσιμο ρόλο στην ανίχνευση απάτης, αξιοποιώντας μαθηματικά μοντέλα και αναλύσεις για τον εντοπισμό μοτίβων, ανωμαλιών και αποκλίσεων από την αναμενόμενη συμπεριφορά σε μεγάλα σύνολα δεδομένων. Αυτές οι μέθοδοι προσφέρουν αρκετά πλεονεκτήματα, αλλά συνοδεύονται επίσης από το δικό τους σύνολο προβληματισμών και προκλήσεων. Ένα από τα κύρια πλεονεκτήματα των στατιστικών μεθόδων έγκειται στην ικανότητά τους να αναλύουν γρήγορα και αποτελεσματικά τεράστιες ποσότητες δεδομένων. Με τη δημιουργία βασικών στατιστικών μοντέλων που αντιπροσωπεύουν την κανονική συμπεριφορά, οι μέθοδοι αυτές μπορούν να επισημάνουν συναλλαγές ή δραστηριότητες που αποκλίνουν σημαντικά από τα αναμενόμενα πρότυπα. Οι στατιστικές τεχνικές, όπως η ανάλυση παλινδρόμησης, η ομαδοποίηση και η ανίχνευση ανωμαλιών, επιτρέπουν στους οργανισμούς να εντοπίζουν παρατυπίες που ενδέχεται να υποδηλώνουν δόλια συμπεριφορά (Bolton, & Hand, 2002).

Η προσαρμοστικότητα των στατιστικών μεθόδων είναι ένα άλλο βασικό πλεονέκτημα. Σε αντίθεση με τα άκαμπτα συστήματα που βασίζονται σε κανόνες, τα

στατιστικά μοντέλα μπορούν να εξελίσσονται και να μαθαίνουν από νέα δεδομένα, γεγονός που τα καθιστά πιο ευέλικτα στις αλλαγές των προτύπων απάτης με την πάροδο του χρόνου. Αυτή η προσαρμοστικότητα είναι ιδιαίτερα πολύτιμη στο δυναμικό τοπίο των χρηματοπιστωτικών συναλλαγών, όπου οι απατεώνες επινοούν συνεχώς νέες τακτικές για να παρακάμπτουν τα μέτρα ανίχνευσης (Lietal., 2008).

Μια κοινή στατιστική προσέγγιση στην ανίχνευση απάτης είναι η χρήση αλγορίθμων μηχανικής μάθησης. Αυτοί οι αλγόριθμοι μπορούν να αναλύσουν ιστορικά δεδομένα για να εντοπίσουν σύνθετες σχέσεις και μοτίβα που μπορεί να είναι ενδεικτικά απάτης. Τα μοντέλα μάθησης με επίβλεψη μπορούν να εκπαιδευτούν σε σύνολα δεδομένων με ετικέτες, ενώ τα μοντέλα μάθησης χωρίς επίβλεψη μπορούν να ανακαλύψουν μοτίβα χωρίς προηγούμενη επισήμανση, καθιστώντας τα αποτελεσματικά για την ανίχνευση άγνωστων προηγουμένως συστημάτων απάτης (O'Kelly, 2004).

Ωστόσο, η αποτελεσματικότητα των στατιστικών μεθόδων εξαρτάται από την ποιότητα και την αντιπροσωπευτικότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση. Εάν τα δεδομένα εκπαίδευσης είναι μεροληπτικά ή ελλιπή, τα στατιστικά μοντέλα ενδέχεται να αποτύχουν να συλλάβουν την πραγματική πολυπλοκότητα των υποκείμενων μοτίβων, οδηγώντας σε ανακριβείς προβλέψεις και αυξημένα ψευδώς θετικά ή αρνητικά αποτελέσματα. Μια άλλη πρόκληση είναι η ερμηνευσιμότητα των στατιστικών μοντέλων. Ενώ οι αλγόριθμοι μηχανικής μάθησης μπορούν να αποκαλύψουν περίπλοκα μοτίβα, οι εσωτερικές λειτουργίες αυτών των μοντέλων συχνά θεωρούνται "μαύρα κουτιά". Η κατανόηση του πώς και γιατί λήφθηκε μια συγκεκριμένη απόφαση από τον αλγόριθμο μπορεί να αποτελέσει πρόκληση, γεγονός που μπορεί να δημιουργήσει εμπόδια όσον αφορά την επεξηγηματικότητα και τη συμμόρφωση, ιδίως σε κλάδους με αυστηρές κανονιστικές απαιτήσεις (Lietal., 2008).

Επιπλέον, οι στατιστικές μέθοδοι μπορεί να δυσκολεύονται να εντοπίσουν ανωμαλίες που είναι λεπτές ή εξαρτώμενες από το πλαίσιο. Οι εξελιγμένοι απατεώνες μπορεί να σχεδιάζουν σκόπιμα δραστηριότητες που μοιάζουν με φυσιολογική συμπεριφορά, καθιστώντας δύσκολο για τα στατιστικά μοντέλα να διακρίνουν τις δόλιες συναλλαγές από τις νόμιμες. Η κατανόηση του πλαισίου, η οποία περιλαμβάνει την εξέταση διαφόρων παραγόντων και σχέσεων, παραμένει μια πρόκληση για τις

στατιστικές μεθόδους, οδηγώντας σε πιθανά ψευδώς θετικά αποτελέσματα ή περιπτώσεις απάτης που δεν έχουν εντοπιστεί (Bolton, & Hand, 2002).

Εν κατακλείδι, οι στατιστικές μέθοδοι, ιδίως εκείνες που βασίζονται στη μηχανική μάθηση, έχουν αποδειχθεί αποτελεσματικά εργαλεία για την καταπολέμηση της απάτης. Η προσαρμοστικότητα τους και η ικανότητά τους να επεξεργάζονται μεγάλα σύνολα δεδομένων τις καθιστούν πολύτιμα περιουσιακά στοιχεία για τους οργανισμούς που επιδιώκουν να μείνουν μπροστά από τις εξελισσόμενες τακτικές απάτης. Ωστόσο, προκλήσεις όπως η ποιότητα των δεδομένων, η ερμηνευσιμότητα και η ανίχνευση λεπτών ανωμαλιών υπογραμμίζουν τη σημασία του συνδυασμού των στατιστικών μεθόδων με άλλες προηγμένες τεχνολογίες και προσεγγίσεις για μια ολοκληρωμένη και ισχυρή στρατηγική ανίχνευσης απάτης.

Κεφάλαιο 2: Εισαγωγή στην τεχνητή νοημοσύνη (TN)

2.1 Ορισμός και επεξήγηση της TN

Με την πιο κοινή του έννοια, ένας ορισμός προσδιορίζει το νόημα ή τη σημασία μιας λέξης ή φράσης. Αν και αυτό ακούγεται απλό, υπάρχουν ακόμη λεπτά ζητήματα που πρέπει να παρατηρηθούν στο πλαίσιο του ορισμού της TN. Πρώτον, ένας ορισμός μπορεί να αφορά είτε τη χρήση μιας λέξης είτε το περιεχόμενο μιας έννοιας που εκφράζεται από μια λέξη, και στις περισσότερες περιπτώσεις η συζήτηση σχετικά με την έννοια της TN αφορά περισσότερο το δεύτερο παρά το πρώτο, αν και αναφέρεται άμεσα το πρώτο. Στην τρέχουσα συζήτηση, αν η λέξη "τεχνητή νοημοσύνη" αντικατασταθεί από τη λέξη "νοημοσύνη υπολογιστών" ή "νοημοσύνη μηχανών", τότε το υποκείμενο πρόβλημα δεν θα άλλαζε πολύ. Το ίδιο ισχύει και αν η έννοια δεν εκφράζεται στα αγγλικά, αλλά σε μια άλλη ανθρώπινη γλώσσα. Εξάλλου, το βασικό ζήτημα δεν έγκειται στην επιλογή των λέξεων, αλλά στην ιδέα που εκφράζεται από αυτές, οπότε η συζήτηση αυτή είναι σε μεγάλο βαθμό ανεξάρτητη από τη γλώσσα.

Όταν η έννοια που πρέπει να εκφραστεί γίνεται σχετικά καλά καθορισμένη, το ποιες λέξεις επιλέγονται για να την εκφράσουν εξακολουθεί να είναι ένα μη τετριμμένο πρόβλημα, αλλά είναι δευτερεύον. Καθορίζοντας τις ικανές και αναγκαίες συνθήκες της, ο ορισμός μιας έννοιας χαράζει τα όριά της και, ως εκ τούτου, ρυθμίζει τη χρήση της στη σκέψη και την επικοινωνία. Ωστόσο, ακόμη και με αυτά τα προφανή πλεονεκτήματα, δεν μπορεί κάποιος να περιμένει ότι κάθε έννοια θα είναι εξ αρχής σαφώς ορισμένη, ακόμη και για τις επιστημονικές έννοιες, επειδή κάθε φαινόμενο που μελετά η επιστήμη εννοιολογείται αρχικά ως μια αόριστη ιδέα και οι έννοιες είναι ρευστές από τη φύση τους (Hofstadter και FARG, 1995). Σε γενικές γραμμές, η ύπαρξη ενός σαφούς ορισμού δεν αποτελεί προϋπόθεση για τη χρήση μιας έννοιας τόσο στην επιστημονική έρευνα όσο και στις συζητήσεις, αν και είναι πράγματι άκρως επιθυμητή.

Στις επιστημονικές συζητήσεις, υπάρχουν δύο τύποι ορισμών: ορισμοί λεξικού, οι οποίοι περιγράφουν την υπάρχουσα χρήση, και ορισμοί εργασίας, οι οποίοι προδιαγράφουν την προτεινόμενη χρήση. Ενώ οι λεξικογραφικοί ορισμοί της Τεχνητής Νοημοσύνης (TN) ευθυγραμμίζονται με το τι έχουν κάνει οι ερευνητές TN, οι ορισμοί εργασίας θέτουν ερευνητικούς στόχους για τα έργα TN. Ένας ορισμός εργασίας πρέπει να ισορροπεί, λαμβάνοντας υπόψη τις κοινές χρήσεις και αποκλείοντας τις ακραίες ερμηνείες. Για τη νοημοσύνη, ιστορικά ανθρωποκεντρική, αναγνωρίζεται η μη ανθρώπινη νοημοσύνη, όπως στις μελέτες για την TN και τα ζώα. Ένας ορισμός εργασίας δεν πρέπει να είναι υπερβολικά ανθρωποκεντρικός, αποφεύγοντας την εξάρτηση από τις ειδικές για τον άνθρωπο ιδιότητες (Hernandez-Orallo, 2017).

Είναι ζωτικής σημασίας να αποφευχθεί ο ορισμός της νοημοσύνης με τη χρήση ασαφών εννοιών όπως "νους" ή "συνείδηση" χωρίς προηγούμενο ορισμό. Οι τυπικοί ορισμοί προτιμώνται για λόγους ακρίβειας, αλλά η εμπειρική φύση της νοημοσύνης σημαίνει ότι η πλήρης τυπικότητα είναι πρόκληση. Η απαίτηση για ακρίβεια στους ορισμούς μπορεί να ικανοποιηθεί μόνο σχετικά, δεδομένης της δυνατότητας διαφορετικών ερμηνειών και της εγγενούς ασάφειας σε ορισμένες έννοιες. Υπάρχει η ανάγκη για ορισμούς εργασίας για την αποτελεσματική καθοδήγηση των ερευνητικών έργων τεχνητής νοημοσύνης και την πρόκληση της εξεύρεσης ισορροπίας μεταξύ ακρίβειας και πρακτικότητας στον ορισμό της νοημοσύνης (Hajek, 2012).

Σύμφωνα με τη Britannica, (2023), Τεχνητή νοημοσύνη (TN), είναι η ικανότητα ενός ψηφιακού υπολογιστή ή ενός ελεγχόμενου από υπολογιστή ρομπότ να εκτελεί εργασίες που συνήθως συνδέονται με ευφυή όντα. Ο όρος εφαρμόζεται συχνά στο έργο της ανάπτυξης συστημάτων που διαθέτουν τις διανοητικές διεργασίες που χαρακτηρίζουν τον άνθρωπο, όπως η ικανότητα να σκέφτεται, να ανακαλύπτει νοήματα, να γενικεύει ή να μαθαίνει από προηγούμενες εμπειρίες. Από την ανάπτυξη του ψηφιακού υπολογιστή τη δεκαετία του 1940, έχει αποδειχθεί ότι οι υπολογιστές μπορούν να προγραμματιστούν για να εκτελούν πολύ σύνθετες εργασίες, όπως η ανακάλυψη αποδείξεων για μαθηματικά θεωρήματα ή το σκάκι, με μεγάλη επιδεξιότητα. Παρόλα αυτά, παρά τις συνεχείς προόδους στην ταχύτητα επεξεργασίας και τη χωρητικότητα της μνήμης των υπολογιστών, δεν υπάρχουν ακόμη προγράμματα που να μπορούν να φτάσουν την πλήρη ανθρώπινη ευελιξία σε ευρύτερους τομείς ή σε εργασίες που απαιτούν πολλές καθημερινές γνώσεις. Από την άλλη πλευρά, ορισμένα προγράμματα έχουν επιτύχει τα επίπεδα επιδόσεων των ανθρώπινων ειδικών και επαγγελματιών στην εκτέλεση ορισμένων συγκεκριμένων εργασιών, έτσι ώστε η τεχνητή νοημοσύνη με αυτή την περιορισμένη έννοια να συναντάται σε εφαρμογές τόσο διαφορετικές όσο η ιατρική διάγνωση, οι μηχανές αναζήτησης μέσω υπολογιστή, η αναγνώριση φωνής ή γραφής και τα chatbots.

2.2 Τύποι TN

2.2.1 Μηχανική μάθηση

Η μηχανική μάθηση είναι ένα υποπεδίο της τεχνητής νοημοσύνης, το οποίο ορίζεται ευρέως ως η ικανότητα μιας μηχανής να μιμείται την ευφυή ανθρώπινη συμπεριφορά. Τα συστήματα τεχνητής νοημοσύνης χρησιμοποιούνται για την εκτέλεση σύνθετων εργασιών με τρόπο παρόμοιο με τον τρόπο που οι άνθρωποι επιλύουν προβλήματα. Στόχος της τεχνητής νοημοσύνης είναι η δημιουργία υπολογιστικών μοντέλων που παρουσιάζουν "ευφυείς συμπεριφορές" όπως οι άνθρωποι. Αυτό σημαίνει μηχανές που μπορούν να αναγνωρίζουν μια οπτική σκηνή, να κατανοούν ένα κείμενο γραμμένο σε φυσική γλώσσα ή να εκτελούν μια ενέργεια στον φυσικό κόσμο. Η μηχανική μάθηση είναι ένας τρόπος για να χρησιμοποιηθεί η TN. Ορίστηκε τη δεκαετία του 1950 από τον πρωτοπόρο της TN Άρθουρ Σάμιουελ ως "το πεδίο

μελέτης που δίνει στους υπολογιστές την ικανότητα να μαθαίνουν χωρίς να προγραμματίζονται ρητά" (Alpaydin, 2021).

Ο ορισμός ισχύει, σύμφωνα με τον Mikey Shulman, λέκτορα στο MIT Sloan και επικεφαλής της μηχανικής μάθησης στην Kensho, η οποία ειδικεύεται στην τεχνητή νοημοσύνη για τις οικονομικές και αμερικανικές κοινότητες πληροφοριών. Συγκρίνει τον παραδοσιακό τρόπο προγραμματισμού των υπολογιστών, ή "λογισμικό 1.0", με τη ζαχαροπλαστική, όπου η συνταγή απαιτεί ακριβείς ποσότητες υλικών και λέει στον αρτοποιό να ανακατέψει για ένα ακριβές χρονικό διάστημα. Ο παραδοσιακός προγραμματισμός απαιτεί ομοίως τη δημιουργία λεπτομερών οδηγιών που πρέπει να ακολουθήσει ο υπολογιστής (Seelam, 2023).

Αλλά σε ορισμένες περιπτώσεις, η συγγραφή ενός προγράμματος για να το ακολουθήσει η μηχανή είναι χρονοβόρα ή αδύνατη, όπως για παράδειγμα η εκπαίδευση ενός υπολογιστή να αναγνωρίζει εικόνες διαφορετικών ανθρώπων. Ενώ οι άνθρωποι μπορούν να κάνουν εύκολα αυτή την εργασία, είναι δύσκολο να μάθει κανείς σε έναν υπολογιστή πώς να το κάνει. Η μηχανική μάθηση ακολουθεί την προσέγγιση να αφήνει τους υπολογιστές να μαθαίνουν να προγραμματίζουν τους εαυτούς τους μέσω της εμπειρίας (Alpaydin, 2021).

Η μηχανική μάθηση ξεκινά με την προετοιμασία δεδομένων, που περιλαμβάνουν διάφορους τύπους, όπως αριθμούς, φωτογραφίες ή κείμενο, και συχνά προέρχονται από διαφορετικές πηγές, όπως τραπεζικές συναλλαγές, εικόνες ή εκθέσεις πωλήσεων. Όσο πιο εκτεταμένο είναι το σύνολο δεδομένων, τόσο καλύτερη είναι η απόδοση του προγράμματος μηχανικής μάθησης. Στη συνέχεια, οι προγραμματιστές επιλέγουν ένα μοντέλο μηχανικής μάθησης, παρέχουν τα δεδομένα για εκπαίδευση και επιτρέπουν στο μοντέλο να μαθαίνει αυτόνομα μοτίβα ή να κάνει προβλέψεις. Οι συνεχείς προσαρμογές, συμπεριλαμβανομένων των αλλαγών των παραμέτρων, μπορούν να γίνουν από ανθρώπινους προγραμματιστές για την ενίσχυση της ακρίβειας (Carleo et al., 2019).

Ένα μέρος των δεδομένων προορίζεται για αξιολόγηση ώστε να αξιολογηθεί η ακρίβεια του μοντέλου όταν εκτίθεται σε νέες πληροφορίες. Οι επιτυχημένοι αλγόριθμοι μηχανικής μάθησης μπορεί να είναι περιγραφικοί, προγνωστικοί ή προδιαγραφικοί, προσφέροντας πληροφορίες για γεγονότα του παρελθόντος,

προβλέποντας μελλοντικά συμβάντα ή προτείνοντας ενέργειες με βάση τα δεδομένα (Alpaydin, 2021).

Η μηχανική μάθηση περιλαμβάνει τρεις υποκατηγορίες:

- **Επιβλεπόμενη μηχανική μάθηση:** Περιλαμβάνει την εκπαίδευση μοντέλων με επισημειωμένα σύνολα δεδομένων, επιτρέποντάς τους να μαθαίνουν και να βελτιώνουν την ακρίβεια με την πάροδο του χρόνου. Για παράδειγμα, ένας αλγόριθμος που εκπαιδεύεται με επισημειωμένες εικόνες σκύλων μπορεί να μάθει να αναγνωρίζει εικόνες σκύλων ανεξάρτητα. Η επιβλεπόμενη μηχανική μάθηση είναι ο πιο συχνά χρησιμοποιούμενος τύπος (Jiang, Gradus, & Rosellini, 2020).
- **Μη εποπτευόμενη μηχανική μάθηση:** Περιλαμβάνει προγράμματα που αναζητούν μοτίβα σε μη επισημειωμένα δεδομένα, ανακαλύπτοντας τάσεις ή μοτίβα που μπορεί να μην αναζητούνται ρητά από τον άνθρωπο. Για παράδειγμα, ένα πρόγραμμα μηχανικής μάθησης χωρίς επίβλεψη μπορεί να εντοπίσει διαφορετικούς τύπους πελατών σε δεδομένα διαδικτυακών πωλήσεων (Cabanesetal., 2019).
- **Ενισχυτική μηχανική μάθηση:** Εκπαιδεύει μηχανές μέσω δοκιμής και σφάλματος, δημιουργώντας ένα σύστημα ανταμοιβής για την καθοδήγηση των βέλτιστων ενεργειών. Χρησιμοποιείται σε σενάρια όπως η εκπαίδευση μοντέλων για να παίζουν παιχνίδια ή η διδασκαλία αυτόνομων οχημάτων για να οδηγούν, παρέχοντας ανατροφοδότηση για τις σωστές αποφάσεις, διευκολύνοντας τη συνεχή μάθηση με την πάροδο του χρόνου (Carleoetal., 2019).

2.2.2 Βαθιά μάθηση

Η βαθιά μάθηση είναι ένας τύπος μηχανικής μάθησης και τεχνητής νοημοσύνης (AI) που μιμείται τον τρόπο με τον οποίο οι άνθρωποι αποκτούν ορισμένους τύπους γνώσης. Τα μοντέλα βαθιάς μάθησης μπορούν να διδαχθούν να εκτελούν εργασίες ταξινόμησης και να αναγνωρίζουν μοτίβα σε φωτογραφίες, κείμενο, ήχο και άλλα

διάφορα δεδομένα. Χρησιμοποιείται επίσης για την αυτοματοποίηση εργασιών που κανονικά θα χρειάζονταν ανθρώπινη νοημοσύνη, όπως η περιγραφή εικόνων ή η μεταγραφή αρχείων ήχου. Η βαθιά μάθηση αποτελεί σημαντικό στοιχείο της επιστήμης των δεδομένων, συμπεριλαμβανομένης της στατιστικής και της προγνωστικής μοντελοποίησης. Είναι εξαιρετικά επωφελής για τους επιστήμονες δεδομένων οι οποίοι είναι επιφορτισμένοι με τη συλλογή, ανάλυση και ερμηνεία μεγάλου όγκου δεδομένων. Η βαθιά μάθηση καθιστά τη διαδικασία αυτή ταχύτερη και ευκολότερη. Ενώ ο ανθρώπινος εγκέφαλος διαθέτει εκατομμύρια διασυνδεδεμένους νευρώνες που συνεργάζονται για να μάθουν πληροφορίες, η βαθιά μάθηση διαθέτει νευρωνικά δίκτυα που κατασκευάζονται από πολλαπλά στρώματα κόμβων λογισμικού που συνεργάζονται. Τα μοντέλα βαθιάς μάθησης εκπαιδεύονται χρησιμοποιώντας ένα μεγάλο σύνολο επισημασμένων δεδομένων και αρχιτεκτονικές νευρωνικών δικτύων. Η βαθιά μάθηση επιτρέπει σε έναν υπολογιστή να μαθαίνει μέσω παραδείγματος (Kelleher, 2019).

2.2.3 Επεξεργασία φυσικής γλώσσας

Η επεξεργασία φυσικής γλώσσας (NLP) είναι ένα διεπιστημονικό υποπεδίο της επιστήμης των υπολογιστών και της γλωσσολογίας. Ασχολείται κυρίως με το να δώσει στους υπολογιστές την ικανότητα να υποστηρίζουν και να χειρίζονται την ανθρώπινη γλώσσα. Περιλαμβάνει την επεξεργασία συνόλων δεδομένων φυσικής γλώσσας, όπως σώματα κειμένων ή σώματα ομιλίας, χρησιμοποιώντας είτε προσεγγίσεις μηχανικής μάθησης βασισμένες σε κανόνες είτε πιθανολογικές (δηλαδή στατιστικές και, πιο πρόσφατα, βασισμένες σε νευρωνικά δίκτυα). Ο στόχος είναι ένας υπολογιστής ικανός να "κατανοεί" το περιεχόμενο των εγγράφων, συμπεριλαμβανομένων των αποχρώσεων της γλώσσας που περιέχεται σε αυτά. Η τεχνολογία μπορεί στη συνέχεια να εξάγει με ακρίβεια πληροφορίες και ιδέες που περιέχονται στα έγγραφα, καθώς και να κατηγοριοποιεί και να οργανώνει τα ίδια τα έγγραφα. Οι προκλήσεις στην επεξεργασία φυσικής γλώσσας περιλαμβάνουν συχνά την αναγνώριση ομιλίας, την κατανόηση φυσικής γλώσσας και την παραγωγή φυσικής γλώσσας (Eisenstein, 2019).

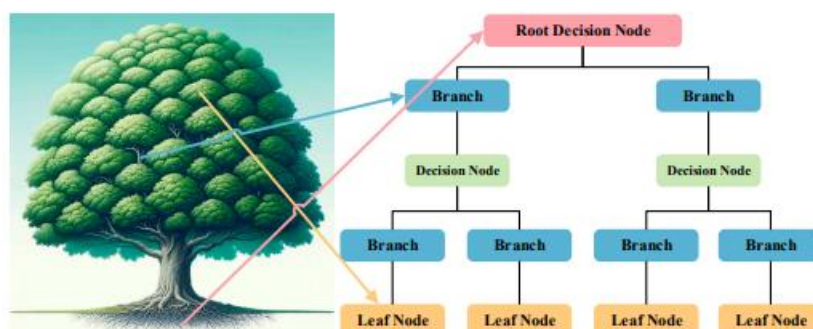
Κεφάλαιο 3: Τεχνικές τεχνητής νοημοσύνης για την ανίχνευση απάτης

3.1 Αλγόριθμοι μηχανικής μάθησης στην ανίχνευση απάτης

Οι παραδοσιακές προσεγγίσεις μηχανικής μάθησης (ML) έχουν διαδραματίσει καθοριστικό ρόλο στον τομέα της ανίχνευσης της απάτης, παρέχοντας αποτελεσματικά μέσα για την ανάλυση και ερμηνεία μοτίβων δεδομένων που υποδεικνύουν δόλιες δραστηριότητες. Αυτές οι μέθοδοι αξιοποιούν τεχνικές μάθησης με επίβλεψη, όπου οι αλγόριθμοι εκπαιδεύονται σε ιστορικά δεδομένα που έχουν επισημανθεί με περιπτώσεις απάτης και μη απάτης. Ο πρωταρχικός στόχος είναι να μπορέσει το σύστημα να γενικεύσει από τα παραδείγματα του παρελθόντος, μαθαίνοντας να διακρίνει μεταξύ φυσιολογικής και ανώμαλης συμπεριφοράς σε νέα

δεδομένα. Η μάθηση με επίβλεψη αποτελεί τον θεμέλιο λίθο των παραδοσιακών προσεγγίσεων ML για την ανίχνευση απάτης (Yang, Varghese, Stephenson, Tu, &Gronsbell, 2023).

Κατά τη διάρκεια της φάσης εκπαίδευσης, ο αλγόριθμος μαθαίνει από ένα σύνολο δεδομένων με ετικέτες, εντοπίζοντας πρότυπα και χαρακτηριστικά που σχετίζονται με γνωστές περιπτώσεις απάτης. Αυτή η γνώση εφαρμόζεται στη συνέχεια σε νέα, μη επισημασμένα δεδομένα για να προβλεφθεί εάν μια συγκεκριμένη συναλλαγή ή δραστηριότητα είναι πιθανό να είναι δόλια. Οι συνήθεις αλγόριθμοι επιβλεπόμενης μάθησης στην ανίχνευση απάτης περιλαμβάνουν δέντρα αποφάσεων(Gkikasetal., 2023), λογιστική παλινδρόμηση (Vo, Jonnagaddala, &Liaw, 2019) και μηχανές διανυσμάτων υποστήριξης. Για παράδειγμα, το διάγραμμα δομής του δέντρου απόφασηςπου παρουσιάζεται στην Εικόνα 1. Η αποτελεσματικότητα αυτών των μοντέλων εξαρτάται από την ποιότητα και την αντιπροσωπευτικότητα των δεδομένων εκπαίδευσης (Mignan, 2019).



Εικόνα 1: διάγραμμα δομής του δέντρου απόφασης (Mignan, 2019).

3.2 Μάθηση χωρίς επίβλεψη για την ανίχνευση ανωμαλιών

Εκτός από τη μάθηση με επίβλεψη, οι παραδοσιακές προσεγγίσεις ML (Zhang, 2010), συχνά ενσωματώνουν τεχνικές μάθησης χωρίς επίβλεψη, ιδίως για την ανίχνευση ανωμαλιών. Η ανώμαλη συμπεριφορά στα δεδομένα, η οποία μπορεί να υποδηλώνει δόλια δραστηριότητα, εντοπίζεται χωρίς τη χρήση επισημασμένων παραδειγμάτων. Οι αλγόριθμοι ομαδοποίησης, όπως ο k-means(Irarraguirre-Villanuevaetal., 2022), και οι μέθοδοι που βασίζονται στην πυκνότητα, όπως τα δάση απομόνωσης (Barbariol, &Susto, 2022), χρησιμοποιούνται συνήθως για τον εντοπισμό ακραίων τιμών ή προτύπων που αποκλίνουν από τον κανόνα. Η μάθηση χωρίς επίβλεψη είναι

πολύτιμη σε σενάρια όπου τα επισημασμένα δεδομένα απάτης είναι σπάνια, παρέχοντας μια συμπληρωματική προσέγγιση στις μεθόδους με επίβλεψη (Lin, Mengaldo, & Maulik, 2023). Η μηχανική των χαρακτηριστικών είναι μια κρίσιμη πτυχή των παραδοσιακών προσεγγίσεων ML στην ανίχνευση απάτης. Η επιλογή και ο μετασχηματισμός των σχετικών χαρακτηριστικών από τα δεδομένα ενισχύουν την ικανότητα του μοντέλου να διακρίνει τα μοτίβα απάτης. Η μηχανική χαρακτηριστικών μπορεί να περιλαμβάνει τη δημιουργία νέων μεταβλητών, τη συγκέντρωση δεδομένων συναλλαγών ή την κανονικοποίηση υφιστάμενων χαρακτηριστικών. Επιπλέον, η ερμηνευσιμότητα των μοντέλων είναι απαραίτητη για την κατανόηση της λογικής πίσω από τις προβλέψεις απάτης. Τα διαφανή μοντέλα, όπως τα δέντρα αποφάσεων (Wang et al., 2023), προσφέρουν πληροφορίες σχετικά με τα χαρακτηριστικά που συμβάλλουν σε μια πρόβλεψη, βοηθώντας τους αναλυτές να κατανοήσουν και να επικυρώσουν τις αποφάσεις του μοντέλου. Παρά την αποτελεσματικότητά τους, οι παραδοσιακές προσεγγίσεις ML για την ανίχνευση απάτης έχουν περιορισμούς. Μπορεί να δυσκολεύονται να προσαρμοστούν στις ταχέως εξελισσόμενες τακτικές απάτης και μπορεί να είναι λιγότερο αποτελεσματικές στο χειρισμό ιδιαίτερα ανισόρροπων συνόλων δεδομένων. Καθώς οι απατεώνες βελτιώνουν συνεχώς τις τεχνικές τους, υπάρχει αυξανόμενη ανάγκη για πιο προηγμένες και προσαρμοστικές μεθόδους. Το μέλλον της ανίχνευσης απάτης έγκειται στην ενσωμάτωση παραδοσιακών προσεγγίσεων με αναδυόμενες τεχνολογίες, όπως η βαθιά μάθηση και οι μέθοδοι συνόλου, για την ενίσχυση της ακρίβειας και της ανθεκτικότητας. Επιπλέον, η αντιμετώπιση προκλήσεων όπως η ερμηνευσιμότητα των μοντέλων και οι ηθικοί προβληματισμοί (Liet al., 2023), παραμένει ένα σημείο εστίασης για τη βελτίωση και την εξέλιξη των παραδοσιακών τεχνικών ML στο πλαίσιο της ανίχνευσης απάτης.

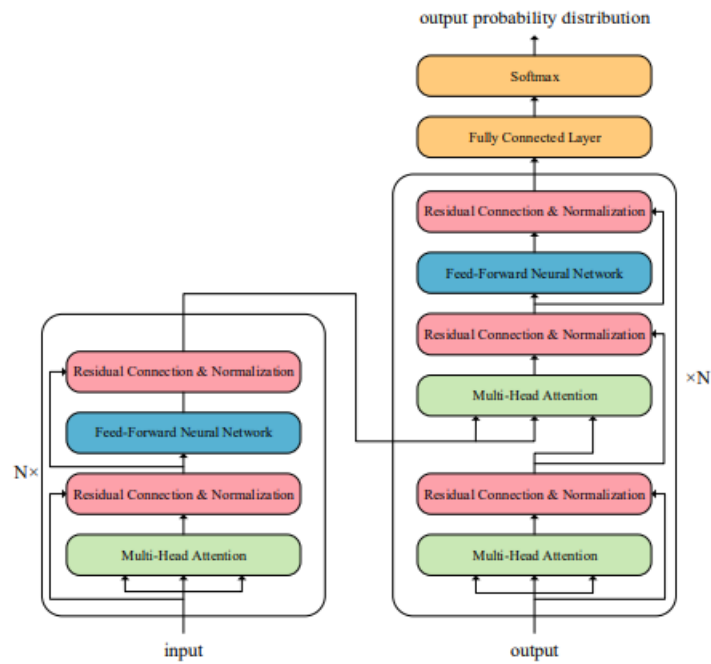
3.3 Βαθιά μάθηση για την ανίχνευση απάτης

Η βαθιά μάθηση έχει αναδειχθεί ως ένα ισχυρό εργαλείο στον τομέα της ανίχνευσης απάτης, αξιοποιώντας σύνθετα νευρωνικά δίκτυα για την αυτόματη εκμάθηση και εξαγωγή περίπλοκων μοτίβων μέσα από μεγάλα σύνολα δεδομένων (Maashi, Alabdullah, & Kouki, 2023). Σε αντίθεση με τις παραδοσιακές μεθόδους μηχανικής μάθησης, οι προσεγγίσεις βαθιάς μάθησης εμβαθύνουν σε πολλαπλά επίπεδα αφαίρεσης (Wang, & Fernandes, 2022; Zhang, 2021), επιτρέποντας την αυτόματη

ανακάλυψη χαρακτηριστικών και αναπαραστάσεων που μπορεί να είναι δύσκολο να καταγραφούν με τη χρήση χειροκίνητης μηχανικής χαρακτηριστικών. Αυτή η προσαρμοστικότητα είναι ιδιαίτερα επωφελής στο πλαίσιο της ανίχνευσης απάτης, όπου τα δόλια μοτίβα μπορεί να είναι δυναμικά και μη γραμμικά (Li, Huang, Liu, & Jiang, 2021).

Τα μοντέλα βαθιάς μάθησης που εφαρμόζονται στην ανίχνευση απάτης συχνά περιλαμβάνουν αρχιτεκτονικές όπως τα βαθιά νευρωνικά δίκτυα, τα νευρωνικά δίκτυα συνελίξεων (CNN), τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) και πιο πρόσφατα τα μοντέλα που βασίζονται σε μετασχηματιστές. Τα βαθιά νευρωνικά δίκτυα είναι ικανά να μαθαίνουν ιεραρχικές αναπαραστάσεις χαρακτηριστικών από δεδομένα συναλλαγών και συμπεριφοράς χρηστών (Barcinal., 2024). Τα CNN είναι αποτελεσματικά για χωρικά δεδομένα, όπως εικόνες που σχετίζονται με την ανίχνευση απάτης (Wu et al., 2023), ενώ τα RNN είναι πολύτιμα για τη σύλληψη χρονικών εξαρτήσεων σε διαδοχικά δεδομένα (Sarkar et al., 2024).

Τα μοντέλα μετασχηματιστών, τα οποία αναπτύχθηκαν αρχικά για την επεξεργασία φυσικής γλώσσας, έχουν δείξει ότι υπόσχονται να συλλάβουν εξαρτήσεις μεγάλης εμβέλειας και πληροφορίες σχετικά με το πλαίσιο που σχετίζονται με την ανίχνευση απάτης (Li, Cao, Xu, Zhu, & Dong, 2024; Kim et al., 2023). Το μοντέλο μετασχηματιστή αποτελείται από δύο κύρια μέρη, τον κωδικοποιητή και τον αποκωδικοποιητή. Το μοντέλο του μετασχηματιστή παρουσιάζεται στην εικόνα 2. Το αριστερό μέρος είναι ο κωδικοποιητής, ο οποίος αποτελείται από πολλαπλά στρώματα μηχανισμών προσοχής και νευρωνικών δικτύων τροφοδότησης προς τα εμπρός και ακολουθεί τη διαδικασία κανονικοποίησης υπολειμμάτων, όπου η έξοδος κάθε στρώματος είναι ένα πολυδιάστατο διάνυσμα και το διάνυσμα εξόδου κάθε στρώματος είναι η είσοδος του επόμενου στρώματος. Δεξιά βρίσκεται ο αποκωδικοποιητής, ο οποίος διαθέτει επίσης τα δύο στρώματα του κωδικοποιητή, αλλά μεταξύ τους υπάρχει η προσοχή κωδικοποιητή-αποκωδικοποιητή, η οποία χρησιμοποιείται για να βοηθήσει τον αποκωδικοποιητή να εστιάσει στα σχετικά μέρη της πρότασης εισόδου (Wang et al., 2023).



Εικόνα 2: Διάγραμμα μοντέλου μετασχηματιστή.

Η εκπαίδευση μοντέλων βαθιάς μάθησης (Zhang, 2023), για την ανίχνευση απάτης απαιτεί σημαντικές ποσότητες επισημασμένων δεδομένων, γεγονός που μπορεί να αποτελέσει πρόκληση λόγω της ανισορροπίας μεταξύ κανονικών και δόλιων περιπτώσεων (Rubaidi, Ammar, & Aouicha, 2022). Για την αντιμετώπιση αυτής της ανισορροπίας χρησιμοποιούνται τεχνικές προεπεξεργασίας δεδομένων, όπως η επαναδειγματοληψία, η υπερδειγματοληψία και η υποδειγματοληψία (Hou, 2018). Επιπλέον, η εκμάθηση μεταφοράς, όπου ένα μοντέλο που έχει προ-εκπαιδευτεί σε ένα μεγάλο σύνολο δεδομένων προσαρμόζεται λεπτομερώς για την ανίχνευση απάτης, μπορεί να βελτιώσει την απόδοση (Sisodia, & Sisodia, 2023). Οι στρατηγικές εκπαίδευσης περιλαμβάνουν επίσης τη χρήση αυτοκωδικοποιητών για μάθηση χαρακτηριστικών χωρίς επίβλεψη και ανίχνευση ανωμαλιών (Chaquet-Ulledemolinsetal., 2022), επιτρέποντας στο μοντέλο να εντοπίζει μοτίβα ενδεικτικά απάτης χωρίς ρητές ετικέτες.

Παρά την επιτυχία τους, οι προσεγγίσεις βαθιάς μάθησης για την ανίχνευση απάτης αντιμετωπίζουν προκλήσεις, συμπεριλαμβανομένης της ερμηνευσιμότητας και της εξηγησιμότητας (Gaoetal., 2023; Probst, 2023). Η εγγενής πολυπλοκότητα των βαθιών νευρωνικών δικτύων μπορεί να καταστήσει δύσκολη την κατανόηση της λογικής πίσω από συγκεκριμένες προβλέψεις, εγείροντας ανησυχίες σχετικά με τη

διαφάνεια των διαδικασιών λήψης αποφάσεων. Οι δεοντολογικοί προβληματισμοί, όπως η μεροληψία στους αλγορίθμους και το ενδεχόμενο διακριτικών αποτελεσμάτων, αποτελούν επίσης κρίσιμες ανησυχίες τις οποίες οι ερευνητές και οι επαγγελματίες του τομέα αντιμετωπίζουν ενεργά για να διασφαλίσουν τη δίκαιη και αμερόληπτη ανίχνευση της απάτης.

Το μέλλον της βαθιάς μάθησης στην ανίχνευση απάτης περιλαμβάνει την αντιμετώπιση των σημερινών προκλήσεων και την ενσωμάτωση με τις παραδοσιακές προσεγγίσεις μηχανικής μάθησης. Διερευνώνται υβριδικά μοντέλα που συνδυάζουν τα πλεονεκτήματα της βαθιάς μάθησης και των παραδοσιακών μεθόδων για να ενισχύσουν την ευρωστία και την ερμηνευσιμότητα (Kimetal., 2019). Περαιτέρω πρόοδοι στην επεξηγήσιμη TN (vanZetten, Ramackers, &Hoos, 2022), τεχνικές κανονικοποίησης (Liu, 2023) και η ανάπτυξη πιο αποτελεσματικών στρατηγικών εκπαίδευσης αναμένεται να προωθήσουν τις προσεγγίσεις βαθιάς μάθησης στο προσκήνιο της ανίχνευσης απάτης, συμβάλλοντας στις συνεχιζόμενες προσπάθειες για τον μετριασμό των εξελισσόμενων απειλών και την προστασία από δόλιες δραστηριότητες σε διάφορους τομείς.

3.4 Κοινά σύνολα δεδομένων ανίχνευσης απάτης

Τα σύνολα δεδομένων υψηλής ποιότητας διαδραματίζουν κρίσιμο ρόλο στην ανάπτυξη, την εκπαίδευση και την αξιολόγηση των μοντέλων ανίχνευσης απάτης. Χρησιμεύουν ως βάση για τη διδασκαλία αλγορίθμων για τη διάκριση μεταξύ νόμιμων και δόλιων μοτίβων εντός των δεδομένων. Τα κοινά σύνολα δεδομένων για την ανίχνευση απάτης παρέχουν στους ερευνητές και τους επαγγελματίες αντιπροσωπευτικά δείγματα πραγματικών σεναρίων, επιτρέποντάς τους να δημιουργήσουν ισχυρά μοντέλα που μπορούν να γενικεύουν καλά σε ποικίλα και εξελισσόμενα πρότυπα απάτης (Xuetal., 2023). Αρκετά ευρέως χρησιμοποιούμενα σύνολα δεδομένων συμβάλλουν καθοριστικά στην προώθηση της έρευνας και της ανάπτυξης στον τομέα της ανίχνευσης απάτης. Το σύνολο δεδομένων για την ανίχνευση απάτης πιστωτικών καρτών, το οποίο διατίθεται σε πλατφόρμες όπως η Kaggle, περιλαμβάνει ανώνυμες συναλλαγές πιστωτικών καρτών, με σημαντική ανισορροπία κλάσεων μεταξύ κανονικών και δόλιων περιπτώσεων (Alharbietal., 2022). Το σύνολο δεδομένων IEEE-CISFraudDetection, το οποίο εισήχθη από το Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE), επικεντρώνεται σε

συναλλαγές ηλεκτρονικού εμπορίου, προσφέροντας ένα ποικίλο σύνολο χαρακτηριστικών για ολοκληρωμένη εκπαίδευση μοντέλων. Ένα άλλο αξιοσημείωτο σύνολο δεδομένων είναι το SyntheticFinancialDatasetsforFraudDetection, το οποίο παρέχει συνθετικά αλλά ρεαλιστικά οικονομικά δεδομένα για την αξιολόγηση της απόδοσης των αλγορίθμων ανίχνευσης απάτης.

Τα κοινά σύνολα δεδομένων για την ανίχνευση απάτης μοιράζονται ορισμένα χαρακτηριστικά που είναι απαραίτητα για την αποτελεσματικότητα των μοντέλων μηχανικής μάθησης. Η μη ισορροπημένη κατανομή κλάσεων, όπου η πλειονότητα των περιπτώσεων δεν είναι δόλιες, αντικατοπτρίζει το σενάριο του πραγματικού κόσμου, αλλά απαιτεί προσεκτικό χειρισμό για την αποφυγή μεροληψίας του μοντέλου. Οι διαχρονικές πτυχές, όπως τα χρονικά μοτίβα στα δεδομένα συναλλαγών, είναι συχνά παρούσες για την αποτύπωση της δυναμικής φύσης των δόλιων δραστηριοτήτων. Τα σύνολα δεδομένων μπορεί επίσης να περιλαμβάνουν χαρακτηριστικά όπως ποσά συναλλαγών, χρονοσφραγίδες και συμπεριφορά χρηστών, παρέχοντας μια ολιστική αναπαράσταση των παραγόντων που συμβάλλουν στην απάτη. Ενώ τα κοινά σύνολα δεδομένων για την ανίχνευση απάτης προσφέρουν πολύτιμους πόρους, οι προκλήσεις εξακολουθούν να υφίστανται. Η ταχέως εξελισσόμενη φύση της απάτης απαιτεί συνεχείς ενημερώσεις και επεκτάσεις των συνόλων δεδομένων ώστε να αντικατοπτρίζουν τα αναδυόμενα πρότυπα και τις τακτικές που χρησιμοποιούν οι απατεώνες. Επιπλέον, η διασφάλιση της ιδιωτικότητας και της ασφάλειας των ευαίσθητων πληροφοριών εντός των συνόλων δεδομένων είναι υψίστης σημασίας, γεγονός που ωθεί στην ανάπτυξη συνθετικών συνόλων δεδομένων που μιμούνται τα χαρακτηριστικά του πραγματικού κόσμου χωρίς να διακυβεύεται η ιδιωτικότητα των ατόμων (Chen et al., 2024). Οι συνεργατικές προσπάθειες της ερευνητικής κοινότητας, της βιομηχανίας και των ρυθμιστικών φορέων είναι ζωτικής σημασίας για την αντιμετώπιση αυτών των προκλήσεων και τη διατήρηση της συνάφειας των κοινών συνόλων δεδομένων ανίχνευσης απάτης στη συνεχή επιδίωξη αποτελεσματικών στρατηγικών πρόληψης και ανίχνευσης της απάτης.

3.5 Νευρωνικά δίκτυα και οι εφαρμογές τους

Τα νευρωνικά δίκτυα είναι ένα υποσύνολο μοντέλων μηχανικής μάθησης εμπνευσμένων από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου.

Αποτελούνται από διασυνδεδεμένους κόμβους ή τεχνητούς νευρώνες, οργανωμένους σε επίπεδα. Κάθε σύνδεση μεταξύ των κόμβων έχει ένα βάρος και τα βάρη αυτά προσαρμόζονται κατά τη διάρκεια της διαδικασίας εκπαίδευσης για τη βελτιστοποίηση της απόδοσης του μοντέλου. Τα νευρωνικά δίκτυα υπερέχουν στην εκμάθηση σύνθετων μοτίβων και σχέσεων στα δεδομένα, γεγονός που τα καθιστά κατάλληλα για εργασίες ανίχνευσης απάτης. Ένα από τα κύρια πλεονεκτήματα των νευρωνικών δικτύων στην ανίχνευση απάτης είναι η ικανότητά τους να προσαρμόζονται στα μεταβαλλόμενα πρότυπα και στις εξελισσόμενες τακτικές που χρησιμοποιούν οι απατεώνες. Τα παραδοσιακά συστήματα που βασίζονται σε κανόνες μπορεί να δυσκολεύονται να συμβαδίσουν με τη δυναμική φύση της απάτης, ενώ τα νευρωνικά δίκτυα μπορούν να μαθαίνουν συνεχώς και να ενημερώνουν τα μοντέλα τους με βάση τα νέα δεδομένα (Yang, & Wang, 2020).

Ένας συνηθισμένος τύπος νευρωνικού δικτύου που χρησιμοποιείται στην ανίχνευση απάτης είναι το νευρωνικό δίκτυο τροφοδότησης. Αυτός ο τύπος δικτύου αποτελείται από ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Το στρώμα εισόδου λαμβάνει χαρακτηριστικά που σχετίζονται με μια συναλλαγή ή δραστηριότητα, όπως το ποσό της συναλλαγής, η τοποθεσία και η συμπεριφορά του χρήστη. Τα κρυφά στρώματα επεξεργάζονται αυτές τις πληροφορίες και το στρώμα εξόδου παράγει μια πρόβλεψη, υποδεικνύοντας αν η συναλλαγή είναι πιθανό να είναι δόλια (Cheng, Wang, Zhang, & Zhang, 2020).

Τα επαναλαμβανόμενα νευρωνικά δίκτυα (RNN) είναι μια άλλη παραλλαγή που είναι ιδιαίτερα αποτελεσματική στη σύλληψη διαδοχικών μοτίβων. Στην ανίχνευση απάτης, οι ακολουθίες συναλλαγών ή ενεργειών του χρήστη μπορούν να αποκαλύψουν ύποπτα μοτίβα που μπορεί να περάσουν απαρατήρητα από τις παραδοσιακές μεθόδους. Τα RNN υπερέχουν στον εντοπισμό τέτοιων χρονικών εξαρτήσεων και είναι ιδιαίτερα χρήσιμα όταν πρόκειται για δεδομένα χρονοσειρών. Μια από τις προκλήσεις στην ανίχνευση απάτης είναι η αντιμετώπιση ανισοβαρών συνόλων δεδομένων, όπου ο αριθμός των κανονικών συναλλαγών υπερτερεί κατά πολύ των δόλιων. Τα νευρωνικά δίκτυα μπορούν να τροποποιηθούν για να το αντιμετωπίσουν αυτό με τη χρήση τεχνικών όπως η υπερδειγματοληψία της μειονοτικής κλάσης ή η προσαρμογή των βαρών των κλάσεων κατά τη διάρκεια της εκπαίδευσης. Με τον τρόπο αυτό διασφαλίζεται ότι το μοντέλο δεν μεροληπτεί προς

την πρόβλεψη της πλειοψηφικής κλάσης και διατηρεί υψηλή ευαισθησία στις περιπτώσεις απάτης (Liu et al., 2020).

Η εφαρμογή των νευρωνικών δικτύων στην ανίχνευση απάτης δεν περιορίζεται σε δεδομένα συναλλαγών. Τα βιομετρικά στοιχεία συμπεριφοράς, όπως η δυναμική των πληκτρολογήσεων και τα μοτίβα κίνησης του ποντικιού, μπορούν επίσης να ενσωματωθούν σε μοντέλα νευρωνικών δικτύων για την ενίσχυση της ακρίβειας. Λαμβάνοντας υπόψη όχι μόνο τα συναλλακτικά χαρακτηριστικά αλλά και τα μοναδικά πρότυπα συμπεριφοράς του χρήστη, τα νευρωνικά δίκτυα μπορούν να διακρίνουν καλύτερα μεταξύ νόμιμων και δόλιων δραστηριοτήτων (Cheng, Wang, Zhang, & Zhang, 2020).

Η ανίχνευση απάτης σε πραγματικό χρόνο είναι ζωτικής σημασίας για την πρόληψη οικονομικών απωλειών. Τα νευρωνικά δίκτυα, με την ικανότητά τους να επεξεργάζονται γρήγορα τις πληροφορίες, είναι κατάλληλα για εφαρμογές πραγματικού χρόνου. Μπορούν να αναλύουν τις εισερχόμενες ροές δεδομένων και να λαμβάνουν στιγμιαίες αποφάσεις, παρέχοντας μια προληπτική προσέγγιση για την πρόληψη της απάτης (Liu et al., 2020).

Εν κατακλείδι, τα νευρωνικά δίκτυα έχουν αποδειχθεί ένα ισχυρό και ευέλικτο εργαλείο στον τομέα της ανίχνευσης απάτης. Η ικανότητά τους να μαθαίνουν πολύπλοκα μοτίβα, να προσαρμόζονται στις εξελισσόμενες απειλές και να επεξεργάζονται δεδομένα σε πραγματικό χρόνο τα καθιστά κατάλληλα για την αντιμετώπιση των προκλήσεων που θέτουν οι σύγχρονες δραστηριότητες απάτης. Καθώς η τεχνητή νοημοσύνη συνεχίζει να εξελίσσεται, τα νευρωνικά δίκτυα είναι πιθανό να διαδραματίσουν ακόμη πιο σημαντικό ρόλο στη διασφάλιση των χρηματοπιστωτικών συστημάτων και των ηλεκτρονικών συναλλαγών από κακόβουλους παράγοντες.

3.6 Αυτοκωδικοποιητές για την ανίχνευση ανωμαλιών

Οι αυτοκωδικοποιητές, ένα υποσύνολο των νευρωνικών δικτύων, έχουν αποκτήσει εξέχουσα θέση σε εργασίες ανίχνευσης ανωμαλιών λόγω της ικανότητάς τους να μαθαίνουν ισχυρές αναπαραστάσεις δεδομένων. Η ανίχνευση ανωμαλιών είναι ζωτικής σημασίας σε διάφορους τομείς, όπως η ανίχνευση απάτης, η ασφάλεια στον κυβερνοχώρο και η βιομηχανική παρακολούθηση. Οι αυτοκωδικοποιητές υπερέρχουν

στον εντοπισμό αποκλίσεων από τα κανονικά πρότυπα με την ανακατασκευή δεδομένων εισόδου και την ανίχνευση ανωμαλιών με βάση τα σφάλματα ανακατασκευής (Pinaya et al., 2020).

Ένας αυτόματος κωδικοποιητής αποτελείται από έναν κωδικοποιητή και έναν αποκωδικοποιητή. Ο κωδικοποιητής συμπιέζει τα δεδομένα εισόδου σε μια αναπαράσταση χαμηλότερης διάστασης (λανθάνων χώρος), ενώ ο αποκωδικοποιητής ανακατασκευάζει τα αρχικά δεδομένα από αυτή την αναπαράσταση. Κατά τη διάρκεια της εκπαίδευσης, ο αυτόματος κωδικοποιητής μαθαίνει να ελαχιστοποιεί τη διαφορά μεταξύ της εισόδου και της ανακατασκευασμένης εξόδου. Η ενδιαφέρουσα πτυχή για την ανίχνευση ανωμαλιών έγκειται στην ικανότητα των αυτοκωδικοποιητών να συλλαμβάνουν τα βασικά χαρακτηριστικά των κανονικών δεδομένων, καθιστώντας τους ευαίσθητους σε αποκλίσεις που μπορεί να υποδηλώνουν ανώμαλη συμπεριφορά. Σε σενάρια ανίχνευσης απάτης, οι αυτοκωδικοποιητές μπορούν να εκπαιδευτούν σε ένα σύνολο δεδομένων που περιλαμβάνει μόνο κανονικές συναλλαγές. Το μοντέλο μαθαίνει να κωδικοποιεί τα κανονικά μοτίβα που ενυπάρχουν στις νόμιμες συναλλαγές. Αφού εκπαιδευτεί, ο αυτόματος κωδικοποιητής μπορεί να ανακατασκευάσει νέες περιπτώσεις και οι περιπτώσεις με σημαντικά σφάλματα ανακατασκευής επισημαίνονται ως πιθανές ανωμαλίες (Bank, Koenigstein, & Giryas, 2023).

Ένα από τα πλεονεκτήματα των αυτόματων κωδικοποιητών είναι η μη εποπτευόμενη φύση τους, καθώς απαιτούν μόνο κανονικά δεδομένα για την εκπαίδευση. Αυτό είναι ιδιαίτερα επωφελές στην ανίχνευση απάτης, όπου οι χαρακτηρισμένες περιπτώσεις απάτης είναι συχνά σπάνιες ή μπορεί να μην αντιπροσωπεύουν όλο το φάσμα των πιθανών μοτίβων απάτης. Οι αυτόματοι κωδικοποιητές μπορούν να προσαρμόζονται στις αναδυόμενες τακτικές απάτης χωρίς ρητές ετικέτες, καθιστώντας τους κατάλληλους για δυναμικά και εξελισσόμενα σενάρια απάτης. Οι μεταβλητοί αυτοκωδικοποιητές (VAE) είναι μια παραλλαγή των αυτοκωδικοποιητών που εισάγουν πιθανοτικά στοιχεία στον λανθάνων χώρο. Οι VAEs όχι μόνο κωδικοποιούν δεδομένα σε ένα λανθάνοντα χώρο, αλλά δημιουργούν επίσης ποικίλα δείγματα κατά τη διαδικασία αποκωδικοποίησης. Αυτός ο στοχαστικός χαρακτήρας επιτρέπει στους VAEs να συλλάβουν την εγγενή αβεβαιότητα στα δεδομένα, καθιστώντας τους πιο αποτελεσματικούς στο χειρισμό σύνθετων και ποικίλων προτύπων, κάτι που είναι απαραίτητο σε εργασίες ανίχνευσης ανωμαλιών (Pinaya et al., 2020).

Οι αυτοκωδικοποιητές δεν περιορίζονται σε δεδομένα σε πίνακες ή απλά χαρακτηριστικά- μπορούν να εφαρμοστούν σε διάφορους τύπους δεδομένων, συμπεριλαμβανομένων εικόνων, κειμένου και δεδομένων χρονοσειρών. Αυτή η ευελιξία τους καθιστά ευέλικτους για διάφορες εφαρμογές, όπως η ανίχνευση ανωμαλιών στην κυκλοφορία δικτύου, ο εντοπισμός δόλιων μοτίβων σε δεδομένα κειμένου ή ο εντοπισμός παρατυπιών σε ενδείξεις αισθητήρων από βιομηχανικό εξοπλισμό. Σε σενάρια πραγματικού κόσμου, η δυναμική φύση των δεδομένων αποτελεί πρόκληση για τις παραδοσιακές μεθόδους ανίχνευσης ανωμαλιών. Οι αυτοκωδικοποιητές, με την ικανότητά τους να προσαρμόζονται στα μεταβαλλόμενα πρότυπα, προσφέρουν μια λύση. Μαθαίνουν συνεχώς και ενημερώνουν τις αναπαραστάσεις τους, καθιστώντας τους αποτελεσματικούς σε σενάρια όπου οι τακτικές απάτης εξελίσσονται με την πάροδο του χρόνου (Bank, Koenigstein, & Giryes, 2023).

Ενώ οι αυτόματοι κωδικοποιητές έχουν επιδείξει επιτυχία στην ανίχνευση ανωμαλιών, είναι σημαντικό να ρυθμίζεται προσεκτικά το μοντέλο και να ορίζεται ένα κατάλληλο κατώφλι για την επισήμανση των ανωμαλιών. Επιπλέον, η ερμηνευσιμότητα των αναπαραστάσεων που μαθαίνονται μπορεί να αποτελέσει πρόκληση, καθιστώντας κρίσιμη την ενσωμάτωση της εμπειρογνομοσύνης του τομέα για την κατανόηση του πλαισίου των ανωμαλιών που ανιχνεύονται (Michelucci, 2022).

Συμπερασματικά, οι αυτοκωδικοποιητές έχουν αναδειχθεί ως ένα ισχυρό εργαλείο για την ανίχνευση ανωμαλιών, ιδίως σε σενάρια ανίχνευσης απάτης. Η μη επιβλεπόμενη φύση της μάθησής τους, η προσαρμοστικότητά τους σε μεταβαλλόμενα πρότυπα και η ικανότητά τους να χειρίζονται διάφορους τύπους δεδομένων τους καθιστούν πολύτιμο πλεονέκτημα για τη διασφάλιση των συστημάτων από αναδυόμενες απειλές και παρατυπίες. Καθώς ο τομέας της τεχνητής νοημοσύνης συνεχίζει να εξελίσσεται, οι αυτοκωδικοποιητές είναι πιθανό να διαδραματίσουν καθοριστικό ρόλο στην ενίσχυση των δυνατοτήτων ανίχνευσης ανωμαλιών σε διάφορους τομείς.

Κεφάλαιο 4: Πραγματικές εφαρμογές της TN στην ανίχνευση απάτης

Ανίχνευση απάτης με πιστωτικές κάρτες:

Οι αλγόριθμοι TN χρησιμοποιούνται εκτενώς στα συστήματα ανίχνευσης απάτης με πιστωτικές κάρτες. Τα συστήματα αυτά αναλύουν μοτίβα συναλλαγών, τη συμπεριφορά των χρηστών και ιστορικά δεδομένα για τον εντοπισμό ανώμαλων δραστηριοτήτων. Τα μοντέλα μηχανικής μάθησης, συμπεριλαμβανομένων των νευρωνικών δικτύων και των δέντρων αποφάσεων, μπορούν να προσαρμόζονται γρήγορα σε νέα μοτίβα δόλιων συναλλαγών, παρέχοντας προστασία σε πραγματικό χρόνο στους καταναλωτές και τα χρηματοπιστωτικά ιδρύματα (Jain, Agrawal, & Kumar, 2020).

Πρόληψη της απάτης στο ηλεκτρονικό εμπόριο:

Οι διαδικτυακές αγορές αντιμετωπίζουν συνεχείς απειλές από δόλιες δραστηριότητες, όπως η κατάληψη λογαριασμών, οι ψεύτικες κριτικές και η απάτη πληρωμών. Τα συστήματα που βασίζονται στην τεχνητή νοημοσύνη αναλύουν τη συμπεριφορά των χρηστών, τα δεδομένα συναλλαγών και τα δακτυλικά αποτυπώματα των συσκευών για τον εντοπισμό ασυνήθιστων μοτίβων. Χρησιμοποιώντας μοντέλα μηχανικής μάθησης, τα συστήματα αυτά μπορούν να εντοπίζουν δυνητικά δόλιες συναλλαγές και να τις αποκλείουν πριν ολοκληρωθούν, εξασφαλίζοντας μια ασφαλή εμπειρία ηλεκτρονικών αγορών (Bao, Hilary, & Ke, 2022).

Ανίχνευση ασφαλιστικής απάτης:

Ο ασφαλιστικός κλάδος αξιοποιεί την TN για τον εντοπισμό δόλιων απαιτήσεων και τη βελτίωση της εκτίμησης κινδύνου. Τα μοντέλα μηχανικής μάθησης αναλύουν ιστορικά δεδομένα απαιτήσεων, εντοπίζοντας μοτίβα που σχετίζονται με δόλιες απαιτήσεις. Η ανάλυση κειμένου και συναισθήματος μπορεί επίσης να εφαρμοστεί

για τη διερεύνηση μη δομημένων δεδομένων, όπως οι περιγραφές απαιτήσεων, για τον εντοπισμό ασυνειών που μπορεί να υποδηλώνουν δόλια δραστηριότητα (Al-Hashedi, &Magalingam, 2021).

Πρόληψη της απάτης στην υγειονομική περίθαλψη:

Η τεχνητή νοημοσύνη διαδραματίζει κρίσιμο ρόλο στον εντοπισμό δόλιων δραστηριοτήτων στον τομέα της υγειονομικής περίθαλψης. Οι ανωμαλίες στα μοτίβα χρέωσης, οι διπλές απαιτήσεις και η ασυνήθιστη συμπεριφορά των ασθενών μπορούν να εντοπιστούν με τη χρήση αλγορίθμων μηχανικής μάθησης. Με τη συνεχή εκμάθηση από ιστορικά δεδομένα, τα συστήματα TN συμβάλλουν στην ελαχιστοποίηση των οικονομικών απωλειών και στη διασφάλιση της ακεραιότητας των διαδικασιών αποζημίωσης της υγειονομικής περίθαλψης(Bao, Hilary, &Ke, 2022).

Τραπεζικές και χρηματοοικονομικές συναλλαγές:

Οι τράπεζες και τα χρηματοπιστωτικά ιδρύματα χρησιμοποιούν TN για την παρακολούθηση των συναλλαγών για πιθανή απάτη. Τα μοντέλα μηχανικής μάθησης αναλύουν ιστορικά συναλλαγών, συμπεριφορές λογαριασμών και γεωγραφικά μοτίβα για τον εντοπισμό ασυνήθιστων δραστηριοτήτων. Τα προηγμένα συστήματα μπορούν επίσης να ανιχνεύσουν απόπειρες κατάληψης λογαριασμού και μη εξουσιοδοτημένης πρόσβασης μαθαίνοντας από τα πρότυπα συμπεριφοράς των χρηστών (Sarmaetal., 2020).

Καταπολέμηση της νομιμοποίησης εσόδων από παράνομες δραστηριότητες (AML):

Η τεχνητή νοημοσύνη συμβάλλει καθοριστικά στις προσπάθειες καταπολέμησης της νομιμοποίησης εσόδων από παράνομες δραστηριότητες, αυτοματοποιώντας την ανάλυση μεγάλου όγκου χρηματοοικονομικών συναλλαγών για τον εντοπισμό δραστηριοτήτων νομιμοποίησης εσόδων από παράνομες δραστηριότητες. Η επεξεργασία φυσικής γλώσσας (NLP) και τα μοντέλα μηχανικής μάθησης βοηθούν στην εξαγωγή σχετικών πληροφοριών από μη δομημένες πηγές δεδομένων, ενισχύοντας την αποτελεσματικότητα των προγραμμάτων AML(Al-Hashedi, &Magalingam, 2021).

Επαλήθευση ταυτότητας:

Τα συστήματα επαλήθευσης ταυτότητας με τεχνητή νοημοσύνη χρησιμοποιούν αναγνώριση προσώπου, βιομετρικά στοιχεία και ανάλυση συμπεριφοράς για την αυθεντικοποίηση των χρηστών. Τα συστήματα αυτά συγκρίνουν δεδομένα σε πραγματικό χρόνο με ιστορικά αρχεία, διασφαλίζοντας ότι οι χρήστες είναι αυτοί που ισχυρίζονται ότι είναι. Αυτό συμβάλλει στην πρόληψη της κλοπής ταυτότητας και της μη εξουσιοδοτημένης πρόσβασης σε ευαίσθητες πληροφορίες (Bao, Hilary, & Ke, 2022).

Ασφάλεια δικτύων και κυβερνοασφάλεια:

Η τεχνητή νοημοσύνη χρησιμοποιείται στην κυβερνοασφάλεια για τον εντοπισμό και την πρόληψη μη εξουσιοδοτημένης πρόσβασης, παραβιάσεων δεδομένων και απειλών στον κυβερνοχώρο. Οι αλγόριθμοι μηχανικής μάθησης αναλύουν μοτίβα κίνησης δικτύου, συμπεριφορές χρηστών και αρχεία καταγραφής συστημάτων για τον εντοπισμό ανωμαλιών που υποδηλώνουν παραβίαση της ασφάλειας. Τα συστήματα ανίχνευσης εισβολών που βασίζονται σε τεχνητή νοημοσύνη μπορούν να ανταποκρίνονται γρήγορα στις εξελισσόμενες απειλές στον κυβερνοχώρο (Al-Hashedi, & Magalingam, 2021).

Κεφάλαιο 5: Οφέλη και ρίσκα της TN στην ανίχνευση απάτης

5.1 Οφέλη

Η χρήση της Τεχνητής Νοημοσύνης (TN) στην ανίχνευση απάτης προσφέρει πλήθος πλεονεκτημάτων, συμβάλλοντας σε ένα πιο ισχυρό και αποτελεσματικό πλαίσιο ασφάλειας. Ένα σημαντικό πλεονέκτημα είναι η ανίχνευση σε πραγματικό χρόνο. Τα συστήματα Τεχνητής Νοημοσύνης τελευταίας τεχνολογίας παρουσιάζουν απaráμιλλη ικανότητα να επεξεργάζονται ταχύως τα εισερχόμενα δεδομένα, επιτρέποντας τον άμεσο εντοπισμό και τον μετριασμό των νέων απειλών. Αυτή η δυναμική ανταπόκριση όχι μόνο ενισχύει τη συνολική ασφάλεια, αλλά χρησιμεύει επίσης ως απόδειξη της αποτελεσματικότητας των διαδικασιών που βασίζονται στην TN (Sina, 2023).

Ένα άλλο σημαντικό πλεονέκτημα είναι η εγγενής ικανότητα της τεχνητής νοημοσύνης να βελτιώνει τις επιδόσεις της με την πάροδο του χρόνου. Καθώς το σύστημα εκτίθεται σε μεγαλύτερο όγκο δεδομένων, οι προγνωστικές του ικανότητες γίνονται πιο εκλεπτυσμένες. Αυτή η συνεχής διαδικασία μάθησης συμβάλλει στη συνεχή βελτίωση της ικανότητας της TN να ανιχνεύει και να προλαμβάνει δόλιες δραστηριότητες. Επιπλέον, οι περιπτώσεις TN που μοιράζονται γνώση μεταξύ τους ενισχύουν τη συλλογική ευφυΐα του συστήματος. Για παράδειγμα, όταν μια περίπτωση Τεχνητής Νοημοσύνης DataDome εντοπίζει ένα νέο μοτίβο απειλής, οι πληροφορίες αυτές διαδίδονται σε όλες τις περιπτώσεις Τεχνητής Νοημοσύνης DataDome παγκοσμίως, ενισχύοντας τη συλλογική άμυνα κατά των αναδυόμενων απειλών (Chiapponietal., 2022).

Ακόμη, η εφαρμογή της τεχνητής νοημοσύνης στην ανίχνευση απάτης μεταφράζεται σε μείωση των χρόνων αντιδραστικής απόκρισης. Με τα συστήματα TN να ελέγχουν και να εντοπίζουν αποτελεσματικά πιθανές απειλές, μειώνεται η ανάγκη για εκτεταμένη χειροκίνητη έρευνα από τους υπαλλήλους. Κατά συνέπεια, το προσωπικό

μπορεί να κατανέμει το χρόνο του πιο στρατηγικά, εστιάζοντας σε έργα που ωθούν τον οργανισμό προς τα εμπρός αντί να ασχολείται με την αντιδραστική ανάλυση απειλών (Sina, 2023).

5.2 Ρίσκα

Η υιοθέτηση της τεχνητής νοημοσύνης για την ανίχνευση της απάτης, ενώ προσφέρει σημαντικά πλεονεκτήματα, δεν είναι απαλλαγμένη από τους εγγενείς κινδύνους και προκλήσεις. Ένας αξιοσημείωτος κίνδυνος είναι η επιμονή της κοινωνικής απάτης. Παρά την αποτελεσματικότητα της τεχνητής νοημοσύνης στην καταπολέμηση των αυτοματοποιημένων απειλών, το πεδίο της κοινωνικής απάτης, που περιλαμβάνει δραστηριότητες όπως το phishing και η κοινωνική μηχανική, παραμένει δύσκολο να αντιμετωπιστεί μόνο μέσω της τεχνητής νοημοσύνης. Αυτού του είδους οι απειλές δεν είναι αυτοματοποιημένες και η ευαισθησία ενός μόνο υπαλλήλου που πέφτει θύμα τέτοιας απάτης μπορεί δυνητικά να θέσει σε κίνδυνο ολόκληρο τον οργανισμό. Κατά συνέπεια, οι οργανισμοί που χρησιμοποιούν TN για την ανίχνευση απάτης πρέπει να αναγνωρίζουν τη συνεχή ανάγκη για ολοκληρωμένη εκπαίδευση των εργαζομένων για την αποτελεσματική αντιμετώπιση αυτών των κινδύνων κοινωνικής μηχανικής (Bao, Hilary, & Ke, 2022).

Ένας άλλος κίνδυνος που συνδέεται με την ανίχνευση απάτης με βάση την TN είναι το φαινόμενο του "μαύρου κουτιού". Η πολυπλοκότητα των αλγορίθμων TN, ιδίως όταν ενσωματώνονται με μηχανική μάθηση και νευρωνικά δίκτυα που μιμούνται τις λειτουργίες του ανθρώπινου εγκεφάλου, μπορεί να καταστήσει δύσκολη την κατανόηση της εσωτερικής λειτουργίας του συστήματος. Αυτή η έλλειψη διαφάνειας εγείρει ανησυχίες σχετικά με την ερμηνευσιμότητα των διαδικασιών λήψης αποφάσεων. Ωστόσο, οι κορυφαίες λύσεις λογισμικού ανίχνευσης απάτης προσπαθούν να αντιμετωπίσουν αυτή την ανησυχία παρέχοντας στους χρήστες ουσιαστικό έλεγχο στην προσαρμογή των κανόνων, εξασφαλίζοντας ένα βαθμό διαφάνειας και προσαρμοστικότητας (Bao, Hilary, & Ke, 2022).

Τα ψευδώς θετικά αποτελέσματα αποτελούν μια μόνιμη πρόκληση στην ανάπτυξη της TN για την ανίχνευση απάτης. Ενώ τα προηγμένα συστήματα TN μειώνουν

σημαντικά την εμφάνιση ψευδώς θετικών αποτελεσμάτων, η πλήρης εξάλειψη παραμένει ανέφικτη. Οι γνήσιοι χρήστες, ιδίως εκείνοι που χρησιμοποιούν ασυνήθιστα προγράμματα περιήγησης και εικονικά ιδιωτικά δίκτυα (VPN), μπορεί περιστασιακά να αντιμετωπίζουν λανθασμένους αποκλεισμούς. Αναγνωρίζοντας αυτόν τον περιορισμό, οι οργανισμοί πρέπει να βρουν μια ισορροπία μεταξύ των αυστηρών μέτρων πρόληψης της απάτης και της ελαχιστοποίησης της ταλαιπωρίας που προκαλούν τα ψευδώς θετικά αποτελέσματα. Οι καλύτερες λύσεις λογισμικού ανίχνευσης απάτης με τεχνητή νοημοσύνη βελτιώνουν συνεχώς τους αλγορίθμους τους για να ενισχύουν την ακρίβεια και να μειώνουν τα ψευδώς θετικά αποτελέσματα σε ένα αποδεκτό ελάχιστο (Sina, 2023).

Συνοψίζοντας, ενώ η ανίχνευση απάτης με τεχνητή νοημοσύνη επιφέρει σημαντικά οφέλη, είναι σημαντικό να αναγνωρίζονται και να αντιμετωπίζονται οι σχετικοί κίνδυνοι. Μια ολοκληρωμένη προσέγγιση που συνδυάζει τις δυνατότητες TN με τη συνεχή εκπαίδευση των εργαζομένων, τις διαφανείς επιλογές προσαρμογής και τη συνεχή βελτίωση των αλγορίθμων είναι ζωτικής σημασίας για τον μετριασμό των προκλήσεων και τη διασφάλιση της αποτελεσματικής ανίχνευσης απάτης, ελαχιστοποιώντας παράλληλα τον αντίκτυπο των ψευδώς θετικών αποτελεσμάτων.

Κεφάλαιο 6: Δεοντολογικές εκτιμήσεις και προκλήσεις

6.1 Προβληματισμοί σχετικά με την προστασία της ιδιωτικής ζωής και των δεδομένων

Η τεχνολογία της τεχνητής νοημοσύνης έχει γίνει διάχυτη τόσο στον προσωπικό όσο και στον επαγγελματικό τομέα, παρουσιάζοντας πληθώρα επιπτώσεων, οι οποίες δεν είναι όλες θετικές. Ο δυνητικός μετασχηματιστικός ρόλος της ΤΝ στην υγειονομική περίθαλψη, για παράδειγμα, είναι προφανής, με ισχυρισμούς ότι οι αλγόριθμοι ΤΝ θα μπορούσαν να επιταχύνουν την ανάλυση ιατρικών εικόνων ταχύτερα από έναν ανθρώπινο ακτινολόγο (Rajpurkaretal., 2018). Ωστόσο, η υλοποίηση αυτών των υποσχέσεων εξαρτάται από την αντιμετώπιση της αλγοριθμικής μεροληψίας και άλλων ηθικών προκλήσεων για την αποφυγή βλάβης των ασθενών. Ειδικότερα, έχουν εντοπιστεί περιπτώσεις όπου ένα σύστημα ΤΝ, το οποίο χρησιμοποιήθηκε για να συστήσει επακόλουθες υπηρεσίες υγειονομικής περίθαλψης, επέδειξε προκατάληψη παραπέμποντας μαύρους ασθενείς σε χαμηλότερο ποσοστό από τους λευκούς ομολόγους τους, παρά τις παρόμοιες διαγνώσεις (Obermeyeretal., 2019).

Τα τελευταία χρόνια, έχει αυξηθεί κατακόρυφα το ενδιαφέρον των κυβερνήσεων και άλλων φορέων για την τεχνολογία αναγνώρισης προσώπου. Ωστόσο, η τεχνολογία αυτή υπόκειται σε έλεγχο, ιδίως όσον αφορά την αναξιοπιστία της στην αναγνώριση ατόμων που δεν είναι λευκοί άνδρες. Η συνακόλουθη απώλεια της ιδιωτικής ζωής λόγω της αναγνώρισης προσώπου έχει γίνει σημαντική ανησυχία, με παράδειγμα τις περιπτώσεις όπου η κυβέρνηση των ΗΠΑ φέρεται να χρησιμοποιήσει την αναγνώριση προσώπου για την ταυτοποίηση διαδηλωτών κατά τη διάρκεια διαδηλώσεων που σχετίζονται με τον Τζορτζ Φλόιντ. Στο Ντιτρόιτ, η κακή χρήση του λογισμικού αναγνώρισης προσώπου οδήγησε στη λανθασμένη σύλληψη ενός μαύρου άνδρα για ένα έγκλημα που δεν διέπραξε (Allynetal., 2020). Επιπλέον, εκφράζονται ανησυχίες σχετικά με τη δυνητική ενσωμάτωση της αναγνώρισης προσώπου στο σύστημα κοινωνικής βαθμολόγησης της Κίνας, μια προοπτική που πολλοί θεωρούν ηθικά

προβληματική. Σε απάντηση, ομάδες για την ελευθερία των πολιτών τάσσονται υπέρ της ρύθμισης της τεχνολογίας αναγνώρισης προσώπου και αρκετές εταιρείες τεχνολογίας έχουν ανακοινώσει την πρόθεσή τους να μην προσφέρουν αυτή την τεχνολογία σε αστυνομικά τμήματα (Greeneetal., 2020). Παρά τις αντιδράσεις αυτές, πολλά ηθικά διλήμματα που σχετίζονται με την τεχνολογία αναγνώρισης προσώπου παραμένουν άλυτα.

Η διάβρωση της ιδιωτικής ζωής εντείνεται με εργαλεία όπως το Clearview AI, που είναι ικανό να σαρώνει πηγές του Διαδικτύου για τις διαδικτυακές φωτογραφίες ενός ατόμου (Duffy, 2020). Δεδομένης της ελεύθερα διαθέσιμης φύσης πολλών διαδικτυακών πληροφοριών, ο προσδιορισμός του ποιος χρησιμοποιεί τέτοια εργαλεία και για ποιους σκοπούς καθίσταται εξαιρετικά δύσκολος, καθώς δεν υπάρχουν οι διασφαλίσεις που συνήθως υπάρχουν σε φυσικές υποδομές.

Μια άλλη πιεστική ηθική ανησυχία είναι το ζήτημα της εμπιστοσύνης στην τεχνολογία τεχνητής νοημοσύνης. Επιστρέφοντας στο παράδειγμα της ιατρικής απεικόνισης, εάν ένας αλγόριθμος TN καθιερωθεί ως "αξιόπιστος", θα μπορούσε δυνητικά να συμπληρώσει ή και να αντικαταστήσει την ανθρώπινη κρίση. Αυτό επεκτείνεται σε σενάρια όπου ο αλγόριθμος χειρίζεται περιπτώσεις πέραν του αρχικού του σχεδιασμού, που μοιάζουν με την "εκτός ετικέτας" χρήση ιατρικών προϊόντων. Σε άλλες περιπτώσεις, οι χρήστες μπορεί να εμπιστευτούν υπερβολικά την TN, περιμένοντας από αυτήν να παρέχει καθοδήγηση πέρα από τον προγραμματισμό της, όπως στις αλληλεπιδράσεις με τα chatbots της θεραπείας. Η υπερεκτίμηση της ικανότητας της τεχνητής νοημοσύνης να μετριάσει τη βλάβη μπορεί να οδηγήσει σε λανθασμένες υποθέσεις, όπως η υπόθεση ότι ένας ρομποτικός εξωσκελετός θα παρέχει προειδοποιήσεις σε επικίνδυνες καταστάσεις, ενώ δεν διαθέτει τέτοιο χαρακτηριστικό (Borensteinetal., 2020).

Αυτά τα παραδείγματα αντιπροσωπεύουν μόνο ένα κλάσμα των ηθικών ζητημάτων που περιβάλλουν την TN (Howard, &Borenstein, 2020). Παρά τις προκλήσεις αυτές, η χρήση της TN συνεχίζει να επεκτείνεται. Κατά συνέπεια, μια παθητική προσέγγιση, με την προσδοκία ότι τα ηθικά προβλήματα θα επιλυθούν μόνα τους, είναι αβάσιμη. Αντίθετα, είναι επιτακτική η ανάγκη μιας μελετητικής και προληπτικής στάσης, η οποία θα εστιάζει στην ανάπτυξη όχι μόνο τεχνικά ικανών εφαρμογών TN, αλλά και

στην καθιέρωση ηθικά ορθών πρακτικών και πολιτικών που θα διέπουν τη χρήση τους.

6.2 Προσπάθειες για την αντιμετώπιση των ηθικών προκλήσεων της Τεχνητής Νοημοσύνης

Έχουν αναδειχθεί πολυάριθμες πρωτοβουλίες για την αντιμετώπιση των ηθικών προκλήσεων που συνδέονται με την τεχνολογία της τεχνητής νοημοσύνης, με διάφορους ενδιαφερόμενους φορείς, όπως ακαδημαϊκά ιδρύματα, κυβερνητικούς φορείς, μη κυβερνητικές οργανώσεις (ΜΚΟ) και παράγοντες της βιομηχανίας να συμβάλλουν ενεργά στη συζήτηση. Τα ακαδημαϊκά ιδρύματα, με παράδειγμα τη Διακήρυξη του Μόντρεαλ, πρωτοστάτησαν στις προσπάθειες για τη διαμόρφωση υπεύθυνης ανάπτυξης της ΤΝ (Gibert, Christophe, Guillaume, 2018). Επαγγελματικοί οργανισμοί, όπως το Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE), έχουν αναλάβει τη σύνταξη έκθεσης για την ηθική των ευφυών συστημάτων και αναπτύσσουν παράλληλα μια σειρά τεχνικών προτύπων για τα συστήματα αυτά. Πολλές εταιρείες, όπως η Google και η Deloitte, τονίζουν δημοσίως τη δέσμευσή τους σε ηθικά ζητήματα, όπως η δικαιοσύνη και η διαφάνεια, μέσω δελτίων τύπου και άλλων ανακοινώσεων. Αυτή η παγκόσμια προσπάθεια έχει δημιουργήσει ένα σημαντικό αποθετήριο εγγράφων ηθικής της τεχνητής νοημοσύνης, το οποίο προκάλεσε λεπτομερείς αναλύσεις (Hagendorff, 2020). Ωστόσο, ο απτός αντίκτυπος αυτών των εγγράφων στην υποκίνηση αλλαγών, είτε με τη μορφή νέων κανονισμών είτε με τη μορφή βιομηχανικών πρακτικών, παραμένει αβέβαιος.

Στις πρόσφατες εξελίξεις περιλαμβάνεται η ίδρυση οργανισμών όπως οι Partnership on AI και AI Now, καθώς και συνεδρίων όπως το ACM FAccT, όλα ρητά προσανατολισμένα στην αντιμετώπιση ζητημάτων που σχετίζονται με τη δεοντολογία της ΤΝ. Φορείς χρηματοδότησης, όπως το Εθνικό Ίδρυμα Επιστημών (NSF), έχουν αφιερώσει υποστήριξη σε πρωτοβουλίες που διερευνούν τη δικαιοσύνη, τη δεοντολογία, τη λογοδοσία και τη διαφάνεια (FEAT) στους τομείς της πληροφορικής, υπογραμμίζοντας την αυξανόμενη σημασία της ηθικής της ΤΝ (National Science Foundation, 2019).

Παρά αυτά τα προληπτικά μέτρα που επιχειρούν να αντιμετωπίσουν τις ηθικές προκλήσεις της ΤΝ, το βασικό ζήτημα παραμένει - αυτό της ανθρώπινης αστοχίας και των συναφών ελλείψεων που επηρεάζουν τον σχεδιασμό και τη χρήση της

τεχνολογίας. Ενώ είναι απίθανο τα άτομα να σχεδιάζουν σκόπιμα ΤΝ με κακόβουλη πρόθεση ή να επιδιώκουν μεροληπτικά αποτελέσματα, δεν τονίζεται ομοιόμορφα η συνεπής ενδοσκόπηση που απαιτείται για τον εντοπισμό των προσωπικών προκαταλήψεων και αξιών που ενσωματώνονται στην τεχνολογία. Η ενθάρρυνση των ατόμων να εξετάζουν κριτικά τις προκαταλήψεις και τις αξίες τους, προάγοντας την αυτογνωσία κατά το σχεδιασμό και την ανάπτυξη της ΤΝ, παραμένει μια θεμελιώδης αλλά και δύσκολη πτυχή της συνεχιζόμενης συζήτησης γύρω από την ηθική της ΤΝ.

6.3 Προώθηση επαγγελματικής νοοτροπίας

Ο συνεχιζόμενος διάλογος υπογραμμίζει ότι πολλά άτομα και οργανισμοί προτείνουν προληπτικά διορθωτικά μέτρα για την αντιμετώπιση των ηθικών προκλήσεων που προκύπτουν από την τεχνητή νοημοσύνη. Ωστόσο, ο εντοπισμός και η εφαρμογή βιώσιμων λύσεων, είτε τεχνικών είτε άλλων, παραμένουν τρομερά καθήκοντα. Μια ουσιώδης πτυχή για τη διαλεύκανση αυτού του πολύπλοκου ζητήματος έγκειται στην εμπέδωση στους προγραμματιστές μιας οξείας επίγνωσης ότι η τεχνολογία που κατασκευάζουν είναι εγγενώς συνυφασμένη με ηθικές διαστάσεις. Οι προγραμματιστές, υπό την επαγγελματική τους ιδιότητα, φέρουν κρίσιμο ρόλο και ευθύνη να ασχοληθούν ενεργά με ηθικές εκτιμήσεις (NationalScienceFoundation, 2019).

Η καθιέρωση μιας αυθεντικής επαγγελματικής νοοτροπίας αρχίζει με την καλλιέργεια ηθικής ευαισθησίας μεταξύ των προγραμματιστών. Αυτό προϋποθέτει την ικανότητα να αναγνωρίζουν ότι η λήψη επαγγελματικών αποφάσεων, συμπεριλαμβανομένων των τεχνικών αποφάσεων, είναι εγγενώς συνυφασμένη με ηθικούς προβληματισμούς. Η απόρριψη της τεχνολογίας ως "ουδέτερης ως προς τις αξίες" αποκρύπτει την πραγματικότητα ότι τα ηθικά ζητήματα είναι θεμελιωδώς ριζωμένα στην επιλογή, το σχεδιασμό, την ανάπτυξη και τη χρήση της τεχνολογίας. Για παράδειγμα, μια φαινομενικά αβλαβής επιλογή που κάνει ο δημιουργός μιας εφαρμογής γνωριμιών, όπως η προσφορά μόνο μιας δυαδικής επιλογής για το φύλο του χρήστη, αντανακλά μια απόφαση που είναι φορτισμένη με αξίες. Όταν η τεχνητή νοημοσύνη ενσωματώνεται στη συνέχεια για να εντοπίσει το καλύτερο ταίρι του χρήστη, δαιμονίζεται ένα σύστημα με εγγενείς προκαταλήψεις που είναι συνυφασμένες με το σχεδιασμό του (Gibert, Christophe, Guillaume, 2018).

Ένα στενά συνδεδεμένο σημείο περιστρέφεται γύρω από την αντίληψη των επαγγελματιών ευθυνών εντός της κοινότητας της TN. Συχνά, οι προγραμματιστές μπορεί να τρέφουν την πεποίθηση, η οποία ενίοτε ενισχύεται μέσω του προγράμματος σπουδών STEM και άλλων καναλιών, ότι η ηθική είναι ένα θέμα που πρέπει να αντιμετωπίσει κάποιος άλλος. Αυτή η προοπτική μπορεί να εκδηλωθεί ως διαχωρισμός των ανησυχιών, με τους προγραμματιστές να σκέφτονται: "Εμείς χειριζόμαστε την τεχνολογία, αφήστε τους δικηγόρους ή τους ηθικολόγους να επιλύσουν τις ηθικές ανησυχίες". Ωστόσο, οι επιλογές που γίνονται κατά τη διάρκεια της διαδικασίας σχεδιασμού δεν έχουν μόνο ηθικές προεκτάσεις αλλά αντικατοπτρίζουν και τις ηθικές αξίες του σχεδιαστή. Αυτές οι επιλογές, είτε πρόκειται για την ιατρική απεικόνιση είτε για την αξιολόγηση της υποτροπής, δεν επηρεάζουν μόνο την τεχνολογία αλλά διαμορφώνουν επίσης τις ζωές των ατόμων και την κοινωνία γενικότερα (Howard, & Borenstein, 2020).

Ο παραλληλισμός με τον ιατρικό τομέα, όπου οι γιατροί δεσμεύονται να τηρούν τον όρκο του Ιπποκράτη, χρησιμεύει ως αξιοσημείωτο παράδειγμα. Αν και ο επαγγελματικός όρκος δεν αποτελεί πανάκεια, αντιπροσωπεύει τη δέσμευση σε ένα κοινωνικό συμβόλαιο μεταξύ ενός επαγγέλματος και του κοινού. Ακόμη και αν δεν διατυπώνεται ρητά, ο όρκος του Ιπποκράτη χρησιμεύει ως υπενθύμιση της ηθικής υποχρέωσης των ιατρών να ενισχύουν τη δημόσια υγεία. Στο πεδίο της τεχνητής νοημοσύνης, όπου οι τεχνολογίες προσδίδουν τόσο οφέλη όσο και βλάβες στο κοινό, τίθεται ένα κρίσιμο ερώτημα: ποιες ηθικές ευθύνες θα πρέπει να αναμένονται από εκείνους που αναπτύσσουν την τεχνολογία; Θα πρέπει οι ευθύνες αυτές να είναι λιγότερο σημαντικές; Ένα κρίσιμο βήμα προς τα εμπρός περιλαμβάνει την ενδυνάμωση των προγραμματιστών TN και της ευρύτερης υπολογιστικής κοινότητας ώστε να κατανοήσουν πλήρως και να ενστερνιστούν αυτές τις ηθικές ευθύνες (Gibert, Christophe, Guillaume, 2018).

6.4 Το μέλλον της ηθικής της Τεχνητής Νοημοσύνης

Η μετασχηματιστική επιρροή της TN στη ζωή του ανθρώπου εκτυλίσσεται με τρόπους που συχνά είναι περίπλοκοι και δύσκολο να προβλεφθούν. Για να οδηγηθεί αυτή η τεχνολογία προς μια πιο κοινωνικά υπεύθυνη πορεία, αναδύεται μια κρίσιμη επιτακτική ανάγκη - η ανάγκη να αφιερωθεί σημαντικός χρόνος και προσοχή στην εκπαίδευση σε θέματα ηθικής της TN. Πέρα από τη σημασία της σταθερής

ενσωμάτωσης της ηθικής στον πυρήνα της ταυτότητάς της από την κοινότητα των υπολογιστών, υπάρχει και μια ρεαλιστική διάσταση, με την εμφάνιση θέσεων εργασίας που εστιάζουν ειδικά στην ηθική της ΤΝ. Ο Lewis (2020) τάσσεται υπέρ της πιθανής υιοθέτησης ενός επικεφαλής υπεύθυνου δεοντολογίας τεχνητής νοημοσύνης εντός των εταιρειών, τονίζοντας την ανάγκη ειλικρινούς δέσμευσης για την ιεράρχηση της δεοντολογίας και όχι για επιφανειακό "πλύσιμο δεοντολογίας".

Η προοπτική να γίνει η ηθική της τεχνητής νοημοσύνης ένας εξειδικευμένος ρόλος εντός των οργανισμών υπογραμμίζει την αυξανόμενη σημασία της ενσωμάτωσης των ηθικών προβληματισμών στον ιστό της ανάπτυξης και της ανάπτυξης της τεχνητής νοημοσύνης. Η πρόταση του Lewis για έναν επικεφαλής υπεύθυνο ηθικής της τεχνητής νοημοσύνης σηματοδοτεί μια στρατηγική κίνηση προς την κατεύθυνση της προώθησης μιας γνήσιας δέσμευσης στις ηθικές αρχές εντός των εταιρειών. Ωστόσο, για να διασφαλιστεί ότι τέτοιες πρωτοβουλίες συμβάλλουν πραγματικά σε ηθικές πρακτικές και δεν χρησιμεύουν απλώς ως βιτρίνα, ένα καίριο βήμα περιλαμβάνει την ανάδειξη της ηθικής σε κεντρική θέση στις εκπαιδευτικές πρωτοβουλίες για την τεχνητή νοημοσύνη.

Με την ενσωμάτωση της ηθικής στον πυρήνα των εκπαιδευτικών προσπαθειών για την τεχνητή νοημοσύνη, υπάρχει μια απτή οδός για την εμπέδωση μιας ολοκληρωμένης κατανόησης των ηθικών προβληματισμών στον εκκολαπτόμενο τομέα της τεχνητής νοημοσύνης. Η προσέγγιση αυτή όχι μόνο ευθυγραμμίζεται με την επιτακτική ανάγκη να ενσωματωθεί η κοινωνική ευθύνη στην ΤΝ, αλλά δημιουργεί επίσης τα θεμέλια για τα άτομα που εισέρχονται στο εργατικό δυναμικό ώστε να ασχοληθούν ενεργά με τις ηθικές προκλήσεις. Με τον τρόπο αυτό, η ενσωμάτωση της ηθικής στην εκπαίδευση της τεχνητής νοημοσύνης μπορεί να χρησιμεύσει ως προληπτικό μέτρο, διαμορφώνοντας μια γενιά επαγγελματιών της τεχνητής νοημοσύνης οι οποίοι δεν είναι μόνο ικανοί να αξιοποιούν τις τεχνολογικές εξελίξεις, αλλά έχουν επίσης επίγνωση των ηθικών διαστάσεων που ενυπάρχουν στην εργασία τους.

Συμπεράσματα

Εν κατακλείδι, η παρούσα διατριβή αξιολογεί κριτικά την εφαρμογή και τη σημασία της τεχνητής νοημοσύνης (TN) στις διαδικασίες ανίχνευσης και διερεύνησης της απάτης, υπογραμμίζοντας τόσο τις πιθανές εξελίξεις όσο και τις εγγενείς προκλήσεις που συνδέονται με την ενσωμάτωση τεχνολογιών TN σε αυτόν τον τομέα. Οι παραδοσιακές μέθοδοι ανίχνευσης απάτης, συμπεριλαμβανομένων των χειροκίνητων διαδικασιών, των συστημάτων που βασίζονται σε κανόνες και των στατιστικών μεθόδων, εξετάζονται διεξοδικά για τους περιορισμούς τους, υπογραμμίζοντας την ανάγκη για πιο εξελιγμένες προσεγγίσεις. Η διερεύνηση των εννοιών της TN, όπως η μηχανική μάθηση, η βαθιά μάθηση και η επεξεργασία φυσικής γλώσσας, θέτει τις βάσεις για μια πιο προσεκτική εξέταση συγκεκριμένων τεχνικών TN που εφαρμόζονται στην ανίχνευση απάτης.

Ενώ οι τεχνικές TN είναι πολλά υποσχόμενες, η διατριβή αξιολογεί τις πρακτικές τους επιπτώσεις. Η συζήτηση σχετικά με τις πραγματικές εφαρμογές της TN στην ανίχνευση απάτης αποκαλύπτει περιπτώσεις όπου οι εν λόγω τεχνολογίες έχουν παίξει καθοριστικό ρόλο. Ωστόσο, είναι ζωτικής σημασίας να αναγνωριστούν τα υφιστάμενα κενά και το ενδεχόμενο ακούσιων συνεπειών, ιδίως σε περιπτώσεις όπου η αλγοριθμική προκατάληψη και τα ψευδώς θετικά αποτελέσματα ενδέχεται να θέσουν σε κίνδυνο την αξιοπιστία των συστημάτων TN. Αναγνωρίζονται τα οφέλη της τεχνητής νοημοσύνης στην ανίχνευση απάτης, αλλά εφαρμόζεται ένας κριτικός φακός για τους συναφείς κινδύνους. Η κοινωνική απάτη, η αλγοριθμική μεροληψία και η επίμονη πρόκληση των ψευδώς θετικών αποτελεσμάτων αναγνωρίζονται ως σημαντικές ανησυχίες που απαιτούν συνεχή προσοχή και στρατηγικές μετριασμού.

Εξετάζονται οι ηθικοί προβληματισμοί κατά την εφαρμογή της τεχνητής νοημοσύνης για την ανίχνευση απάτης, με έμφαση στις ανησυχίες για την προστασία της ιδιωτικής ζωής και των δεδομένων. Αναγνωρίζονται οι προσπάθειες για την αντιμετώπιση αυτών των ηθικών προκλήσεων, αλλά η διατριβή υπογραμμίζει την ανάγκη για μια πιο ολοκληρωμένη και προληπτική προσέγγιση στη διαμόρφωση ηθικών πλαισίων. Η

συζήτηση σχετικά με την προώθηση μιας επαγγελματικής νοοτροπίας στην κοινότητα της τεχνητής νοημοσύνης αναγνωρίζει τα βήματα που έχουν γίνει, αλλά τονίζει τις συνεχιζόμενες προκλήσεις και την επιτακτική ανάγκη για συνεχή επαγρύπνηση.

Η μελλοντική πορεία της ηθικής της τεχνητής νοημοσύνης εξετάζεται με μια προσεκτική προοπτική, αναγνωρίζοντας ότι καθώς οι τεχνολογίες της τεχνητής νοημοσύνης εξελίσσονται, ενδέχεται να προκύψουν νέες ηθικές προκλήσεις. Η διατριβή ενθαρρύνει μια κριτική στάση απέναντι στο εξελισσόμενο τοπίο της ΤΝ στην ανίχνευση της απάτης, τονίζοντας τη σημασία του συνεχούς ηθικού ελέγχου, των κανονιστικών πλαισίων και της διαφάνειας.

Συμπερασματικά, η παρούσα διατριβή υπογραμμίζει τις μετασχηματιστικές δυνατότητες της ΤΝ στην ανίχνευση απάτης, διατηρώντας παράλληλα μια κριτική προοπτική όσον αφορά τους περιορισμούς, τους κινδύνους και τα ηθικά ζητήματα. Καθώς οι τεχνολογίες τεχνητής νοημοσύνης συνεχίζουν να εξελίσσονται, είναι επιτακτική ανάγκη να πλοηγηθούμε σε αυτές τις εξελίξεις με σύνεση, προωθώντας υπεύθυνες πρακτικές που ευθυγραμμίζονται με τα ηθικά πρότυπα και τις κοινωνικές αξίες. Τα συμπεράσματα που προσφέρονται στο παρόν συμβάλλουν στην κατανόηση της πολύπλευρης αλληλεπίδρασης μεταξύ της ΤΝ, της ανίχνευσης απάτης και των ηθικών προβληματισμών, ανοίγοντας το δρόμο για τεκμηριωμένες συζητήσεις και μελλοντικές ερευνητικές προσπάθειες σε αυτό το δυναμικό πεδίο.

Βιβλιογραφία

- Albrecht, W. S. (2003). *FrmtedE.ramination*. U.S.: Thomson South-Western.
- Alharbi, A., Alshammari, M., Okon, O. D., Alabrah, A., Rauf, H. T., Alyami, H., et al. (2022). A Novel text2IMG Mechanism of Credit Card Fraud Detection: A Deep Learning Approach. *Electronics, 11*.
- Al-Hashedi, K. G., &Magalingam, P. (2021). Financial fraud detection applying data mining techniques: A comprehensive review from 2009 to 2019. *Computer Science Review, 40*, 100402.
- Allyn, B. (2020). 'The Computer Got It Wrong': how facial recognition led to false arrest of black man. NPR. URL: <https://www.npr.org/2020/06/24/882683463/the-computer-got-it-wrong-how-facial-recognition-led-to-a-false-arrest-in-michig>.
- Alpaydin, E. (2021). *Machine learning*. Mit Press.
- Association of Certified Fraud Examiners. (2020). Report to the Nations: 2020 Global Study on Occupational Fraud and Abuse.
- Aziz, L. A. R., &Andriansyah, Y. (2023). The Role Artificial Intelligence in Modern Banking: An Exploration of AI-Driven Approaches for Enhanced Fraud Prevention, Risk Management, and Regulatory Compliance. *Reviews of Contemporary Business Analytics, 6*(1), 110-132.
- Bank, D., Koenigstein, N., &Giryes, R. (2023). Autoencoders. *Machine learning for data science handbook: data mining and knowledge discovery handbook, 353-374*.
- Bao, Y., Hilary, G., & Ke, B. (2022). Artificial intelligence and fraud detection. *Innovative Technology at the Interface of Finance and Operations: Volume I, 223-247*.
- Barbariol, T., &Susto, G. A. (2022). TiWS-iForest: Isolation forest in weakly supervised and tiny ML scenarios. *Information Sciences, 610*, 126-143.
- Barcin, T., Yucel, M. A., Ersan, R. H., Alagoz, M. A., Dogen, A., Burmaoglu, S., et al. (2024). Deep learning approach to the discovery of novel bisbenzazole derivatives for antimicrobial effect. *Journal of Molecular Structure, 1295*.
- Bhasin, M. L. (2016). The role of technology in combatting bank frauds: perspectives and prospects. *Ecoforum Journal, 5*(2).
- Bolton, R. J., & Hand, D. J. (2002). Statistical fraud detection: A review. *Statistical science, 17*(3), 235-255.
- Borenstein, J., Mahajan, H. P., Wagner, A. R., & Howard, A. (2020). Trust and pediatric exoskeletons: a comparative study of clinician and parental perspectives.

IEEE Transactions on Technology and Society, 1(2), 83–88. DOI: <https://doi.org/10.1109/TTS.2020.2974990>.

Britannica. (2023). artificial intelligence. <https://www.britannica.com/technology/artificial-intelligence>

Cabanes, Y., Barbaresco, F., Arnaudon, M., & Bigot, J. (2019, June). Non-supervised machine learning algorithms for radar clutter high-resolution Doppler segmentation and pathological clutter analysis. In *2019 20th International Radar Symposium (IRS)* (pp. 1-10). IEEE.

Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., ... & Zdeborová, L. (2019). Machine learning and the physical sciences. *Reviews of Modern Physics*, 91(4), 045002.

Chaquet-Ulldemolins, J., Gimeno-Blanes, F.-J., Moral-Rubio, S., Muñoz-Romero, S., & Rojo-Álvarez, J.-L. (2022). On the Black-Box Challenge for Fraud Detection Using Machine Learning (II): Nonlinear Analysis through Interpretable Autoencoders. *Applied Sciences*, 12.

Chen, J., Chen, Q., Jiang, F., Guo, X., Sha, K., & Wang, Y. (2024). SCN_GNN: A GNN-based fraud detection algorithm combining strong node and graph topology information. *Expert Systems with Applications*, 237.

Cheng, D., Wang, X., Zhang, Y., & Zhang, L. (2020). Graph neural network for fraud detection via spatial-temporal attention. *IEEE Transactions on Knowledge and Data Engineering*, 34(8), 3800-3813.

Chiapponi, E., Dacier, M., Thonnard, O., Fangar, M., Mattsson, M., & Rigal, V. (2022, June). An industrial perspective on web scraping characteristics and open issues. In *2022 52nd Annual IEEE/IFIP International Conference on Dependable Systems and Networks-Supplemental Volume (DSN-S)* (pp. 5-8). IEEE.

Clements, L. H., & Knudstrup, M. (2016). Which fraud investigation procedures are most often performed? an exploratory study. *Journal of Forensic & Investigative Accounting*, 8(2), 168-178.

Coram, P., Ferguson, C., & Moroney, R. (2006). The importance of internal audit in fraud detection. *Research Journal*.

Duffy, C. (2020). The ACLU sues Clearview AI, calling the tool an 'unprecedented violation' of privacy rights. CNN.com. URL: <https://www.cnn.com/2020/05/28/tech/clearview-ai-aclu-lawsuit/index.html>.

Eisenstein, J. (2019). *Introduction to natural language processing*. MIT press.

Fawcett, T., & Provost, F. (1997). Adaptive fraud detection. *Data mining and knowledge discovery*, 1(3), 291-316.

Gao, Y., Liu, J., Li, W., Hou, M., Li, Y., & Zhao, H. (2023). Augmented Grad-CAM++: Super-Resolution Saliency Maps for Visual Interpretation of Deep Neural Network. *Electronics*, 12.

Gibert, M., Christophe, M., Guillaume, C. (2018). Montréal Declaration of Responsible AI: 2018 Overview of International Recommendations for AI Ethics. University of Montréal. URL: <https://www.montrealdeclaration-responsibleai.com/reports-of-montreal-declaration>.

Gkikas, D. C., Theodoridis, P. K., Theodoridis, T., & Gkikas, M. C. (2023). Finding Good Attribute Subsets for Improved Decision Trees Using a Genetic Algorithm Wrapper; a Supervised Learning Application in the Food Business Sector for Wine Type Classification. *Informatics*, 10.

Greene, J. (2020). Microsoft won't sell police its facial-recognition technology, following similar moves by Amazon and IBM. The Washington Post. URL: <https://www.washingtonpost.com/technology/2020/06/11/microsoft-facial-recognition/>.

Hajek, A. (2012). Interpretations of Probability. In Zalta, E. N., ed., *The Stanford Encyclopedia of Philosophy*. Winter 2012 edition. URL: <http://plato.stanford.edu/archives/win2012/entries/probability-interpret/>,

Hajek, P. (2019, May). Interpretable fuzzy rule-based systems for detecting financial statement fraud. In *IFIP international conference on artificial intelligence applications and innovations* (pp. 425-436). Cham: Springer International Publishing.

Hagendorff, T. (2020). The ethics of AI Ethics: an evaluation of guidelines. *Minds and Machines*. DOI: <https://doi.org/10.1007/s11023-020-09517-8>.

Hassan, S. W. U., Kiran, S., Gul, S., Khatatbeh, I. N., & Zainab, B. (2023). The perception of accountants/auditors on the role of corporate governance and information technology in fraud detection and prevention. *Journal of Financial Reporting and Accounting*.

Hernandez-Orallo, J. (2017). *The Measure of All Minds: Evaluating Natural and Artificial Intelligence*. Cambridge: Cambridge University Press.

Hofstadter, D. R., and FARG. (1995). *Fluid Concepts and Creative Analogies: Computer Models of the Fundamental Mechanisms of Thought*. New York: Basic Books.

Hou, X.-X. (2018). Voxelwise detection of cerebral microbleed in CADASIL patients by leaky rectified linear unit and early stopping. *Multimedia Tools and Applications*, 77, 21825-21845.

Howard, A., & Borenstein, J. (2020). AI, robots, and ethics in the age of COVID-19. MIT Sloan Management Review. URL: <https://sloanreview.mit.edu/article/ai-robots-and-ethics-in-the-age-of-covid-19/>.

Huang, S. Y., Lin, C. C., Chiu, A. A., & Yen, D. C. (2017). Fraud detection using fraud triangle risk factors. *Information Systems Frontiers*, 19, 1343-1356.

Information Systems Audit and Control Association (ISACA). (2004). *IS Standards, Guidelines and Procedures for Auditing and Control Professionals* [online]. Available

from the World Wide Web: (<http://www.isaca.org/ContentGroups/Standards2/StandardsBooklet21july04.pdf>).

Iparraquirre-Villanueva, O., Guevara-Ponce, V., Sierra-Linan, F., Beltozar-Clemente, S., & Cabanillas-Carbonell, M. (2022). Sentiment Analysis of Tweets using Unsupervised Learning Techniques and the K-Means Algorithm. *International Journal of Advanced Computer Science and Applications*, 13.

Jain, V., Agrawal, M., & Kumar, A. (2020, June). Performance analysis of machine learning algorithms in credit cards fraud detection. In *2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)* (pp. 86-88). IEEE.

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: a brief primer. *Behavior Therapy*, 51(5), 675-687.

Kelleher, J. D. (2019). *Deep learning*. MIT press.

Kim, Y., & Vasarhelyi, M. A. (2012). A model to detect potentially fraudulent/abnormal wires of an insurance company: An unsupervised rule-based approach. *Journal of Emerging Technologies in Accounting*, 9(1), 95-110.

Kim, G. B., Kim, J. Y., Lee, J. A., Norsigian, C. J., Palsson, B. O., & Lee, S. Y. (2023). Functional annotation of enzyme-encoding genes using deep learning with transformer layers. *Nature Communications*, 14, 7370.

Kim, E., Lee, J., Shin, H., Yang, H., Cho, S., Nam, S.-k., et al. (2019). Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning. *Expert Systems with Applications*, 128, 214-224.

Kolman, M. R. (1999). What constitutes fraud? *The Internal Auditor* 56 (3): 88.

Kumar, Y. (2022). AI techniques in blockchain technology for fraud detection and prevention. *Security engineering for embedded and cyber-physical systems*, 207-224.

Lewis, N. (2020). Why having a chief AI Officer should matter to HR. SHRM. URL: <https://www.shrm.org/ResourcesAndTools/hr-topics/technology/Pages/Why-Having-a-Chief-AI-Officer-Should-Matter-to-HR.aspx>.

Li, J., Huang, K. Y., Jin, J., & Shi, J. (2008). A survey on statistical methods for health care fraud detection. *Health care management science*, 11, 275-287.

Li, X., Sale, M., Nieforth, K., Bigos, K. L., Craig, J., Wang, F., et al. (2023). pyDarwin: A Machine Learning Enhanced Automated Nonlinear Mixed-effect Model Selection Toolbox. *Clinical Pharmacology & Therapeutics*, 2023.

Li, Z., Huang, M., Liu, G., & Jiang, C. (2021). A hybrid method with dynamic weighted entropy for handling the problem of class imbalance with overlap in credit card fraud detection. *Expert Systems with Applications*, 175.

Liu, Y. (2023). Design of XGBoost prediction model for financial operation fraud of listed companies. *International Journal of System Assurance Engineering and Management*, 14, 2354-2364.

- Liu, Z., Dou, Y., Yu, P. S., Deng, Y., & Peng, H. (2020, July). Alleviating the inconsistency problem of applying graph neural network to fraud detection. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval* (pp. 1569-1572).
- Li, Y., Cao, J., Xu, Y., Zhu, L., & Dong, Z. Y. (2024). Deep learning based on Transformer architecture for power system short-term voltage stability assessment with class imbalance. *Renewable and Sustainable Energy Reviews*, 189.
- Lin, S., Mengaldo, G., & Maulik, R. (2023). Online data-driven changepoint detection for high-dimensional dynamical systems. *Chaos*, 33.
- Maashi, M., Alabduallah, B., & Kouki, F. (2023). Sustainable Financial Fraud Detection Using GarraRufa Fish Optimization Algorithm with Ensemble Deep Learning. *Sustainability*, 15.
- Michelucci, U. (2022). An introduction to autoencoders. *arXiv preprint arXiv:2201.03898*.
- Mignan, A. (2019). A preliminary text classification of the precursory accelerating seismicity corpus: inference on some theoretical trends in earthquake predictability research from 1988 to 2018. *Journal of Seismology*, 23, 771-785.
- National Science Foundation. (2019). Artificial intelligence (AI) at NSF. URL: <https://www.nsf.gov/cise/ai.jsp>.
- Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447–453. DOI: <https://doi.org/10.1126/science.aax2342>.
- O'Kelly, M. (2004). Using statistical techniques to detect fraud: a test case. *Pharmaceutical Statistics: The Journal of Applied Statistics in the Pharmaceutical Industry*, 3(4), 237-246.
- Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.
- Pinaya, W. H. L., Vieira, S., Garcia-Dias, R., & Mechelli, A. (2020). Autoencoders. In *Machine learning* (pp. 193-208). Academic Press.
- Probst, D. (2023). An explainability framework for deep learning on chemical reactions exemplified by enzyme-catalyzed reaction classification. *Journal of Cheminformatics*, 15, 113.
- Rajpurkar, P., Irvin, J., Ball, R.L., Zhu, K., Yang, B., Mehta, H., et al. (2018). Deep learning for chest radiograph diagnosis: a retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med*, 15(11), e1002686. DOI: <https://doi.org/10.1371/journal.pmed.1002686>.
- Rubaidi, Z. S., Ammar, B. B., & Aouicha, M. B. (2022). Fraud Detection Using Large-scale Imbalance Dataset. *International Journal on Artificial Intelligence Tools*, 31.

- Sarkar, M. R., Anavatti, S. G., Dam, T., Ferdous, M. M., Tahtali, M., Ramasamy, S., et al. (2024). GATE: A guided approach for time series ensemble forecasting. *Expert Systems with Applications*, 235.
- Sarma, D., Alam, W., Saha, I., Alam, M. N., Alam, M. J., & Hossain, S. (2020, July). Bank fraud detection using community detection algorithm. In *2020 second international conference on inventive research in computing applications (ICIRCA)* (pp. 642-646). IEEE.
- Seelam, S. A. R. (2023). Comparative Analytics on Chilli Plant Disease using Machine Learning Techniques.
- Sina, A. (2023). Open AI and its Impact on Fraud Detection in Financial Industry. *Journal of Knowledge Learning and Science Technology ISSN: 2959-6386 (Online)*, 2(3), 263-281.
- Sisodia, D., & Sisodia, D. S. (2023). A transfer learning framework towards identifying behavioral changes of fraudulent publishers in pay-per-click model of online advertising for click fraud detection. *Expert Systems with Applications*, 232.
- Statement on Auditing Standards (SAS) No. 99- Consideration of Fraud in a Financial Statement Audit.
- van Zetten, W., Ramackers, G. J., & Hoos, H. H. (2022). Increasing trust and fairness in machine learning applications within the mortgage industry. *Machine Learning with Applications*, 10.
- Vo, K., Jonnagaddala, J., & Liaw, S. T. (2019). Statistical supervised meta-ensemble algorithm for medical record linkage. *Journal of Biomedical Informatics*, 95, 103220.
- Wang, Y., Liu, Y., Zhao, J., & Zhang, Q. (2023). A Low-Complexity Fast CU Partitioning Decision Method Based on Texture Features and Decision Trees. *Electronics*, 12.
- Wang, S.-H., & Fernandes, S. (2022). AVNC: Attention-based VGG-style network for COVID-19 diagnosis by CBAM. *IEEE Sensors Journal*, 22, 17431-17438.
- Wang, L., Ye, M., Lu, Y., Qiu, Q., Niu, Z., Shi, H., et al. (2023). A combined encoder-transformer-decoder network for volumetric segmentation of adrenal tumors. *Biomedical Engineering Online*, 22, 106.
- Webster's Ninth New Collegiate Dictionary. (1990). Springfield, MA: Merriam-Webster.
- Wells, J. T. (2005). Principles of Fraud Examination. Hoboken, NJ: John Wiley & Sons.
- West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, 57, 47-66.
- Wu, X., Jiang, S., Li, G., Liu, S., Metcalfe, B., Chen, L., et al. (2023). Deep Learning With Convolutional Neural Networks for Motor Brain-Computer Interfaces Based on

Stereo-Electroencephalography (SEEG). *IEEE Journal of Biomedical and Health Informatics*, 27, 2387-2398.

Xu, B., Wang, Y., Liao, X., & Wang, K. (2023). Efficient fraud detection using deep boosting decision trees. *Decision Support Systems*, 175.

Yang, S., Varghese, P., Stephenson, E., Tu, K., & Gronsbell, J. (2023). Machine learning approaches for electronic health records phenotyping: A methodical review. *Journal of the American Medical Informatics Association*, 30, 367-381.

Yang, G. R., & Wang, X. J. (2020). Artificial neural networks for neuroscientists: a primer. *Neuron*, 107(6), 1048-1070.

Zhang, Y. (2010). Color Image Enhancement based on HVS and PCNN. *SCIENCE CHINA Information Sciences*, 53, 1963-1976.

Zhang, Y.-D. (2021). MIDCAN: A multiple input deep convolutional attention network for Covid-19 diagnosis based on chest CT and chest X-ray. *Pattern Recognition Letters*, 150, 8-16.

Zhang, Y. (2023). Deep learning in food category recognition. *Information Fusion*, 98, 101859.