2026-01

þÿ § Á ® Ã · Ä · Â ¤ µ Ç ½ · Ä ® Â • ¿ · ¼ ¿ Ã Î

þÿ Ä · ½ ‘ ½ ¯ Ç ½ µ Å Ã · R e d F l a g s Ã

þÿ ˜ µ ¿ ´ Ì Ã ¹ ¿ Â , ‘ ½ Ä É ½ Ì À ¿ Å » ¿ Â

þÿ œ µ Ä ± À Ä Å Ç ¹ ± º Ì Á Ì ³ Á ± ¼ ¼ ± Ã Ä · ½ • ³ º » · ¼ ± Ä ¿ » ¿ ³ ¹ º ® › ¿ ³ ¹ Ã Ä ¹ º ® º ± ¹ Ä ± § Á · ¼ ±

þÿ £ Ç ¿ » ® Ÿ ¹ º ¿ ½ ¿ ¼ ¹ º Î ½ • À ¹ Ã Ä · ¼ Î ½ º ± ¹ ” ¹ ¿ ¯ º · Ã · Â , ± ½ µ À ¹ Ã Ä ® ¼ ¹ ¿ • µ ¬ À ¿ » ¹ Â

ΣΧΟΛΗ
Οικονομικών Επιστημών

ΤΜΗΜΑ
Λογιστικής και Χρηματοοικονομικής

Χρήση της Τεχνητής Νοημοσύνης για την Ανίχνευση "Red Flags" σε Οργανισμούς

Διατριβή η οποία υποβλήθηκε προς απόκτηση εξ αποστάσεως μεταπτυχιακού τίτλου σπουδών στην Εγκληματολογική Λογιστική στο Πανεπιστήμιο Νεάπολις

Αντωνόπουλος Θεοδόσιος
Επιβλέπουσα καθηγήτρια : κα Καταραχιά Ανδρονίκη

ΙΑΝΟΥΑΡΙΟΣ 2026

Πνευματικά δικαιώματα

Ονοματεπώνυμο Φοιτητή: Αντωνόπουλος Θεοδόσιος
Τίτλος Μεταπτυχιακής Διατριβής: Χρήση της Τεχνητής Νοημοσύνης για την Ανίχνευση "Red Flags" σε Οργανισμούς
Η παρούσα Μεταπτυχιακή Διατριβή εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση εξ αποστάσεως μεταπτυχιακού τίτλου στο Πανεπιστήμιο Νεάπολις και εγκρίθηκε
στις …… από τα μέλη της Εξεταστικής Επιτροπής.
Εξεταστική Επιτροπή:
Πρώτος επιβλέπων: Καταραχιά Ανδρονίκη
Μέλος Εξεταστικής Επιτροπής:
Μέλος Εξεταστικής Επιτροπής:

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

Ο Αντωνόπουλος Θεοδόσιος, γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο «Χρήση της Τεχνητής Νοημοσύνης για την Ανίχνευση “Red Flags” σε Οργανισμούς», 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

Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει τον ρόλο της Τεχνητής Νοημοσύνης (ΤΝ) στην ανίχνευση red flags και στη διαχείριση κινδύνων απάτης, εστιάζοντας στις τεχνολογικές, ελεγκτικές και κανονιστικές διαστάσεις. Αρχικά παρουσιάζεται η εξέλιξη της ΤΝ και οι σύγχρονες εφαρμογές της στον εσωτερικό έλεγχο, ενώ αναλύονται διεξοδικά τα μοντέλα μηχανικής μάθησης, τα συστήματα επεξεργασίας φυσικής γλώσσας και οι αλγόριθμοι ανίχνευσης ανωμαλιών που χρησιμοποιούνται για την αναγνώριση ύποπτων μοτίβων σε λογιστικά και λειτουργικά δεδομένα. Στη συνέχεια εξετάζονται τα κριτήρια επιλογής και αξιολόγησης των αλγορίθμων, οι δείκτες απόδοσης, καθώς και τα ζητήματα ερμηνευσιμότητας, διαφάνειας και συμμόρφωσης με το κανονιστικό πλαίσιο, συμπεριλαμβανομένων του GDPR και του AI Act. Κεντρικό συμπέρασμα της εργασίας είναι ότι, παρά τη σημαντική συμβολή της ΤΝ στην αύξηση της αποτελεσματικότητας και της ακρίβειας των διαδικασιών ελέγχου, η ανθρώπινη εποπτεία παραμένει αναντικατάστατη. Η μελέτη καταλήγει ότι η ΤΝ δεν αντικαθιστά τον εσωτερικό έλεγχο αλλά τον ενισχύει, υπό την προϋπόθεση ότι εφαρμόζεται σε περιβάλλον με σαφείς μηχανισμούς εποπτείας, διαφάνειας και λογοδοσίας. Η επιτυχής ενσωμάτωσή της προϋποθέτει στοχευμένη επιλογή μοντέλων, επαρκή εκπαίδευση των ανθρώπων που τα χρησιμοποιούν και συνεχή παρακολούθηση της λειτουργίας τους. Σε αυτό το πλαίσιο, η συνεργασία ανθρώπου και μηχανής αναδεικνύεται ως κρίσιμος παράγοντας για τη βιώσιμη και ασφαλή αξιοποίηση της ΤΝ στην ανίχνευση των red flags και στη διασφάλιση της ακεραιότητας των οργανισμών.

Λέξεις κλειδιά: τεχνητή νοημοσύνη, ανίχνευση απάτης, red flags, ανθρώπινη εποπτεία

Abstract

This master's thesis examines the role of Artificial Intelligence (AI) in the detection of red flags and the management of fraud-related risks, focusing on technological, auditing, and regulatory dimensions. Initially, the evolution of AI and its contemporary applications in internal auditing are presented, followed by an analysis of machine learning models, natural language processing systems, and anomaly detection algorithms used to identify suspicious patterns in accounting and operational data. Subsequently, the study explores the criteria for selecting and evaluating algorithms, performance metrics, and issues related to interpretability, transparency, and compliance with the regulatory framework, including the GDPR and the AI Act. A key finding of the thesis is that, despite the significant contribution of AI to improving the efficiency and accuracy of audit processes, human oversight remains indispensable. The study concludes that AI does not replace internal auditing but enhances it, provided that it is implemented within a framework characterized by clear mechanisms of supervision, transparency, and accountability. Successful integration requires careful model selection, adequate training of the professionals involved, and continuous monitoring of system performance. Within this context, human–machine collaboration emerges as a critical factor for the sustainable and secure use of AI in red flag detection and in safeguarding organizational integrity.

Keywords: artificial intelligence, fraud detection, red flags, human supervision

Ευχαριστίες

Θα ήθελα να εκφράσω τις πιο ειλικρινείς μου ευχαριστίες στην οικογένειά μου, η οποία στάθηκε δίπλα μου σε όλη τη διάρκεια των σπουδών μου και της συγγραφής της παρούσας διπλωματικής εργασίας. Η υπομονή, η στήριξη και η εμπιστοσύνη τους αποτέλεσαν για εμένα σταθερό σημείο αναφοράς και πολύτιμη πηγή δύναμης. Χωρίς την καθημερινή τους ενθάρρυνση και την αμέριστη κατανόησή τους, η ολοκλήρωση αυτής της προσπάθειας δεν θα ήταν εφικτή. Σε αυτούς οφείλω ένα μεγάλο μέρος της πορείας μου και τους ευχαριστώ βαθιά για όλα όσα μου έχουν προσφέρει. Επίσης, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στην επιβλέπων καθηγήτρια μου, κυρία Καταραχιά και την κυρία Καραγιαννοπούλου, για την καθοδήγηση την υποστήριξη τους σε όλη τη διάρκεια της εκπόνησης της διπλωματικής μου.

ΠΕΡΙΛΗΨΗ V

ABSTRACT VI

ΕΥΧΑΡΙΣΤΙΕΣ VII

ΕΙΣΑΓΩΓΗ I

ΚΕΦΑΛΑΙΟ 1 - ΕΝΝΟΙΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ ΕΣΩΤΕΡΙΚΟΥ ΕΛΕΓΧΟΥ ΚΑΙ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ 1

1.1 Ορισμος του Εσωτερικου Ελεγχου 1
1.1.1 Ανεξαρτησία και Αντικειμενικότητα 2
1.1.2 Αξιολόγηση Διαδικασιών 2
1.1.3 Διαχείριση Κινδύνων 3
1.1.4 Συμμόρφωση 3
1.1.5 Εταιρική Διακυβέρνηση 4
1.1.6 Προσθήκη Αξίας και Συνεχής Βελτίωση 4
1.2 Ορισμος της Απατης 5
1.2.1 Το Τρίγωνο της Απάτης (Fraud Triangle) 5
1.2.2 Το Διαμάντι της Απάτης (Fraud Diamond) 6
1.2.3 Πρόληψη και Αντιμετώπιση της Απάτης 6
1.3 Red Flags: Ορισμοι, κατηγοριες και εφαρμογες 7
1.4 Ορισμοι και Ιστορικη εξελιξη της Τεχνητης Νοημοσυνης 9
1.4.1 Ορισμός της Τεχνητής Νοημοσύνης 9
1.4.2 Ιστορική Εξέλιξη της Τεχνητής Νοημοσύνης 10
1.4.3 Βασικές Κατηγορίες και Τεχνολογίες της ΤΝ 11
1.4.4 Προκλήσεις και Ηθικές Διαστάσεις 12

ΚΕΦΑΛΑΙΟ 2 - ΣΥΓΧΡΟΝΕΣ ΤΑΣΕΙΣ ΣΤΗΝ ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ (ΑΙ) ΚΑΙ ΤΟΝ ΕΣΩΤΕΡΙΚΟ ΕΛΕΓΧΟ 13

2.1 Συγχρονες μελετες για την εφαρμογη της Τεχνητης Νοημοσυνης στον εσωτερικο ελεγχο 13
2.1.1 Cognitive Auditing και Predictive Analytics 14
2.2 Ερμηνευσιμοτητα και ζητηματα εμπιστοσυνης στην Τεχνητη Νοημοσυνη 14
2.2.1 Η έννοια της ερμηνευσιμότητας στην ΤΝ 15
2.2.2 Ζητήματα εμπιστοσύνης σε συστήματα ΤΝ 15
Εμπιστοσύνη στην τεχνολογία 15
Εμπιστοσύνη στον οργανισμό 16
Ηθική εμπιστοσύνη 16
Εμπιστοσύνη στη διαδικασία της λήψης αποφάσεων 16
Αλγοριθμικη μεροληψια 16
2.2.3 Η Explainable AI (XAI) ως λύση στο πρόβλημα εμπιστοσύνης 16
2.2.4 Ο ρόλος της ανθρώπινης εποπτείας (human-in-the-loop) 17
2.2.5 Το πρόβλημα των Hallucinations στα συστήματα ΤΝ και οι επιπτώσεις στην ανίχνευση red flags 17
2.3 Υβριδικα μοντελα AI–Human στον εσωτερικο ελεγχο 18
2.4 Οι "Agents" στην Τεχνητη Νοημοσυνη και ο Ρολος τους στην Ανιχνευση Red Flags 21

ΚΕΦΑΛΑΙΟ 3 - ΜΟΝΤΕΛΑ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ – ΕΙΔΗ, ΛΕΙΤΟΥΡΓΙΑ ΚΑΙ ΕΦΑΡΜΟΓΕΣ 23

3.1 Natural Language Processing (NLP) και ο ρολος του στην ανιχνευση απατης και red flags 25
3.2 Μηχανικη Μαθηση (Machine Learning) 26
3.3 Βαθια Μαθηση (Deep Learning) 26
3.4 Τα Δεντρα Αποφασεων (Decision Trees) και η χρηση τους στην ανιχνευση εταιρικων red flags 26

ΚΕΦΑΛΑΙΟ 4 - ΔΙΑΔΙΚΑΣΙΑ ΑΝΑΠΤΥΞΗΣ ΣΥΣΤΗΜΑΤΟΣ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ ΓΙΑ ΤΗΝ ΑΝΙΧΝΕΥΣΗ ΕΤΑΙΡΙΚΩΝ RED FLAGS 28

4.1 Προσδιορισμος Περιοχων Κινδυνου και Χαρτογραφηση Red Flags 28

4.2 Δεδομενα, πηγες και ποιοτητα πληροφοριας στην ανιχνευση Red Flags με ΤΝ 29
4.3 Επιλογη μοντελων Τεχνητης Νοημοσυνης στην ανιχνευση Red Flags 31
4.4 Εσωτερικη η διαδικτυακη λειτουργια συστηματων ΤΝ για την ανιχνευση red flags: πλεονεκτηματα, μειονεκτηματα και επιλογη 32
4.4.1 Εσωτερική λειτουργία (On-Premise Systems) *32*
4.4.2 Διαδικτυακή λειτουργία (Cloud-Based Systems) *34*
4.4.3 Υβριδική προσέγγιση *35*
4.5 Εκπαιδευση και αξιολογησης Μοντελων ΤΝ 35
Ανάγκη συνεχούς επανεκπαίδευσης *36*
4.6 Δεικτες αξιολογησης μοντελων ΤΝ για ανιχνευση εταιρικων red flags 36
4.7 Ζητηματα δεοντολογιας, GDPR και Governance στα συστηματα ΤΝ για ανιχνευση εταιρικων Red Flags 37
4.7.1 Προκαταλήψεις (bias), διαφάνεια και ανθρώπινη εποπτεία *38*
4.7.2 GDPR: Προσωπικά δεδομένα και αρχές επεξεργασίας *38*
4.7.3 AI Act και υποχρεώσεις για τα συστήματα υψηλού κινδύνου *39*
4.7.4 Σύνδεση με τον ελεγκτικό κίνδυνο *39*
4.8 Διασφαλιση της ερμηνειας των αποτελεσματων και της νομιμοτητας στα συστηματα ΤΝ ανιχνευσης Red Flags 39
4.8.1 Η ανάγκη για ερμηνευσιμότητα στα συστήματα red-flag detection *40*
4.8.2 Τεχνικές Explainable AI (XAI) και η εφαρμογή τους στον εσωτερικό έλεγχο *40*
4.8.3 Ερμηνεία αποτελεσμάτων στο πλαίσιο εσωτερικού ελέγχου *40*
4.8.4 Νομιμότητα και λογοδοσία (accountability) *40*
4.9 Παρακολουθηση λειτουργιας και συντηρηση συστηματων ΤΝ στην ανιχνευση εταιρικων Red Flags 41
4.9.1 Η ανάγκη συνεχούς παρακολούθησης των μοντέλων *41*
4.9.2 Κατηγορίες monitoring: performance, data και concept drift *41*
α) Performance monitoring *41*
β) Data drift monitoring *41*
γ) Concept drift monitoring *42*
4.9.3 Διαδικασίες συνεχούς συντήρησης (maintenance cycle) *42*
4.9.4 Επανεκπαίδευση (retraining) *42*
4.9.5 Αναβάθμιση μοντέλου (model updates) *42*
ΣΥΜΠΕΡΑΣΜΑΤΑ **42**
ΒΙΒΛΙΟΓΡΑΦΙΑ **44**

Εισαγωγή

Τα τελευταία χρόνια και ειδικά στο πλαίσιο της ψηφιακής εποχής που διανύουμε, η Τεχνητή Νοημοσύνη (Artificial Intelligence – AI) έχει καταστεί αναπόσπαστο κομμάτι της καθημερινότητας του ανθρώπου, με εφαρμογή σε πολλαπλά διαφορετικά πεδία. Υπό αυτό το πρίσμα, η ταχεία ανάπτυξη της Τεχνητής Νοημοσύνης έχει, μεταξύ άλλων, αναδείξει νέες δυνατότητες στον χώρο του εσωτερικού ελέγχου και της εγκληματολογικής λογιστικής.

Στο σύγχρονο επιχειρηματικό περιβάλλον, οι οργανισμοί καλούνται να διαχειριστούν αυξημένους κινδύνους που αφορούν την απάτη, την παραποίηση των χρηματοοικονομικών καταστάσεων και την παραβίαση των κανονιστικών πλαισίων. Οι παραδοσιακές ελεγκτικές τεχνικές, οι οποίες στηρίζονται κυρίως σε δειγματοληπτικούς ελέγχους, αποδεικνύονται συχνά ανεπαρκείς στην ανίχνευση σύνθετων ή καλά κρυμμένων μοτίβων απάτης. Η εισαγωγή της ΤΝ, και ειδικότερα των αλγορίθμων μηχανικής μάθησης (Machine Learning – ML) και βαθιάς μάθησης (Deep Learning – DL), επιτρέπει την ανάλυση τεράστιων όγκων δεδομένων με μεγάλη ταχύτητα και ακρίβεια, αλλά το σημαντικότερο, σε πραγματικό χρόνο.

Παράλληλα, η δυνατότητα αξιοποίησης τόσο δομημένων (structured) όσο και αδόμητων (unstructured) δεδομένων, όπως οικονομικές καταστάσεις, αρχεία ERP, ακόμα και κείμενα, δημιουργεί νέες ευκαιρίες στον εντοπισμό “red flags”. Τα “red flags” αποτελούν κρίσιμους δείκτες πιθανής απάτης ή παρατυπιών, και η έγκαιρη αναγνώρισή τους ενισχύει σημαντικά την αποτελεσματικότητα του εσωτερικού ελέγχου.

Ωστόσο, η εφαρμογή της ΤΝ στον τομέα αυτό δημιουργεί και πολλές σημαντικές προκλήσεις. Η ερμηνευσιμότητα (Explainability) των αλγορίθμων, η συμμόρφωση με το κανονιστικό πλαίσιο (π.χ. GDPR), καθώς και η ανάγκη να διατηρηθεί η ανθρώπινη κρίση ως μέρος του ελέγχου, συνιστούν ζητήματα που απαιτούν ιδιαίτερη προσοχή.

Η παρούσα εργασία φιλοδοξεί να διερευνήσει πώς τα συστήματα ΤΝ μπορούν να συμβάλουν στην ανίχνευση των “red flags”, συγκρίνοντας την αποτελεσματικότητά τους με τις παραδοσιακές μεθόδους, και να προτείνει ένα πλαίσιο ενσωμάτωσης της τεχνολογίας αυτής στον εσωτερικό έλεγχο.

Κεφάλαιο 1 - Εννοιολογική προσέγγιση εσωτερικού ελέγχου και Τεχνητής Νοημοσύνης

1.1 Ορισμός του Εσωτερικού Ελέγχου

Ο εσωτερικός έλεγχος είναι μια ανεξάρτητη και αντικειμενική λειτουργία που προσφέρει τόσο διασφάλιση όσο και συμβουλευτικές υπηρεσίες. Βασικός του σκοπός είναι να προσθέτει αξία και να βελτιώνει τις λειτουργίες ενός οργανισμού (Καπετάνιου, 2023). Δεν πρόκειται απλώς για έναν έλεγχο «τυπικότητας» — είναι μια οργανωμένη και πειθαρχημένη διαδικασία που βοηθά τον οργανισμό να πετύχει τους στόχους του, αξιολογώντας και ενισχύοντας την αποτελεσματικότητα της διαχείρισης κινδύνων, των εσωτερικών ελέγχων και της εταιρικής διακυβέρνησης (Λογοθέτης, 2019). Μέσω μιας συστηματικής και πειθαρχημένης προσέγγισης, αξιολογεί κρίσιμους τομείς όπως είναι η συμμόρφωση με το θεσμικό πλαίσιο, συμβάλλοντας έτσι ουσιαστικά στη βελτίωση των διαδικασιών και της λογοδοσίας (Γιαννάκης, 2018). Με απλά λόγια, ο εσωτερικός έλεγχος είναι σαν ένας εσωτερικός σύμβουλος που εξετάζει πώς λειτουργεί η επιχείρηση, εντοπίζει τι μπορεί να γίνει καλύτερα και προτείνει λύσεις, ώστε να βοηθήσει την επιχείρηση να λειτουργεί πιο αποδοτικά, πιο υπεύθυνα και με μεγαλύτερη ασφάλεια. Γίνεται συνεπώς κατανοητό ότι ο εσωτερικός έλεγχος αποτελεί έναν από τους βασικότερους μηχανισμούς διακυβέρνησης, διαφάνειας και αποτελεσματικής λειτουργίας κάθε οργανισμού δεδομένου ότι λειτουργεί ως μια συστηματική, ανεξάρτητη και αντικειμενική διαδικασία αξιολόγησης με στόχο να παρέχει διαβεβαίωση στη διοίκηση ότι οι λειτουργίες, οι διαδικασίες και τα συστήματα της επιχείρησης λειτουργούν αποδοτικά και σύμφωνα με τους κανονισμούς και τους στόχους της. Όπως αναφέρει χαρακτηριστικά το Institute of Internal Auditors (IIA), ο εσωτερικός έλεγχος δεν περιορίζεται απλώς στον εντοπισμό λαθών ή παρατυπιών· αποτελεί μια προστιθέμενη αξία στη λειτουργία του οργανισμού, συμβάλλοντας στη βελτίωση της εταιρικής διακυβέρνησης, στη διαχείριση κινδύνων και στη στρατηγική λήψη αποφάσεων (IIA, 2017).

Λαμβάνοντας υπόψη την έννοια, τα χαρακτηριστικά και τον ρόλο του εσωτερικού ελέγχου συνάγεται ότι τέσσερις κρίσιμοι παράγοντες μπορούν να επηρεάσουν την αποτελεσματικότητα αυτού:

- η ανεξαρτησία των ελεγκτών
- η επάρκεια και η εξειδίκευση του προσωπικού
- η ποιότητα των ελεγκτικών διαδικασιών
- η υποστήριξη από την ανώτατη διοίκηση (Γιαννάκης, 2018).

Ωστόσο, η εφαρμογή του εσωτερικού ελέγχου στην πράξη αντιμετωπίζει και προκλήσεις, όπως η έλλειψη πληροφοριακών συστημάτων, η υποστελέχωση και η απουσία καταγεγραμμένων διαδικασιών. Υπό αυτό το πρίσμα, η τεχνολογία μπορεί να λειτουργήσει ως καταλύτης, ενισχύοντας την ακρίβεια και την ταχύτητα των ελέγχων, αρκεί να συνοδεύεται από κατάλληλη εκπαίδευση και υποστήριξη.

Τι περιλαμβάνει ο Εσωτερικός Έλεγχος

Η δομή και το περιεχόμενο του εσωτερικού ελέγχου μπορούν να κατανοηθούν μέσα από έξι κύριους άξονες:

1.1.1 Ανεξαρτησία και Αντικειμενικότητα

Η ανεξαρτησία αποτελεί θεμέλιο του εσωτερικού ελέγχου, καθώς εξασφαλίζει ότι οι ελεγκτές μπορούν να εκτελούν τα καθήκοντά τους χωρίς επιρροές, συγκρούσεις συμφερόντων ή πιέσεις από τη διοίκηση. Σύμφωνα με τον Λογοθέτη (2019), η ανεξαρτησία συνδέεται άρρηκτα με την αντικειμενικότητα: ο εσωτερικός ελεγκτής δεν πρέπει να εμπλέκεται στις λειτουργικές δραστηριότητες που αξιολογεί, ώστε να διατηρεί αμερόληπτη κρίση. Αυτό επιτρέπει την ύπαρξη μιας «καθαρής οπτικής» των διαδικασιών, χωρίς προσωπικές προκαταλήψεις ή διοικητικές επιρροές.

Η ανεξαρτησία έχει δύο διαστάσεις:

I. Οργανωτική ανεξαρτησία, που διασφαλίζει ότι το τμήμα εσωτερικού ελέγχου αναφέρεται απευθείας στο διοικητικό συμβούλιο ή στην επιτροπή ελέγχου, και όχι στη διοίκηση.
II. Προσωπική ανεξαρτησία, που αναφέρεται στην ακεραιότητα, την επαγγελματική ηθική και την ικανότητα του ελεγκτή να εκφέρει γνώμη χωρίς φόβο ή εύνοια.

Όπως υπογραμμίζει το COSO (Committee of Sponsoring Organizations of the Treadway Commission, 2013), χωρίς ανεξαρτησία ο εσωτερικός έλεγχος χάνει τη λειτουργική του αξία, καθώς η αντικειμενικότητα είναι αυτή που καθιστά τα πορίσματά του αξιόπιστα. Παράλληλα, η ανεξαρτησία δεν σημαίνει απομόνωση: οι ελεγκτές οφείλουν να συνεργάζονται με τα διοικητικά στελέχη, αλλά να διατηρούν καθαρό διαχωρισμό ρόλων και ευθυνών.

1.1.2 Αξιολόγηση Διαδικασιών

Ένα από τα βασικότερα καθήκοντα του εσωτερικού ελέγχου είναι η αξιολόγηση των διαδικασιών της επιχείρησης. Ο ελεγκτής εξετάζει όλες τις επιμέρους λειτουργίες -από τις λογιστικές και διοικητικές μέχρι τις λειτουργικές και τεχνολογικές- προκειμένου να διαπιστώσει εάν αυτές επιτελούνται αποτελεσματικά, αποδοτικά και σύμφωνα με τις πολιτικές της εταιρείας. Η αξιολόγηση αυτή περιλαμβάνει τη μελέτη της διαχείρισης κινδύνων, της πρόληψης απάτης, της τήρησης διαδικασιών και της αποδοτικότητας των συστημάτων. Για παράδειγμα, σε ένα περιβάλλον χρηματοοικονομικής δραστηριότητας, ο εσωτερικός έλεγχος ελέγχει αν οι μηχανισμοί πληρωμών και λογιστικών εγγραφών λειτουργούν σωστά, αν υπάρχουν δικλείδες ασφαλείας κατά της απάτης και αν οι διαδικασίες εγκρίσεων είναι τεκμηριωμένες (Καπετάνιου, 2023).
Πέραν της επιφανειακής εξέτασης, ο σύγχρονος εσωτερικός έλεγχος υιοθετεί μια προσέγγιση βασισμένη στον κίνδυνο (risk-based approach), αξιολογώντας πρώτα τις περιοχές που παρουσιάζουν μεγαλύτερη πιθανότητα σφάλματος ή παρατυπίας (IIA, 2017). Αυτή η προσέγγιση επιτρέπει στους ελεγκτές να κατανέμουν αποτελεσματικά τους πόρους τους και να εστιάζουν στις κρίσιμες λειτουργίες που επηρεάζουν περισσότερο τη στρατηγική της επιχείρησης.

1.1.3 Διαχείριση Κινδύνων

Η διαχείριση κινδύνων βρίσκεται στον πυρήνα του εσωτερικού ελέγχου. Οι ελεγκτές δεν λειτουργούν μόνο ως «επιθεωρητές», αλλά και ως σύμβουλοι που βοηθούν τον οργανισμό να εντοπίζει, να αναλύει και να αντιμετωπίζει πιθανούς κινδύνους πριν αυτοί εξελιχθούν σε προβλήματα. Όπως επισημαίνει η Καπετάνιου (2023), οι εσωτερικοί ελεγκτές παρέχουν διαβεβαίωση ότι οι μηχανισμοί διαχείρισης κινδύνου λειτουργούν αποτελεσματικά, ενώ προτείνουν βελτιώσεις όπου αυτό απαιτείται.

Η διαδικασία αυτή περιλαμβάνει:

- τον εντοπισμό των κινδύνων, όπως είναι οι οικονομικοί, οι επιχειρησιακοί, οι τεχνολογικοί και οι κανονιστικοί κίνδυνοι,
- την ανάλυση και την αξιολόγηση των κινδύνων, με βάση την πιθανότητα εμφάνισης τους και το μέγεθος των συνεπειών τους,
- την εφαρμογή ελεγκτικών μηχανισμών για τη μείωση ή την εξάλειψη των κινδύνων, και
- την παρακολούθηση και την αναθεώρηση των μηχανισμών αυτών σε τακτά χρονικά διαστήματα.

Το πλαίσιο COSO προσφέρει ένα διεθνώς αναγνωρισμένο πρότυπο που οι εσωτερικοί ελεγκτές χρησιμοποιούν για να συνδέσουν τη διαχείριση κινδύνων με τη στρατηγική και τη λήψη αποφάσεων. Μέσω αυτού του πλαισίου, ο εσωτερικός έλεγχος συμβάλλει στη διαμόρφωση μιας κουλτούρας πρόληψης και όχι απλώς αντίδρασης.

1.1.4 Συμμόρφωση

Η συμμόρφωση με το νομοθετικό και κανονιστικό πλαίσιο είναι απαραίτητη για τη βιωσιμότητα κάθε οργανισμού. Ο εσωτερικός έλεγχος διαδραματίζει κρίσιμο ρόλο στη διασφάλιση ότι οι δραστηριότητες της επιχείρησης πραγματοποιούνται σύμφωνα με τους νόμους, τους κανονισμούς, τις εσωτερικές πολιτικές και τα πρότυπα δεοντολογίας (Καπετάνιου, 2023).

Στην πράξη αυτό σημαίνει ότι οι ελεγκτές εξετάζουν αν η εταιρεία συμμορφώνεται με:

- τις εθνικές και ευρωπαϊκές νομοθεσίες (όπως GDPR, φορολογικοί κανόνες, περιβαλλοντικές ρυθμίσεις),
- τους κώδικες δεοντολογίας και τις εσωτερικές πολιτικές,
- τις διεθνείς λογιστικές πρακτικές και τα πρότυπα χρηματοοικονομικής αναφοράς.

Η συμμόρφωση, ωστόσο, δεν αποτελεί απλώς «έλεγχο τήρησης κανόνων». Ο εσωτερικός έλεγχος λειτουργεί και ως μηχανισμός διασφάλισης της φήμης και της εμπιστοσύνης, καθώς προλαμβάνει νομικές κυρώσεις και αρνητική δημοσιότητα που μπορούν να πλήξουν τον οργανισμό (IIA, 2017). Η συμμόρφωση είναι, συνεπώς, στοιχείο εταιρικής υπευθυνότητας και καλής διακυβέρνησης.

1.1.5 Εταιρική Διακυβέρνηση

Η εταιρική διακυβέρνηση αφορά το σύστημα κανόνων, αρχών και διαδικασιών με βάση τα οποία διοικείται και ελέγχεται ένας οργανισμός. Ο εσωτερικός έλεγχος συμβάλλει ουσιαστικά στη διαφάνεια, τη λογοδοσία και τη χρηστή διοίκηση, λειτουργώντας ως σύνδεσμος μεταξύ της διοίκησης, των μετόχων και της επιτροπής ελέγχου.

Όπως αναφέρει το πλαίσιο του OECD (Organisation for Economic Co-operation and Development, 2015), η καλή εταιρική διακυβέρνηση απαιτεί μηχανισμούς ελέγχου που διασφαλίζουν ότι η διοίκηση δρα προς το συμφέρον της επιχείρησης και των μετόχων. Επομένως, ο εσωτερικός έλεγχος αποτελεί αναπόσπαστο μέρος αυτού του συστήματος καθώς ενισχύει τη διαφάνεια παρέχοντας αντικειμενικές αναφορές προς τη διοίκηση, ενθαρρύνει τη λογοδοσία καταγράφοντας ευρήματα και εισηγήσεις, και προωθεί την ακεραιότητα και την εμπιστοσύνη στο σύνολο της εταιρικής κουλτούρας. Στην πράξη, ο ρόλος του εσωτερικού ελέγχου δεν περιορίζεται σε τεχνικά ζητήματα· αποτελεί συγχρόνως εργαλείο στρατηγικής διακυβέρνησης που βοηθά τη διοίκηση να λαμβάνει τεκμηριωμένες αποφάσεις, βασισμένες σε πραγματικά δεδομένα και όχι σε εικασίες (Λογοθέτης, 2019).

1.1.6 Προσθήκη Αξίας και Συνεχής Βελτίωση

Ο τελικός στόχος του εσωτερικού ελέγχου είναι η προσθήκη αξίας στον οργανισμό. Ο ελεγκτής δεν είναι απλώς ελεγκτής λαθών, αλλά σύμβουλος βελτίωσης. Μέσα από τις διαπιστώσεις του προτείνει διορθωτικά μέτρα, αναδιοργάνωση διαδικασιών και αξιοποίηση νέων τεχνολογιών με στόχο την αύξηση της αποδοτικότητας και τη μείωση κόστους (Λογοθέτης, 2019).

Η προσθήκη αξίας επιτυγχάνεται μέσω της βελτιστοποίησης των διαδικασιών και του ανασχεδιασμού τους ώστε να μειώνονται οι περιττές γραφειοκρατικές λειτουργίες, μέσω της αναγνώρισης ευκαιριών για αυτοματοποίηση από την υιοθέτηση καινούργιων πληροφοριακών συστημάτων (π.χ. data analytics), μέσω της προώθησης μιας εταιρικής κουλτούρας ηθικής και διαφάνειας και, τέλος, μέσω της ενίσχυσης της αποτελεσματικότητας των λειτουργιών, μέσα από την παρακολούθηση δεικτών απόδοσης (KPIs).

Σύμφωνα με το IIA (2017), ο εσωτερικός έλεγχος προσθέτει αξία όταν βοηθά την επιχείρηση να επιτυγχάνει τους στρατηγικούς της στόχους, να αξιοποιεί τους πόρους της με σύνεση και να καλλιεργεί περιβάλλον υπευθυνότητας και συνεχούς μάθησης. Με αυτόν τον τρόπο, ο εσωτερικός έλεγχος δεν θεωρείται πλέον «μηχανισμός ελέγχου», αλλά μοχλός ανάπτυξης.

Συνοψίζοντας, ο εσωτερικός έλεγχος περιλαμβάνει ένα ευρύ φάσμα δραστηριοτήτων που δεν περιορίζονται μόνο στην ανίχνευση παρατυπιών. Συνδέεται με τη στρατηγική, τη διαχείριση κινδύνων, την ηθική κουλτούρα και τη βελτίωση της αποτελεσματικότητας. Η ανεξαρτησία του τον καθιστά αξιόπιστο· η αξιολόγηση διαδικασιών τον καθιστά αποτελεσματικό· η διαχείριση κινδύνων τον καθιστά αναγκαίο· η συμμόρφωση τον καθιστά θεματοφύλακα της νομιμότητας· η εταιρική διακυβέρνηση τον καθιστά θεμέλιο της διαφάνειας· και η προσθήκη αξίας τον καθιστά πολύτιμο στρατηγικό εταίρο για κάθε οργανισμό.

Αυτό που αξίζει εδώ να σημειωθεί είναι ότι ο ρόλος αυτός, σύμφωνα με τις διεθνείς προδιαγραφές (IIA, COSO, OECD), είναι διαρκώς εξελισσόμενος. Στο σύγχρονο επιχειρηματικό περιβάλλον, ο εσωτερικός έλεγχος πρέπει να προσαρμόζεται στις

τεχνολογικές αλλαγές, στους κανονιστικούς μετασχηματισμούς και στις αυξημένες απαιτήσεις για διαφάνεια. Έτσι, ο εσωτερικός έλεγχος παραμένει ένας ζωντανός μηχανισμός που συνδυάζει έλεγχο, συμβουλευτική και στρατηγική διορατικότητα. Λειτουργεί εν ολίγοις ως ένας απαραίτητος πυλώνας βιώσιμης εταιρικής λειτουργίας.

1.2 Ορισμός της Απάτης

Η απάτη αποτελεί ένα από τα πιο κρίσιμα ζητήματα που απασχολούν τις σύγχρονες επιχειρήσεις, καθώς μπορεί να οδηγήσει σε σημαντικές οικονομικές απώλειες, απώλεια αξιοπιστίας και υπονόμευση της εταιρικής κουλτούρας. Σύμφωνα με τη Φαρμάκη (2023), η απάτη ορίζεται ως σκόπιμη ενέργεια ή παράλειψη, με στόχο την απόκτηση αθέμιτου οικονομικού ή άλλου οφέλους εις βάρος ενός τρίτου μέρους, μέσω μεθόδων όπως η παραπλάνηση, η παραποίηση στοιχείων, η απόκρυψη πληροφοριών ή η παραβίαση της σχέσης εμπιστοσύνης μεταξύ τους. Ο ορισμός αυτός ευθυγραμμίζεται με εκείνον της Association of Certified Fraud Examiners (ACFE, 2024), η οποία περιγράφει την απάτη ως «οποιαδήποτε σκόπιμη πράξη ή παράλειψη που αποσκοπεί στην εξαπάτηση άλλου μέρους, με αποτέλεσμα την οικονομική ή άλλη ζημία».

Είναι σημαντικό να τονιστεί ότι η απάτη δεν ταυτίζεται με την αμέλεια ή το ανθρώπινο σφάλμα. Αντίθετα, προϋποθέτει προμελετημένη πρόθεση και συνειδητή απόφαση του δράστη να ενεργήσει με δόλο. Έτσι, η απάτη μπορεί να λάβει πολλές μορφές: οικονομική απάτη (π.χ. παραποίηση οικονομικών καταστάσεων), λειτουργική απάτη (π.χ. ψευδή τιμολόγια), απάτη προμηθειών, απάτη σε συστήματα πληροφορικής, ακόμα και εσωτερική απάτη (internal fraud), όπου οι ίδιοι οι εργαζόμενοι ή τα διοικητικά στελέχη εκμεταλλεύονται την πρόσβασή τους σε ευαίσθητα δεδομένα (Καπετάνιου, 2023).

Η διεθνής βιβλιογραφία καταδεικνύει ότι η απάτη έχει σημαντικό κοινωνικό και οικονομικό κόστος. Σύμφωνα με την ACFE (2024), οι επιχειρήσεις χάνουν κατά μέσο όρο το 5% των εσόδων τους ετησίως λόγω φαινομένων απάτης. Το ποσοστό αυτό, σε παγκόσμια κλίμακα, μεταφράζεται σε εκατοντάδες δισεκατομμύρια δολάρια.

Θεωρίες και Μοντέλα Απάτης

Η κατανόηση των θεωρητικών προσεγγίσεων της απάτης είναι καθοριστικής σημασίας για τον εσωτερικό έλεγχο και, κατ’ επέκταση, για την ανάπτυξη συστημάτων ανίχνευσης red flags μέσω της Τεχνητής Νοημοσύνης (AI). Η βιβλιογραφία έχει αναπτύξει διάφορα μοντέλα που επιχειρούν να εξηγήσουν τους ψυχολογικούς, κοινωνικούς και οργανωσιακούς παράγοντες που ωθούν ένα άτομο να διαπράξει απάτη.

Παρακάτω παρουσιάζονται αδρομερώς δύο από τα βασικότερα μοντέλα ερμηνείας της απάτης, το Τρίγωνο της Απάτης και το Διαμάντι της Απάτης.

1.2.1 Το Τρίγωνο της Απάτης (Fraud Triangle)

Το Τρίγωνο της Απάτης, που προτάθηκε από τον Αμερικανό κοινωνιολόγο Donald Cressey (1953), αποτελεί το πλέον κλασικό και ευρέως αποδεκτό θεωρητικό μοντέλο ερμηνείας της απάτης. Ο Cressey υποστήριξε ότι η απάτη εκδηλώνεται όταν συνυπάρχουν τρεις αναγκαίες συνθήκες:

1. Πίεση (Pressure) – Πρόκειται για την εσωτερική ή εξωτερική ανάγκη που ωθεί ένα άτομο να παραβεί τους κανόνες. Συνήθως σχετίζεται με οικονομικές δυσκολίες, χρέη, υπερβολικές προσδοκίες από τη διοίκηση ή ακόμα και με την επιθυμία διατήρησης κοινωνικού status (Cressey, 1953).
2. Ευκαιρία (Opportunity) – Δημιουργείται όταν ο δράστης έχει τη δυνατότητα να διαπράξει απάτη χωρίς να εντοπιστεί, συνήθως λόγω ανεπαρκών εσωτερικών ελέγχων ή αδυναμιών στα πληροφοριακά συστήματα (Φαρμάκη, 2023).
3. Εξορθολογισμός (Rationalization) – Αναφέρεται στην ψυχολογική διαδικασία μέσω της οποίας ο δράστης δικαιολογεί τις ενέργειές του, θεωρώντας ότι η πράξη του δεν είναι πραγματικά επιζήμια ή ότι είναι προσωρινή και «αναγκαία».

Η θεωρία αυτή αποτελεί τη βάση για τις σύγχρονες στρατηγικές πρόληψης κατά της απάτης, καθώς οι επιχειρήσεις μπορούν να σχεδιάσουν μέτρα ελέγχου που μειώνουν τις ευκαιρίες (μέσω ισχυρών μηχανισμών εποπτείας), περιορίζουν τις πιέσεις (με σωστή διαχείριση των κινήτρων) και ενισχύουν την ηθική κουλτούρα (για να αποδυναμωθεί ο εξορθολογισμός).

1.2.2 Το Διαμάντι της Απάτης (Fraud Diamond)

Οι Wolfe και Hermanson (2004) πρότειναν την εξέλιξη του Τριγώνου της Απάτης, προσθέτοντας ένα τέταρτο στοιχείο, την Ικανότητα (Capability). Το μοντέλο ονόματι Διαμάντι της Απάτης το οποίο εισήγαγαν, αναγνωρίζει ότι, εκτός από την πίεση, την ευκαιρία και τον εξορθολογισμό, χρειάζεται και η δεξιότητα ή η θέση ισχύος του ατόμου για να εκτελέσει και να συγκαλύψει μια απάτη.

Η ικανότητα σχετίζεται με χαρακτηριστικά όπως:

- η γνώση των συστημάτων και των διαδικασιών,
- η πρόσβαση σε κρίσιμες πληροφορίες,
- η εμπειρία στον χειρισμό ελεγκτικών μηχανισμών, και
- η ψυχολογική αντοχή στο να διαπράττει κάποιος πράξεις με υψηλό ρίσκο.

Η θεωρία αυτή καθίσταται ιδιαίτερα χρήσιμη στη σημερινή ψηφιακή εποχή, όπου οι απάτες συχνά εκτελούνται από άτομα με τεχνικές δεξιότητες και πρόσβαση σε ευαίσθητα πληροφοριακά συστήματα. Εδώ είναι που η Τεχνητή Νοημοσύνη αποκτά καίριο ρόλο: μέσω αλγορίθμων ανίχνευσης ανωμαλιών, machine learning μοντέλων και data analytics, μπορεί να εντοπίσει ύποπτα μοτίβα που υποδεικνύουν την ύπαρξη αυτών των «ικανοτήτων» και των «ευκαιριών» σε πραγματικό χρόνο (Καπετάνιου, 2023).

1.2.3 Πρόληψη και Αντιμετώπιση της Απάτης

Η πρόληψη και η αντιμετώπιση της απάτης απαιτούν ολιστική προσέγγιση, που συνδυάζει τις πολιτικές, τους εσωτερικούς ελέγχους και τα σύγχρονα τεχνολογικά εργαλεία. Όπως τονίζει η Φαρμάκη (2023), ο ρόλος του εσωτερικού ελέγχου είναι θεμελιώδης: όχι μόνο για τον εντοπισμό της απάτης αλλά κυρίως για τη δημιουργία περιβάλλοντος εμπιστοσύνης και διαφάνειας.

Η Τεχνητή Νοημοσύνη έχει αρχίσει να μεταμορφώνει τον τρόπο με τον οποίον οι οργανισμοί προσεγγίζουν την πρόληψη απάτης. Μέσω αλγορίθμων μηχανικής μάθησης (machine learning) και ανάλυσης μεγάλων δεδομένων (big data analytics), είναι πλέον δυνατό να εντοπιστούν "red flags", δηλαδή πρότυπα ή συμπεριφορές που αποκλίνουν από το φυσιολογικό και υποδηλώνουν πιθανό δόλο. Ως παραδείγματα μπορούμε να αναφέρουμε την ανάλυση συναλλαγών σε πραγματικό χρόνο για τον εντοπισμό ύποπτων μοτίβων ή την εξόρυξη δεδομένων (data mining) από τιμολόγια, προμήθειες, εσωτερικά emails, ακόμη και την εφαρμογή predictive analytics για την πρόβλεψη κινδύνων απάτης. Τα συστήματα αυτά μπορούν να λειτουργούν είτε προληπτικά, εντοπίζοντας πρώιμα σημάδια κινδύνου, είτε αντιδραστικά, βοηθώντας στη διερεύνηση ήδη υφιστάμενων περιστατικών. Κατά αυτό τον τρόπο, ο ρόλος του εσωτερικού ελέγχου μεταβάλλεται, από την παραδοσιακή «χειροκίνητη» προσέγγιση, σε data-driven auditing, όπου η ανθρώπινη κρίση συμπληρώνεται από την ανάλυση τεχνητής νοημοσύνης (Daniel E. et al., 2025).

Συμπερασματικά, η κατανόηση των θεωριών απάτης παρέχει το θεωρητικό υπόβαθρο πάνω στο οποίο μπορούν να αναπτυχθούν ευφυή συστήματα ανίχνευσης "red flags". Με τη χρήση της Τεχνητής Νοημοσύνης, οι επιχειρήσεις έχουν τη δυνατότητα όχι μόνο να εντοπίζουν απάτες με μεγαλύτερη ακρίβεια αλλά και να δημιουργούν προληπτικά συστήματα ηθικής συμμόρφωσης και διακυβέρνησης. Η σύζευξη θεωρίας, εσωτερικού ελέγχου και τεχνολογίας συνιστά πλέον τη νέα εποχή στην καταπολέμηση της απάτης.

1.3 Red Flags: Ορισμοί, κατηγορίες και εφαρμογές

Η έννοια των "red flags" (κόκκινες σημαίες) έχει πλέον αποκτήσει κεντρική θέση στην πρακτική και θεωρητική προσέγγιση του εσωτερικού ελέγχου, της επίβλεψης των εταιρικών διαδικασιών και της ανίχνευσης απάτης. Όταν χρησιμοποιούμε τον όρο red flag εννοούμε ένα σημάδι, ένα προειδοποιητικό σήμα ή ένδειξη που να υποδεικνύει ότι κάτι δεν πάει καλά και ότι ενδέχεται να απαιτείται περαιτέρω διερεύνηση. Μολονότι η ύπαρξη ενός red flag δεν αποδεικνύει κατ' ανάγκην ότι έχει διαπραχθεί απάτη, λειτουργεί ως πρώιμος δείκτης που αξίζει προσοχής και δεν πρέπει να αγνοείται (Corporate Finance Institute, 2024).

Η θεωρητική αποσαφήνιση του όρου είναι κρίσιμη: σύμφωνα με έρευνα, τα red flags ορίζονται ως «συνθήκες ή περιστάσεις που αποκλίνουν από τη φυσιολογική λειτουργία και μπορούν να επιφέρουν υποψία ή είναι σημάδια κινδύνου ότι έχει ή πρόκειται να διαπραχθεί απάτη» (Munteanu et al., 2024). Επισημαίνεται ότι τα red flags λειτουργούν ως "δείκτες" για τις πρώτες φάσεις μιας απάτης. Για τον λόγο αυτό οι ελεγκτές τα εκλαμβάνουν ως «αποτυπώματα» (fraud fingerprints) που χρήζουν εξερεύνησης και επαλήθευσης (Munteanu et al., 2024).

Στην πράξη, η αξιολόγηση και η κατηγοριοποίηση των red flags αποτελεί βασικό στοιχείο για την ανάπτυξη συστημάτων προειδοποίησης, συστημάτων εσωτερικού ελέγχου και τεχνολογικών εργαλείων ανάλυσης κινδύνου. Σε αυτό το πλαίσιο είναι σκόπιμο να αναλυθούν οι βασικές κατηγορίες των red flags, καθώς και οι εφαρμογές τους τόσο στο χρηματοοικονομικό πεδίο όσο και στο ευρύτερο περιβάλλον εταιρικής διακυβέρνησης και ελέγχου.

Σε μια πρώτη κατηγορία, μπορούμε να διακρίνουμε τα διοικητικά ή οργανωσιακά red flags. Αυτά περιλαμβάνουν ενδείξεις που σχετίζονται με την κουλτούρα της εταιρείας, τον διαχωρισμό καθηκόντων, την εποπτεία της διοίκησης, με ασαφείς διαδικασίες ή με συγκεντρωτική εξουσία. Για παράδειγμα, η έλλειψη ισχυρού συστήματος εσωτερικού

ελέγχου ή η υπερβολικά μεγάλη εξουσία σε ένα άτομο μπορεί να δημιουργεί ευκαιρίες για κατάχρηση (Grabosky P. & Duffield G., 2001). Ουσιαστικά, τα red flags αυτού του τύπου δηλώνουν ότι το πλαίσιο οργάνωσης και διοίκησης μπορεί να επιτρέπει λάθη ή δόλιες πράξεις χωρίς να μπορούν να ανιχνευτούν.

Δεύτερη κατηγορία συνιστούν τα λειτουργικά ή διαδικαστικά red flags. Εδώ εντοπίζονται ενδείξεις σε επίπεδο διαδικασιών, συστημάτων πληροφορικής, διεκπεραίωσης συναλλαγών ή τιμολόγησης. Συγκεκριμένα, τέτοιου είδους παραδείγματα περιλαμβάνουν πολλές αλλαγές στο προσωπικό ελέγχου, απουσία διαχωρισμού καθηκόντων, εμπλοκή προμηθευτών που είναι συγγενικά πρόσωπα, υπερβολικές εγκρίσεις χωρίς τεκμηρίωση, τιμολόγια που πληρώνονται χωρίς έγκριση κ.ά. (Χαρίτου Σ. χ.χ).

Η τρίτη κατηγορία εντοπίζει τα προσωπικά ή συμπεριφορικά red flags, τα οποία σχετίζονται άμεσα με τον άνθρωπο-δράστη ή με το περιβάλλον του. Ενδείξεις αυτών αποτελούν τα ξαφνικά αποκτηθέντα πολυτελή αγαθά χωρίς αντίστοιχο εισόδημα, η υπέρμετρη ενασχόληση με οικονομικά προβλήματα ή τζόγο, η άρνηση άδειας ή διακοπής εργασίας λόγο του φόβου ότι θα προδοθεί, η άνευ προηγουμένου προσωπική συμπεριφορά που αποκλίνει από το σύνηθες (Χαρίτου Σ., χ.χ.). Εδώ λοιπόν εντοπίζεται ότι «η υπερβολικά υψηλή προσωπική πίεση», «η αύξηση προσωπικών χρεών» και «η αλλαγή του τρόπου ζωής του εργαζόμενου» αποτελούν red flags προσωπικού χαρακτήρα.

Μια τέταρτη κατηγορία αφορά τα χρηματοοικονομικά ή λογιστικά red flags. Σε αυτήν εντάσσονται οι αποκλίσεις σε οικονομικούς δείκτες, οι αλλαγές στη λογιστική πολιτική χωρίς σαφή αιτιολόγηση, η απόκρυψη υποχρεώσεων, η ξαφνική βελτίωση κερδών χωρίς αντίστοιχη αύξηση των δραστηριοτήτων και η υπερβολική εξάρτηση της διοίκησης από τους οικονομικούς στόχους. Για παράδειγμα, η μελέτη "Auditing the Risk of Financial Fraud Using the Red Flags Technique" αποδεικνύει ότι η παρακολούθηση δεικτών όπως της ρευστότητας ή της αποδοτικότητας μπορεί να λειτουργήσει ως red flag για απάτες (Munteanu et al., 2024). Τα red flags αυτού του τύπου αποτελούν το βασικό πεδίο εφαρμογής σε ελέγχους οικονομικών καταστάσεων.

Τέλος, μπορεί να προστεθεί μια κατηγορία τεχνολογικών ή πληροφοριακών red flags, που αφορά συστήματα πληροφορικής και τεχνολογίας με ελλιπή πρόσβαση ή έλεγχο, διαδοχικές αλλαγές στο προσωπικό IT, απουσία ιστορικού ελέγχου συναλλαγών, χρήση δύο λογιστικών βιβλίων, ασυνήθιστους συνδυασμούς προμηθευτών/εργολάβων με κοινά στοιχεία (Ehlermann-Cache, 2015). Αυτές οι ενδείξεις έχουν αποκτήσει ιδιαίτερη σημασία στον 21^{ο} αιώνα, δεδομένου του όγκου δεδομένων και της εξάπλωσης των ψηφιακών συστημάτων.

Εντούτοις, ο ρόλος των red flags δεν περιορίζεται στην απλή αναγνώριση. Κατ' ουσίαν, αποτελούν εργαλείο προειδοποίησης και βάση για την ανάλυση κινδύνου. Στην πράξη, οι εσωτερικοί και οι εξωτερικοί ελεγκτές χρησιμοποιούν «λίστες με τις διάφορες κατηγορίες» και συστήματα «ηλεκτρονικής προειδοποίησης» για να εντοπίσουν περιοχές υψηλού κινδύνου. Η γνώση και η αντίληψη των red flags ενισχύει την κριτική αντιμετώπιση σύμφωνα με την οποία ο ελεγκτής δεν αποδέχεται ως δεδομένα τα στοιχεία, αλλά τα εξετάζει κριτικά, λαμβάνοντας υπόψη την πιθανότητα να είναι ασυνεπή ή λανθασμένα. Κατέχοντας τις απαραίτητες αυτές γνώσεις, μπορεί να δώσει την δέουσα προσοχή και διερευνητικότητα για να διακρίνει τις καταστάσεις που μπορεί να

υποδηλώνουν ουσιώδες σφάλμα από αυτές που μπορεί να υποδηλώνουν απάτη (Munteanu et al., 2024).
Ένα χαρακτηριστικό παράδειγμα red flag σε συμβάσεις ή προμήθειες, μπορεί να είναι η τιμή που φαίνεται υπερβολικά μεγάλη ή υπερβολικά μικρή σε σχέση με παραπλήσιες συμβάσεις (Χαρίτου Σ., χ.χ.). Αν σε συνδυασμό υπάρχει κοινός διαχειριστής, αλλαγές προμηθευτών χωρίς λόγο κ..λπ, τότε ο ελεγκτής μπορεί να σχεδιάσει περαιτέρω έρευνα. Στον χρηματοοικονομικό έλεγχο, ένα red flag μπορεί να είναι ότι η εταιρεία αυξάνει τα κέρδη της πολύ περισσότερο από τον κλάδο χωρίς αντίστοιχη αύξηση πωλήσεων – και συνδυάζεται με υψηλή μόχλευση και συχνές αλλαγές της λογιστικής πολιτικής (Ehlermann-Cache, 2015).
Είναι σημαντικό να διευκρινιστεί ότι η ύπαρξη red flags δεν σημαίνει αυτομάτως απάτη. Αντιθέτως, σηματοδοτεί ότι υπάρχει ένδειξη για έρευνα και διασταύρωση. Ο ελεγκτής, λοιπόν, πρέπει να αποκτήσει ένα σύστημα αξιολόγησης των red flags, να σταθμίσει την πιθανότητα κινδύνου και να λάβει αποφάσεις για περαιτέρω διερεύνηση ή όχι. Επίσης, σε ένα σύγχρονο πλαίσιο, η Τεχνητή Νοημοσύνη μπορεί να επιταχύνει και να αυτοματοποιήσει την ανίχνευση red flags καθώς και την αξιολόγησή τους, όπως θα αναλύσουμε περεταίρω παρακάτω.
Εν κατακλείδι, η έννοια των red flags αποτελεί βασικό πυλώνα στην πρόληψη και την ανίχνευση της απάτης. Η κατηγοριοποίησή τους (διοικητικά, διαδικαστικά, προσωπικά, χρηματοοικονομικά, τεχνολογικά) επιτρέπει στον οργανισμό να εφαρμόζει δομημένες στρατηγικές ελέγχου και ανάπτυξης συστημάτων έγκαιρης προειδοποίησης, ενώ η εφαρμογή τους τόσο στην ανάλυση οικονομικών καταστάσεων όσο και στη διαχείριση προμηθειών, συστημάτων πληροφορικής και ανθρώπινου δυναμικού τα καθιστά εργαλείο όχι μόνο ανίχνευσης, αλλά και πρόληψης.

1.4 Ορισμοί και Ιστορική εξέλιξη της Τεχνητής Νοημοσύνης

Η Τεχνητή Νοημοσύνη (ΤΝ) αποτελεί έναν από τους πιο ραγδαία αναπτυσσόμενους τομείς της σύγχρονης επιστήμης και της τεχνολογίας, με εφαρμογές που εκτείνονται από την ιατρική και τη ρομποτική έως τη λογιστική, τη χρηματοοικονομική ανάλυση και την ανίχνευση απάτης. Η έννοια της ΤΝ δεν είναι νέα, ωστόσο, η έκρηξη των τεχνολογιών και των εφαρμογών της τα τελευταία χρόνια έχει αναδιαμορφώσει τον τρόπο με τον οποίο οι οργανισμοί αντιλαμβάνονται και αξιοποιούν τα δεδομένα τους.
Η μελέτη της ΤΝ και των λειτουργιών της στην παρούσα εργασία είναι κρίσιμη, καθώς συνδέεται άμεσα με τη διαδικασία εντοπισμού red flags, δηλαδή ενδείξεων ύποπτων συναλλαγών ή συμπεριφορών που μπορεί να υποδηλώνουν απάτη.

1.4.1 Ορισμός της Τεχνητής Νοημοσύνης

Η Τεχνητή Νοημοσύνη ορίζεται ως «ο επιστημονικός κλάδος που ασχολείται με τη δημιουργία συστημάτων ικανών να εκτελούν εργασίες που απαιτούν ανθρώπινη νοημοσύνη, όπως η μάθηση, η λήψη αποφάσεων, η αντίληψη και η κατανόηση της φυσικής γλώσσας» (European Commission, 2019).
Ο ορισμός αυτός υποδηλώνει ότι η ΤΝ δεν περιορίζεται σε μια μεμονωμένη τεχνολογία, αλλά περιλαμβάνει ένα σύνολο τεχνικών και μεθόδων, όπως είναι:

- η Μηχανική μάθηση (Machine Learning)

- τα Νευρωνικά δίκτυα (Neural Networks)
- η Επεξεργασία φυσικής γλώσσας (Natural Language Processing)
- τα Εξειδικευμένα συστήματα (Expert Systems)
- η Ρομποτική.

Σύμφωνα με την Ευρωπαϊκή Επιτροπή, η ΤΝ στοχεύει στην προσομοίωση των γνωστικών λειτουργιών του ανθρώπινου εγκεφάλου, ενώ η εξέλιξή της βασίζεται σε τρεις κύριες αρχές: την αναπαράσταση της γνώσης, τη μάθηση και την αυτοβελτίωση των συστημάτων μέσω της εμπειρίας.

1.4.2 Ιστορική Εξέλιξη της Τεχνητής Νοημοσύνης

Η απαρχή της ΤΝ τοποθετείται χρονικά στη δεκαετία του 1940, όταν οι επιστήμονες άρχισαν να εξετάζουν τη δυνατότητα δημιουργίας μηχανών ικανών να «σκέφτονται». Ο Alan Turing (1950) με το κλασικό άρθρο του *"Computing Machinery and Intelligence"* έθεσε το θεμέλιο ερώτημα: «Μπορούν οι μηχανές να σκεφτούν;». Την ίδια περίοδο, ο John von Neumann σχεδίασε την αρχιτεκτονική των υπολογιστών, που αποτέλεσε τη βάση για όλα τα μελλοντικά συστήματα ΤΝ.

Το 1956, στο συνέδριο του Dartmouth College, οι John McCarthy, Marvin Minsky, Claude Shannon και Nathan Rochester εισήγαγαν επίσημα τον όρο Artificial Intelligence, σηματοδοτώντας την έναρξη μιας νέας επιστήμης. Τα πρώτα προγράμματα ΤΝ, όπως το Logic Theorist των Newell και Simon (1956), έδειξαν ότι οι υπολογιστές μπορούσαν να επιλύουν προβλήματα που απαιτούσαν συλλογιστική λογική.

Κατά τη δεκαετία του 1960, η ΤΝ γνώρισε έντονη ανάπτυξη και υπερβολικές προσδοκίες. Εμφανίστηκαν τα πρώτα εξειδικευμένα συστήματα, όπως το ELIZA (1966), που προσομοίωνε συνομιλία ανθρώπου με μηχανή. Ωστόσο, τα τεχνολογικά όρια της εποχής και η έλλειψη υπολογιστικής ισχύος οδήγησαν στο λεγόμενο "AI Winter", δηλαδή μια περίοδο περιορισμένου ενδιαφέροντος και χρηματοδότησης.

Η δεκαετία του 1980 σηματοδοτήθηκε από την αναβίωση των νευρωνικών δικτύων, κυρίως λόγω του αλγορίθμου οπισθοδιάδοσης (backpropagation) που ανέπτυξαν οι Rumelhart και Hinton το 1986. Αυτή η τεχνική, που περιγράφηκε στο άρθρο «Learning Representations by Back-propagating Errors», κατέστη θεμελιώδης για την εκπαίδευση των σύγχρονων νευρωνικών δικτύων και την ανάπτυξη της βαθιάς μάθησης (deep learning). Στα τέλη της δεκαετίας του 1990 οι αλγόριθμοι μηχανικής μάθησης έθεσαν τα θεμέλια της σύγχρονης ΤΝ.

Η σημερινή εποχή χαρακτηρίζεται από εκρηκτική αύξηση δεδομένων, υψηλή υπολογιστική ισχύ και νέες αρχιτεκτονικές βαθιάς μάθησης. Η ανάπτυξη της τεχνολογίας GPU (τα γνωστά chip από τις κάρτες γραφικών), σε συνδυασμό με την πρόοδο στη συλλογή και την ανάλυση δεδομένων, οδήγησε σε εντυπωσιακές εφαρμογές όπως είναι η αναγνώριση εικόνων και φωνής, η αυτόνομη οδήγηση, η ανάλυση κινδύνου, η πρόβλεψη απάτης και, τέλος, τα συστήματα σύστασης (recommendation systems) τα οποία είναι αλγόριθμοι που χρησιμοποιούν δεδομένα όπως προηγούμενες αγορές, ιστορικό προβολής και αλληλεπιδράσεις για να προβλέψουν και να προτείνουν σχετικά στοιχεία, περιεχόμενο ή υπηρεσίες στους χρήστες. Οι παραπάνω αλγόριθμοι είναι απαραίτητοι για τη διαχείριση του συντριπτικού αριθμού επιλογών σε πλατφόρμες όπως το Amazon, το Netflix και το YouTube, βοηθώντας τους χρήστες να ανακαλύψουν νέα πράγματα και να

βελτιώσουν την εμπειρία τους. Αυτά τα συστήματα αποτελούν βασική εφαρμογή της μηχανικής μάθησης.

1.4.3 Βασικές Κατηγορίες και Τεχνολογίες της ΤΝ

Η Τεχνητή Νοημοσύνη διακρίνεται σε δύο βασικές κατηγορίες (European Commission, 2019):

1. Ασθενής ή περιορισμένη ΤΝ (Narrow AI): «είναι συστήματα που μπορούν να εκτελούν μόνο μία δραστηριότητα ή μικρό αριθμό συγκεκριμένων δραστηριοτήτων» και εφαρμόζεται σε συγκεκριμένα καθήκοντα, όπως η ταξινόμηση εικόνων ή η πρόβλεψη τιμών.
2. Γενική ή ισχυρή ΤΝ (General AI): πρόκειται για συστήματα που θα μπορούσαν να εκτελέσουν οποιαδήποτε γνωστική λειτουργία ενός ανθρώπου — στόχος που παραμένει μέχρι σήμερα θεωρητικός.

Στο πλαίσιο της λογιστικής και των εσωτερικών ελέγχων, η Ασθενής ΤΝ είναι η κυρίαρχη μορφή, καθώς περιλαμβάνει αλγορίθμους ανίχνευσης ανωμαλιών (anomaly detection) και μοντέλα πρόβλεψης που εντοπίζουν ύποπτα πρότυπα συναλλαγών ή red flags.
Η ΤΝ μετασχηματίζει ριζικά τη λογιστική επιστήμη και τον εσωτερικό έλεγχο. Οι σύγχρονοι ελεγκτικοί μηχανισμοί χρησιμοποιούν πλέον αλγορίθμους μηχανικής μάθησης για την ανίχνευση ασυνήθιστων λογιστικών εγγραφών, την παρακολούθηση των ροών πληρωμής, τον εντοπισμό πιθανών συγκρούσεων συμφερόντων και τη δημιουργία "προφίλ κινδύνου" για κάθε συναλλαγή. Η ενσωμάτωση της ΤΝ στη διαδικασία ελέγχου συμβάλλει στη μείωση του ανθρώπινου λάθους, στην αύξηση της ταχύτητας ανάλυσης και κυρίως στην πρόληψη της απάτης, καθώς μπορεί να εντοπίσει μοτίβα που δύσκολα αντιλαμβάνεται ο άνθρωπος. Επιπλέον, η ανάλυση μεγάλου όγκου δεδομένων (Big Data Analytics) επιτρέπει την επεξεργασία τεράστιων όγκων πληροφοριών από διαφορετικές πηγές όπως είναι τα λογιστικά βιβλία, τα τιμολόγια, οι τραπεζικές κινήσεις, οι επικοινωνίες, που πάντα έχουν ως στόχο τον εντοπισμό red flags. (Αντωνιάδου, 2025).

Σύμφωνα με τον Γενίτσαρη (2024), τα συστήματα ΤΝ μπορούν να εντοπίσουν μοτίβα απάτης μέσω:

- αλγορίθμων ταξινόμησης, που αναγνωρίζουν αν μια συναλλαγή είναι "κανονική" ή "ύποπτη",
- αλγορίθμων συσταδοποίησης (clustering), μιας θεμελιώδους τεχνικής ανάλυσης δεδομένων και μηχανικής μάθησης χωρίς επίβλεψη (unsupervised learning), η οποία έχει ως στόχο την ομαδοποίηση ενός συνόλου αντικειμένων με βάση την ομοιότητά τους ώστε να εντοπίζει αποκλίσεις,
- ανάλυσης συναισθήματος, που εφαρμόζεται σε εταιρικές επικοινωνίες για εντοπισμό πιθανών συγκαλύψεων.

Οι αλγόριθμοι αυτοί τροφοδοτούνται από ιστορικά δεδομένα απάτης και μαθαίνουν να αναγνωρίζουν χαρακτηριστικά που επαναλαμβάνονται. Έτσι, οι ελεγκτές μπορούν να εστιάσουν σε υψηλού κινδύνου συναλλαγές, εξοικονομώντας χρόνο και πόρους.

1.4.4 Προκλήσεις και Ηθικές Διαστάσεις

Παρά τα ποικίλα οφέλη, η ενσωμάτωση της ΤΝ δημιουργεί και προκλήσεις, όπως:

- η έλλειψη διαφάνειας (black box problem,) στους αλγορίθμους, δηλαδή της αδυναμίας του συστήματος να εξηγήσει πως κατέληξε σε ένα συμπέρασμα,
- τα ζητήματα προστασίας προσωπικών δεδομένων, και
- η αντικατάσταση ανθρώπινων ρόλων με αυτοματοποιημένα συστήματα.

Οι ηθικές αυτές προκλήσεις απαιτούν νέες προσεγγίσεις στη διακυβέρνηση της ΤΝ, ώστε να διασφαλίζεται ότι οι τεχνολογίες αυτές χρησιμοποιούνται υπεύθυνα και προς όφελος της κοινωνίας (European Commission, 2019).

Η Τεχνητή Νοημοσύνη εξελίσσεται από μια θεωρητική ιδέα σε εργαλείο στρατηγικής σημασίας για κάθε οργανισμό. Από τα πρώτα βήματα του Turing έως τα σημερινά συστήματα ανίχνευσης red flags, η ΤΝ έχει αποδείξει ότι μπορεί να ενισχύσει τη διαφάνεια, τη λογοδοσία και την αποδοτικότητα στον εσωτερικό έλεγχο.
Η μελλοντική πρόκληση έγκειται στη σωστή αξιοποίησή της, με σεβασμό στην ηθική, την ασφάλεια και την ανθρώπινη κρίση.

Κεφάλαιο 2 - Σύγχρονες τάσεις στην Τεχνητή Νοημοσύνη (AI) και τον εσωτερικό έλεγχο

2.1 Σύγχρονες μελέτες για την εφαρμογή της Τεχνητής Νοημοσύνης στον εσωτερικό έλεγχο

Η τελευταία δεκαετία χαρακτηρίζεται από σημαντικές εξελίξεις στον χώρο της Τεχνητής Νοημοσύνης (ΤΝ), οι οποίες επηρεάζουν ριζικά το αντικείμενο του εσωτερικού ελέγχου. Η μετάβαση από τις παραδοσιακές, χειροκίνητες διαδικασίες ελέγχου σε προηγμένα συστήματα ανάλυσης δεδομένων και αλγοριθμικής υποστήριξης έχει μεταμορφώσει τον τρόπο με τον οποίο οι οργανισμοί εντοπίζουν κινδύνους, αξιολογούν διαδικασίες και προλαμβάνουν φαινόμενα απάτης και παρατυπιών (Vasarhelyi, Kogan & Tuttle, 2015). Παράλληλα, η αύξηση του όγκου και της πολυπλοκότητας των εταιρικών δεδομένων καθιστά απαραίτητη τη χρήση αλγορίθμων Μηχανικής Μάθησης (Machine Learning – ML), Εξόρυξης Δεδομένων (Data Mining) και Επεξεργασίας Φυσικής Γλώσσας (Natural Language Processing – NLP) στην καθημερινή πρακτική του ελεγκτικού έργου (Deloitte, 2024).

Οι σύγχρονες ελεγκτικές υπηρεσίες αξιοποιούν την ΤΝ όχι απλώς για να εκτελέσουν ταχύτερα τις υπάρχουσες διαδικασίες, αλλά για να μεταβάλουν τη φιλοσοφία του ελέγχου από μια εκ των υστέρων διαδικασία επαλήθευσης σε μια συνεχή, δυναμική και προληπτική λειτουργία (Alles et al, 2014). Με βάση ερευνητικά δεδομένα, πάνω από το 70% των μεγάλων διεθνών οργανισμών επενδύουν σε τεχνολογίες ΤΝ για την ενίσχυση της διαχείρισης κινδύνων (risk management) και της κανονιστικής συμμόρφωσης (compliance).

Σήμερα, οι εφαρμογές της ΤΝ στον εσωτερικό έλεγχο περιλαμβάνουν:

- συνεχή παρακολούθηση (continuous auditing),
- ανάλυση ανωμαλιών (anomaly detection),
- εντοπισμό ύποπτων συναλλαγών,
- αξιολόγηση προμηθευτών και πελατών,
- αποτύπωση και προσπάθεια ερμηνείας των διαφόρων συμπεριφορών με χρήση NLP (natural language prossesing),
- αυτόματη αναγνώριση red flags σε λογιστικά και λειτουργικά δεδομένα καθώς και σε δεδομένα επικοινωνίας (πχ e-mails).

Ειδικά σε διεθνείς οργανισμούς, η ΤΝ λειτουργεί ως μια «δεύτερη γραμμή άμυνας» στο πλαίσιο της εταιρικής διακυβέρνησης, ενισχύοντας τους μηχανισμούς κανονιστικής συμμόρφωσης με τρόπο που δεν ήταν εφικτός στο παρελθόν (Warren, Moffitt & Byrnes, 2015). Από τα παραπάνω, αναδεικνύονται οι σημαντικές εφαρμογές της ΤΝ στον εντοπισμό red flags σε πολλαπλές λειτουργίες ενός οργανισμού.

Κάποια από τα πιο ανεπτυγμένα πεδία έρευνας αφορούν την εφαρμογή της ΤΝ για τον εντοπισμό ύποπτων συναλλαγών, την πρόβλεψη λογιστικών παρατυπιών και την ανίχνευση αλλοίωσης των οικονομικών καταστάσεων.

Μελέτες όπως των Ngai et al. (2011) καταδεικνύουν ότι μοντέλα ταξινόμησης (classification), νευρωνικά δίκτυα κ.λπ., μπορούν να εντοπίζουν μοτίβα απάτης με ακρίβεια υψηλότερη από εκείνη των παραδοσιακών ελεγκτικών μεθόδων.
Στην Ελλάδα, η μελέτη της Φαρμάκη (2023) επιβεβαιώνει ότι η χρήση ΤΝ σε τραπεζικούς και χρηματοοικονομικούς οργανισμούς οδηγεί σε μείωση λαθών και ταχύτερο εντοπισμό παρατυπιών. Όπως αναφέρθηκε ανωτέρω όμως, τα διάφορα μοντέλα ΤΝ μπορούν να προχωρήσουν ένα βήμα παραπέρα από αυτό, προσφέροντας ανάλυση και ερμηνεία των e-mail, οδηγώντας έτσι στον εντοπισμό κάποιας ασυνήθιστης επικοινωνίας, μέσω για παράδειγμα της καταγραφής του συναισθηματικού τόνου, τον εντοπισμό πιέσεων, την ανακάλυψη χειριστικών συμπεριφορών ή κάποιας απόπειρας συγκάλυψης.
Αυτή η προσέγγιση επιτρέπει τον εντοπισμό red flags σε ανθρώπινες συμπεριφορές και εταιρική κουλτούρα, τα οποία οι παραδοσιακοί έλεγχοι πολλές φορές αδυνατούν να αξιολογήσουν.

2.1.1 Cognitive Auditing και Predictive Analytics

Ο όρος Cognitive Auditing περιγράφει την εφαρμογή συστημάτων ΤΝ που «κατανοούν» πρότυπα, ερμηνεύουν διαδικασίες και έτσι παράγουν αυτόματα υποδείξεις προς τους ελεγκτές. Η Deloitte (2024) αναφέρει ότι πλέον οι ελεγκτές λαμβάνουν διαγνωστικές αναφορές**,** όπου η ΤΝ εξηγεί ποια συναλλαγή θεωρεί red flag, ποιες μεταβλητές συνέβαλαν και πόσο σοβαρό θεωρεί τον κίνδυνο.
Παράλληλα, τα Predictive Analytics σχετίζονται με μια διαδικασία η οποία αξιοποιώντας δεδομένα, επιτρέπει τον εντοπισμό ανερχόμενων κινδύνων πριν αυτοί εκδηλωθούν.
Ένα από τα μεγαλύτερα πλεονεκτήματα της ΤΝ είναι ότι υποστηρίζει, επικουρεί αλλά δεν αντικαθιστά την ανθρώπινη κρίση. Ως εκ τούτου μπορεί να χρησιμοποιηθεί ως μηχανισμός πρώιμης ειδοποίησης, ως σύστημα φίλτρου για μεγάλο όγκο δεδομένων ακόμη και ως εργαλείο προτεραιοποίησης κινδύνων. Εντούτοις, οι ελεγκτές παραμένουν οι τελικοί υπεύθυνοι για την ερμηνεία, την επιβεβαίωση και τη λήψη αποφάσεων. Έτσι, επιτυγχάνεται ένα υβριδικό μοντέλο λειτουργίας, το οποίο θα αναλυθεί περαιτέρω σε επόμενο κεφάλαιο.
Οι σύγχρονες μελέτες, όμως, συμφωνούν και στο γεγονός ότι η ΤΝ φέρνει και προκλήσεις που αφορούν θέματα ποιότητας δεδομένων, την ανάγκη για explainability (ερμηνευσιμότητα, δηλαδή το πώς και το γιατί η ΤΝ έφτασε σε ένα συμπέρασμα), θέματα δεοντολογίας που προκύπτουν από τη χρήση της, και την ανάγκη για περεταίρω εκπαίδευση των ελεγκτών.

2.2 Ερμηνευσιμότητα και ζητήματα εμπιστοσύνης στην Τεχνητή Νοημοσύνη

Η ραγδαία ενσωμάτωση συστημάτων Τεχνητής Νοημοσύνης (ΤΝ) στον εσωτερικό έλεγχο έχει δημιουργήσει νέα δεδομένα ως προς τη λήψη αποφάσεων, την αναγνώριση κινδύνων και την ανίχνευση εταιρικών red flags. Ωστόσο, η εφαρμογή τέτοιων συστημάτων συνοδεύεται από μια σειρά ηθικών, τεχνικών και λειτουργικών προκλήσεων, με κυρίαρχη τη δυσκολία ερμηνείας των αλγοριθμικών αποφάσεων και το συνακόλουθο πρόβλημα της εμπιστοσύνης των ελεγκτών σε αυτά τα συστήματα (Lipton, 2018).
Η ερμηνευσιμότητα και η εμπιστοσύνη αποτελούν πυλώνες της ελεγκτικής επιστήμης. Ο εσωτερικός ελεγκτής οφείλει όχι μόνο να εντοπίζει αποκλίσεις αλλά και να τεκμηριώνει

επαρκώς κάθε κρίση ή σύσταση που διατυπώνει (IIA, 2020). Επομένως, η εισαγωγή αλγοριθμικών μοντέλων που λειτουργούν ως «μαύρα κουτιά» κινδυνεύει να υπονομεύσει τη θεσμική λογοδοσία και τη νομιμότητα του ελεγκτικού έργου, εάν δεν συνοδεύεται από κατάλληλες πρακτικές διαφάνειας (European Commission, 2021).
Στην παρούσα ενότητα αναλύονται σε βάθος οι παράμετροι της ερμηνευσιμότητας της ΤΝ, οι παράγοντες που επηρεάζουν την εμπιστοσύνη των ελεγκτών, οι κίνδυνοι αλγοριθμικής μεροληψίας και οι σύγχρονες προσεγγίσεις Explainable AI (XAI) που συμβάλλουν στη γεφύρωση του χάσματος ανθρώπου-μηχανής.

2.2.1 Η έννοια της ερμηνευσιμότητας στην ΤΝ

Η ερμηνευσιμότητα αναφέρεται στη δυνατότητα του χρήστη να κατανοήσει τους λόγους για τους οποίους ένα μοντέλο ΤΝ κατέληξε σε μια συγκεκριμένη απόφαση (Lipton, 2018). Στο πλαίσιο του εσωτερικού ελέγχου, η ερμηνευσιμότητα αποτελεί κρίσιμη προϋπόθεση, διότι ο ελεγκτής οφείλει να εξηγήσει:

- γιατί ένα σύστημα χαρακτήρισε μια συναλλαγή ως red flag,
- ποιες μεταβλητές επηρέασαν την απόφαση,
- με ποιον τρόπο η απόφαση επηρεάζει την αξιολόγηση του κινδύνου,
- ποια μέτρα πρέπει να ληφθούν.

Σε αντίθεση με άλλους τομείς της επιχειρησιακής λειτουργίας, ο εσωτερικός έλεγχος είναι βαθιά θεσμικός και ρυθμισμένος. Το Ινστιτούτο Εσωτερικών Ελεγκτών (IIA) απαιτεί από τους επαγγελματίες να είναι σε θέση να αιτιολογούν πλήρως τα ευρήματα και τις κρίσεις τους (IIA, 2017). Η αδιαφάνεια ενός αλγορίθμου μπορεί να δημιουργήσει προβλήματα σε τυχόν δικαστικές διαδικασίες, επιθεωρήσεις ρυθμιστικών αρχών, διάφορες διοικητικές αποφάσεις, ίσως και στην εσωτερική διακυβέρνηση εξ ολοκλήρου. Επιπλέον, χωρίς επαρκή ερμηνευσιμότητα, ένας ελεγκτής δεν μπορεί να αξιολογήσει την αξιοπιστία μιας πρόβλεψης ούτε να διακρίνει εάν ένα red flag οφείλεται σε πραγματική ανωμαλία ή σε τεχνικό σφάλμα.
Τυχόν χαμηλή ερμηνευσιμότητα ενδέχεται, μεταξύ άλλων, να οδηγήσει σε αυξημένα false positives με αποτέλεσμα την εργασιακή υπερφόρτωση των ελεγκτών, πολλαπλά false negatives με αποτέλεσμα τον μη εντοπισμό κρίσιμων red flags, περιορισμένη δυνατότητα λογοδοσίας, δυσπιστία εναντίον του συστήματος ΤΝ εκ μέρους των εργαζομένων και των ομάδων management, και, τέλος, σε ρίσκο τυχόν παραβίασης της νομοθεσίας (π.χ. του GDPR). Συνεπώς, η ερμηνευσιμότητα δεν συνιστά απλώς τεχνική απαίτηση αλλά πυλώνα λειτουργικής και ηθικής νομιμοποίησης του συστήματος ΤΝ.

2.2.2 Ζητήματα εμπιστοσύνης σε συστήματα ΤΝ

Η εμπιστοσύνη είναι πολυδιάστατη και μπορεί να διακριθεί στα εξής:

Εμπιστοσύνη στην τεχνολογία

Αφορά το κατά πόσο ο ελεγκτής πιστεύει ότι το σύστημα λειτουργεί σωστά, παράγει σταθερά αποτελέσματα, έχει εκπαιδευτεί με αξιόπιστα δεδομένα και όλα αυτά χωρίς να επηρεάζεται από τεχνικά σφάλματα.

Εμπιστοσύνη στον οργανισμό
Σχετίζεται με το αν ο οργανισμός χρησιμοποιεί υπεύθυνα την ΤΝ, με την κουλτούρα διαφάνειας των αποτελεσμάτων που αυτή παράγει καθώς και την ύπαρξη θεσμικών πολιτικών για τη λειτουργία της.

Ηθική εμπιστοσύνη
Η ηθική εμπιστοσύνη αναφέρεται στη δικαιοσύνη του συστήματος. Αυτή αφορά το εάν το σύστημα είναι δίκαιο. Δηλαδή, δεν εισάγει διακρίσεις και δεν λειτουργεί μεροληπτικά. Για παράδειγμα, να έχει εκπαιδευτεί ώστε να αποφεύγει να συμπεριλαμβάνει στους ελέγχους τα τιμολόγια των εξόδων του CEO.

Εμπιστοσύνη στη διαδικασία της λήψης αποφάσεων
Οι χρήστες δυσπιστούν απέναντι σε black-box μοντέλα. Οι ελεγκτές εμπιστεύονται περισσότερο ένα σύστημα, όταν μπορούν να δουν τα δεδομένα που το «οδήγησαν» σε μια απόφαση, όταν υπάρχει ανθρώπινη επικύρωση (human-in-the-loop) και όταν η ΤΝ εξηγεί τη συλλογιστική της πορεία.

Αλγοριθμική μεροληψία
Η αλγοριθμική μεροληψία (bias) είναι ένας από τους πιο σημαντικούς κινδύνους της ΤΝ. Ένας αλγόριθμος μπορεί να ενισχύσει υφιστάμενες ανισότητες, εάν εκπαιδευτεί με ελλιπή δεδομένα, με μεροληπτικά ιστορικά δεδομένα ή δεδομένα που αντικατοπτρίζουν παλαιές νοοτροπίες. Τέτοια παραδείγματα μεροληψίας σε red flags μπορεί να περιλαμβάνουν τη συχνότερη στοχοποίηση των εργαζομένων ενός συγκεκριμένου τμήματος, την εσφαλμένη σύνδεση συγκεκριμένων ρόλων με παρατυπίες κ.λπ.
Η μεροληψία αυτή μπορεί να έχει σοβαρές συνέπειες, ιδίως βάσει του GDPR και του EU AI Act.

2.2.3 Η Explainable AI (XAI) ως λύση στο πρόβλημα εμπιστοσύνης
Η Explainable AI αναπτύχθηκε για να κάνει τα συστήματα ΤΝ πιο κατανοητά, πιο διαφανή, πιο επεξηγήσιμα και πιο αξιόπιστα. Τούτο επιτυγχάνεται μέσα από τις τεχνικές XAI που χρησιμοποιούνται στον εσωτερικό έλεγχο. Παρακάτω δίνονται μερικά παραδείγματα τέτοιων τεχνικών:

LIME και SHAP values
Εξηγούν ποια χαρακτηριστικά συνέβαλαν περισσότερο σε μια απόφαση. Π.χ. «Το τιμολόγιο θεωρήθηκε ύποπτο επειδή:

(1) ήταν 48% πάνω από τη μέση τιμή,

(2) εκδόθηκε εκτός ωραρίου».

Decision rules extraction
Μετατρέπει συστήματα ΤΝ που έχουν το πρόβλημα του black-box σε Explainable συστήματα μέσα από κανόνες που θέτει ο διαχειριστής.

Attention mechanisms
Χρησιμοποιούνται σε NLP για να δείξουν ποιες λέξεις ή προτάσεις οδήγησαν στο αποτέλεσμα.

Τα οφέλη της XAI για τον εσωτερικό έλεγχο μπορούν να απαριθμηθούν ως εξής:

- μειώνει την αβεβαιότητα,
- ενισχύει τη λογοδοσία,
- βοηθά τους ελεγκτές να εξηγήσουν τα όποια ευρήματά τους στη διοίκηση,
- προστατεύει τον οργανισμό από νομικές προκλήσεις,
- μειώνει τον κίνδυνο κακής χρήσης ή κατάχρησης των αλγορίθμων.

2.2.4 Ο ρόλος της ανθρώπινης εποπτείας (human-in-the-loop)

Καμία σύγχρονη μελέτη δεν υποστηρίζει την πλήρη αυτοματοποίηση των ελεγκτικών αποφάσεων. Η Ευρωπαϊκή Επιτροπή (2021) επιβάλλει την ύπαρξη ανθρώπινης εποπτείας στα συστήματα υψηλού κινδύνου, όπως είναι η ανίχνευση red flags. Αυτό συμβαίνει διότι η ανθρώπινη εποπτεία μπορεί να διορθώσει τυχόν αλγοριθμικά σφάλματα, διασφαλίζει τη δίκαιη χρήση, ενισχύει την τεκμηρίωση, παρέχει κρίση και ηθικό πλαίσιο, αποτρέπει την τυφλή εμπιστοσύνη στην ΤΝ και ενσωματώνει την εμπειρία των ελεγκτών. Αυτό είναι βασικό μέρος των υβριδικών AI–Human συστημάτων, που θα αναλυθούν σε επόμενο υποκεφάλαιο.
Η εμπιστοσύνη, όμως, δεν αφορά μόνο τους ελεγκτές αλλά και τους εργαζόμενους. Σύμφωνα με μελέτες (Floridi et al., 2021), οι εργαζόμενοι εμπιστεύονται περισσότερο τα συστήματα ΤΝ όταν υπάρχει διαφάνεια στη χρήση τους, όταν γνωρίζουν πότε και γιατί παρακολουθούνται συγκεκριμένα δεδομένα, όταν υπάρχει εγγύηση ότι δεν αξιολογούνται άδικα, όταν υπάρχουν διαδικασίες διόρθωσης των σφαλμάτων και όταν τα συστήματα χρησιμοποιούνται για προστασία και όχι τιμωρία. Με βάση όλα τα παραπάνω, η ΤΝ μπορεί να λειτουργήσει ως ένα αξιόπιστο εργαλείο ενίσχυσης της ελεγκτικής λειτουργίας, χωρίς να διακυβεύεται η δεοντολογία, η διαφάνεια και η εταιρική νομιμοποίηση.

2.2.5 Το πρόβλημα των Hallucinations στα συστήματα ΤΝ και οι επιπτώσεις στην ανίχνευση red flags

Ένα από τα σημαντικότερα προβλήματα αξιοπιστίας των σύγχρονων μοντέλων Τεχνητής Νοημοσύνης, ιδιαίτερα των μεγάλων γλωσσικών μοντέλων (LLMs), είναι το φαινόμενο των «hallucinations», δηλαδή η παραγωγή απαντήσεων που φαίνονται συνεκτικές και πειστικές αλλά είναι λανθασμένες, ανακριβείς ή πλήρως επινοημένες (Rotman et al., 2022). Τα hallucinations δεν αποτελούν απλώς τεχνικά σφάλματα, συνδέονται με την εσωτερική λειτουργία των μοντέλων, τα οποία προβλέπουν επόμενες λέξεις ή μοτίβα με βάση στατιστικές πιθανότητες και όχι πραγματική κατανόηση περιεχομένου. Αυτό σημαίνει ότι ένα LLM μπορεί να «επινοήσει» στοιχεία, γεγονότα, συσχετίσεις ή αιτιολογήσεις όταν δεν διαθέτει επαρκή πληροφορία ή όταν τα δεδομένα εκπαίδευσης είναι ασαφή, ανεπαρκή ή μεροληπτικά (Zhang et al., 2015). Στην επιστημονική βιβλιογραφία το φαινόμενο αυτό θεωρείται συστημικό και όχι μεμονωμένο, καθώς τα μοντέλα αυτά δεν έχουν αληθινή πρόσβαση σε πραγματική γνώση, παρά μόνο σε

γλωσσικά μοτίβα, γεγονός που οδηγεί σε εσφαλμένες προβλέψεις όταν καλούνται να απαντήσουν σε εξειδικευμένες ή κρίσιμες ερωτήσεις.
Οι αιτίες εμφάνισης hallucinations είναι πολλαπλές. Πρώτον, συνδέονται με την ποιότητα και το εύρος των δεδομένων εκπαίδευσης. Εάν τα δεδομένα περιέχουν λάθη, παρωχημένες πληροφορίες ή ανακρίβειες, το μοντέλο τα αναπαράγει σε νέα συμφραζόμενα, παράγοντας πειστικές αλλά λανθασμένες απαντήσεις (Κιορτσή et al., 2024). Δεύτερον, οφείλονται σε υπερβολικά «δημιουργική» συμπεριφορά των μοντέλων. Τρίτον, συνδέονται με την έλλειψη επαρκούς γνώσης σε εξειδικευμένα πεδία, όπου το μοντέλο αναπληρώνει τα κενά με γενικεύσεις. Τέλος, τα hallucinations εμφανίζονται συχνότερα όταν το μοντέλο καλείται να συνδυάσει δεδομένα πολλαπλών πηγών ή να παραγάγει συμπεράσματα σε περιβάλλοντα που απαιτούν υψηλή ακρίβεια, όπως οι διαδικασίες εσωτερικού ελέγχου.
Στο πλαίσιο της ανίχνευσης εταιρικών red flags, το πρόβλημα των hallucinations αποκτά ιδιαίτερη σημασία, διότι μπορεί να οδηγήσει σε σοβαρά ελεγκτικά σφάλματα. Ένα σύστημα ΤΝ που εντοπίζει ανωμαλίες σε λογιστικές εγγραφές, e-mails, αρχεία προμηθειών ή συναλλαγές μπορεί να παραγάγει false positives εάν «φανταστεί» συσχετίσεις που δεν υφίστανται, ή false negatives εάν δεν αναγνωρίσει πραγματικά ύποπτα μοτίβα επειδή το εσωτερικό του μοντέλο είναι ελλιπές ή παραπλανημένο. Χαρακτηριστικό παράδειγμα είναι η χρήση LLMs για την ανάλυση εταιρικής επικοινωνίας: αν το μοντέλο παρερμηνεύσει το ύφος, τις εκφράσεις ή τις συχνότητες επικοινωνίας, μπορεί να δημιουργήσει ανύπαρκτες ενδείξεις πίεσης, σύγκρουσης συμφερόντων ή απόπειρας συγκάλυψης, οι οποίες θα οδηγήσουν σε άσκοπες ελεγκτικές ενέργειες (Κιορτσή et al., 2024). Αντίστοιχα, σε περιβάλλοντα AML ή continuous auditing, ένα hallucinated pattern μπορεί να οδηγήσει σε υπερβολική αύξηση παρατηρήσεων.
Από τα παραπάνω συνάγεται ότι οι εν λόγω «παραισθήσεις» συνιστούν αυξημένο ελεγκτικό κίνδυνο, καθώς υπονομεύουν την αξιοπιστία των συστημάτων ΤΝ και ενδέχεται να οδηγήσουν σε λανθασμένες αποφάσεις που επηρεάζουν εργαζόμενους, προμηθευτές, οικονομικές αναφορές ή συμμόρφωση με κανονισμούς. Το AI Act της Ευρωπαϊκής Ένωσης επισημαίνει ρητά την ανάγκη ελαχιστοποίησης των συστημικών λαθών και απαιτεί ανθρώπινη εποπτεία και μηχανισμούς ελέγχου για συστήματα υψηλού κινδύνου, όπως αυτά που χρησιμοποιούνται στην ανίχνευση απάτης (European Commission, 2024).
Για τον λόγο αυτό καθίσταται αναγκαία η ενσωμάτωση υβριδικών μοντέλων HITL (Human-in-the-Loop). Η ανθρώπινη εποπτεία δεν αποτελεί μόνο νομική υποχρέωση αλλά και λειτουργική ανάγκη: ο ελεγκτής εξετάζει τα ευρήματα της ΤΝ, αξιολογεί το επιχειρησιακό πλαίσιο, εντοπίζει περιπτώσεις αλγοριθμικών επινοήσεων και εξασφαλίζει ότι μια ανωμαλία αποτελεί πράγματι red flag και όχι hallucination. Η HITL προσέγγιση συμβάλλει στη μείωση false positives, ενισχύει την ακρίβεια και αποτελεί βασική δικλείδα ασφαλείας απέναντι σε συστημικά σφάλματα των αλγορίθμων. Αυτός ο συνδυασμός ανθρώπινης κρίσης και υπολογιστικής ισχύος θεωρείται πλέον η βέλτιστη πρακτική στα σύγχρονα περιβάλλοντα εσωτερικού ελέγχου (Κιορτσή et al., 2024).

2.3 Υβριδικά μοντέλα AI–Human στον εσωτερικό έλεγχο

Η ταχεία ανάπτυξη της Τεχνητής Νοημοσύνης (ΤΝ) τα τελευταία χρόνια έχει δημιουργήσει νέες συνθήκες για τη λειτουργία του εσωτερικού ελέγχου και ειδικότερα για τον τρόπο με τον οποίο εντοπίζονται και αξιολογούνται τα εταιρικά red flags. Παρά

την αδιαμφισβήτητη πρόοδο των αλγοριθμικών μοντέλων, η πλήρης αυτοματοποίηση των διαδικασιών ελέγχου παραμένει όχι μόνο ανέφικτη αλλά και ανεπιθύμητη, δεδομένων των ηθικών, θεσμικών και λειτουργικών απαιτήσεων που διέπουν τον ελεγκτικό ρόλο. Για τον λόγο αυτό, η διεθνής βιβλιογραφία και οι πρακτικές της αγοράς συγκλίνουν στη χρήση υβριδικών μοντέλων συνεργασίας ανθρώπου-μηχανής, τα οποία συνδυάζουν την υπολογιστική ισχύ της ΤΝ με την κρίση, την εμπειρία και τη δεοντολογία του ανθρώπου (Sathe, 2025).

Βασική αρχή των υβριδικών συστημάτων αποτελεί η συμπληρωματικότητα. Τα αλγοριθμικά συστήματα είναι εξαιρετικά στην αναγνώριση μοτίβων, στη διαχείριση τεράστιων όγκων δεδομένων και στην παραγωγή ταχέων προβλέψεων, αλλά αδυνατούν να κατανοήσουν πλήρως τα συμφραζόμενα, τις ιδιαιτερότητες μιας επιχειρησιακής διαδικασίας και τη βαθύτερη πρόθεση πίσω από μια ανθρώπινη συμπεριφορά. Αντίθετα, οι ελεγκτές διαθέτουν θεσμική γνώση, επαγγελματική κρίση, ηθική ευθύνη και κατανόηση του πλαισίου. Καθίσταται, λοιπόν, σαφές ότι ο συνδυασμός των δύο οδηγεί σε υψηλότερη ποιότητα, ακρίβεια και αξιοπιστία στην ανίχνευση κινδύνων και red flags σε σύγκριση με τη χρήση αποκλειστικά ανθρώπινων ή αποκλειστικά αλγοριθμικών διαδικασιών (Sathe, 2025).

Η θεωρητική βάση των υβριδικών συστημάτων συνδέεται στενά με την έννοια της ενισχυμένης ευφυΐας (Augmented Intelligence). Σε αντίθεση με την κλασική προσέγγιση της ΤΝ (Artificial Intelligence), η Augmented Intelligence υποστηρίζει ότι ο στόχος δεν είναι η αντικατάσταση της ανθρώπινης σκέψης αλλά η ενίσχυσή της μέσω της αλγοριθμικής υποστήριξης. Ο άνθρωπος παραμένει ο τελικός λήπτης αποφάσεων, ενώ η ΤΝ προσφέρει δεδομένα, προβλέψεις, αναλύσεις και ενδείξεις που βοηθούν στη λήψη πιο τεκμηριωμένων και αποτελεσματικών ελεγκτικών συμπερασμάτων. Η Ευρωπαϊκή Επιτροπή, μέσα από τις αρχές της αξιόπιστης ΤΝ, επισημαίνει ότι τα συστήματα που χρησιμοποιούνται για κρίσιμες διαδικασίες όπως η αξιολόγηση εργαζομένων, ο εντοπισμός απάτης ή η παρακολούθηση συναλλαγών θεωρούνται «υψηλού κινδύνου» και απαιτούν υποχρεωτική ανθρώπινη εποπτεία (European Commission, 2019).

Στο πλαίσιο του εσωτερικού ελέγχου, η ανάγκη για ανθρώπινη εποπτεία είναι ακόμη πιο επιτακτική. Ο ελεγκτής δεν είναι ένας απλός παρατηρητής αλλά ένας επαγγελματίας που οφείλει να αιτιολογεί κάθε κρίση του με βάση τα διεθνή ελεγκτικά πρότυπα, να διασφαλίζει τη συμμόρφωση με το ρυθμιστικό πλαίσιο και να υποστηρίζει έναν οργανισμό στη διαχείριση κινδύνων, στην προστασία της περιουσίας του και στη διασφάλιση της αξιοπιστίας της πληροφόρησης (ΙΙΑ, 2017). Όταν μια αλγοριθμική διαδικασία εντοπίζει ένα πιθανό red flag, ο ρόλος του ανθρώπου είναι να εξετάσει τα δεδομένα, να αξιολογήσει τις εναλλακτικές εξηγήσεις, να ελέγξει τη λογική του μοντέλου και να αποφασίσει εάν πράγματι υπάρχει κίνδυνος ή αν πρόκειται για ψευδή ένδειξη.

Ένα από τα σημαντικότερα οφέλη της χρήσης υβριδικών συστημάτων είναι η μείωση των false positives και false negatives, τα οποία αποτελούν βασικό πρόβλημα στα μοντέλα αυτόματης ανίχνευσης ανωμαλιών. Τα συστήματα ΤΝ συχνά εντοπίζουν ανωμαλίες που δεν σχετίζονται απαραίτητα με απάτη, αλλά με φυσιολογικές διαφοροποιήσεις ή με δομικές ιδιαιτερότητες μιας επιχειρησιακής διαδικασίας. Η παρουσία ανθρώπου στον κύκλο λήψης αποφάσεων επιτρέπει τη διόρθωση αυτών των σφαλμάτων, αποτρέποντας την εργασιακή υπερφόρτωση των ελεγκτών με άσκοπες

ειδοποιήσεις και διασφαλίζοντας ότι οι πραγματικά κρίσιμες ενδείξεις δεν θα αγνοηθούν (Vasarhelyi, Kogan & Tuttle, 2015).

Τα υβριδικά μοντέλα διακρίνονται σε τρεις βασικές μορφές εποπτείας. Η πρώτη είναι το human-in-the-loop, όπου ο άνθρωπος συμμετέχει άμεσα στη λήψη αποφάσεων. Το σύστημα ΤΝ παράγει αρχικά αποτελέσματα ή ενδείξεις, αλλά η τελική κρίση γίνεται αποκλειστικά από τον άνθρωπο. Η μορφή αυτή είναι η πιο κατάλληλη στα μοντέλα red flags που επηρεάζουν εργαζομένους, προμηθευτές, οικονομικές καταστάσεις ή ζητήματα δεοντολογίας. Η δεύτερη μορφή είναι το human-on-the-loop, όπου ο άνθρωπος λειτουργεί ως επιβλέπων και έχει τη δυνατότητα να επεμβαίνει μόνο όταν κάτι φαίνεται λανθασμένο ή όταν ενεργοποιούνται ορισμένα κριτήρια κινδύνου. Τέλος, το human-in-command αφορά ένα επίπεδο πιο διοικητικό, στο οποίο ο άνθρωπος έχει τον πλήρη θεσμικό και νομικό έλεγχο των αποφάσεων και είναι υπεύθυνος για την τελική έγκριση των αποτελεσμάτων ή των ενεργειών που προκύπτουν από τα συστήματα ΤΝ (Wachter, Mittelstadt , 2019).

Στον τομέα της ανίχνευσης red flags, τα υβριδικά μοντέλα αποδεικνύονται ιδιαίτερα αποτελεσματικά. Για παράδειγμα, στα συστήματα NLP που εξετάζουν εταιρικές επικοινωνίες, η ΤΝ μπορεί να εντοπίσει ασυνήθιστη συχνότητα στην επικοινωνία μεταξύ δυο μερών, απότομες αλλαγές ύφους ή φράσεις που υποδηλώνουν πίεση ή προσπάθεια συγκάλυψης. Ωστόσο, η τελική ερμηνεία τέτοιων ενδείξεων απαιτεί ανθρώπινη κρίση, διότι πολλές από αυτές τις γλωσσικές διαφοροποιήσεις μπορεί να είναι αθώες ή να σχετίζονται με περιστασιακά συμβάντα που δεν αποτελούν πραγματικό κίνδυνο. Αντίστοιχα, σε συστήματα ML που εντοπίζουν ύποπτες συναλλαγές με βάση στατιστικές αποκλίσεις, η παρουσία του ανθρώπου διασφαλίζει ότι το σύστημα δεν θα χαρακτηρίσει ως ύποπτες, συναλλαγές που είναι απολύτως νόμιμες αλλά σπάνιες λόγω της φύσης της επιχειρησιακής δραστηριότητας.

Η λειτουργική αρχιτεκτονική ενός υβριδικού συστήματος στον εσωτερικό έλεγχο βασίζεται σε ένα μοντέλο συνεχούς αλληλεπίδρασης ανθρώπου και ΤΝ. Το σύστημα συλλέγει δεδομένα από πολλαπλές πηγές, αναλύει μοτίβα, εντοπίζει ανωμαλίες, εκπέμπει ειδοποιήσεις και παρουσιάζει στον ελεγκτή μια συνοπτική εικόνα των ενδεχόμενων κινδύνων. Ο ελεγκτής εξετάζει τα ευρήματα, προσθέτει στοιχεία που το σύστημα δεν λαμβάνει υπόψη του, διατυπώνει συμπεράσματα και είτε προχωρά σε περαιτέρω έρευνα είτε απορρίπτει την ανωμαλία ως μη κρίσιμη. Επιπλέον, η ανθρώπινη διάδραση με τα συστήματα ΤΝ μπορεί να χρησιμοποιηθεί για την εκ νέου εκπαίδευση των αλγορίθμων, συμβάλλοντας έτσι στη βελτίωσή τους.

Εκτός των παραπάνω, η ανάγκη ανθρώπινης εποπτείας ενισχύεται από το ζήτημα της αλγοριθμικής μεροληψίας. Τα συστήματα ΤΝ εκπαιδεύονται με βάση τα δεδομένα που τους παρέχονται και, ως εκ τούτου, μπορεί να αναπαραγάγουν υφιστάμενες ανισότητες ή εσφαλμένες τάσεις που υπήρχαν στο παρελθόν. Χωρίς τον άνθρωπο να αξιολογεί τα αποτελέσματα, ένας αλγόριθμος μπορεί να στοχεύει συστηματικά συγκεκριμένες ομάδες εργαζομένων ή τμημάτων, δημιουργώντας αδικίες και επηρεάζοντας αρνητικά την εταιρική κουλτούρα. Η συνεχής συμμετοχή του ανθρώπου επιτρέπει την αναγνώριση τέτοιων συμπεριφορών, εξασφαλίζοντας ότι η διαδικασία παραμένει δίκαιη και δεοντολογικά συμβατή.

Οι πρακτικές εφαρμογές υβριδικών συστημάτων είναι πολυάριθμες. Στον τραπεζικό τομέα, τα συστήματα ανίχνευσης ξεπλύματος χρήματος εφαρμόζουν ML (μηχανική μάθηση) για την ανάλυση χιλιάδων συναλλαγών ανά λεπτό, αλλά η τελική έγκριση των

αποτελεσμάτων γίνεται από εξειδικευμένους αναλυτές. Σε ελεγκτικές εταιρείες, η ΤΝ χρησιμοποιείται για τον έλεγχο μεγάλων όγκων λογιστικών εγγραφών, τον εντοπισμό μοτίβων και την υποστήριξη των ελεγκτών στην εκτίμηση κινδύνου και στην επιλογή περιοχών που απαιτούν λεπτομερέστερο έλεγχο (Deloitte, 2024). Στον δημόσιο τομέα, υβριδικά μοντέλα υποστηρίζουν την ανίχνευση υπερκοστολογήσεων ή ύποπτων διαγωνισμών, όμως η τελική απόφαση αποτελεί και πάλι ευθύνη της αρμόδιας υπηρεσίας.

Παρά τα σημαντικά οφέλη, τα υβριδικά μοντέλα δημιουργούν προκλήσεις. Η υπερβολική εμπιστοσύνη στην ΤΝ μπορεί να μειώσει την κριτική ικανότητα των ελεγκτών, καθιστώντας τους λιγότερο δραστήριους στη διερεύνηση κινδύνων. Η λειτουργική πολυπλοκότητα τέτοιων συστημάτων απαιτεί σημαντικούς πόρους και εξειδικευμένο προσωπικό. Επιπλέον, ελλοχεύει ο κίνδυνος η ανθρώπινη εποπτεία να περιοριστεί τυπικά, χωρίς ουσιαστική εμπλοκή, εάν η διοίκηση δεν διασφαλίσει κατάλληλη κουλτούρα ευθύνης και λογοδοσίας.

Το μέλλον των υβριδικών συστημάτων στον εσωτερικό έλεγχο αναμένεται να περιλαμβάνει ακόμη πιο δυναμικές και προσαρμοστικές μορφές συνεργασίας. Οι αλγόριθμοι θα γίνονται πιο διαφανείς, χάρη στις τεχνικές Explainable AI, ενώ η ανθρώπινη εποπτεία θα γίνεται πιο πολύπλοκη και πιο κρίσιμη. Η ενσωμάτωση AI auditing, δηλαδή ο έλεγχος των ίδιων των αλγορίθμων ως μέρος του ελέγχου, θα δημιουργήσει νέες προκλήσεις αλλά και νέες ευκαιρίες για την ενίσχυση της αξιοπιστίας και της διαφάνειας (Floridi et al., 2021).

Συνοψίζοντας, τα υβριδικά συστήματα συνεργασίας ανθρώπου-μηχανής αποτελούν τον πιο αποτελεσματικό τρόπο αξιοποίησης της ΤΝ στον εσωτερικό έλεγχο. Μολονότι η ΤΝ προσφέρει ισχυρές αναλυτικές δυνατότητες και υποστηρίζει τον εντοπισμό κινδύνων και red flags, η ανθρώπινη κρίση παραμένει αναντικατάστατη. Ο συνδυασμός των δύο οδηγεί σε ένα ολοκληρωμένο και δεοντολογικά ορθό πλαίσιο λήψης αποφάσεων, το οποίο ενισχύει την εταιρική διακυβέρνηση, μειώνει την πιθανότητα λαθών, και αυξάνει την αξιοπιστία και την αποδοχή των συστημάτων ΤΝ από τους εργαζομένους και τους ελεγκτές.

2.4 Οι “Agents” στην Τεχνητή Νοημοσύνη και ο Ρόλος τους στην Ανίχνευση Red Flags

Στο πεδίο της Τεχνητής Νοημοσύνης (ΤΝ), ο όρος agent αποτελεί μια κεντρική και θεμελιώδη έννοια, καθώς περιγράφει μια αυτόνομη ή ημιαυτόνομη οντότητα που μπορεί να αντιλαμβάνεται το περιβάλλον της, να λαμβάνει αποφάσεις και να ενεργεί με στόχο τη βελτιστοποίηση ενός συγκεκριμένου δείκτη απόδοσης (Russell & Norvig, 2021). Η λειτουργία ενός agent διακρίνεται σε τρία βασικά στάδια: αντίληψη περιβάλλοντος, λήψη αποφάσεων και εκτέλεση δράσεων. Σε αντίθεση με παραδοσιακά λογισμικά συστήματα, τα οποία λειτουργούν αυστηρά βάσει προκαθορισμένων κανόνων, ένας agent διαθέτει τη δυνατότητα προσαρμογής και μάθησης μέσα από την αλληλεπίδραση με το περιβάλλον, γεγονός που καθιστά την έννοια κρίσιμη για σύγχρονες εφαρμογές της ΤΝ.

Ένα σημαντικό χαρακτηριστικό των agents είναι η αυτονομία. Όσο λιγότερο απαιτούν ανθρώπινη παρέμβαση για να λάβουν αποφάσεις, τόσο υψηλότερο είναι το επίπεδο αυτονομίας τους. Ωστόσο, η αυτονομία δεν είναι μόνο τεχνικό χαρακτηριστικό, φέρει και σοβαρές ηθικές και λειτουργικές προεκτάσεις, ειδικά σε εφαρμογές όπως η ανίχνευση

απάτης, η εταιρική συμμόρφωση και ο εσωτερικός έλεγχος. Η ανάγκη για διαφάνεια, λογοδοσία και ορθή τεκμηρίωση σημαίνει ότι ακόμη και πολύ εξελιγμένοι agents πρέπει να λειτουργούν μέσα σε πλαίσιο ανθρώπινης εποπτείας (Wachter, Mittelstadt, 2019).

Η έννοια του agent συνδέεται άμεσα με την ανίχνευση εταιρικών red flags. Ένας agent μπορεί να παρακολουθεί αδιάκοπα συναλλαγές, αλληλεπιδράσεις, συμπεριφορές χρηστών και λειτουργικές διαδικασίες, αναγνωρίζοντας μοτίβα που αποκλίνουν από το κανονικό. Οι agents χρησιμοποιούνται σε συστήματα anomaly detection, NLP και ML, όπου αναλύουν δεδομένα όπως λογιστικές εγγραφές, αρχεία προμηθειών, email, αρχεία πρόσβασης και δείκτες κινδύνου. Με βάση αυτό το σύνολο πληροφοριών, ένας agent μπορεί να ενεργοποιήσει ειδοποιήσεις (alerts) όταν εντοπίσει ενδείξεις που μοιάζουν με γνωστά σχήματα απάτης ή έντονα αποκλίνοντα μοτίβα συμπεριφοράς (Ngai et al., 2011).

Για παράδειγμα, σε ένα σύστημα AML (Anti-Money Laundering), ο agent παρακολουθεί χιλιάδες συναλλαγές σε πραγματικό χρόνο. Εφόσον εντοπίσει απότομη αύξηση δραστηριότητας σε λογαριασμό με χαμηλό ιστορικό, ή συναλλαγές διαφορετικές από το προφίλ του χρήστη, θα ενεργοποιήσει ένα red flag και θα ενημερώσει τον ανθρώπινο ελεγκτή. Σε ένα σύστημα NLP, ο agent μπορεί να αναλύει εταιρική επικοινωνία και να εντοπίζει ύποπτες φράσεις που υποδηλώνουν πιέσεις, συγκάλυψη, προσπάθεια αποφυγής διαδικασιών ή ασυνήθιστη οικειότητα μεταξύ υπαλλήλων και προμηθευτών. Αυτές οι λειτουργίες ενδέχεται να συνδέονται με περιπτώσεις απάτης, σύγκρουσης συμφερόντων ή παραβίασης εσωτερικών κανόνων.

Οι agents παίζουν καθοριστικό ρόλο στην αποφυγή false positives και false negatives. Ένας agent μπορεί να λειτουργεί με μεγάλη ακρίβεια όσον αφορά την ανίχνευση μοτίβων, αλλά δεν έχει την ικανότητα να κατανοήσει τα συμφραζόμενα, όπως άτυπες επαγγελματικές πρακτικές ή ειδικές συνθήκες που μπορεί να εξηγούν μια ανωμαλία χωρίς να υποδηλώνει απάτη. Ο άνθρωπος, αντιθέτως, μπορεί να καταλάβει ότι μια φαινομενικά ύποπτη ενέργεια είναι απλώς μια προσωρινή απόκλιση λόγω αλλαγών στην παραγωγή, στον όγκο εργασίας ή σε εξωτερικούς παράγοντες. Η συνδυαστική λειτουργία agent–ελεγκτή αποτελεί επομένως καθοριστικό στοιχείο της σύγχρονης ελεγκτικής πρακτικής (Alles et al, 2014).

Σημαντική εφαρμογή των agents βρίσκεται επίσης σε multi-agent συστήματα, όπου πολλοί agents συνεργάζονται και ανταλλάσσουν πληροφορίες. Ένα multi-agent σύστημα μπορεί να περιλαμβάνει διαφορετικούς agents: έναν agent NLP που αναλύει επικοινωνίες, έναν agent ML που εντοπίζει ανωμαλίες στα λογιστικά αρχεία, έναν agent risk scoring που υπολογίζει δείκτες κινδύνου και έναν agent monitoring που συγκρίνει τις αποκλίσεις με ιστορικά δεδομένα. Η συνεργασία όλων αυτών παράγει έναν πιο ολοκληρωμένο μηχανισμό ανίχνευσης red flags από ό,τι θα μπορούσε να επιτύχει κάθε agent ξεχωριστά (Stone & Veloso, 2000).

Συνολικά, οι agents αποτελούν βασικούς μηχανισμούς της σύγχρονης ΤΝ, με καθοριστικό ρόλο στην ανίχνευση εταιρικών red flags. Λειτουργούν ως συστήματα που παρακολουθούν, αναλύουν, μαθαίνουν και υποδεικνύουν πιθανούς κινδύνους, ενώ παράλληλα εντάσσονται σε ένα ευρύτερο πλαίσιο ανθρώπινης εποπτείας και διακυβέρνησης, διασφαλίζοντας ότι η τελική απόφαση παραμένει ορθολογική, δίκαιη και σύμφωνη με τις αρχές του εσωτερικού ελέγχου.

Κεφάλαιο 3 - Μοντέλα Τεχνητής Νοημοσύνης – Είδη, Λειτουργία και Εφαρμογές

Η Τεχνητή Νοημοσύνη (ΤΝ) δεν αποτελεί μία ενιαία τεχνολογία, αλλά ένα σύνολο μεθόδων και μοντέλων που βασίζονται σε διαφορετικές μαθηματικές και στατιστικές προσεγγίσεις. Η έκφραση "Μοντέλα ΤΝ" αναφέρεται σε υπολογιστικά συστήματα που έχουν εκπαιδευτεί σε τεράστιους όγκους δεδομένων για να μπορούν να εκτελούν συγκεκριμένες εργασίες που απαιτούν νοημοσύνη.
Ένα μοντέλο ΤΝ είναι ουσιαστικά το "προϊόν" της διαδικασίας εκπαίδευσης. Μπορεί να θεωρηθεί ως μια μαθηματική αναπαράσταση των μοτίβων και των σχέσεων που βρήκε ο αλγόριθμος στα δεδομένα της εκπαίδευσης.

- **Αλγόριθμος:** Είναι το σύνολο των κανόνων και των διαδικασιών που χρησιμοποιούνται για την εκμάθηση
- **Δεδομένα:** Είναι οι πληροφορίες (κείμενα, εικόνες, ήχοι) που χρησιμοποιούνται για να "μάθει" το μοντέλο (π.χ., εκατομμύρια άρθρα για να μάθει τη γλώσσα).
- **Μοντέλο:** Είναι το εκπαιδευμένο αποτέλεσμα. Είναι έτοιμο να πάρει μια είσοδο (π.χ., μια ερώτηση) και να παράγει μια έξοδο (π.χ., μια απάντηση ή μια εικόνα) προβλέποντας την πιθανότερη σωστή απάντηση βάσει όσων έχει μάθει.

Η λειτουργία ενός μοντέλου ΤΝ περιλαμβάνει τρία βασικά στάδια:

1. **Μοντελοποίηση (Modeling):** Αναπτύσσεται ένας σύνθετος αλγόριθμος ή ένα σύνολο αλγορίθμων (όπως ένα νευρωνικό δίκτυο) για την ανάλυση δεδομένων.
2. **Εκπαίδευση:** Ο αλγόριθμος τροφοδοτείται με ένα τεράστιο σύνολο δεδομένων. Μέσω επαναληπτικών διαδικασιών, το μοντέλο προσαρμόζει τις εσωτερικές του παραμέτρους (τα λεγόμενα βάρη) για να ελαχιστοποιήσει το σφάλμα του και να μάθει να αναγνωρίζει μοτίβα.
    - Παράδειγμα: Σε ένα μοντέλο αναγνώρισης εικόνων, το μοντέλο "μαθαίνει" ποια χαρακτηριστικά (γραμμές, χρώματα) αντιστοιχούν σε μια γάτα και ποια σε έναν σκύλο.
3. **Εφαρμογή (Deployment), Πρόβλεψη/Δημιουργία:** Αφού ολοκληρωθεί η εκπαίδευση, το μοντέλο μπορεί να λάβει νέα δεδομένα που δεν έχει ξαναδεί και να εφαρμόσει τα μοτίβα που έμαθε για να κάνει μια πρόβλεψη ή να δημιουργήσει νέο περιεχόμενο.
    - Παράδειγμα: Αν του δώσεις την εικόνα μιας γάτας, θα προβλέψει με υψηλή πιθανότητα ότι είναι "γάτα".

Τα μοντέλα ΤΝ διαφέρουν ανάλογα με την εργασία για την οποία εκπαιδεύονται. Μερικά από αυτά παρατίθενται παρακάτω:

- **Μεγάλα Γλωσσικά Μοντέλα (LLMs):** Όπως το ChatGPT, το Gemini ή το Copilot. Εκπαιδεύονται σε τεράστια σύνολα κειμένου και μπορούν να

κατανοήσουν, να μεταφράσουν, να απαντήσουν σε ερωτήσεις και να δημιουργήσουν κείμενο.

- **Μοντέλα Παραγωγής Εικόνας (Generative AI):** Όπως το DALL-E. Εκπαιδεύονται σε εικόνες και κείμενα και μπορούν να δημιουργήσουν πρωτότυπες εικόνες από μια περιγραφή σε κείμενο.
- **Μοντέλα Πρόβλεψης:** Χρησιμοποιούνται για να προβλέπουν αποτελέσματα (π.χ., αν μια πιστωτική κάρτα έχει κλαπεί, ή ποια μετοχή θα ανέβει).

Πίνακας 1. Τύποι μοντέλων, στόχοι και παραδείγματα

Τύπος Μοντέλου	**Στόχος**	**Παραδείγματα Εφαρμογών**
Μοντέλα Μηχανικής Μάθησης (ML)	Αναγνώριση μοτίβων και λήψη αποφάσεων από δεδομένα.	Πρόβλεψη τιμών κατοικιών, ανάλυση πιστωτικού κινδύνου.
Μοντέλα Βαθιάς Μάθησης (Deep Learning)	Πολύπλοκες εργασίες χρησιμοποιώντας πολυεπίπεδα νευρωνικά δίκτυα.	Αναγνώριση προσώπου, επεξεργασία φυσικής γλώσσας (NLP).
Μεγάλα Γλωσσικά Μοντέλα (LLMs)	Κατανόηση και παραγωγή ανθρώπινης γλώσσας (κείμενο).	Chatbots (π.χ., ChatGPT, Gemini), μετάφραση, περίληψη κειμένων.
Γενετική AI (Generative AI)	Δημιουργία νέου περιεχομένου (όχι μόνο πρόβλεψη).	Παραγωγή εικόνων από κείμενο (Text-to-Image), σύνθεση μουσικής, δημιουργία κώδικα.
Μοντέλα Όρασης Υπολογιστών (Computer Vision)	Κατανόηση και επεξεργασία εικόνων και βίντεο.	Αυτόνομα οχήματα, ιατρική διάγνωση από ακτινογραφίες.

Πηγή: Ιδιοκατασκευή

Τα μοντέλα αυτά έχουν ως κοινό στόχο την ανάλυση δεδομένων, την εκμάθηση προτύπων και την πρόβλεψη ή αναγνώριση συμπεριφορών. Στην πράξη, η επιλογή του κατάλληλου μοντέλου εξαρτάται από τη φύση των δεδομένων, τον όγκο τους, την ανάγκη ερμηνευσιμότητας και την ακρίβεια που απαιτείται. Η κατανόηση των διαφόρων μοντέλων είναι κρίσιμη για την εφαρμογή τους στην ανίχνευση red flags, καθώς κάθε μοντέλο έχει διαφορετική ικανότητα να «μαθαίνει» από δεδομένα και να εντοπίζει ανωμαλίες ή ύποπτα μοτίβα.

3.1 Natural Language Processing (NLP) και ο ρόλος του στην ανίχνευση απάτης και red flags

Η επεξεργασία της φυσικής γλώσσας (Natural Language Processing, NLP) αποτελεί έναν κρίσιμο άξονα της σύγχρονης τεχνητής νοημοσύνης, με την ικανότητα να μετατρέπει το ανθρώπινο λόγο, γραπτό ή προφορικό, σε δεδομένα που μπορούν να αναλυθούν από υπολογιστικά συστήματα για την εξαγωγή αξιόπιστων συμπερασμάτων. Στο πλαίσιο της ανίχνευσης εταιρικής απάτης και των red flags ,δηλαδή ενδείξεων που υποδεικνύουν πιθανή δόλια συμπεριφορά ή παρατυπία, το NLP αναδεικνύεται ως ένα εργαλείο με ιδιαίτερη δυναμική. Οι λύσεις που βασίζονται στο NLP επιτρέπουν την κατανόηση και ανάλυση μη δομημένων στοιχείων, όπως είναι τα μηνύματα ηλεκτρονικού ταχυδρομείου, σημειώσεις εργαζομένων, συμβάσεις, τιμολόγια ή επικοινωνίες με τους πελάτες, τα οποία παραδοσιακά είναι δυσκολότερα στην ανίχνευση και την ερμηνεία από τα υπάρχοντα συμβατικά συστήματα ελέγχου.

Η λειτουργία του NLP στηρίζεται σε τρεις βασικές διαστάσεις: την κατανόηση της γλώσσας (π.χ. αναγνώριση όρων, οντοτήτων, σχέσεων), την επεξεργασία της (π.χ. ανάλυση συναισθήματος, ανίχνευση ανωμαλιών στο ύφος ή στο περιεχόμενο) και τη μετατροπή της σε μορφή που μπορεί να αξιοποιηθεί στην ανάλυση κινδύνου και στην εξακρίβωση ύποπτων μοτίβων. Για παράδειγμα, η ανάλυση συναισθήματος (sentiment analysis) μπορεί να αναδείξει ύποπτη ενδοεταιρική επικοινωνία, ενώ μέθοδοι όπως η αναγνώριση οντότητας (named entity recognition) εντοπίζουν άτομα, εταιρείες ή οικονομικά ποσά που εμφανίζονται σε μη αναμενόμενα πλαίσια. Τα συστήματα NLP εφαρμόζονται πλέον με επιτυχία και στον τομέα της ανίχνευσης απάτης (IBM, 2025).

Η χρήση του NLP στην ανίχνευση απάτης βασίζεται στο γεγονός ότι πολλές δόλιες συναλλαγές ή επικοινωνίες αφήνουν πίσω τους «γλωσσικά ίχνη» ή ασυνήθιστα μοτίβα, είτε στο περιεχόμενο είτε στο ύφος (IBM, 2025). Έτσι, το NLP μπορεί να εντοπίσει red flags όπως: ασυνήθιστα μεγάλες ή περίπλοκες φράσεις σε τιμολόγια, επαναλαμβανόμενες επικοινωνίες με συγκεκριμένα μοτίβα, χρήση όρων που συνήθως δεν συνδέονται με νόμιμες συναλλαγές, ή αποκλίσεις στο ύφος της επικοινωνίας εργαζομένου πριν από ύποπτη δραστηριότητα.

Η εφαρμογή της επεξεργασίας φυσικής γλώσσας στις εταιρικές διαδικασίες μπορεί να δημιουργήσει σημαντικά πλεονεκτήματα: ταχύτητα ανάλυσης μεγάλου όγκου δεδομένων, δυνατότητα επεξεργασίας μη δομημένων στοιχείων που συχνά περιέχουν σημαντικές πληροφορίες απάτης, και βελτιωμένη ακρίβεια εντοπισμού ύποπτων μοτίβων (BioCatch, 2024). Για παράδειγμα, μελέτη της BioCatch αναφέρει ότι το 72 % των τραπεζών που εφαρμόζουν λύσεις τεχνητής νοημοσύνης χρησιμοποιούν NLP για την ανίχνευση χρηματοοικονομικών εγκλημάτων.

Ωστόσο, δεν λείπουν οι προκλήσεις. Πρώτον, η ποιότητα των δεδομένων μη δομημένου κειμένου είναι κρίσιμη, η ύπαρξη κακογραμμένων, ανορθόγραφων ή ατελών εγγράφων δυσχεραίνει την εξαγωγή αξιόπιστων χαρακτηριστικών. Δεύτερον, η επεξηγηματικότητα των μοντέλων NLP (ιδιαίτερα αυτών που βασίζονται σε βαθιά μάθηση) είναι μειωμένη, γεγονός που δυσκολεύει την αποδοχή των ευρημάτων από τους ελεγκτές και την διοίκηση. Τρίτον, ανακύπτουν θέματα ιδιωτικότητας και συμμόρφωσης, η επεξεργασία επικοινωνιών ή εσωτερικών emails χρήζει προσεκτικής διαχείρισης σύμφωνα με τους κανονισμούς (π.χ. GDPR). Τέλος, η κακή χρήση του NLP μπορεί να οδηγήσει σε ψευδώς θετικά ή αρνητικά αποτελέσματα, που δεν ανταποκρίνονται σε πραγματική απάτη.

Παρά τις προκλήσεις, η ενσωμάτωση του NLP στην ανίχνευση red flags και εταιρικής απάτης φαίνεται να αποτελεί φυσική εξέλιξη για τους οργανισμούς που επιδιώκουν βιωσιμότητα, διαφάνεια και συμμόρφωση. Σε ελληνικό πλαίσιο, όπως καταγράφεται σε μεταπτυχιακή εργασία του Πανεπιστημίου Μακεδονίας, η ενσωμάτωση τεχνολογιών όπως η επεξεργασία φυσικής γλώσσας έχει αποδειχθεί ότι ενισχύει την αποτελεσματικότητα των ελεγκτικών μηχανισμών και συμβάλλει στον έγκαιρο εντοπισμό παρατυπιών (Γενίτσαρης, 2024). Είναι σαφές ότι σε ένα περιβάλλον όπου οι απάτες γίνονται πιο πολύπλοκες και πολυδιάστατες, το NLP παρέχει ένα «διευρυμένο μάτι» στα μη δομημένα δεδομένα, εκεί όπου κατά κανόνα οι παραδοσιακές διαδικασίες ελέγχου δεν είχαν ορατότητα.

3.2 Μηχανική Μάθηση (Machine Learning)

Η Μηχανική Μάθηση αποτελεί έναν κλάδο της Τεχνητής Νοημοσύνης που επιτρέπει στα συστήματα να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους χωρίς να είναι ρητά προγραμματισμένα. Τα μοντέλα μαθαίνουν μοτίβα, συσχετίσεις και ανωμαλίες μέσα από μεγάλο όγκο παρατηρήσεων και μπορούν να χρησιμοποιηθούν για ταξινόμηση, πρόβλεψη ή ανίχνευση αποκλίσεων. Στο πλαίσιο της ανίχνευσης red flags, η μηχανική μάθηση προσφέρει μεθόδους όπως δέντρα αποφάσεων, και αλγορίθμους ανίχνευσης ανωμαλιών, επιτρέποντας την αναγνώριση ύποπτων συμπεριφορών με υψηλότερη ακρίβεια από τις παραδοσιακές μεθόδους. Η αξία της έγκειται κυρίως στην ικανότητά της να εντοπίζει κρυφές δομές σε δεδομένα και να προσαρμόζεται σε δυναμικά περιβάλλοντα (El Naqa, 2015)

3.3 Βαθιά Μάθηση (Deep Learning)

Η Βαθιά Μάθηση αποτελεί υποκατηγορία της μηχανικής μάθησης και βασίζεται στη χρήση πολυεπίπεδων νευρωνικών δικτύων που μαθαίνουν απευθείας από αδόμητα ή ημιδομημένα δεδομένα. Τα συστήματα deep learning χρησιμοποιούν διαδοχικά επίπεδα επεξεργασίας για να εξάγουν υψηλού επιπέδου αναπαραστάσεις χαρακτηριστικών, γεγονός που τα καθιστά ιδανικά για πολύπλοκα προβλήματα, όπως ανάλυση κειμένου, ανίχνευση ανωμαλιών σε μεγάλους όγκους δεδομένων και πρόβλεψη προτύπων συμπεριφοράς. Στην ανίχνευση red flags, τα νευρωνικά δίκτυα μπορούν να εντοπίσουν δύσκολα μοτίβα απάτης, να αναλύσουν περιγραφές συναλλαγών, ηλεκτρονική αλληλογραφία και λογιστικά δεδομένα με τρόπο που υπερβαίνει τις δυνατότητες των παραδοσιακών μοντέλων. Παρά την ισχυρή τους απόδοση, απαιτούν προσεκτική ρύθμιση, αυξημένη υπολογιστική ισχύ και ισχυρή ανθρώπινη εποπτεία για την αποφυγή λαθών ή παραπλανητικών ενδείξεων (LeCun et al, 2015)

3.4 Τα Δέντρα Αποφάσεων (Decision Trees) και η χρήση τους στην ανίχνευση εταιρικών red flags

Τα Δέντρα Αποφάσεων (Decision Trees) αποτελούν έναν από τους πιο διαδεδομένους και κατανοητούς αλγορίθμους μηχανικής μάθησης, ιδιαίτερα χρήσιμους σε περιβάλλοντα όπου απαιτείται ερμηνευσιμότητα και διαφάνεια. Πρόκειται για μοντέλα ταξινόμησης που βασίζονται σε μια ιεραρχική δομή, η οποία μοιάζει με δέντρο: κάθε εσωτερικός κόμβος αντιστοιχεί σε μια συνθήκη (κανόνα), κάθε κλαδί σε μια απόφαση που προκύπτει από τη συνθήκη, και κάθε φύλλο σε μια τελική πρόβλεψη (Quinlan, 2014).

Το μοντέλο χωρίζει τα δεδομένα σε υποσύνολα με τρόπο που μειώνει την ανομοιογένεια. Έτσι δημιουργείται μια σειρά από κανόνες του «ναι/όχι» που οδηγούν στην τελική απόφαση. Το γεγονός ότι κάθε απόφαση είναι ορατή και διαφανής καθιστά τα δέντρα αποφάσεων ιδιαιτέρως ελκυστικά για τον εσωτερικό έλεγχο, όπου απαιτείται η δυνατότητα τεκμηρίωσης και αιτιολόγησης κάθε ευρήματος.
Στην ανίχνευση εταιρικών red flags, τα Decision Trees είναι αποτελεσματικά επειδή μπορούν να ενσωματώσουν πολύπλοκες σχέσεις μεταξύ μεταβλητών χωρίς να απαιτείται προηγμένη παραμετροποίηση. Μπορούν, για παράδειγμα, να εντοπίσουν ύποπτα μοτίβα όπως:

- συχνές χειροκίνητες τροποποιήσεις λογιστικών εγγραφών,
- αιφνίδιες αλλαγές στη συμπεριφορά προμηθευτών,
- συναλλαγές που αποκλίνουν σημαντικά από τον μέσο όρο ή τα ιστορικά πρότυπα,
- εργαζόμενους με χρονικά ασυνήθιστα επίπεδα πρόσβασης σε συστήματα.

Επιπλέον, ο οπτικός χαρακτήρας των Δέντρων Αποφάσεων βοηθά τους ελεγκτές να κατανοήσουν γιατί μια συναλλαγή χαρακτηρίστηκε ως ύποπτη, κάτι που δεν είναι πάντα εφικτό σε πιο πολύπλοκα μοντέλα όπως τα νευρωνικά δίκτυα. Έτσι, τα Decision Trees αποτελούν στοιχείο-κλειδί σε υβριδικά μοντέλα AI–Human, καθώς προσφέρουν ένα επίπεδο ερμηνευσιμότητας που ενισχύει τόσο την αξιοπιστία όσο και την αποδοχή του συστήματος από τον ελεγκτικό μηχανισμό.

Κεφάλαιο 4 - Διαδικασία Ανάπτυξης Συστήματος Τεχνητής Νοημοσύνης για την Ανίχνευση Εταιρικών Red Flags

4.1 Προσδιορισμός Περιοχών Κινδύνου και Χαρτογράφηση Red Flags

Το πρώτο και πιο κρίσιμο βήμα στην ανάπτυξη ενός συστήματος ανίχνευσης πιθανών αποκλίσεων, παρατυπιών ή φαινομένων απάτης με τη χρήση Τεχνητής Νοημοσύνης (ΤΝ) αποτελεί ο προσδιορισμός των περιοχών κινδύνου και η χαρτογράφηση των red flags. Σε έναν σύγχρονο οργανισμό, ο εσωτερικός έλεγχος καλείται να λειτουργήσει σε ένα περιβάλλον μεγάλης πολυπλοκότητας, όπου οι κίνδυνοι εμφανίζονται σε πολλαπλές διαδικασίες, υπό διαφορετικές μορφές και με ποικίλες επιπτώσεις. Για τον λόγο αυτό, η συστηματική χαρτογράφηση των περιοχών κινδύνου και των ενδείξεων που μπορεί να συνδέονται με ύποπτες ή μη κανονικές συμπεριφορές αποτελεί θεμέλιο της ελεγκτικής στρατηγικής αλλά και αναγκαία προϋπόθεση για την ορθή εφαρμογή των μοντέλων ΤΝ.

Ο εντοπισμός των περιοχών κινδύνου περιλαμβάνει την αναγνώριση όλων των λειτουργιών και διαδικασιών όπου υπάρχει πιθανότητα εμφάνισης απάτης, λανθασμένης πληροφόρησης, καταχρήσεων ή μη συμμόρφωσης. Σύμφωνα με το Committee of Sponsoring Organizations of the Treadway Commission (COSO), η διαδικασία αναγνώρισης κινδύνων πρέπει να λαμβάνει υπόψη τόσο εξωτερικούς όσο και εσωτερικούς παράγοντες, καθώς και το επίπεδο των υφιστάμενων δικλίδων ελέγχου (COSO, 2017). Επισημαίνεται, ότι η ταχύτητα των αλλαγών, η ψηφιοποίηση των διαδικασιών και η αυξανόμενη εξάρτηση από δεδομένα δημιουργούν νέους κινδύνους, φαινόμενο στο οποίο η έγκαιρη και ακριβής αναγνώριση red flags αποκτά ιδιαίτερη αξία (Kogan et al., 2014).

Η χαρτογράφηση red flags αποτελεί μια δομημένη διαδικασία μέσω της οποίας ένα σύνολο ενδείξεων συνδέεται με συγκεκριμένους τύπους κινδύνων.

Η διαδικασία χαρτογράφησης των red flags επεκτείνεται με τη χρήση ΤΝ, καθώς τα μοντέλα ΤΝ μπορούν να αναλύουν δεδομένα που πριν ήταν πρακτικά αδύνατο να εξεταστούν χειρωνακτικά. Τα μοντέλα NLP μπορούν να εξαγάγουν μοτίβα από e-mails, αρχεία συνομιλιών ή έγγραφα, τα μοντέλα machine learning μπορούν να εντοπίζουν αριθμητικές ανωμαλίες και κρυφές συσχετίσεις στα λογιστικά στοιχεία, ενώ τα συστήματα anomaly detection μπορούν να επισημαίνουν αποκλίσεις που θα περνούσαν απαρατήρητες από τον άνθρωπο. Η συμβολή της ΤΝ έγκειται στο ότι μπορεί να μετατρέψει «σήματα» που δεν είναι προφανή σε ενδείξεις κινδύνου με υψηλή πιθανότητα αξιοπιστίας.

Ωστόσο, απαιτεί προσοχή για δύο βασικούς λόγους. Πρώτον, τα συστήματα ΤΝ εντοπίζουν στατιστικές αποκλίσεις αλλά όχι αναγκαστικά απάτη. Ένα red flag αποτελεί απλώς ένδειξη και όχι απόδειξη παρατυπίας. Δεύτερον, τα μοντέλα ΤΝ επηρεάζονται από τα δεδομένα εκπαίδευσης και από τον τρόπο δόμησης των αλγορίθμων. Αυτό σημαίνει ότι μπορεί να προκύψουν τόσο false positives όσο και false negatives. Για παράδειγμα, μια απότομη μεταβολή στην παραγωγή μπορεί να θεωρηθεί λανθασμένα ύποπτη, ενώ μια προσεκτικά καλυμμένη απάτη μπορεί να μη γίνει αντιληπτή χωρίς τον κατάλληλο αλγόριθμο. Σε αυτό το πλαίσιο, η ανθρώπινη εποπτεία και η χρήση υβριδικών συστημάτων αποτελούν απαραίτητες προϋποθέσεις για την αξιοπιστία των αποτελεσμάτων (Warren, Moffitt & Byrnes, 2015).

Τέλος, ο προσδιορισμός περιοχών κινδύνου και η χαρτογράφηση red flags διαδραματίζουν σημαντικό ρόλο στη διασφάλιση της συμμόρφωσης με το κανονιστικό πλαίσιο. Τα μοντέλα ΤΝ πρέπει να ευθυγραμμίζονται με τις απαιτήσεις του GDPR και του AI Act, ιδιαιτέρως όσον αφορά την επεξήγηση αποτελεσμάτων, την επεξεργασία προσωπικών δεδομένων και την ύπαρξη ανθρώπινης εποπτείας σε συστήματα υψηλού κινδύνου. Η σωστή χαρτογράφηση των red flags διευκολύνει τη δημιουργία τέτοιων μηχανισμών συμμόρφωσης, καθιστώντας το σύστημα πιο διαφανές και αξιόπιστο τόσο για τους ελεγκτές όσο και για τη διοίκηση.
Συνολικά, ο προσδιορισμός των περιοχών κινδύνου και η χαρτογράφηση των red flags αποτελούν θεμέλιο της διαδικασίας ανίχνευσης παρατυπιών μέσω ΤΝ. Η διαδικασία απαιτεί συνδυασμό τεχνικής κατανόησης, επιχειρησιακής γνώσης και ελεγκτικής εμπειρίας, ενώ η χρήση ΤΝ ενισχύει σημαντικά την ικανότητα των ελεγκτών να εντοπίζουν κρίσιμες αποκλίσεις. Σε αυτό το πλαίσιο, η υιοθέτηση ενός ολοκληρωμένου συστήματος που συνδυάζει ανθρώπινη επίβλεψη και αλγοριθμική ανάλυση επιτρέπει τη δημιουργία ενός πιο αποτελεσματικού και αξιόπιστου πλαισίου εσωτερικού ελέγχου.

4.2 Δεδομένα, πηγές και ποιότητα πληροφορίας στην ανίχνευση Red Flags με ΤΝ

Η ανάπτυξη οποιουδήποτε συστήματος Τεχνητής Νοημοσύνης για την ανίχνευση εταιρικών red flags προϋποθέτει την ύπαρξη κατάλληλων δεδομένων, τόσο ως προς τη διαθεσιμότητα και τη δομή τους όσο και ως προς την ποιότητα και την αξιοπιστία τους. Στον εσωτερικό έλεγχο, τα δεδομένα δεν αποτελούν απλώς «πρώτη ύλη» για τις αλγοριθμικές μεθόδους αλλά θεμέλιο στοιχείο πάνω στο οποίο στηρίζονται τα ελεγκτικά συμπεράσματα και η αποτύπωση του κινδύνου. Αμφότερες, η διεθνής και η ελληνική βιβλιογραφία αναδεικνύουν ότι, χωρίς επαρκή και ποιοτικά δεδομένα, ακόμη και τα πιο εξελιγμένα μοντέλα ΤΝ αδυνατούν να παραγάγουν αξιόπιστα και χρήσιμα αποτελέσματα για τον εσωτερικό έλεγχο και την ανίχνευση απάτης (Papadakis et al, 2020).
Τα δεδομένα που αξιοποιεί ένας οργανισμός για σκοπούς εσωτερικού ελέγχου και ανίχνευσης red flags προέρχονται από πολλαπλές πηγές. Τα βασικά λογιστικά και οικονομικά στοιχεία αντλούνται από τα λογιστικά πληροφοριακά συστήματα (ERP, λογιστικές εφαρμογές, υποσυστήματα τιμολόγησης, μισθοδοσίας κ.λπ.). Η ποσότητα και η λεπτομέρεια των δεδομένων που παράγονται από αυτά τα συστήματα έχουν αυξηθεί εντυπωσιακά τα τελευταία χρόνια, δημιουργώντας νέες δυνατότητες αλλά και προκλήσεις για τον εσωτερικό έλεγχο (Τσατσούλας, 2022).
Εκτός από τα λογιστικά δεδομένα, σημαντικό ρόλο παίζουν και τα λειτουργικά δεδομένα (π.χ. κινήσεις αποθήκης, παραγγελίες, κινήσεις πελατών), τα δεδομένα προμηθειών (συμβάσεις, προσφορές, εγκρίσεις), τα δεδομένα ανθρώπινου δυναμικού (άδειες, υπερωρίες, μετακινήσεις, μεταβολές αποδοχών), καθώς και δεδομένα από συστήματα πρόσβασης και e-mail. Η αξιοποίηση όλων αυτών των πηγών δημιουργεί μια πολυδιάστατη εικόνα της λειτουργίας του οργανισμού, στην οποία μπορούν να «κουμπώσουν» τεχνικές εξόρυξης δεδομένων και μηχανικής μάθησης για τον εντοπισμό ασυνήθιστων μοτίβων που συχνά συνδέονται με red flags (Λαφτσίδης, 2008· Papadakis et al, 2020).
Για την ανίχνευση απάτης μέσω εξόρυξης δεδομένων επισημαίνεται ότι η σωστή προετοιμασία των δεδομένων (data preprocessing) και η επιλογή των κατάλληλων μεταβλητών έχουν καθοριστική σημασία για την απόδοση των μοντέλων (Κούγιας, 2022). Σχετικές διπλωματικές εργασίες που εξετάζουν την ανίχνευση απάτης στο λιανικό εμπόριο ή σε οικονομικές οντότητες, δείχνουν ότι η επιτυχία αλγορίθμων ταξινόμησης

και ανίχνευσης ανωμαλιών εξαρτάται σε μεγάλο βαθμό από το πώς έχουν επιλεγεί και καθοριστεί τα δεδομένα, καθώς και από το αν οι μεταβλητές που χρησιμοποιούνται αντανακλούν πραγματικούς επιχειρησιακούς κινδύνους και όχι απλώς τεχνικές ή στατιστικές ιδιαιτερότητες.

Το ζήτημα της ποιότητας των δεδομένων αναδεικνύεται ως κεντρικός παράγοντας επιτυχίας ή αποτυχίας τέτοιων συστημάτων. Για τη σχέση μεταξύ ποιότητας δεδομένων και χρηματοοικονομικής πληροφόρησης επισημαίνεται ότι ανακριβή, ελλιπή ή μη συνεπή δεδομένα, αποτελούν σημαντικό εμπόδιο στη λήψη αξιόπιστων αποφάσεων, τόσο από τη διοίκηση όσο και από τους ελεγκτές (Τσατσούλας, 2022). Η ποιότητα των δεδομένων αξιολογείται συνήθως με βάση κριτήρια όπως η ακρίβεια, η πληρότητα, η συνέπεια, η εγκυρότητα και η καταλληλόλητα για τον συγκεκριμένο σκοπό. Αν, για παράδειγμα, τα δεδομένα των προμηθειών είναι ελλιπή ως προς τα στοιχεία τιμών, όρων πληρωμής ή ταυτοποίησης προμηθευτών, τότε ένα σύστημα ΤΝ θα δυσκολευτεί να εντοπίσει red flags, όπως ασυνήθιστες αυξήσεις τιμών ή σύγκρουση συμφερόντων.

Όπως αποτυπώνεται σε πρόσφατες μεταπτυχιακές εργασίες για τον εσωτερικό έλεγχο και τις νέες τεχνολογίες, σε πολλούς οργανισμούς (ιδίως μικρού και μεσαίου μεγέθους) η χρήση ΤΝ συναντά εμπόδια ακριβώς επειδή τα δεδομένα είναι διάσπαρτα σε διαφορετικά συστήματα χωρίς ενιαία πρότυπα και χωρίς επαρκή τεκμηρίωση (Γκεβρέκη, 2024· Τσουνιάς, 2023· Λαζαρίδου, 2024).

Ένα ακόμη κρίσιμο ζήτημα αφορά την τεκμηρίωση της προέλευσης και του μετασχηματισμού των δεδομένων. Τα δεδομένα περνούν συχνά από πολλαπλά στάδια επεξεργασίας. Αν αυτά δεν τεκμηριώνονται επαρκώς, τότε ο εσωτερικός ελεγκτής δεν μπορεί να γνωρίζει με βεβαιότητα πώς προέκυψε μια συγκεκριμένη ένδειξη, ποια ανεπεξέργαστα δεδομένα χρησιμοποιήθηκαν και ποιοι μετασχηματισμοί εφαρμόστηκαν. Η έλλειψη αυτή μπορεί να υπονομεύσει τόσο τη δυνατότητα ελέγχου του μοντέλου όσο και τη συμμόρφωση με τα κανονιστικά πλαίσια, όπως ο GDPR και το AI Act (Κουτσουνάκη, 2020).

Η σχέση ποιότητας δεδομένων και red flags κινείται σε δύο επίπεδα: από τη μία, κακή ποιότητα δεδομένων μπορεί να οδηγήσει σε ψευδείς ενδείξεις (false positives), καθώς το σύστημα ΤΝ εντοπίζει ανωμαλίες που είναι στην πραγματικότητα λάθη καταχώρισης, καθυστερημένες ενημερώσεις ή ασυμβατότητες μεταξύ συστημάτων, από την άλλη, μπορεί να οδηγήσει και σε false negatives, καθώς πραγματικά ύποπτες συμπεριφορές «χάνονται» μέσα σε θόρυβο ή παραμορφώνονται από μη έγκυρα δεδομένα. Έρευνες για τη χρήση data analytics στην ανίχνευση απάτης υπογραμμίζουν ότι η αξιοπιστία των συμπερασμάτων εξαρτάται σε κρίσιμο βαθμό από την ποιότητα των δεδομένων εισόδου αλλά και από τη συνεχή παρακολούθηση και βελτίωση αυτής της ποιότητας (Sidauruk, 2024).

Σε πρακτικό επίπεδο, η διασφάλιση της ποιότητας των δεδομένων περιλαμβάνει μια σειρά από ενέργειες όπως είναι ο ορισμός υπεύθυνων για κάθε σύνολο δεδομένων, η καθιέρωση διαδικασιών ελέγχου κατά την εισαγωγή και ενημέρωση των δεδομένων, η τακτική εκτέλεση ελέγχων συνέπειας και πληρότητας, καθώς και η υλοποίηση διαδικασιών για τη διόρθωση τυχόν σφαλμάτων. Επιπλέον, απαιτείται συνεργασία μεταξύ των τμημάτων της πληροφορικής, των οικονομικών υπηρεσιών, των προμηθειών, του HR και του εσωτερικού ελέγχου, ώστε να συμφωνηθούν κοινά πρότυπα. Οι οργανισμοί που επενδύουν συστηματικά στη βελτίωση της ποιότητας των δεδομένων

τους δημιουργούν παράλληλα το κατάλληλο υπόβαθρο για τη χρήση πιο σύνθετων και αξιόπιστων μοντέλων ΤΝ (Supervizor blog, 2025· Solix, 2024).
Τέλος, πρέπει να αναφερθεί ότι η ποιότητα των δεδομένων συνδέεται και με τη δεοντολογική διάσταση της χρήσης της ΤΝ. Όταν τα δεδομένα είναι ελλιπή ή μεροληπτικά, τα μοντέλα ΤΝ μπορεί να «μάθουν» στρεβλές σχέσεις, οδηγώντας σε άδικη στοχοποίηση συγκεκριμένων εργαζομένων, τμημάτων ή προμηθευτών. Αυτό δημιουργεί όχι μόνο επιχειρησιακό και ελεγκτικό κίνδυνο αλλά και νομικές και ηθικές συνέπειες, ιδιαίτερα όταν τα αποτελέσματα της ΤΝ χρησιμοποιούνται για αποφάσεις που επηρεάζουν άμεσα ανθρώπους. Επομένως, η διασφάλιση ότι τα δεδομένα είναι αντιπροσωπευτικά, ορθά και απαλλαγμένα από συστηματικές προκαταλήψεις αποτελεί βασική προϋπόθεση για τη δίκαιη και υπεύθυνη χρήση της ΤΝ στον εσωτερικό έλεγχο (Τζίφας, 2025).

4.3 Επιλογή μοντέλων Τεχνητής Νοημοσύνης στην ανίχνευση Red Flags

Η επιλογή των κατάλληλων μοντέλων Τεχνητής Νοημοσύνης (ΤΝ) αποτελεί καθοριστικό παράγοντα για την αποτελεσματική ανίχνευση red flags. Στο πλαίσιο του εσωτερικού ελέγχου, η ΤΝ δεν χρησιμοποιείται απλώς για την αυτοματοποίηση διαδικασιών αλλά για την αναγνώριση πρότυπων και αποκλίσεων που συχνά υποδηλώνουν αυξημένο κίνδυνο απάτης ή μη συμμόρφωσης. Για τον λόγο αυτό, η επιλογή μοντέλου δεν μπορεί να βασίζεται μόνο στη στατιστική του απόδοση, αλλά πρέπει να λαμβάνει υπόψη τη φύση των δεδομένων, το είδος των red flags, τις απαιτήσεις επεξηγησιμότητας, το επιχειρησιακό πλαίσιο και τα όρια της ανθρώπινης εποπτείας (Brown-Liburd & Vasarhelyi, 2015).
Τα πιο συνηθισμένα μοντέλα που χρησιμοποιούνται στην ανίχνευση απάτης ανήκουν στις κατηγορίες της επιβλεπόμενης μάθησης (supervised learning) και της μη επιβλεπόμενης μάθησης (unsupervised learning). Στην επιβλεπόμενη μάθηση περιλαμβάνονται αλγόριθμοι όπως τα Decision Trees (δέντρα αποφάσεων), Gradient Boosting Machines (GBM), Support Vector Machines και τα νευρωνικά δίκτυα. Οι μέθοδοι αυτές απαιτούν ιστορικά δεδομένα στα οποία υπάρχουν «ετικέτες» που δείχνουν ποιες περιπτώσεις αποτέλεσαν απάτη και ποιες όχι. Έρευνες δείχνουν ότι τα δέντρα αποφάσεων εμφανίζουν υψηλή απόδοση στην επεξεργασία δεδομένων απάτης, επειδή μπορούν να διαχειριστούν πολύπλοκες σχέσεις μεταξύ αυτών των δεδομένων και να ενσωματώσουν πολλές μεταβλητές (Ngai et al., 2011· West & Bhattacharya, 2016).
Στον εσωτερικό έλεγχο, ωστόσο, η επιβλεπόμενη μάθηση παρουσιάζει περιορισμούς. Η απάτη εξελίσσεται διαρκώς, με αποτέλεσμα τα ιστορικά δεδομένα να μην απεικονίζουν πάντα τα νέα μοτίβα. Αυτό οδηγεί στην αυξημένη χρήση μεθόδων μη επιβλεπόμενης μάθησης. Τα μοντέλα αυτά δεν απαιτούν ετικέτες και μπορούν να εντοπίσουν παραβάσεις ή απότομες μεταβολές στα δεδομένα. Σε περιβάλλοντα όπου τα red flags δεν έχουν ξεκάθαρα χαρακτηριστικά ή όπου η συμπεριφορά που θεωρείται «κανονική» μεταβάλλεται συχνά, οι τεχνικές αυτές είναι καταλληλότερες (Goldstein & Uchida, 2016).
Μια τρίτη κατηγορία αφορά τα υβριδικά μοντέλα που συνδυάζουν supervised και unsupervised μάθηση. Τα συστήματα αυτά μπορούν να εντοπίζουν νέες μορφές απάτης ενώ ταυτόχρονα αξιοποιούν την ιστορική γνώση των ελεγκτών. Έτσι, εμφανίζονται ολοένα και περισσότερα παραδείγματα υβριδικών μεθόδων που εφαρμόζονται σε δεδομένα τραπεζών, ασφαλιστικών εταιρειών και λογιστικών συστημάτων,

επιτυγχάνοντας υψηλότερη ακρίβεια σε σχέση με τις παραδοσιακές μεθόδους (Phua et al., 2010).
Ανεξαρτήτως κατηγορίας μοντέλου, η επιλογή πρέπει να βασίζεται όχι μόνο στην ακρίβεια αλλά και στις ανάγκες επεξηγησιμότητας και διαφάνειας. Στον εσωτερικό έλεγχο, η επεξηγησιμότητα είναι απαραίτητη, καθώς οι ελεγκτές χρειάζεται να αιτιολογήσουν τα ευρήματά τους στη διοίκηση και σε εποπτικές αρχές. Για τον λόγο αυτό, μοντέλα «μαύρου κουτιού» όπως τα deep neural networks παρουσιάζουν δυσκολίες: μπορεί να επιτυγχάνουν υψηλή ακρίβεια αλλά αδυνατούν να παρέχουν κατανοητές εξηγήσεις για το πώς έφτασαν σε ένα συμπέρασμα. Τα explainable AI frameworks (όπως LIME και SHAP) χρησιμοποιούνται συχνά για να κάνουν πιο κατανοητές τις αποφάσεις τέτοιων μοντέλων (Lundberg & Lee, 2017).
Επιπλέον, η επιλογή μοντέλου πρέπει να λαμβάνει υπόψη το ενδεχόμενο παραγωγής ψευδών red flags. Μοντέλα που είναι ιδιαίτερα ευαίσθητα μπορούν να εντοπίζουν υπερβολικά πολλές «ύποπτες» περιπτώσεις (false positives), κάτι που επιβαρύνει τον εσωτερικό έλεγχο και οδηγεί σε μειωμένη εμπιστοσύνη των χρηστών στο σύστημα. Αντίθετα, μοντέλα υπερβολικά αυστηρά μπορεί να παραβλέψουν πραγματικές ενδείξεις απάτης (false negatives), γεγονός που αυξάνει σημαντικά τον λειτουργικό κίνδυνο. Η ισορροπία αυτή αποτελεί κρίσιμο σημείο στην επιλογή μοντέλου (West & Bhattacharya, 2016).
Στην πράξη, ο συνδυασμός πολλών μοντέλων μπορεί να προσφέρει καλύτερη συνολική απόδοση και σταθερότητα. Επιπλέον, οι οργανισμοί τείνουν να ενσωματώνουν «κανόνες» βασισμένους στην ελεγκτική εμπειρία δημιουργώντας έτσι ένα υβριδικό μοντέλο κανόνων–αλγορίθμων που ενισχύει την ποιότητα των αποτελεσμάτων και μειώνει τον κίνδυνο ψευδών ενδείξεων. Στο πλαίσιο της εταιρικής απάτης, τέτοια συστήματα εμφανίζουν αξιοσημείωτη αποτελεσματικότητα λόγω της δυνατότητάς τους να συνδυάζουν μαθηματικά πρότυπα και ανθρώπινη κρίση.

4.4 Εσωτερική ή διαδικτυακή λειτουργία συστημάτων ΤΝ για την ανίχνευση red flags: πλεονεκτήματα, μειονεκτήματα και επιλογή

Η απόφαση σχετικά με το πού και πώς θα λειτουργεί ένα σύστημα Τεχνητής Νοημοσύνης για την ανίχνευση red flags -δηλαδή εάν θα βρίσκεται εντός του εσωτερικού δικτύου (on-premise) ή στο διαδίκτυο (cloud-based), αποτελεί κρίσιμο παράγοντα επιτυχίας.
Η επιλογή αυτή επηρεάζει την ασφάλεια των δεδομένων, την απόδοση, την προσβασιμότητα, το κόστος συντήρησης και τον τρόπο εκπαίδευσης των μοντέλων ΤΝ. Τα συστήματα που εντοπίζουν red flags βασίζονται σε συνεχή ροή δεδομένων από ποικίλες πηγές (ERP, λογιστικά συστήματα, HR, email, CRM κ.λπ.), επομένως η υποδομή φιλοξενίας επηρεάζει άμεσα την αξιοπιστία και τη λειτουργικότητα της ΤΝ (ENISA, 2015).

4.4.1 Εσωτερική λειτουργία (On-Premise Systems)

Η εσωτερική εγκατάσταση (on-premise) αναφέρεται σε συστήματα ΤΝ που λειτουργούν εντός του εσωτερικού δικτύου του οργανισμού, σε ιδιόκτητους διακομιστές ή ελεγχόμενα data centers. Τα δεδομένα συλλέγονται, αποθηκεύονται, δέχονται επεξεργασία και αναλύονται χωρίς να μεταφέρονται εκτός του οργανισμού.

Πλεονεκτήματα

1. **Αυξημένη ασφάλεια και έλεγχος δεδομένων:** Τα δεδομένα παραμένουν εντός του οργανισμού, μειώνοντας τον κίνδυνο διαρροής ή μη εξουσιοδοτημένης πρόσβασης (ENISA, 2015). Αυτό είναι ιδιαίτερα κρίσιμο σε οργανισμούς που διαχειρίζονται ευαίσθητα οικονομικά ή προσωπικά δεδομένα (π.χ. τράπεζες, δημόσιοι φορείς).
2. **Συμμόρφωση με κανονιστικά πλαίσια:** Η τοπική αποθήκευση διευκολύνει τη συμμόρφωση με το GDPR και άλλες εθνικές νομοθεσίες περί προστασίας δεδομένων (European Commission, 2021).
3. **Προσαρμοστικότητα και αυτονομία:** Οι οργανισμοί έχουν πλήρη έλεγχο στην παραμετροποίηση, στη διαχείριση και στην ενημέρωση των συστημάτων, χωρίς εξάρτηση από εξωτερικούς παρόχους (Γενίτσαρης, 2023).
4. **Μειωμένη εξάρτηση από το διαδίκτυο:** Τα συστήματα συνεχίζουν να λειτουργούν ακόμα και αν διακοπεί η σύνδεση στο Internet, εξασφαλίζοντας έτσι σταθερότητα.

Μειονεκτήματα

1. **Υψηλό αρχικό κόστος εγκατάστασης:** Η δημιουργία υποδομής (servers, αποθηκευτικοί χώροι, ασφάλεια, τεχνική ομάδα) απαιτεί σημαντικές επενδύσεις.
2. **Περιορισμένη υπολογιστική ισχύς:** Οι τοπικοί διακομιστές έχουν φυσικά όρια σε επεξεργαστική ισχύ και αποθηκευτικό χώρο. Αυτό μπορεί να περιορίσει τη χρήση σύνθετων μοντέλων ΤΝ όπως είναι τα Deep Learning.
3. **Ανάγκη εξειδικευμένου προσωπικού:** Η συντήρηση, ενημέρωση και εκπαίδευση του συστήματος απαιτεί ομάδα data scientists, IT administrators και ειδικούς ασφάλειας.
4. **Περιορισμένη δυνατότητα εκπαίδευσης σε εξωτερικά δεδομένα:** Η εκπαίδευση μοντέλων ΤΝ σε νέα δεδομένα (π.χ. διεθνή πρότυπα απάτης) είναι δύσκολη, αφού τα δεδομένα παραμένουν εσωτερικά.

Εκπαίδευση, αναβάθμιση και συντήρηση

Η εκπαίδευση των μοντέλων γίνεται τοπικά (offline training), με χρήση των διαθέσιμων εταιρικών δεδομένων. Οι αναβαθμίσεις του συστήματος απαιτούν χειροκίνητες ενημερώσεις από το τμήμα πληροφορικής.

Η συντήρηση περιλαμβάνει:

- ενημέρωση των μοντέλων,
- έλεγχο ακεραιότητας των δεδομένων,
- ανανέωση των κανόνων που αφορούν τα red flags.

Αυτό το μοντέλο λειτουργίας είναι ιδανικό όταν προέχει η ασφάλεια και ο έλεγχος, παρά η ταχύτητα ή η εξωτερική διασύνδεση.

4.4.2 Διαδικτυακή λειτουργία (Cloud-Based Systems)

Στην cloud-based λειτουργία, το σύστημα φιλοξενείται σε υποδομές τρίτων παρόχων (π.χ. AWS, Azure, Google Cloud). Η επεξεργασία των δεδομένων πραγματοποιείται σε απομακρυσμένους διακομιστές με πρόσβαση μέσω του διαδικτύου.

Πλεονεκτήματα

1. **Υπολογιστική ισχύς:** Το cloud παρέχει πρακτικά απεριόριστους πόρους για την επεξεργασία μεγάλου όγκου δεδομένων και την εκπαίδευση πολύπλοκων μοντέλων.
2. **Ευκολία αναβάθμισης:** Οι πάροχοι cloud ενημερώνουν αυτόματα το λογισμικό, επιτρέποντας πρόσβαση σε σύγχρονες εκδόσεις εργαλείων ΤΝ.
3. **Συνεργασία και προσβασιμότητα:** Τα δεδομένα είναι διαθέσιμα σε πραγματικό χρόνο σε ελεγκτές, αναλυτές και διοικητικά στελέχη, ακόμη και από διαφορετικές τοποθεσίες.
4. **Συνεχής εκπαίδευση μοντέλων:** Το cloud υποστηρίζει τη συνεχή εκπαίδευση των μοντέλων καθώς προστίθενται διαρκώς νέα δεδομένα.

Μειονεκτήματα

1. **Αυξημένοι κίνδυνοι ασφάλειας και ιδιωτικότητας:** Η αποστολή δεδομένων στο διαδίκτυο αυξάνει τον κίνδυνο υποκλοπής ή μη εξουσιοδοτημένης πρόσβασης, παρά τις σύγχρονες τεχνικές κρυπτογράφησης (ENISA, 2015).
2. **Νομικοί περιορισμοί:** Σε ορισμένους τομείς (π.χ. δημόσιοι οργανισμοί, υγεία, χρηματοπιστωτικά ιδρύματα), η αποθήκευση δεδομένων εκτός εθνικού εδάφους απαγορεύεται (European Commission, 2021).
3. **Εξάρτηση από τον πάροχο:** Η αξιοπιστία, η διαθεσιμότητα και η απόδοση εξαρτώνται από τις υπηρεσίες του παρόχου cloud.
4. **Κόστος λειτουργίας σε βάθος χρόνου:** Αν και η αρχική εγκατάσταση είναι φθηνότερη, η συνεχής χρήση cloud μπορεί να οδηγήσει σε υψηλό λειτουργικό κόστος λόγω αποθήκευσης και λοιπών χρεώσεων.

Εκπαίδευση, αναβάθμιση και συντήρηση

Στην περίπτωση του cloud, η εκπαίδευση των μοντέλων μπορεί να γίνεται συνδυαστικά, δηλαδή με δεδομένα του οργανισμού και δεδομένα που θα είναι διαθέσιμα διαδικτυακά.
Οι αναβαθμίσεις είναι αυτόματες και η συντήρηση ανατίθεται στον πάροχο. Επιπλέον, τα μοντέλα μπορούν να μαθαίνουν από άλλα δίκτυα χωρίς να ανταλλάσσουν πραγματικά δεδομένα, διασφαλίζοντας τη συμμόρφωση με την ιδιωτικότητα (Kairouz, et al. 2021).

Πότε συμφέρει η κάθε επιλογή

- **Εσωτερικό δίκτυο (On-Premise):** Όταν ο οργανισμός χειρίζεται ευαίσθητα δεδομένα, όπως λογιστικές εγγραφές, προσωπικά στοιχεία ή εμπιστευτικές συναλλαγές. Ιδανικό για:
    - Δημόσιους οργανισμούς,
    - Τράπεζες,

- Νοσοκομεία,
- Επιχειρήσεις με αυστηρά κανονιστικά πλαίσια (π.χ. GDPR, SOX).
-

- **Cloud περιβάλλον:** Προτιμάται όταν ο οργανισμός χρειάζεται μεγάλη υπολογιστική ισχύ, ευελιξία και συνεχή ενημέρωση. Ιδανικό για:
  - Ιδιωτικές εταιρείες,
  - Πολυεθνικούς οργανισμούς,
  - Startups που εφαρμόζουν data-driven auditing,
  - Εταιρίες με υπάλληλους σε διαφορετικές χώρες.

4.4.3 Υβριδική προσέγγιση

Η βέλτιστη πρακτική για πολλούς οργανισμούς είναι ο συνδυασμός και των δύο προσεγγίσεων.

Σε ένα υβριδικό μοντέλο, τα ευαίσθητα δεδομένα παραμένουν εντός του οργανισμού, ενώ η ανάλυση μεγάλου όγκου ή η εκπαίδευση μοντέλων πραγματοποιείται σε ασφαλές cloud περιβάλλον (Kairouz, et al. 2021).

Έτσι τα μοντέλα μαθαίνουν συλλογικά χωρίς να χρειάζεται η ανταλλαγή δεδομένων, ενώ η ανάλυσή τους γίνεται τοπικά, αλλά τα αποτελέσματά της συγχρονίζονται σε όλα τα ενδιαφερόμενα μέρη, από όπου και αν έχουν πρόσβαση.

Η λύση αυτή προσφέρει ισορροπία ανάμεσα στην ιδιωτικότητα, την επεξεργαστική ισχύ και την αποδοτικότητα, καθιστώντας την πλέον κατάλληλη για την ανίχνευση red flags σε σύγχρονα περιβάλλοντα (Γενίτσαρης, 2023).

4.5 Εκπαίδευση και αξιολόγησης Μοντέλων ΤΝ

Η ανάπτυξη ενός αξιόπιστου συστήματος Τεχνητής Νοημοσύνης για την ανίχνευση red flags απαιτεί μια συστηματική και ολοκληρωμένη διαδικασία εκπαίδευσης και αξιολόγησης των μοντέλων. Το πλαίσιο αυτό δεν περιορίζεται στο τεχνικό μέρος της εκπαίδευσης αλλά καλύπτει τον πλήρη κύκλο ζωής του μοντέλου, από τη συλλογή και προετοιμασία των δεδομένων μέχρι τη συνεχή παρακολούθηση της λειτουργίας και την επαναξιολόγηση της απόδοσης. Η ποιότητα της διαδικασίας εκπαίδευσης επηρεάζει καθοριστικά την ικανότητα του συστήματος να εντοπίζει ακριβώς τους κινδύνους που ενδιαφέρουν τον εσωτερικό έλεγχο (Ngai et al., 2011).

Κατά την εκπαίδευση, το μοντέλο μαθαίνει να αναγνωρίζει τα μοτίβα που ξεχωρίζουν τις φυσιολογικές συναλλαγές από τις δυνητικά ύποπτες. Αυτό είναι ιδιαίτερα δύσκολο επειδή τα δεδομένα πρέπει να είναι σαφώς διαχωρισμένα, ώστε το σύστημα να μπορέσει να εκπαιδευτεί σωστά. Κάτι που πολλές φορές είναι δύσκολο να διαχωρίσει και ο άνθρωπος εάν έχει ελλιπείς πληροφορίες (Elkan, 2001).

Το επόμενο στάδιο είναι μια διαδικασία επικύρωσης η οποία βοηθά στον εντοπισμό υπερπροσαρμογής (overfitting), δηλαδή της τάσης του μοντέλου να μαθαίνει υπερβολικά καλά τα δεδομένα εκπαίδευσης αλλά να αποτυγχάνει στα πραγματικά δεδομένα. Η διαδικασία αυτή συμβάλλει στην εξισορρόπηση της απόδοσης και στη δημιουργία μοντέλων που γενικεύουν επαρκώς τα νέα δεδομένα. Τα μοντέλα που εκπαιδεύονται χωρίς διαδικασίες επικύρωσης εμφανίζουν συχνά δραματική μείωση της αποτελεσματικότητάς τους όταν εφαρμοστούν σε πραγματικές συναλλαγές (West & Bhattacharya, 2016).

Υστέρα έχουμε την αξιολόγηση των μοντέλων ΤΝ στην ανίχνευση red flags, η οποία απαιτεί τη χρήση κατάλληλων δεικτών απόδοσης (He & Garcia, 2009).

Οι σημαντικότεροι από αυτούς τους δείκτες είναι:

- **Precision**: το ποσοστό των εντοπισμένων red flags που είναι πραγματικά ύποπτα.
- **Recall (ευαισθησία)**: το ποσοστό των πραγματικών red flags που εντοπίστηκαν.
- **ROC–AUC και PR–AUC**: καμπύλες που δείχνουν την ικανότητα του μοντέλου να διαχωρίζει ύποπτες από μη ύποπτες συναλλαγές, ειδικά σε περιβάλλον σπάνιων γεγονότων.

Ανάγκη συνεχούς επανεκπαίδευσης

Στα συστήματα ανίχνευσης απάτης, οι τακτικές των δραστών εξελίσσονται διαρκώς. Αυτό δημιουργεί το φαινόμενο του concept drift**,** δηλαδή της σταδιακής αλλαγής των μοτίβων στα δεδομένα. Τα μοντέλα που δεν επανεκπαιδεύονται ανά τακτά χρονικά διαστήματα χάνουν σημαντικό μέρος της απόδοσής τους. Για αυτό χρειάζεται:

- συνεχής παρακολούθηση των δεικτών απόδοσης,
- σύγκριση νέων συναλλαγών με τα αρχικά πρότυπα,
- μηχανισμοί online learning ή retraining,
- τακτικές ενημερώσεις του συστήματος.

Ένα μοντέλο που είχε εκπαιδευτεί σε παλαιά δεδομένα μπορεί να αδυνατεί να εντοπίσει νέους τρόπους παράκαμψης των διαδικασιών.

Οι ελεγκτές πρέπει να συμμετέχουν ενεργά στη διαδικασία αξιολόγησης των αλγορίθμων, παρέχοντας feedback σε false positives και false negatives, ώστε το σύστημα να βελτιώνεται συνεχώς (Brown-Liburd & Vasarhelyi, 2015). Η συνεργασία αυτή δημιουργεί έναν κύκλο ανατροφοδότησης που αυξάνει την αποτελεσματικότητα του συστήματος (human–AI feedback loop).

Η απόδοση ενός μοντέλου ΤΝ πρέπει να παρακολουθείται σε πραγματικό χρόνο ή σε τακτική βάση. Όπως δείχνουν οι πρόσφατες μελέτες, η μείωση της απόδοσης μπορεί να μην εμφανιστεί ξαφνικά αλλά σταδιακά, καθιστώντας την παρακολούθηση κρίσιμη λειτουργία (Goldstein & Uchida, 2016). Για τον λόγο αυτό, η παρακολούθηση πρέπει να συνδέεται με τις ελεγκτικές διαδικασίες, ώστε τυχόν αλλαγές στην απόδοση να εντοπίζονται και να επιλύονται άμεσα από τους υπεύθυνους.

4.6 Δείκτες αξιολόγησης μοντέλων ΤΝ για ανίχνευση εταιρικών red flags

Η αξιολόγηση των μοντέλων Τεχνητής Νοημοσύνης που χρησιμοποιούνται για την ανίχνευση red flags είναι ένα κρίσιμο στάδιο, καθώς από αυτήν εξαρτάται το κατά πόσο τα παραγόμενα αποτελέσματα μπορούν να θεωρηθούν αξιόπιστα και χρήσιμα για τον εσωτερικό έλεγχο. Σε αντίθεση με άλλες εφαρμογές της ΤΝ, όπου μπορεί να γίνει ανεκτό ένα γενικό ποσοστό σφάλματος, στην ανίχνευση απάτης και παρατυπιών τα λάθη έχουν άμεσο οικονομικό, νομικό και ηθικό κόστος. Για τον λόγο αυτό, κάποιοι κλασικοί δείκτες, όπως είναι η συνολική ακρίβεια (accuracy), δεν επαρκούν, εδώ απαιτείται ένα πιο λεπτομερές πλαίσιο δεικτών που λαμβάνει υπόψη την ανισορροπία των δεδομένων,

τη βαρύτητα των σφαλμάτων, τη σταθερότητα της απόδοσης και ταυτόχρονα, τον βαθμό ερμηνευσιμότητας των μοντέλων (Ngai et al., 2011· West & Bhattacharya, 2016).
Στην καρδιά της αξιολόγησης βρίσκονται τέσσερις βασικές κατηγορίες αποτελεσμάτων: τα true positives (περιπτώσεις όπου το μοντέλο σωστά εντόπισε ένα red flag), τα true negatives (σωστά μη-ύποπτες περιπτώσεις), τα false positives (λανθασμένα ύποπτες περιπτώσεις) και τα false negatives (πραγματικές ύποπτες περιπτώσεις που το μοντέλο δεν εντόπισε). Στο πλαίσιο του εσωτερικού ελέγχου, κάθε μία από αυτές τις κατηγορίες έχει συγκεκριμένη επιχειρησιακή σημασία: τα false positives αυξάνουν τον φόρτο εργασίας, προκαλούν κόπωση και μπορεί να υπονομεύσουν την εμπιστοσύνη στο σύστημα, τα false negatives, αντίθετα, οδηγούν σε μη ανίχνευση πραγματικών περιπτώσεων απάτης ή σοβαρών παρατυπιών, με ενδεχομένως πολύ μεγαλύτερο κόστος.
Για αυτόν τον λόγο, χρησιμοποιούνται συστηματικά οι δείκτες precision και recall. Ο precision (θετική προβλεπτική αξία) εκφράζει το ποσοστό των περιπτώσεων που χαρακτηρίστηκαν ως red flags και αποδείχθηκαν πράγματι ύποπτες. Σε ελεγκτικούς όρους, υψηλός precision σημαίνει ότι ο ελεγκτής έχει μεγάλη πιθανότητα να ασχολείται με κάτι ουσιαστικά ύποπτο, άρα μειώνεται ο κίνδυνος να σπαταλά χρόνο σε αβάσιμα ευρήματα. Ο recall εκφράζει το ποσοστό των πραγματικών red flags που εντοπίστηκαν από το μοντέλο. Υψηλός recall σημαίνει ότι το σύστημα «βλέπει» μεγάλο μέρος των πραγματικών περιστατικών απάτης ή απόπειρας παράκαμψης των δικλίδων ασφαλείας. Στην πράξη, ο recall συνδέεται περισσότερο με την προστασία του οργανισμού από μη ανιχνευμένες απάτες, ενώ ο precision συνδέεται με τη λειτουργική αποδοτικότητα του ελεγκτικού μηχανισμού (He & Garcia, 2009).
Ένας ακόμη δείκτης που ενδιαφέρει ιδιαίτερα τον εσωτερικό έλεγχο αλλά δεν είναι τόσο «ποσοτικός», είναι η ερμηνευσιμότητα (interpretability). Η ερμηνευσιμότητα δεν είναι απλώς ιδιότητα του μοντέλου αλλά και κριτήριο αξιολόγησης: πόσο εύκολα μπορεί ο ελεγκτής να καταλάβει γιατί ένα συγκεκριμένο περιστατικό χαρακτηρίστηκε ως red flag (Lipton, 2016· Doshi-Velez & Kim, 2017). Μοντέλα όπως τα απλά δέντρα αποφάσεων έχουν το πλεονέκτημα ότι μπορούν να παρουσιαστούν και να εξηγηθούν εύκολα στη διοίκηση ή στους ρυθμιστικούς φορείς, ενώ πολύπλοκα νευρωνικά δίκτυα, παρότι ενδέχεται να έχουν καλύτερη απόδοση σε κάποιες περιπτώσεις, είναι δύσκολο να γίνουν πλήρως κατανοητά.
Στο πλαίσιο της ανίχνευσης red flags, η ερμηνευσιμότητα λειτουργεί και ως δείκτης εμπιστοσύνης. Αν οι ελεγκτές κατανοούν γιατί το σύστημα χαρακτηρίζει μια κίνηση ως ύποπτη, είναι πιο πιθανό να εμπιστευτούν το εργαλείο και να το ενσωματώσουν στη δουλειά τους. Αντίθετα, όταν το σύστημα εμφανίζει «μαύρα κουτιά» χωρίς επαρκείς εξηγήσεις, υπάρχει κίνδυνος είτε να αγνοούνται οι ειδοποιήσεις είτε να γίνεται τυφλή αποδοχή τους χωρίς κριτική σκέψη, αμφότερα εξίσου επικίνδυνα από ελεγκτική άποψη (Brown-Liburd & Vasarhelyi, 2015).

4.7 Ζητήματα δεοντολογίας, GDPR και Governance στα συστήματα ΤΝ για ανίχνευση εταιρικών Red Flags

Η χρήση της Τεχνητής Νοημοσύνης στην ανίχνευση εταιρικών red flags δημιουργεί σημαντικές τεχνικές και ελεγκτικές δυνατότητες αλλά ταυτόχρονα εγείρει κρίσιμα ζητήματα δεοντολογίας, διαφάνειας και συμμόρφωσης με το νομοθετικό πλαίσιο. Τα συστήματα αυτά επεξεργάζονται δεδομένα εργαζομένων, συναλλασσόμενων και διαδικασιών του οργανισμού, συχνά σε πραγματικό χρόνο. Για τον λόγο αυτό,

θεωρούνται συστήματα υψηλού κινδύνου, κατά την έννοια του υπό διαμόρφωση EU AI Act**,** και υπόκεινται σε αυστηρές απαιτήσεις διακυβέρνησης, ελέγχου και διαφάνειας (Wachter & Mittelstadt, 2019).

4.7.1 Προκαταλήψεις (bias), διαφάνεια και ανθρώπινη εποπτεία

Η δεοντολογία αποτελεί τον πρώτο πυλώνα αξιολόγησης. Τα μοντέλα ΤΝ ενδέχεται να παραγάγουν προκαταλήψεις**,** δηλαδή συστηματικές αποκλίσεις εις βάρος συγκεκριμένων ομάδων ή προτύπων συμπεριφοράς. Το bias μπορεί να προκύψει από τα ίδια τα δεδομένα (historical bias), από τον τρόπο εκπαίδευσης (algorithmic bias) ή από ερμηνευτικά λάθη των χρηστών (interpretation bias). Στον εσωτερικό έλεγχο, το bias μπορεί να οδηγήσει σε λάθος ταυτοποίηση ύποπτων συναλλαγών, άδικη στοχοποίηση συγκεκριμένων εργαζομένων, τμημάτων ή χωρών και τελικά, απώλεια αξιοπιστίας τόσο του συστήματος όσο και του ίδιου του ελεγκτικού μηχανισμού.

Η διαφάνεια αποτελεί δεύτερη κρίσιμη παράμετρο. Τα μοντέλα black-box, ειδικά τα deep learning συστήματα, είναι δύσκολο να εξηγηθούν. Η έλλειψη ερμηνευσιμότητας μπορεί να οδηγήσει σε έλλειψη εμπιστοσύνης των ελεγκτών και της διοίκησης.

Τρίτος δεοντολογικός άξονας είναι η ύπαρξη επαρκούς ανθρώπινης εποπτείας (human oversight). Τα ευρωπαϊκά πλαίσια δεοντολογίας (High-Level Expert Group on AI) αναφέρουν ως βασική προϋπόθεση ότι η ΤΝ πρέπει να λειτουργεί ως εργαλείο υποβοήθησης και όχι αντικατάστασης κρίσιμων εταιρικών αποφάσεων (Floridi et al., 2021). Στην πράξη, αυτό σημαίνει ότι τα συστήματα red-flag detection πρέπει να επιτρέπουν την ανθρώπινη επιβεβαίωση των αποτελεσμάτων, τη δυνατότητα αναθεώρησης και την καταγραφή των παρεμβάσεων των ελεγκτών. Σε αντίθετη περίπτωση, η απόλυτη αυτοματοποίηση οδηγεί σε ανεξέλεγκτο ρίσκο.

4.7.2 GDPR: Προσωπικά δεδομένα και αρχές επεξεργασίας

Η ανίχνευση red flags βασίζεται στην επεξεργασία μεγάλων όγκων δεδομένων που συχνά περιλαμβάνουν προσωπικά δεδομένα εργαζομένων, πελατών ή προμηθευτών. Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) επηρεάζει άμεσα τη χρήση της ΤΝ σε τέτοιες εφαρμογές. Ο κανονισμός επιβάλλει αρχές όπως:

- ελαχιστοποίηση των δεδομένων που χρησιμοποιούνται στα απολύτως απαραίτητα (data minimization),
- περιορισμό του σκοπού για τον οποίο θα χρησιμοποιηθούν (purpose limitation),
- ακρίβεια και επικαιροποίηση δεδομένων,
- ασφάλεια της επεξεργασίας (integrity & confidentiality),
- λογοδοσία (accountability), και
- δικαιώματα πρόσβασης και ένστασης του υποκειμένου του οποίου τα δεδομένα χρησιμοποιούνται.

Το άρθρο 22 του GDPR απαγορεύει αποφάσεις που βασίζονται αποκλειστικά σε αυτοματοποιημένη επεξεργασία, όταν έχουν σημαντικές επιπτώσεις στο υποκείμενο. Στην πράξη, αυτό σημαίνει ότι η τελική αξιολόγηση πρέπει να γίνεται από άνθρωπο, ενισχύοντας την ανάγκη για HITL (human-in-the-loop) προσέγγιση στα συστήματα red-flag detection.

4.7.3 AI Act και υποχρεώσεις για τα συστήματα υψηλού κινδύνου

Ο EU AI Act κατατάσσει τα συστήματα που σχετίζονται με την αξιολόγηση της αξιοπιστίας, της συμμόρφωσης ή της επαγγελματικής συμπεριφοράς ως συστήματα υψηλού κινδύνου (high-risk). Τα συστήματα ανίχνευσης απάτης εμπίπτουν σε αυτή την κατηγορία λόγω της επίδρασής τους σε λειτουργίες ελέγχου και διακυβέρνησης.

Οι βασικές υποχρεώσεις περιλαμβάνουν:

- ισχυρή προστασία των δεδομένων**,**
- τεκμηρίωση για τον τρόπο λειτουργίας**,**
- διασφάλιση της σωστής ερμηνείας**,**
- συνεχή παρακολούθηση για πιθανές αστοχίες**,**
- σχεδιασμό με γνώμονα την ασφάλεια και την ανθεκτικότητα.

Οι απαιτήσεις αυτές έχουν άμεση επιχειρησιακή σημασία: δεν μπορούν πλέον οι οργανισμοί να εφαρμόζουν μοντέλα ΤΝ στον έλεγχο χωρίς τεκμηρίωση και διαδικασίες ελέγχου ποιότητας, καθώς αυτό θα τους εκθέσει σε νομικό και κανονιστικό κίνδυνο (European Commission, 2021).

4.7.4 Σύνδεση με τον ελεγκτικό κίνδυνο

Ο δεοντολογικός και κανονιστικός σχεδιασμός συνδέεται άμεσα με τον ελεγκτικό κίνδυνο. Ειδικότερα:

- Το bias οδηγεί σε αναποτελεσματική κατανομή πόρων στον έλεγχο.
- Η χαμηλή διαφάνεια οδηγεί σε κίνδυνο ανακριβών συμπερασμάτων.
- Τα συστήματα χωρίς HITL αντίκεινται στο άρθρο 22 GDPR και αυξάνουν τον κίνδυνο μη αποδεκτών αποφάσεων.
- Η αδυναμία επίσημης τεκμηρίωσης, οδηγεί στον έλεγχο από τις ρυθμιστικές αρχές.

Με άλλα λόγια, η δεοντολογία και η συμμόρφωση δεν είναι δευτερεύοντα ζητήματα, αλλά θεμέλιο της αξιοπιστίας του συστήματος ΤΝ στον εσωτερικό έλεγχο.

4.8 Διασφάλιση της ερμηνείας των αποτελεσμάτων και της νομιμότητας στα συστήματα ΤΝ ανίχνευσης Red Flags

Η αποτελεσματική χρήση συστημάτων Τεχνητής Νοημοσύνης για την ανίχνευση εταιρικών red flags απαιτεί όχι μόνο υψηλή απόδοση αλλά και διασφάλιση ότι τα αποτελέσματα του μοντέλου είναι ερμηνεύσιμα**,** τεκμηριώσιμα και νομικά αποδεκτά. Σε περιβάλλοντα εσωτερικού ελέγχου, η ερμηνεία των αποφάσεων έχει καθοριστική σημασία: οι ελεγκτές και τα στελέχη πρέπει να μπορούν να κατανοήσουν γιατί ένα σύστημα "υποπτεύεται" μια συναλλαγή ή μια συμπεριφορά. Η έλλειψη ερμηνευσιμότητας υπονομεύει την εμπιστοσύνη, δημιουργεί κανονιστικούς κινδύνους και δυσχεραίνει τη σύνδεση των αποτελεσμάτων με τις ελεγκτικές διαδικασίες (Doshi-Velez & Kim, 2017).

4.8.1 Η ανάγκη για ερμηνευσιμότητα στα συστήματα red-flag detection

Η ερμηνευσιμότητα (interpretability) αναφέρεται στον βαθμό στον οποίο μπορεί ένας άνθρωπος να κατανοήσει την αιτιολόγηση πίσω από την πρόβλεψη ενός μοντέλου. Στα συστήματα red flags, τα αποτελέσματα επηρεάζουν τις αποφάσεις του εσωτερικού ελέγχου, την εμπλοκή των εργαζομένων στις διαδικασίες διερεύνησης, την αξιολόγηση των επιχειρησιακών κινδύνων και την επικοινωνία με τις ρυθμιστικές αρχές. Για αυτόν τον λόγο, δεν είναι αποδεκτό τα συστήματα να λειτουργούν ως "μαύρα κουτιά" (black box) χωρίς δυνατότητα παροχής εξηγήσεων (Lipton, 2016). Στην εταιρική απάτη, η κατανόηση των παραγόντων που οδηγούν σε χαρακτηρισμό μιας συναλλαγής ως ύποπτης είναι αναπόσπαστο μέρος του ελέγχου.

4.8.2 Τεχνικές Explainable AI (XAI) και η εφαρμογή τους στον εσωτερικό έλεγχο

Έχει αναπτυχθεί πλήθος μεθόδων ερμηνευσιμότητας, γνωστών ως Explainable AI (XAI). Οι πιο διαδεδομένες τεχνικές είναι:

- **LIME (Local Interpretable Model-Agnostic Explanations)**, που παρέχει τοπικές εξηγήσεις για μεμονωμένες προβλέψεις
- **SHAP (SHapley Additive exPlanations)**, το οποίο βασίζεται στη θεωρία παιγνίων για να ποσοτικοποιήσει τη συνεισφορά κάθε χαρακτηριστικού στη λήψη απόφασης (Lundberg & Lee, 2017).
- **Partial Dependence Plots (PDPs)** και **Feature Importance Scores**, που εξηγούν τις συνολικές τάσεις ενός μοντέλου.

Στον εσωτερικό έλεγχο, αυτές οι τεχνικές καθιστούν δυνατό να καταλάβει ο ελεγκτής ποια χαρακτηριστικά συμβάλλουν στην εμφάνιση ενός red flag, πώς συγκρίνεται μια συναλλαγή με "κανονικά" πρότυπα και ποιοι παράγοντες αυξάνουν ή μειώνουν τον υπολογισμένο κίνδυνο. Η XAI λειτουργεί, έτσι, ως γέφυρα μεταξύ του αλγορίθμου και της ανθρώπινης κρίσης, επιτρέποντας την ελεγκτική τεκμηρίωση.

4.8.3 Ερμηνεία αποτελεσμάτων στο πλαίσιο εσωτερικού ελέγχου

Η ερμηνεία δεν είναι στατική διαδικασία. Ο ελεγκτής πρέπει να ελέγχει:

- αν τα αποτελέσματα ευθυγραμμίζονται με τις δικλίδες ελέγχου,
- αν υπάρχουν επιχειρησιακοί λόγοι που εξηγούν μια ανωμαλία,
- αν το μοντέλο υπερ-αντιδρά σε συγκεκριμένες κατηγορίες συναλλαγών,
- αν υπάρχει κίνδυνος false positives ή false negatives.

Η ερμηνεία συνδέεται άμεσα με την έννοια της ελεγκτικής επαγγελματικής κρίσης (professional judgement). Το μοντέλο υποδεικνύει, όμως ο ελεγκτής αποφασίζει. Όπως σημειώνουν οι Brown-Liburd & Vasarhelyi (2015), η ΤΝ μπορεί να ενισχύσει τη διαδικασία ελέγχου αλλά δεν μπορεί να την υποκαταστήσει.

4.8.4 Νομιμότητα και λογοδοσία (accountability)

Η νομιμότητα των συστημάτων red-flag detection δεν αφορά μόνο τα δεδομένα αλλά και τις αλγοριθμικές διαδικασίες. Οι οργανισμοί πρέπει να τεκμηριώνουν:

- τη διαδικασία εκπαίδευσης του μοντέλου,
- τις μετρικές απόδοσης,
- τον τρόπο αξιολόγησης της ερμηνευσιμότητας,
- τον τρόπο ενσωμάτωσης της ανθρώπινης εποπτείας,
- τη συμμόρφωση με τον GDPR και το AI Act.

Σύμφωνα με το AI Act, η λογοδοσία είναι κρίσιμη γιατί σε περίπτωση νομικών ή άλλων αμφισβητήσεων, πρέπει να είναι δυνατό να αποδειχθεί ότι ο οργανισμός ενήργησε κατά τρόπο νόμιμο και διαφανή.

4.9 Παρακολούθηση λειτουργίας και συντήρηση συστημάτων ΤΝ στην ανίχνευση εταιρικών Red Flags

Η λειτουργία ενός συστήματος Τεχνητής Νοημοσύνης για την ανίχνευση εταιρικών red flags δεν ολοκληρώνεται με την αρχική εκπαίδευση και την εγκατάστασή του στον οργανισμό. Αντίθετα, η πραγματική πολυπλοκότητα ξεκινά μετά την ένταξή του (deployment), όπου το μοντέλο αλληλεπιδρά με πραγματικά δεδομένα, δυναμικά επιχειρησιακά περιβάλλοντα και μεταβαλλόμενες μορφές κινδύνου. Επισημαίνεται ότι η συνεχής παρακολούθηση και συντήρηση (model monitoring & maintenance) είναι εξίσου κρίσιμες διαδικασίες με την αρχική μοντελοποίηση, ιδιαίτερα σε εφαρμογές ανίχνευσης απάτης και anomaly detection (Sculley et al., 2015).

4.9.1 Η ανάγκη συνεχούς παρακολούθησης των μοντέλων

Τα συστήματα ΤΝ που λειτουργούν σε περιβάλλοντα εσωτερικού ελέγχου αντιμετωπίζουν μία θεμελιώδη πρόκληση: τα δεδομένα και οι συμπεριφορές των χρηστών αλλάζουν**.** Αυτό το φαινόμενο είναι γνωστό ως concept drift και έχει σημαντική επίδραση στην αξιοπιστία των συστημάτων ανίχνευσης red-flag. Οι απατεώνες προσαρμόζουν τις τακτικές τους ακριβώς με στόχο να παρακάμπτουν τα υπάρχοντα συστήματα εντοπισμού. Αυτό σημαίνει ότι ένα μοντέλο που λειτουργούσε άριστα κατά την αρχική του φάση μπορεί να υποβαθμιστεί σημαντικά με την πάροδο του χρόνου. Στο πλαίσιο του εσωτερικού ελέγχου, τέτοιες μεταβολές συνδέονται άμεσα με ελεγκτικό κίνδυνο, καθώς μειώνουν την ικανότητα του συστήματος να εντοπίζει πραγματικά red flags.

4.9.2 Κατηγορίες monitoring: performance, data και concept drift

Η παρακολούθηση ενός συστήματος ΤΝ περιλαμβάνει τρεις βασικές κατηγορίες:

α) Performance monitoring

Αφορά την παρακολούθηση των δεικτών απόδοσης (precision, recall). Ακόμα και μικρές μεταβολές μπορεί να υποδεικνύουν κάποια αστοχία.

β) Data drift monitoring

Αφορά την παρακολούθηση της εισροής δεδομένων για να εντοπιστούν αλλαγές στις μεταβλητές που συνυπολογίζει το σύστημα για να καταλήξει σε κάποιο συμπέρασμα, στις συσχετίσεις των δεδομένων, στα μοτίβα που εντοπίζει καθώς και στους στατιστικούς δείκτες ποιότητας. Οι αλλαγές αυτές μπορεί να οφείλονται σε τεχνικά

προβλήματα, σε αλλαγές στην επιχειρησιακή διαδικασία ή ακόμα και σε απόπειρες χειραγώγησης του συστήματος.

γ) Concept drift monitoring
Συνδέεται με αλλαγές στη σχέση μεταξύ εισόδου και εξόδου, δηλαδή με αλλαγές στη βάση της συμπεριφοράς της απάτης. Για τον εσωτερικό έλεγχο, η έγκαιρη ανίχνευση drift είναι κρίσιμη. Διαφορετικά, ένα σύστημα μπορεί να συνεχίζει να λειτουργεί «τυπικά» αλλά να χάνει πραγματικά red flags (Gama et al, 2014).

4.9.3 Διαδικασίες συνεχούς συντήρησης (maintenance cycle)
Η συντήρηση ενός συστήματος ΤΝ περιλαμβάνει έναν πλήρη κύκλο ζωής (ML lifecycle), που με βάση τους Sculley et al. (2015) είναι συχνά πολύ πιο περίπλοκος από τον κύκλο ανάπτυξης. Τα βασικά βήματα είναι:

4.9.4 Επανεκπαίδευση (retraining)
Γίνεται είτε περιοδικά είτε όταν εντοπιστεί drift. Οι οργανισμοί συχνά επιλέγουν είτε μηνιαίο κύκλο επανεκπαίδευσης είτε τριμηνιαίο. Πολλές φορές όμως μπορεί να χρειαστεί επανεκπαίδευση μετά τον εντοπισμό ενός συμβάντος όπου το σύστημα είχε αστοχία, Επίσης μπορεί να χρειαστεί και μετά την αλλαγή κάποιας διαδικασίας ή νομοθεσίας.

4.9.5 Αναβάθμιση μοντέλου (model updates)
Αφορά την βελτίωση των χαρακτηριστικών, την ενσωμάτωση νέων πηγών δεδομένων και των ενημερώσεων ασφαλείας.

Συμπεράσματα

Η παρούσα εργασία ανέδειξε ότι η Τεχνητή Νοημοσύνη (ΤΝ) μπορεί να μετασχηματίσει ουσιαστικά τον τρόπο με τον οποίο οι οργανισμοί ανιχνεύουν red flags, αξιολογούν κινδύνους και εντοπίζουν ενδείξεις απάτης. Η ανάλυση των σύγχρονων τεχνικών μηχανικής μάθησης, των μεθόδων επεξεργασίας της φυσικής γλώσσας, των αλγορίθμων ανίχνευσης ανωμαλιών και των υβριδικών πλαισίων AI–Human έδειξε ότι η ΤΝ διαθέτει τη δυνατότητα να ενισχύσει την αποτελεσματικότητα, την ακρίβεια και την ταχύτητα της ελεγκτικής διαδικασίας, προσφέροντας ένα επίπεδο ανάλυσης που δεν ήταν εφικτό με τα παραδοσιακά εργαλεία.
Ωστόσο, η μελέτη καταλήγει με σαφήνεια στο συμπέρασμα ότι, παρά τις τεχνικές προόδους, η ανθρώπινη εποπτεία παραμένει απαραίτητη. Κανένα μοντέλο ΤΝ, ανεξάρτητα από την πολυπλοκότητα ή τη μαθησιακή του ικανότητα, δεν μπορεί να αναλάβει πλήρως τη λειτουργία της ελεγκτικής κρίσης. Τα αποτελέσματα των μοντέλων επηρεάζονται από προκαταλήψεις στα δεδομένα, από concept drift, από τεχνικούς περιορισμούς και από συστημική αβεβαιότητα. Τα φαινόμενα αυτά καθιστούν σαφές ότι η ΤΝ, χωρίς ενεργή συμμετοχή ανθρώπινων ελεγκτών, δεν μπορεί να εγγυηθεί τη συμμόρφωση, τη διαφάνεια και την ακρίβεια που απαιτείται σε περιβάλλοντα εταιρικής διακυβέρνησης.

Η ανά χείρας εργασία, τονίζει ότι η αλληλεπίδραση μεταξύ ανθρώπου και μηχανής δεν είναι συμπληρωματική μόνο σε λειτουργικό επίπεδο αλλά και σε στρατηγικό. Η ΤΝ μπορεί να εντοπίζει μοτίβα, αποκλίσεις και ύποπτες συμπεριφορές αλλά ο άνθρωπος είναι αυτός που ερμηνεύει το πλαίσιο, αξιολογεί τις συνθήκες, κρίνει τις προθέσεις και κατανοεί τις οργανωσιακές ιδιαιτερότητες. Τα μοντέλα ΤΝ διευκολύνουν το έργο του ελεγκτή αλλά δεν μπορούν να αντικαταστήσουν τη βαθιά γνώση του περιβάλλοντος, τη διορατικότητα και την επαγγελματική κρίση που απαιτούνται στις τελικές αποφάσεις.
Η διερεύνηση των ζητημάτων δεοντολογίας, των απαιτήσεων του GDPR και του υπό διαμόρφωση AI Act κατέδειξε επίσης ότι η ανθρώπινη εποπτεία δεν αποτελεί απλώς βέλτιστη πρακτική αλλά νομική υποχρέωση. Η ανάγκη για επεξηγησιμότητα, τεκμηρίωση, λογοδοσία, καθιστά αδιανόητη την πλήρη αυτοματοποίηση. Τα συστήματα που χαρακτηρίζουν συναλλαγές, εργαζομένους ή προμηθευτές ως ύποπτους πρέπει να λειτουργούν υπό στενή ανθρώπινη παρακολούθηση, διασφαλίζοντας ότι οι αποφάσεις λαμβάνονται δίκαια, με σεβασμό στα δικαιώματα των υποκειμένων και με γνώμονα τη διασφάλιση του οργανισμού.
Συνολικά, η εργασία υποστηρίζει ότι το μέλλον της ανίχνευσης red flags βρίσκεται στα υβριδικά μοντέλα AI–Human, όπου η ΤΝ αναλαμβάνει την ανάλυση μεγάλων όγκων δεδομένων, την αναγνώριση προτύπων και την αυτόματη ειδοποίηση, ενώ ο άνθρωπος διατηρεί την τελική ευθύνη της λήψης αποφάσεων, την ερμηνεία των ευρημάτων, την αξιολόγηση της επιχειρησιακής πραγματικότητας και την προσαρμογή του συστήματος στις αλλαγές του περιβάλλοντος.
Τα υβριδικά μοντέλα προσφέρουν το καλύτερο αποτέλεσμα: συνδυάζουν την ταχύτητα και την ανάλυση της ΤΝ με την κρίση, την ενσυναίσθηση, τη γνώση και την ηθική ευθύνη του ανθρώπου. Επιτρέπουν, έτσι, τη δημιουργία ενός συστήματος ελέγχου που είναι όχι μόνο αποτελεσματικό αλλά και δίκαιο, με διαφάνεια και συμβατό με τις αρχές της εταιρικής διακυβέρνησης.
Τελικά, η ΤΝ δεν αντικαθιστά τον εσωτερικό έλεγχο αλλά τον ενισχύει. Η προστιθέμενη αξία προκύπτει από τη συνεργασία ανθρώπου και μηχανής, από ένα πλαίσιο όπου οι αλγόριθμοι λειτουργούν ως εργαλεία, όχι ως αυτόνομοι κριτές. Η επιτυχημένη ενσωμάτωση της ΤΝ στις διαδικασίες ανίχνευσης red flags εξαρτάται από τον βαθμό στον οποίο οι οργανισμοί αντιλαμβάνονται αυτή τη συνεργατική σχέση και οικοδομούν συστήματα που αξιοποιούν τη δύναμη της τεχνολογίας χωρίς να απεμπολούν την ανθρώπινη κρίση.

Βιβλιογραφία

- Association of Certified Fraud Examiners (ACFE). (2024). Occupational Fraud 2024: A Report to the Nations.
- Brown-Liburd, H., & Vasarhelyi, M. (2015). Big Data and Audit Evidence. *Journal of Emerging Technologies in Accounting, 12*(1), 1–16.
- Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2013). *Internal Control — Integrated Framework.*
- Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2017). *Enterprise Risk Management — Integrating with Strategy and Performance.*
- Corporate Finance Institute (CFI). (2024). *Fraud red flags*. https://corporatefinanceinstitute.com/resources/esg/fraud-red-flags/ Corporate Finance Institute(accessed 25.09.2025)
- Cressey, D. R. (1953). *Other People's Money: A Study in the Social Psychology of Embezzlement.*
- Daniel E. O'Leary, Vernon J. Richardson, Marcia Weidenmier Watson; Data-Driven Audits: Audit Analytic Platforms and General Ledger Analytic Tools. *Current Issues in Auditing* 1 May 2025; 19 (1): A1–A9. https://doi.org/10.2308/CIIA-2023-027. (accessed 20.10.2025)
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv:1702.08608*. https://arxiv.org/abs/1702.08608
- Ehlermann-Cache, N. (2015). *Red flags: Indicators that may help prevent, detect and investigate malpractices.*https://rai-see.org/php_sets/uploads/2015/07/Red-flags_indicators_that_may_help_prevent_detect_and_investigate_malpractices-Mrs_Nicola_Ehlermann-Cache.pdf Regional Anti-Corruption Initiative (accessed 25.11.2025)
- El Naqa, I., & Murphy, M. J. (2015). What is machine learning?. In *Machine learning in radiation oncology: theory and applications* (pp. 3-11). Cham: Springer International Publishing.
- Elkan, C. (2001). The foundations of cost-sensitive learning. In *IJCAI '01*.
- ENISA. (2015). *Cloud Security Guide for SMEs.* European Union Agency for Cybersecurity.
- European Commission. (2019). *Ethics Guidelines for Trustworthy AI.* https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (accessed 25.11.2025)
- European Commission. (2021). *Proposal for a Regulation laying down harmonised rules on Artificial Intelligence (AI Act).* https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206(accessed 15.09.2025)

- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Rossi, F. (2021). An ethical framework for a good AI Society: Opportunities, risks, principles, and recommendations. Ethics. *Governance, and Policies in Artificial Intelligence*.
- Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM computing surveys (CSUR)*, *46*(4), 1-37.
- Goldstein, M., & Uchida, S. (2016). A comparative evaluation of unsupervised anomaly detection algorithms for multivariate data. *PloS one*, *11*(4), e0152173.
- Grabosky P. & Duffield G, (2001). Red flags of fraud. Trends & issues in crime and criminal justice no. 200. Canberra: Australian Institute of Criminology. https://www.aic.gov.au/publications/tandi/tandi200 Australian Institute of Criminology (accessed 15.11.2025)
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284.
- https://www.biocatch.com/ai-fraud-financial-crime-survey (accessed 11.11.2025)
- https://www.google.gr/books/edition/Security_and_Privacy_in_Cloud_Based_AI/TIY7EQAAQBAJ?hl=el&gbpv=0 (accessed 27.11.2025)
- https://dart.deloitte.com/USDART/pdf/5560dc0b-61a0-4f81-b53b-360f7bd041b8 (accessed 22.11.2025)
- https://www.ibm.com/think/topics/natural-language-processing (accessed 20.11.2025)
- Institute of Internal Auditors (IIA). (2017). *International Standards for the Professional Practice of Internal Auditing.*
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., ... & Zhao, S. (2021). Advances and open problems in federated learning. *Foundations and trends® in machine learning*, *14*(1–2), 1-210.
- Kogan, A., Alles, M. G., Vasarhelyi, M. A., & Wu, J. (2014). Design and evaluation of a continuous data level auditing system. Auditing: A Journal of Practice & Theory, 33(4), 221-245.
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. nature, 521(7553), 436-444.
- Lipton, Z. C. (2016). The mythos of model interpretability. *Communications of the ACM* (as arXiv:1606.03490). https://arxiv.org/abs/1606.03490(accessed 20.10.2025)
- Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.
- Munteanu, V., Zuca, M. R., Horaicu, A., Florea, L. A., Poenaru, C. E., & Anghel, G. (2024). Auditing the risk of financial fraud using the red flags technique. *Applied Sciences*, *14*(2), 757.
- Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems*, *50*(3), 559-569.
- Organisation for Economic Co-operation and Development (OECD). (2015). *G20/OECD Principles of Corporate Governance.*

- Phua, C., Lee, V., Smith, K., & Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119*.
- Quinlan, J. R. (2014). *C4. 5: programs for machine learning*. Elsevier.
- Russell, S., Norvig, P., Popineau, F., Miclet, L., & Cadet, C. (2021). Intelligence artificielle: une approche moderne (4e édition). Pearson France.
- Sathe, A. (2025). Forensic Accounting in the Age of AI: Detecting Fraud with Data Anomalies. International Journal of Social Impact, 10(3), 492-503. DIP: 18.02.053/20251003, DOI: 10.25215/2455/1003053
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., ... & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in neural information processing systems*, *28*.
- Shah Samarth Dr. Vikhyat Singhal (2025) Security and Privacy in Cloud-Based AI. DeepMisti Publication.
- Sidauruk, D. L. (2024). Data analytics in fraud prevention and detection by Government internal supervisory apparatuses at ministries/institutions/local governments: a mixed-method study. *Asia Pacific Fraud Journal*, *9*(2), 241-260.
- Solix. (2024). *Σημασία Ποιότητας δεδομένων*. (Solix Technologies, Inc.) accesed 27/11/2025
- Stone, P., & Veloso, M. (2000). Multiagent systems: A survey from a machine learning perspective. Autonomous Robots, 8(3), 345-383.
- Supervizor. (2025). *How Internal Audit Can Work Smarter, Not Harder with Analytics*. (supervizor.com) accesed 27/11/2025
- Turing, A. M. (1950). Computing machinery and intelligence.
- Vasarhelyi, M. A., Kogan, A., & Tuttle, B. M. (2015). Big data in accounting: An overview. *Accounting Horizons*, *29*(2), 381-396.
- Voigt, P., & Von dem Bussche, A. (2017). *The EU General Data Protection Regulation (GDPR)*. Springer
- Wachter, S., & Mittelstadt, B. (2019). A right to reasonable inferences: Re-thinking data protection law in the age of Big Data and AI. *Columbia Business Law Review*, *2019*(2), 494
- Warren, J. D., Moffitt, K. C., & Byrnes, P. (2015). How big data will change accounting. *Accounting horizons*, *29*(2), 397-407.
- West, J., & Bhattacharya, M. (2016). Intelligent financial fraud detection: a comprehensive review. *Computers & security*, *57*, 47-66.
- Wolfe, David T., and Dana R. Hermanson. *"The Fraud Diamond: Considering the Four Elements of Fraud*." CPA Journal 74.12 (2004).
- Zhang, J., Yang, X., & Appelbaum, D. (2015). Toward effective big data analysis in continuous auditing. *Accounting Horizons*, *29*(2), 469-476.
- Αντωνιάδου, Π. (2025). *Ανίχνευση τραπεζικής απάτης με χρήση μηχανικής μάθησης.*
- Γενίτσαρης, Ν. (2024). *Η ανίχνευση της απάτης με την χρήση νέων τεχνολογιών.*
- Γιαννάκης, Π. (2018). Ο εσωτερικός έλεγχος στην κεντρική δημόσια διοίκηση ως μηχανισμός προαγωγής της αποτελεσματικότητας της διοίκησης: η περίπτωση της Ελλάδας. Πανεπιστήμιο Αιγαίου.
- Γκεβρέκη, Κ. (2024). *Εσωτερικός έλεγχος και νέες τεχνολογίες* (Μεταπτυχιακή διατριβή). Πανεπιστήμιο Μακεδονίας.

- Διγενάκης Π., (2022). *Μεθοδολογίες τεχνητής νοημοσύνης για τον εντοπισμό της χρηματοοικονομικής απάτης: Μια βιβλιογραφική ανασκόπηση* (Master's thesis, Technical University of Crete (Greece).
- Καπετάνιου, Ε. (2023). *Η επίδραση του εσωτερικού ελέγχου στην αποτελεσματική διαχείριση κινδύνου*. Μεταπτυχιακή εργασία, Πανεπιστήμιο Μακεδονίας
- Κιορτσή, Π., Ιγγλεζάκης, Ι., Φείδας, Χ., Κοσμόπουλος, Δ., & Θωμόπουλος, Γ. (2024). *Η ρύθμιση των Μεγάλων Γλωσσικών Μοντέλων (LLMs) στον Κανονισμό (ΕΕ) 2024/1689 για την τεχνητή νοημοσύνη και τον ΓΚΠΔ*. Επιθεώρηση Δικαίου Πληροφορικής, 5(1).
- Κούγιας, Κ. Ν. (2022). *Ανίχνευση και πρόβλεψη απάτης στο λιανικό εμπόριο μέσω εξόρυξης δεδομένων*(Μεταπτυχιακή διατριβή). Πανεπιστήμιο Πειραιώς.
- Κουτσουνάκη, Α. (2020). *Ελεγκτική στην ελληνική δημόσια διοίκηση με τη χρήση τεχνολογιών πληροφορικής και επικοινωνιών καθώς και προηγμένων μεθόδων ανάλυσης* (Μεταπτυχιακή διατριβή)..
- Λαζαρίδου, Κ. (2024). *Αξιοποίηση της τεχνητής νοημοσύνης στον εσωτερικό έλεγχο* (Μεταπτυχιακή διατριβή). Πανεπιστήμιο Μακεδονίας.
- Λαφτσίδης, Μ. Σ. (2008). *Επισκόπηση της εξόρυξης δεδομένων* , Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης.
- Λογοθέτης, Α. (2019). *Ο εσωτερικός έλεγχος στο δημόσιο τομέα. Μεταπτυχιακή διατριβή*, Πανεπιστήμιο Αιγαίου.
- Τερζούδη, Χ. (2021). *Ανίχνευση απάτης σε οικονομικές οντότητες: Μια προσέγγιση με επίκεντρο τον εσωτερικό έλεγχο* (Μεταπτυχιακή διατριβή). Πανεπιστήμιο Μακεδονίας.
- Τζίφας Ι. (2025) *Ο Ρόλος της Τεχνητής Νοημοσύνης στον Εσωτερικό και Εξωτερικό Έλεγχο των Ανωνύμων Εταιρειών στην Ελλάδα,* Ανοικτό Πανεπιστήμιο.
- Τσατσούλας, Ι. (2022). *Συνεισφορά των λογιστικών πληροφοριακών συστημάτων στη λειτουργία των επιχειρήσεων* (Μεταπτυχιακή διατριβή). Πανεπιστήμιο Μακεδονίας.
- Τσουνιάς, Μ. (2023). *Αυτοματισμός και Τεχνητή Νοημοσύνη στον Εσωτερικό Έλεγχο* (Μεταπτυχιακή διατριβή). Πανεπιστήμιο Πειραιώς.
- Φαρμάκη, Λ. (2023). *Ο εσωτερικός έλεγχος και ο ρόλος του στην πρόληψη και αποτροπή της λογιστικής απάτης.* Μεταπτυχιακή εργασία, Πανεπιστήμιο Μακεδονίας.
- Χαρίτου, Στ. (χ.χ.) *Ενδείξεις απάτης / διαφθοράς στη δημόσια διοίκηση*. https://eclass.ekdd.gr/esdda/modules/document/file.php/KST_BEID_PDOKF111/Χαρίτου%20Στέλλα%20-%20Ενδείξεις%20Απάτης%20%26%20Διαφθοράς%20στη%20Δημόσια%20Διοίκηση.pdf (accessed 10.11.2025)