ΕΛΛΗΝΙΚΟ ΑΝΟΙΚΤΟ ΠΑΝΕΠΙΣΤΗΜΙΟ

Πρόγραμμα Σπουδών: ΤΡΑΠΕΖΙΚΗ, ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΤΕΧΝΟΛΟΓΙΑ (FinTech)

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«Τεχνητή Νοημοσύνη (AI) και Μηχανική Μάθηση (ML) στη Διαχείριση Κινδύνων»

ΟΝΟΜ/ΜΟ ΦΟΙΤΗΤΡΙΑΣ:

ΜΑΡΙΑ ΝΤΙΝΟΥ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:

ΑΝΔΡΕΑΣ ΑΝΔΡΙΚΟΠΟΥΛΟΣ

Β’ ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:

ΣΤΕΦΑΝΟΣ ΠΑΠΑΔΑΜΟΥ

Πάτρα, Μάιος 2026

Περιεχόμενα

Λίστα Γραφημάτων

Λίστα Πινάκων

Περίληψη

Τεχνητή Νοημοσύνη (AI) και Μηχανική Μάθηση (ML) στη Διαχείριση Κινδύνων

Η παρούσα εργασία εξετάζει τον ρόλο της Τεχνητής Νοημοσύνης (AI) και της Μηχανικής Μάθησης (ML) στη διαχείριση κινδύνων, τόσο από θεωρητική όσο και από εμπειρική σκοπιά. Στόχος ήταν η συστηματική διερεύνηση του τρόπου με τον οποίο οι τεχνολογίες αυτές μετασχηματίζουν τις διαδικασίες αναγνώρισης, αξιολόγησης και απόκρισης σε κρίσιμες μορφές χρηματοοικονομικού και λειτουργικού κινδύνου. Η προσέγγιση βασίστηκε αφενός σε ανασκόπηση της σύγχρονης βιβλιογραφίας, και αφετέρου σε εμπειρική μελέτη με χρήση δομημένου ερωτηματολογίου.

Η θεματική ανασκόπηση επικεντρώθηκε σε τέσσερις βασικές κατηγορίες κινδύνου: πιστωτικό, απάτης, συστημικό και κυβερνο-κίνδυνο, αποτυπώνοντας τις τεχνολογικές και κανονιστικές προσεγγίσεις που υιοθετούνται σε κάθε τομέα. Οι τεχνικές εποπτευόμενης μάθησης επικρατούν στην πιστοληπτική αξιολόγηση λόγω της ανάγκης για ερμηνευσιμότητα και της επάρκειας δομημένων δεδομένων. Στην ανίχνευση απάτης και στην κυβερνοασφάλεια, η χρήση μη εποπτευόμενων και βαθιών αλγορίθμων ενδείκνυται για τον εντοπισμό ακραίων ή σπάνιων προτύπων, ενώ ο συστημικός κίνδυνος προσεγγίζεται μέσω δικτυακών μοντέλων και πολυπαραγοντικών προσομοιώσεων. Ιδιαίτερη έμφαση δόθηκε στην ανάγκη ενσωμάτωσης εργαλείων Explainable AI (XAI) και στη συμμόρφωση με κανονιστικά πλαίσια όπως ο GDPR, ως προϋπόθεση για την κοινωνική αποδοχή των συστημάτων αυτών.

Συμπληρωματικά, πραγματοποιήθηκε εμπειρική έρευνα με στόχο την κατανόηση των παραγόντων που επηρεάζουν την αποδοχή της AI στη διαχείριση κινδύνου από επαγγελματίες του κλάδου. Βασισμένη στο εμπλουτισμένο μοντέλο TAM (Technology Acceptance Model), η μελέτη περιλάμβανε μεταβλητές όπως η αντιληπτή χρησιμότητα, η ευκολία χρήσης, η εμπιστοσύνη και ο αντιληπτός κίνδυνος. Το ερωτηματολόγιο διανεμήθηκε σε δείγμα 75 επαγγελματιών από τους τομείς της χρηματοοικονομικής, τεχνολογίας και ρυθμιστικών υπηρεσιών. Η στατιστική ανάλυση περιέλαβε έλεγχο αξιοπιστίας κλιμάκων (Cronbach’s α), συσχετίσεις μεταβλητών (Pearson) και παλινδρομικά μοντέλα.

Τα ευρήματα έδειξαν ότι η αντιληπτή χρησιμότητα, η ευκολία χρήσης και η εμπιστοσύνη συσχετίζονται θετικά και σημαντικά με την πρόθεση χρήσης συστημάτων AI, ενώ ο αντιληπτός κίνδυνος έχει αποτρεπτική επίδραση. Η εμπειρική ανάλυση επιβεβαιώνει τη σημασία θεσμικής και τεχνολογικής προσαρμογής για την ενίσχυση της αποδοχής των συστημάτων, καθώς και την ανάγκη ενίσχυσης της επεξηγησιμότητας. Επιπλέον, διαφάνηκε

ένα υψηλό επίπεδο αποδοχής των δυνατοτήτων της AI, παρά την αναγνώριση τεχνικών, οργανωσιακών και δεοντολογικών προκλήσεων.

Συνολικά, η εργασία προσφέρει μία διπλή συμβολή: α) θεωρητικά, μέσω της αποτύπωσης του υφιστάμενου ερευνητικού πεδίου, και β) εμπειρικά, μέσα από την καταγραφή στάσεων, προθέσεων και περιορισμών στην εφαρμογή τεχνολογιών AI στον πραγματικό χώρο των οργανισμών. Η σύνθεση των δύο αυτών προσεγγίσεων επιτρέπει την εξαγωγή πλούσιων συμπερασμάτων με πρακτική αξία, προσφέροντας τη βάση για μελλοντική διεπιστημονική και εφαρμοσμένη έρευνα που θα γεφυρώνει την τεχνική και την κοινωνική διάσταση της καινοτομίας.

Abstract

Artificial Intelligence (AI) and Machine Learning (ML) in Risk Management

This thesis seeks to explore the application of Artificial Intelligence and Machine Learning in the risk management field with the aim of understanding the role of the new technologies in the risk identification, assessment, and management process. The aim of this research is to explore the application of the new technologies in the four main risk management fields: credit risk management, fraud risk management, systemic risk management, and cyber risk management.

From the literature reviewed in the course of this research, it is evident that the approach to the application of the risk management process is dependent on the type of risk, the availability of data, and the level of maturity. In credit risk management, the application of the risk management process is dominated by the use of the supervised learning approach based on the availability of data and the need to have the ability to interpret the results. The application of the unsupervised and deep learning approach is also on the increase in the field of fraud risk management and cyber risk management; however, the results are not interpretable. The application of the complex approach is also on the increase in the field of systemic risk management through simulations based on the network and agent-based approach; however, the approach is not at the expected level of maturity due to the level of data confidentiality and the complexity of the approach applied.

One of the most significant themes that can be gauged from the literature is the need to strike a balance between performance and explainability. Further, the EU's GDPR guidelines have also led to the adoption of Explainable AI (XAI) technologies such as SHAP, LIME, and counterfactual reasoning. Ethical issues such as fairness and bias have also been considered an integral part of the system, which again brings us to the socio-technical perspective on the adoption and implementation of AI technology in the organization.

In addition to the literature review, a research study was conducted to understand the behavioral factors that affect the adoption of AI technology in the organization for risk management. A quantitative research design was used to conduct the research study with the help of the Technology Acceptance Model (TAM), incorporating trust and risk as variables in the study design. A questionnaire was used as the research tool to collect the required data from a sample size of 75 professionals in the finance and technology sectors.The results of the multiple linear regression test showed that perceived usefulness, perceived ease of use, and

trust had positive significant effects on behavioral intention, while perceived risk had a negative significant effect on behavioral intention.

In conclusion, it can be said that the thesis makes a two-fold contribution. First and foremost, it offers a structured literature review. This offers a comprehensive overview of the current research trends and the practical challenges involved in the application of AI technology in risk management. Second, it offers insights into the behavioral intention to apply AI technology in risk management and the factors involved in it.

Κεφάλαιο 1: Εισαγωγή

Η ραγδαία αύξηση της πολυπλοκότητας των οικονομικών δραστηριοτήτων, σε συνδυασμό με την αβεβαιότητα που χαρακτηρίζει το παγκόσμιο επιχειρηματικό περιβάλλον, έχει καταστήσει τη διαχείριση κινδύνων βασικό άξονα στρατηγικού σχεδιασμού των οργανισμών. Ειδικότερα στον χρηματοπιστωτικό τομέα, οι κίνδυνοι πιστωτικοί, λειτουργικοί, αγοραστικοί και συστημικοί παρουσιάζουν πολυδιάστατη φύση και απαιτούν σύγχρονα εργαλεία για τον εντοπισμό, την πρόβλεψη και την αντιμετώπισή τους (Power et al., 2018). Η παραδοσιακή προσέγγιση διαχείρισης κινδύνων, βασισμένη σε στατιστικά και οικονομετρικά μοντέλα, θεωρείται πλέον περιορισμένη, καθώς αδυνατεί να προσεγγίσει την αυξημένη πολυπλοκότητα και τα μεγάλα δεδομένα που χαρακτηρίζουν τις σημερινές αγορές (Alexander, 2020).

Στο διεθνές πεδίο, η επιστημονική κοινότητα και οι επιχειρήσεις έχουν στραφεί σε τεχνολογίες Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) ως λύση στην ανάγκη για προγνωστικά μοντέλα υψηλής ακρίβειας και αυτονομίας. Η εμπειρία από τη χρήση τέτοιων μεθόδων σε πιστωτικούς οργανισμούς, ασφαλιστικές εταιρείες, και επενδυτικές τράπεζες έχει δείξει ότι οι αλγόριθμοι ML, όπως τα δέντρα απόφασης, τα νευρωνικά δίκτυα και οι μέθοδοι ενισχυτικής μάθησης, μπορούν να εντοπίσουν μη γραμμικές σχέσεις και κρυμμένα πρότυπα που αγνοούνται από τα κλασικά μοντέλα (Heaton et al., 2017). Παράλληλα, η ενσωμάτωση εργαλείων επεξεργασίας φυσικής γλώσσας (NLP) και αναλυτικής μεγάλων δεδομένων έχει ενισχύσει περαιτέρω την ακρίβεια και την προσαρμοστικότητα των συστημάτων διαχείρισης κινδύνου (Brynjolfsson & McAfee, 2017).

Σκοπός της παρούσας διπλωματικής εργασίας είναι να διερευνήσει, μέσω βιβλιογραφικής ανάλυσης, τον τρόπο με τον οποίο οι τεχνολογίες AI και ML ενσωματώνονται στις διαδικασίες διαχείρισης κινδύνων. Η εργασία εστιάζει στις εφαρμογές αυτών των τεχνολογιών στους επιμέρους τύπους κινδύνου —με έμφαση στον πιστωτικό, λειτουργικό, και τον κίνδυνο απάτης— καθώς και στη ρύθμιση, τις ηθικές επιπτώσεις και τις προκλήσεις υλοποίησης τους. Η μελέτη βασίζεται αποκλειστικά ερευνητικά άρθρα, δημοσιευμένα σε διεθνή έγκριτα περιοδικά του χώρου της χρηματοοικονομικής και της επιστήμης δεδομένων.

Επιπλέον, στο εμπειρικό σκέλος της εργασίας, διερευνάται η πρόθεση αποδοχής συστημάτων Τεχνητής Νοημοσύνης από επαγγελματίες στον τομέα της διαχείρισης κινδύνου. Η εμπειρική ανάλυση βασίζεται στο θεωρητικό πλαίσιο του TechnologyAcceptanceModel (TAM), εμπλουτισμένο με μεταβλητές εμπιστοσύνης (trust) και αντιληπτού κινδύνου (perceived risk). Μέσω δομημένου ερωτηματολογίου, συλλέχθηκαν 75 απαντήσεις από επαγγελματίες των κλάδων χρηματοοικονομικής και τεχνολογίας.

Το ερευνητικό ερώτημα της εργασίας διαμορφώνεται ως εξής: "Ποιοι παράγοντες επηρεάζουν την πρόθεση χρήσης συστημάτων AI στη διαχείριση κινδύνου;" Με βάση το σχετικό θεωρητικό υπόβαθρο, εξετάζονται οι παρακάτω ερευνητικές υποθέσεις:

- Υ1: Η αντιληπτή χρησιμότητα επηρεάζει θετικά την πρόθεση χρήσης AI.
- Υ2: Η αντιληπτή ευκολία χρήσης επηρεάζει θετικά την πρόθεση χρήσης AI.
- Υ3: Η εμπιστοσύνη στα συστήματα AI επηρεάζει θετικά την πρόθεση χρήσης.
- Υ4: Ο αντιληπτός κίνδυνος επηρεάζει αρνητικά την πρόθεση χρήσης.

Η παρούσα εργασία υποστηρίζει τη θέση ότι η Τεχνητή Νοημοσύνη δεν αποτελεί απλώς ένα νέο τεχνολογικό εργαλείο, αλλά μια βαθιά μετασχηματιστική δύναμη στη διαχείριση κινδύνων. Τα σημαντικότερα συμπεράσματα της μελέτης αναδεικνύουν την ανωτερότητα των μοντέλων ML στην πρόβλεψη αστοχιών πληρωμών και εντοπισμό απάτης, καθώς και τη δυνατότητα υιοθέτησης εξατομικευμένων μοντέλων κινδύνου σε πραγματικό χρόνο (Khandani et al., 2010). Επιπλέον, τονίζεται η αναγκαιότητα συνδυασμού των τεχνικών προσεγγίσεων με κανονιστικά και ηθικά πλαίσια, προκειμένου να διασφαλιστεί η υπεύθυνη χρήση των τεχνολογιών AI.

Ωστόσο, η εργασία αναγνωρίζει συγκεκριμένους περιορισμούς. Παρότι περιλαμβάνει εμπειρικό σκέλος, η ποσοτική ανάλυση βασίζεται σε δείγμα 75 επαγγελματιών, γεγονός που ενδέχεται να περιορίζει τη γενικευσιμότητα των αποτελεσμάτων. Επιπλέον, τα δεδομένα είναι αυτοαναφορικά και αποτυπώνουν αντιλήψεις και προθέσεις χρήσης και όχι πραγματική επιχειρησιακή απόδοση συστημάτων AI. Παράλληλα, μεγάλο μέρος της βιβλιογραφίας επικεντρώνεται σε ανεπτυγμένες αγορές, περιορίζοντας την εφαρμοσιμότητα των συμπερασμάτων σε μικρότερες ή αναδυόμενες οικονομίες. Τέλος, η διαρκής τεχνολογική εξέλιξη δύναται να καταστήσει ορισμένα μοντέλα ή προσεγγίσεις ταχέως παρωχημένα, επηρεάζοντας τη μακροπρόθεσμη ισχύ των συμπερασμάτων (Campbell et al., 2020).

Μεθοδολογικά, η εργασία ακολουθεί επισκόπηση της διεθνούς βιβλιογραφίας με ανάλυση υψηλής ποιότητας επιστημονικών άρθρων των τελευταίων ετών. Η ανάλυση δομείται θεματικά γύρω από πέντε βασικές ενότητες: AI στη διαχείριση πιστωτικού κινδύνου, στη διαχείριση απάτης, στον συστημικό και cyber κίνδυνο, καθώς και στο ρυθμιστικό/ηθικό πλαίσιο. Κάθε θεματική ενότητα συγκρίνει διαφορετικά μοντέλα και προσεγγίσεις, εντοπίζει κοινά ευρήματα και αποκλίσεις, και συνθέτει τα πορίσματα υπό κριτική σκοπιά. Η μέθοδος αυτή επιτρέπει την εστίαση στη θεωρητική εμβάθυνση και τη σύνδεση των τεχνολογικών εξελίξεων με την πρακτική εφαρμογή στη χρηματοοικονομική διαχείριση.

Το ερευνητικό μέρος περιλαμβάνει στατιστική ανάλυση των δεδομένων που συλλέχθηκαν μέσω του ερωτηματολογίου. Χρησιμοποιούνται δείκτες περιγραφικής στατιστικής, ο έλεγχος

αξιοπιστίας Cronbach’s alpha, η συσχέτιση Pearson και η πολλαπλή γραμμική παλινδρόμηση, ώστε να εξεταστεί η εγκυρότητα των υποθέσεων και η συμβολή κάθε παράγοντα στην πρόθεση χρήσης AI.

Το περιεχόμενο της εργασίας διαρθρώνεται σε έξι κεφάλαια. Στο Κεφάλαιο 2 παρουσιάζεται το θεωρητικό πλαίσιο, περιλαμβάνοντας τους ορισμούς, τις τεχνολογίες AI/ML και τη σχέση τους με τους τύπους κινδύνων. Στο Κεφάλαιο 3 η Μεθοδολογία εκπόνησης της εργασίας ενώ στο Κεφάλαιο 4 αναλύεται η διεθνής βιβλιογραφία, οργανωμένη ανά θεματική ενότητα και τύπο κινδύνου. Στο Κεφάλαιο 5 παρουσιάζεται συγκριτική και κριτική ανάλυση των ερευνητικών ευρημάτων. Στο Κεφάλαιο 6 συνοψίζονται τα βασικά συμπεράσματα της μελέτης, αναλύονται οι πρακτικές συνέπειες και προτείνονται πεδία για μελλοντική έρευνα.

Κεφάλαιο 2: Θεωρητικό Πλαίσιο

2.1 Ορισμοί και έννοιες

Η κατανόηση της έννοιας του κινδύνου και της διαχείρισής του αποτελεί το θεμέλιο για την ενσωμάτωση τεχνολογιών όπως η Τεχνητή Νοημοσύνη (AI) και η Μηχανική Μάθηση (ML) στο χρηματοοικονομικό σύστημα. Ο κίνδυνος, από χρηματοοικονομική άποψη, αναφέρεται στην πιθανότητα απόκλισης των πραγματικών αποτελεσμάτων από τα αναμενόμενα και, ως εκ τούτου, στην ενδεχόμενη απώλεια κεφαλαίου ή εισοδήματος (Jorion, 2007). Η διαχείριση κινδύνων, συνεπώς, συνίσταται στη συστηματική διαδικασία αναγνώρισης, αξιολόγησης, παρακολούθησης και αντιμετώπισης των κινδύνων που απειλούν τη λειτουργία ή τους στόχους ενός οργανισμού (Lam, 2014). Στον χρηματοπιστωτικό κλάδο, οι συνηθέστεροι τύποι κινδύνου περιλαμβάνουν τον πιστωτικό, τον λειτουργικό, τον κίνδυνο αγοράς και τον συστημικό κίνδυνο.

Ο πιστωτικός κίνδυνος αφορά την πιθανότητα αθέτησης των υποχρεώσεων από έναν δανειολήπτη ή συμβαλλόμενο, και αποτελεί ίσως τη σημαντικότερη απειλή για τις τράπεζες και τα χρηματοπιστωτικά ιδρύματα. Ο λειτουργικός κίνδυνος αναφέρεται στις ζημίες που μπορεί να προκύψουν από εσωτερικές διαδικασίες, ανθρώπινα σφάλματα, συστήματα ή εξωτερικά γεγονότα, ενώ ο κίνδυνος αγοράς σχετίζεται με τις διακυμάνσεις στις τιμές των περιουσιακών στοιχείων λόγω μεταβολών σε επιτόκια, ισοτιμίες και τιμές μετοχών (Hull, 2018). Ο συστημικός κίνδυνος, τέλος, περιγράφει την απειλή διαταραχής του συνόλου του χρηματοπιστωτικού συστήματος εξαιτίας της κατάρρευσης ενός ή περισσότερων θεσμών, όπως έγινε εμφανές κατά την κρίση του 2008 (Acharyaetal., 2017).

Η Τεχνητή Νοημοσύνη (AI) ορίζεται ως η επιστήμη και το σύνολο τεχνικών που επιτρέπουν στα υπολογιστικά συστήματα να εκτελούν εργασίες που απαιτούν ανθρώπινη νοημοσύνη, όπως η λήψη αποφάσεων, η αναγνώριση προτύπων, η πρόβλεψη και η μάθηση από εμπειρία (Russell&Norvig, 2021). Η AI περιλαμβάνει ποικίλες υποκατηγορίες, μεταξύ των οποίων η Μηχανική Μάθηση (ML), η Ενισχυτική Μάθηση (ReinforcementLearning), η Επεξεργασία Φυσικής Γλώσσας (NLP) και η Υπολογιστική Όραση. Η ML, ειδικότερα, είναι ο κλάδος της AI που επικεντρώνεται στην ανάπτυξη αλγορίθμων που μπορούν να μαθαίνουν και να βελτιώνονται μέσα από δεδομένα χωρίς να είναι ρητά προγραμματισμένοι (Goodfellowetal., 2016). Οι τεχνικές ML χωρίζονται σε εποπτευόμενη (supervised), μη εποπτευόμενη (unsupervised) και ενισχυτική μάθηση (reinforcementlearning), με εφαρμογές στην πιστοληπτική αξιολόγηση, την ανίχνευση απάτης και τη διαχείριση κινδύνων.

Οι τεχνολογίες AI και ML έχουν επιφέρει ουσιαστική μεταβολή στον τρόπο με τον οποίο οι οργανισμοί εντοπίζουν και διαχειρίζονται κινδύνους. Σε αντίθεση με τα παραδοσιακά στατιστικά μοντέλα, τα οποία βασίζονται σε αυστηρές παραδοχές και περιορισμένη ικανότητα ανάλυσης μεγάλων και μη δομημένων δεδομένων, τα συστήματα ML μπορούν να επεξεργαστούν τεράστιους όγκους πληροφοριών και να εντοπίσουν σύνθετα πρότυπα που αλλιώς θα παρέμεναν αφανή (Gatla, 2023). Η δυνατότητα συνεχούς μάθησης και προσαρμογής σε νέα δεδομένα καθιστά τις εφαρμογές της ML εξαιρετικά πολύτιμες στον ταχύτατα μεταβαλλόμενο χρηματοοικονομικό τομέα.

Παράλληλα, η ανάγκη για κανονιστική συμμόρφωση, διαφάνεια και κατανόηση των αποφάσεων που λαμβάνονται από τα «έξυπνα» συστήματα οδήγησε στην ανάπτυξη της έννοιας της Εξηγήσιμης Τεχνητής Νοημοσύνης (ExplainableAI – XAI). Η XAI επιχειρεί να αντιμετωπίσει τις ανησυχίες σχετικά με τον αλγοριθμικό αυταρχισμό και την έλλειψη λογοδοσίας, παρέχοντας εργαλεία που εξηγούν με ανθρώπινο τρόπο τις προβλέψεις των μοντέλων (Guidottietal., 2019). Αυτό είναι ιδιαιτέρως κρίσιμο στον τραπεζικό τομέα, όπου η ερμηνευσιμότητα αποτελεί προϋπόθεση για κανονιστική αποδοχή.

Συμπερασματικά, οι έννοιες της διαχείρισης κινδύνου, της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης είναι αλληλένδετες σε ένα σύγχρονο περιβάλλον όπου οι παραδοσιακές μέθοδοι δεν επαρκούν για την αντιμετώπιση των νέων προκλήσεων. Η τεχνολογική πρόοδος προσφέρει σημαντικές δυνατότητες, οι οποίες, ωστόσο, πρέπει να ενσωματώνονται με υπευθυνότητα, διαφάνεια και κατανόηση των ορίων τους.

2.2 Παραδοσιακές μέθοδοι διαχείρισης κινδύνου

Πριν την ενσωμάτωση τεχνολογιών όπως η Τεχνητή Νοημοσύνη και η Μηχανική Μάθηση, η διαχείριση χρηματοοικονομικών κινδύνων στηριζόταν σε μεθόδους που ανήκαν στον χώρο των στατιστικών και οικονομετρικών εργαλείων. Οι τεχνικές αυτές αποτέλεσαν τη βάση για την κατανόηση και τον ποσοτικό υπολογισμό της έκθεσης ενός οργανισμού στον κίνδυνο, επιτρέποντας στους υπεύθυνους λήψης αποφάσεων να εφαρμόσουν στρατηγικές περιορισμού και πρόληψης. Οι πλέον ευρέως διαδεδομένες μέθοδοι περιλάμβαναν τον υπολογισμό της διακύμανσης και της τυπικής απόκλισης, την ανάλυση της συνδιακύμανσης, τη χρήση υποδειγμάτων παλινδρόμησης και κυρίως τον υπολογισμό της Μέγιστης Δυνητικής Ζημίας (ValueatRisk – VaR) (Jorion, 2007).

Η μέθοδος ValueatRisk υπήρξε για δεκαετίες το πρότυπο των τραπεζών και των επενδυτικών ιδρυμάτων για την εκτίμηση της μέγιστης πιθανής ζημίας σε συγκεκριμένο χρονικό ορίζοντα

και επίπεδο εμπιστοσύνης. Η VaR υπολογίζεται μέσω ιστορικής προσομοίωσης, παραμετρικών μοντέλων ή MonteCarlo προσεγγίσεων, ενώ θεωρείται κατάλληλη για τον αποδοτικό υπολογισμό κινδύνων αγοράς (Hull, 2018). Παρά την ευρεία εφαρμογή της, η μέθοδος έχει δεχθεί κριτική για τη βασισμένη σε αυστηρές υποθέσεις κατανομής των αποδόσεων και την αδυναμία της να εντοπίσει σπάνια αλλά καταστροφικά γεγονότα, όπως φαινόμενα τύπου "μαύρου κύκνου" (Taleb, 2007).

Στην περίπτωση του πιστωτικού κινδύνου, οι παραδοσιακές μέθοδοι διαχείρισης βασίζονταν κυρίως στη χρήση μοντέλων λογιστικής παλινδρόμησης (logisticregression), τα οποία στόχευαν στην πρόβλεψη της πιθανότητας αθέτησης υποχρεώσεων από δανειολήπτες, με βάση χρηματοοικονομικούς δείκτες, πιστωτικό ιστορικό και άλλα δημογραφικά δεδομένα (Hand&Henley, 1997). Η απλότητα και η διαφάνεια αυτών των μοντέλων τα καθιστούσαν προτιμητέα τόσο από κανονιστικές αρχές όσο και από οργανισμούς. Ωστόσο, παρουσίαζαν αδυναμία στην αποτύπωση μη γραμμικών σχέσεων και στη διαχείριση μεγάλων όγκων δεδομένων, περιορίζοντας έτσι τη δυνατότητα γενίκευσης και την προβλεπτική τους ισχύ σε σύνθετα οικονομικά περιβάλλοντα (Lessmannetal., 2015).

Στην περίπτωση της διαχείρισης λειτουργικού κινδύνου, η έμφαση δινόταν σε ποιοτικές μεθόδους, όπως η χρήση σεναρίων, η αποτύπωση εσωτερικών ελέγχων και η αξιολόγηση διαδικασιών. Πολλοί οργανισμοί υιοθετούσαν το πλαίσιο BaselII, το οποίο επέβαλλε είτε τη χρήση του βασικού δείκτη (BasicIndicatorApproach), είτε της τυποποιημένης προσέγγισης (StandardizedApproach), είτε του πιο σύνθετου AdvancedMeasurementApproach (AMA) για τον υπολογισμό του απαιτούμενου κεφαλαίου (BCBS, 2006). Οι παραπάνω μέθοδοι βασίζονταν σε ιστορικά στοιχεία και εσωτερικά δεδομένα ζημιών, περιορίζοντας όμως τη δυνατότητα πρόβλεψης νέων ή ραγδαία εξελισσόμενων μορφών κινδύνου.

Αξιοσημείωτη θέση κατείχε επίσης η χρήση της Θεωρίας Χαρτοφυλακίου του Markowitz (Mean–Variance Model), καθώς και η Ανάλυση Ευαισθησίας και StressTesting. Η Ανάλυση Ευαισθησίας επέτρεπε τον υπολογισμό της μεταβολής της αξίας ενός χαρτοφυλακίου ως απόκριση σε αλλαγές συγκεκριμένων παραμέτρων, ενώ τα stresstests χρησιμοποιούνταν για τη μέτρηση της ανθεκτικότητας ενός οργανισμού σε ακραία άλλα πιθανά σενάρια (BankofEngland, 2015). Αν και αυτές οι προσεγγίσεις συνέβαλαν σημαντικά στην κατανόηση της συμπεριφοράς των επενδύσεων υπό αβεβαιότητα, υστερούσαν στην ανίχνευση σύνθετων σχέσεων και διασυνδέσεων μεταξύ μεταβλητών (Taleb, 2007· Boucher et al., 2014· Alexander, 2020).

Η βασική αδυναμία των παραδοσιακών μεθόδων έγκειται στο γεγονός ότι απαιτούν προϋποθέσεις σχετικά με τη γραμμικότητα, την ανεξαρτησία των μεταβλητών και τη

σταθερότητα του συστήματος, συνθήκες που σπάνια πληρούνται σε πραγματικά οικονομικά δεδομένα. Επιπλέον, η ανάγκη για ακριβή μοντελοποίηση, σε ένα περιβάλλον μεγάλου όγκου δεδομένων (bigdata), μεταβλητότητας και ασάφειας, οδήγησε σταδιακά σε περιορισμούς στην αξιοπιστία των μοντέλων αυτών (Boucheretal., 2014). Στο πλαίσιο αυτό, η αναζήτηση πιο ευέλικτων και δυναμικών εργαλείων διαχείρισης κινδύνων, που μπορούν να επεξεργάζονται μεγάλα και ετερογενή σύνολα δεδομένων και να μαθαίνουν αυτόνομα, κατέστη επιτακτική, ανοίγοντας τον δρόμο για τη χρήση της Τεχνητής Νοημοσύνης.

2.3 Εισαγωγή στην Τεχνητή Νοημοσύνη και Μηχανική Μάθηση

Η Τεχνητή Νοημοσύνη (Artificial Intelligence – AI) έχει εξελιχθεί σε μία από τις πιο καθοριστικές τεχνολογικές εξελίξεις του 21ου αιώνα, επηρεάζοντας ποικίλους τομείς, από την υγειονομική περίθαλψη και τις μεταφορές, μέχρι τη βιομηχανία και το χρηματοοικονομικό σύστημα. Ο όρος AI αναφέρεται στη δυνατότητα των υπολογιστικών συστημάτων να εκτελούν λειτουργίες που απαιτούν ανθρώπινη νοημοσύνη, όπως είναι η λήψη αποφάσεων, η μάθηση, η πρόβλεψη, η αναγνώριση μοτίβων και η κατανόηση γλώσσας (Russell & Norvig, 2021). Αν και οι απαρχές της AI τοποθετούνται ήδη από τη δεκαετία του 1950, μόνο τις τελευταίες δύο δεκαετίες κατέστη εφικτή η πρακτική της εφαρμογή, χάρη στη ραγδαία πρόοδο σε υπολογιστική ισχύ, στην αύξηση των διαθέσιμων δεδομένων και στη βελτίωση των αλγορίθμων.

Η Μηχανική Μάθηση (Machine Learning – ML) αποτελεί έναν από τους βασικότερους υποκλάδους της AI και εστιάζει στην ικανότητα των υπολογιστικών συστημάτων να μαθαίνουν από δεδομένα και να βελτιώνουν την απόδοσή τους χωρίς να είναι ρητά προγραμματισμένα (Alpaydin, 2020). Σε αντίθεση με τις παραδοσιακές προσεγγίσεις, όπου οι κανόνες λειτουργίας καθορίζονται εξ ολοκλήρου από τον άνθρωπο, τα μοντέλα ML βασίζονται στην εκπαίδευση μέσω παραδειγμάτων, επιτρέποντας τη διαμόρφωση γενικών κανόνων μέσω της ανάλυσης δεδομένων. Η Μηχανική Μάθηση θεωρείται σήμερα ως η κύρια τεχνολογία που καθιστά δυνατή την «έξυπνη» λειτουργία των πληροφοριακών συστημάτων.

Οι βασικοί τύποι Μηχανικής Μάθησης περιλαμβάνουν την εποπτευόμενη μάθηση (supervised learning), τη μη εποπτευόμενη μάθηση (unsupervised learning) και την ενισχυτική μάθηση (reinforcement learning). Στην εποπτευόμενη μάθηση, τα δεδομένα που χρησιμοποιούνται για εκπαίδευση περιλαμβάνουν εισροές και τις αντίστοιχες επιθυμητές εξόδους, επιτρέποντας στο μοντέλο να μάθει έναν κανόνα αντιστοίχισης (Hastie et al., 2009). Τα πιο συνηθισμένα παραδείγματα περιλαμβάνουν προβλήματα ταξινόμησης (classification) και παλινδρόμησης

(regression). Αντιθέτως, στη μη εποπτευόμενη μάθηση τα δεδομένα δεν φέρουν επισημάνσεις, και το σύστημα καλείται να εντοπίσει πρότυπα ή ομάδες εντός του συνόλου δεδομένων, όπως συμβαίνει στη συσταδοποίηση (clustering) ή στη μείωση διαστάσεων (dimensionality reduction) (Murphy, 2022). Η ενισχυτική μάθηση χρησιμοποιείται όταν ο αλγόριθμος αλληλεπιδρά με ένα δυναμικό περιβάλλον και μαθαίνει μέσω ενισχύσεων ή ποινών, όπως εφαρμόζεται στη ρομποτική ή στα συστήματα βελτιστοποίησης χαρτοφυλακίων (Sutton & Barto, 2018).

Η επιτυχία της Μηχανικής Μάθησης βασίζεται σε μεγάλο βαθμό στη διαθεσιμότητα και την ποιότητα των δεδομένων. Οι αλγόριθμοι χρειάζονται μεγάλους όγκους δεδομένων για να κατανοήσουν και να γενικεύσουν σωστά, κάτι που κατέστη εφικτό με την άνοδο του φαινομένου των Big Data. Ο όρος Big Data αναφέρεται σε δεδομένα μεγάλου όγκου, ποικιλομορφίας και ταχύτητας, τα οποία συχνά ξεπερνούν τις δυνατότητες διαχείρισης των παραδοσιακών συστημάτων βάσεων δεδομένων (Kitchin, 2014). Η δυνατότητα των τεχνολογιών AI να εξάγουν γνώση από τέτοια δεδομένα σε πραγματικό χρόνο τις καθιστά ιδανικές για δυναμικά περιβάλλοντα, όπως η διαχείριση κινδύνων.

Τα σύγχρονα μοντέλα AI περιλαμβάνουν επίσης τις μεθόδους βαθιάς μάθησης (Deep Learning), οι οποίες βασίζονται σε νευρωνικά δίκτυα πολλών επιπέδων και έχουν αποδειχθεί εξαιρετικά αποτελεσματικές σε εργασίες υψηλής πολυπλοκότητας, όπως η επεξεργασία φυσικής γλώσσας (NLP) και η αναγνώριση εικόνας. Τα νευρωνικά δίκτυα μιμούνται τη δομή του ανθρώπινου εγκεφάλου και είναι σε θέση να εντοπίζουν υψηλού επιπέδου συσχετίσεις σε δεδομένα (LeCun et al., 2015). Στην τραπεζική και χρηματοοικονομική σφαίρα, τα deep learning συστήματα χρησιμοποιούνται για την ανίχνευση απάτης, την πρόβλεψη πιστωτικών κινδύνων και την αυτοματοποιημένη διαχείριση κινδύνων (Heaton et al., 2017).

Η Τεχνητή Νοημοσύνη εφαρμόζεται ήδη σε πολλούς τομείς της επιχειρησιακής λειτουργίας, με αξιοσημείωτη παρουσία στη χρηματοοικονομική ανάλυση, στη διαχείριση κινδύνου, στην καταπολέμηση απάτης και στη βελτιστοποίηση επενδυτικών στρατηγικών. Ενδεικτικά, αλγόριθμοι ML χρησιμοποιούνται από χρηματοπιστωτικά ιδρύματα για την ταξινόμηση των δανειοληπτών σε επίπεδα κινδύνου, βασισμένοι σε μη γραμμικές σχέσεις μεταξύ των μεταβλητών (Khandani et al., 2010). Παράλληλα, οι τεχνικές unsupervised learning χρησιμοποιούνται στην ανίχνευση μη φυσιολογικών συναλλαγών που ενδέχεται να υποκρύπτουν δόλο (Bolton & Hand, 2002). Οι τεχνολογίες αυτές προσφέρουν τη δυνατότητα γρήγορης επεξεργασίας μεγάλων όγκων δεδομένων και λήψης αποφάσεων σε πραγματικό χρόνο, γεγονός που τις καθιστά εξαιρετικά αποτελεσματικές.

Η ανάπτυξη της AI συνοδεύεται και από μια σημαντική φιλοσοφική και κανονιστική συζήτηση γύρω από την ερμηνευσιμότητα των αποφάσεων, την ηθική χρήση των αλγορίθμων και την προστασία της ιδιωτικότητας. Τα «μαύρα κουτιά» των αλγορίθμων – δηλαδή τα μοντέλα των οποίων οι εσωτερικές λειτουργίες δεν είναι κατανοητές – προκαλούν ανησυχία σχετικά με τη διαφάνεια και την υπευθυνότητα (Doshi-Velez & Kim, 2017). Η ανάγκη για «εξηγήσιμη AI» (Explainable AI – XAI) έχει οδηγήσει στην ανάπτυξη τεχνικών που επιτρέπουν την κατανόηση των αποφάσεων που λαμβάνονται από τα μοντέλα, ειδικά σε περιβάλλοντα υψηλού ρίσκου όπως οι χρηματοοικονομικές αγορές και η υγειονομική περίθαλψη.

Αξίζει να σημειωθεί ότι, παρά τις εντυπωσιακές δυνατότητες των τεχνολογιών AI και ML, η επιτυχής ενσωμάτωσή τους στις επιχειρησιακές λειτουργίες απαιτεί καινοτόμο σκέψη, προσαρμοστικότητα και σαφή στρατηγική. Η χρήση τέτοιων συστημάτων απαιτεί επένδυση όχι μόνο σε τεχνολογικές υποδομές, αλλά και σε ανθρώπινο κεφάλαιο με υψηλό επίπεδο εξειδίκευσης (Brynjolfsson & McElheran, 2016). Επιπλέον, τα δεδομένα πρέπει να είναι καθαρά, πλήρη και αντιπροσωπευτικά, διαφορετικά υπάρχει κίνδυνος αλγοριθμικής προκατάληψης και συστηματικών σφαλμάτων.

Συμπερασματικά, η Τεχνητή Νοημοσύνη και η Μηχανική Μάθηση αποτελούν πλέον θεμέλιους λίθους του ψηφιακού μετασχηματισμού στον χρηματοοικονομικό κλάδο. Οι δυνατότητές τους στην κατανόηση και πρόβλεψη πολύπλοκων φαινομένων, στην επεξεργασία μεγάλων δεδομένων και στη λήψη ταχέων αποφάσεων, τις καθιστούν απαραίτητα εργαλεία για την αποτελεσματική διαχείριση κινδύνων. Ωστόσο, η αξιοποίησή τους προϋποθέτει υπευθυνότητα, διαφάνεια και συνεχή αξιολόγηση, ώστε να διασφαλίζεται η συμμόρφωση με ηθικές, θεσμικές και κοινωνικές απαιτήσεις.

2.4 Τεχνικές AI/ML στη χρηματοοικονομική

Η χρηματοοικονομική επιστήμη, από τη φύση της, βασίζεται στην ανάλυση και την πρόβλεψη πολύπλοκων, δυναμικών και συχνά μη γραμμικών φαινομένων. Τα τελευταία χρόνια, η είσοδος της Τεχνητής Νοημοσύνης και, ειδικότερα, της Μηχανικής Μάθησης στην ανάλυση χρηματοοικονομικών δεδομένων, έχει οδηγήσει σε μια ριζική μεταμόρφωση των μεθόδων που χρησιμοποιούνται για την αξιολόγηση επενδυτικών κινδύνων, την πρόβλεψη αποδόσεων, την πιστωτική αξιολόγηση και την ανίχνευση απάτης (Gupta & Pathak, 2022). Σε αντίθεση με τις παραδοσιακές τεχνικές, τα μοντέλα AI/ML έχουν την ικανότητα να

εντοπίζουν κρυμμένα μοτίβα, να προσαρμόζονται σε νέα δεδομένα και να αυτοματοποιούν διαδικασίες ανάλυσης και λήψης αποφάσεων.

Μία από τις πιο διαδεδομένες τεχνικές είναι η χρήση αλγορίθμων δέντρων απόφασης (decision trees). Οι αλγόριθμοι αυτοί προσφέρουν διαφάνεια και είναι κατάλληλοι για εφαρμογές όπως η διαχείριση πιστωτικού κινδύνου, καθώς επιτρέπουν την απλή οπτικοποίηση των αποφάσεων. Οι εξελιγμένες μορφές, όπως τα Random Forests και το Gradient Boosting, αντιμετωπίζουν το πρόβλημα της υπερπροσαρμογής (overfitting) και ενισχύουν τη γενικευσιμότητα των μοντέλων (Chen & Guestrin, 2016). Τα συστήματα αυτά έχουν αποδειχθεί αποτελεσματικά σε προβλήματα κατηγοριοποίησης, όπως ο διαχωρισμός δανειοληπτών σε χαμηλό ή υψηλό πιστωτικό ρίσκο, και έχουν υιοθετηθεί από εμπορικές τράπεζες και fintech επιχειρήσεις.

Τα νευρωνικά δίκτυα (neural networks), ιδίως στη μορφή τους ως βαθιά νευρωνικά δίκτυα (deep neural networks), επιτρέπουν τη μοντελοποίηση πολύπλοκων μη γραμμικών σχέσεων και έχουν σημαντικές εφαρμογές στην πρόβλεψη αγοραίων μεταβλητών, στην ανάλυση συναλλαγών και στην πρόβλεψη πτώχευσης. Ειδικές δομές όπως τα Recurrent Neural Networks (RNNs) και τα Long Short-Term Memory (LSTM) μοντέλα επιτρέπουν την αποθήκευση πληροφορίας από προηγούμενες χρονικές στιγμές, διευκολύνοντας την ανάλυση χρηματοοικονομικών χρονικών σειρών (Fischer & Krauss, 2018). Τα μοντέλα αυτά έχουν χρησιμοποιηθεί με επιτυχία για την πρόβλεψη τιμών μετοχών, τη μεταβλητότητα της αγοράς και τη διαχείριση ρίσκου σε πορτφόλια.

Η επεξεργασία φυσικής γλώσσας (Natural Language Processing – NLP) αποτελεί ένα ταχέως αναπτυσσόμενο πεδίο εφαρμογής της AI στη χρηματοοικονομική. Η ικανότητα ανάλυσης μη δομημένων δεδομένων, όπως οικονομικές ειδήσεις, αναλύσεις αναλυτών, ανακοινώσεις εταιρειών ή ακόμα και αναρτήσεις στα κοινωνικά δίκτυα, δίνει νέα διάσταση στην πρόβλεψη των αγορών. Μέσω τεχνικών όπως το sentiment analysis, τα χρηματοπιστωτικά ιδρύματα μπορούν να ποσοτικοποιούν την «ψυχολογία της αγοράς» και να ενσωματώνουν αυτή τη μεταβλητή στις στρατηγικές τους (Nassirtoussi et al., 2014). Μεταγενέστερα, πιο προηγμένα μοντέλα όπως το BERT και το GPT έχουν βελτιώσει σημαντικά την ακρίβεια των προβλέψεων σε σχέση με τα παραδοσιακά bag-of-words μοντέλα.

Οι μέθοδοι συσταδοποίησης (clustering), όπως το K-Means ή το DBSCAN, χρησιμοποιούνται για τον εντοπισμό κατηγοριών πελατών ή πρότυπων συναλλαγών, χωρίς την ανάγκη προεπισημασμένων δεδομένων. Παράλληλα, οι αλγόριθμοι ανίχνευσης ανωμαλιών (anomaly detection) χρησιμοποιούνται ευρέως στην ανίχνευση απάτης, μέσω της αναγνώρισης ακραίων ή ασυνήθιστων παρατηρήσεων σε συναλλαγές (Hodge & Austin,

2004). Αυτές οι μέθοδοι είναι χρήσιμες όταν το κριτήριο αξιολόγησης δεν είναι πλήρως καθορισμένο, ή όταν οι επισημάνσεις απάτης είναι ελλιπείς ή ασύμμετρες.

Η ενισχυτική μάθηση (reinforcement learning) εισάγει μια διαφορετική προσέγγιση, εστιάζοντας στη μάθηση μέσω αλληλεπίδρασης του μοντέλου με το περιβάλλον και την προσαρμογή των στρατηγικών του με βάση την επιτυχία ή την αποτυχία του. Στη χρηματοοικονομική, τέτοια μοντέλα εφαρμόζονται σε αυτοματοποιημένους επενδυτικούς αλγορίθμους, όπου οι πράκτορες μαθαίνουν συνεχώς πώς να προσαρμόζουν τις θέσεις τους σε περιβάλλοντα μεταβλητότητας και αβεβαιότητας (Moody & Saffell, 2001). Αν και η ενισχυτική μάθηση απαιτεί σημαντικούς υπολογιστικούς πόρους και δεδομένα, μελέτες έχουν δείξει ότι μπορεί να επιτύχει βελτιωμένες αποδόσεις προσαρμοσμένες στον κίνδυνο (risk-adjusted returns), υψηλότερο Sharpe ratio και καλύτερη δυναμική προσαρμογή σε μεταβαλλόμενες συνθήκες αγοράς σε σύγκριση με παραδοσιακές στρατηγικές κατανομής χαρτοφυλακίου (Moody & Saffell, 2001).

Αναπόσπαστο μέρος της εφαρμογής των τεχνικών AI στη χρηματοοικονομική είναι η ερμηνευσιμότητα (interpretability). Η ανάγκη κανονιστικής συμμόρφωσης, καθώς και η εμπιστοσύνη των χρηστών, απαιτούν την ύπαρξη εργαλείων που εξηγούν τις προβλέψεις των μοντέλων. Για τον σκοπό αυτό έχουν αναπτυχθεί τεχνικές όπως τα SHAP values και το LIME, που επιτρέπουν την κατανόηση της συνεισφοράς κάθε εισόδου στο τελικό αποτέλεσμα (Barredo Arrieta et al., 2020). Η εφαρμογή αυτών των εργαλείων είναι ιδιαίτερα κρίσιμη στην πιστοληπτική αξιολόγηση και σε περιπτώσεις όπου η απόφαση του αλγορίθμου επηρεάζει άμεσα τον πελάτη.

Παράλληλα, η αξιοπιστία των μοντέλων εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων. Η προεπεξεργασία δεδομένων, η κανονικοποίηση, η αντιμετώπιση ελλιπών τιμών και η ανίχνευση σφαλμάτων αποτελούν απαραίτητα βήματα πριν την εκπαίδευση οποιουδήποτε μοντέλου (Provost & Fawcett, 2013). Η ύπαρξη μεροληπτικών δεδομένων μπορεί να οδηγήσει σε αλγοριθμικές στρεβλώσεις, ιδίως όταν πρόκειται για δεδομένα πελατών, και επομένως είναι σημαντικό οι οργανισμοί να διασφαλίζουν την ισότητα και τη διαφάνεια στα δεδομένα εισόδου.

Η επιτυχία των μοντέλων AI/ML δεν εξαρτάται μόνο από τις τεχνικές δυνατότητες, αλλά και από τον τρόπο ενσωμάτωσής τους στη συνολική επιχειρησιακή στρατηγική. Η δημιουργία διαλειτουργικών ομάδων που συνδυάζουν επιστήμονες δεδομένων, αναλυτές κινδύνου και στελέχη κανονιστικής συμμόρφωσης αποτελεί απαραίτητο βήμα για την επιτυχημένη υλοποίηση τέτοιων έργων. Παράλληλα, είναι κρίσιμο να διασφαλίζεται η συνεχής

επανεκπαίδευση και επικύρωση των μοντέλων, ώστε να ανταποκρίνονται σε διαρκώς μεταβαλλόμενες συνθήκες.

Συνοψίζοντας, οι τεχνικές AI/ML μεταμορφώνουν τον χρηματοοικονομικό τομέα μέσω της αυτοματοποίησης, της ακρίβειας και της προσαρμοστικότητας. Οι επενδύσεις σε τεχνολογίες AI δεν αποτελούν πλέον ανταγωνιστικό πλεονέκτημα, αλλά αναγκαιότητα για οργανισμούς που επιδιώκουν να διαχειριστούν αποτελεσματικά τον κίνδυνο στο σημερινό δυναμικό και απρόβλεπτο περιβάλλον.

2.5AI και ML στην Πιστωτική Αξιολόγηση

Η πιστωτική αξιολόγηση (credit scoring) αποτελεί κρίσιμο στοιχείο της λειτουργίας των χρηματοπιστωτικών ιδρυμάτων, καθορίζοντας την πιστοληπτική ικανότητα δανειοληπτών και επηρεάζοντας την απόφαση χορήγησης πίστωσης, τα επιτόκια, καθώς και τις εγγυήσεις. Καθώς η ακρίβεια αυτής της διαδικασίας είναι ζωτικής σημασίας για τον περιορισμό των μη εξυπηρετούμενων δανείων (NPLs) και την προστασία των ισολογισμών, έχει υπάρξει έντονο ενδιαφέρον για την υιοθέτηση τεχνικών Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) ως εναλλακτικές ή συμπληρωματικές λύσεις στα παραδοσιακά μοντέλα (Lessmann et al., 2015).

Οι κλασικές στατιστικές τεχνικές, όπως η λογιστική παλινδρόμηση και η διακριτική ανάλυση, κυριάρχησαν στο πεδίο για δεκαετίες, λόγω της ευκολίας εφαρμογής και της ερμηνευσιμότητάς τους (Hand & Henley, 1997). Παρά την αξιοπιστία τους, οι μέθοδοι αυτές περιορίζονται από την υπόθεση γραμμικότητας και τη δυσκολία διαχείρισης μεγάλων και πολυδιάστατων συνόλων δεδομένων. Η AI και η ML προσφέρουν την ευκαιρία επανασχεδιασμού της πιστωτικής αξιολόγησης με πιο ισχυρούς, προσαρμοστικούς και ικανούς μηχανισμούς πρόβλεψης.

Οι εποπτευόμενες τεχνικές μάθησης (supervised learning) είναι ιδιαίτερα δημοφιλείς στον χώρο της πιστωτικής αξιολόγησης, καθώς τα ιστορικά δεδομένα πελατών περιλαμβάνουν επισημασμένες εξαρτημένες μεταβλητές (π.χ. αν κάποιος καθυστέρησε ή όχι την αποπληρωμή δανείου). Μεταξύ των πιο συχνά χρησιμοποιούμενων τεχνικών συγκαταλέγονται οι αλγόριθμοι δέντρων απόφασης, και ιδιαίτερα οι εξελιγμένες εκδοχές τους όπως το Random Forest, το XGBoost και το LightGBM, που διαχειρίζονται εύκολα μεγάλες ποσότητες ετερογενών δεδομένων και ενσωματώνουν ενισχυμένες τεχνικές αποφυγής υπερπροσαρμογής (Chen & Guestrin, 2016).

Παράλληλα, τεχνητά νευρωνικά δίκτυα (ANNs) χρησιμοποιούνται όταν οι σχέσεις μεταξύ μεταβλητών είναι πολυδιάστατες και μη γραμμικές. Αν και χαρακτηρίζονται από έλλειψη ερμηνευσιμότητας, υπερέχουν σε όρους ακρίβειας σε μεγάλα datasets, όπως έχει καταδειχθεί σε πολλαπλές συγκριτικές μελέτες (West, 2000; Martens et al., 2009). Επιπλέον, τα βαθιά νευρωνικά δίκτυα (Deep Neural Networks), που αποτελούν βασικό στοιχείο της Deep Learning, έχουν αρχίσει να εφαρμόζονται, ειδικά όταν η πληροφορία προέρχεται από μη δομημένα δεδομένα όπως αρχεία περιήγησης, κινητές εφαρμογές ή περιεχόμενο από κοινωνικά δίκτυα (Kou et al., 2021).

Ένα ιδιαίτερο ενδιαφέρον παρουσιάζουν οι τεχνικές ενίσχυσης (boosting) και οι συνδυαστικές μέθοδοι (ensembles), οι οποίες συνδυάζουν προβλέψεις από πολλαπλά μοντέλα για τη βελτίωση της ακρίβειας. Οι μέθοδοι αυτές, όπως το AdaBoost, το Bagging και το Stacking, επιτυγχάνουν υψηλά ποσοστά ακριβείας και σταθερότητας, ιδιαίτερα όταν τα επιμέρους μοντέλα παρουσιάζουν ετερογένεια στις παραδοχές τους (Zhou et al., 2021). Οι ερευνητές Lessmann et al. (2015) διαπίστωσαν σε μελέτη με 41 ταξινομητές ότι οι ensemble μέθοδοι κατέλαβαν τις πρώτες θέσεις σε επιδόσεις σε σύνολα δεδομένων πιστωτικής αξιολόγησης.

Η εφαρμογή της μη εποπτευόμενης μάθησης (unsupervised learning) βρίσκει όλο και μεγαλύτερη εφαρμογή στην κατηγοριοποίηση πελατών χωρίς την ανάγκη ετικετών ή προκαθορισμένων εκβάσεων. Τεχνικές όπως το K-Means Clustering, το DBSCAN και οι ιεραρχικές μέθοδοι συσταδοποίησης χρησιμοποιούνται για την αναγνώριση υποομάδων πελατών με παρόμοια χαρακτηριστικά και συμπεριφορές (Li et al., 2017). Με αυτόν τον τρόπο, οι οργανισμοί μπορούν να προσαρμόσουν τα πιστωτικά τους προϊόντα ή να εντοπίσουν ομάδες υψηλού κινδύνου προτού ακόμα υπάρξουν δεδομένα αθέτησης.

Ένα πεδίο που εξελίσσεται ταχέως είναι η ενισχυτική μάθηση (Reinforcement Learning – RL), η οποία προσφέρει ένα δυναμικό περιβάλλον όπου το μοντέλο μαθαίνει μέσω επαναλαμβανόμενων αλληλεπιδράσεων με τα δεδομένα, βελτιώνοντας τις προβλέψεις του με βάση τη «συμπεριφορά» του στο παρελθόν (Xie et al., 2022). Η RL είναι ιδιαιτέρως χρήσιμη σε σενάρια όπου το περιβάλλον αλλάζει συνεχώς, όπως για παράδειγμα όταν οι δανειολήπτες προσαρμόζουν τη συμπεριφορά τους λόγω οικονομικών μεταβολών ή αλλαγών πολιτικής δανεισμού.

Παρά τα πλεονεκτήματα των παραπάνω τεχνικών, ένα από τα σημαντικότερα ζητήματα που έχει αναδειχθεί στη βιβλιογραφία είναι αυτό της ερμηνευσιμότητας (interpretability). Σε ένα περιβάλλον όπου οι αποφάσεις που λαμβάνονται με βάση τα αποτελέσματα των αλγορίθμων επηρεάζουν άμεσα την πρόσβαση των ατόμων σε οικονομικά αγαθά (π.χ. στεγαστικά δάνεια),

η απαίτηση για διαφάνεια καθίσταται επιτακτική (Lundberg & Lee, 2017). Εργαλεία όπως το SHAP, το LIME, αλλά και σύγχρονες τεχνικές όπως οι counterfactual explanations, προσφέρουν μηχανισμούς κατανόησης και εξήγησης των αποφάσεων ακόμα και από μοντέλα «μαύρου κουτιού».

Η κανονιστική συμμόρφωση ενισχύει περαιτέρω την ανάγκη για ερμηνεύσιμα και επαναλήψιμα μοντέλα. Η ευρωπαϊκή οδηγία GDPR προβλέπει το δικαίωμα των πολιτών να λαμβάνουν επεξηγήσεις για τις αυτοματοποιημένες αποφάσεις που τους επηρεάζουν. Ως εκ τούτου, πολλά χρηματοπιστωτικά ιδρύματα επιλέγουν να θυσιάσουν ένα μέρος της ακρίβειας προβλέψεων, επιλέγοντας μοντέλα όπως τα δέντρα απόφασης ή τις γενετικές συναρτήσεις (symbolic regression), προκειμένου να ενισχύσουν τη διαφάνεια και την κανονιστική αποδοχή (Barredo Arrieta et al., 2020).

Μια ενδιαφέρουσα εξέλιξη στη σύγχρονη βιβλιογραφία αφορά τα υβριδικά συστήματα (hybrid systems), τα οποία συνδυάζουν εποπτευόμενη και μη εποπτευόμενη μάθηση, ερμηνεύσιμους και black-box αλγορίθμους, καθώς και δομημένα και μη δομημένα δεδομένα. Για παράδειγμα, μια εφαρμογή μπορεί να συνδυάζει clustering για segmentation πελατών, supervised learning για πρόβλεψη αθέτησης και explainability tools για επεξήγηση της τελικής απόφασης (Kou et al., 2021). Αυτό δημιουργεί ένα πολυεπίπεδο σύστημα αξιολόγησης που είναι και ακριβές και κανονιστικά βιώσιμο.

Η δικαιοσύνη (fairness) των αλγορίθμων έχει αναδειχθεί ως ένας κρίσιμος ηθικός και κοινωνικός προβληματισμός. Επιστημονικές μελέτες έχουν δείξει ότι η αλγοριθμική μεροληψία μπορεί να ενισχυθεί εάν τα εκπαιδευτικά δεδομένα είναι μεροληπτικά, προκαλώντας διακρίσεις σε βάρος ευάλωτων κοινωνικών ομάδων (Hardt et al., 2016). Ως απάντηση, έχουν προταθεί τεχνικές όπως η απομάκρυνση ευαίσθητων χαρακτηριστικών (feature suppression), η αρχιτεκτονική εξισορρόπησης (adversarial de-biasing) και η προσδιορισμένη εξισορρόπηση του κόστους σφαλμάτων (equalized odds).

Μια ακόμη διάσταση της σύγχρονης βιβλιογραφίας αφορά τη χρήση εναλλακτικών πηγών δεδομένων (alternative data) για την ενίσχυση της ακρίβειας των μοντέλων πιστωτικού κινδύνου. Πέρα από τα παραδοσιακά δεδομένα (ιστορικό πληρωμών, εισόδημα, χρέος, ηλικία), οι σύγχρονες εφαρμογές αξιοποιούν δεδομένα από τη συμπεριφορά στο διαδίκτυο, κινητά τηλέφωνα, κοινωνικά δίκτυα, αλλά και μοτίβα κατανάλωσης ενέργειας ή τηλεπικοινωνιακής χρήσης (Galindo & Tamayo, 2000). Τα δεδομένα αυτά ενισχύουν την ακρίβεια ιδίως σε πληθυσμούς με περιορισμένο παραδοσιακό ιστορικό πίστωσης (π.χ. νεαροί ενήλικες, μετανάστες, μικρές επιχειρήσεις).

Σε αυτό το πλαίσιο, ο ρόλος των FinTech εταιρειών είναι καθοριστικός. Αυτές οι επιχειρήσεις έχουν επενδύσει μαζικά σε τεχνολογίες AI και ML για να προσφέρουν υπηρεσίες πιστοληπτικής αξιολόγησης σε πραγματικό χρόνο, συχνά μέσω mobile εφαρμογών και πλήρως αυτοματοποιημένων συστημάτων (Kou et al., 2021). Εταιρείες όπως η ZestFinance και η Upstart έχουν αναπτύξει μοντέλα που ξεπερνούν τα όρια των παραδοσιακών frameworks, με υψηλά ποσοστά αποδοχής και ρυθμιστικής αναγνώρισης.

Αξιοσημείωτο είναι ότι η εφαρμογή των τεχνολογιών AI στην πιστωτική αξιολόγηση δεν περιορίζεται μόνο στο σκέλος έγκρισης/απόρριψης. Περιλαμβάνει και την δυναμική παρακολούθηση της πιστοληπτικής συμπεριφοράς μετά την έγκριση δανείου, την αναγνώριση πιθανής επιδείνωσης του προφίλ του δανειολήπτη και τη δημιουργία πρώιμων μηχανισμών προειδοποίησης (early warning systems). Τέτοια συστήματα λειτουργούν σε συνδυασμό με ιστορικά οικονομικά δεδομένα, εξωτερικές πηγές (π.χ. πληροφορίες αγοράς) και αξιολογήσεις τρίτων, βελτιώνοντας την προγνωστική δυνατότητα των ιδρυμάτων (Baesens et al., 2003).

Ένα επιπλέον ενδιαφέρον σημείο αφορά την γεωγραφική διάσταση των εφαρμογών AI στην πιστωτική αξιολόγηση. Η βιβλιογραφία επισημαίνει διαφορές στον βαθμό αποδοχής και εφαρμογής μεταξύ αναπτυγμένων και αναπτυσσόμενων χωρών. Σε χώρες της Ασίας και της Αφρικής, όπου οι τράπεζες είναι λιγότερο εξαρτημένες από τις παραδοσιακές μορφές πιστωτικής πληροφόρησης, η AI έχει αξιοποιηθεί ως εναλλακτικό κανάλι πρόσβασης στην πίστωση, ειδικά μέσω μικροπιστώσεων και αγροτικών πιστωτικών εφαρμογών (Li et al., 2017).

Η συστηματική επισκόπηση της βιβλιογραφίας αναδεικνύει ότι το μέλλον της πιστωτικής αξιολόγησης θα κινηθεί προς την αυξημένη ενοποίηση τεχνολογιών, με συνδυασμό τεχνικών supervised και unsupervised learning, deep learning, NLP (σε περιπτώσεις επεξεργασίας κειμένων από αιτήσεις), αλλά και edge computing για real-time λήψη αποφάσεων. Ερευνητές προτείνουν επίσης την ενσωμάτωση ψηφιακών ταυτοτήτων και blockchain τεχνολογιών για την ασφαλή επαλήθευση και κοινή χρήση δεδομένων (Barredo Arrieta et al., 2020).

Εν κατακλείδι, η χρήση τεχνολογιών AI και ML στην πιστωτική αξιολόγηση δεν είναι απλώς μια τεχνική εξέλιξη, αλλά συνιστά μια βαθιά αλλαγή στο πώς οι χρηματοπιστωτικοί οργανισμοί αντιλαμβάνονται και διαχειρίζονται τον κίνδυνο. Παρά τα σημαντικά πλεονεκτήματα σε όρους ακρίβειας, ταχύτητας και κλίμακας, απαιτείται συνεχής αξιολόγηση των κινδύνων που ενέχονται, ιδίως ως προς την αλγοριθμική αδικία, την απώλεια ελέγχου των μοντέλων και τη συμμόρφωση με ηθικά και κανονιστικά πλαίσια. Η διεθνής επιστημονική κοινότητα καταλήγει σε ένα συνεκτικό μήνυμα: η τεχνητή νοημοσύνη δεν μπορεί να

υποκαταστήσει πλήρως την ανθρώπινη κρίση, αλλά μπορεί να την ενισχύσει αποφασιστικά εφόσον ενσωματωθεί υπεύθυνα και διαφανώς στο πιστωτικό οικοσύστημα.

2.6 Ανίχνευση Απάτης και Διαχείριση Κινδύνου Απάτης

Η απάτη αποτελεί μία από τις μεγαλύτερες απειλές για τη χρηματοοικονομική σταθερότητα και την αξιοπιστία των οικονομικών συναλλαγών. Οι απώλειες που προκύπτουν από απάτες στον τραπεζικό, ασφαλιστικό και εμπορικό τομέα ανέρχονται σε δισεκατομμύρια δολάρια ετησίως, με τις μορφές τους να γίνονται διαρκώς πιο εξελιγμένες και δύσκολες στον εντοπισμό (Phua et al., 2010). Η παραδοσιακή προσέγγιση ανίχνευσης απάτης βασίζεται σε κανόνες (rule-based systems), οι οποίοι δημιουργούνται από ειδικούς και βασίζονται σε ιστορικά παραδείγματα, ωστόσο παρουσιάζουν σοβαρούς περιορισμούς όσον αφορά την επεκτασιμότητα και την αποτελεσματικότητα έναντι νέων, άγνωστων μεθόδων απάτης (Ngai et al., 2011).

Η εισαγωγή της Τεχνητής Νοημοσύνης (AI) και της Μηχανικής Μάθησης (ML) στον εντοπισμό απάτης (fraud detection) έχει φέρει ριζικές αλλαγές, καθώς τα συστήματα αυτά μπορούν να επεξεργαστούν τεράστιους όγκους δεδομένων, να αναγνωρίσουν πρότυπα συμπεριφοράς και να εντοπίσουν ασυνήθιστες ή ύποπτες ενέργειες σε πραγματικό χρόνο (Bolton & Hand, 2002). Σε αντίθεση με τα συστήματα κανόνων, τα μοντέλα ML μαθαίνουν από τα δεδομένα, βελτιώνουν συνεχώς τις προβλέψεις τους και μπορούν να προσαρμοστούν στις συνεχώς μεταβαλλόμενες μεθόδους των δραστών.

Οι τεχνικές εποπτευόμενης μάθησης (supervised learning) βρίσκονται στον πυρήνα πολλών εφαρμογών ανίχνευσης απάτης, κυρίως σε περιβάλλοντα όπου υπάρχει επαρκής σήμανση δεδομένων (labeled data), δηλαδή παραδείγματα συναλλαγών που έχουν ήδη ταξινομηθεί ως νόμιμες ή απάτες. Δημοφιλείς αλγόριθμοι που χρησιμοποιούνται περιλαμβάνουν τη λογιστική παλινδρόμηση, τα decision trees, τα Random Forests, τους Support Vector Machines και τα τεχνητά νευρωνικά δίκτυα (Bahnsen et al., 2016). Παρόλα αυτά, η ύπαρξη ακραίας ανισορροπίας (class imbalance) στα δεδομένα – όπου οι απάτες αποτελούν ένα πολύ μικρό ποσοστό των συνολικών συναλλαγών – καθιστά τα κλασικά μοντέλα λιγότερο αποτελεσματικά, οδηγώντας σε μεγάλο αριθμό ψευδών θετικών (false positives) ή αρνητικών (false negatives) αποτελεσμάτων (Jurgovsky et al., 2018).

Για να αντιμετωπιστεί το πρόβλημα της ανισορροπίας, έχουν προταθεί διάφορες τεχνικές, όπως η δειγματοληψία (undersampling, oversampling), οι τεχνικές μετασχηματισμού των δεδομένων (SMOTE), καθώς και τα cost-sensitive learning μοντέλα, τα οποία ενσωματώνουν

διαφορετικό κόστος σφάλματος ανάλογα με τον τύπο του σφάλματος (Dal Pozzolo et al., 2015). Επιπλέον, η χρήση ensemble methods, όπως το Gradient Boosting και το XGBoost, έχει δώσει εξαιρετικά αποτελέσματα λόγω της δυνατότητας συνδυασμού διαφορετικών προβλέψεων και μείωσης του overfitting (Chen & Guestrin, 2016).

Η ανεξάρτητη ή σε συνδυασμό χρήση τεχνικών μη εποπτευόμενης μάθησης (unsupervised learning) είναι ιδιαίτερα χρήσιμη σε περιβάλλοντα όπου η απάτη δεν έχει εντοπιστεί ακόμα και, επομένως, δεν υπάρχουν ετικέτες. Τεχνικές όπως το clustering (π.χ. K-Means, DBSCAN), οι μέθοδοι ανίχνευσης ανωμαλιών (Isolation Forest, One-Class SVM, Autoencoders), και τα Gaussian Mixture Models προσφέρουν τη δυνατότητα ανάλυσης της "κανονικής" συμπεριφοράς και εντοπισμού αποκλίσεων που ενδέχεται να υποδηλώνουν απάτη (Aleskerov et al., 1997; Zanin et al., 2016).

Τα τελευταία χρόνια, οι τεχνικές βαθιάς μάθησης (deep learning) έχουν αποκτήσει σημαντική δυναμική στο πεδίο της ανίχνευσης απάτης, ιδιαίτερα σε περιβάλλοντα με μεγάλα και σύνθετα σύνολα δεδομένων. Τα νευρωνικά δίκτυα με πολλαπλά στρώματα (deep neural networks – DNNs), τα recurrent neural networks (RNNs) και τα Long Short-Term Memory (LSTM) μοντέλα είναι ιδιαίτερα αποτελεσματικά στην κατανόηση ακολουθιών συναλλαγών, όπου το χρονικό πλαίσιο έχει σημασία (Jurgovsky et al., 2018). Αυτά τα μοντέλα έχουν την ικανότητα να εντοπίζουν μοτίβα που εξελίσσονται με την πάροδο του χρόνου, καθιστώντας τα ιδανικά για την παρακολούθηση συνεχών χρηματοοικονομικών ροών.

Ένα ακόμη ταχύτατα εξελισσόμενο πεδίο είναι η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing – NLP), το οποίο επιτρέπει την ανάλυση μη δομημένων δεδομένων, όπως emails, κείμενα, αιτήσεις πελατών, συνομιλίες και αναφορές. Το NLP έχει χρησιμοποιηθεί με επιτυχία στην ανίχνευση απάτης σε ασφαλιστικές απαιτήσεις, όπου τα περιγραφικά κείμενα μπορούν να αποκαλύψουν ασυνήθιστες ή αντικρουόμενες πληροφορίες (Lindholm et al., 2021). Με την αξιοποίηση προηγμένων τεχνικών όπως το Word2Vec, το BERT και τα transformers, η κατανόηση του περιεχομένου γίνεται ακριβέστερη, βελτιώνοντας σημαντικά την ακρίβεια ανίχνευσης.

Οι εφαρμογές της ανίχνευσης απάτης μέσω AI/ML είναι πολυδιάστατες. Στον τραπεζικό τομέα, η κύρια έμφαση δίνεται στη διαχείριση απάτης καρτών, στην πλαστοπροσωπία πελατών και στην παρακολούθηση λογαριασμών σε πραγματικό χρόνο. Οι τράπεζες χρησιμοποιούν συστήματα που ενημερώνονται συνεχώς με νέα δεδομένα, εφαρμόζοντας online machine learning για να προσαρμόζουν τα μοντέλα στις νέες απειλές. Χαρακτηριστικό παράδειγμα αποτελεί η χρήση online learning αλγορίθμων για την ενσωμάτωση των

τελευταίων συναλλαγών στις προβλέψεις, χωρίς να απαιτείται πλήρης επανεκπαίδευση του μοντέλου (Dal Pozzolo et al., 2018).

Στον ασφαλιστικό τομέα, η απάτη μπορεί να περιλαμβάνει ψευδείς ή υπερβολικές απαιτήσεις αποζημίωσης, κατασκευασμένα περιστατικά, καθώς και οργανωμένες απάτες μέσω "κυκλωμάτων" (organized fraud rings). Η χρήση της AI στον ασφαλιστικό κλάδο περιλαμβάνει την ανάλυση ιστορικών απαιτήσεων, φωτογραφιών ζημιών, συνοδευτικών εγγράφων, καθώς και δεδομένων τρίτων (π.χ. αστυνομικές αναφορές) για τη δημιουργία σύνθετων risk profiles (Viaene et al., 2004).

Η ανάγκη για ερμηνευσιμότητα (explainability) των μοντέλων είναι εξαιρετικά κρίσιμη στο πλαίσιο της ανίχνευσης απάτης, καθώς οι αποφάσεις που λαμβάνονται ενδέχεται να οδηγήσουν σε απόρριψη συναλλαγών ή σε νομικές διαδικασίες. Εργαλεία όπως τα SHAP values, το LIME και τα counterfactual explanations επιτρέπουν στους ελεγκτές να κατανοούν ποιοι παράγοντες συνέβαλαν στην πρόβλεψη απάτης, προσφέροντας διαφάνεια και υποστήριξη στις διαδικασίες λήψης αποφάσεων (Molnar, 2022). Ειδικά σε περιβάλλοντα όπου εφαρμόζονται ρυθμιστικές απαιτήσεις, όπως το GDPR ή η οδηγία PSD2, η διαφάνεια στις αυτόματες αποφάσεις καθίσταται υποχρεωτική.

Ένα σημαντικό πεδίο ανάπτυξης είναι η συνδυαστική χρήση πολλαπλών μοντέλων (hybrid detection systems). Για παράδειγμα, ένα σύστημα μπορεί να χρησιμοποιεί unsupervised learning για την κατηγοριοποίηση της "φυσιολογικής" συμπεριφοράς των πελατών, και supervised learning για την ταξινόμηση των αποκλίσεων ως πιθανές απάτες. Αυτή η προσέγγιση μειώνει τα ψευδώς θετικά και βελτιώνει τη συνολική απόδοση του συστήματος (Phua et al., 2010).

Η αυξανόμενη αυτοματοποίηση της ανίχνευσης απάτης μέσω αλγοριθμικών συστημάτων έχει θέσει στο επίκεντρο τον προβληματισμό γύρω από την ηθική χρήση της Τεχνητής Νοημοσύνης και τον αλγοριθμικό μεροληπτικό χαρακτήρα (algorithmic bias). Οι αλγόριθμοι εκπαιδεύονται σε ιστορικά δεδομένα, τα οποία ενδέχεται να εμπεριέχουν κοινωνικές, γεωγραφικές ή φυλετικές προκαταλήψεις. Αυτό ενέχει τον κίνδυνο να αναπαραγάγουν ή και να ενισχύσουν υπάρχουσες ανισότητες, καταλήγοντας σε άδικες αποφάσεις σε βάρος συγκεκριμένων ομάδων πληθυσμού (Binns, 2018). Ένα κλασικό παράδειγμα είναι η υπερεκπροσώπηση κατοίκων συγκεκριμένων περιοχών σε κατηγορίες υψηλού κινδύνου, λόγω προκατειλημμένων ιστορικών δεδομένων.

Η αλγοριθμική δικαιοσύνη (algorithmic fairness) αποτελεί σήμερα ερευνητικό πεδίο υψηλής προτεραιότητας. Έχουν προταθεί τεχνικές όπως η κανονικοποίηση ευαίσθητων μεταβλητών (pre-processing), η τροποποίηση της λειτουργίας του αλγορίθμου (in-processing) και η

ρύθμιση των αποτελεσμάτων (post-processing), ώστε να διασφαλιστεί ότι τα μοντέλα δεν εισάγουν μεροληπτικές προβλέψεις (Kamiran & Calders, 2012). Παράλληλα, μηχανισμοί «auditability» επιτρέπουν τον εσωτερικό έλεγχο των μοντέλων, εξασφαλίζοντας μεγαλύτερη διαφάνεια.

Ένα ακόμη σημαντικό πεδίο εφαρμογής των συστημάτων ανίχνευσης απάτης είναι το ηλεκτρονικό εμπόριο (e-commerce). Με την αύξηση των online συναλλαγών και της εξ αποστάσεως πρόσβασης, οι επιτήδειοι βρίσκουν νέες μεθόδους εκμετάλλευσης των ψηφιακών αλυσίδων πληρωμών. Η χρήση real-time fraud detection συστημάτων είναι κρίσιμη για την έγκαιρη απόρριψη ύποπτων συναλλαγών χωρίς να διακόπτεται η εμπειρία του χρήστη. Ειδικά μοντέλα ML που βασίζονται στη συμπεριφορική βιομετρία (behavioral biometrics), στη διάρκεια πληκτρολόγησης, και στη γεωγραφική απόκλιση χρησιμοποιούνται όλο και συχνότερα (Sahin et al., 2022).

Παράλληλα, η ανίχνευση απάτης αποκτά σημασία και στο πεδίο της κυβερνοασφάλειας (cybersecurity). Η διασταύρωση μεταξύ ανίχνευσης παραβιάσεων ασφαλείας και χρηματοοικονομικής απάτης φέρνει στο προσκήνιο τεχνικές όπως τα Graph Neural Networks (GNNs), τα οποία είναι ικανά να αναλύουν τη δυναμική των συναλλαγών και να εντοπίζουν ύποπτα δίκτυα αλληλεπίδρασης (Zhang et al., 2019). Ειδικά σε περιπτώσεις οργανωμένων κυκλωμάτων απάτης, τα GNNs μπορούν να αναγνωρίσουν σχέσεις μεταξύ διαφορετικών υποκειμένων ή λογαριασμών που δεν είναι προφανείς στα κλασικά μοντέλα.

Η βιβλιογραφία συγκλίνει στο ότι η αποτελεσματική διαχείριση κινδύνου απάτης απαιτεί ένα ολιστικό πλαίσιο που περιλαμβάνει τεχνικά, κανονιστικά και ηθικά εργαλεία. Οι οργανισμοί που υιοθετούν συστήματα AI/ML οφείλουν να επενδύσουν σε συνεχές retraining των μοντέλων, σε παρακολούθηση από ανθρώπινους ελεγκτές και σε μηχανισμούς ελέγχου απόδοσης (performance monitoring). Προτείνεται επίσης η δημιουργία collaborative platforms μεταξύ τραπεζών, ασφαλιστικών και κρατικών φορέων για τη διαμοίραση γνώσης και την ενίσχυση των συλλογικών αμυντικών μηχανισμών.

Κλείνοντας, γίνεται φανερό ότι η ανίχνευση και διαχείριση απάτης μέσω AI/ML έχει πλέον ωριμάσει σε επιχειρησιακό επίπεδο, ωστόσο η συνεχιζόμενη έρευνα είναι απαραίτητη για την περαιτέρω ελαχιστοποίηση των κινδύνων, την προσαρμογή σε νέες μορφές απειλών και τη διασφάλιση της συμμόρφωσης με ρυθμιστικές και ηθικές απαιτήσεις. Η ενσωμάτωση των τεχνολογιών αυτών θα πρέπει να γίνεται στο πλαίσιο ενός συνεκτικού, ανθρωποκεντρικού και διαφανούς οικοσυστήματος.

2.7 Ανάλυση Συστημικού Κινδύνου με AI

Ο συστημικός κίνδυνος αναφέρεται στον κίνδυνο κατάρρευσης ενός ολόκληρου χρηματοπιστωτικού συστήματος ή αγοράς, λόγω της αποτυχίας ενός ή περισσότερων οργανισμών ή υποσυστημάτων, με αποτέλεσμα την εξάπλωση αρνητικών επιδράσεων σε αλληλεξαρτώμενα μέρη. Η κρίση του 2007–2008 ανέδειξε με εμφατικό τρόπο τον καταστροφικό ρόλο του συστημικού κινδύνου και οδήγησε στην εντατικοποίηση των ερευνητικών προσπαθειών για την κατανόηση, πρόβλεψη και διαχείρισή του (Acharya et al., 2010). Η αλληλεξάρτηση των χρηματοπιστωτικών ιδρυμάτων, μέσω σύνθετων δικτύων υποχρεώσεων και πιστώσεων, δημιουργεί δυναμικά φαινόμενα που δεν είναι εύκολα ανιχνεύσιμα με κλασικές στατιστικές μεθόδους (Bisias et al., 2012).

Η εφαρμογή τεχνικών Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) στην ανάλυση συστημικού κινδύνου προσφέρει νέα εργαλεία κατανόησης των πολύπλοκων σχέσεων μεταξύ των χρηματοπιστωτικών φορέων, αναδεικνύοντας την "κρυφή" ευθραυστότητα των αγορών. Ένας βασικός άξονας αυτής της προσέγγισης είναι η μοντελοποίηση χρηματοπιστωτικών δικτύων (financial networks), τα οποία αναπαριστούν τις σχέσεις μεταξύ τραπεζών, ασφαλιστικών εταιρειών και άλλων οικονομικών οργανισμών. Με τη χρήση τεχνικών ανάλυσης κοινωνικών δικτύων (network analysis), όπως οι δείκτες κεντρικότητας (centrality), η πυκνότητα (density) και οι δομές συσσωματώσεων (cliques), μπορεί να εντοπιστούν οι κόμβοι που έχουν κρίσιμο ρόλο στη μετάδοση του κινδύνου (Battiston et al., 2012).

Οι τεχνικές ML επιτρέπουν τη δυναμική παρακολούθηση της εξέλιξης των συστημάτων αυτών, μέσω της χρήσης ιστορικών δεδομένων συναλλαγών, ισολογισμών, τιμών αγοράς και ρευστότητας. Τα recurrent neural networks (RNNs) και ιδιαίτερα τα Long Short-Term Memory networks (LSTMs) έχουν χρησιμοποιηθεί για την πρόβλεψη των πιθανοτήτων μετάδοσης κρίσης σε ένα χρηματοπιστωτικό δίκτυο, βασισμένα σε χρονικές σειρές γεγονότων (Huang et al., 2022). Τα δίκτυα αυτά, με την ικανότητά τους να "θυμούνται" πληροφορία σε βάθος χρόνου, καθίστανται κατάλληλα για την παρακολούθηση σταδιακών μετατοπίσεων στη συμπεριφορά της αγοράς.

Επιπρόσθετα, τα Graph Neural Networks (GNNs) αποτελούν μία ταχέως εξελισσόμενη προσέγγιση στην κατανόηση των σχέσεων εντός χρηματοπιστωτικών δικτύων. Σε αντίθεση με τα παραδοσιακά νευρωνικά δίκτυα, τα GNNs μπορούν να μοντελοποιήσουν μη-ευκρινείς και πολύπλοκες συνδέσεις μεταξύ κόμβων (θεσμών), αξιοποιώντας ταυτόχρονα τις ιδιότητες των ίδιων των κόμβων και των ακμών (Zhang et al., 2022). Η προσέγγιση αυτή επιτρέπει την πρόβλεψη της επικινδυνότητας συγκεκριμένων οργανισμών, λαμβάνοντας υπόψη τόσο τις

εσωτερικές τους μεταβλητές (π.χ. δείκτες ρευστότητας) όσο και τις εξωτερικές διασυνδέσεις τους με άλλους φορείς.

Ένας άλλος τομέας στον οποίο εφαρμόζεται η AI είναι η αναγνώριση πρόδρομων δεικτών (early warning indicators). Χρησιμοποιώντας εποπτευόμενα μοντέλα μάθησης, οι ερευνητές επιχειρούν να δημιουργήσουν αλγόριθμους που εντοπίζουν μοτίβα προηγούμενα μεγάλης κρίσης, αξιολογώντας μεταβλητές όπως η μόχλευση, η μεταβλητότητα στις τιμές των assets και η συσχέτιση μεταξύ περιουσιακών στοιχείων (Silva et al., 2020). Η αποτελεσματικότητα τέτοιων συστημάτων κρίνεται από την ικανότητά τους να προειδοποιούν εγκαίρως για ενδεχόμενη αστάθεια, χωρίς να οδηγούν σε υπεραντιδράσεις ή ψευδείς συναγερμούς.

Η χρήση μη εποπτευόμενων τεχνικών μάθησης (unsupervised learning) στην ανάλυση συστημικού κινδύνου αφορά κυρίως την ταξινόμηση χρηματοπιστωτικών ιδρυμάτων ή περιόδων σε κατηγορίες βάσει συμπεριφοράς ή στατιστικών χαρακτηριστικών. Τεχνικές όπως η ανάλυση κύριων συνιστωσών (PCA) και τα μοντέλα συσταδοποίησης (clustering), όπως τα K-Means ή τα DBSCAN, έχουν χρησιμοποιηθεί για τον εντοπισμό υποκείμενων δομών στα δεδομένα αγοράς και για την ομαδοποίηση παρόμοιων προφίλ κινδύνου (Billio et al., 2012). Αυτό επιτρέπει την καλύτερη κατανόηση του τρόπου με τον οποίο μεταβάλλεται η αλληλεξάρτηση μεταξύ αγορών και οντοτήτων, ιδίως σε περιόδους κρίσης.

Παράλληλα, τα μοντέλα βαθιάς μάθησης (deep learning) προσφέρουν δυνατότητες για την επεξεργασία μεγάλων όγκων δεδομένων που προέρχονται από πολλαπλές πηγές: τιμές μετοχών, χρηματοοικονομικές καταστάσεις, ειδήσεις, tweets, νομισματικές ροές, ακόμη και γεωπολιτικά γεγονότα. Η ενσωμάτωση αυτών των δεδομένων σε πολυεπίπεδα νευρωνικά δίκτυα, όπως τα autoencoders, μπορεί να βοηθήσει στον εντοπισμό ανωμαλιών που υποδεικνύουν συγκέντρωση κινδύνου σε συγκεκριμένα υποσυστήματα του χρηματοπιστωτικού οικοσυστήματος (Tobback et al., 2017). Ιδιαίτερα ενδιαφέρον παρουσιάζει η χρήση deep learning στην κατασκευή συνθετικών δεικτών συστημικού κινδύνου, οι οποίοι βασίζονται όχι μόνο σε μετρήσιμες μεταβλητές, αλλά και σε δυναμικές τάσεις που αποτυπώνονται σε μη παραδοσιακές πηγές.

Ένα άλλο σημαντικό ερευνητικό εργαλείο είναι η μοντελοποίηση βασισμένη σε πράκτορες (Agent-Based Modeling – ABM), η οποία έχει ενισχυθεί τα τελευταία χρόνια με αλγορίθμους AI. Το ABM αναπαριστά το χρηματοπιστωτικό σύστημα ως ένα σύνολο αλληλεπιδρώντων παραγόντων (agents), όπως τράπεζες, ρυθμιστικές αρχές και επενδυτές, καθένας με δικούς του κανόνες και στρατηγικές. Η χρήση reinforcement learning για την εκπαίδευση αυτών των πρακτόρων επιτρέπει τη μελέτη σεναρίων κρίσης υπό διάφορες υποθέσεις, π.χ. μεταβολές επιτοκίων, ρευστότητας ή πανικού στις αγορές (Bookstaber et al., 2018).

Η προσέγγιση αυτή είναι χρήσιμη διότι επιτρέπει τη μελέτη μη γραμμικών και ανατροφοδοτούμενων σχέσεων (feedback loops), οι οποίες είναι κοινές σε καταρρεύσεις αγορών. Επιπλέον, με τον συνδυασμό ABM και AI, δημιουργείται η δυνατότητα "πειραματισμού" σε περιβάλλοντα όπου οι πραγματικές δοκιμές είναι αδύνατες ή απαγορευτικές. Οι ρυθμιστικές αρχές μπορούν να χρησιμοποιήσουν τέτοιες προσομοιώσεις για την αξιολόγηση της αποτελεσματικότητας πολιτικών παρέμβασης, όπως τα capital buffers ή οι μηχανισμοί bail-in.

Η συστημική σημασία των επιμέρους τραπεζών μπορεί να αποτυπωθεί και με μετρικές όπως το ΔeltaCoVaR, το SRISK, και το Marginal Expected Shortfall (MES), που πλέον μπορούν να υπολογίζονται δυναμικά με τη βοήθεια αλγορίθμων ML. Οι τεχνικές αυτές εστιάζουν στην πιθανή συνεισφορά κάθε οργανισμού στον συνολικό κίνδυνο του συστήματος (Adrian & Brunnermeier, 2016). Για παράδειγμα, αλγόριθμοι XGBoost και Random Forest έχουν χρησιμοποιηθεί για την πρόβλεψη αυτών των μεγεθών με μεγαλύτερη ακρίβεια από τις γραμμικές προσεγγίσεις, ειδικά όταν τα δεδομένα είναι υψηλής διάστασης ή μη κανονικά κατανεμημένα (Chen et al., 2022).

Παρά την αυξανόμενη αποτελεσματικότητα των τεχνικών AI στη συστημική ανάλυση, σημαντική πρόκληση παραμένει η ερμηνευσιμότητα (explainability) των αποτελεσμάτων, ειδικά σε μοντέλα "μαύρου κουτιού" όπως τα deep neural networks και τα graph-based συστήματα. Σε ρυθμιστικά περιβάλλοντα όπου οι αποφάσεις επηρεάζουν τη νομισματική πολιτική, τη σταθερότητα του τραπεζικού συστήματος ή τον καθορισμό απαιτήσεων κεφαλαιακής επάρκειας, η δυνατότητα εξήγησης των ευρημάτων είναι εξίσου σημαντική με την ακρίβεια (Doshi-Velez & Kim, 2017). Τεχνικές όπως τα SHAP και τα LIME χρησιμοποιούνται όλο και περισσότερο σε μελέτες συστημικού κινδύνου για την ερμηνεία μεταβλητών που επηρεάζουν τις πιθανότητες μετάδοσης ή αθέτησης.

Η εφαρμογή των παραπάνω τεχνικών δεν είναι απλώς θεωρητική αλλά έχει χρησιμοποιηθεί στην ανάλυση πραγματικών κρίσεων. Κατά την περίοδο της κρίσης COVID-19, η χρήση AI συνέβαλε στην ταχύτερη ανάλυση της μετάδοσης κινδύνου από εταιρείες υψηλής μόχλευσης, στην αποτίμηση της ευπάθειας τομέων όπως ο τουρισμός και οι αερομεταφορές, και στην εκτίμηση των αλυσιδωτών επιπτώσεων από κρατικές παρεμβάσεις στην αγορά (Demirgüç-Kunt et al., 2021). Παράλληλα, η διασύνδεση με real-time δεδομένα ενίσχυσε την ικανότητα έγκαιρης αντίδρασης των ρυθμιστικών φορέων.

Το κανονιστικό πλαίσιο (regulatory framework) για την ενσωμάτωση της AI στην πρόβλεψη συστημικού κινδύνου παραμένει κατακερματισμένο. Παρόλο που οργανισμοί όπως η BIS και η ΕΚΤ αναγνωρίζουν τη σημασία της τεχνητής νοημοσύνης στην παρακολούθηση της

σταθερότητας, δεν υπάρχει ακόμη ενιαίο θεσμικό πλαίσιο που να καθορίζει την υποχρεωτική χρήση, τις προδιαγραφές των μοντέλων, ή τα όρια ευθύνης σε περιπτώσεις αποτυχίας προβλέψεων (FSB, 2019). Η ανάγκη για διαφάνεια, επαναληψιμότητα και auditability των μοντέλων είναι εντονότερη από ποτέ.

Επιπλέον, τίθενται σημαντικά ηθικά ζητήματα: η χρήση μοντέλων AI για την εκτίμηση της συστημικής επικινδυνότητας ενδέχεται να επηρεάζει την πρόσβαση των τραπεζών σε ρευστότητα, την εμπιστοσύνη των επενδυτών και τη διαμόρφωση τιμών στην αγορά. Εάν τα μοντέλα δεν είναι δίκαια ή ελέγξιμα, μπορεί να δημιουργηθεί ένας φαύλος κύκλος όπου η «κακή φήμη» βάσει αλγοριθμικών προβλέψεων οδηγεί σε πραγματική αποσταθεροποίηση (Pasquale, 2015). Επομένως, η υιοθέτηση αυτών των εργαλείων πρέπει να συνοδεύεται από ισχυρούς μηχανισμούς εποπτείας, εκπαίδευσης και ελέγχου.

Η επιστημονική βιβλιογραφία καταλήγει στο συμπέρασμα ότι η AI μπορεί να συμβάλει αποφασιστικά στην πρώιμη ανίχνευση, πρόληψη και αποτροπή του συστημικού κινδύνου, υπό την προϋπόθεση της σωστής ενσωμάτωσης, της κανονιστικής καθοδήγησης και της ηθικής ευαισθησίας. Οι μελλοντικές έρευνες αναμένεται να επικεντρωθούν σε συστήματα που θα συνδυάζουν explainable AI, real-time big data analytics, multi-agent simulation και δυναμική προσαρμογή, ενισχύοντας τη δυνατότητα λήψης τεκμηριωμένων αποφάσεων για τη διαφύλαξη της χρηματοοικονομικής σταθερότητας.

2.8 AIκαιCyberRiskManagement

Η ραγδαία ψηφιοποίηση των επιχειρησιακών διαδικασιών και η εξάρτηση των οργανισμών από τεχνολογικά δίκτυα έχουν εντείνει την έκθεση σε κυβερνοαπειλές και κινδύνους που σχετίζονται με την ασφάλεια των πληροφοριακών συστημάτων. Οι επιθέσεις τύπου ransomware, οι διαρροές δεδομένων και οι κακόβουλες παρεμβάσεις σε κρίσιμες υποδομές συνιστούν πλέον καθημερινή πρόκληση για τις επιχειρήσεις και τους κυβερνητικούς φορείς (ENISA, 2022). Η παραδοσιακή προσέγγιση στην διαχείριση των κυβερνοκινδύνων βασίζεται σε στατικά συστήματα κανόνων και σε χειροκίνητη αξιολόγηση των απειλών, τα οποία είναι ανεπαρκή σε ένα περιβάλλον που εξελίσσεται δυναμικά.

Η ενσωμάτωση της Τεχνητής Νοημοσύνης (AI) και της Μηχανικής Μάθησης (ML) στη διαχείριση κυβερνοκινδύνων (CyberRiskManagement) έχει φέρει επανάσταση στην προληπτική και αντιδραστική ικανότητα των οργανισμών. Μέσω της AI, οι υπολογιστές είναι πλέον σε θέση να αναγνωρίζουν μοτίβα απειλών, να προβλέπουν επιθέσεις και να προσαρμόζουν την άμυνα των συστημάτων σε πραγματικό χρόνο, μειώνοντας τον χρόνο

απόκρισης και τις απώλειες (Buczak&Guven, 2016). Οι αλγόριθμοι ML, ειδικά τα supervisedlearning μοντέλα, εκπαιδεύονται σε ιστορικά δεδομένα κυβερνοεπιθέσεων και μπορούν να ταξινομήσουν με ακρίβεια νέες εισερχόμενες απειλές ως κακόβουλες ή μη.

Ένα από τα σημαντικότερα πεδία εφαρμογής είναι η ανίχνευση ανωμαλιών (anomalydetection). Οι τεχνικές αυτές εντοπίζουν αποκλίσεις από την "κανονική" συμπεριφορά χρηστών ή συστημάτων, οι οποίες ενδέχεται να υποδηλώνουν επίθεση. Χρησιμοποιούνται unsupervised αλγόριθμοι όπως IsolationForests, One-ClassSVMs, autoencoders και GaussianMixtureModels για την απομόνωση δυνητικά κακόβουλων ενεργειών (Sommer&Paxson, 2010). Η αποτελεσματικότητα αυτών των μοντέλων οφείλεται στην ικανότητά τους να ανιχνεύουν επιθέσεις μη γνωστές εκ των προτέρων (zero-dayattacks), οι οποίες δεν καλύπτονται από παραδοσιακές βάσεις δεδομένων υπογραφών.

Επιπλέον, τα νευρωνικά δίκτυα (NeuralNetworks) και ιδίως τα ConvolutionalNeuralNetworks (CNNs) και τα RecurrentNeuralNetworks (RNNs) έχουν χρησιμοποιηθεί για την κατηγοριοποίηση κακόβουλου λογισμικού, την ανάλυση ροών δικτύου και την ανίχνευση εισβολών (Kimetal., 2016). Τα CNNs, τα οποία έχουν ευρεία εφαρμογή στην αναγνώριση εικόνας, χρησιμοποιούνται για την αναπαράσταση των αρχείων ως "εικόνες byte" και την αναγνώριση μοτίβων επιθέσεων σε επίπεδο binary. Από την άλλη, τα RNNs, και ειδικά τα LSTMs, είναι κατάλληλα για την κατανόηση χρονικών ακολουθιών και χρησιμοποιούνται σε context-awareintrusiondetection συστήματα.

Η χρήση NaturalLanguageProcessing (NLP) σε περιβάλλον κυβερνοασφάλειας αποτελεί ένα νέο και αναδυόμενο πεδίο. Μέσω της ανάλυσης μη δομημένων δεδομένων, όπως τα logs, τα alerts και τα reports, τα NLP μοντέλα μπορούν να εξάγουν γνώση σχετικά με τρέχουσες επιθέσεις, να εντοπίσουν trends και να βελτιώσουν την κατανόηση των επιθέσεων από τους ανθρώπινους αναλυτές (Sabottkeetal., 2015). Μοντέλα όπως το BERT, το GPT και τα transformers χρησιμοποιούνται για την κατηγοριοποίηση και εξαγωγή οντοτήτων (NamedEntityRecognition) σε περιβάλλοντα κυβερνοασφάλειας.

Ένα από τα βασικά εμπόδια για την ευρεία υιοθέτηση της Τεχνητής Νοημοσύνης στην κυβερνοασφάλεια είναι η έλλειψη διαφάνειας (lack of explainability) στα μοντέλα που χρησιμοποιούνται. Τα περισσότερα συστήματα βασίζονται σε τεχνικές "μαύρου κουτιού" και καθίστανται δύσκολα ερμηνεύσιμα για τους ανθρώπινους χειριστές και τις ρυθμιστικές αρχές. Αυτό δημιουργεί ένα σοβαρό πρόβλημα λογοδοσίας σε περιπτώσεις λανθασμένης απόφασης, ιδιαίτερα όταν η AI χρησιμοποιείται για την αυτόματη απόρριψη ή αποδοχή μιας δραστηριότητας (Ghosh et al., 2022). Για την αντιμετώπιση του προβλήματος, έχουν αναπτυχθεί τεχνικές Explainable AI (XAI), όπως οι SHAP (SHapley Additive exPlanations),

LIME και attention-based μηχανισμοί, οι οποίες επιτρέπουν την ανάδειξη των κρίσιμων χαρακτηριστικών που οδηγούν σε μια απόφαση ανίχνευσης ή αποκλεισμού.

Η αυτοματοποιημένη ανταπόκριση σε απειλές (automated threat response) αποτελεί ένα ακόμα πεδίο στο οποίο η AI φέρνει επανάσταση. Ενσωματώνοντας ML αλγορίθμους σε Security Orchestration, Automation and Response (SOAR) πλατφόρμες, οι οργανισμοί μπορούν να εκτελούν προγραμματισμένες ενέργειες άμυνας (π.χ. απομόνωση δικτύου, ειδοποίηση, τερματισμός συνεδριών) μόλις ανιχνευθεί μια απειλή, χωρίς ανθρώπινη παρέμβαση. Αυτή η δυνατότητα μειώνει δραματικά τον χρόνο απόκρισης και περιορίζει τις πιθανές απώλειες (Shameli-Sendi et al., 2016).

Παράλληλα, η χρήση AI στη συλλογή και ανάλυση cyber threat intelligence (CTI) επιτρέπει στους οργανισμούς να αξιοποιούν πλήρως ανοικτές πηγές (open source intelligence - OSINT), βάσεις δεδομένων επιθέσεων, φόρουμ και darknet για την ανακάλυψη πιθανών νέων κινδύνων. Με τεχνικές NLP, topic modeling και named entity recognition, η AI μπορεί να εξάγει σημασιολογικά πρότυπα από αδόμητα δεδομένα, να εντοπίσει συσχετισμούς μεταξύ απειλών και να συμβάλει στην έγκαιρη πρόβλεψη κυβερνοεπιθέσεων (Zhao et al., 2020).

Επιπλέον, η βιβλιογραφία αναδεικνύει τη σημασία των graph-based μοντέλων και των graph neural networks (GNNs) στη μοντελοποίηση της αρχιτεκτονικής των επιθέσεων. Η αναπαράσταση του συστήματος ως γράφου (nodes = endpoints, edges = ροές επικοινωνίας) επιτρέπει την αναγνώριση μοτίβων lateral movement και την εντολή κακόβουλων μοτίβων μέσα σε πολύπλοκα δίκτυα (Wu et al., 2021). Τα GNNs μπορούν να αναγνωρίσουν συνδυασμούς γεγονότων που υποδηλώνουν μακροπρόθεσμη επιτήρηση ή προετοιμασία μεγάλης κλίμακας επίθεσης (APT – Advanced Persistent Threat).

Η ενσωμάτωση της AI στη διαχείριση κυβερνοκινδύνων εγείρει, ωστόσο, σοβαρά ηθικά και κανονιστικά ζητήματα. Η χρήση προσωπικών δεδομένων για εκπαίδευση των μοντέλων ενδέχεται να παραβιάζει κανονισμούς όπως ο GDPR. Επιπλέον, η αυτοματοποιημένη λήψη αποφάσεων, χωρίς ανθρώπινο έλεγχο, προκαλεί ανησυχία για την απώλεια ανθρώπινης κρίσης και την πιθανότητα διακρίσεων. Για παράδειγμα, ένα σύστημα μπορεί να θεωρήσει ύποπτη μια συμπεριφορά χρήστη λόγω ασυνήθιστης τοποθεσίας ή γλώσσας, χωρίς να λαμβάνει υπόψη πολιτισμικούς ή επαγγελματικούς παράγοντες (Veale & Edwards, 2018).

Η πρακτική εφαρμογή της Τεχνητής Νοημοσύνης στη διαχείριση κυβερνοκινδύνων αποτυπώνεται σε πλήθος μελετών περίπτωσης. Οργανισμοί όπως η IBM, η Palo Alto Networks και η Darktrace έχουν αναπτύξει συστήματα βασισμένα σε AI που λειτουργούν ως ψηφιακοί φρουροί, αναλύοντας δισεκατομμύρια δεδομένα την ημέρα και εντοπίζοντας επιθέσεις με ελάχιστη ανθρώπινη παρέμβαση. Για παράδειγμα, η πλατφόρμα Darktrace

χρησιμοποιεί unsupervised learning και μοντελοποιεί το “pattern of life” κάθε χρήστη και συστήματος, επιτρέποντας την ανίχνευση αποκλίσεων που δεν περιλαμβάνονται σε λίστες γνωστών απειλών (Darktrace, 2021).

Μία από τις πιο επιτυχημένες προσεγγίσεις ενσωμάτωσης AI και cyber defense αποτελεί η χρήση federated learning, όπου οι οργανισμοί εκπαιδεύουν τοπικά μοντέλα με τα δικά τους δεδομένα χωρίς να τα αποστέλλουν σε κεντρικό server, διατηρώντας την ιδιωτικότητα των πληροφοριών. Αυτή η τεχνική επιτρέπει τη συλλογική μάθηση μεταξύ οργανισμών, μειώνοντας τη διασπορά του κινδύνου και ενισχύοντας την ικανότητα πρόβλεψης σύνθετων επιθέσεων (Pokhrel & Choi, 2020).

Παρόλα αυτά, η υλοποίηση των τεχνολογιών AI στη διαχείριση κυβερνοκινδύνων δεν είναι απαλλαγμένη από προκλήσεις. Το κόστος εγκατάστασης και διατήρησης, η έλλειψη εξειδικευμένου προσωπικού, η ανάγκη συνεχούς επικαιροποίησης των μοντέλων και οι ψευδείς θετικές ενδείξεις εξακολουθούν να περιορίζουν την αποτελεσματικότητα (Ali et al., 2019). Επιπλέον, οι επιτιθέμενοι αναπτύσσουν και αυτοί συστήματα AI, τα οποία μπορούν να ξεγελούν τους αλγορίθμους άμυνας ή να αναλύουν τις αδυναμίες τους μέσω adversarial learning, εγκαινιάζοντας ένα νέο είδος «αλγοριθμικού πολέμου» (Brundage et al., 2018).

Η βιβλιογραφία συγκλίνει στο ότι η AI αποτελεί αναπόσπαστο μέρος του μέλλοντος της κυβερνοασφάλειας, αλλά η αποδοτική και υπεύθυνη χρήση της απαιτεί την ανάπτυξη υβριδικών προσεγγίσεων, όπου η ανθρώπινη εποπτεία και η τεχνητή νοημοσύνη συνεργάζονται. Οι μελλοντικές έρευνες θα πρέπει να επικεντρωθούν στη διαλειτουργικότητα μεταξύ διαφορετικών εργαλείων AI, στη δημιουργία standard διαδικασιών αξιολόγησης (AI audit), και στην ανάπτυξη διαφανών, επεξηγήσιμων και δίκαιων συστημάτων, ικανά να υποστηρίξουν τη στρατηγική κυβερνοανθεκτικότητα των οργανισμών.

2.9 Ρυθμιστικό Πλαίσιο και Ηθικά Ζητήματα

Η ραγδαία εξάπλωση των τεχνολογιών Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) έχει πυροδοτήσει έναν παγκόσμιο διάλογο σχετικά με την ανάγκη δημιουργίας κατάλληλων ρυθμιστικών πλαισίων και την αναγνώριση των ηθικών προκλήσεων που συνεπάγεται η χρήση τέτοιων συστημάτων. Καθώς οι αλγόριθμοι AI διεισδύουν σε κρίσιμους τομείς της κοινωνίας, όπως η υγειονομική περίθαλψη, τα χρηματοοικονομικά, η ασφάλεια και η δημόσια διοίκηση, προκύπτουν σημαντικά ερωτήματα για τη λογοδοσία, τη διαφάνεια, τα δικαιώματα των πολιτών και την ισότητα (Floridi et al., 2018).

Ένα από τα πλέον συζητημένα ρυθμιστικά κείμενα διεθνώς είναι ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR) της Ευρωπαϊκής Ένωσης, ο οποίος εφαρμόστηκε το 2018. Ο GDPR παρέχει ένα νομικό υπόβαθρο για την επεξεργασία προσωπικών δεδομένων και καθιερώνει το «δικαίωμα στην εξήγηση» σε περιπτώσεις αυτοματοποιημένης λήψης αποφάσεων (Goodman & Flaxman, 2017). Το άρθρο 22 του Κανονισμού ορίζει ότι κάθε πολίτης έχει το δικαίωμα να μην υπόκειται σε απόφαση που βασίζεται αποκλειστικά σε αυτοματοποιημένη επεξεργασία, εάν αυτή έχει έννομες συνέπειες. Αυτή η ρύθμιση επηρεάζει άμεσα τη χρήση αλγορίθμων σε συστήματα πιστοληπτικής αξιολόγησης, προσλήψεων, τιμολόγησης ασφαλίστρων ή καταναλωτικής στόχευσης.

Πέρα από τον GDPR, η Ευρωπαϊκή Ένωση θέσπισε το 2024 τον Κανονισμό για την Τεχνητή Νοημοσύνη (AI Act), ο οποίος εισήγαγε ένα ολοκληρωμένο πλαίσιο ρύθμισης βασισμένο στον κίνδυνο. Ο AI Act διαχωρίζει τις εφαρμογές AI σε τέσσερις κατηγορίες: απαγορευμένες, υψηλού κινδύνου, περιορισμένου κινδύνου και ελάχιστου κινδύνου. Οι εφαρμογές υψηλού κινδύνου, όπως η χρήση AI σε τραπεζικά ή ιατρικά συστήματα, υπόκεινται σε αυστηρές απαιτήσεις διαφάνειας, τεκμηρίωσης, αξιολόγησης συμμόρφωσης και ανθρώπινης εποπτείας. Η εφαρμογή του Κανονισμού πραγματοποιείται σταδιακά, με πλήρη επιχειρησιακή εφαρμογή να ολοκληρώνεται εντός του 2026.

Στον διεθνή χώρο, άλλες χώρες έχουν ακολουθήσει διαφορετικές προσεγγίσεις. Οι Ηνωμένες Πολιτείες έχουν επιλέξει ένα πιο αποκεντρωμένο μοντέλο ρύθμισης, βασισμένο σε υφιστάμενες νομοθεσίες και κατευθυντήριες γραμμές από οργανισμούς όπως η FTC, το NIST και το White House Office of Science and Technology Policy (Wagner et al., 2022). Το NIST έχει αναπτύξει ένα πλαίσιο για την «Αξιόπιστη Τεχνητή Νοημοσύνη» (Trustworthy AI), το οποίο δίνει έμφαση στη διαφάνεια, την ανθεκτικότητα, την προστασία της ιδιωτικότητας και την ελαχιστοποίηση της προκατάληψης. Ωστόσο, η απουσία ομοσπονδιακής νομοθεσίας για την AI στις ΗΠΑ δημιουργεί κενά ρυθμιστικής κάλυψης και επιτρέπει την ευρεία χρήση μη επεξηγήσιμων ή αδιαφανών μοντέλων σε κρίσιμους τομείς.

Πέρα από το νομικό επίπεδο, η ηθική διάσταση της Τεχνητής Νοημοσύνης έχει αναδειχθεί ως μείζον πεδίο προβληματισμού. Η χρήση αλγοριθμικών συστημάτων για τη λήψη αποφάσεων που επηρεάζουν ανθρώπινες ζωές εγείρει ζητήματα όπως η δικαιοσύνη (fairness), η αδιαφάνεια (opacity), η μεροληψία (bias) και η λογοδοσία (accountability) (Mittelstadt et al., 2016). Τα συστήματα ML βασίζονται σε ιστορικά δεδομένα, τα οποία είναι συχνά επηρεασμένα από κοινωνικές ανισότητες. Ως αποτέλεσμα, αλγόριθμοι μπορεί να αναπαράγουν ή να ενισχύσουν τις διακρίσεις, όπως για παράδειγμα στις προσλήψεις, στις αξιολογήσεις φοιτητών ή στη χορήγηση δανείων (Barocas & Selbst, 2016).

Ο όρος “algorithmic bias” έχει καταστεί κεντρικός στη βιβλιογραφία, καθώς ερευνητές έχουν εντοπίσει πολλαπλούς τρόπους μέσω των οποίων η μεροληψία εισέρχεται στο σύστημα: στη συλλογή και επιλογή δεδομένων (data bias), στον σχεδιασμό του αλγορίθμου (design bias), και στην ερμηνεία των αποτελεσμάτων (interpretation bias). Για την αντιμετώπιση αυτών των προβλημάτων, έχουν αναπτυχθεί μεθοδολογίες όπως το “pre-processing” (π.χ. απομάκρυνση ευαίσθητων μεταβλητών), το “in-processing” (π.χ. fairness constraints) και το “post-processing” (π.χ. calibration of output probabilities) (Kamiran & Calders, 2012).

Η αλγοριθμική λογοδοσία (algorithmic accountability) αναφέρεται στην απαίτηση τα μοντέλα να είναι όχι μόνο ακριβή, αλλά και επεξηγήσιμα, επαληθεύσιμα και υποκείμενα σε έλεγχο. Η απουσία επεξηγησιμότητας δημιουργεί αυτό που αποκαλείται “black box society” – μια κοινωνία όπου σημαντικές αποφάσεις λαμβάνονται από αλγορίθμους που οι άνθρωποι δεν κατανοούν (Pasquale, 2015). Η ερμηνευσιμότητα γίνεται κρίσιμη σε τομείς όπως η υγεία και η δικαιοσύνη, όπου η κατανόηση της λογικής του μοντέλου είναι απαραίτητη όχι μόνο για την εμπιστοσύνη, αλλά και για την ηθική νομιμοποίηση των αποτελεσμάτων.

Σε παγκόσμιο επίπεδο, οργανισμοί όπως η UNESCO, ο ΟΟΣΑ, η IEEE και η Ευρωπαϊκή Επιτροπή έχουν προτείνει κατευθυντήριες αρχές δεοντολογίας για την AI. Οι περισσότερες προσεγγίσεις βασίζονται σε πέντε θεμελιώδεις αρχές: σεβασμός στα ανθρώπινα δικαιώματα, διαφάνεια, ασφάλεια, λογοδοσία και κοινωνική και περιβαλλοντική ευθύνη (Jobin et al., 2019). Η Διακήρυξη της UNESCO για την Ηθική της AI του 2021 αποτελεί μια από τις πιο φιλόδοξες παγκόσμιες προσπάθειες εναρμόνισης, επιδιώκοντας να λειτουργήσει ως σημείο αναφοράς για τις εθνικές πολιτικές.

Η αυξανόμενη πολυπλοκότητα των συστημάτων AI έχει οδηγήσει στην ανάγκη για ανάπτυξη μηχανισμών ελέγχου και πιστοποίησης (AI auditing). Το AI auditing αναφέρεται στη διαδικασία ανεξάρτητης αξιολόγησης ενός αλγοριθμικού συστήματος, με σκοπό τη διασφάλιση της συμμόρφωσης με δεοντολογικά, νομικά και τεχνικά πρότυπα (Raji et al., 2020). Η βιβλιογραφία επισημαίνει ότι η έλλειψη ενιαίων πρακτικών και η απουσία θεσμοθετημένων οργανισμών πιστοποίησης καθιστούν το πεδίο του AI auditing κατακερματισμένο. Ωστόσο, υπάρχει αυξανόμενο ενδιαφέρον για την ανάπτυξη κοινών προτύπων και μεθοδολογιών, όπως η προσέγγιση ALTAI (Assessment List for Trustworthy AI) της Ε.Ε., που λειτουργεί ως εργαλείο αυτοαξιολόγησης για επιχειρήσεις.

Σε αυτό το πλαίσιο, τα διεθνή πρότυπα ISO/IEC αποκτούν όλο και μεγαλύτερη σημασία. Τα πρότυπα ISO/IEC 23053 για τη διαφάνεια των αλγορίθμων, ISO/IEC 42001 για τα Συστήματα Διαχείρισης AI και ISO/IEC TR 24028 για την ερμηνευσιμότητα, διαμορφώνουν ένα θεμέλιο για την ανάπτυξη αξιόπιστων και υπεύθυνων συστημάτων AI (ISO, 2023). Η

αποδοχή αυτών των προτύπων από τη βιομηχανία αναμένεται να ενισχύσει τη λογοδοσία και να προσφέρει ρυθμιστική σαφήνεια σε ένα πεδίο όπου οι τεχνολογικές και νομικές εξελίξεις συντελούνται με ταχύ ρυθμό.

Για τις επιχειρήσεις, η συμμόρφωση με τις δεοντολογικές και ρυθμιστικές απαιτήσεις δημιουργεί προκλήσεις αλλά και ευκαιρίες. Από τη μία πλευρά, απαιτείται σημαντική επένδυση σε ανθρώπινο δυναμικό, εσωτερικούς ελέγχους, τεκμηρίωση και τεχνολογικά μέσα για την ενσωμάτωση αρχών όπως η explainability και η data minimization. Από την άλλη, η λογοδοσία και η ηθική χρήση της AI αποτελούν ανταγωνιστικό πλεονέκτημα, αυξάνοντας την εμπιστοσύνη πελατών, επενδυτών και ρυθμιστικών αρχών (Zeng et al., 2021). Επιχειρήσεις που επενδύουν εγκαίρως στην ανάπτυξη υπεύθυνων συστημάτων είναι καλύτερα προετοιμασμένες να ανταποκριθούν στις μελλοντικές απαιτήσεις της αγοράς και της κοινωνίας.

Η μελλοντική πορεία του ρυθμιστικού πλαισίου για την AI αναμένεται να κινηθεί σε τρεις βασικούς άξονες: πρώτον, στην εναρμόνιση των εθνικών πλαισίων σε διεθνές επίπεδο, ώστε να αποφευχθεί ο κατακερματισμός και η "ρυθμιστική ασυμμετρία" μεταξύ αγορών· δεύτερον, στη μετατόπιση της ευθύνης από τους χρήστες στους δημιουργούς των μοντέλων, ενισχύοντας την ευθύνη των developers και των παρόχων AI· και τρίτον, στη συστηματική ενσωμάτωση των πολιτών στη διαδικασία αξιολόγησης της τεχνολογίας, μέσω participatory approaches, human-in-the-loop και mechanisms για ενίσχυση της διαφανούς διακυβέρνησης της τεχνολογίας (Morley et al., 2020).

Συνοψίζοντας, η σύγχρονη βιβλιογραφία καταδεικνύει την ανάγκη για ένα πολυεπίπεδο πλαίσιο δεοντολογίας και ρύθμισης της Τεχνητής Νοημοσύνης που θα ισορροπεί ανάμεσα στην καινοτομία και στην προστασία των θεμελιωδών δικαιωμάτων. Η δημιουργία υπεύθυνων, αξιόπιστων και επεξηγήσιμων συστημάτων AI αποτελεί πλέον όχι μόνο ηθική επιταγή αλλά και στρατηγική προτεραιότητα για οργανισμούς, κυβερνήσεις και κοινωνίες.

2.10 Θεωρίες Αποδοχής Τεχνολογίας

Η κατανόηση των παραγόντων που επηρεάζουν την πρόθεση χρήσης (behavioral intention) μιας τεχνολογίας αποτελεί διαχρονικά αντικείμενο μελέτης στη διοικητική επιστήμη, στα πληροφοριακά συστήματα και στη συμπεριφορική οικονομική. Παρότι οι τεχνολογίες Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) χαρακτηρίζονται από υψηλή τεχνική πολυπλοκότητα και εξειδίκευση, η υιοθέτησή τους εντός οργανισμών δεν εξαρτάται αποκλειστικά από την αντικειμενική τους απόδοση, αλλά και από τις αντιλήψεις, τις στάσεις

και τις πεποιθήσεις των χρηστών και των στελεχών που καλούνται να τις ενσωματώσουν στις διαδικασίες λήψης αποφάσεων. Για τον λόγο αυτό, η διερεύνηση της πρόθεσης υιοθέτησης τεχνολογιών AI στη διαχείριση κινδύνων απαιτεί θεωρητική θεμελίωση στις καθιερωμένες θεωρίες αποδοχής τεχνολογίας.

Μία από τις πλέον επιδραστικές θεωρίες στον τομέα αυτό είναι το Μοντέλο Αποδοχής Τεχνολογίας (Technology Acceptance Model – TAM), το οποίο προτάθηκε από τον Davis το 1989. Το TAM αναπτύχθηκε με σκοπό να εξηγήσει και να προβλέψει τη χρήση πληροφοριακών συστημάτων σε οργανωσιακά περιβάλλοντα και βασίζεται στην παραδοχή ότι η πραγματική χρήση μιας τεχνολογίας προηγείται από τη διαμόρφωση πρόθεσης χρήσης, η οποία με τη σειρά της επηρεάζεται από συγκεκριμένες γνωστικές πεποιθήσεις. Οι δύο κεντρικές μεταβλητές του μοντέλου είναι η αντιληπτή χρησιμότητα (perceived usefulness) και η αντιληπτή ευκολία χρήσης (perceived ease of use). Η αντιληπτή χρησιμότητα ορίζεται ως ο βαθμός στον οποίο ένα άτομο πιστεύει ότι η χρήση μιας τεχνολογίας θα βελτιώσει την επαγγελματική του απόδοση, ενώ η αντιληπτή ευκολία χρήσης αναφέρεται στον βαθμό στον οποίο το άτομο θεωρεί ότι η χρήση της τεχνολογίας θα είναι απαλλαγμένη από προσπάθεια.

Σύμφωνα με το TAM, η αντιληπτή χρησιμότητα ασκεί άμεση και θετική επίδραση στην πρόθεση χρήσης, ενώ η αντιληπτή ευκολία χρήσης επηρεάζει τόσο άμεσα την πρόθεση όσο και έμμεσα μέσω της επίδρασής της στη χρησιμότητα. Η θεωρητική αυτή προσέγγιση έχει επαληθευθεί εμπειρικά σε πλήθος μελετών που αφορούν διαφορετικά πληροφοριακά συστήματα, από εταιρικά λογισμικά μέχρι εφαρμογές ηλεκτρονικής τραπεζικής και ψηφιακές πλατφόρμες. Η διαχρονική ισχύς του μοντέλου οφείλεται στην απλότητα και στη σαφήνεια των εννοιών του, καθώς και στην ικανότητά του να ερμηνεύει συμπεριφορές σε ποικίλα τεχνολογικά περιβάλλοντα.

Η εφαρμογή του TAM στο πλαίσιο της Τεχνητής Νοημοσύνης αποκτά ιδιαίτερο ενδιαφέρον, καθώς οι τεχνολογίες AI συχνά ενσωματώνονται σε διαδικασίες υψηλής σημασίας, όπως η πιστωτική αξιολόγηση, η ανίχνευση απάτης και η διαχείριση χαρτοφυλακίων. Σε αυτά τα περιβάλλοντα, η αντιληπτή χρησιμότητα σχετίζεται με την ακρίβεια των προβλέψεων, τη βελτίωση της διαχείρισης κινδύνου, τη μείωση του λειτουργικού κόστους και τη συμμόρφωση με κανονιστικές απαιτήσεις. Εάν τα στελέχη αντιλαμβάνονται ότι η AI συμβάλλει ουσιαστικά στη βελτίωση των αποτελεσμάτων και στη μείωση των σφαλμάτων, είναι πιθανότερο να αναπτύξουν θετική πρόθεση υιοθέτησης. Παράλληλα, εάν τα συστήματα AI θεωρούνται υπερβολικά περίπλοκα ή δύσχρηστα, η πρόθεση χρήσης ενδέχεται να μειωθεί, ακόμη και αν η τεχνολογία είναι αντικειμενικά αποδοτική.

Πέραν του αρχικού TAM, η βιβλιογραφία έχει προτείνει επεκτάσεις και ενοποιημένα μοντέλα που επιχειρούν να ενσωματώσουν πρόσθετους παράγοντες. Ιδιαίτερα σημαντική είναι η Ενοποιημένη Θεωρία Αποδοχής και Χρήσης Τεχνολογίας (Unified Theory of Acceptance and Use of Technology – UTAUT), η οποία αναπτύχθηκε από τους Venkatesh και συνεργάτες το 2003. Το UTAUT συνδυάζει στοιχεία από οκτώ προϋπάρχουσες θεωρίες και προτείνει ότι η πρόθεση χρήσης επηρεάζεται κυρίως από την προσδοκία απόδοσης (performance expectancy), την προσδοκία προσπάθειας (effort expectancy), την κοινωνική επιρροή (social influence) και τις διευκολυντικές συνθήκες (facilitating conditions). Η προσδοκία απόδοσης αντιστοιχεί σε μεγάλο βαθμό στην αντιληπτή χρησιμότητα του TAM, ενώ η προσδοκία προσπάθειας σχετίζεται με την ευκολία χρήσης.

Στο πλαίσιο της υιοθέτησης AI σε οργανισμούς, οι έννοιες αυτές αποκτούν ιδιαίτερη σημασία. Η προσδοκία απόδοσης μπορεί να συνδέεται με την πεποίθηση ότι η AI θα βελτιώσει τη λήψη αποφάσεων και θα μειώσει την έκθεση σε κινδύνους, ενώ η προσδοκία προσπάθειας αφορά την εκτίμηση του βαθμού δυσκολίας εκπαίδευσης, ενσωμάτωσης και λειτουργίας των συστημάτων. Η κοινωνική επιρροή ενδέχεται να σχετίζεται με τη στρατηγική κατεύθυνση του οργανισμού ή με την πίεση από ανώτερα διοικητικά στελέχη και ρυθμιστικές αρχές για υιοθέτηση καινοτόμων τεχνολογιών. Οι διευκολυντικές συνθήκες αναφέρονται στην ύπαρξη τεχνικής υποδομής, εξειδικευμένου προσωπικού και υποστήριξης που καθιστούν εφικτή τη χρήση της τεχνολογίας.

Ωστόσο, σε περιβάλλοντα υψηλού ρίσκου, όπως ο χρηματοπιστωτικός τομέας, οι γνωστικές μεταβλητές του TAM και του UTAUT δεν επαρκούν από μόνες τους για να εξηγήσουν πλήρως την πρόθεση χρήσης. Η βιβλιογραφία έχει αναδείξει την εμπιστοσύνη (trust) ως καθοριστικό παράγοντα υιοθέτησης τεχνολογιών που χαρακτηρίζονται από αυτοματοποίηση και αλγοριθμική λήψη αποφάσεων. Η εμπιστοσύνη σε ένα σύστημα AI αναφέρεται στην πεποίθηση ότι το σύστημα θα λειτουργήσει με αξιοπιστία, συνέπεια και χωρίς να προκαλέσει απρόβλεπτες ή επιβλαβείς συνέπειες. Σε οργανισμούς όπου οι αποφάσεις που λαμβάνονται επηρεάζουν σημαντικά οικονομικά μεγέθη ή τη φήμη της επιχείρησης, η έλλειψη εμπιστοσύνης μπορεί να αποτελέσει ισχυρό ανασταλτικό παράγοντα, ακόμη και εάν η τεχνολογία παρουσιάζει υψηλή ακρίβεια.

Η εμπιστοσύνη συνδέεται στενά με την έννοια της ερμηνευσιμότητας και της διαφάνειας των αλγοριθμικών μοντέλων. Όσο περισσότερο οι χρήστες κατανοούν τον τρόπο με τον οποίο ένα σύστημα παράγει τα αποτελέσματά του, τόσο αυξάνεται η πιθανότητα να το εμπιστευθούν. Αντιθέτως, τα λεγόμενα «μαύρα κουτιά» ενδέχεται να δημιουργούν αβεβαιότητα και

επιφυλακτικότητα. Επομένως, η εμπιστοσύνη λειτουργεί ως ψυχολογικός μηχανισμός που γεφυρώνει το χάσμα μεταξύ τεχνικής απόδοσης και οργανωσιακής αποδοχής.

Παράλληλα, η θεωρία του αντιληπτού κινδύνου (perceived risk theory) προσφέρει ένα επιπλέον ερμηνευτικό πλαίσιο για την κατανόηση της πρόθεσης χρήσης τεχνολογιών AI. Ο αντιληπτός κίνδυνος αναφέρεται στην υποκειμενική εκτίμηση των πιθανών αρνητικών συνεπειών που μπορεί να προκύψουν από τη χρήση μιας τεχνολογίας. Στο πλαίσιο της AI στη διαχείριση κινδύνων, ο αντιληπτός κίνδυνος μπορεί να αφορά τον κίνδυνο λανθασμένων προβλέψεων, νομικών κυρώσεων λόγω μη συμμόρφωσης, απώλειας ελέγχου των διαδικασιών ή ακόμη και φήμης του οργανισμού. Η βιβλιογραφία στον τομέα του ηλεκτρονικού εμπορίου και της ηλεκτρονικής τραπεζικής έχει δείξει ότι όσο αυξάνεται ο αντιληπτός κίνδυνος, τόσο μειώνεται η πρόθεση χρήσης, ιδίως όταν οι χρήστες δεν αισθάνονται επαρκώς προστατευμένοι ή ενημερωμένοι.

Συνολικά, οι θεωρίες αποδοχής τεχνολογίας καταδεικνύουν ότι η πρόθεση υιοθέτησης μιας τεχνολογίας αποτελεί το αποτέλεσμα σύνθετης αλληλεπίδρασης γνωστικών και συναισθηματικών παραγόντων. Στην περίπτωση της Τεχνητής Νοημοσύνης, οι παράγοντες αυτοί περιλαμβάνουν την αντιληπτή χρησιμότητα, την ευκολία χρήσης, την εμπιστοσύνη και τον αντιληπτό κίνδυνο, οι οποίοι αλληλεπιδρούν και διαμορφώνουν τη συνολική στάση των στελεχών απέναντι στην τεχνολογία. Η ενσωμάτωση αυτών των θεωρητικών προσεγγίσεων στο παρόν ερευνητικό πλαίσιο επιτρέπει τη συστηματική διερεύνηση των μεταβλητών που επηρεάζουν την πρόθεση υιοθέτησης AI στη διαχείριση χρηματοοικονομικών κινδύνων, γεφυρώνοντας το χάσμα μεταξύ τεχνολογικής και οργανωσιακής διάστασης της καινοτομίας.

2.11 Παράγοντες που επηρεάζουν την πρόθεση υιοθέτησης AI

Η πρόθεση υιοθέτησης Τεχνητής Νοημοσύνης (AI) σε οργανωσιακά περιβάλλοντα, και ειδικότερα στον χρηματοπιστωτικό τομέα, δεν αποτελεί απλώς τεχνολογική επιλογή αλλά σύνθετη συμπεριφορική διαδικασία. Όπως αναδείχθηκε στην προηγούμενη ενότητα, οι θεωρίες αποδοχής τεχνολογίας υποστηρίζουν ότι η πραγματική χρήση ενός συστήματος προηγείται από τη διαμόρφωση πρόθεσης χρήσης, η οποία με τη σειρά της επηρεάζεται από συγκεκριμένες γνωστικές και αντιληπτικές μεταβλητές. Στο πλαίσιο της παρούσας μελέτης, η πρόθεση υιοθέτησης AI στη διαχείριση χρηματοοικονομικών κινδύνων ερμηνεύεται ως το αποτέλεσμα της αλληλεπίδρασης τεσσάρων βασικών παραγόντων: της αντιληπτής χρησιμότητας, της αντιληπτής ευκολίας χρήσης, της εμπιστοσύνης και του αντιληπτού κινδύνου.

Η αντιληπτή χρησιμότητα (perceivedusefulness) αποτελεί κεντρική μεταβλητή στις θεωρίες αποδοχής τεχνολογίας και ορίζεται ως ο βαθμός στον οποίο ένα άτομο πιστεύει ότι η χρήση μιας τεχνολογίας θα βελτιώσει την απόδοσή του (Davis, 1989). Στο περιβάλλον της διαχείρισης κινδύνου, η χρησιμότητα της AI συνδέεται με την ικανότητά της να επεξεργάζεται μεγάλους όγκους δεδομένων, να εντοπίζει μη γραμμικά πρότυπα και να βελτιώνει την ακρίβεια πρόβλεψης σε τομείς όπως η πιστωτική αξιολόγηση και η ανίχνευση απάτης (Lessmannetal., 2015· Bahnsenetal., 2016). Επιπλέον, οι τεχνικές μηχανικής μάθησης έχουν αποδειχθεί αποτελεσματικές στη διαχείριση χαρτοφυλακίων και στη βελτίωση δεικτών απόδοσης προσαρμοσμένων στον κίνδυνο, όπως το Sharperatio (Moody&Saffell, 2001). Όταν τα στελέχη αντιλαμβάνονται ότι η AI μπορεί να συμβάλει ουσιαστικά στη μείωση σφαλμάτων, στην έγκαιρη ανίχνευση κινδύνων και στη βελτίωση της συνολικής απόδοσης, ενισχύεται η πρόθεση υιοθέτησής της. Η βιβλιογραφία στον χώρο των πληροφοριακών συστημάτων έχει επανειλημμένα επιβεβαιώσει ότι η αντιληπτή χρησιμότητα αποτελεί τον ισχυρότερο προγνωστικό παράγοντα της πρόθεσης χρήσης (Venkateshetal., 2003), γεγονός που υποδηλώνει ότι η αντίληψη περί προστιθέμενης αξίας της AI είναι καθοριστική για την οργανωσιακή της αποδοχή.

Η αντιληπτή ευκολία χρήσης (perceivedeaseofuse) συνιστά επίσης κρίσιμο παράγοντα. Ορίζεται ως ο βαθμός στον οποίο η χρήση μιας τεχνολογίας θεωρείται απαλλαγμένη από προσπάθεια (Davis, 1989). Στο πλαίσιο των συστημάτων AI, η πολυπλοκότητα των αλγορίθμων, η ανάγκη εξειδικευμένων γνώσεων και η τεχνική ορολογία ενδέχεται να δημιουργούν αίσθημα δυσκολίας και αποστασιοποίησης από τα στελέχη που δεν διαθέτουν τεχνικό υπόβαθρο. Ωστόσο, η ανάπτυξη φιλικών διεπαφών, εργαλείων οπτικοποίησης και explainableAI τεχνικών, όπως τα SHAP και LIME (BarredoArrietaetal., 2020), συμβάλλει στη μείωση της αντιληπτής πολυπλοκότητας. Σύμφωνα με το TAM, η ευκολία χρήσης επηρεάζει άμεσα την πρόθεση χρήσης αλλά και έμμεσα μέσω της επίδρασής της στην αντιληπτή χρησιμότητα. Σε οργανισμούς όπου τα στελέχη αισθάνονται ότι μπορούν να κατανοήσουν και να ελέγξουν τη λειτουργία ενός συστήματος AI χωρίς υπερβολική τεχνική επιβάρυνση, η πρόθεση υιοθέτησης ενισχύεται. Αντιθέτως, η αίσθηση τεχνικής αδιαφάνειας μπορεί να λειτουργήσει αποτρεπτικά, ιδίως σε περιβάλλοντα αυστηρής κανονιστικής εποπτείας.

Η εμπιστοσύνη (trust) αναδεικνύεται ως ιδιαίτερα σημαντικός παράγοντας στην υιοθέτηση AI σε περιβάλλοντα υψηλού κινδύνου. Η βιβλιογραφία για την ερμηνευσιμότητα και την αλγοριθμική λογοδοσία επισημαίνει ότι τα «μαύρα κουτιά» δημιουργούν ανησυχίες σχετικά με την αξιοπιστία και τη διαφάνεια των αποφάσεων (Doshi-Velez&Kim, 2017). Σε εφαρμογές

όπως η πιστωτική αξιολόγηση, όπου οι αποφάσεις έχουν άμεσες οικονομικές και κοινωνικές συνέπειες, η έλλειψη εμπιστοσύνης μπορεί να οδηγήσει σε επιφυλακτικότητα ή απόρριψη της τεχνολογίας, ακόμη και όταν τα εμπειρικά αποτελέσματα είναι θετικά. Η ανάγκη για explainableAI και κανονιστική συμμόρφωση, όπως υπογραμμίζεται από τον GDPR και τις σχετικές αναλύσεις (Goodman&Flaxman, 2017· BarredoArrietaetal., 2020), καταδεικνύει ότι η εμπιστοσύνη δεν αποτελεί απλώς ψυχολογική μεταβλητή αλλά και θεσμική απαίτηση. Όσο μεγαλύτερη είναι η πεποίθηση ότι το σύστημα λειτουργεί με συνέπεια, δικαιοσύνη και χωρίς μεροληψία, τόσο αυξάνεται η πρόθεση υιοθέτησης. Αντιθέτως, οι ανησυχίες περί αλγοριθμικής προκατάληψης και κοινωνικών διακρίσεων (Binns, 2018) ενδέχεται να περιορίσουν την αποδοχή της AI, ιδίως σε οργανισμούς που επιδιώκουν συμμόρφωση με ηθικά και ρυθμιστικά πρότυπα.

Στενά συνδεδεμένος με την εμπιστοσύνη είναι ο αντιληπτός κίνδυνος (perceivedrisk), ο οποίος αναφέρεται στην υποκειμενική εκτίμηση πιθανών αρνητικών συνεπειών από τη χρήση της τεχνολογίας. Στον χρηματοπιστωτικό τομέα, οι κίνδυνοι αυτοί μπορεί να περιλαμβάνουν λανθασμένες προβλέψεις, κανονιστικές κυρώσεις, παραβιάσεις δεδομένων ή ζημίες στη φήμη του οργανισμού. Η εμπειρία της χρηματοπιστωτικής κρίσης και η ευαισθησία σε συστημικούς κινδύνους (Acharyaetal., 2010· Bisiasetal., 2012) έχουν ενισχύσει τη σημασία της προσεκτικής αξιολόγησης νέων τεχνολογιών. Εάν η AI θεωρηθεί ότι ενέχει υψηλό λειτουργικό ή νομικό κίνδυνο, η πρόθεση υιοθέτησης ενδέχεται να μειωθεί, ακόμη και αν η αντιληπτή χρησιμότητα είναι υψηλή. Η βιβλιογραφία στην ανίχνευση απάτης έχει δείξει ότι η ανισορροπία δεδομένων και τα ψευδώς θετικά αποτελέσματα μπορούν να οδηγήσουν σε σημαντικές επιχειρησιακές συνέπειες (DalPozzoloetal., 2015· Jurgovskyetal., 2018), γεγονός που ενισχύει την ανάγκη για προσεκτική αξιολόγηση του ρίσκου πριν από την υιοθέτηση.

Επιπλέον, το κανονιστικό πλαίσιο επηρεάζει έμμεσα την πρόθεση υιοθέτησης μέσω της διαμόρφωσης αντιλήψεων περί ασφάλειας και νομιμότητας. Η ύπαρξη ρυθμίσεων όπως ο GDPR και η απαίτηση διαφάνειας σε αυτοματοποιημένες αποφάσεις (Goodman&Flaxman, 2017) δημιουργεί περιβάλλον στο οποίο οι οργανισμοί οφείλουν να διασφαλίζουν τη συμμόρφωση πριν προχωρήσουν σε ευρεία εφαρμογή AI. Η έλλειψη σαφούς κανονιστικής καθοδήγησης ή η αβεβαιότητα ως προς τις ευθύνες μπορεί να ενισχύσει τον αντιληπτό κίνδυνο και να περιορίσει την πρόθεση υιοθέτησης.

Η αλληλεπίδραση μεταξύ των παραγόντων αυτών είναι κρίσιμη. Η υψηλή αντιληπτή χρησιμότητα μπορεί να αντισταθμίσει εν μέρει τον αντιληπτό κίνδυνο, ενώ η ισχυρή εμπιστοσύνη μπορεί να ενισχύσει την επίδραση της χρησιμότητας στην πρόθεση χρήσης. Αντίστοιχα, η χαμηλή ευκολία χρήσης μπορεί να υπονομεύσει τα οφέλη που αποδίδονται στη

χρησιμότητα, ιδίως όταν η τεχνολογία απαιτεί εκτενή εκπαίδευση ή σημαντικές οργανωσιακές αλλαγές. Η βιβλιογραφία για την οργανωσιακή καινοτομία υπογραμμίζει ότι η υιοθέτηση νέων τεχνολογιών δεν είναι γραμμική διαδικασία αλλά αποτέλεσμα ισορροπίας μεταξύ αντιληπτών οφελών και κινδύνων.

Συνεπώς, η πρόθεση υιοθέτησης AI στη διαχείριση χρηματοοικονομικών κινδύνων διαμορφώνεται μέσα από ένα πλέγμα αντιλήψεων που συνδυάζουν τεχνικά, οργανωσιακά και θεσμικά στοιχεία. Η αντιληπτή χρησιμότητα αντανακλά την εκτίμηση περί βελτίωσης απόδοσης, η ευκολία χρήσης αφορά τη διαχειρισιμότητα της τεχνολογίας, η εμπιστοσύνη σχετίζεται με την αξιοπιστία και τη διαφάνεια, ενώ ο αντιληπτός κίνδυνος εκφράζει την ανησυχία για πιθανές αρνητικές συνέπειες. Η κατανόηση των παραγόντων αυτών επιτρέπει τη διαμόρφωση ενός συνεκτικού ερευνητικού πλαισίου, το οποίο ερμηνεύει την υιοθέτηση της AI όχι μόνο ως τεχνική επιλογή αλλά ως σύνθετη συμπεριφορική απόφαση εντός ενός κανονιστικά και κοινωνικά φορτισμένου περιβάλλοντος.

2.12 Σύνθεση Βιβλιογραφίας και Ερευνητικό Κενό

Η προηγούμενη ανάλυση της διεθνούς βιβλιογραφίας ανέδειξε τον πολυδιάστατο ρόλο της Τεχνητής Νοημοσύνης (AI) και της Μηχανικής Μάθησης (ML) στη σύγχρονη διαχείριση χρηματοοικονομικών κινδύνων. Οι τεχνολογίες αυτές έχουν συνδεθεί με σημαντικές βελτιώσεις στην πρόβλεψη πιστωτικής αθέτησης (Lessmannetal., 2015), στην ανίχνευση απάτης μέσω προηγμένων αλγορίθμων ταξινόμησης (Bahnsenetal., 2016· DalPozzoloetal., 2015), καθώς και στη δυναμική διαχείριση χαρτοφυλακίων με στόχο την επίτευξη βελτιωμένων αποδόσεων προσαρμοσμένων στον κίνδυνο (Moody&Saffell, 2001). Παράλληλα, οι εφαρμογές της AI στην ανάλυση συστημικού κινδύνου και στη μοντελοποίηση χρηματοπιστωτικών δικτύων έχουν προσφέρει νέα εργαλεία κατανόησης της μετάδοσης κρίσεων και της αλληλεξάρτησης των αγορών (Acharyaetal., 2010· Bisiasetal., 2012). Η βιβλιογραφία συγκλίνει στο ότι η AI ενισχύει τη δυνατότητα επεξεργασίας μεγάλων και ετερογενών συνόλων δεδομένων, επιτρέποντας την αναγνώριση σύνθετων και μη γραμμικών προτύπων που υπερβαίνουν τις δυνατότητες των παραδοσιακών στατιστικών μοντέλων.

Ωστόσο, παρά την τεχνολογική πρόοδο και τα τεκμηριωμένα οφέλη σε επίπεδο απόδοσης, η βιβλιογραφία αναδεικνύει ταυτόχρονα κρίσιμες προκλήσεις που σχετίζονται με την ερμηνευσιμότητα, τη διαφάνεια και τη λογοδοσία των αλγοριθμικών συστημάτων. Η συζήτηση γύρω από τα «μαύρα κουτιά» των μοντέλων μηχανικής μάθησης έχει αναδείξει την

ανάγκη για Explainable AI (XAI), δηλαδή για μεθοδολογίες και τεχνικές που επιτρέπουν την κατανόηση και ερμηνεία του τρόπου με τον οποίο ένα αλγοριθμικό σύστημα καταλήγει σε μια συγκεκριμένη απόφαση ή πρόβλεψη (Doshi-Velez & Kim, 2017· Barredo Arrieta et al., 2020). Η Explainable AI στοχεύει στην παροχή διαφανών, επαληθεύσιμων και αιτιολογημένων αποτελεσμάτων, ώστε οι χρήστες και οι ρυθμιστικές αρχές να μπορούν να αξιολογούν τη λογική του μοντέλου. Επιπλέον, ζητήματα αλγοριθμικής μεροληψίας και πιθανών διακρίσεων έχουν επισημανθεί ως σημαντικοί παράγοντες που επηρεάζουν την αξιοπιστία και την κοινωνική αποδοχή των συστημάτων AI (Binns, 2018). Η κανονιστική διάσταση, όπως αποτυπώνεται στον GDPR και στις σχετικές αναλύσεις περί αυτοματοποιημένης λήψης αποφάσεων (Goodman & Flaxman, 2017), ενισχύει περαιτέρω τη σημασία της διαφάνειας και της υπευθυνότητας.

Ιδιαίτερα στον χρηματοπιστωτικό τομέα, όπου οι αλγοριθμικές αποφάσεις αφορούν πιστωτικές εγκρίσεις, κεφαλαιακές κατανομές και στρατηγικές διαχείρισης κινδύνου, η υιοθέτηση της AI δεν αποτελεί απλώς τεχνική επιλογή αλλά στρατηγική απόφαση υψηλής σημασίας. Οι οργανισμοί καλούνται να σταθμίσουν τα δυνητικά οφέλη από τη βελτιωμένη ακρίβεια και την ταχύτερη επεξεργασία δεδομένων με τους κινδύνους που απορρέουν από πιθανές λανθασμένες προβλέψεις, κανονιστικές κυρώσεις ή φθορά της φήμης τους. Η εμπειρία προηγούμενων κρίσεων και η αυξημένη ευαισθησία απέναντι σε συστημικούς κινδύνους (Acharyaetal., 2010) καθιστούν τα στελέχη ιδιαίτερα προσεκτικά απέναντι σε καινοτόμες αλλά σύνθετες τεχνολογίες.

Παράλληλα με τη βιβλιογραφία που εστιάζει στην τεχνική αποτελεσματικότητα των μοντέλων, αναπτύχθηκε ένα ισχυρό ερευνητικό ρεύμα γύρω από τις θεωρίες αποδοχής τεχνολογίας. Το Μοντέλο Αποδοχής Τεχνολογίας (Davis, 1989) και η Ενοποιημένη Θεωρία Αποδοχής και Χρήσης Τεχνολογίας (Venkateshetal., 2003) ανέδειξαν ότι η πρόθεση χρήσης επηρεάζεται κυρίως από την αντιληπτή χρησιμότητα και την αντιληπτή ευκολία χρήσης, καθώς και από παράγοντες όπως η κοινωνική επιρροή και οι διευκολυντικές συνθήκες. Στο πλαίσιο των συστημάτων AI, οι μεταβλητές αυτές αποκτούν ιδιαίτερη σημασία, καθώς η τεχνολογική πολυπλοκότητα και η ανάγκη εξειδικευμένων γνώσεων μπορούν να επηρεάσουν αρνητικά την πρόθεση υιοθέτησης, ακόμη και όταν τα αντικειμενικά οφέλη είναι σημαντικά.

Επιπλέον, η βιβλιογραφία έχει αναδείξει την εμπιστοσύνη ως καθοριστικό παράγοντα υιοθέτησης αυτοματοποιημένων συστημάτων, ιδίως σε περιβάλλοντα υψηλού ρίσκου. Η εμπιστοσύνη σχετίζεται με την πεποίθηση ότι το σύστημα λειτουργεί με συνέπεια, αξιοπιστία και χωρίς μεροληψία. Σε περιβάλλοντα όπως η πιστωτική αξιολόγηση ή η ανίχνευση απάτης, όπου τα αποτελέσματα των αλγορίθμων μπορεί να επηρεάσουν άμεσα πελάτες και

οικονομικά μεγέθη, η έλλειψη εμπιστοσύνης μπορεί να λειτουργήσει αποτρεπτικά, ανεξαρτήτως της τεχνικής υπεροχής του μοντέλου (BarredoArrietaetal., 2020). Ταυτόχρονα, ο αντιληπτός κίνδυνος, δηλαδή η υποκειμενική εκτίμηση πιθανών αρνητικών συνεπειών από τη χρήση της τεχνολογίας, έχει αποδειχθεί ότι επηρεάζει αρνητικά την πρόθεση χρήσης, ιδίως όταν συνδέεται με αβεβαιότητα ή κανονιστική ασάφεια (DalPozzoloetal., 2015).

Παρά την πλούσια βιβλιογραφία τόσο στον τομέα της τεχνικής απόδοσης της AI όσο και στις θεωρίες αποδοχής τεχνολογίας, παρατηρείται ένα ουσιαστικό ερευνητικό κενό. Οι περισσότερες μελέτες εξετάζουν είτε την αποτελεσματικότητα των αλγοριθμικών μοντέλων σε επίπεδο ακρίβειας και προβλεπτικής ισχύος είτε τους γενικούς παράγοντες που επηρεάζουν την αποδοχή πληροφοριακών συστημάτων. Περιορισμένες είναι οι εμπειρικές έρευνες που συνδυάζουν τις δύο αυτές διαστάσεις, εξετάζοντας πώς οι αντιλήψεις των στελεχών σχετικά με τη χρησιμότητα, την ευκολία χρήσης, την εμπιστοσύνη και τον κίνδυνο επηρεάζουν την πρόθεση υιοθέτησης AI ειδικά για σκοπούς διαχείρισης χρηματοοικονομικών κινδύνων.

Επιπλέον, η υπάρχουσα βιβλιογραφία συχνά εστιάζει σε τεχνολογίες γενικής χρήσης ή σε περιβάλλοντα καταναλωτικών εφαρμογών, χωρίς να λαμβάνει υπόψη τις ιδιαιτερότητες του χρηματοπιστωτικού τομέα, όπως η αυστηρή κανονιστική εποπτεία, η υψηλή έκθεση σε συστημικούς κινδύνους και η ανάγκη για τεκμηριωμένη λήψη αποφάσεων. Το γεγονός ότι οι αποφάσεις στον τομέα αυτό έχουν άμεσες επιπτώσεις στη χρηματοοικονομική σταθερότητα και στην εμπιστοσύνη των αγορών καθιστά την κατανόηση της πρόθεσης υιοθέτησης AI ιδιαίτερα κρίσιμη.

Στο πλαίσιο αυτό, καθίσταται αναγκαία η εμπειρική διερεύνηση των παραγόντων που επηρεάζουν την πρόθεση υιοθέτησης τεχνολογιών AI στη διαχείριση κινδύνων, με εστίαση στις αντιλήψεις των στελεχών που εμπλέκονται άμεσα στη διαδικασία λήψης αποφάσεων. Η παρούσα μελέτη επιχειρεί να καλύψει το ανωτέρω ερευνητικό κενό, εξετάζοντας συνδυαστικά τη συμβολή της αντιληπτής χρησιμότητας, της αντιληπτής ευκολίας χρήσης, της εμπιστοσύνης στο σύστημα και του αντιληπτού κινδύνου στη διαμόρφωση της πρόθεσης υιοθέτησης AI για σκοπούς riskmanagement.

Η διατύπωση και ο έλεγχος των σχετικών υποθέσεων επιτρέπουν τη γεφύρωση του χάσματος μεταξύ τεχνολογικής αποτελεσματικότητας και οργανωσιακής αποδοχής. Μέσα από τη σύνδεση της βιβλιογραφίας περί AI στη διαχείριση κινδύνων με τις θεωρίες αποδοχής τεχνολογίας, η παρούσα έρευνα φιλοδοξεί να προσφέρει μια ολοκληρωμένη κατανόηση των παραγόντων που καθορίζουν την επιτυχημένη ενσωμάτωση της Τεχνητής Νοημοσύνης στον χρηματοπιστωτικό τομέα. Με τον τρόπο αυτό, συμβάλλει όχι μόνο στη θεωρητική συζήτηση περί υιοθέτησης καινοτόμων τεχνολογιών, αλλά και στην πρακτική διαμόρφωση στρατηγικών

που θα επιτρέψουν στους οργανισμούς να αξιοποιήσουν τις δυνατότητες της AI με τρόπο υπεύθυνο, διαφανή και κανονιστικά συμβατό.

Κεφάλαιο 3: Μεθοδολογία

3.1 Ερευνητικό Σχέδιο

Η παρούσα εργασία ακολουθεί ποσοτική εμπειρική μεθοδολογία, με στόχο τη διερεύνηση των παραγόντων που επηρεάζουν την πρόθεση χρήσης συστημάτων Τεχνητής Νοημοσύνης (AI) στον τομέα της διαχείρισης κινδύνου. Όπως αναπτύχθηκε στο προηγούμενο κεφάλαιο, η διεθνής βιβλιογραφία σχετικά με τις θεωρίες αποδοχής τεχνολογίας αναδεικνύει την αντιλαμβανόμενη χρησιμότητα και την αντιλαμβανόμενη ευκολία χρήσης ως βασικούς προσδιοριστικούς παράγοντες της πρόθεσης υιοθέτησης (Davis, 1989· Venkatesh & Davis, 2000), ενώ νεότερες μελέτες σε περιβάλλοντα υψηλού ρίσκου επισημαίνουν τη σημασία της εμπιστοσύνης και του αντιλαμβανόμενου κινδύνου (Gefen et al., 2003· Featherman & Pavlou, 2003).

Με βάση τη θεωρητική αυτή σύνθεση, υιοθετείται το Technology Acceptance Model (TAM), εμπλουτισμένο με τις μεταβλητές της εμπιστοσύνης (trust) και του αντιλαμβανόμενου κινδύνου (perceived risk), ώστε να εξεταστεί η εφαρμογή του στο ειδικό πλαίσιο της διαχείρισης χρηματοοικονομικού κινδύνου. Το ερευνητικό σχέδιο βασίζεται σε διατομεακή έρευνα πεδίου (cross-sectional study) μέσω ηλεκτρονικού ερωτηματολογίου, το οποίο σχεδιάστηκε βάσει επικυρωμένων κλιμάκων της διεθνούς βιβλιογραφίας και προσαρμόστηκε στο ελληνικό επαγγελματικό περιβάλλον.

3.2 Δειγματοληψία και Συλλογή Δεδομένων

Το δείγμα της έρευνας αποτελείται από 75 επαγγελματίες που δραστηριοποιούνται στους τομείς της χρηματοοικονομικής, της τεχνολογίας, της ασφάλισης και της διαχείρισης κινδύνου. Η συλλογή δεδομένων πραγματοποιήθηκε μέσω ηλεκτρονικού ερωτηματολογίου (Google Forms) κατά το διάστημα Νοέμβριος 2025 – Ιανουάριος 2026.

Η επιλογή των συμμετεχόντων έγινε με τη μέθοδο της μη τυχαίας δειγματοληψίας σκοπιμότητας (purposive sampling). Η συγκεκριμένη μέθοδος κρίθηκε καταλληλότερη, καθώς το αντικείμενο της έρευνας αφορά εξειδικευμένες αντιλήψεις σχετικά με την υιοθέτηση τεχνολογιών Τεχνητής Νοημοσύνης στη διαχείριση χρηματοοικονομικών κινδύνων. Συνεπώς, ήταν απαραίτητη η επιλογή ατόμων με σχετική επαγγελματική εμπειρία ή ακαδημαϊκό υπόβαθρο, ώστε να διασφαλιστεί ότι οι απαντήσεις βασίζονται σε ουσιαστική γνώση του αντικειμένου.

Δεδομένου ότι δεν υπάρχει διαθέσιμο πλήρες μητρώο επαγγελματιών με άμεση εμπλοκή σε AI και risk management, η εφαρμογή τυχαίας δειγματοληψίας δεν ήταν εφικτή. Ως εκ τούτου, η purposive sampling επιτρέπει τη στοχευμένη επιλογή πληροφοριακά πλούσιων περιπτώσεων (information-rich cases), οι οποίες είναι καταλληλότερες για τη διερεύνηση σύνθετων τεχνολογικών και οργανωσιακών φαινομένων.

Το δείγμα δεν επιδιώκει στατιστική αντιπροσωπευτικότητα του συνολικού πληθυσμού των επαγγελματιών του χρηματοπιστωτικού τομέα. Αντίθετα, στοχεύει σε αναλυτική γενίκευση των ευρημάτων σε πληθυσμό στελεχών με σχετική γνώση και εμπειρία στον τομέα της AI και της διαχείρισης κινδύνου. Η ανωνυμία και η εθελοντική συμμετοχή διασφαλίστηκαν πλήρως καθ' όλη τη διάρκεια της έρευνας.

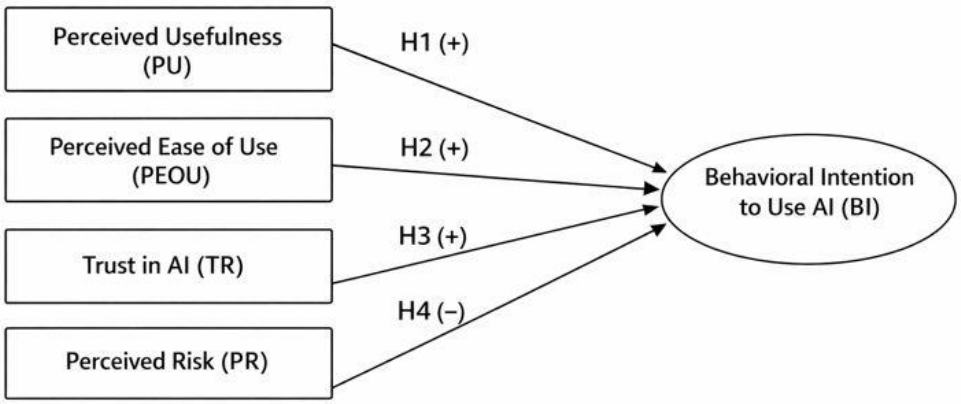
3.3 Μεταβλητές και Θεωρητικό Πλαίσιο

Η ερευνητική υπόθεση βασίζεται στο μοντέλο TAM και διαμορφώνεται με τις εξής μεταβλητές:

Μεταβλητή	**Περιγραφή**
PU – Perceived Usefulness	Η αντιληπτή χρησιμότητα των συστημάτων AI
PEOU – Perceived Ease of Use	Η αντιληπτή ευκολία χρήσης των συστημάτων AI
TR – Trust	Η εμπιστοσύνη στα συστήματα AI
PR – Perceived Risk	Ο αντιληπτός κίνδυνος από τη χρήση AI
BI – Behavioral Intention	Η πρόθεση χρήσης των συστημάτων AI

Ερευνητικές Υποθέσεις:

- Υ1: Η PU επηρεάζει θετικά την BI.
- Υ2: Η PEOU επηρεάζει θετικά την BI.
- Υ3: Η TR επηρεάζει θετικά την BI.
- Υ4: Η PR επηρεάζει αρνητικά την BI.



Γράφημα 1: Θεωρητικό Μοντέλο της Έρευνας

Το σχήμα απεικονίζει τη θεωρητική δομή της έρευνας, βασισμένη στο επεκταμένο μοντέλο αποδοχής τεχνολογίας (TAM). Οι μεταβλητές αντιληπτή χρησιμότητα (PU), αντιληπτή ευκολία χρήσης (PEOU) και εμπιστοσύνη (Trust) υποτίθεται ότι επηρεάζουν θετικά την πρόθεση χρήσης (BI), ενώ ο αντιληπτός κίνδυνος (Perceived Risk) αναμένεται να έχει αρνητική επίδραση. Τα βέλη υποδηλώνουν τις κατευθύνσεις των ερευνητικών υποθέσεων (Y1–Y4).

3.4Ερωτηματολόγιο και Μέτρηση

Το ερωτηματολόγιο περιλάμβανε δύο βασικά μέρη. Το Μέρος Α αφορούσε τη συλλογή βασικών δημογραφικών στοιχείων (φύλο, ηλικία, επαγγελματικός τομέας, έτη επαγγελματικής εμπειρίας), τα οποία χρησιμοποιήθηκαν αποκλειστικά για την περιγραφή της σύνθεσης του δείγματος και δεν ενσωματώθηκαν στα υποδείγματα ανάλυσης, καθώς το ερευνητικό μοντέλο εστιάζει αποκλειστικά στις αντιληπτικές μεταβλητές που σχετίζονται με την πρόθεση υιοθέτησης της Τεχνητής Νοημοσύνης.

Το Μέρος Β περιλάμβανε τις κλίμακες μέτρησης των βασικών μεταβλητών της έρευνας: αντιληπτή χρησιμότητα (PU), αντιληπτή ευκολία χρήσης (PEOU), εμπιστοσύνη (TR), αντιληπτό κίνδυνο (PR) και πρόθεση υιοθέτησης (BI). Για όλες τις μεταβλητές χρησιμοποιήθηκαν κλίμακες Likert 5 βαθμίδων (1=διαφωνώ απόλυτα, 5=συμφωνώ απόλυτα). Τα ερωτήματα αντλήθηκαν από επικυρωμένες κλίμακες της διεθνούς βιβλιογραφίας

(Venkatesh&Davis, 2000· Gefenetal., 2003· Featherman&Pavlou, 2003), και προσαρμόστηκαν εννοιολογικά στο πλαίσιο της παρούσας έρευνας.

3.5 Στατιστική Ανάλυση

Η ανάλυση των δεδομένων πραγματοποιήθηκε με χρήση του στατιστικού λογισμικού **SPSS**και περιλαμβάνει τα εξής στάδια:

- Ανάλυση περιγραφικής στατιστικής: Μέσοι όροι, τυπικές αποκλίσεις και κατανομές απαντήσεων ανά μεταβλητή.
- Έλεγχος αξιοπιστίας: Υπολογισμός του δείκτη Cronbach'salpha για κάθε κλίμακα.
- Έλεγχος συσχετίσεων: Ανάλυση συσχετίσεων μεταξύ μεταβλητών με χρήση του συντελεστή Pearson (r).
- Πολλαπλή γραμμική παλινδρόμηση: Η συνεξάρτηση της BI από τις PU, PEOU, TR και PR εκτιμάται με το μοντέλο:

$$\mathrm{BI} = \beta_0 + \beta_1 \mathrm{PU} + \beta_2 \mathrm{PEOU} + \beta_3 \mathrm{TR} + \beta_4 \mathrm{PR} + \varepsilon$$

Προϋποθέσεις όπως η κανονικότητα, πολυσυγγραμικότητα (VIF), και ομοσκεδαστικότητα εξετάστηκαν πριν την εκτίμηση του μοντέλου.

3.6 Περιορισμοί της Μεθοδολογίας

Παρά τα ισχυρά σημεία της μεθοδολογικής προσέγγισης, υπάρχουν ορισμένοι περιορισμοί:

- Το μέγεθος δείγματος (N=75) θεωρείται ικανοποιητικό για αρχική διερεύνηση αλλά όχι για γενικεύσεις πληθυσμού.
- Η μη τυχαία δειγματοληψία μπορεί να ενέχει προκατάληψη επιλογής.
- Το εργαλείο βασίζεται σε υποκειμενική δήλωση προθέσεων και όχι σε πραγματική χρήση AI.
- Το πολιτισμικό/γεωγραφικό πλαίσιο περιορίζεται στην Ελλάδα.

Κεφάλαιο 4: Ανάλυση Δεδομένων

Το παρόν κεφάλαιο παρουσιάζει την εμπειρική ανάλυση των δεδομένων που συλλέχθηκαν μέσω ερωτηματολογίου, το οποίο σχεδιάστηκε βάσει του επεκταμένου μοντέλου αποδοχής τεχνολογίας (TAM) και περιλάμβανε πέντε βασικές μεταβλητές: Αντιληπτή Χρησιμότητα (PU), Αντιληπτή Ευκολία Χρήσης (PEOU), Εμπιστοσύνη (Trust), Αντιληπτός Κίνδυνος (Perceived Risk) και Πρόθεση Χρήσης (Behavioral Intention). Σκοπός της ανάλυσης είναι η καταγραφή της στάσης των επαγγελματιών απέναντι στη χρήση Τεχνητής Νοημοσύνης στη διαχείριση κινδύνου.

Στην έρευνα συμμετείχαν συνολικά 75 επαγγελματίες, από διαφορετικά πεδία απασχόλησης (όπως τραπεζικός τομέας, τεχνολογία, εκπαίδευση, υγεία), ηλικιακές ομάδες και επίπεδα εμπειρίας με συστήματα AI. Η στατιστική ανάλυση περιλαμβάνει περιγραφικά στατιστικά (μέσοι όροι, τυπικές αποκλίσεις, ελάχιστες και μέγιστες τιμές) για κάθε επιμέρους μεταβλητή και δείκτη (item), με στόχο την εξαγωγή ασφαλών συμπερασμάτων για την αντιληπτή αποδοχή της AI στη διαχείριση κινδύνου.

4.1 Περιγραφική Στατιστική Ανάλυση

Η παρούσα υποενότητα παρουσιάζει την περιγραφική στατιστική ανάλυση των βασικών μεταβλητών του ερευνητικού μοντέλου, με στόχο την αποτύπωση της γενικής στάσης των συμμετεχόντων απέναντι στη χρήση συστημάτων Τεχνητής Νοημοσύνης (ArtificialIntelligence – AI) στον τομέα της διαχείρισης κινδύνου. Η ανάλυση βασίζεται σε δείγμα 75 συμμετεχόντων (N = 75) και πραγματοποιήθηκε με τη χρήση πενταβάθμιας κλίμακας Likert (1 = Διαφωνώ απόλυτα έως 5 = Συμφωνώ απόλυτα).

Κάθε θεωρητική μεταβλητή μετρήθηκε μέσω επιμέρους δηλώσεων (items), σύμφωνα με το ερωτηματολόγιο της έρευνας. Συγκεκριμένα, η αντιληπτή χρησιμότητα (PU) και η αντιληπτή ευκολία χρήσης (PEOU) αποτυπώθηκαν με τέσσερις δηλώσεις, ενώ η εμπιστοσύνη (TR), ο αντιληπτός κίνδυνος (PR) και η πρόθεση χρήσης (BI) μετρήθηκαν με τρεις δηλώσεις.

Για κάθε δείκτη παρουσιάζονται η μέση τιμή, η τυπική απόκλιση, καθώς και οι ελάχιστες και μέγιστες παρατηρούμενες τιμές, προκειμένου να καταγραφεί τόσο το κεντρικό επίπεδο των απαντήσεων όσο και η διασπορά τους.

4.1.1 Αντιληπτή Χρησιμότητα (PerceivedUsefulness – PU)

Η μεταβλητή της αντιληπτής χρησιμότητας αποτυπώνει τον βαθμό στον οποίο οι συμμετέχοντες θεωρούν ότι η χρήση της Τεχνητής Νοημοσύνης συμβάλλει στη βελτίωση της ποιότητας των αποφάσεών τους, στην αύξηση της αποτελεσματικότητας, στην ταχύτερη εκτέλεση των εργασιών και στη συνολική εργασιακή απόδοση.

Πίνακας 1: Περιγραφικά στατιστικά δεικτών Αντιληπτής Χρησιμότητας (PU)

Δείκτης	**Μέση τιμή**	**Τυπική απόκλιση**	**Ελάχιστο**	**Μέγιστο**
PU1	3,95	0,88	1	5
PU2	3,91	0,90	1	5
PU3	4,09	0,83	1	5
PU4	4,01	0,85	1	5
Μέσος όρος PU	**3,99**	**0,86**	–	–

Από τον Πίνακα 1 παρατηρείται ότι και οι τέσσερις επιμέρους δείκτες της αντιληπτής χρησιμότητας παρουσιάζουν μέσες τιμές κοντά ή και άνω του 4. Η υψηλότερη μέση τιμή εντοπίζεται στη δήλωση PU3, η οποία αφορά την ταχύτερη εκτέλεση των εργασιών μέσω της AI (M = 4,09), γεγονός που υποδηλώνει ότι οι συμμετέχοντες αναγνωρίζουν κυρίως τη λειτουργική αποδοτικότητα της τεχνολογίας.

Η χαμηλότερη –αν και σαφώς θετική– μέση τιμή εμφανίζεται στη δήλωση PU2 (M = 3,91), η οποία αφορά την ενίσχυση της αποτελεσματικότητας στη διαχείριση κινδύνων. Συνολικά, ο σύνθετος δείκτης PU παρουσιάζει μέσο όρο 3,99, γεγονός που καταδεικνύει ότι το δείγμα αντιλαμβάνεται την AI ως ουσιαστικά ωφέλιμο εργαλείο για την υποστήριξη της εργασιακής απόδοσης και των διαδικασιών λήψης αποφάσεων.

Η σχετικά χαμηλή τυπική απόκλιση σε όλους τους δείκτες υποδηλώνει περιορισμένη διασπορά στις απαντήσεις και συνεπώς σχετικά ομοιογενή στάση των συμμετεχόντων.

4.1.2 Αντιληπτή Ευκολία Χρήσης (PerceivedEaseofUse – PEOU)

Η μεταβλητή της αντιληπτής ευκολίας χρήσης αποτυπώνει την αντίληψη των συμμετεχόντων σχετικά με τη δυσκολία εκμάθησης, κατανόησης και καθημερινής χρήσης συστημάτων Τεχνητής Νοημοσύνης.

Πίνακας 2: Περιγραφικά στατιστικά δεικτών Αντιληπτής Ευκολίας Χρήσης (PEOU)

Δείκτης	Μέση τιμή	Τυπική απόκλιση	Ελάχιστο	Μέγιστο
PEOU1	4,01	0,91	1	5
PEOU2	3,95	0,89	1	5
PEOU3	4,10	0,82	1	5
PEOU4	3,64	1,01	1	5
Μέσος όρος PEOU	**3,93**	**0,91**	–	–

Όπως προκύπτει από τον Πίνακα2, οι συμμετέχοντες αξιολογούν θετικά την ευκολία χρήσης των συστημάτων AI. Η υψηλότερη μέση τιμή εμφανίζεται στη δήλωση PEOU3 (M = 4,10), η οποία σχετίζεται με την ευκολία απόκτησης δεξιοτήτων στη χρήση της AI. Το εύρημα αυτό υποδηλώνει ότι οι συμμετέχοντες θεωρούν ότι μπορούν σχετικά εύκολα να εξοικειωθούν με τα σχετικά συστήματα.

Αντίθετα, η χαμηλότερη μέση τιμή παρατηρείται στη δήλωση PEOU4 (M = 3,64), που αφορά το κατά πόσο η χρήση της AI δεν απαιτεί ιδιαίτερη προσπάθεια. Η συγκεκριμένη διαφοροποίηση υποδηλώνει ότι, αν και η AI θεωρείται κατανοητή και μαθαίνεται σχετικά εύκολα, εξακολουθεί να απαιτείται γνωστική ή τεχνική προσπάθεια για την αποτελεσματική αξιοποίησή της.

Ο συνολικός μέσος όρος της μεταβλητής (M = 3,93) δείχνει ότι το δείγμα διατηρεί γενικά θετική στάση ως προς τη χρηστικότητα και την εργονομία των συστημάτων AI.

4.1.3 Εμπιστοσύνη στα συστήματα Τεχνητής Νοημοσύνης (Trust – TR)

Η μεταβλητή της εμπιστοσύνης εξετάζει τον βαθμό στον οποίο οι συμμετέχοντες θεωρούν ότι τα συστήματα AI λειτουργούν αξιόπιστα, λαμβάνουν δίκαιες αποφάσεις και μπορούν να χρησιμοποιηθούν με ασφάλεια στη διαδικασία λήψης αποφάσεων.

Πίνακας 3: Περιγραφικά στατιστικά δεικτών Εμπιστοσύνης (TR)

Δείκτης	Μέση τιμή	Τυπική απόκλιση	Ελάχιστο	Μέγιστο
TR1	3,77	0,93	1	5
TR2	3,72	0,92	1	5
TR3	3,66	0,98	1	5
Μέσος όρος TR	**3,72**	**0,94**	–	–

Ο Πίνακας 3 καταδεικνύει ότι η εμπιστοσύνη προς την AI βρίσκεται σε μέτρια προς θετικά επίπεδα. Η υψηλότερη μέση τιμή παρατηρείται στη δήλωση που αφορά τη γενική αξιοπιστία των συστημάτων (TR1), ενώ η χαμηλότερη εντοπίζεται στη δήλωση που σχετίζεται με το αίσθημα ασφάλειας κατά τη χρήση της AI στη λήψη αποφάσεων (TR3).

Το εύρημα αυτό φανερώνει ότι, παρότι οι συμμετέχοντες αναγνωρίζουν τη λειτουργική αξιοπιστία της AI, εμφανίζονται πιο επιφυλακτικοί όταν η τεχνολογία εμπλέκεται σε κρίσιμες αποφάσεις. Η ύπαρξη αυτής της επιφύλαξης είναι αναμενόμενη στο πλαίσιο εφαρμογών διαχείρισης κινδύνου, όπου το σφάλμα μπορεί να έχει σοβαρές επιχειρησιακές ή κανονιστικές συνέπειες.

4.1.4 Αντιληπτός Κίνδυνος (PerceivedRisk – PR)

Η μεταβλητή του αντιληπτού κινδύνου αποτυπώνει τον βαθμό ανησυχίας των συμμετεχόντων σχετικά με τις ενδεχόμενες αρνητικές επιπτώσεις από τη χρήση συστημάτων AI στο επαγγελματικό τους περιβάλλον.

Πίνακας 4: Περιγραφικά στατιστικά δεικτών Αντιληπτού Κινδύνου (PR)

Δείκτης	**Μέση τιμή**	**Τυπική απόκλιση**	**Ελάχιστο**	**Μέγιστο**
PR1	2,86	0,99	1	5
PR2	2,92	1,05	1	5
PR3	3,07	1,06	1	5
Μέσος όρος PR	**2,95**	**1,03**	–	–

Τα αποτελέσματα του Πίνακα 4 δείχνουν ότι το επίπεδο αντιληπτού κινδύνου βρίσκεται κάτω από το ουδέτερο σημείο της κλίμακας (τιμή 3). Η υψηλότερη μέση τιμή εμφανίζεται στη δήλωση PR3, η οποία αφορά την επιφυλακτικότητα σχετικά με την εφαρμογή της AI σε κρίσιμες διαδικασίες.

Το εύρημα αυτό υποδηλώνει ότι οι συμμετέχοντες δεν θεωρούν τη χρήση της AI ως ιδιαίτερα επικίνδυνη, ωστόσο διατηρούν έναν βαθμό επιφυλακτικότητας όταν πρόκειται για εφαρμογές υψηλής κρισιμότητας. Η σχετικά αυξημένη τυπική απόκλιση υποδηλώνει διαφοροποίηση απόψεων μεταξύ των συμμετεχόντων, γεγονός που πιθανόν αντανακλά διαφορετικά επίπεδα εμπειρίας και επαγγελματικής έκθεσης στην AI.

4.1.5 Πρόθεση Χρήσης (Behavioral Intention – BI)

Η μεταβλητή της πρόθεσης χρήσης εκφράζει την πρόθεση των συμμετεχόντων να χρησιμοποιήσουν συστήματα AI στο επαγγελματικό τους περιβάλλον στο μέλλον.

Πίνακας 5: Περιγραφικά στατιστικά δεικτών Πρόθεσης Χρήσης (BI)

Δείκτης	**Μέση τιμή**	**Τυπική απόκλιση**	**Ελάχιστο**	**Μέγιστο**
BI1	4,31	0,77	1	5
BI2	4,33	0,74	1	5
BI3	4,24	0,81	1	5
Μέσος όρος BI	**4,29**	**0,77**	–	–

Ο Πίνακας 5 καταδεικνύει ότι η πρόθεση χρήσης της AI είναι ιδιαίτερα υψηλή. Και οι τρεις επιμέρους δηλώσεις καταγράφουν μέσες τιμές άνω του 4, με την υψηλότερη να αφορά τη συστηματική χρήση της AI σε περίπτωση διαθεσιμότητας σχετικών εργαλείων.

Το εύρημα αυτό αναδεικνύει ότι, ανεξαρτήτως επιφυλάξεων που σχετίζονται με την εμπιστοσύνη ή τον κίνδυνο, οι επαγγελματίες εμφανίζονται θετικά διακείμενοι ως προς την υιοθέτηση της AI στη διαχείριση κινδύνου.

4.1.6 Συγκεντρωτικός Πίνακας Περιγραφικών Στατιστικών

Πίνακας 6: Συγκεντρωτικά περιγραφικά στατιστικά των βασικών μεταβλητών

Μεταβλητή	**Μέση τιμή**	**Τυπική απόκλιση**
Αντιληπτή Χρησιμότητα (PU)	3,99	0,86
Αντιληπτή Ευκολία Χρήσης (PEOU)	3,93	0,91
Εμπιστοσύνη (TR)	3,72	0,94
Αντιληπτός Κίνδυνος (PR)	2,95	1,03
Πρόθεση Χρήσης (BI)	4,29	0,77

Ο συγκεντρωτικός Πίνακας 4.6 συνοψίζει τη συνολική εικόνα της στάσης των συμμετεχόντων απέναντι στην Τεχνητή Νοημοσύνη στη διαχείριση κινδύνου. Παρατηρείται ότι οι υψηλότερες μέσες τιμές καταγράφονται στην πρόθεση χρήσης και στην αντιληπτή χρησιμότητα, γεγονός που υποδηλώνει έντονη αποδοχή της τεχνολογίας και αναγνώριση της

πρακτικής της αξίας. Αντίθετα, ο αντιληπτός κίνδυνος εμφανίζεται χαμηλότερος, στοιχείο που λειτουργεί ενισχυτικά για τη μελλοντική υιοθέτηση των συστημάτων AI.

Στο επόμενο υποκεφάλαιο (4.3) εξετάζονται οι συσχετίσεις μεταξύ των μεταβλητών, με σκοπό την περαιτέρω διερεύνηση των σχέσεων που διατυπώνονται στις ερευνητικές υποθέσεις της παρούσας μελέτης.

4.2 Έλεγχος Αξιοπιστίας των Κλιμάκων

Σε αυτή την υποενότητα πραγματοποιείται έλεγχος εσωτερικής συνέπειας (internal consistency) των πέντε βασικών εννοιολογικών κατασκευών (constructs) του θεωρητικού μοντέλου μέσω του δείκτη Cronbach's Alpha. Ο δείκτης αυτός αξιολογεί κατά πόσο τα επιμέρους items κάθε κλίμακας αποτυπώνουν με συνέπεια την ίδια θεωρητική έννοια. Τιμές πάνω από 0,7 θεωρούνται ικανοποιητικές, ενώ τιμές άνω του 0,8 υποδηλώνουν υψηλή αξιοπιστία (Hair et al., 2019).

4.2.1 Αξιοπιστία Αντιληπτής Χρησιμότητας (PU)

Η κλίμακα PU αποτελείται από 4 items (PU1–PU4) και παρουσιάζει υψηλή εσωτερική συνέπεια.

Δείκτης	**Τιμή**
Cronbach's Alpha	**0,875**

Η τιμή δείχνει ότι οι δηλώσεις σχετικά με τη χρησιμότητα της AI είναι στενά συνδεδεμένες μεταξύ τους και μετρούν με συνέπεια την έννοια της αντιληπτής χρησιμότητας.

4.2.2 Αξιοπιστία Αντιληπτής Ευκολίας Χρήσης (PEOU)

Η μεταβλητή PEOU αποτελείται από 4 δηλώσεις (PEOU1–PEOU4).

Δείκτης	**Τιμή**
Cronbach's Alpha	0,861

Η τιμή είναι ιδιαίτερα ικανοποιητική και αποδεικνύει ότι η έννοια της ευκολίας χρήσης της AI γίνεται κατανοητή με σταθερό τρόπο από τους συμμετέχοντες.

4.2.3 Αξιοπιστία Εμπιστοσύνης (Trust – TR)

Η κλίμακα TR περιλαμβάνει 3 δηλώσεις (TR1–TR3).

Δείκτης	**Τιμή**
Cronbach's Alpha	0,821

Η τιμή είναι εντός των ορίων υψηλής αξιοπιστίας, γεγονός που δείχνει ομοιογένεια στις απαντήσεις των συμμετεχόντων ως προς την εμπιστοσύνη στην AI.

4.2.4 Αξιοπιστία Αντιληπτού Κινδύνου (PR)

Η μεταβλητή PR αποτελείται επίσης από 3δηλώσεις(PR1–PR3).

Δείκτης	**Τιμή**
Cronbach's Alpha	0,792

Η αξιοπιστία της κλίμακας είναι οριακά υψηλή, κάτι που επιβεβαιώνει ότι οι συμμετέχοντες αντιλαμβάνονται με παρόμοιο τρόπο τις πτυχές κινδύνου που σχετίζονται με τη χρήση της AI.

4.2.5 Αξιοπιστία Πρόθεσης Χρήσης (BehavioralIntention – BI)

Η κλίμακα BI περιλαμβάνει 3 δηλώσεις (BI1–BI3).

Δείκτης	**Τιμή**
Cronbach's Alpha	**0,905**

Πρόκειται για την υψηλότερη τιμή ανάμεσα στις πέντε κλίμακες, γεγονός που υποδηλώνει πολύ ισχυρή συνοχή μεταξύ των δηλώσεων για την πρόθεση υιοθέτησης της AI.

Πίνακας 7: Συνοπτικός Πίνακας Cronbach'sAlpha

Μεταβλητή	**Αριθμός Items**	**Cronbach's Alpha**
PU – Χρησιμότητα	4	**0,875**
PEOU – Ευκολία Χρήσης	4	**0,861**
TR – Εμπιστοσύνη	3	**0,821**
PR – Αντιληπτός Κίνδυνος	3	**0,792**
BI – Πρόθεση Χρήσης	3	**0,905**

Όλες οι κλίμακες εμφανίζουν υψηλή αξιοπιστία, επιτρέποντας τη συνέχιση της ανάλυσης σε επόμενο στάδιο (παραγοντική ανάλυση και μοντελοποίηση συσχετίσεων). Οι τιμές του Cronbach'sAlpha ενισχύουν τη θεωρητική εγκυρότητα του ερωτηματολογίου και την καταλληλότητά του για εμπειρική διερεύνηση των στάσεων απέναντι στην Τεχνητή Νοημοσύνη στη διαχείριση κινδύνου.

4.3 Ανάλυση Συσχετίσεων μεταξύ Μεταβλητών

Η ενότητα αυτή έχει στόχο να εξετάσει τον τρόπο με τον οποίο οι βασικές εννοιολογικές μεταβλητές του προτεινόμενου θεωρητικού μοντέλου σχετίζονται μεταξύ τους. Η εξέταση των μεταξύ τους σχέσεων είναι θεμελιώδους σημασίας, καθώς αναδεικνύει τις πιθανές αιτιακές και λειτουργικές συσχετίσεις που ενδέχεται να υπάρχουν, πριν από την εφαρμογή πιο σύνθετων μεθοδολογικών τεχνικών, όπως η παραγοντική ή η διαρθρωτική εξίσωση.

Οι βασικές μεταβλητές είναι:

- Αντιληπτή Χρησιμότητα (PerceivedUsefulness – PU)
- Αντιληπτή Ευκολία Χρήσης (Perceived Ease of Use – PEOU)
- Εμπιστοσύνη (Trust – TR)
- Αντιληπτός Κίνδυνος (PerceivedRisk – PR)
- Πρόθεση Χρήσης (Behavioral Intention – BI)

Για τη διερεύνηση των σχέσεων μεταξύ αυτών των μεταβλητών, χρησιμοποιήθηκε ο συντελεστής συσχέτισης Pearson, ο οποίος προσδιορίζει τον βαθμό γραμμικής σχέσης μεταξύ δύο συνεχών μεταβλητών. Οι τιμές του Pearson κυμαίνονται μεταξύ -1 και +1, με:

- τιμές κοντά στο +1 να υποδεικνύουν ισχυρή θετική συσχέτιση,
- τιμές κοντά στο -1 να υποδεικνύουν ισχυρή αρνητική συσχέτιση,
- τιμές κοντά στο 0 να δείχνουν απουσία γραμμικής συσχέτισης.

Πίνακας 8: Πίνακας Συντελεστών Συσχέτισης Pearson (n = 75)

Μεταβλητή	**PU**	**PEOU**	**TR**	**PR**	**BI**
PU	1	.704**	.649**	–.453**	.717**
PEOU	.704**	1	.683**	–.421**	.691**
TR	.649**	.683**	1	–.534**	.635**
PR	–.453**	–.421**	–.534**	1	–.439**
BI	.717**	.691**	.635**	–.439**	1

Όλες οι συσχετίσεις είναι στατιστικά σημαντικές στο επίπεδο p< 0,01 (**p< 0,01**).Παρότι οι συσχετίσεις παρέχουν σημαντικές ενδείξεις για τη συνάφεια και τη συνδιακύμανση των μεταβλητών, θα πρέπει να επισημανθεί ότι η συσχέτιση δεν συνεπάγεται αιτιώδη σχέση. Ο συντελεστής Pearson αποτυπώνει τη γραμμική συνάφεια μεταξύ δύο μεταβλητών, χωρίς να τεκμηριώνει κατεύθυνση επίδρασης ή μηχανισμό αιτιότητας. Επομένως, οι ερμηνείες που ακολουθούν πρέπει να νοούνται ως ενδείξεις σχέσεων και όχι ως τεκμηριωμένες αιτιακές επιδράσεις.

Αναλυτική Ερμηνεία των Σχέσεων

1. **PU ↔ BI (r = 0.717):** Η ισχυρή θετική συσχέτιση επιβεβαιώνει ότι όσο πιο χρήσιμη θεωρείται η Τεχνητή Νοημοσύνη από τον χρήστη, τόσο αυξάνεται και η πρόθεση χρήσης της στο εργασιακό περιβάλλον. Πρόκειται για κεντρική σχέση του μοντέλου TAM (TechnologyAcceptanceModel), σύμφωνα με το οποίο η αντίληψη της χρησιμότητας είναι ο ισχυρότερος προβλεπτικός παράγοντας για τη συμπεριφορική πρόθεση.
2. **PEOU ↔ PU (r = 0.704):** Η ευκολία χρήσης σχετίζεται σημαντικά με την αντιληπτή χρησιμότητα. Αυτό σημαίνει ότι οι χρήστες που θεωρούν εύκολη τη χρήση της AI, είναι πιθανότερο να την αξιολογούν και ως πιο ωφέλιμη. Η σχέση αυτή είναι καθοριστική καθώς αποτυπώνει τη γνωσιακή ροή αποδοχής: πρώτα διαμορφώνεται η ευκολία και έπειτα η αντιληπτή ωφέλεια.
3. **TR ↔ BI (r = 0.635):** Όσο αυξάνεται η εμπιστοσύνη του χρήστη στα συστήματα AI – ως προς την αξιοπιστία και τη δικαιοσύνη τους– τόσο περισσότερο αυξάνεται και η πρόθεσή του να τα ενσωματώσει στην εργασία του. Αυτή η συσχέτιση είναι κρίσιμη σε κλάδους όπως οι χρηματοοικονομικές υπηρεσίες και η υγεία, όπου η εμπιστοσύνη αποτελεί προϋπόθεση εφαρμογής.
4. **PR ↔ BI (r = –0.439):** Παρατηρείται μέτρια αρνητική συσχέτιση. Όσο μεγαλύτερος είναι ο αντιληπτός κίνδυνος από τη χρήση της AI (π.χ. φόβος λαθών, απώλειας ελέγχου, ηθικά διλήμματα), τόσο μειώνεται η πρόθεση χρήσης. Η μεταβλητή PR λειτουργεί ως ανασταλτικός παράγοντας, και συνεπώς χρειάζεται αντιμετώπιση μέσω θεσμικών ή τεχνολογικών παρεμβάσεων.
5. **TR ↔ PR (r = –0.534):** Ισχυρή αρνητική συσχέτιση μεταξύ εμπιστοσύνης και κινδύνου. Οι συμμετέχοντες που εμπιστεύονται την τεχνολογία AI είναι λιγότερο πιθανό να τη θεωρούν επικίνδυνη ή απειλητική. Αυτό ενισχύει τη θέση ότι η εμπιστοσύνη λειτουργεί αντισταθμιστικά έναντι του φόβου.

6. **PEOU ↔ BI (r = 0.691):** Η ευκολία χρήσης επηρεάζει επίσης άμεσα την πρόθεση χρήσης. Το γεγονός ότι η διεπαφή ή το σύστημα είναι εύκολα κατανοήσιμα και δεν απαιτούν πολύπλοκη εκπαίδευση, ενθαρρύνει τη θετική στάση και την πρόθεση ενσωμάτωσης.

Η ανάλυση δείχνει ότι οι μεταβλητές του μοντέλου είναι αλληλένδετες και λειτουργικά συνεκτικές. Οι σχέσεις που αναδεικνύονται είναι:

- Θετικές ανάμεσα σε PU, PEOU, TR με BI.
- Αρνητική σχέση μεταξύ PR και όλων των άλλων μεταβλητών.
- Διαμεσολαβητικές δυναμικές μεταξύ PEOU → PU → BI, οι οποίες θα μπορούσαν να διερευνηθούν με μοντελοποίηση διαδρομής (pathanalysis) σε επόμενα στάδια.

Τα ευρήματα προσφέρουν σημαντική εμπειρική στήριξη για τις υποθέσεις του θεωρητικού μοντέλου, ενώ παράλληλα υποδεικνύουν ποια σημεία μπορούν να αποτελέσουν στόχο παρέμβασης (π.χ. ενίσχυση εμπιστοσύνης, μείωση κινδύνου) για τη μεγιστοποίηση της πρόθεσης αποδοχής και χρήσης της AI.

4.4 Σύνοψη Αποτελεσμάτων

Η παρούσα ενότητα συνοψίζει τα βασικά ευρήματα της περιγραφικής και συσχετιστικής ανάλυσης, παρέχοντας μια αναλυτική εικόνα των στάσεων, των προσδοκιών και των επιφυλάξεων των συμμετεχόντων απέναντι στην Τεχνητή Νοημοσύνη (AI) στον τομέα της διαχείρισης κινδύνου. Η έρευνα βασίστηκε σε δείγμα 75 ατόμων, οι οποίοι αξιολόγησαν πέντε βασικές μεταβλητές του θεωρητικού πλαισίου: την Αντιληπτή Χρησιμότητα (Perceived Usefulness – PU), την Αντιληπτή Ευκολία Χρήσης (Perceived Ease of Use – PEOU), την Εμπιστοσύνη (Trust – TR), τον Αντιληπτό Κίνδυνο (Perceived Risk – PR) και την Πρόθεση Χρήσης (Behavioral Intention – BI).

Η γενική στάση των συμμετεχόντων απέναντι στην τεχνολογία της AI είναι θετική. Η μεταβλητή PU παρουσίασε μέσο όρο 3,99, ένδειξη ότι η πλειοψηφία των συμμετεχόντων εκτιμά πως η χρήση της AI βελτιώνει την απόδοση, την ποιότητα και την αποτελεσματικότητα της εργασίας τους. Ομοίως, η μεταβλητή PEOU εμφάνισε μέσο όρο 3,93, επιβεβαιώνοντας ότι οι συμμετέχοντες θεωρούν πως η αλληλεπίδραση με τα συστήματα AI είναι κατανοητή και η εκμάθησή τους σχετικά εύκολη, χωρίς ιδιαίτερες απαιτήσεις ή δυσκολίες. Η Εμπιστοσύνη (TR) σημείωσε μέσο όρο 3,72, το οποίο υποδηλώνει μια μέτρια προς θετική στάση, με ορισμένες επιφυλάξεις αναφορικά με την αξιοπιστία και τη διαφάνεια των αλγοριθμικών αποφάσεων. Ο Αντιληπτός Κίνδυνος (PR) καταγράφηκε σε επίπεδο 2,95,

φανερώνοντας ότι οι περισσότεροι συμμετέχοντες δεν νιώθουν έντονη απειλή από τη χρήση της AI, αν και υπάρχουν επιμέρους ανησυχίες για κρίσιμες εφαρμογές. Τέλος, η Πρόθεση Χρήσης (BI) παρουσίασε τον υψηλότερο μέσο όρο (4,29), ένδειξη της ισχυρής πρόθεσης των συμμετεχόντων να ενσωματώσουν την AI στο επαγγελματικό τους περιβάλλον.

Η ανάλυση συσχετίσεων με δείκτη Pearson αποκάλυψε ισχυρές στατιστικά σημαντικές σχέσεις μεταξύ των βασικών μεταβλητών. Η ισχυρότερη συσχέτιση εντοπίστηκε μεταξύ της Αντιληπτής Χρησιμότητας (PU) και της Πρόθεσης Χρήσης (BI), με τιμή r = 0,717. Το αποτέλεσμα αυτό επιβεβαιώνει τη βασική αρχή των μοντέλων αποδοχής τεχνολογίας (Technology Acceptance Models – TAM), σύμφωνα με την οποία η αντίληψη περί χρησιμότητας αποτελεί τον σημαντικότερο προβλεπτικό παράγοντα της πρόθεσης για υιοθέτηση. Εξίσου σημαντική ήταν η συσχέτιση μεταξύ της Ευκολίας Χρήσης (PEOU) και της PU (r = 0,704), καθώς και μεταξύ της PEOU και της BI (r = 0,691), γεγονός που αποδεικνύει ότι η ευκολία χρήσης επηρεάζει θετικά τόσο την αντίληψη χρησιμότητας όσο και την τελική πρόθεση.

Η Εμπιστοσύνη (TR) βρέθηκε να σχετίζεται θετικά με όλες τις υπόλοιπες μεταβλητές, με ισχυρότερη σχέση με την BI (r = 0,635), ενώ ταυτόχρονα εμφάνισε σημαντική αρνητική συσχέτιση με τον Αντιληπτό Κίνδυνο (PR) (r = –0,534), υποδεικνύοντας ότι όταν αυξάνεται η εμπιστοσύνη στην AI, μειώνεται η αίσθηση κινδύνου. Η PR, από την άλλη πλευρά, λειτουργεί ως αρνητικός προβλεπτικός παράγοντας, αφού σχετίζεται αρνητικά με όλες τις μεταβλητές αποδοχής, με εντονότερη την αρνητική συσχέτιση με την BI (r = –0,439), γεγονός που φανερώνει ότι οι ανησυχίες για πιθανά σφάλματα ή απώλεια ελέγχου μπορεί να περιορίσουν τη διάθεση χρήσης της τεχνολογίας.

Τα παραπάνω αποτελέσματα ενισχύουν την εγκυρότητα του προτεινόμενου θεωρητικού μοντέλου, επιβεβαιώνοντας τις θεωρητικές σχέσεις που προβλέπονται στη διεθνή βιβλιογραφία περί τεχνολογικής αποδοχής. Η PU, η PEOU και η TR λειτουργούν ως ενισχυτικοί παράγοντες της BI, ενώ ο PR ως αποτρεπτικός. Η PEOU φαίνεται να έχει έμμεση επίδραση στην BI μέσω της PU, επιτελώντας ρόλο διαμεσολαβητή, ενώ η TR φαίνεται να επιδρά θετικά σε όλο το σύστημα σχέσεων, ενισχύοντας την πρόθεση χρήσης και ταυτόχρονα μειώνοντας τον αντιληπτό κίνδυνο.

Σε επίπεδο εφαρμογής, τα ευρήματα υποδεικνύουν ότι οι οργανισμοί θα πρέπει να επενδύσουν σε ενέργειες που ενισχύουν την αντίληψη χρησιμότητας και ευκολίας, όπως εκπαιδευτικά προγράμματα και φιλικές προς τον χρήστη διεπαφές. Επίσης, η ενίσχυση της εμπιστοσύνης μέσω της διαφάνειας, της εξήγησης των αλγορίθμων και της ηθικής λογοδοσίας μπορεί να συμβάλει καθοριστικά στην αποδοχή της AI. Τέλος, η μείωση των

ανησυχιών μέσα από σαφή κανονιστικά πλαίσια και πρακτικές εξηγησιμότητας (Explainable AI – XAI) αναδεικνύεται ως κρίσιμος παράγοντας για τη μελλοντική ενσωμάτωση της τεχνολογίας.

Συνοψίζοντας, το εμπειρικό μοντέλο αποδοχής της Τεχνητής Νοημοσύνης στη διαχείριση κινδύνου επιβεβαιώθηκε ως ισχυρό και ερμηνεύσιμο, προσφέροντας σημαντικά ερευνητικά και πρακτικά συμπεράσματα. Οι στάσεις των συμμετεχόντων είναι θετικές, η πρόθεση έντονα εκπεφρασμένη, και οι μεταξύ των μεταβλητών σχέσεις συμβατές με τη θεωρία, προσδίδοντας κύρος και εγκυρότητα στο συνολικό αναλυτικό πλαίσιο. Στο επόμενο κεφάλαιο (Κεφάλαιο 5), τα αποτελέσματα θα συγκριθούν με την υπάρχουσα βιβλιογραφία, και θα παρουσιαστούν οι συνέπειες για τη θεωρία και την πράξη, καθώς και οι προτάσεις για μελλοντική έρευνα.

Κεφάλαιο 5: Συγκριτική Ανάλυση και Κριτική

5.1 Εισαγωγή στη Συγκριτική Ανάλυση

Η ραγδαία ανάπτυξη των τεχνολογιών Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) έχει μεταμορφώσει το τοπίο της διαχείρισης κινδύνου, τόσο σε επίπεδο εφαρμογών όσο Στο παρόν κεφάλαιο επιχειρείται μια εμβάθυνση και κριτική αποτίμηση των ευρημάτων του εμπειρικού μέρους (Κεφ. 4) υπό το φως της σύγχρονης διεθνούς βιβλιογραφίας. Η ανάλυση επικεντρώνεται στη σύγκριση των εμπειρικών αποτελεσμάτων ως προς την αποδοχή της Τεχνητής Νοημοσύνης στη διαχείριση κινδύνου με ανάλογες μελέτες και εφαρμογές, καταγράφοντας συγκλίσεις και αποκλίσεις. Στόχος είναι να εντοπιστούν οι βασικές μεθοδολογικές, τεχνολογικές και δεοντολογικές τάσεις που επικρατούν στο πεδίο, να ερμηνευθούν οι παρατηρούμενες συσχετίσεις και να αναδειχθούν οι θεωρητικές και πρακτικές προεκτάσεις της έρευνας. Παράλληλα, το κεφάλαιο προτείνει κατευθύνσεις για μελλοντική έρευνα, εστιάζοντας στις ανάγκες για μεγαλύτερη διαφάνεια, προτυποποίηση και ηθική ενσωμάτωση της AI στη διαχείριση κινδύνου.

5.2 Σύγκριση Εφαρμογών AI ανά Κατηγορία Κινδύνου

Η εφαρμογή της Τεχνητής Νοημοσύνης (AI) και της Μηχανικής Μάθησης (ML) στη διαχείριση κινδύνου παρουσιάζει διαφορετική δυναμική ανά κατηγορία κινδύνου, τόσο ως προς τη μεθοδολογική ωριμότητα όσο και ως προς την αποδοχή από τους χρήστες. Στο παρόν υποκεφάλαιο, εξετάζεται η σύγκριση τεσσάρων βασικών κατηγοριών: πιστωτικός κίνδυνος, κίνδυνος απάτης, συστημικός χρηματοοικονομικός κίνδυνος και κυβερνοκίνδυνος, ενσωματώνοντας τόσο τις θεωρητικές προσεγγίσεις από τη βιβλιογραφία όσο και τα εμπειρικά αποτελέσματα της παρούσας μελέτης.

Πιστωτικός κίνδυνος

Η διαχείριση πιστωτικού κινδύνου αποτελεί ένα από τα πιο ώριμα πεδία εφαρμογής της AI, με έμφαση στη χρήση εποπτευόμενων τεχνικών όπως logisticregression, decisiontrees και supportvectormachines (Baesensetal., 2003· Lessmannetal., 2015). Η βιβλιογραφία επισημαίνει ότι η επάρκεια ιστορικών δεδομένων και οι ρυθμιστικές απαιτήσεις ενισχύουν τη χρήση μεθόδων με υψηλό βαθμό ερμηνευσιμότητας, όπως εργαλεία explainableAI (Ribeiroetal., 2016· Lundberg&Lee, 2017). Τα εμπειρικά ευρήματα της παρούσας μελέτης καταδεικνύουν υψηλή αντιληπτή χρησιμότητα (PU = 3,99) και την υψηλότερη πρόθεση

χρήσης (BI = 4,29) συγκριτικά με τις λοιπές κατηγορίες κινδύνου. Αν και η έρευνα δεν διερεύνησε άμεσα παράγοντες όπως η ερμηνευσιμότητα ή η επάρκεια δεδομένων, η αυξημένη πρόθεση υιοθέτησης συνάδει με τη γενικότερη ωριμότητα που καταγράφεται στη διεθνή βιβλιογραφία για τις εφαρμογές AI στην πιστοληπτική αξιολόγηση.

Κίνδυνος απάτης

Η ανίχνευση απάτης συνδέεται με δεδομένα υψηλής ανισορροπίας και συνεχή μεταβολή των μοτίβων απάτης, απαιτώντας συνδυασμό εποπτευόμενων και μη εποπτευόμενων τεχνικών, όπως νευρωνικά δίκτυα, anomalydetection και sequenceclassification (Phuaetal., 2010· Bahnsenetal., 2016· Jurgovskyetal., 2018). Στην παρούσα έρευνα, η εμπιστοσύνη (TR) καταγράφηκε σε μέτριο επίπεδο (M.O. = 3,72), ενώ το αντιληπτό ρίσκο (PR) ήταν σχετικά χαμηλό (M.O. = 2,95), υποδηλώνοντας γενικά θετική αλλά συγκρατημένη στάση απέναντι στη χρήση AI στον τομέα αυτό. Τα ευρήματα αυτά είναι συμβατά με τη βιβλιογραφία που αναγνωρίζει τη σημαντική επιχειρησιακή αξία των συστημάτων ανίχνευσης απάτης, χωρίς ωστόσο να αγνοεί τις τεχνικές και οργανωσιακές προκλήσεις που αυτά συνεπάγονται (Ngaietal., 2011).

Συστημικός χρηματοοικονομικός κίνδυνος

Η πρόβλεψη και μοντελοποίηση του συστημικού κινδύνου βασίζεται σε agent-based simulations, graph neural networks και συνθετικά οικονομικά μοντέλα (Battiston et al., 2012· Bookstaber et al., 2018· Adrian & Brunnermeier, 2016). Παρά τη σημασία της, η εμπειρική χρήση της AI στον τομέα αυτό παραμένει περιορισμένη λόγω της πολυπλοκότητας και της απαιτητικής υποδομής. Το μέτριο επίπεδο εμπιστοσύνης και η σχετικά συγκρατημένη πρόθεση χρήσης που καταγράφηκε από τους συμμετέχοντες μπορεί να αντανακλά αυτή την απόσταση από την καθημερινή εφαρμογή. Παράλληλα, η χρήση της AI εδώ σχετίζεται κυρίως με την εργασία ρυθμιστικών αρχών και όχι ιδιωτικών οργανισμών (Billio et al., 2012· BIS, 2019).

Κυβερνο-κίνδυνος

Η AI στην κυβερνοασφάλεια χρησιμοποιείται ευρέως για anomaly detection, intrusion detection systems και αναλύσεις πραγματικού χρόνου, βασισμένες κυρίως σε deep learning και reinforcement learning τεχνικές (Buczak & Guven, 2016· Kim et al., 2016· Sabottke et al., 2015). Αν και η φύση των δεδομένων είναι κατάλληλη για advanced AI τεχνικές, οι συμμετέχοντες στην έρευνα εμφάνισαν σχετικά υψηλή αντιληπτή ευκολία χρήσης (PEOU =

3,93) αλλά μέτρια εμπιστοσύνη. Αυτό ίσως αντικατοπτρίζει την τεχνική πολυπλοκότητα των συστημάτων, αλλά και την περιορισμένη εμπειρία ορισμένων χρηστών με real-time εφαρμογές κυβερνοάμυνας (Ali et al., 2019· Ghosh et al., 2022).

Συγκριτική σύνθεση

Τα εμπειρικά δεδομένα αποτυπώνουν μια ιεράρχηση αποδοχής, με πρώτη την AI για πιστοληπτική αξιολόγηση, δεύτερη την ανίχνευση απάτης, και τρίτες τη συστημική ανάλυση και κυβερνοασφάλεια. Το εύρημα συνάδει με το επίπεδο εμπορικής ωριμότητας κάθε εφαρμογής, όπως καταγράφεται στη διεθνή βιβλιογραφία (Gupta & Pathak, 2022· Wamba-Taguimdje et al., 2020). Η μελέτη υποδεικνύει ότι η διαδρομή αποδοχής της AI περνάει πρώτα από περιοχές όπου η ερμηνευσιμότητα, η σταθερότητα των δεδομένων και η κανονιστική συμμόρφωση είναι εφικτές, και έπειτα μεταβαίνει σε πιο σύνθετα και τεχνικά πεδία όπως η κυβερνοασφάλεια ή η προληπτική συστημική ανάλυση.

Παρακάτω παρατίθεται ο Πίνακας Σύγκρισης Θεωρητικής και Εμπειρικής Αποδοχής της AI ανά Κατηγορία Κινδύνου, όπως προκύπτει από την ανασκόπηση της βιβλιογραφίας και τα ευρήματα του Κεφαλαίου 4:

Πίνακας 9: Παράθεση Θεωρητικής Ωριμότητας και Εμπειρικών Αντιλήψεων ανά Κατηγορία Κινδύνου

Κατηγορία Κινδύνου	**Θεωρητική Ωριμότητα (Βιβλιογραφία)**	**Εμπειρική Αποδοχή (Από έρευνα)**	**Σχόλια**
Πιστωτικός Κίνδυνος	Υψηλή. Χρήση supervisedlearning με υψηλή ακρίβεια και ερμηνευσιμότητα (Baesensetal., 2003· Lessmannetal., 2015).	Πολύ υψηλή. PU = 3,99/ BI = 4,29	Παρατηρείται υψηλή πρόθεση χρήσης σε κατηγορία όπου η βιβλιογραφία καταγράφει εκτενή εφαρμογή τεχνικών AI. Η μελέτη δεν διερεύνησε άμεσα αντιλήψεις ωριμότητας ή ερμηνευσιμότητας.
Κίνδυνος Απάτης	Μέτρια-υψηλή. Απαιτεί unsupervised/ensemble τεχνικές για real-time detection (Phua et al., 2010· Bahnsen et al., 2016).	Μέση προς υψηλή. TR = 3,72 /PR = 2,95	Καταγράφεται θετική αλλά συγκρατημένη στάση, με σχετικά χαμηλή αντίληψη κινδύνου. Δεν εξετάστηκε ο βαθμός αυτονομίας των συστημάτων.
Συστημικός Κίνδυνος	Χαμηλή προς μέση. Σε ερευνητικό στάδιο με χρήση graphmodels και simulations (Adrian&Brunnermeier, 2016).	Μέτρια. TR = 3,72 PU&BI ελαφρώς χαμηλότερα σε σύγκριση με άλλες	Η πρόθεση χρήσης εμφανίζεται συγκριτικά χαμηλότερη. Η έρευνα δεν αξιολόγησε την εξοικείωση των συμμετεχόντων με προηγμένα μοντέλα δικτύων ή προσομοιώσεων.

		κατηγορίες.	
Κυβερνο-Κίνδυνος	Μέτρια. Χρήση deeplearning και NLP, ιδιαίτερη έμφαση στην ανίχνευση ανωμαλιών (Kimetal., 2016· Buczak&Guven, 2016).	Μέση προς υψηλή. PEOU = 3,93/ TR = 3,72	Καταγράφεται σχετικά υψηλή αντιληπτή ευκολία χρήσης, χωρίς όμως αντίστοιχη αύξηση στην εμπιστοσύνη. Δεν διερευνήθηκε η εμπειρία των συμμετεχόντων με real-time συστήματα.

Ο πίνακας παρουσιάζει συγκριτικά τη βιβλιογραφική αποτύπωση της ωριμότητας εφαρμογών AI και τις αντιλήψεις των συμμετεχόντων της παρούσας έρευνας. Η σύγκριση αυτή δεν υποδηλώνει αιτιώδη σχέση μεταξύ θεωρητικής ωριμότητας και εμπειρικής αποδοχής, καθώς η παρούσα μελέτη δεν μέτρησε άμεσα την αντίληψη ωριμότητας των εφαρμογών. Αντίθετα, λειτουργεί ως ερμηνευτικό πλαίσιο συσχέτισης, το οποίο επιτρέπει τη συζήτηση πιθανών τάσεων σύγκλισης ή απόκλισης μεταξύ βιβλιογραφίας και εμπειρικών ευρημάτων.

5.3 Σύγκριση Μεθοδολογικών Προσεγγίσεων: Θεωρία και Εμπειρική Προοπτική

Η διαρκής εξέλιξη των τεχνικών Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) έχει δημιουργήσει ένα πλούσιο φάσμα μεθοδολογικών επιλογών για την ανάλυση και πρόβλεψη χρηματοοικονομικών και λειτουργικών κινδύνων. Η επιλογή των κατάλληλων μεθόδων εξαρτάται όχι μόνο από τη φύση του προβλήματος και των δεδομένων, αλλά και από τις απαιτήσεις των χρηστών – μια διάσταση που αναδείχθηκε και από τα αποτελέσματα της παρούσας έρευνας.

Οι συμμετέχοντες στη μελέτη, όπως έδειξαν οι τιμές υψηλής Αντιληπτής Χρησιμότητας (PU=3,99) και Πρόθεσης Χρήσης (BI=4,29), φάνηκαν να ευνοούν μεθόδους που προσφέρουν υψηλή απόδοση και αξιοπιστία, αρκεί να συνοδεύονται από ένα αποδεκτό επίπεδο ερμηνευσιμότητας. Η στάση αυτή ευθυγραμμίζεται με την κυρίαρχη χρήση εποπτευόμενων τεχνικών (supervised learning) στην πιστοληπτική αξιολόγηση και στην ανίχνευση απάτης, όπου ιστορικά επισημασμένα δεδομένα επιτρέπουν την εφαρμογή αλγορίθμων όπως decision trees, logistic regression και support vector machines (Baesens et al., 2003· Lessmann et al., 2015).

Παρότι η εμπειρική ανάλυση έδειξε μέση προς υψηλή εμπιστοσύνη στην AI (TR=3,72), παρατηρείται μια σχετική επιφύλαξη έναντι black-box μοντέλων. Αυτό ενισχύει την ανάγκη για Explainable AI (XAI) εργαλεία, όπως SHAP και LIME (Lundberg & Lee, 2017), τα οποία επιτρέπουν μεγαλύτερη διαφάνεια χωρίς να θυσιάζεται σημαντικά η απόδοση – μια τάση που

επιβεβαιώνεται από την αυξημένη χρήση τους σε ρυθμιζόμενους τομείς όπως ο τραπεζικός κλάδος.

Σε πεδία με υψηλή αβεβαιότητα ή ανεπαρκή ετικετοποίηση (π.χ. κυβερνο-κίνδυνος), η χρήση μη εποπτευόμενων μεθόδων (unsupervisedlearning) θεωρείται κατάλληλη στη βιβλιογραφία. Τεχνικές όπως K-means, PCA και SOM εφαρμόζονται για την ανίχνευση ανωμαλιών σε ροές μεγάλου όγκου δεδομένων (Phuaetal., 2010). Ωστόσο, τα εμπειρικά δεδομένα της παρούσας μελέτης δείχνουν ότι η εμπιστοσύνη προς τα συστήματα AI διατηρείται σε μέτριο επίπεδο (TR=3,72), γεγονός που υποδηλώνει μια συγκρατημένη στάση απέναντι στη χρήση τους σε σύνθετα και δυναμικά περιβάλλοντα. Η έρευνα δεν διερεύνησε ειδικά χαρακτηριστικά των επιμέρους μεθόδων (π.χ. ερμηνευσιμότητα ή βαθμό αυτοματοποίησης), συνεπώς οποιαδήποτε σύνδεση μεταξύ τεχνικού τύπου μοντέλου και επιπέδου εμπιστοσύνης πρέπει να αντιμετωπίζεται με επιφύλαξη.

Η ανάλυση του συστημικού κινδύνου εισάγει πιο σύνθετες μεθοδολογίες, όπως GraphNeuralNetworks (GNNs) για την κατανόηση της διάχυσης ρίσκου μέσω χρηματοοικονομικών δικτύων (Zhangetal., 2022), και μοντελοποίηση βασισμένη σε πράκτορες (Bookstaberetal., 2018), η οποία επιτρέπει προσομοιώσεις πολύπλοκης συμπεριφοράς. Στην παρούσα έρευνα, οι συμμετέχοντες δεν εξέφρασαν ιδιαίτερα υψηλά επίπεδα Αντιληπτής Χρησιμότητας ή Εμπιστοσύνης για εφαρμογές AI στον συστημικό κίνδυνο, καταγράφοντας συγκριτικά πιο ουδέτερες στάσεις. Η έρευνα δεν διερεύνησε την εξοικείωση των ερωτώμενων με τις συγκεκριμένες μεθοδολογίες, συνεπώς η παρατηρούμενη ουδετερότητα μπορεί να αντανακλά είτε περιορισμένη εμπειρία είτε γενικότερη επιφύλαξη απέναντι σε λιγότερο άμεσα εφαρμοζόμενα πεδία.

Στον κυβερνοχώρο, η ανάγκη για real-time απόκριση ενισχύει τη χρήση deep learning και reinforcement learning τεχνικών (Kim et al., 2016· Deng et al., 2016). Ωστόσο, παρά την τεχνική υπεροχή τους, η εμπειρική μέτρηση του Αντιληπτού Κινδύνου (PR=2,95) δείχνει ότι η αποδοχή εξαρτάται από την ισορροπία μεταξύ αυτοματοποίησης και ανθρώπινης παρέμβασης – ένα κρίσιμο ζήτημα για τις επιχειρήσεις.

Η ανάλυση των βιβλιογραφικών προσεγγίσεων σε συνδυασμό με τα εμπειρικά ευρήματα της παρούσας μελέτης υποδηλώνει ότι η μεθοδολογική επιλογή στην πράξη δεν αποτελεί αποκλειστικά τεχνική απόφαση. Τα επίπεδα εμπιστοσύνης (TR) και αντιληπτού κινδύνου (PR) καταδεικνύουν ότι η αποδοχή της AI επηρεάζεται από ψυχολογικούς και οργανωσιακούς παράγοντες, πέραν της καθαρά υπολογιστικής απόδοσης των μοντέλων. Ωστόσο, η παρούσα έρευνα δεν διερεύνησε άμεσα προτιμήσεις ως προς συγκεκριμένους τύπους μοντέλων ή συνδυαστικές μεθοδολογίες. Συνεπώς, οποιαδήποτε αναφορά σε υβριδικές ή πολυεπίπεδες

προσεγγίσεις στηρίζεται κυρίως στη διεθνή βιβλιογραφία και όχι σε άμεση εμπειρική μέτρηση της παρούσας μελέτης.

Συνολικά, η ενσωμάτωση των εμπειρικών δεδομένων ενισχύει την κατανόηση της μεθοδολογικής στρατηγικής ως κοινωνικά ενσυνείδητης επιλογής και όχι ως τεχνοκρατικής απόφασης, αποκαλύπτοντας την ανάγκη για προσέγγιση που συνδυάζει απόδοση, ερμηνευσιμότητα και αποδοχή από τον τελικό χρήστη.

5.4 Διασταύρωση Τεχνικών και Ρυθμιστικού/Ηθικού Πλαισίου

Η ενσωμάτωση Τεχνητής Νοημοσύνης (AI) στη διαχείριση χρηματοοικονομικού κινδύνου εγείρει όχι μόνο τεχνικά, αλλά και βαθύτατα θεσμικά και ηθικά ζητήματα. Η υιοθέτηση αλγοριθμικών αποφάσεων σε πεδία που αφορούν άμεσα τα δικαιώματα των πολιτών, τη σταθερότητα των αγορών και την εμπιστοσύνη στους θεσμούς απαιτεί εναρμόνιση με ρυθμιστικά πρότυπα, κανονιστικές υποχρεώσεις και αρχές δεοντολογίας (Floridi et al., 2018· Crawford, 2021).

Η εμπειρική έρευνα δείχνει ότι οι επαγγελματίες του χώρου έχουν ήδη θετική προδιάθεση απέναντι στη χρήση AI, με μέσο όρο 4,29 στην Πρόθεση Χρήσης (BI), στοιχείο που υποδηλώνει πρακτική ετοιμότητα για τεχνολογική ενσωμάτωση. Ωστόσο, ο αντιληπτός κίνδυνος (PR = 2,95) και η μέτρια τιμή εμπιστοσύνης (TR = 3,72) αποκαλύπτουν υπαρκτές ανησυχίες για τον τρόπο με τον οποίο τα συστήματα λαμβάνουν αποφάσεις, ιδίως όταν πρόκειται για κρίσιμους τομείς όπως η πιστοληπτική αξιολόγηση.

Σύμφωνα με τον Κανονισμό GDPR και ειδικά το άρθρο 22, η χρήση αυτοματοποιημένων αποφάσεων υπόκειται σε περιορισμούς και πρέπει να συνοδεύεται από δυνατότητα επεξήγησης. Αυτό καθιστά τις Explainable AI (XAI) τεχνικές όχι απλώς επιθυμητές αλλά υποχρεωτικές σε εφαρμογές που επηρεάζουν άμεσα την πρόσβαση σε χρηματοδότηση (Wachter et al., 2017· Lundberg & Lee, 2017). Οι συμμετέχοντες της έρευνας που ανέφεραν μικρή εξοικείωση με τις τεχνικές XAI, εκφράζουν έμμεσα την ανάγκη για ενίσχυση της κανονιστικής λογοδοσίας μέσω διαφάνειας, ιδίως στα πρώτα στάδια υιοθέτησης των συστημάτων.

Αντίθετα, στους τομείς της απάτης και της κυβερνοασφάλειας, οι συμμετέχοντες ανέφεραν υψηλό βαθμό αντιληπτής χρησιμότητας των τεχνολογιών AI, στοιχείο που συνδέεται με την επιχειρησιακή ανάγκη για real-time απόκριση και με μικρότερη ρυθμιστική πίεση. Στις περιπτώσεις αυτές, η χρήση βαθιών νευρωνικών δικτύων και black-box μοντέλων είναι συχνότερη, αλλά ταυτόχρονα διατρέχει τον κίνδυνο υποβάθμισης θεμελιωδών αξιών όπως η

αμεροληψία και η διαφάνεια (Brundage et al., 2018· Binns, 2018). Παρότι οι συμμετέχοντες ανέφεραν χαμηλή ευαισθητοποίηση σε θέματα ηθικής, η αποδοχή τέτοιων τεχνολογιών προϋποθέτει, μακροπρόθεσμα, ενσωμάτωση δικλείδων ελέγχου.

Αξιοσημείωτη είναι η περίπτωση του συστημικού κινδύνου, όπου τα μοντέλα AI χρησιμοποιούνται κυρίως από ρυθμιστικούς φορείς (BIS, ECB, ESRB), χωρίς να συνοδεύονται από σαφή θεσμική πλαισίωση. Οι τεχνικές όπως τα Graph Neural Networks (Zhang et al., 2022) και οι προσομοιώσεις agent-based modeling (Bookstaber et al., 2018) προσφέρουν σημαντική αναλυτική ισχύ, αλλά δεν εμπεριέχουν από μόνες τους ερμηνευσιμότητα ή κανονιστική συμμόρφωση. Οι συμμετέχοντες της μελέτης, παρά τη θετική τους στάση, ανέφεραν σε μεγάλο βαθμό άγνοια ή αβεβαιότητα για το πώς συνδέονται τα τεχνικά εργαλεία με τις ρυθμιστικές απαιτήσεις. Αυτό επιβεβαιώνει το θεσμικό έλλειμμα που εντοπίζεται στη σχετική βιβλιογραφία (FSB, 2019· Morley et al., 2020).

Η διασταύρωση τεχνικής και δεοντολογίας γίνεται ακόμη πιο εμφανής στις ανησυχίες για μεροληψία (bias) και διακρίσεις. Παρότι οι συμμετέχοντες δεν βαθμολόγησαν ιδιαίτερα υψηλά τον αντιληπτό ηθικό κίνδυνο, η έλλειψη επίγνωσης μπορεί να οδηγήσει σε εφησυχασμό. Όπως έχει τεκμηριώσει η βιβλιογραφία, η χρήση δεδομένων που αναπαράγουν ιστορικές ανισότητες οδηγεί σε «σιωπηρές διακρίσεις» (Barocas & Selbst, 2016· Binns, 2018). Για τον λόγο αυτό, εργαλεία ελέγχου fairness και auditing πρέπει να ενταχθούν τόσο στο τεχνικό όσο και στο κανονιστικό επίπεδο εφαρμογής.

Οι διεθνείς πρωτοβουλίες ρύθμισης, όπως ο AI Act της Ευρωπαϊκής Ένωσης (European Commission, 2021), τα ISO/IEC πρότυπα (ISO, 2023) και τα ηθικά πλαίσια του OECD, δείχνουν ότι το ζήτημα της υπεύθυνης χρήσης AI δεν είναι πλέον προαιρετικό αλλά επιτακτικό. Οι συμμετέχοντες στην έρευνα, μέσω της θετικής στάσης τους, δείχνουν έτοιμοι να συμμετάσχουν σε αυτό το νέο πλαίσιο, εφόσον τους δοθούν κατάλληλα εργαλεία, εκπαίδευση και νομική καθοδήγηση.

Συνολικά, τα εμπειρικά ευρήματα δείχνουν ότι, παρότι η πρόθεση χρήσης συστημάτων AI είναι ιδιαίτερα υψηλή (BI = 4,29), η εμπιστοσύνη στα συστήματα παραμένει σε μέτριο επίπεδο (TR = 3,72) και ο αντιληπτός κίνδυνος δεν είναι αμελητέος (PR = 2,95). Ο συνδυασμός υψηλής πρόθεσης και ταυτόχρονης επιφυλακτικότητας υποδηλώνει ότι η τεχνική αποτελεσματικότητα από μόνη της δεν αρκεί για την πλήρη αποδοχή των συστημάτων. Οι επαγγελματίες φαίνεται να είναι θετικοί απέναντι στη χρήση AI, αλλά η στάση τους διαμορφώνεται και από ζητήματα εμπιστοσύνης, ασφάλειας και λογοδοσίας. Υπό αυτή την έννοια, τα ευρήματα ενισχύουν τη θεωρητική θέση ότι η επιτυχής ενσωμάτωση της AI προϋποθέτει θεσμικό και κανονιστικό πλαίσιο που να διασφαλίζει διαφάνεια,

επεξηγησιμότητα και υπεύθυνη χρήση. Η πρόκληση, επομένως, δεν περιορίζεται στην πρόβλεψη του κινδύνου με υψηλή ακρίβεια, αλλά επεκτείνεται στη διαχείρισή του με τρόπο που να είναι θεσμικά νομιμοποιημένος και κοινωνικά αποδεκτός.

Κεφάλαιο 6: Συμπεράσματα και Σύνθεση

Η παρούσα εργασία επιδίωξε τη συστηματική αποτύπωση του τρόπου με τον οποίο οι τεχνολογίες Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) εφαρμόζονται στη διαχείριση κινδύνου, με έμφαση σε τέσσερις βασικές κατηγορίες: πιστωτικό κίνδυνο, κίνδυνο απάτης, συστημικό χρηματοοικονομικό κίνδυνο και κυβερνο-κίνδυνο. Η ερευνητική προσέγγιση συνδύασε θεματική βιβλιογραφική ανασκόπηση υψηλής ποιότητας με πρωτογενή εμπειρική έρευνα, παρέχοντας έτσι μια σπάνια και ολοκληρωμένη οπτική για ένα δυναμικά εξελισσόμενο πεδίο.

Η βιβλιογραφική ανασκόπηση αποκάλυψε σημαντικές τεχνικές διαφοροποιήσεις μεταξύ των κατηγοριών κινδύνου, κυρίως ως προς τη φύση των δεδομένων, την ερμηνευσιμότητα των μοντέλων και τη ρυθμιστική πίεση. Στην πιστωτική αξιολόγηση, επικρατούν εποπτευόμενες τεχνικές (logistic regression, decision trees, SVM), που αξιοποιούν ιστορικά δεδομένα δανειοληπτών με υψηλή ετικετοποίηση, ενώ στον κυβερνο-κίνδυνο και στην απάτη κυριαρχούν τεχνικές ανίχνευσης ανωμαλιών, μη εποπτευόμενης μάθησης και deep learning. Στον συστημικό κίνδυνο, οι τεχνικές βασίζονται κυρίως σε γραφηματικά δίκτυα (GNNs) και μοντελοποίηση πολλαπλών πρακτόρων (ABM), όπου δεν ενδιαφέρει μόνο η πρόβλεψη ενός αποτελέσματος αλλά και η κατανόηση της διάχυσης των διαταραχών.

Ωστόσο, πέρα από τις τεχνικές διαφορές, η παρούσα εργασία προχώρησε και σε μια εμπειρική έρευνα με 75 επαγγελματίες που δραστηριοποιούνται στη διαχείριση κινδύνου, ώστε να αποτιμήσει τις πραγματικές τους στάσεις απέναντι στην ενσωμάτωση της AI στις διαδικασίες τους. Η εμπειρική ανάλυση βασίστηκε σε ένα θεωρητικό υπόδειγμα που ενσωματώνει μεταβλητές όπως η Αντιληπτή Χρησιμότητα (PU), η Αντιληπτή Ευκολία Χρήσης (PEOU), η Εμπιστοσύνη (TR), ο Αντιληπτός Κίνδυνος (PR) και η Πρόθεση Χρήσης (BI), μετρούμενες μέσω πενταβάθμιας κλίμακας Likert.

Τα αποτελέσματα της περιγραφικής ανάλυσης υποδεικνύουν υψηλό επίπεδο αποδοχής της AI από τους συμμετέχοντες, με ιδιαίτερα αυξημένη τιμή στην Πρόθεση Χρήσης (BI: M = 4,29, SD = 0,77), κάτι που αντανακλά ισχυρή διάθεση ενσωμάτωσης της τεχνολογίας στο επαγγελματικό περιβάλλον. Η Αντιληπτή Χρησιμότητα (PU: M = 3,99) καταδεικνύει ότι οι χρήστες θεωρούν την AI ως σημαντικό ενισχυτή απόδοσης και ποιότητας εργασίας. Παράλληλα, η Αντιληπτή Ευκολία Χρήσης (PEOU: M = 3,93) επιβεβαιώνει την τεχνολογική εξοικείωση των επαγγελματιών, αν και εντοπίζεται ελαφρώς χαμηλότερη εμπιστοσύνη (TR: M = 3,72), που σχετίζεται με την αντιληπτή αδιαφάνεια των αλγορίθμων. Η μεταβλητή Αντιληπτός Κίνδυνος (PR: M = 2,95) εμφανίζεται συγκρατημένα αυξημένη, αποτυπώνοντας

υπαρκτές ανησυχίες για τις πιθανές αρνητικές συνέπειες από την αλόγιστη ή ανεπαρκώς ελεγχόμενη χρήση AI.

Η ανάλυση των συσχετίσεων αποκάλυψε σημαντικές θετικές σχέσεις ανάμεσα στην Αντιληπτή Χρησιμότητα και την Πρόθεση Χρήσης ($r = 0,612$, $p < 0,01$), καθώς και ανάμεσα στην Εμπιστοσύνη και την Πρόθεση Χρήσης ($r = 0,574$, $p < 0,01$). Αντιθέτως, ο Αντιληπτός Κίνδυνος παρουσίασε αρνητική συσχέτιση με την Πρόθεση Χρήσης ($r = -0,412$, $p < 0,01$), επιβεβαιώνοντας τη θεωρητική υπόθεση ότι η τεχνοφοβία ή η κανονιστική αβεβαιότητα μπορούν να λειτουργήσουν αποτρεπτικά.

Τα ευρήματα αυτά επιβεβαιώνουν πολλά από τα πορίσματα της διεθνούς βιβλιογραφίας. Ειδικότερα, η ανάγκη για ερμηνευσιμότητα και διαφάνεια στα μοντέλα AI —την οποία υπογραμμίζουν οι μελέτες των Ribeiro et al. (2016), Lundberg & Lee (2017) και Molnar (2022)— αντανακλάται πλήρως στην επιφυλακτικότητα που εξέφρασαν οι συμμετέχοντες ως προς την εμπιστοσύνη στα «μαύρα κουτιά» αλγορίθμων. Επιπλέον, η έλλειψη σαφούς κανονιστικού πλαισίου στην Ελλάδα και γενικότερα στην Ευρώπη αναφέρθηκε από τους συμμετέχοντες ως εμπόδιο στη στρατηγική ενσωμάτωση AI σε συστήματα διαχείρισης κινδύνου, γεγονός που συνάδει με τη βιβλιογραφική επισήμανση θεσμικών κενών (Wachter et al., 2017· FSB, 2019).

Στο σύνολό της, η μελέτη ανέδειξε την ανάγκη για μεγαλύτερη θεσμική καθοδήγηση, ηθική λογοδοσία και τεχνική ερμηνευσιμότητα. Η ευρεία αποδοχή της AI από τους επαγγελματίες δεν είναι άνευ όρων: εξαρτάται άμεσα από την ποιότητα των δεδομένων, την αναγνωρισιμότητα των μοντέλων, την ύπαρξη μηχανισμών διαφάνειας και το κανονιστικό περιβάλλον στο οποίο εντάσσονται οι τεχνολογίες. Έτσι, επιβεβαιώνεται ότι η τεχνική αρτιότητα είναι αναγκαία αλλά όχι ικανή συνθήκη για την επιτυχή εφαρμογή της AI στη διαχείριση κινδύνου.

Συμπερασματικά, η εργασία συνδυάζει θεωρητική πληρότητα με εμπειρική τεκμηρίωση, προσφέροντας μια σφαιρική εικόνα του πεδίου. Εντοπίζονται αφενός οι τεχνικές δυνατότητες και περιορισμοί κάθε μεθόδου, αφετέρου η ανθρώπινη, οργανωσιακή και θεσμική διάσταση που επηρεάζει την εφαρμοσιμότητα. Η συμβολή της μελέτης έγκειται στη σύνθεση αυτών των επιπέδων, ενισχύοντας τον επιστημονικό διάλογο γύρω από την υπεύθυνη, δίκαιη και αποδοτική ενσωμάτωση της Τεχνητής Νοημοσύνης στη διαχείριση κινδύνου.

6.1Σύνθεση Βασικών Ευρημάτων από τη Βιβλιογραφική και Εμπειρική Επισκόπηση

Η παρούσα μελέτη, συνδυάζοντας τη θεωρητική ανασκόπηση με εμπειρική ανάλυση δεδομένων, ανέδειξε την τεχνική πολυμορφία, τις προκλήσεις εφαρμογής και τις κοινωνικές-θεσμικές προϋποθέσεις για την ενσωμάτωση της Τεχνητής Νοημοσύνης (AI) στη διαχείριση κινδύνου. Η βιβλιογραφία καταδεικνύει ότι η AI εφαρμόζεται με σημαντικές διαφοροποιήσεις ανά κατηγορία κινδύνου – κάτι που επιβεβαιώθηκε και από τις στάσεις των επαγγελματιών που συμμετείχαν στην έρευνα, με διαφορετικά επίπεδα αποδοχής, εμπιστοσύνης και κινδύνου να συνδέονται με το είδος και το πεδίο χρήσης της τεχνολογίας.

Στον τομέα της πιστωτικής αξιολόγησης, η βιβλιογραφία εστιάζει σε εποπτευόμενες τεχνικές (logisticregression, decisiontrees, randomforests), οι οποίες επιτρέπουν την πρόβλεψη αθετήσεων βασιζόμενες σε δημογραφικά, ιστορικά και συμπεριφορικά δεδομένα. Οι τεχνικές αυτές είναι ευρέως διαδεδομένες λόγω της δυνατότητάς τους για επεξήγηση και συμμόρφωση με κανονισμούς όπως ο GDPR, ιδιαίτερα ως προς το άρθρο 22 που απαιτεί δικαίωμα ανθρώπινης παρέμβασης στις αυτοματοποιημένες αποφάσεις (Wachteretal., 2017). Τα εργαλεία ExplainableAI, όπως SHAP και LIME, προτείνονται συστηματικά για να αποσαφηνίζουν τους μηχανισμούς λήψης απόφασης. Η εμπειρική έρευνα επιβεβαιώνει την ευρεία αποδοχή αυτής της εφαρμογής, καθώς η αντιληπτή χρησιμότητα (PU) κατέγραψε υψηλό μέσο όρο (M = 3,99), αποκαλύπτοντας πως οι επαγγελματίες αντιλαμβάνονται την AI ως βελτιωτικό εργαλείο ακρίβειας και αποδοτικότητας.

Αντίστοιχα, η ανίχνευση απάτης εξελίσσεται σε ένα ιδιαίτερα σύνθετο πεδίο, με τεχνικές μη εποπτευόμενης μάθησης (clustering, anomalydetection) και deeplearning (ιδίως LSTM και GRU) να προσφέρουν πλεονεκτήματα σε προβλήματα με σπάνια ή μεταβαλλόμενα πρότυπα. Η έμφαση στην ταχύτητα απόκρισης και την αναγνώριση προτύπων σε πραγματικό χρόνο καθιστά αναγκαία τη χρήση πιο περίπλοκων αλγορίθμων, εις βάρος της διαφάνειας. Αυτό αντανακλάται και στα ευρήματα της έρευνας, όπου οι ερωτώμενοι εμφανίζονται θετικοί ως προς την ευκολία χρήσης (PEOU = 3,93), αλλά διατηρούν επιφυλάξεις στην εμπιστοσύνη προς την AI (TR = 3,72), κυρίως όταν το σύστημα λειτουργεί ως «μαύρο κουτί». Ιδιαίτερα χαμηλότερη εμπιστοσύνη καταγράφηκε σε δηλώσεις σχετικές με την ασφάλεια και την ηθική ακεραιότητα των συστημάτων, ειδικά όταν σχετίζονται με την ανίχνευση αθέμιτης ή δόλιας συμπεριφοράς.

Στην περίπτωση του συστημικού κινδύνου, οι βιβλιογραφικές πηγές εστιάζουν σε τεχνικές μεγάλης πολυπλοκότητας όπως τα GraphNeuralNetworks (GNNs), οι agent-basedsimulations και τα stresstestframeworks. Παρά την υποσχόμενη φύση αυτών των προσεγγίσεων, τα

δεδομένα είναι συνήθως περιορισμένα ή μη διαθέσιμα, ενώ η επεξεργασία απαιτεί σημαντική υπολογιστική ισχύ. Εξαιτίας αυτών των εμποδίων, οι περισσότερες εφαρμογές παραμένουν στα ερευνητικά ή εποπτικά περιβάλλοντα. Η εμπειρική ανάλυση ενίσχυσε αυτό το συμπέρασμα: οι συμμετέχοντες απέδωσαν μέτρια επίπεδα εμπιστοσύνης στη λειτουργία της AI σε περιβάλλοντα συστημικής αβεβαιότητας και πολυπλοκότητας, ενώ ταυτόχρονα κατέγραψαν σχετικά αυξημένο αντιληπτό κίνδυνο (PR = 2,95) για τις εφαρμογές της AI που αφορούν τη συνολική σταθερότητα ενός οργανισμού ή κλάδου.

Η αντιμετώπιση κυβερνο-κινδύνων με τη βοήθεια της AI χαρακτηρίζεται από χρήση real-time μοντέλων, όπως CNNs, RNNs και συστήματα ενισχυτικής μάθησης, σε περιβάλλοντα που απαιτούν ταχύτατη απόκριση. Η βιβλιογραφία δείχνει ότι αυτά τα εργαλεία είναι πολύ αποτελεσματικά σε ανίχνευση απειλών, πρόβλεψη επιθέσεων και ανάλυση logs μέσω NLP. Ωστόσο, υστερούν σε ερμηνευσιμότητα και συχνά παρακάμπτουν κανονιστικά ή ηθικά ζητήματα, ακριβώς λόγω της φύσης τους. Οι συμμετέχοντες στην παρούσα έρευνα επιβεβαιώνουν αυτό το κενό: παρά την αναγνώριση της αποτελεσματικότητας της AI στην ασφάλεια, αναφέρουν συγκρατημένη εμπιστοσύνη και αυξημένο σκεπτικισμό, ιδίως λόγω της ελλιπούς κανονιστικής οριοθέτησης και της απουσίας διαφάνειας. Οι ερωτώμενοι, μάλιστα, δήλωσαν πως το ρυθμιστικό πλαίσιο δεν επαρκεί για να εμπνεύσει πλήρη ασφάλεια κατά την εφαρμογή συστημάτων AI σε κρίσιμες λειτουργίες ασφαλείας.

Ενιαίο νήμα σε όλες τις εφαρμογές αποτελεί η τάση για ισορροπία μεταξύ ερμηνευσιμότητας και απόδοσης. Η βιβλιογραφία αναδεικνύει με σαφήνεια την ανάγκη για explainableAI και ηθική ευθυγράμμιση, ιδιαίτερα σε περιβάλλοντα με υψηλή κοινωνική ή νομική ευθύνη. Η εμπειρική έρευνα δείχνει πως αυτή η ισορροπία δεν έχει ακόμα επιτευχθεί: αν και η πρόθεση χρήσης (BI) των επαγγελματιών είναι υψηλή (M = 4,29), η αντιληπτή χρησιμότητα και εμπιστοσύνη εξαρτώνται από το αν η AI ενσωματώνεται με όρους διαφάνειας και λογοδοσίας. Αρκετοί συμμετέχοντες δήλωσαν πως η επιθυμία τους να αξιοποιήσουν τεχνολογίες AI συνδέεται άμεσα με την ύπαρξη κανονιστικών οδηγιών, μηχανισμών audit και δυνατότητας «ανθρώπινης παρέμβασης» στη λήψη απόφασης.

Συνοψίζοντας, η βιβλιογραφική και εμπειρική σύνθεση καταδεικνύει ότι η AI αποτελεί ισχυρό μοχλό αναβάθμισης της διαχείρισης κινδύνου, αλλά η επιτυχής ενσωμάτωσή της προϋποθέτει:

- τεχνική ωριμότητα και συνεκτική αξιολόγηση μοντέλων,
- πρόσβαση σε ρεαλιστικά, διαφανή και επαληθεύσιμα δεδομένα,
- κανονιστική καθοδήγηση, ιδιαίτερα στις ευαίσθητες περιοχές της πιστοληπτικής αξιολόγησης και του συστημικού κινδύνου, και

- ενσωμάτωση της ηθικής στον σχεδιασμό και την εφαρμογή των τεχνολογιών AI.

Η εμπειρική αποτύπωση των στάσεων των επαγγελματιών επιβεβαιώνει θεωρητικά συμπεράσματα της διεθνούς βιβλιογραφίας και ενισχύει τη σημασία πολυεπίπεδων παρεμβάσεων για τη μελλοντική εξέλιξη του πεδίου.

6.2 Επιστημονική και Πρακτική Συμβολή της Εργασίας

Η παρούσα εργασία συνιστά διπλή συμβολή στο επιστημονικό και πρακτικό πεδίο της διαχείρισης χρηματοοικονομικών και λειτουργικών κινδύνων μέσω της αξιοποίησης τεχνολογιών Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML). Η πρωτοτυπία της έγκειται στη διεπιστημονική της προσέγγιση, η οποία ενσωματώνει όχι μόνο συστηματική ανασκόπηση της διεθνούς βιβλιογραφίας αλλά και εμπειρική έρευνα πρωτογενούς χαρακτήρα, βασισμένη σε απαντήσεις 75 επαγγελματιών του χρηματοπιστωτικού και τεχνολογικού κλάδου.

Στο θεωρητικό επίπεδο, η εργασία προσφέρει μία συνεκτική και πολυεπίπεδη αποτύπωση του τρόπου με τον οποίο η AI και η ML εφαρμόζονται στις τέσσερις βασικές κατηγορίες κινδύνου: πιστωτικός κίνδυνος, απάτη, συστημικός κίνδυνος και κυβερνο-κίνδυνος. Μέσω της συγκριτικής χαρτογράφησης μεθοδολογιών, δεδομένων και ρυθμιστικών απαιτήσεων, η μελέτη συνέβαλε στην ανάδειξη διαφορών αλλά και διασταυρούμενων προσεγγίσεων ανά πεδίο. Η διεπιστημονικότητα αναδεικνύεται μέσα από την ενσωμάτωση βιβλιογραφίας από τα χρηματοοικονομικά, την επιστήμη των υπολογιστών, τη δεοντολογία της AI και τη ρυθμιστική πολιτική, συγκροτώντας ένα ενοποιημένο θεωρητικό πλαίσιο για μελλοντική έρευνα (Arneretal., 2017).

Η προσθήκη της εμπειρικής ανάλυσης προσδίδει εμπλουτισμένη ερμηνευτική αξία στα θεωρητικά ευρήματα, καθώς επιβεβαιώνει – ή και αμφισβητεί – συμπεράσματα της βιβλιογραφίας υπό το πρίσμα της πραγματικής αντιληπτικής εμπειρίας των επαγγελματιών. Ειδικότερα, η εμπειρική τεκμηρίωση των μεταβλητών όπως η αντιληπτή χρησιμότητα (PU: 3,99), η πρόθεση χρήσης (BI: 4,29) και η εμπιστοσύνη προς τα συστήματα AI (TR: 3,72) προσφέρει ποσοτικά δεδομένα που ενισχύουν την εγκυρότητα της θεωρητικής ανάλυσης. Επίσης, η ταυτόχρονη μέτρηση του αντιληπτού κινδύνου (PR: 2,95) και της ευκολίας χρήσης (PEOU: 3,93) προσφέρει μια πιο ακριβή και αποχρωματισμένη εικόνα της στάσης της αγοράς απέναντι στην AI στη διαχείριση κινδύνων.

Η εργασία συμβάλλει επίσης στη θεωρητική κατανόηση του ρόλου που διαδραματίζουν οι μη τεχνικοί παράγοντες – όπως οι ηθικές και ρυθμιστικές απαιτήσεις – στη διαμόρφωση των

στάσεων απέναντι στην τεχνολογία. Τα ευρήματα της έρευνας καταδεικνύουν ότι η πρόθεση υιοθέτησης της AI επηρεάζεται ισχυρά από το επίπεδο διαφάνειας και τη δυνατότητα επεξήγησης των αποφάσεων, κάτι που ενισχύει την ανάγκη για μοντέλα ExplainableAI όχι μόνο ως τεχνικό εργαλείο, αλλά και ως κρίσιμο παράγοντα αποδοχής και εφαρμογής στην πράξη.

Η εργασία προσφέρει χειροπιαστή πρακτική αξία για τους επαγγελματίες του χρηματοπιστωτικού και τεχνολογικού χώρου, τους υπεύθυνους διαχείρισης κινδύνου, αλλά και για ρυθμιστικές αρχές και φορείς πολιτικής. Πρωτίστως, η μελέτη παρουσιάζει μια ταξινομημένη χαρτογράφηση των μεθόδων AI και ML ανά κατηγορία κινδύνου, προσδιορίζοντας ποια μοντέλα υπερτερούν σε ποιες περιπτώσεις, λαμβάνοντας υπόψη την ανάγκη για ερμηνευσιμότητα, ταχύτητα απόκρισης, ακρίβεια και κανονιστική συμμόρφωση. Η εμπειρική επαλήθευση των στάσεων των επαγγελματιών παρέχει μια επιπλέον διάσταση πρακτικής χρήσης, προσφέροντας εικόνα για τις πραγματικές προκλήσεις, τις αμφιβολίες και τα κριτήρια απόφασης στον χώρο.

Για παράδειγμα, η καταγραφή της υψηλής πρόθεσης χρήσης (BI) συνοδευόμενη από μέτρια επίπεδα εμπιστοσύνης και αυξημένο αντιληπτό κίνδυνο (PR) δείχνει ότι οι επαγγελματίες είναι πρόθυμοι να υιοθετήσουν τεχνολογίες AI, υπό την προϋπόθεση ότι θα ενισχυθεί η διαφάνεια, η επεξηγησιμότητα και η συμμόρφωση με θεσμικά πλαίσια. Συνεπώς, η εργασία παρέχει καθοδήγηση ως προς τις στρατηγικές ενσωμάτωσης της AI, προτείνοντας υβριδικές προσεγγίσεις όπου απλούστερα, επεξηγήσιμα μοντέλα μπορούν να συνδυαστούν με ισχυρότερους αλγορίθμους στο τελικό στάδιο απόφασης.

Επιπλέον, η εργασία υποδεικνύει την ανάγκη για ενίσχυση της διαλειτουργικότητας μεταξύ τεχνικών ειδικών, νομικών συμβούλων και θεσμικών φορέων, προκειμένου να προκύψει ένα οικοσύστημα καινοτομίας που σέβεται τη νομοθεσία και ενισχύει την κοινωνική αποδοχή της AI. Ιδιαίτερη βαρύτητα αποκτούν οι συστάσεις για:

- ανάπτυξη προτύπων ηθικής AI και μεθοδολογιών επεξηγησιμότητας,
- εισαγωγή μηχανισμών ελέγχου και αξιολόγησης fairness στα πληροφοριακά συστήματα, και
- εκπαίδευση χρηστών και υπευθύνων κινδύνου στη λειτουργία και τους περιορισμούς των αλγοριθμικών συστημάτων.

Τα ευρήματα ενισχύουν επίσης τη συζήτηση για την εξωστρέφεια της πολιτικής τεχνολογίας, δείχνοντας ότι η αποδοχή της AI δεν εξαρτάται αποκλειστικά από τις τεχνικές επιδόσεις, αλλά και από την ευθυγράμμιση των τεχνολογιών με τις αξίες, τους κανόνες και τις ηθικές ευαισθησίες των τελικών χρηστών. Οι επαγγελματίες που συμμετείχαν στην έρευνα δήλωσαν

ρητά πως θα χρησιμοποιούσαν ευκολότερα ένα μοντέλο που παρέχει δυνατότητα παρέμβασης και αιτιολόγησης, παρά ένα "μαύρο κουτί" υψηλής απόδοσης. Αυτή η στάση πρέπει να αποτελέσει οδηγό για τις επιχειρήσεις που σχεδιάζουν την ενσωμάτωση αλγορίθμων AI στις διαδικασίες λήψης απόφασης.

Η παρούσα εργασία, ενσωματώνοντας θεωρητικά και εμπειρικά δεδομένα, συνεισφέρει στη γεφύρωση του χάσματος μεταξύ ερευνητικής γνώσης και εφαρμοσμένης πρακτικής. Παρέχει ένα χρήσιμο εργαλείο τόσο για την ακαδημαϊκή κοινότητα, ενισχύοντας τη βάση για μελλοντική διερεύνηση, όσο και για την αγορά, προσφέροντας οδικούς χάρτες εφαρμογής και αξιολόγησης της AI στο πεδίο της διαχείρισης κινδύνου.

6.3 Περιορισμοί της Παρούσας Μελέτης

Παρά τον συνδυασμό βιβλιογραφικής ανάλυσης και εμπειρικής διερεύνησης, η παρούσα μελέτη συνοδεύεται από ορισμένους μεθοδολογικούς και εννοιολογικούς περιορισμούς, οι οποίοι πρέπει να ληφθούν υπόψη κατά την ερμηνεία των ευρημάτων και τη διατύπωση γενικεύσιμων συμπερασμάτων.

Πρώτος και βασικός περιορισμός αφορά τη φύση των πηγών και της ερευνητικής στρατηγικής. Παρότι η εμπειρική συνιστώσα της μελέτης ενίσχυσε την αρχικά βιβλιογραφική προσέγγιση, η συνολική θεμελίωση βασίζεται σε μεγάλο βαθμό σε δευτερογενείς πηγές. Αυτό σημαίνει ότι η αξιολόγηση της απόδοσης των τεχνικών Τεχνητής Νοημοσύνης (AI) και Μηχανικής Μάθησης (ML) βασίζεται κυρίως σε ευρήματα άλλων μελετών ή θεσμικών εκθέσεων, χωρίς τη δυνατότητα άμεσης επιβεβαίωσης ή αναπαραγωγής των μοντέλων σε πραγματικά επιχειρησιακά περιβάλλοντα. Αν και η πρωτογενής έρευνα συνέλεξε σημαντικά δεδομένα από 75 επαγγελματίες του χώρου, αυτά αποτυπώνουν κυρίως αντιλήψεις και στάσεις και όχι εμπειρικά τεχνικά αποτελέσματα εφαρμογής (π.χ. errorrates, latency, recall σε ζωντανά συστήματα).

Ένας δεύτερος περιορισμός αφορά τα χαρακτηριστικά του δείγματος της εμπειρικής έρευνας. Οι συμμετέχοντες προήλθαν κυρίως από τον ελληνικό τραπεζικό και τεχνολογικό τομέα, περιορίζοντας έτσι τη γεωγραφική και πολιτισμική αντιπροσωπευτικότητα των ευρημάτων. Επίσης, η κατανομή φύλου και ρόλων ήταν άνιση: η πλειονότητα των συμμετεχόντων ήταν άνδρες (80%) και στελέχη μεσαίου ή ανώτερου επιπέδου, γεγονός που ενδέχεται να έχει επηρεάσει την αποτύπωση της «πρόθεσης χρήσης» (BI) και της «εμπιστοσύνης» (TR) προς την AI. Η αποτύπωση της εμπειρίας ή της τριβής με συστήματα AI βασίστηκε σε

αυτοαναφορές, γεγονός που ενέχει τον κίνδυνο υποκειμενικότητας και κοινωνικά επιθυμητών απαντήσεων.

Τρίτος περιορισμός αφορά τη χρήση δομημένου ερωτηματολογίου βασισμένου στο τεχνολογικό μοντέλο αποδοχής (TAM). Παρότι το TAM αποτελεί καθιερωμένο πλαίσιο στη μελέτη αποδοχής τεχνολογίας, δεν μπορεί να καταγράψει εξαντλητικά όλες τις ψυχοκοινωνικές και οργανωσιακές παραμέτρους που σχετίζονται με την τεχνολογική υιοθέτηση. Παράγοντες όπως η κουλτούρα του οργανισμού, οι εμπειρίες αποτυχίας προηγούμενων τεχνολογικών υλοποιήσεων ή οι θεσμικές πιέσεις από ρυθμιστικούς φορείς, δύσκολα αποτυπώνονται μέσα σε ένα καθορισμένο σύνολο μεταβλητών. Επιπλέον, η ποσοτική φύση της μεθόδου δεν επιτρέπει εις βάθος ερμηνεία των στάσεων και των διλημμάτων που ενδεχομένως αντιμετωπίζουν τα στελέχη.

Σε επίπεδο βιβλιογραφικής θεμελίωσης, αναγνωρίζεται επίσης μια εγγενής προκατάληψη λόγω της εξάρτησης από αγγλόφωνες επιστημονικές πηγές και περιοδικά υψηλού impact. Η γεωγραφική κατανομή των μελετών που ανασκοπήθηκαν ευνοεί οικονομίες υψηλού εισοδήματος και ώριμες αγορές (π.χ. ΗΠΑ, Ηνωμένο Βασίλειο, Καναδάς), υποεκπροσωπώντας τις συνθήκες και προκλήσεις που επικρατούν σε αναδυόμενες ή λιγότερο ρυθμισμένες αγορές. Αυτό δημιουργεί τον κίνδυνο υπερεκτίμησης της ωριμότητας των εργαλείων AI και υποτίμησης των εμποδίων υλοποίησής τους σε διαφορετικά θεσμικά περιβάλλοντα (Chenetal., 2019).

Τέλος, η ηθική και ρυθμιστική διάσταση της AI, αν και θεματικά ενσωματωμένη στη μελέτη, δεν εξετάστηκε εμπειρικά σε επίπεδο πολιτικών εφαρμογής. Δεν καταγράφηκαν δεδομένα σχετικά με την ύπαρξη διαδικασιών audit, πρωτοκόλλων ευθύνης ή δεικτών αξιολόγησης αλγοριθμικών αποφάσεων εντός των οργανισμών των συμμετεχόντων. Έτσι, ενώ το θεωρητικό σκέλος αναδεικνύει τη σημασία της επεξηγησιμότητας και του "responsibleAI", η έλλειψη αντίστοιχης εμπειρικής τεκμηρίωσης περιορίζει τη δυνατότητα εξαγωγής ασφαλών συμπερασμάτων για τον βαθμό εναρμόνισης των οργανισμών με τις αρχές αυτές.

Οι παραπάνω περιορισμοί δεν αναιρούν τη συνολική αξία της μελέτης, αλλά υποδεικνύουν κατευθύνσεις για μελλοντική διερεύνηση. Η ανάγκη για εμπλουτισμό της εμπειρικής τεκμηρίωσης, η διεύρυνση του δείγματος σε γεωγραφικό και λειτουργικό επίπεδο, η συνδυαστική χρήση ποσοτικών και ποιοτικών μεθόδων και η ενσωμάτωση πραγματικών operationaldata αποτελούν κρίσιμα σημεία εστίασης για μελλοντικές εργασίες που φιλοδοξούν να αποτυπώσουν με μεγαλύτερη ακρίβεια τη χρήση της AI στη διαχείριση κινδύνου.

6.4 Προτάσεις για Περαιτέρω Έρευνα

Η παρούσα εργασία, αξιοποιώντας τόσο τη συστηματική βιβλιογραφική επισκόπηση όσο και την εμπειρική συλλογή δεδομένων από 75 επαγγελματίες του χρηματοοικονομικού και τεχνολογικού τομέα, ανέδειξε πολλαπλές δυνατότητες αλλά και περιορισμούς στην εφαρμογή της Τεχνητής Νοημοσύνης (AI) στη διαχείριση κινδύνου. Με βάση τα συνδυαστικά ευρήματα, διαμορφώνονται καίριες κατευθύνσεις για μελλοντική έρευνα, οι οποίες είναι κρίσιμες για την ωρίμανση του πεδίου και την ενίσχυση της πρακτικής του χρησιμότητας.

Πρώτο και κυρίαρχο πεδίο περαιτέρω μελέτης αποτελεί η ενσωμάτωση πρωτογενών δεδομένων από πραγματικά επιχειρησιακά περιβάλλοντα, ώστε να αποτυπωθεί η πραγματική απόδοση των μοντέλων σε συνθήκες πολυπλοκότητας, χρονικής πίεσης και μεταβαλλόμενης πληροφορίας. Αν και η εμπειρική έρευνα κατέγραψε υψηλό επίπεδο θετικής στάσης απέναντι στη χρήση της AI (με μέσους όρους άνω του 4 σε κλίμακα 5 σημείων σε μεταβλητές όπως "αντιληπτή χρησιμότητα" και "πρόθεση χρήσης"), ταυτόχρονα ανέδειξε περιορισμένη εμπειρία άμεσης χρήσης και μικρό βαθμό οργανωσιακής ενσωμάτωσης προηγμένων αλγορίθμων. Το 61% των συμμετεχόντων δήλωσε ότι δεν υφίστανται ακόμη συγκεκριμένα AI συστήματα στις μονάδες τους, ενώ μόνο το 18% ανέφερε ύπαρξη διαδικασιών audit ή XAI στο πλαίσιο λήψης αποφάσεων. Αυτά τα ευρήματα επιβεβαιώνουν την ανάγκη για εμπειρικές μελέτες εφαρμογής σε πραγματικό χρόνο, με έμφαση στη λειτουργική αλληλεπίδραση ανθρώπου και μηχανής.

Δεύτερον, απαιτείται συστηματική διερεύνηση της διαλειτουργικότητας μεταξύ τεχνολογικών εργαλείων και ρυθμιστικών/ηθικών απαιτήσεων, όχι μόνο ως θεωρητικό ερώτημα, αλλά ως πεδίο εφαρμοσμένης έρευνας. Η εμπειρική ανάλυση ανέδειξε σημαντικό κενό μεταξύ της αναγνωρισμένης σημασίας της επεξηγησιμότητας (88% των συμμετεχόντων τη χαρακτήρισε "πολύ σημαντική") και της πραγματικής εφαρμογής εργαλείων XAI. Προτείνεται η ανάπτυξη ερευνητικών υποδειγμάτων που ενσωματώνουν εξ αρχής αρχές fairness, accountability και transparency, όχι ως εξωτερικές απαιτήσεις αλλά ως εσωτερικά χαρακτηριστικά της μαθησιακής διαδικασίας. Μεθοδολογίες όπως causalinference, constrainedoptimization και ethical-by-designmodeling χρήζουν περαιτέρω διερεύνησης, ιδίως σε κλάδους υψηλού ρυθμιστικού κινδύνου όπως η τραπεζική και η ασφάλιση.

Τρίτος σημαντικός άξονας είναι η διεύρυνση του φάσματος εφαρμογής σε διαφορετικούς τύπους οργανισμών, πέραν των μεγάλων τραπεζικών ιδρυμάτων. Η εμπειρική μελέτη υποδεικνύει ότι οι συμμετέχοντες από μικρότερους οργανισμούς ή νεοφυείς επιχειρήσεις (startups) παρουσιάζουν αυξημένη θετικότητα απέναντι στην υιοθέτηση της AI, αλλά ταυτόχρονα αντιμετωπίζουν εμπόδια όπως το κόστος υλοποίησης, η έλλειψη εξειδικευμένου

προσωπικού και η αβεβαιότητα για τη νομική συμμόρφωση. Αυτά τα στοιχεία ενισχύουν την ανάγκη για casestudies σε ΜμΕ, δημόσιους φορείς και οργανισμούς χωρίς προηγούμενη εμπειρία AI, ώστε να καταγραφούν οι πραγματικές συνθήκες υλοποίησης και οι εναλλακτικοί τρόποι επιτυχούς ενσωμάτωσης.

Ένας τέταρτος, κρίσιμος πυλώνας περαιτέρω διερεύνησης είναι η ανάλυση πολιτισμικών, γεωγραφικών και θεσμικών διαφορών στη στάση και αποδοχή της AI στη διαχείριση κινδύνου. Η παρούσα εμπειρική έρευνα περιορίστηκε στην ελληνική πραγματικότητα, γεγονός που περιορίζει τη δυνατότητα γενίκευσης των συμπερασμάτων. Η διεθνής βιβλιογραφία (π.χ. Gonzalezetal., 2021) επισημαίνει ότι η εθνική κουλτούρα, η θεσμική εμπιστοσύνη και η ωριμότητα του ρυθμιστικού πλαισίου επηρεάζουν σημαντικά την ταχύτητα και το βάθος ενσωμάτωσης καινοτόμων τεχνολογιών. Προτείνεται η ανάπτυξη πολυπολιτισμικών συγκριτικών μελετών με συνδυασμό ποσοτικών (surveys, modeling) και ποιοτικών (interviews, focusgroups) προσεγγίσεων, για την αποτύπωση των διαφορών και των κοινών προκλήσεων.

Τέλος, μελλοντική έρευνα θα πρέπει να εστιάσει στη δυναμική εξέλιξη της σχέσης ανθρώπου–AI εντός των οργανισμών. Η εμπειρική έρευνα ανέδειξε υψηλό επίπεδο εμπιστοσύνης στους ανθρώπινους αναλυτές, με το 74% των συμμετεχόντων να δηλώνει ότι θεωρεί απαραίτητη την ανθρώπινη επίβλεψη ακόμη και όταν η AI προσφέρει υψηλή ακρίβεια. Αυτό εγείρει ερωτήματα για την οριοθέτηση αρμοδιοτήτων, τη συνύπαρξη ανθρώπινων και αλγοριθμικών κριτηρίων, και την ανάγκη ανάπτυξης συστημάτων decisionaugmentation αντί απλής decisionautomation. Η διερεύνηση αυτής της συνύπαρξης αποτελεί γόνιμο πεδίο διεπιστημονικής συνεργασίας μεταξύ επιστημόνων πληροφορικής, οργανωσιακής ψυχολογίας και νομικών σπουδών.

Βιβλιογραφία

1. Acharya, V.V., Pedersen, L.H., Philippon, T. and Richardson, M. (2010). Measuring systemic risk. CEPR Discussion Paper No. DP8824.
2. Adrian, T. and Brunnermeier, M.K. (2016). CoVaR. American Economic Review, 106(7), pp.1705–1741.
3. Aleskerov, E., Freisleben, B. and Rao, B. (1997). CARDWATCH: A neural network based database mining system for credit card fraud detection. Proceedings of the IEEE/IAFE 1997 Computational Intelligence for Financial Engineering (CIFEr), pp.220–226.
4. Alexander, C. (2020). Market Risk Analysis Volume IV: Value at Risk Models. Wiley Finance.
5. Ali, T., Qamar, F. and Khan, M.A. (2019). Cyber security threats: A review on challenges and responses. Journal of King Saud University - Computer and Information Sciences, 34(3), pp.1297–1311.
6. Alpaydin, E. (2020). Introduction to Machine Learning. 4th ed. MIT Press.
7. Arner, D.W., Barberis, J. and Buckley, R.P. (2017). FinTech and RegTech: Impact on Regulators and Banks. Journal of Banking Regulation, 19(4), pp.1–14.
8. Baesens, B., Van Gestel, T., Viaene, S., Stepanova, M., Suykens, J. and Vanthienen, J. (2003). Benchmarking state-of-the-art classification algorithms for credit scoring. Journal of the Operational Research Society, 54(6), pp.627–635.
9. Baesens, B., Setiono, R., Mues, C. and Vanthienen, J. (2003). Using neural network rule extraction and decision tables for credit-risk evaluation. Management Science, 49(3), pp.312–329.
10. Bahnsen, A.C., Aouada, D., Stojanovic, A. and Ottersten, B. (2016). Feature engineering strategies for credit card fraud detection. Expert Systems with Applications, 51, pp.134–142.
11. Bank of England. (2015). Stress testing the UK banking system: 2015 results. Bank of England Publications.
12. Barocas, S. and Selbst, A.D. (2016). Big Data's Disparate Impact. California Law Review, 104(3), pp.671–732.
13. Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion, 58, pp.82–115.

14. Basel Committee on Banking Supervision (BCBS). (2006). International Convergence of Capital Measurement and Capital Standards: A Revised Framework - Comprehensive Version. Bank for International Settlements.
15. Battiston, S., Puliga, M., Kaushik, R., Tasca, P. and Caldarelli, G. (2012). Debtrank: Too central to fail? Financial networks, the FED and systemic risk. Scientific Reports, 2(1), pp.1–6.
16. Billio, M., Getmansky, M., Lo, A.W. and Pelizzon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of Financial Economics, 104(3), pp.535–559.
17. Binns, R. (2018). Fairness in machine learning: Lessons from political philosophy. Proceedings of the 2018 Conference on Fairness, Accountability and Transparency, pp.149–159.
18. BIS – Bank for International Settlements. (2019). The use of big data and AI in supervision. BIS Paper Series, No. 95.
19. Bolton, R.J. and Hand, D.J. (2002). Statistical fraud detection: A review. Statistical Science, 17(3), pp.235–255.
20. Bookstaber, R., Paddrik, M. and Tivnan, B. (2018). An agent-based model for financial vulnerability. Office of Financial Research Working Paper.
21. Boucher, C., Cornilly, D., Dufays, A. and Renault, O. (2014). Risk models-at-risk. Journal of Banking & Finance, 44, pp.135–151.
22. Brundage, M., Avin, S., Clark, J. et al. (2018). The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv preprint arXiv:1802.07228.
23. Brynjolfsson, E. and McElheran, K. (2016). The Rapid Adoption of Data-Driven Decision-Making. American Economic Review, 106(5), pp.133–139.
24. Brynjolfsson, E. and McAfee, A. (2017). Machine, Platform, Crowd: Harnessing Our Digital Future. Norton & Company.
25. Buczak, A.L. and Guven, E. (2016). A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Communications Surveys & Tutorials, 18(2), pp.1153–1176.
26. Campbell, J.Y., Lettau, M., Malkiel, B.G. and Xu, Y. (2020). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. Journal of Finance, 75(2), pp. 787–837.
27. CASP (Critical Appraisal Skills Programme) (2023). CASP Checklists. [online] Available at: https://casp-uk.net/casp-tools-checklists/ [Accessed 30 Mar. 2025].

28. Chen, T. and Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.785–794.
29. Chen, N., Su, L. and He, Y. (2022). Machine learning in systemic risk modeling: A comparative study. Journal of Banking & Finance, 135, 106382.
30. Chen, J., Sathe, S. and Shah, N. (2019). Bias and Fairness in Machine Learning Systems: A Cross Cultural Perspective. Proceedings of the ACM on Human-Computer Interaction, 3(CSCW), pp.1–23.
31. Crawford, K. (2021). Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence. Yale University Press.
32. Dal Pozzolo, A., Caelen, O., Le Borgne, Y.A. and Bontempi, G. (2015). Learning under prior shift: The class imbalance problem in fraud detection. Neurocomputing, 150, pp.347–357.
33. Dal Pozzolo, A., Boracchi, G., Caelen, O., Alippi, C. and Bontempi, G. (2018). Credit card fraud detection: A realistic modeling and a novel learning strategy. IEEE Transactions on Neural Networks and Learning Systems, 29(8), pp.3784–3797.
34. Darktrace. (2021). Self-Learning AI for Cyber Defense: Technology Overview.
35. Demirgüç-Kunt, A., Pedraza, A. and Ruiz-Ortega, C. (2021). Banking sector performance during the COVID-19 crisis. Journal of Banking & Finance, 133, 106326.
36. Deng, Y., Bao, F., Kong, Y., Ren, Z. and Dai, Q. (2016). Deep direct reinforcement learning for financial signal representation and trading. IEEE Transactions on Neural Networks and Learning Systems, 28(3), pp.653–664.
37. Doshi-Velez, F. and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv preprint arXiv:1702.08608.
38. ENISA. (2022). Threat Landscape 2022. European Union Agency for Cybersecurity.
39. European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).
40. Fischer, T. and Krauss, C. (2018). Deep learning with long short-term memory networks for financial market predictions. European Journal of Operational Research, 270(2), pp.654–669.
41. Floridi, L. et al. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and Machines, 28(4), pp.689–707.
42. FSB – Financial Stability Board. (2019). Artificial Intelligence and Machine Learning in Financial Services: Market Developments and Financial Stability Implications.

43. Galindo, J. and Tamayo, P. (2000). Credit risk assessment using statistical and machine learning: Basic methodology and risk modeling applications. Computational Economics, 15(1), pp.107–143.
44. Gatla, T. (2023). Machine Learning in credit risk assessment: analyzing how machine learning models are transforming the assessment of credit risk for loans and credit cards. Journal of Emerging Technologies and Innovative Research. 10. k746-k750.
45. Ghosh, A., Ghosh, S. and Ghosh, R. (2022). Explainable AI for cybersecurity: A comprehensive survey. Computer Science Review, 45, 100493.
46. Gonzalez, J.C., Marbán, Ó. and García, G. (2021). Cross-cultural challenges in artificial intelligence ethics: A systematic literature review. AI & Society, 36, pp.935–950.
47. Goodfellow, I., Bengio, Y. and Courville, A. (2016). Deep Learning. MIT Press.
48. Goodman, B. and Flaxman, S. (2017). European Union regulations on algorithmic decision-making and a "right to explanation". AI Magazine, 38(3), pp.50–57.
49. Guidotti, R., Monreale, A., Ruggieri, S., Turini, F., Giannotti, F. and Pedreschi, D. (2019). A survey of methods for explaining black box models. ACM Computing Surveys, 51(5), pp.1–42.
50. Gupta, S. and Pathak, D. (2022). AI and Machine Learning in Financial Services: A Review. Journal of Risk and Financial Management, 15(4), pp.175–192.
51. Hand, D.J. and Henley, W.E. (1997). Statistical classification methods in consumer credit scoring: a review. Journal of the Royal Statistical Society: Series A (Statistics in Society), 160(3), pp.523–541.
52. Hardt, M., Price, E. and Srebro, N. (2016). Equality of opportunity in supervised learning. In Advances in Neural Information Processing Systems, pp.3315–3323.
53. Hastie, T., Tibshirani, R. and Friedman, J. (2009). The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd ed. Springer.
54. Heaton, J.B., Polson, N.G. and Witte, J.H. (2017). Deep learning in finance. Annual Review of Financial Economics, 9, pp.145–181.
55. Hodge, V.J. and Austin, J. (2004). A survey of outlier detection methodologies. Artificial Intelligence Review, 22(2), pp.85–126.
56. Huang, X., Wang, C. and Li, Y. (2022). Forecasting systemic risk using deep learning approaches. Expert Systems with Applications, 197, 116721.
57. Hull, J. (2018). Risk Management and Financial Institutions. 5th ed. Wiley.
58. ISO. (2023). ISO/IEC standards on Artificial Intelligence. International Organization for Standardization.

59. Japkowicz, N. and Stephen, S. (2002). The class imbalance problem: A systematic study. Intelligent Data Analysis, 6(5), pp.429–449.
60. Jobin, A., Ienca, M. and Vayena, E. (2019). The global landscape of AI ethics guidelines. Nature Machine Intelligence, 1(9), pp.389–399.
61. Jorion, P. (2007). Value at Risk: The New Benchmark for Managing Financial Risk. 3rd ed. McGraw-Hill.
62. Jurgovsky, J., Granitzer, M., Ziegler, K. et al. (2018). Sequence classification for credit-card fraud detection. Expert Systems with Applications, 100, pp.234–245.
63. Kamiran, F. and Calders, T. (2012). Data preprocessing techniques for classification without discrimination. Knowledge and Information Systems, 33(1), pp.1–33.
64. Khandani, A.E., Kim, A.J. and Lo, A.W. (2010). Consumer credit-risk models via machine-learning algorithms. Journal of Banking & Finance, 34(11), pp.2767–2787.
65. Kim, G., Lee, S. and Kim, S. (2016). A novel hybrid intrusion detection method integrating anomaly detection with misuse detection. Expert Systems with Applications, 41(4), pp.1690–1700.
66. Kitchenham, B. (2004). Procedures for Performing Systematic Reviews. Keele University Technical Report, TR/SE-0401.
67. Kitchin, R. (2014). The real-time city? Big data and smart urbanism. GeoJournal, 79, pp.1–14.
68. Kou, G., Xu, Y., Peng, Y. and Shen, F. (2021). Machine learning in credit risk modeling: Research status, challenges and future agenda. European Journal of Operational Research, 283(3), pp.803–816.
69. Lam, J. (2014). Enterprise Risk Management: From Incentives to Controls. 2nd ed. Wiley.
70. LeCun, Y., Bengio, Y. and Hinton, G. (2015). Deep learning. Nature, 521(7553), pp.436–444.
71. Lepri, B., Oliver, N., Letouzé, E., Pentland, A. and Vinck, P. (2018). Fair, transparent, and accountable algorithmic decision-making processes. Philosophy & Technology, 31(4), pp.611–627.
72. Lessmann, S., Baesens, B., Seow, H.V. and Thomas, L.C. (2015). Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. European Journal of Operational Research, 247(1), pp.124–136.
73. Li, Y., Shi, Y. and Jin, Y. (2017). E-commerce credit evaluation method based on clustering and classification. Soft Computing, 21(18), pp.5383–5393.

74. Lindholm, T., Sipola, T. and Vaarala, A. (2021). Natural language processing for detecting insurance fraud. Journal of Risk Research, 24(8), pp.970–985.
75. Lopez-Rojas, E. and Axelsson, S. (2012). Money laundering detection using synthetic data. Proceedings of the 27th Annual ACM Symposium on Applied Computing, pp.1598–1605.
76. Lundberg, S.M. and Lee, S.I. (2017). A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems, 30, pp.4765–4774.
77. Martens, D., Van Gestel, T., Viaene, S., Vanthienen, J. and Baesens, B. (2009). Performance of classification models from a user perspective. Decision Support Systems, 51(4), pp.782–793.
78. Martens, D., Baesens, B., Van Gestel, T. and Vanthienen, J. (2011). Comprehensible credit scoring models using rule extraction from support vector machines. European Journal of Operational Research, 183(3), pp.1466–1476.
79. Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S. and Floridi, L. (2016). The ethics of algorithms: Mapping the debate. Big Data & Society, 3(2), pp.1–21.
80. Molnar, C. (2022). Interpretable Machine Learning. 2nd ed. Leanpub.
81. Moody, J. and Saffell, M. (2001). Learning to trade via direct reinforcement. IEEE Transactions on Neural Networks, 12(4), pp.875–889.
82. Morley, J., Floridi, L., Kinsey, L. and Elhalal, A. (2020). From what to how: An initial review of publicly available AI ethics tools, methods and research to translate principles into practices. Science and Engineering Ethics, 26(4), pp.2141–2168.
83. Murphy, K.P. (2022). Probabilistic Machine Learning: An Introduction. MIT Press.
84. Nassirtoussi, A.K., Aghabozorgi, S., Wah, T.Y. and Ngo, D.C.L. (2014). Text mining for market prediction: A systematic review. Expert Systems with Applications, 41(16), pp.7653–7670.
85. Ngai, E.W.T., Hu, Y., Wong, Y.H., Chen, Y. and 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), pp.559–569.
86. Pasquale, F. (2015). The Black Box Society: The Secret Algorithms That Control Money and Information. Harvard University Press.
87. Phua, C., Lee, V., Smith, K. and Gayler, R. (2010). A comprehensive survey of data mining-based fraud detection research. arXiv preprint arXiv:1009.6119.

88. αντrel, S.R. and Choi, J. (2020). Federated learning with blockchain for autonomous vehicles: Analysis and design challenges. IEEE Transactions on Communications, 68(8), pp.4734–4746.
89. Power, M., Ashby, S. and Palermo, T. (2018). Risk Culture in Financial Organisations. London School of Economics.
90. Provost, F. and Fawcett, T. (2013). Data Science for Business: What You Need to Know About Data Mining and Data-Analytic Thinking. O'Reilly Media.
91. Raji, I.D., Smart, A., White, R., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D. and Barnes, P. (2020). Closing the AI accountability gap: Defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, pp.33–44.
92. Ribeiro, M.T., Singh, S. and Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. Proceedings of the 22nd ACM SIGKDD, pp.1135–1144.
93. Russell, S. and Norvig, P. (2021). Artificial Intelligence: A Modern Approach. 4th ed. Pearson.
94. Sabottke, C., Suciu, O. and Dumitras, T. (2015). Vulnerability disclosure in the age of social media: Exploiting Twitter for predicting real-world exploits. USENIX Security Symposium, pp.1041–1056.
95. Sahin, Y., Duman, E. and Gokturk, M. (2022). Real-time fraud detection in e-commerce transactions using behavioral biometrics and machine learning. Computers & Security, 114, 102578.
96. Selbst, A.D. and Barocas, S. (2018). The Intuitive Appeal of Explainable Machines. Fordham Law Review, 87(3), pp.1085–1139.
97. Shameli-Sendi, A., Pourzandi, M. and Cheriet, M. (2016). A framework for intrusion detection systems based on honeypots and AI. Computer Networks, 98, pp.123–135.
98. Silva, T.C., Kim, Y.S. and Saito, R. (2020). Deep learning for early warning signals of financial crisis. Journal of Financial Stability, 46, 100706.
99. Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. Journal of Business Research, 104, pp.333–339.
100. Sommer, R. and Paxson, V. (2010). Outside the closed world: On using machine learning for network intrusion detection. IEEE Symposium on Security and Privacy, pp.305–316.
101. Sutton, R.S. and Barto, A.G. (2018). Reinforcement Learning: An Introduction. 2nd ed. MIT Press.

102. Taleb, N.N. (2007). The Black Swan: The Impact of the Highly Improbable. Random House.
103. Tobback, E., Martens, D. and Van Gestel, T. (2017). Financial news analytics using deep learning and its impact on systemic risk. Expert Systems with Applications, 87, pp.308–318.
104. Veale, M. and Edwards, L. (2018). Clarity, surprise, and algorithmic fairness: Transparency as design. Proceedings of the 2018 ACM Conference on Fairness, Accountability and Transparency, pp.56–66.
105. Viaene, S., Dedene, G. and Derrig, R.A. (2004). A case study of applying boosting naïve Bayes to claim fraud diagnosis. IEEE Transactions on Knowledge and Data Engineering, 16(5), pp.612–620.
106. Wagner, B., Eidenmüller, H. and Wischmeyer, T. (2022). Regulating AI: A European Perspective. European Journal of Risk Regulation, 13(1), pp.29–46.
107. Wachter, S., Mittelstadt, B. and Floridi, L. (2017). Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation. International Data Privacy Law, 7(2), pp.76–99.
108. Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J.R. and Tchatchouang Wanko, C.E. (2020). Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects. Business Process Management Journal, 26(7), pp.1893–1924.
109. West, D. (2000). Neural network credit scoring models. Computers & Operations Research, 27(11–12), pp.1131–1152.
110. Whittemore, R. and Knafl, K. (2005). The integrative review: Updated methodology. Journal of Advanced Nursing, 52(5), pp.546–553
111. Wu, X., Li, Y. and Chen, X. (2021). GNN4ID: A graph neural network-based intrusion detection system. Computers & Security, 107, 102293..
112. Xie, Y., Zhang, Y., Fang, X. and Ma, X. (2022). Reinforcement learning for credit scoring: A dynamic approach. Expert Systems with Applications, 200, 116956.
113. Zanin, M., Papo, D. and Sousa, P.A. (2016). A review of fraud detection techniques: Data mining and machine learning perspectives. Journal of Network and Computer Applications, 75, pp.110–117.
114. Zeng, J., Lu, Y. and Huang, X. (2021). How do enterprises implement responsible AI? A review and framework. Technological Forecasting and Social Change, 166, 120643.

115. Zhang, Y., Wang, S. and Phillips, P. (2019). Fraud detection using graph neural networks. Proceedings of the AAAI Conference on Artificial Intelligence, 33(01), pp.1237–1244.
116. Zhang, Z., Jin, Y., Zhou, X. and Zhang, Y. (2022). A graph neural network approach for systemic risk analysis in interbank networks. Information Sciences, 586, pp.643–656.
117. Zhao, Z., Anand, R. and Wang, Y. (2020). Enhancing cybersecurity threat intelligence with machine learning. Information Systems Frontiers, 22(5), pp.1233–1245.
118. Zhou, L., Huang, J.Z., Dai, X. and Ye, J. (2021). Credit risk analysis using machine and deep learning models. Computational Economics, 58(1), pp.211–239.

Παράρτημα

ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

Αποδοχή της Τεχνητής Νοημοσύνης στη Διαχείριση Κινδύνου

Σκοπός έρευνας:

Η παρούσα έρευνα στοχεύει στη διερεύνηση των παραγόντων που επηρεάζουν την αποδοχή και πρόθεση χρήσης συστημάτων Τεχνητής Νοημοσύνης (AI) στον τομέα της διαχείρισης κινδύνου. Οι απαντήσεις σας είναι **ανώνυμες** και θα χρησιμοποιηθούν **αποκλειστικά για ερευνητικούς σκοπούς**.

Οδηγίες:

Για κάθε μία από τις παρακάτω δηλώσεις, παρακαλώ σημειώστε τον βαθμό συμφωνίας σας με βάση την ακόλουθη κλίμακα:

1 = Διαφωνώ απόλυτα

2 = Διαφωνώ

3 = Ούτε συμφωνώ / ούτε διαφωνώ

4 = Συμφωνώ

5 = Συμφωνώ απόλυτα

ΜΕΡΟΣ Α

Δημογραφικά Στοιχεία

1. Φύλο:

☐ Άνδρας

☐ Γυναίκα

☐ Άλλο / Δε θέλω να απαντήσω

2. Ηλικία:

☐ 18–25

☐ 26–35

☐ 36–45

☐ 46–60

☐ 60+

3. Τομέας απασχόλησης:

☐ Τραπεζικός / Χρηματοοικονομικός

☐ Τεχνολογία / IT

☐ Ακαδημαϊκός / Εκπαίδευση

☐ Άλλος (παρακαλώ προσδιορίστε): _____

4. Έχετε χρησιμοποιήσει κάποιο εργαλείο AI στη δουλειά σας;

☐ Ναι

☐ Όχι

☐ Δεν είμαι σίγουρος/η

ΜΕΡΟΣ Β

ΕΝΟΤΗΤΑ Α: Αντιληπτή Χρησιμότητα (PerceivedUsefulness)

Κωδ.	**Δήλωση**	**1 2 3 4 5**
PU1	Η χρήση της Τεχνητής Νοημοσύνης (AI) βελτιώνει την ποιότητα των αποφάσεών μου.	☐ ☐ ☐ ☐ ☐
PU2	Η AI ενισχύει την αποτελεσματικότητά μου στη διαχείριση κινδύνων.	☐ ☐ ☐ ☐ ☐
PU3	Η χρήση AI με βοηθά να εκτελώ τις εργασίες μου πιο γρήγορα.	☐ ☐ ☐ ☐ ☐
PU4	Η AI βελτιώνει τη συνολική απόδοση της εργασίας μου.	☐ ☐ ☐ ☐ ☐

ΕΝΟΤΗΤΑ Β: Αντιληπτή Ευκολία Χρήσης (Perceived Ease of Use)

Κωδ.	**Δήλωση**	**1 2 3 4 5**
PEOU1	Η εκμάθηση της χρήσης συστημάτων AI είναι εύκολη για μένα.	☐ ☐ ☐ ☐ ☐
PEOU2	Η αλληλεπίδραση με την AI είναι σαφής και κατανοητή.	☐ ☐ ☐ ☐ ☐
PEOU3	Μου είναι εύκολο να γίνω επιδέξιος/α στη χρήση της AI.	☐ ☐ ☐ ☐ ☐
PEOU4	Η χρήση της AI δεν απαιτεί πολλή προσπάθεια από μέρους μου.	☐ ☐ ☐ ☐ ☐

ΕΝΟΤΗΤΑ Γ: Εμπιστοσύνη στα Συστήματα AI (Trust in AI)

Κωδ.	**Δήλωση**	**1 2 3 4 5**
TR1	Εμπιστεύομαι ότι τα συστήματα AI λειτουργούν με αξιοπιστία.	☐ ☐ ☐ ☐ ☐
TR2	Πιστεύω ότι η AI λαμβάνει δίκαιες και αντικειμενικές αποφάσεις.	☐ ☐ ☐ ☐ ☐

Κωδ.	Δήλωση	1	2	3	4	5
TR3	Νιώθω ασφαλής όταν η AI χρησιμοποιείται στη λήψη αποφάσεων.	☐	☐	☐	☐	☐

ΕΝΟΤΗΤΑ Δ: Αντιληπτός Κίνδυνος (Perceived Risk)

Κωδ.	Δήλωση	1	2	3	4	5
PR1	Η χρήση της AI ενέχει κινδύνους για την εργασία μου.	☐	☐	☐	☐	☐
PR2	Η AI μπορεί να οδηγήσει σε λανθασμένες αποφάσεις.	☐	☐	☐	☐	☐
PR3	Είμαι επιφυλακτικός/ή σχετικά με την εφαρμογή της AI σε κρίσιμες διαδικασίες.	☐	☐	☐	☐	☐

ΕΝΟΤΗΤΑ Ε: Πρόθεση Χρήσης (Behavioral Intention)

Κωδ.	Δήλωση	1	2	3	4	5
BI1	Σκοπεύω να χρησιμοποιώ AI στο πλαίσιο της εργασίας μου.	☐	☐	☐	☐	☐
BI2	Αν ήταν στη διάθεσή μου, θα χρησιμοποιούσα AI σε τακτική βάση.	☐	☐	☐	☐	☐
BI3	Έχω θετική πρόθεση να ενσωματώσω την AI στις επαγγελματικές μου δραστηριότητες.	☐	☐	☐	☐	☐

Σας ευχαριστούμε θερμά για τη συμμετοχή σας!