



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ
ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
“ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ & ΥΠΗΡΕΣΙΕΣ”

**Οι ηθικές προεκτάσεις από την υιοθέτηση της τεχνητής
νοημοσύνης στον τραπεζικό τομέα – Μελέτη περίπτωσης της Apple
Card**

Από
Ηλία Αλεξάκη
Υποβάλλεται
για την εκπλήρωση των προϋποθέσεων λήψης
Μεταπτυχιακού Διπλώματος
στην ειδίκευση «Πληροφοριακής Διακυβέρνησης»
του ΠΜΣ “Πληροφορικά Συστήματα & Υπηρεσίες”
στο
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
Απρίλιος 2026

Επιβλέπων: Μιχαήλ Φιλιππάκης
Ακαδημαϊκή Θέση: Καθηγητής

Πανεπιστήμιο Πειραιώς. Κάτοχος όλων των δικαιωμάτων
University of Piraeus. All rights reserved.
Συγγραφέας / Author: Ηλίας Αλεξάκης

ΣΕΛΙΔΑ ΕΓΚΥΡΟΤΗΤΑΣ

Όνοματεπώνυμο Φοιτητή: Ηλίας Αλεξάκης

Τίτλος Μεταπτυχιακής Διπλωματικής Εργασίας: Οι ηθικές προεκτάσεις από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα – Μελέτη περίπτωσης της Apple Card

Η παρούσα Μεταπτυχιακή Διπλωματική Εργασία υποβάλλεται ως μερική εκπλήρωση των απαιτήσεων του Προγράμματος Μεταπτυχιακών Σπουδών “Πληροφοριακά Συστήματα & Υπηρεσίες” του Τμήματος Ψηφιακών Συστημάτων του Πανεπιστημίου Πειραιώς και εγκρίθηκε στις 08/04/2026 από τα μέλη της Εξεταστικής Επιτροπής.

Εξεταστική Επιτροπή

Επιβλέπων (Τμήμα Ψηφιακών Συστημάτων): Μιχαήλ Φιλιππάκης, Καθηγητής

Μέλος Εξεταστικής Επιτροπής: Νικόλαος Γεωργόπουλος, Καθηγητής

Μέλος Εξεταστικής Επιτροπής: Ευαγγελία Κοπανάκη, Καθηγήτρια

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΑΥΘΕΝΤΙΚΟΤΗΤΑΣ

Ο Ηλίας Αλεξάκης γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο «Οι ηθικές προεκτάσεις από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα – Μελέτη περίπτωσης της Apple Card», αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει, έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές παραπομπές και αναφορές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Επιπλέον δηλώνω υπεύθυνα ότι η συγκεκριμένη Μεταπτυχιακή Διπλωματική Εργασία έχει συγγραφεί από εμένα προσωπικά και δεν έχει υποβληθεί ούτε έχει αξιολογηθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου. Σε κάθε περίπτωση, αναληθούς ή ανακριβούς δηλώσεως, υπόκειμαι στις συνέπειες που προβλέπονται τις διατάξεις που προβλέπει η Ελληνική και Κοινωνική Νομοθεσία περί πνευματικής ιδιοκτησίας.

Ο ΔΗΛΩΝ

Όνοματεπώνυμο: Ηλίας Αλεξάκης

Αριθμός Μητρώου: ME2488

Υπογραφή:

Ευχαριστίες

Η φοίτησή μου στο Πρόγραμμα Μεταπτυχιακών Σπουδών και η εκπόνηση της παρούσας εργασίας αποτέλεσαν για εμένα μια σημαντική ακαδημαϊκή αλλά και προσωπική πρόκληση. Η ολοκλήρωση αυτού του κύκλου δεν θα ήταν δυνατή χωρίς την αμέριστη στήριξη, κατανόηση και υπομονή της συζύγου μου, στην οποία εκφράζω τις πιο θερμές μου ευχαριστίες για τη διαρκή της παρουσία και συμβολή σε όλη αυτή τη διαδρομή.

Παράλληλα, θα ήθελα να ευχαριστήσω εγκάρδια την οικογένειά μου, που μου παρείχε όλα τα εφόδια, ώστε να μπορώ σήμερα να ολοκληρώνω με επιτυχία αυτό το σημαντικό κεφάλαιο της ζωής μου.

Πίνακας Περιεχομένων

Περίληψη	6
Abstract	7
1. Εισαγωγή.....	8
1.1 Αντικείμενο, σκοπός και ερευνητικά ερωτήματα	8
1.2 Μεθοδολογία έρευνας	9
1.3 Διάρθρωση της εργασίας.....	10
2. Εννοιολογικό πλαίσιο της τεχνητής νοημοσύνης.....	12
2.1 Ορισμός της ανθρώπινης νοημοσύνης.....	12
2.2 Σύντομη ιστορική αναδρομή στην τεχνητή νοημοσύνη.....	14
2.3 Ορισμός της Τεχνητής Νοημοσύνης	16
2.4 Ταξινόμηση της τεχνητής νοημοσύνης.....	18
2.4.1 Ταξινόμηση βάσει του επιπέδου ευφυΐας	18
2.4.2 Ταξινόμηση βάσει του εύρους δυνατοτήτων	19
2.5 Τεχνολογίες της τεχνητής νοημοσύνης	20
2.5.1 Μηχανική μάθηση	21
2.5.2 Βαθιά μάθηση.....	23
2.5.3 Παραγωγική τεχνητή νοημοσύνη	25
2.5.4 Επεξεργασία φυσικής γλώσσας.....	26
3. Κανονιστικό και ηθικό πλαίσιο για την τεχνητή νοημοσύνη.....	28
3.1 Κανονιστικό πλαίσιο για την τεχνητή νοημοσύνη.....	28
3.2 Βασικές αρχές ηθικής και συστάσεις για την τεχνητή νοημοσύνη	30
3.2.1 Η έννοια της ηθικής.....	30
3.2.2 Επισκόπηση των αρχών τεχνητής νοημοσύνης του ΟΟΣΑ.....	32
3.2.3 Κατευθυντήριες γραμμές της Ευρωπαϊκής Επιτροπής	36
3.2.4 Σύσταση για την ηθική της τεχνητής νοημοσύνης από την UNESCO.....	40
4. Η τεχνητή νοημοσύνη στον τραπεζικό τομέα.....	46
4.1 Ψηφιακός μετασχηματισμός στον τραπεζικό τομέα.....	46
4.2 Χρήσεις της τεχνητής νοημοσύνης στον τραπεζικό τομέα.....	47
4.2.1 Εξυπηρέτηση πελατών και εξατομίκευση υπηρεσιών	49
4.2.2 Πρόληψη απάτης.....	50

4.2.3 Πίστωση και αξιολόγηση κινδύνου	51
4.2.4 Ανάλυση αγοράς και διαχείριση επενδύσεων	52
4.2.5 Ανάπτυξη και ασφάλεια πληροφοριακών συστημάτων	53
4.3 Οφέλη από την υιοθέτηση της τεχνητής νοημοσύνης για τον τραπεζικό τομέα.....	53
4.3.1 Λειτουργική και οικονομική αποδοτικότητα	54
4.3.2 Λήψη αποφάσεων	55
4.3.3 Εμπειρία πελάτη	56
4.4 Προκλήσεις και περιορισμοί από την υιοθέτηση της τεχνητής νοημοσύνης για τον τραπεζικό τομέα.....	56
5. Ανάλυση ηθικών ζητημάτων από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα.....	60
5.1 Ιδιωτικότητα και προστασία δεδομένων	60
5.2 Μεροληψία των συστημάτων τεχνητής νοημοσύνης	62
5.2.1 Τύποι μεροληψίας	63
5.2.2 Μεροληψία στα συστήματα τεχνητής νοημοσύνης του τραπεζικού τομέα.....	64
5.2.3 Τρόποι αντιμετώπισης.....	65
5.3 Λογοδοσία και εποπτεία της τεχνητής νοημοσύνης	66
5.4 Διαφάνεια και εξηγησιμότητα της τεχνητής νοημοσύνης	67
5.5 Επιπτώσεις στην απασχόληση από την τεχνητή νοημοσύνη	68
6. Μελέτη περίπτωσης της Apple Card	72
6.1 Χαρακτηριστικά της Apple Card	72
6.2 Αναφορές για μεροληψία.....	73
6.3 Ευρήματα από την έρευνα της ρυθμιστικής αρχής.....	74
6.4 Γενικές παρατηρήσεις από τη ρυθμιστική αρχή	76
7. Συζήτηση	78
8. Συμπεράσματα.....	81
Βιβλιογραφία	84

Περίληψη

Η παρούσα εργασία εξετάζει τις ηθικές προεκτάσεις από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα. Η ολοκληρωμένη προσέγγιση του αντικειμένου επιτυγχάνεται μέσω της παρουσίασης των βασικών χρήσεων της τεχνητής νοημοσύνης στις τραπεζικές λειτουργίες, δεδομένου ότι τα ηθικά διλήμματα δεν είναι δυνατόν να αποτιμηθούν ανεξάρτητα από τα πεδία εφαρμογής της τεχνητής νοημοσύνης.

Η επισκόπηση της σχετικής βιβλιογραφίας καταδεικνύει πως τα συστήματα τεχνητής νοημοσύνης χρησιμοποιούνται ευρέως σε κρίσιμους τομείς, όπως η εξυπηρέτηση πελατών και η εξατομίκευση υπηρεσιών, η πρόληψη και η ανίχνευση απάτης, η αξιολόγηση πιστοληπτικής ικανότητας και η διαχείριση κινδύνου, η ανάλυση αγοράς και η διαχείριση επενδύσεων, καθώς και η ασφάλεια και ανάπτυξη των τραπεζικών πληροφοριακών συστημάτων. Η αξιοποίηση της τεχνητής νοημοσύνης σε αυτούς τους τομείς εγείρει κρίσιμα ηθικά ζητήματα που αφορούν την ιδιωτικότητα και την προστασία των δεδομένων, τη μεροληψία στις αλγοριθμικές αποφάσεις, τη λογοδοσία και την ανάγκη ουσιαστικής ανθρώπινης εποπτείας, τη διαφάνεια και εξηγησιμότητα των συστημάτων, καθώς και τις πιθανές επιπτώσεις στην απασχόληση στον τραπεζικό κλάδο.

Στο πλαίσιο εμβάθυνσης του αντικειμένου, εξετάζεται η περίπτωση της Apple Card η οποία αναδεικνύει, κυρίως, ζητήματα σχετικά με τη λογοδοσία, την ανθρώπινη εποπτεία, τη διαφάνεια και την εξηγησιμότητα των αλγοριθμικών συστημάτων. Από την ανάλυση προκύπτει ότι η τυπική συμμόρφωση με τις εκάστοτε κανονιστικές απαιτήσεις δεν επαρκεί για την ουσιαστική αντιμετώπιση των ηθικών προκλήσεων. Αντιθέτως, απαιτείται η ανάπτυξη ολοκληρωμένων στρατηγικών για την προληπτική και υπεύθυνη διαχείρισή τους μέσω της θεσμοθέτησης πελατοκεντρικών μηχανισμών διακυβέρνησης και ελέγχου της τεχνητής νοημοσύνης. Τέλος, προτείνεται η υιοθέτηση υβριδικών ρυθμιστικών προσεγγίσεων, οι οποίες επιτρέπουν την προσαρμογή στις τεχνολογικές εξελίξεις, διατηρώντας ταυτόχρονα τη ρυθμιστική συνέπεια και τη διασφάλιση των ηθικών αρχών που διέπουν τη λειτουργία του χρηματοπιστωτικού συστήματος.

Abstract

This paper examines the ethical implications stemming from the adoption of artificial intelligence within the banking sector. A comprehensive approach to the subject is achieved through the presentation of the main applications of artificial intelligence in banking operations, as ethical dilemmas cannot be assessed independently of the specific domains in which artificial intelligence is implemented.

A review of the relevant literature reveals that artificial intelligence systems are extensively utilized in key areas such as customer service and service personalization, fraud prevention and detection, credit scoring and risk management, market analysis and investment management, as well as the development and security of banking information systems. The implementation of artificial intelligence across these domains raises significant ethical concerns related to privacy and data protection, algorithmic bias, accountability and the necessity of meaningful human oversight, system transparency and explainability, and the potential effects on employment within the banking industry.

The study further examines the case of the Apple Card, which illustrates challenges associated, mainly, with accountability, human supervision, transparency and the explainability of algorithmic systems. The analysis suggests that mere regulatory compliance is insufficient to adequately address the complex ethical challenges arising from artificial intelligence adoption. Instead, the development of comprehensive strategies for the proactive and responsible management of such challenges is required, including the introduction of customer-centric AI governance and control mechanisms. Finally, the adoption of hybrid regulatory approaches is proposed, enabling adaptation to technological developments while maintaining regulatory consistency and safeguarding the ethical principles that underpin the functioning of the financial system.

1. Εισαγωγή

Η παρουσίαση του ChatGPT το 2022 αποτέλεσε σημείο καμπής στη δημόσια συζήτηση γύρω από την τεχνητή νοημοσύνη, χωρίς ωστόσο να σηματοδοτεί την αφετηρία της σχετικής έρευνας. Η συστηματική επιστημονική ενασχόληση και η ανάπτυξη αντίστοιχων συστημάτων ξεκινά ήδη από τη δεκαετία του 1950, μέσα από προσπάθειες θεωρητικής θεμελίωσης και πρακτικής εφαρμογής. Σήμερα, η τεχνητή νοημοσύνη αναγνωρίζεται ως μετασχηματιστική τεχνολογία γενικού σκοπού, με δυνητικό αντίκτυπο συγκρίσιμο με εκείνον της αξιοποίησης του ηλεκτρισμού και της εμφάνισης του διαδικτύου (Condoleezza Rice κ.ά., 2025).

Ο τραπεζικός τομέας αποτελεί ένα από τους πρώτους τομείς που υιοθέτησαν συστηματικά την τεχνητή νοημοσύνη, επιδιώκοντας την αξιοποίηση των πλεονεκτημάτων της σε ένα περιβάλλον έντονου ανταγωνισμού και αυξημένων ρυθμιστικών απαιτήσεων. Η χρήση τέτοιων συστημάτων υπόσχεται τον επαναπροσδιορισμό του τρόπου με τον οποίο οι πελάτες αλληλοεπιδρούν με τα χρηματοπιστωτικά ιδρύματα, οδηγώντας στην ανάπτυξη καινοτόμων προϊόντων και υπηρεσιών, στην ενίσχυση της ασφάλειας των συναλλαγών, στη βελτιωμένη προστασία των δεδομένων και στην αποτελεσματικότερη πρόληψη φαινομένων οικονομικής απάτης. Αναμένεται η παροχή οικονομικότερων υπηρεσιών, η βελτιστοποίηση των επιχειρησιακών διαδικασιών και σημαντική εξοικονόμηση πόρων (Ridzuan κ.ά., 2024).

1.1 Αντικείμενο, σκοπός και ερευνητικά ερωτήματα

Το αντικείμενο της παρούσας μεταπτυχιακής εργασίας είναι η παρουσίαση και ανάλυση των ηθικών προεκτάσεων που απορρέουν από την υιοθέτηση των συστημάτων τεχνητής νοημοσύνης στον τραπεζικό τομέα. Η επιλογή του συγκεκριμένου θέματος βασίζεται στη μετασχηματιστική δυναμική της τεχνητής νοημοσύνης και στην ολοένα αυξανόμενη εξάρτηση των τραπεζικών ιδρυμάτων από συστήματα αυτού του τύπου για κρίσιμες λειτουργίες, όπως η εξυπηρέτηση πελατών, η πρόληψη απάτης, η αξιολόγηση πιστωτικού κινδύνου, η ανάλυση αγορών, η διαχείριση επενδύσεων και η ανάπτυξη και ασφάλεια των τραπεζικών πληροφοριακών συστημάτων. Η ταχεία υιοθέτηση τέτοιων συστημάτων εντείνει τους κινδύνους αναπαραγωγής ανισοτήτων, αδιαφάνειας και υπονόμησης θεμελιωδών δικαιωμάτων, ιδίως όταν απουσιάζουν επαρκείς μηχανισμοί ελέγχου, εποπτείας και διακυβέρνησης.

Η εργασία έχει σκοπό να αναδείξει τις ηθικές προεκτάσεις από την υιοθέτηση της τεχνητής στον τραπεζικό κλάδο, μέσα από την αναλυτική παρουσίασή τους, ενώ για την πληρέστερη κατανόηση των σχετικών εννοιών και προκλήσεων επιλέχθηκε η μελέτη περίπτωσης της Apple Card, η οποία αποτελεί το πρώτο χρηματοπιστωτικό προϊόν της Apple σε συνεργασία με την Goldman Sachs και συνιστά χαρακτηριστικό παράδειγμα χρήσης αλγοριθμικών συστημάτων σε διαδικασίες πίστωσης.

Η εργασία επιδιώκει να απαντήσει σε δύο ερωτήματα:

- Πώς ο τραπεζικός τομέας χρησιμοποιεί την τεχνητή νοημοσύνη και ποια είναι τα οφέλη και οι προκλήσεις που απορρέουν;
- Ποια είναι τα ηθικά ζητήματα που προκύπτουν από τη χρήση της τεχνητής νοημοσύνης στον τραπεζικό τομέα;

Η ανάλυση των ηθικών παραμέτρων της τεχνητής νοημοσύνης προϋποθέτει την κατανόηση των επιμέρους χρήσεων της, καθώς τα ηθικά διλήμματα δεν μπορούν να αποτιμηθούν ανεξάρτητα από τα πεδία εφαρμογής και τους τρόπους ενσωμάτωσής τους στις τραπεζικές πρακτικές. Ως εκ τούτου, το πρώτο ερευνητικό ερώτημα λειτουργεί ως σκαλοπάτι για την απάντηση του δεύτερου ερωτήματος, το οποίο αποτελεί και τον κεντρικό άξονα της εργασίας.

1.2 Μεθοδολογία έρευνας

Η παρούσα μεταπτυχιακή εργασία υιοθετεί μία θεωρητική μεθοδολογική προσέγγιση, εστιάζοντας στην αναλυτική βιβλιογραφική επισκόπηση, καθώς και στην παράθεση μίας στοχευμένης μελέτης περίπτωσης. Οι πηγές που αξιοποιήθηκαν περιλαμβάνουν επιστημονικά βιβλία αναγνωρισμένων συγγραφέων, δημοσιεύσεις και άρθρα από διεθνώς έγκριτες ακαδημαϊκές βάσεις δεδομένων, με την πλειονότητά τους να έχει δημοσιευθεί την τελευταία πενταετία. Επιπλέον, συμπεριλήφθηκαν επίσημες εκθέσεις διεθνών οργανισμών, σχετικές νομοθετικές πράξεις που αφορούν την τεχνητή νοημοσύνη, καθώς και περιορισμένος αριθμός έγκριτων δημοσιογραφικών άρθρων.

Η αναζήτηση της βιβλιογραφίας πραγματοποιήθηκε με τη χρήση των μηχανών αναζήτησης γενικού περιεχομένου Google και Bing, των ακαδημαϊκών αποθετηρίων Google Scholar και Science Direct και του εργαλείου τεχνητής νοημοσύνης Perplexity. Οι βασικές λέξεις κλειδιά που χρησιμοποιήθηκαν τόσο στα ελληνικά όσο και στα αγγλικά είναι:

- τεχνητή νοημοσύνη
- τραπεζικός τομέας
- ψηφιακός μετασχηματισμός
- ηθικές προεκτάσεις/ηθικές προκλήσεις

Από το εργαλείο Perplexity ζητήθηκαν αποκλειστικά πηγές από αναγνωρισμένα ακαδημαϊκά περιοδικά, διεθνείς οργανισμούς και ακαδημαϊκά αποθετήρια, οι οποίες στη συνέχεια ελέγχθηκαν σχολαστικά από τον συγγραφέα ως προς την εγκυρότητα, τη θεματική σχετικότητα και την ακαδημαϊκή τους αποδοχή. Πηγές χαμηλής αξιοπιστίας, ασαφούς προέλευσης ή μη δημοσιευμένες σε ή από διεθνώς αναγνωρισμένα fora απορρίφθηκαν, εξασφαλίζοντας έτσι την επιστημονική υπευθυνότητα και τεκμηρίωση. Επιπλέον, χρησιμοποιήθηκε για να βελτιώσει τη συντακτική δομή ορισμένων αποσπασμάτων της εργασίας. Μετά τη χρήση του, ο συγγραφέας επανεξέτασε και επεξεργάστηκε την εργασία και αναλαμβάνει την πλήρη ευθύνη για το περιεχόμενό της.

Η μεθοδολογία αυτή κρίθηκε κατάλληλη για την επίτευξη των στόχων της εργασίας, επιτρέποντας τη συστηματική σύνθεση της βιβλιογραφίας και παρουσίασης της μελέτης περίπτωσης για την Apple Card.

1.3 Διάρθρωση της εργασίας

Η εργασία οργανώνεται σε επτά κύρια κεφάλαια, τα οποία δομούν σταδιακά το πλήρες πλαίσιο απαντώντας στα ερευνητικά ερωτήματα που έχουν τεθεί.

Στο Κεφάλαιο 2 παρουσιάζεται το εννοιολογικό πλαίσιο της τεχνητής νοημοσύνης, με έμφαση να δίνεται στους ορισμούς της ανθρώπινης και της τεχνητής νοημοσύνης, στην ιστορική αναδρομή στην τεχνητή νοημοσύνη, στις ταξινομήσεις της και τις βασικές τεχνολογίες που χρησιμοποιούνται από τον τραπεζικό τομέα.

Στο Κεφάλαιο 3 παρουσιάζεται το νομοθετικό και ηθικό πλαίσιο για την τεχνητή νοημοσύνη. Γίνεται αναφορά στις διαφορετικές κανονιστικές προσεγγίσεις, βάσει αρχών και βάσει κανόνων, χρησιμοποιώντας ως παραδείγματα τη νομοθεσία EU AI Act και την προσέγγιση των ΗΠΑ. Στα ηθικά πλαίσια παρουσιάζονται οι βασικές ηθικές αρχές και συστάσεις διεθνών αναγνωρισμένων οργανισμών όπως ο ΟΟΣΑ, η Ευρωπαϊκή Επιτροπή και η UNESCO.

Στο Κεφάλαιο 4 εξετάζεται η τεχνητή νοημοσύνη στον τραπεζικό τομέα, με έμφαση στις διαφορετικές χρήσεις της, τα οφέλη που προκύπτουν από αυτές αλλά και τις προκλήσεις και τους περιορισμούς που αναδύονται.

Στο Κεφάλαιο 5 παρουσιάζονται αναλυτικά τα βασικά ηθικά ζητήματα από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα με αναφορά σε θεωρητικά πλαίσια και διεθνείς πρακτικές.

Στο Κεφάλαιο 6 αναπτύσσεται η μελέτη περίπτωσης της Apple Card, όπου παρουσιάζονται τα χαρακτηριστικά του προϊόντος, οι αναφορές για μεροληψία στις αποφάσεις του πιστωτικού αλγορίθμου τεχνητής νοημοσύνης, τα ευρήματα της ρυθμιστικής αρχής και οι γενικές παρατηρήσεις για τη χρήση της τεχνητής νοημοσύνης σε πιστωτικές υπηρεσίες.

Στο Κεφάλαιο 7 ακολουθεί η Συζήτηση, όπου παρουσιάζονται και αναλύονται τα ευρήματα από την περίπτωση της Apple Card σε σύγκριση με το θεωρητικό πλαίσιο που παρουσιάστηκε στην εργασία.

Τέλος, στο Κεφάλαιο 8 παρουσιάζονται τα συμπεράσματα της εργασίας, συνοψίζονται οι βασικές απαντήσεις στα ερευνητικά ερωτήματα και διατυπώνονται προτάσεις για μελλοντική έρευνα και για τη διαμόρφωση πολιτικών διακυβέρνησης γύρω από την ηθική χρήση της τεχνητής νοημοσύνης στον τραπεζικό τομέα.

2. Εννοιολογικό πλαίσιο της τεχνητής νοημοσύνης

Η τεχνητή νοημοσύνη αποτελεί μια θεμελιώδη και μετασχηματιστική τεχνολογία, αντίστοιχης σημασίας με την εμφάνιση του διαδικτύου, καθώς ενισχύει την εξέλιξη άλλων επιστημονικών κλάδων και διαθέτει τη δυναμική να μεταμορφώσει τις κοινωνίες, τις οικονομίες και τα πολιτικά συστήματα σε παγκόσμιο επίπεδο (Condoleezza Rice κ.ά., 2025). Στο παρόν κεφάλαιο αναλύεται το εννοιολογικό πλαίσιο της τεχνητής νοημοσύνης, αρχικά μέσα από τον προσδιορισμό της ανθρώπινης νοημοσύνης. Ακολουθεί μια ιστορική αναδρομή της εξέλιξης του πεδίου και μια πιο εκτενής προσέγγιση του ορισμού της τεχνητής νοημοσύνης. Τέλος, παρουσιάζονται οι βασικές ταξινομήσεις και οι συναφείς τεχνολογίες, διαμορφώνοντας το θεωρητικό υπόβαθρο για την εις βάθος διερεύνηση των ηθικών προεκτάσεων της στον τραπεζικό τομέα.

2.1 Ορισμός της ανθρώπινης νοημοσύνης

Η ανθρώπινη νοημοσύνη αποτελεί ένα από τα πιο ενδιαφέροντα θέματα συζήτησης του τομέα της ψυχολογίας (Gignac & Szodorai, 2024). Οι ορισμοί της τείνουν να διαφέρουν σε ακριβείς όρους, ωστόσο υπάρχει σύγκλιση γύρω από ορισμένες βασικές έννοιες. Συγκεκριμένα, η ανθρώπινη νοημοσύνη θα μπορούσε να ορισθεί ως η ικανότητα ενός ατόμου να συλλογίζεται, να επιλύει προβλήματα, να κατανοεί έννοιες και να μαθαίνει αποτελεσματικά. Επιπλέον, η νοημοσύνη αναφέρεται στην ικανότητα ενός ατόμου να προσαρμόζεται στο περιβάλλον, να αλλάζει και να χρησιμοποιεί τις διαθέσιμες πληροφορίες για να λαμβάνει τεκμηριωμένες αποφάσεις (Morandín-Ahuerma, 2022). Προκειμένου ο ορισμός να θεωρηθεί επιστημονικά αποδεκτός, οφείλει να επιτυγχάνει μία λεπτή ισορροπία μεταξύ γενίκευσης και εννοιολογικής ακρίβειας, ώστε να διευκολύνει την ενδελεχή θεωρητική κατανόηση του τι συνιστά και τι δε συνιστά νοημοσύνη (Gignac & Szodorai, 2024). Η Αμερικανική Ψυχολογική Εταιρεία ορίζει την νοημοσύνη ως την ικανότητα να αντλεί κάποιος πληροφορίες, να μαθαίνει από την εμπειρία, να προσαρμόζεται στο περιβάλλον, να κατανοεί και να χρησιμοποιεί σωστά τη σκέψη και τη λογική (American Psychological Association, 2018). Ο Wechsler περιέγραψε τη νοημοσύνη ως τη συνολική ή παγκόσμια ικανότητα του ατόμου να ενεργεί σκόπιμα, να σκέφτεται λογικά και να αντιμετωπίζει αποτελεσματικά το περιβάλλον του (Wechsler, 1944). Αντίστοιχα, ο Gignac όρισε αρχικά την ανθρώπινη νοημοσύνη ως τη μέγιστη ικανότητα ενός ατόμου να επιτύχει έναν καινούργιο

στόχο χρησιμοποιώντας αντιληπτικό-γνωστικές διεργασίες. Με τον ορισμό αυτό υπογραμμίστηκαν τρία σημαντικά χαρακτηριστικά:

- πρώτον, η νοημοσύνη ενός ατόμου εξετάζεται στο πλαίσιο της μέγιστης ικανότητάς του να επιλύει καινοτόμα προβλήματα, και όχι της τυπικής εκδήλωσης της νοημοσύνης του·
- δεύτερον, η ουσία της ανθρώπινης νοημοσύνης συνδέεται στενά με την εφαρμογή της σε καινοτόμα πλαίσια για την επίλυση προβλημάτων που ένα άτομο δεν έχει αντιμετωπίσει στο παρελθόν·
- τρίτον, η ανθρώπινη νοημοσύνη υποστηρίζεται από αντιληπτικές-γνωστικές λειτουργίες οι οποίες, σε βασικό επίπεδο, περιλαμβάνουν μια σειρά ψυχικών διεργασιών, όπως η προσοχή, η οπτική αντίληψη, η ακουστική αντίληψη και η αισθητηριακή ολοκλήρωση (Gignac & Szodorai, 2024).

Αναγνωρίζοντας την ανάγκη βελτίωσης του αρχικού του ορισμού, ο Gignac όρισε εκ νέου την ανθρώπινη νοημοσύνη ως τη μέγιστη ικανότητα ενός ατόμου να ολοκληρώσει μια νέα τυποποιημένη δοκιμασία με αληθή βαθμολογία χρησιμοποιώντας αντιληπτικές-γνωστικές διεργασίες. Ο νέος ορισμός υπογραμμίζει την ανάγκη ύπαρξης μιας τυποποιημένης δοκιμασίας με σαφείς οδηγίες και διαδικασίες στην οποία το άτομο δεν έχει εκτεθεί νωρίτερα, ενώ η αληθής βαθμολόγηση υπογραμμίζει τη σημασία της αντικειμενικής αξιολόγησης και ερμηνείας των απαντήσεων πάνω σε επαληθεύσιμα κριτήρια για τα οποία υπάρχει ευρεία συναίνεση (Gignac & Szodorai, 2024).

Ένα ακόμα σημαντικό ζήτημα αποτέλεσε η ύπαρξη ή μη διαφορετικών τύπων νοημοσύνης. Η συγκεκριμένη θεωρία διατυπώθηκε για πρώτη φορά από τον Αμερικανό ψυχολόγο Howard Gardner, ο οποίος αναφέρθηκε στην ύπαρξη διαφορετικών ειδών νοημοσύνης, όπως η λογικομαθηματική νοημοσύνη, η λεκτική-γλωσσική νοημοσύνη, η σωματική-κινησιολογική νοημοσύνη κ.α. Ωστόσο, στην πράξη έχει αποδειχθεί πώς η νοημοσύνη είναι ενιαία και απλά εφαρμόζεται σε διαφορετικά πλαίσια (Morandín-Ahuerma, 2022). Η κατανόηση της έννοιας της ανθρώπινης νοημοσύνης και των χαρακτηριστικών της καθίσταται αναγκαία για την ορθότερη προσέγγιση της τεχνητής νοημοσύνης, των διαφορετικών ορισμών και των θεμελιωδών διαφορών με την ανθρώπινη νοημοσύνη.

2.2 Σύντομη ιστορική αναδρομή στην τεχνητή νοημοσύνη

Η έννοια της τεχνητής νοημοσύνης εισήχθη για πρώτη φορά τη δεκαετία του 1950, όταν ο Alan Turing δημοσίευσε το άρθρο «Computing Machinery and Intelligence». Στο πρώτο κεφάλαιο του έργου του, ο Turing θέτει το ερώτημα αν μια μηχανή μπορεί να σκεφτεί και προχωρώντας στην ανάλυση αυτού του ερωτήματος το επαναδιατυπώνει μέσα από ένα πείραμα που αποκαλεί «παιχνίδι μίμησης» (Turing, 1950). Ο σκοπός του παιχνιδιού, το οποίο είναι ευρύτερα γνωστό ως «τεστ του Turing», έγκειται στο κατά πόσο μια μηχανή επιδεικνύει ανθρώπινη νοημοσύνη. Κατά τη διεξαγωγή του τεστ, ένας ανακριτής απευθύνει ερωτήσεις σε έναν άνθρωπο και σε έναν υπολογιστή, χωρίς να γνωρίζει την ταυτότητα του καθενός, και καλείται με βάση τις απαντήσεις τους να αναγνωρίσει τη μηχανή. Αν ο ανακριτής αποτύχει να τη διακρίνει, τότε τεκμαίρεται ότι ο υπολογιστής εμφανίζει ανθρώπινη νοημοσύνη (Morandín-Ahuerma, 2022; Turing, 1950).

Το πρώτο βασικό πρόγραμμα τεχνητής νοημοσύνης αναπτύχθηκε από τον Christopher Strachey το 1951, ενώ μία βελτιωμένη εκδοχή του το 1952 ήταν σε θέση να παίζει το παιχνίδι της ντάμας με επαρκή αποτελεσματικότητα. Λίγο αργότερα, ο Strachey δημιούργησε ένα νέο πρόγραμμα το οποίο προσομοίαζε έναν υποθετικό κόσμο στον οποίο υπήρχαν οκτώ μόνο καταστάματα· στο πλαίσιο αυτό, όταν το πρόγραμμα λάμβανε εντολή να αγοράσει ένα προϊόν, διερευνούσε συστηματικά όλα τα καταστάματα έως ότου εντοπίσει το επιθυμητό αντικείμενο. Κατά την επανάληψη της διαδικασίας, όταν το ίδιο προϊόν απαιτούνταν εκ νέου, η μηχανή προσέγγιζε άμεσα το κατάστημα όπου το είχε προηγουμένως εντοπίσει, καταδεικνύοντας στοιχειώδη ικανότητα απομνημόνευσης και προσαρμοστικής μάθησης, λειτουργίες παρόμοιες με αυτές του ανθρώπινου εγκεφάλου (Ahmed, 2022).

Το 1955, ο καθηγητής του πανεπιστημίου Stanford, John McCarthy, εισήγαγε για πρώτη φορά τον όρο «Τεχνητή Νοημοσύνη», ορίζοντάς την ως την επιστήμη και τη μηχανική της κατασκευής ευφύων μηχανών (Taylor κ.ά., 2025). Τον όρο αυτό παρουσίασε στο συνέδριο του Dartmouth το 1956, το οποίο αποτελεί ορόσημο στην απαρχή της πρώτης περιόδου ανάπτυξης της τεχνητής νοημοσύνης, η οποία διήρκεσε έως το 1970. Η περίοδος αυτή διακρίνεται από τον διάχυτο ενθουσιασμό που επικρατούσε, καθώς για πρώτη φορά οι υπολογιστές μπορούσαν να λύνουν προβλήματα όπως οι άνθρωποι, παρουσιάζοντας δείγματα ανθρώπινης ευφυΐας. Η ερευνητική κοινότητα συμερίστηκε τον ενθουσιασμό, διατυπώνοντας ιδιαίτερα αισιόδοξες προβλέψεις για το μέλλον, ωστόσο σύντομα η τεχνητή

νοημοσύνη αντιμετώπισε τεχνολογικά εμπόδια, κυρίως, λόγω των περιορισμών της υπολογιστικής ισχύος, της μνήμης και της ταχύτητας επεξεργασίας. Άμεση συνέπεια αποτέλεσε η δραματική μείωση της χρηματοδότησης και των δραστηριοτήτων της τεχνητής νοημοσύνης τόσο στη βιομηχανία όσο και στον ακαδημαϊκό τομέα, που οδήγησε στον πρώτο χειμώνα της τεχνητής νοημοσύνης (Delipetrev κ.ά., 2020).

Η δεύτερη περίοδος της τεχνητής νοημοσύνης εκτείνεται χρονικά από το 1970 έως το 1990. Κατά τη διάρκεια αυτής της περιόδου, το ενδιαφέρον στράφηκε προς τα λεγόμενα «εξειδικευμένα συστήματα» (expert systems), τα οποία βασίζονταν στην ιδέα της μετατροπής της ανθρώπινης εμπειρογνωμοσύνης σε μορφή κατάλληλη για επεξεργασία από υπολογιστή, ώστε να καταστεί δυνατή η διάδοση της μέσω λογισμικού σε πολλαπλούς προσωπικούς υπολογιστές. Παρά τη σημαντική διάδοση τέτοιων συστημάτων τη δεκαετία του 1990, η διαθέσιμη τεχνολογία αποδείχθηκε ανεπαρκής, καθώς δε διέθετε την εκφραστική ικανότητα που απαιτούνταν για την αποτύπωση του εύρους της ανθρώπινης γνώσης. Επιπρόσθετα, η διαδικασία απόκτησης και κωδικοποίησης της εξειδικευμένης γνώσης αποτέλεσε ένα από τα κυριότερα εμπόδια στην περαιτέρω ανάπτυξή τους, γεγονός που οδήγησε τελικά στον δεύτερο χειμώνα της τεχνητής νοημοσύνης και στην κατάρρευση ενός κλάδου που τότε αποτιμάτο σε περίπου 500 εκατομμύρια δολάρια (Delipetrev κ.ά., 2020).

Η τρίτη περίοδος της τεχνητής νοημοσύνης ξεκίνησε τη δεκαετία του 1990 και συνεχίζεται μέχρι και σήμερα. Η έναρξη της περιόδου χαρακτηρίστηκε από την αξιοσημείωτη αύξηση της υπολογιστικής ισχύος και της διαθεσιμότητας δεδομένων. Κατά τη διάρκεια αυτής της περιόδου, ο τομέας της μηχανικής μάθησης (machine learning) γνώρισε ραγδαία ανάπτυξη, ενώ τη δεκαετία του 2010 ακολούθησε η εδραίωση της βαθιάς μάθησης (deep learning), η οποία ενίσχυσε τόσο το εύρος όσο και την απόδοση των εφαρμογών τεχνητής νοημοσύνης (Crisanto κ.ά., 2024; N. Kaur κ.ά., 2020). Η τεχνητή νοημοσύνη άρχισε να αντιμετωπίζει σύνθετα προβλήματα, προσφέροντας αποδοτικές λύσεις σε ένα ευρύ φάσμα πεδίων, όπως η εξόρυξη δεδομένων, η βιομηχανική ρομποτική, τα τραπεζικά συστήματα, η ιατρική διάγνωση, τα συστήματα προτάσεων και οι μηχανές αναζήτησης. Η ανάπτυξη μιας ενιαίας μαθηματικής και υπολογιστικής γλώσσας διευκόλυνε τη διεπιστημονική συνεργασία με καθιερωμένους επιστημονικούς κλάδους και συνέβαλε αποτελεσματικά στη διαμόρφωση

ενός περισσότερο συνεκτικού και μεθοδολογικά ακριβούς επιστημονικού πλαισίου για την τεχνητή νοημοσύνη (Delipetrev κ.ά., 2020).

2.3 Ορισμός της Τεχνητής Νοημοσύνης

Παρά τη σχετικά μακρά ιστορία της τεχνητής νοημοσύνης, δεν υπάρχει ακόμη ένας ενιαίος, καθολικά αποδεκτός ορισμός της. Έγκυρες ακαδημαϊκές, θεσμικές και επιχειρηματικές πηγές έχουν δώσει τους δικούς τους ορισμούς, οι οποίοι μπορούν να χωριστούν σε τρεις κυρίως κατηγορίες:

- i. Ορισμοί που βασίζονται στην επίτευξη στόχων σε ένα σύνθετο περιβάλλον
- ii. Ορισμοί που βασίζονται στη μάθηση και προσαρμογή
- iii. Ορισμοί που βασίζονται στη μίμηση των ανθρώπινων ικανοτήτων (Weitzel, 2025).

Οι ορισμοί της τεχνητής νοημοσύνης που θεμελιώνονται στην επίτευξη στόχων εντός σύνθετων ή μεταβαλλόμενων περιβαλλόντων αποδεικνύονται ανεπαρκείς, δεδομένου ότι οι ίδιοι οι στόχοι μεταβάλλονται διαρκώς. Το φαινόμενο αυτό έχει αποτυπωθεί στη βιβλιογραφία ως *φαινόμενο της «τεχνητής νοημοσύνης»* και αποδίδεται συχνά με τη φράση *«τεχνητή νοημοσύνη είναι ό,τι δεν έχουν ακόμη κατορθώσει οι μηχανές»*. Η διαθέσιμη βιβλιογραφία επιβεβαιώνει την ύπαρξη του φαινομένου, καταδεικνύοντας ότι, όταν ένα υπολογιστικό σύστημα κατορθώνει να εκτελέσει επιτυχώς μια δεδομένη εργασία, η τελευταία παύει να θεωρείται ένδειξη νοημοσύνης και αντιμετωπίζεται ως απλή υπολογιστική λειτουργία (Weitzel, 2025).

Οι ορισμοί που επικεντρώνονται στη μάθηση και την προσαρμογή αντιμετωπίζουν δύο διακριτές προκλήσεις. Πρώτον, υπόκεινται στα ίδια μεταβαλλόμενα κριτήρια και σκοπούς που προαναφέρθηκαν. Δεύτερον, κάθε σφάλμα ερμηνεύεται ως ένδειξη αποτυχίας προσαρμογής, με αποτέλεσμα συστήματα που επιδεικνύουν βασικά λάθη να θεωρούνται ανεπαρκή ως προς την ικανότητά τους να μαθαίνουν ή να προσαρμόζονται. Ως εκ τούτου, τέτοια συστήματα δεν ανταποκρίνονται επαρκώς στον συγκεκριμένο ορισμό (Weitzel, 2025).

Τέλος, η προσέγγιση που καθορίζει την τεχνητή νοημοσύνη με γνώμονα τον βαθμό ομοιότητάς της προς τις ανθρώπινες γνωστικές ή συμπεριφορικές ικανότητες, αν και λειτουργική στο πλαίσιο της ερευνητικής ανάλυσης, αποδεικνύεται προβληματική από την οπτική της κανονιστικής συμμόρφωσης. Η σύνδεση της τεχνητής νοημοσύνης με τα χαρακτηριστικά του ανθρώπου ενδέχεται να οδηγήσει στη διαμόρφωση πολιτικών

υπερβολικά ανεκτών για τα αδύναμα, ανθρωπομορφικά μοντέλα και ταυτόχρονα ανεπαρκώς περιεκτικών για τα ισχυρότερα, τα οποία δεν παρουσιάζουν εμφανή ανθρώπινα γνωρίσματα (Weitzel, 2025).

Στο πλαίσιο της παρούσας εργασίας, για τον ορισμό της τεχνητής νοημοσύνης ο συγγραφέας παραθέτει τους ορισμούς που προέρχονται από τον Οργανισμό Οικονομικής Συνεργασίας και Ανάπτυξης (ΟΟΣΑ) και από τη νομοθεσία της Ευρωπαϊκής Ένωσης σχετικά με την τεχνητή νοημοσύνη (EU AI Act). Ο ορισμός του ΟΟΣΑ λόγω του μη δεσμευτικού του χαρακτήρα και της ευρείας αποδοχής του, αποτελεί τη βάση για τη διαμόρφωση ηθικών αρχών και εθελοντικών πλαισίων πολιτικών διακυβέρνησης της τεχνητής νοημοσύνης (Crisanto κ.ά., 2024). Αντίθετα, ο ορισμός της Ευρωπαϊκής Ένωσης είναι νομικά δεσμευτικός και εντάσσεται σε ένα υποχρεωτικό κανονιστικό πλαίσιο που διέπει τη λειτουργία των συστημάτων τεχνητής νοημοσύνης εντός της (European Commission, 2025; Regulation (EU) 2024/1689, 2024).

Σύμφωνα με τον ΟΟΣΑ, ως σύστημα τεχνητής νοημοσύνης ορίζεται *«ένα μηχανικό σύστημα το οποίο, για ρητούς ή άρρητους στόχους, συνάγει από τα δεδομένα εισόδου που λαμβάνει τον τρόπο παραγωγής αποτελεσμάτων, όπως προβλέψεις, περιεχόμενο, συστάσεις ή αποφάσεις, τα οποία μπορούν να επηρεάσουν φυσικά ή εικονικά περιβάλλοντα. Διαφορετικά συστήματα τεχνητής νοημοσύνης διαφοροποιούνται ως προς τα επίπεδα αυτονομίας και προσαρμοστικότητας μετά την ενεργοποίησή τους»* (OECD, 2024b). Ο κανονισμός της Ευρωπαϊκής Ένωσης, αντιστοίχως, ορίζει ως σύστημα τεχνητής νοημοσύνης *«ένα μηχανικό σύστημα το οποίο έχει σχεδιαστεί ώστε να λειτουργεί με διάφορα επίπεδα αυτονομίας και μπορεί να επιδεικνύει προσαρμοστικότητα μετά την ενεργοποίησή του, και το οποίο, για ρητούς ή άρρητους στόχους, συνάγει από τις εισόδους που λαμβάνει τον τρόπο με τον οποίο θα παράγει εξόδους, όπως προβλέψεις, περιεχόμενο, συστάσεις ή αποφάσεις, που μπορούν να επηρεάσουν φυσικά ή εικονικά περιβάλλοντα»* (European Commission, 2025; Regulation (EU) 2024/1689, 2024). Οι δύο ορισμοί, αν και δεν είναι ταυτόσημοι, προσεγγίζουν την τεχνητή νοημοσύνη ως ένα μηχανικό σύστημα που βασίζεται στη χρήση δεδομένων για την παραγωγή αποτελεσμάτων και χαρακτηρίζεται από διαφορετικά επίπεδα αυτονομίας και ικανότητας προσαρμογής στο περιβάλλον λειτουργίας του.

Η απουσία ενός παγκοσμίως αποδεκτού ορισμού της τεχνητής νοημοσύνης δυσχεραίνει την κατανόηση των επιμέρους περιπτώσεων χρήσης της στον διεθνή χρηματοπιστωτικό τομέα και περιορίζει τη δυνατότητα εντοπισμού τομέων αυξημένου ρίσκου. Για τον λόγο αυτό, οι

διεθνείς πρωτοβουλίες συνεργασίας μεταξύ δημόσιου και ιδιωτικού τομέα καλούνται να επικεντρωθούν στην ανάπτυξη ενός κοινά αποδεκτού λεξιλογίου για την τεχνητή νοημοσύνη, προωθώντας παράλληλα τη σταδιακή εναρμόνιση των εννοιολογικών και κανονιστικών της προσεγγίσεων (Crisanto κ.ά., 2024).

2.4 Ταξινόμηση της τεχνητής νοημοσύνης

Η τεχνητή νοημοσύνη αποτελεί κομμάτι της επιστήμης των υπολογιστών και έχει ως στόχο την κατασκευή ευφυών μηχανών που είναι ικανές να εκτελούν δραστηριότητες που συνήθως απαιτούν ανθρώπινη νοημοσύνη, όπως η οπτική αντίληψη, η αναγνώριση ομιλίας, η λήψη αποφάσεων και η επεξεργασία φυσικής γλώσσας (M. Kaur, 2024; Ridzuan κ.ά., 2024). Οι εφαρμογές και οι συσκευές που είναι εξοπλισμένες με αυτή μπορούν να ενεργούν ανεξάρτητα, αντικαθιστώντας την ανάγκη για ανθρώπινη παρέμβαση (Stryker & Kavlakoglu, 2024). Τα συστήματα τεχνητής νοημοσύνης μαθαίνουν και βελτιώνονται μέσω της έκθεσής τους σε τεράστιες ποσότητες δεδομένων, εντοπίζοντας μοτίβα και σχέσεις που οι άνθρωποι ενδέχεται να παραβλέπουν. Αυτά τα δεδομένα χρησιμεύουν ως εκπαιδευτικό υλικό, με την ποιότητα και την ποσότητά τους να είναι κρίσιμη για την απόδοση της τεχνητής νοημοσύνης (Google Cloud, χ.χ.).

Η τεχνητή νοημοσύνη ταξινομείται με πολλαπλούς τρόπους, καθώς διαφορετικοί μελετητές, ερευνητές και οργανισμοί υιοθετούν διαφορετικά κριτήρια. Εντούτοις, η ομαδοποίηση βάσει του επιπέδου ευφυΐας και του εύρους δυνατοτήτων συνιστά μία γενικά αποδεκτή προσέγγιση και χρησιμοποιείται συχνά ως αφετηρία για την κατανόηση των επιμέρους τύπων συστημάτων τεχνητής νοημοσύνης (Google Cloud, χ.χ.; Rashid & Kausik, 2024).

2.4.1 Ταξινόμηση βάσει του επιπέδου ευφυΐας

Η ταξινόμηση βάσει του επιπέδου ευφυΐας διακρίνει την τεχνητή νοημοσύνη στη:

- στενή τεχνητή νοημοσύνη (Artificial Narrow Intelligence - ANI),
- γενική τεχνητή νοημοσύνη (Artificial General Intelligence - AGI) και
- υπερνοημοσύνη ή υπερευφυΐα (Artificial Super Intelligence - ASI).

Η στενή τεχνητή νοημοσύνη, ή και αδύναμη (weak), αποτελεί τον κυρίαρχο τύπο τεχνητής νοημοσύνης που συναντάται σήμερα. Είναι εκπαιδευμένη στην εκτέλεση συγκεκριμένων εργασιών, ταχύτερα και αποτελεσματικότερα από τον άνθρωπο, ωστόσο αδυνατεί να εκτελέσει εργασίες εκτός του προκαθορισμένου πλαισίου λειτουργίας της. Ενδεικτικά

παραδείγματα αποτελούν οι ψηφιακοί βοηθοί Siri και Alexa της Apple και Amazon αντίστοιχα (IBM Data and AI Team, 2025; A. Kaur, 2025).

Η γενική τεχνητή νοημοσύνη, ή και ισχυρή (strong), αποτελεί προς ώρας μια θεωρητική έννοια, αφού παρά τις σημαντικές ερευνητικές προσπάθειες, εξακολουθούν να υπάρχουν άλυτα τεχνικά και ηθικά ζητήματα. Δυνητικά είναι ικανή να αξιοποιεί προηγούμενες γνώσεις και δεξιότητες για την εκτέλεση νέων εργασιών σε διαφορετικά πλαίσια, χωρίς την ανάγκη ανθρώπινης εκπαίδευσης των υποκείμενων μοντέλων. Αυτή η ικανότητα, της επιτρέπει την αναπαραγωγή των γνωστικών λειτουργιών του ανθρώπου σε ένα ευρύ φάσμα δραστηριοτήτων (IBM Data and AI Team, 2025; A. Kaur, 2025).

Τέλος, η υπερνοημοσύνη ή υπερευφυΐα συνιστά ένα θεωρητικό πλαίσιο στο οποίο τα συστήματα τεχνητής νοημοσύνης ξεπερνούν την ανθρώπινη νοημοσύνη σε όλους τους τομείς. Αυτό το είδος τεχνητής νοημοσύνης θα διαθέτει δυνατότητες που ξεπερνούν αυτές των ανθρώπων, καθώς τα συστήματα αυτά θα έχουν εξελιχθεί πέρα από το σημείο κατανόησης των ανθρώπινων συναισθημάτων και εμπειριών, ώστε να αισθάνονται συναισθήματα, να έχουν ανάγκες και να διαθέτουν δικές τους πεποιθήσεις και επιθυμίες. Ο τομέας αυτός εγείρει βαθιά φιλοσοφικά και ηθικά ζητήματα, ιδίως όσον αφορά τον έλεγχο και την αρμονία με τα ανθρώπινα ιδανικά (IBM Data and AI Team, 2025; A. Kaur, 2025).

2.4.2 Ταξινόμηση βάσει του εύρους δυνατοτήτων

Η ταξινόμηση βάσει του εύρους δυνατοτήτων διακρίνει την τεχνητή νοημοσύνη σε τέσσερις βασικές κατηγορίες:

- τα ρομπότ ή αντιδραστικές μηχανές (robots or reactive machines),
- τα συστήματα περιορισμένης μνήμης (limited memory),
- τα συστήματα θεωρίας του νου (theory of mind) και
- τα συστήματα αυτογνωσίας (self-awareness).

Οι αντιδραστικές μηχανές είναι συστήματα τεχνητής νοημοσύνης σχεδιασμένα για την εκτέλεση αυστηρά καθορισμένων εργασιών. Δεδομένου ότι δεν μπορούν να ανακαλέσουν προηγούμενα αποτελέσματα ή αποφάσεις, λειτουργούν μόνο με τα διαθέσιμα δεδομένα. Η συγκεκριμένη μορφή νοημοσύνης εδράζεται στη μαθηματική και στατιστική ανάλυση, επιτρέποντας την επεξεργασία τεράστιων όγκων δεδομένων με σκοπό την παραγωγή

φαινομενικά ευφών αποτελεσμάτων. Ενδεικτικό παράδειγμα τέτοιου είδους νοημοσύνης αποτελεί ο υπερυπολογιστής Deep Blue (IBM Data and AI Team, 2025).

Τα συστήματα περιορισμένης μνήμης διαθέτουν τη δυνατότητα αξιοποίησης ιστορικών εμπειριών για τη λήψη μελλοντικών αποφάσεων. Ωστόσο, δεν είναι σε θέση να αποθηκεύουν τα δεδομένα προηγούμενων εμπειριών με σκοπό τη μακροπρόθεσμη χρήση. Η τεχνητή νοημοσύνη περιορισμένης μνήμης δύναται να βελτιώσει την απόδοσή της μέσω της εκπαίδευσης των υποκείμενων μοντέλων με περισσότερα δεδομένα στην πάροδο του χρόνου. Χαρακτηριστικά παραδείγματα αυτής της κατηγορίας αποτελούν τα αυτόνομα οχήματα και τα chatbot, τα οποία χρησιμοποιούν βραχυπρόθεσμη μνήμη για να βελτιώσουν τη λειτουργικότητά τους (IBM Data and AI Team, 2025; A. Kaur, 2025).

Τα συστήματα θεωρίας του νου αναφέρονται σε μια μορφή τεχνητής νοημοσύνης που επί του παρόντος βρίσκεται σε φάση έρευνας. Στόχος τους είναι η κατανόηση των ανθρώπινων συναισθημάτων, πεποιθήσεων και προθέσεων, προκειμένου να διευκολυνθεί η αλληλεπίδραση ανθρώπου-μηχανής και να ενισχυθεί ο βαθμός εξατομίκευσης (IBM Data and AI Team, 2025).

Τέλος, τα συστήματα αυτογνωσίας αντιπροσωπεύουν το κορυφαίο ορόσημο προόδου της τεχνητής νοημοσύνης, καθώς αναμένεται να επιδεικνύουν στοιχεία συνείδησης και ταυτότητας. Αν και το συγκεκριμένο επίπεδο τεχνητής νοημοσύνης παραμένει θεωρητικό προκαλεί σημαντικά ηθικά και φιλοσοφικά διλήμματα σχετικά με τα δικαιώματα και την αυτονομία των μηχανών (A. Kaur, 2025).

2.5 Τεχνολογίες της τεχνητής νοημοσύνης

Η τεχνητή νοημοσύνη δεν είναι μια μεμονωμένη τεχνολογία, αλλά ένας όρος που περιλαμβάνει πολλές διαφορετικές επιστημονικές ειδικότητες, όπως την επιστήμη των υπολογιστών, την ανάλυση δεδομένων, τη στατιστική, τη μηχανική υλικού και λογισμικού, τη γλωσσολογία ακόμη και τη φιλοσοφία και την ψυχολογία (Google Cloud, χ.χ.). Το ερώτημα σχετικά με το ποια επιμέρους πεδία ανήκουν στην τεχνητή νοημοσύνη αποτελεί αντικείμενο συνεχούς συζήτησης, ενώ τα όρια μεταξύ αυτών των πεδίων είναι συχνά μη ορατά (Condoleezza Rice κ.ά., 2025). Στα επόμενα υποκεφάλαια θα παρουσιαστούν συνοπτικά τα βασικά τεχνολογικά πεδία της τεχνητής νοημοσύνης που χρησιμοποιούνται στον χρηματοοικονομικό τομέα (Marple κ.ά., 2023).

2.5.1 Μηχανική μάθηση

Η μηχανική μάθηση αποτελεί ένα θεμελιώδες τμήμα της τεχνητής νοημοσύνης και προϋποθέτει την πρόσβαση σε μεγάλο όγκο δεδομένων και μεγάλη υπολογιστική ισχύ (Condoleezza Rice κ.ά., 2025). Περιλαμβάνει αλγορίθμους που έχουν τη δυνατότητα εντοπισμού μοτίβων εντός των δεδομένων εκπαίδευσης και τη συνεπακόλουθη δημιουργία συμπερασμάτων και υποθέσεων σε νέα δεδομένα. Αυτή η ικανότητα αναγνώρισης μοτίβων επιτρέπει στα μοντέλα μηχανικής μάθησης να λαμβάνουν αποφάσεις ή να κάνουν προβλέψεις χωρίς ρητές, προγραμματισμένες οδηγίες (Bergmann, 2025; Google Cloud, χ.χ.; Stryker & Kavlakoglu, 2024).

Η μηχανική μάθηση μπορεί να ταξινομηθεί σε τρεις βασικές κατηγορίες βάσει του τρόπου εκπαίδευσης των μοντέλων (Bergmann, 2025). Συγκεκριμένα:

➤ Εποπτευόμενη μηχανική μάθηση (Supervised machine learning)

Η εποπτευόμενη μηχανική μάθηση αναφέρεται στη διαδικασία εκπαίδευσης ενός μοντέλου μέσω της παροχής δεδομένων, τα οποία έχουν επισημανθεί με τις κατάλληλες ετικέτες, ώστε αργότερα να είναι σε θέση να προσδιορίσει πρότυπα και σχέσεις σε νέα άγνωστα δεδομένα. Τα δεδομένα αυτά παρέχουν το πλαίσιο ή την «βασική αλήθεια» για το μοντέλο, διδάσκοντας το ρητά στο πώς να αναγνωρίζει τις σχέσεις μεταξύ των χαρακτηριστικών και των ετικετών. Όσο περισσότερα δεδομένα εισόδου τροφοδοτούνται στο μοντέλο, τόσο αυτό προσαρμόζει τις παραμέτρους του έως ότου προσαρμοστεί πλήρως. Για παράδειγμα, ένα μοντέλο μπορεί να εκπαιδευτεί σε χιλιάδες εικόνες σκύλων, έτσι ώστε αργότερα να μπορεί να αναγνωρίσει σκύλους σε νέες εικόνες τις οποίες δεν έχει ξανά επεξεργαστεί (Bergmann, 2025; Google Cloud, 2025b; Maple κ.ά., 2023).

Πολλά μοντέλα εποπτευόμενης μηχανικής μάθησης βασίζονται σε τεχνικές όπως η γραμμική και η λογιστική παλινδρόμηση, ενώ άλλες ευρέως χρησιμοποιούμενες προσεγγίσεις, όπως τα δέντρα αποφάσεων, τα τυχαία δάση και τα νευρωνικά δίκτυα, εφαρμόζονται σε ένα μεγάλο φάσμα πεδίων, συμπεριλαμβανομένων των χρηματοοικονομικών προβλέψεων, της διαχείρισης χαρτοφυλακίων, της αξιολόγησης πιστοληπτικού κινδύνου και της ανίχνευσης απάτης (Bergmann, 2025; Google Cloud, 2025b; Maple κ.ά., 2023). Τα νευρωνικά δίκτυα, ειδικότερα, είναι εμπνευσμένα από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου και αποτελούνται από διασυνδεδεμένα επίπεδα κόμβων, τα οποία συνεργάζονται για την επεξεργασία πολύπλοκων και ογκωδών συνόλων δεδομένων, επιτυγχάνοντας την

αναγνώριση σύνθετων προτύπων και συσχετίσεων (Google Cloud, χ.χ.; Stryker & Kavlakoglu, 2024).

Η εποπτευόμενη μηχανική μάθηση προσφέρει σημαντικά πλεονεκτήματα, ωστόσο παρουσιάζει και ορισμένους περιορισμούς. Η ανάπτυξη και εφαρμογή των μοντέλων αυτών απαιτεί εξειδικευμένο προσωπικό, αφού η ανθρώπινη εποπτεία για την εκπαίδευση και την επικύρωση των αποτελεσμάτων τους είναι αναγκαία. Η ανάγκη ύπαρξης εξειδικευμένου προσωπικού ταυτόχρονα με την ανάγκη επισήμανσης των δεδομένων εκπαίδευσης χειροκίνητα, κάνει την ανάπτυξη αυτών των μοντέλων ακριβή και χρονοβόρα. Τα μοντέλα εποπτευόμενης μάθησης παρουσιάζουν περιορισμένη ευελιξία απέναντι σε νέα ή μη δομημένα δεδομένα και ενδέχεται να υπόκεινται σε προκαταλήψεις λόγω λανθασμένης σήμανσης ή ιστορικών ανισοτήτων στα σύνολα δεδομένων που χρησιμοποιήθηκαν στο στάδιο της εκπαίδευσης. Τέλος, η υπερπροσαρμογή των μοντέλων αποτελεί συχνό ζήτημα, καθώς η υπερβολική εξειδίκευση στα δεδομένα εκπαίδευσης μπορεί να μειώσει την ικανότητα γενίκευσης του μοντέλου (Belcic & Stryker, 2025).

➤ Μη εποπτευόμενη μηχανική μάθηση (Unsupervised machine learning)

Η μη εποπτευόμενη μηχανική μάθηση αναφέρεται στη διαδικασία εκπαίδευσης ενός μοντέλου που μπορεί να διακρίνει μοτίβα, εξαρτήσεις και συσχετίσεις στα δεδομένα, χωρίς να έχει εκπαιδευτεί σε κάποιο συγκεκριμένο πρότυπο (πχ. αναγνώριση σκύλων). Το μοντέλο μαθαίνει από τα δεδομένα χωρίς ανθρώπινη παρέμβαση, εξ' ου και το μη εποπτευόμενη, και τα κατηγοριοποιεί σε ομάδες με βάση τα χαρακτηριστικά τους. Για την επίτευξη αυτού του σκοπού μπορεί να χρησιμοποιηθούν τεχνικές, όπως η ομαδοποίηση (clustering), η μείωση των διαστάσεων των δεδομένων ή η ανίχνευση ανωμαλιών (Bergmann, 2025; Google Cloud, 2025b; Maple κ.ά., 2023).

Η μη εποπτευόμενη μάθηση εγκυμονεί προκλήσεις ως προς την ερμηνεία των αποτελεσμάτων της, σε σύγκριση με την εποπτευόμενη, διότι τα πρότυπα και οι συσχετίσεις που αναδύονται συχνά παραμένουν ασαφή. Η απουσία προκαθορισμένων προτύπων ή κλάσεων αναφοράς καθιστά την αξιολόγηση και επικύρωση των ευρημάτων ιδιαίτερως πολύπλοκη διαδικασία, καθώς στερείται ενός προτύπου σύγκρισης. Τα μοντέλα αυτά συνήθως απαιτούν περισσότερους υπολογιστικούς πόρους, λόγω της ανάγκης επεξεργασίας εκτενέστερων συνόλων δεδομένων και της υιοθέτησης πιο σύνθετων αλγορίθμων, ενώ η έλλειψη δομικών περιορισμών την καθιστά πιο ευάλωτη σε φαινόμενα υπερπροσαρμογής,

όπου το μοντέλο προσαρμόζεται υπερβολικά σε θόρυβο ή τυχαίες ιδιαιτερότητες των δεδομένων (Marple κ.ά., 2023).

Στον τραπεζικό τομέα η μη εποπτευόμενη μάθηση βρίσκει ένα ευρύ φάσμα εφαρμογής σε πεδία, όπως η ανίχνευση απάτης, η τμηματοποίηση πελατών, η βελτιστοποίηση χαρτοφυλακίου, η ανάλυση πιστωτικού κινδύνου και η ανάλυση αγορών (Marple κ.ά., 2023).

➤ Ενισχυτική μάθηση (Reinforcement learning)

Η ενισχυτική μάθηση αναφέρεται στη διαδικασία ανάπτυξης και εκπαίδευσης ενός μοντέλου το οποίο μαθαίνει στην πράξη να εκτελεί μια καθορισμένη εργασία μέσω ενός βρόγχου ανατροφοδότησης έως ότου η απόδοσή του βρίσκεται εντός του καθορισμένου επιθυμητού εύρους. Ο στόχος της ενισχυτικής μάθησης έγκειται στη συστηματική βελτιστοποίηση των παραμέτρων του μοντέλου, με απώτερο σκοπό τη μεγιστοποίηση της μακροπρόθεσμης αθροιστικής αξίας των ενεργειών (Bergmann, 2025; Google Cloud, 2025b; Marple κ.ά., 2023). Η ενισχυτική μάθηση είναι λιγότερο διαδεδομένη σε σύγκριση με τις άλλες προσεγγίσεις, ωστόσο, η μοναδικότητά της την καθιστά αναντικατάστατη. Στον τραπεζικό τομέα μπορεί να χρησιμοποιηθεί για την εκτέλεση χρηματιστηριακών συναλλαγών και τη δυναμική τιμολόγηση (Marple κ.ά., 2023).

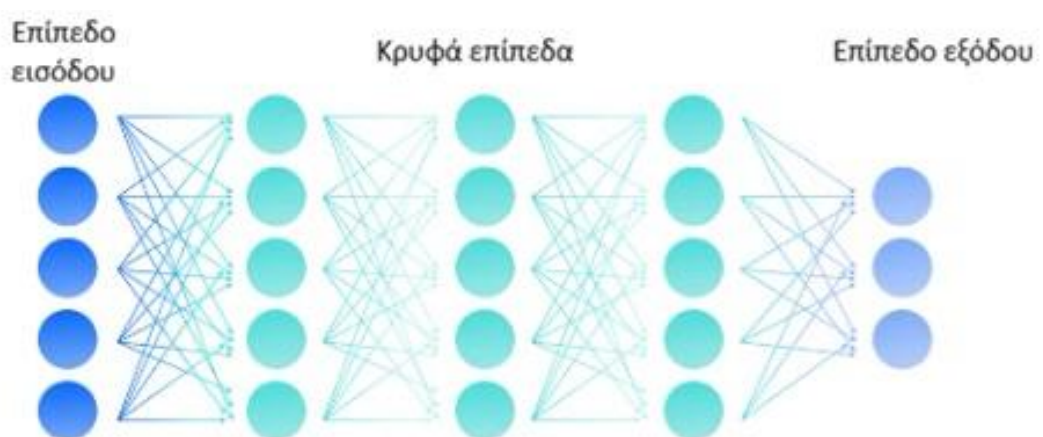
2.5.2 Βαθιά μάθηση

Οι όροι της τεχνητής νοημοσύνης, της μηχανικής μάθησης και της βαθιάς μάθησης (deep learning) συχνά χρησιμοποιούνται συνώνυμα, παρόλο που διαφέρουν εννοιολογικά. Ένας ενδεικτικός τρόπος για την καλύτερη κατανόηση των όρων είναι ο παραλληλισμός τους με μια κούκλα ματριόσκα, όπου κάθε επίπεδο εμπεριέχει το επόμενο. Η τεχνητή νοημοσύνη αντιπροσωπεύει την αρχική κούκλα, εντός της οποίας βρίσκεται η μηχανική μάθηση. Στο εσωτερικό της, βρίσκεται η βαθιά μάθηση, ενώ εντός της εντάσσεται η παραγωγική τεχνητή νοημοσύνη (Generative AI – Gen AI) (Hirz κ.ά., 2024; Stryker & Kavlakoglu, 2024). Μια απλούστερη αναπαράσταση παρατίθεται στην εικόνα 1 (Stryker & Kavlakoglu, 2024):



Εικόνα 1 Τεχνητή νοημοσύνης, Μηχανική και Βαθιά μάθηση

Η βαθιά μάθηση χρησιμοποιεί πολλαπλά νευρωνικά δίκτυα τα οποία προσομοιώνουν με μεγαλύτερη ακρίβεια την πολύπλοκη διαδικασία λήψης αποφάσεων του ανθρώπινου εγκεφάλου. Σε αυτή την προσέγγιση τα νευρωνικά δίκτυα περιλαμβάνουν ένα επίπεδο εισόδου, συνήθως εκατοντάδες κρυφά ενδιάμεσα επίπεδα και ένα επίπεδο εξόδου, σε αντίθεση με τα νευρωνικά δίκτυα της μηχανικής μάθησης, τα οποία συνήθως έχουν μόνο ένα ή δύο κρυφά ενδιάμεσα επίπεδα (Stryker & Kavliakoglu, 2024). Αναπαράσταση ενός απλού νευρωνικού δικτύου προώθησης με 3 κρυφά επίπεδα παρατίθεται στην εικόνα 3 (IBM, 2025).



Εικόνα 2. Αναπαράσταση απλοποιημένου νευρωνικού δικτύου προώθησης

Κάθε επίπεδο ενός νευρωνικού δικτύου περιλαμβάνει πολλαπλούς κόμβους, οι οποίοι είναι ικανοί να επεξεργάζονται εισερχόμενες πληροφορίες. Το αποτέλεσμα της επεξεργασίας κάθε κόμβου μεταβιβάζεται ως είσοδος στους κόμβους του επόμενου επιπέδου, μέσω μιας διαδοχικής διαδικασίας που οδηγεί στο τελικό επίπεδο, όπου προκύπτει το συνολικό αποτέλεσμα του δικτύου. Οι συνδέσεις μεταξύ των κόμβων χαρακτηρίζονται από διαφοροποιημένη βαρύτητα, η οποία αποτελεί μια παράμετρο που βελτιστοποιείται κατά τη φάση εκπαίδευσης του μοντέλου (Google Cloud, 2025a; IBM, 2025).

Στο πλαίσιο της βαθιάς μάθησης, οι συνδέσεις των κόμβων είναι μη γραμμικές, επιτρέποντας στα νευρωνικά δίκτυα να μοντελοποιούν σύνθετα μοτίβα στα δεδομένα. Αυτό το χαρακτηριστικό αποτελεί τη σημαντικότερη διαφοροποίηση μεταξύ των μοντέλων βαθιάς μάθησης και των πολύπλοκων γραμμικών μοντέλων παλινδρόμησης. Τέλος, ο αριθμός των επιπέδων, ο αριθμός των κόμβων κάθε επιπέδου και οι παράμετροι ενεργοποίησης των κόμβων κάθε επιπέδου αποτελούν «υπερπαραμέτρους» που πρέπει να καθοριστούν συγκεκριμένα για κάθε μοντέλο πριν από την εκπαίδευση του (IBM, 2025).

2.5.3 Παραγωγική τεχνητή νοημοσύνη

Η παραγωγική τεχνητή νοημοσύνη χρησιμοποιεί εξελιγμένα μοντέλα βαθιάς μάθησης τόσο για να αναλύει όσο και για να δημιουργεί νέο περιεχόμενο. Έχει την ικανότητα να μαθαίνει μοτίβα και πρότυπα μέσα από τεράστιες ποσότητες δεδομένων, ενώ στη συνέχεια τα χρησιμοποιεί για να παράγει εντελώς νέο, πρωτότυπο περιεχόμενο με βάση τις αντίστοιχες εισαγωγές (Google Cloud, χ.χ.; Hirz κ.ά., 2024). Στην πλειονότητα των περιπτώσεων η παραγωγική τεχνητή νοημοσύνη λειτουργεί σε τρεις διακριτές φάσεις: τη φάση της εκπαίδευσης, τη φάση της ρύθμισης και τη φάση της παραγωγής (Stryker & Scapicchio, 2025).

Κατά την αρχική φάση εκπαίδευσης, δημιουργείται ένα θεμελιώδες μοντέλο, το οποίο λειτουργεί ως βάση για πολλαπλές εφαρμογές παραγωγικής τεχνητής νοημοσύνης. Τα πιο διαδεδομένα πρότυπα περιλαμβάνουν τα μεγάλα γλωσσικά μοντέλα (Large Language Models - LLMs) για παραγωγή κειμένου, καθώς και εξειδικευμένα μοντέλα για εικόνες, βίντεο, ήχο και άλλες μορφές περιεχομένου. Η υλοποίηση αυτών των μοντέλων βασίζεται στη χρήση νευρωνικών δικτύων βαθιάς μάθησης, εκπαιδευόμενα σε τεράστιους όγκους ακατέργαστων, μη δομημένων και μη επισημασμένων δεδομένων, μια διαδικασία που απαιτεί εντατική υπολογιστική ισχύ, καθιστώντας την ιδιαίτερα χρονοβόρα και δαπανηρή (Stryker & Scapicchio, 2025).

Μετά την αρχική εκπαίδευση του θεμελιώδους μοντέλου, απαιτείται η προσαρμογή του για μια συγκεκριμένη εφαρμογή, διαδικασία που μπορεί να πραγματοποιηθεί με διάφορες τεχνικές. Μία προσέγγιση αποτελεί η λεπτομερής προ εκπαίδευση, κατά την οποία το μοντέλο τροφοδοτείται με δεδομένα, που έχουν επισημανθεί με τις κατάλληλες ετικέτες, ερωτήματα ή προτροπές που αναμένεται να δεχθεί στην πράξη, μαζί με τις αντίστοιχες ορθές απαντήσεις. Μία ακόμη προσέγγιση είναι η ενισχυτική μάθηση με ανθρώπινη ανατροφοδότηση όπου οι χρήστες αξιολογούν την ακρίβεια ή τη συνάφεια των παραγόμενων εξόδων, διορθώνοντας ή σχολιάζοντας τα αποτελέσματα, ώστε το μοντέλο να βελτιώνει σταδιακά την απόδοσή του (Stryker & Scarpicchio, 2025).

Τέλος, κατά τη φάση της παραγωγής, αξιολόγησης και επαναρύθμισης, οι προγραμματιστές και οι χρήστες αξιολογούν τακτικά τα παραγόμενα αποτελέσματα και διενεργούν περαιτέρω προσαρμογές στο μοντέλο, ακόμα και σε εβδομαδιαία βάση, με στόχο την ενίσχυση της ακρίβειας και της συνάφειας. Είναι σημαντικό να αναφερθεί ότι το αρχικό θεμελιώδες μοντέλο ενημερώνεται λιγότερο συχνά, μια φορά ετησίως ή ανά δεκαοκτώ μήνες (Stryker & Scarpicchio, 2025).

2.5.4 Επεξεργασία φυσικής γλώσσας

Η επεξεργασία φυσικής γλώσσας (Natural Language Processing – NLP) αποτελεί έναν σημαντικό κλάδο της τεχνητής νοημοσύνης, ο οποίος επιτρέπει στα υπολογιστικά συστήματα να κατανοούν, να ερμηνεύουν και να παράγουν φυσική γλώσσα, τόσο σε προφορική όσο και σε γραπτή μορφή (Condoleezza Rice κ.ά., 2025). Τεχνολογίες αυτής της κατηγορίας αξιοποιούνται ήδη σε πληθώρα εφαρμογών της καθημερινής ζωής, όπως στις μηχανές αναζήτησης, στα chatbot εξυπηρέτησης πελατών μέσω φωνητικών εντολών, στα συστήματα πλοήγησης GPS, καθώς και σε ψηφιακούς βοηθούς όπως η Alexa και η Siri, οι οποίοι ανταποκρίνονται σε λεκτικά αιτήματα και ερωτήσεις χρηστών (Stryker & Holdsworth, 2025).

Η επεξεργασία φυσικής γλώσσας συνδυάζει τις μεθόδους της υπολογιστικής γλωσσολογίας, οι οποίες αξιοποιούν τα εργαλεία της επιστήμης των δεδομένων για τη μελέτη και ανάλυση της γλώσσας και του λόγου, με τις τεχνικές της μηχανικής και βαθιάς μάθησης. Στο πλαίσιο αυτό, διακρίνονται δύο βασικοί τύποι ανάλυσης της γλώσσας: η συντακτική ανάλυση και η σημασιολογική ανάλυση. Η πρώτη ασχολείται με τη δομή και τη σύνταξη των λέξεων, των προτάσεων και φράσεων, μέσω της εφαρμογής προκαθορισμένων γραμματικών κανόνων για τον προσδιορισμό του νοήματος. Η σημασιολογική ανάλυση αξιοποιεί τα αποτελέσματα της

συντακτικής προκειμένου να αποδώσει και να ερμηνεύσει το περιεχόμενο των λέξεων στο πλαίσιο της συνολικής δομής της πρότασης (Stryker & Holdsworth, 2025).

Ορισμένοι από τους πιο ευρέως χρησιμοποιούμενους αλγορίθμους επεξεργασίας φυσικής γλώσσας (NLP) περιλαμβάνουν τις ακόλουθες λειτουργίες (Stryker & Holdsworth, 2025):

- Προ επεξεργασία κειμένου (Text Preprocessing): Περιλαμβάνει τον καθαρισμό και τη μετατροπή των δεδομένων κειμένου σε μορφή κατάλληλη για ανάλυση. Η διαδικασία αυτή συμβάλλει στη βελτίωση της ποιότητας των δεδομένων και στη μείωση του όγκου τους.
- Ανάλυση συναισθήματος (Sentiment Analysis): Χρησιμοποιείται για τον εντοπισμό και την αξιολόγηση του συναισθηματικού υποβάθρου ενός κειμένου.
- Αναγνώριση οντοτήτων (Named Entity Recognition – NER): Χρησιμοποιείται για τον εντοπισμό και την ταξινόμηση οντοτήτων, όπως προσώπων, οργανισμών ή τοποθεσιών, από μη δομημένα δεδομένα κειμένου, με βάση προκαθορισμένες κατηγορίες.
- Μοντελοποίηση θεμάτων (Topic Modeling): Αναφέρεται στη διαδικασία εντοπισμού και ομαδοποίησης θεματικών ενοτήτων μέσα από μεγάλο όγκο κειμένων, με σκοπό την κατανόηση των βασικών θεμάτων που αναδύονται στα δεδομένα.
- Μηχανική μετάφραση (Machine Translation): Εστιάζει στη μετάφραση κειμένου από μία γλώσσα σε μία άλλη. Παρά τη σημαντική πρόοδο στον τομέα αυτό, υφίστανται προκλήσεις σχετιζόμενες με τη διατήρηση της εννοιολογικής και γραμματικής ακρίβειας.
- Αναγνώριση ομιλίας (Speech Recognition): Συνίσταται στη μετατροπή προφορικού λόγου σε γραπτό κείμενο.

Η επεξεργασία φυσικής γλώσσας συμβάλλει καθοριστικά στην αυτοματοποίηση και βελτιστοποίηση των επιχειρηματικών διαδικασιών, προσφέροντας λύσεις που ενισχύουν την παραγωγικότητα και βρίσκουν εφαρμογή σε τομείς όπως οι χρηματοοικονομικές υπηρεσίες μέσω εργαλείων ανάλυσης δεδομένων και εξυπηρέτησης πελατών (Marple κ.ά., 2023).

3. Κανονιστικό και ηθικό πλαίσιο για την τεχνητή νοημοσύνη

Οι απόψεις σχετικά με τη ρύθμιση της τεχνητής νοημοσύνης διαφέρουν σημαντικά, καθώς οι προσπάθειες αντιμετώπισης ηθικών ζητημάτων, νομοθετικών κενών και θεμάτων ιδιωτικότητας έχουν φέρει στο προσκήνιο διαφορετικές προσεγγίσεις ανά την υφήλιο. Ένα εκτενές σύνολο κατευθυντήριων οδηγιών, πλαισίων και αρχών, το οποίο συμπληρώνεται διαρκώς σε εθνικό και περιφερειακό επίπεδο, έχει αναπτυχθεί και υιοθετηθεί από κυβερνήσεις, επιχειρήσεις, κοινοπραξίες, περιφερειακούς και διεθνείς οργανισμούς και ποικίλα φόρουμ που συγκεντρώνουν πολυάριθμους ενδιαφερόμενους φορείς (UN. Advisory Body on Artificial Intelligence, 2024; Vuković κ.ά., 2025).

3.1 Κανονιστικό πλαίσιο για την τεχνητή νοημοσύνη

Οι προσεγγίσεις δικαιοδοτικής πολιτικής για τη διαχείριση της τεχνητής νοημοσύνης διακρίνονται κυρίως σε δύο κατηγορίες: αυτές που βασίζονται σε μη δεσμευτικές αρχές και αυτές που βασίζονται σε κανόνες (Crisanto κ.ά., 2024; KPMG, 2024). Τα δικαιοδοτικά όργανα που επιλέγουν την πρώτη προσέγγιση βασίζονται σε μεγάλο βαθμό σε μη δεσμευτικές αρχές ή/και εθελοντικές δεσμεύσεις που υποστηρίζονται από γενικά τεχνικά πρότυπα ή/και διατομεακούς κανονισμούς. Η προσέγγιση αυτή αναγνωρίζει τους κινδύνους και τις προκλήσεις που εμπεριέχει η τεχνητή νοημοσύνη, θεωρώντας όμως πρόωγη την αυστηρή ρύθμιση της. Αντίθετα, τα δικαιοδοτικά όργανα που επιλέγουν μια προσέγγιση βασισμένη σε κανόνες έχουν εκδώσει ή βρίσκονται στη διαδικασία θέσπισης νομοθεσίας για την τεχνητή νοημοσύνη. Αυτή η προσέγγιση στοχεύει στην παροχή κανονιστικής σαφήνειας, προωθώντας την ασφαλή ανάπτυξη της τεχνολογίας, και στην κατοχύρωση νομικών εξουσιών για την επιβολή κυρώσεων σε περιπτώσεις παράνομης χρήσης της τεχνητής νοημοσύνης (Crisanto κ.ά., 2024).

Παράδειγμα της πρώτης προσέγγισης αποτέλεσαν αρχικά οι ΗΠΑ με την εκτελεστική εντολή 14110¹ (Executive Order 14110, 2023), η οποία δημιούργησε ένα πλαίσιο για την τεχνητή νοημοσύνη σε ομοσπονδιακό επίπεδο. Με την αλλαγή κυβέρνησης τον Ιανουάριο του 2025,

¹ Executive Order 14110 on Safe, Secure, and Trustworthy Artificial Intelligence, 2023

εκδόθηκε η εκτελεστική εντολή 14119² (Executive Order 14119, 2025) η οποία ανακάλεσε πλήρως την 14110. Η νέα κυβέρνηση των ΗΠΑ υιοθέτησε μια απορρυθμιστική νομοθετική πολιτική δίνοντας προτεραιότητα στην καινοτομία και την παγκόσμια τεχνολογική ηγεμονία για την χώρα, με λιγότερους υποχρεωτικούς κανόνες, βασιζόμενη σε μεγάλο βαθμό στους υφιστάμενους ομοσπονδιακούς νόμους που συμπληρώνονται από νομοθετικά πλαίσια σε πολιτειακό επίπεδο. Η αμερικανική προσέγγιση είναι πιο ευέλικτη αλλά κατακερματισμένη, με λιγότερη έμφαση στις προληπτικές απαγορεύσεις ή συγκεκριμένες ταξινομήσεις για τα συστήματα τεχνητής νοημοσύνης (Schettenhelm, 2025).

Αντίθετα, η Ευρωπαϊκή Ένωση επέλεξε την δημιουργία ενός ολοκληρωμένου πλαισίου κανόνων. Ο κανονισμός 2024/1689³ (Regulation (EU) 2024/1689, 2024), γνωστός και ως EU AI Act, αποτελεί το πρώτο στοχευμένο νομοθετικό πλαίσιο για την τεχνητή νοημοσύνη. Ο κανονισμός ταξινομεί τα συστήματα τεχνητής νοημοσύνης βάσει του κινδύνου που ενέχουν σε: απαγορευμένα, υψηλού κινδύνου, περιορισμένου κινδύνου και ελάχιστου κινδύνου. Απαγορευμένα θεωρούνται τα συστήματα που χρησιμοποιούνται για κοινωνική βαθμολόγηση, χειραγώγηση της κοινής γνώμης, ζωντανή ταυτοποίηση ανθρώπων κ.ά., με στόχο την προστασία των θεμελιωδών δικαιωμάτων των ανθρώπων και των κοινωνικών αξιών. Ο κανονισμός εστιάζει κυρίως στα συστήματα υψηλού κινδύνου, τα οποία υπόκεινται σε εκτενείς ρυθμιστικές απαιτήσεις εξαιτίας του σημαντικού αντίκτυπου στην ασφάλεια και στα ατομικά δικαιώματα. Τα συστήματα περιορισμένου κινδύνου υπόκεινται σε πιο χαλαρές κανονιστικές υποχρεώσεις, υποχρεώνοντας τους δημιουργούς και διαχειριστές τους να ενημερώνουν τους τελικούς χρήστες για την χρήση τεχνητής νοημοσύνης, όπως στην περίπτωση συνομιλίας με chatbot. Τέλος, οι εφαρμογές τεχνητής νοημοσύνης που χαρακτηρίζονται από ελάχιστο κίνδυνο, που αποτελούν την πλειονότητα των τεχνολογιών στην ενιαία αγορά της Ευρωπαϊκής Ένωσης, όπως τα βιντεοπαιχνίδια, παραμένουν κατά κανόνα χωρίς ρυθμιστικό καθεστώς, αν και η εικόνα αυτή μπορεί να μεταβληθεί βάσει των τεχνολογικών εξελίξεων (European Parliament, 2025; Regulation (EU) 2024/1689, 2024).

Η διαθέσιμη βιβλιογραφία φανερώνει ότι το δίπολο «ρύθμιση βάσει αρχών» ή «ρύθμιση βάσει κανόνων» καθίσταται ολοένα και πιο ανεπαρκές για την αντιμετώπιση της πολυπλοκότητας και την αποτελεσματική εποπτεία της τεχνητής νοημοσύνης. Καμία από τις

² Executive Order 14119 - Removing Barriers to American Leadership in Artificial Intelligence, 2025

³ Regulation (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 13 June 2024

δύο προσεγγίσεις, ανεξάρτητα, δεν ανταποκρίνεται επαρκώς στις σύγχρονες απαιτήσεις διακυβέρνησης, υποδηλώνοντας την ανάγκη για υβριδικές ρυθμιστικές προσεγγίσεις που θα μπορούν να προσαρμόζονται στις τεχνολογικές αλλαγές, διατηρώντας παράλληλα τη ρυθμιστική συνέπεια. Η αποτελεσματική λειτουργία ενός τέτοιου πλαισίου προϋποθέτει υψηλό επίπεδο διεθνούς συντονισμού, προκειμένου να διασφαλιστεί η συνοχή μεταξύ διαφορετικών κρατών και θεσμικών οργάνων. Ωστόσο, οι διαφοροποιήσεις ως προς τον βαθμό οικονομικής ανάπτυξης, τη θεσμική ικανότητα και τις κοινωνικές προτεραιότητες ενδέχεται να παρεμποδίσουν την ομοιόμορφη εφαρμογή των κανονιστικών ρυθμίσεων. Σε αυτό το πλαίσιο, καθίσταται σαφές ότι η διαμόρφωση παγκοσμίως συνεκτικών ρυθμιστικών πλαισίων απαιτεί τη λεπτή εξισορρόπηση μεταξύ ενιαίων προτύπων και περιφερειακής ευελιξίας, με στόχο τόσο την ενίσχυση της αποτελεσματικότητας όσο και τη διασφάλιση της νομιμοποίησης των ρυθμιστικών παρεμβάσεων (Νυκονιά κ.ά., 2025).

3.2 Βασικές αρχές ηθικής και συστάσεις για την τεχνητή νοημοσύνη

Ένας ακόμη τομέας ενασχόλησης της διεθνούς κοινότητας αφορά την ανάπτυξη πλαισίων δεοντολογίας για τον σχεδιασμό, την υιοθέτηση και τη χρήση συστημάτων τεχνητής νοημοσύνης. Στο πλαίσιο της παρούσας εργασίας, επιλέχθηκε η συνοπτική παρουσίαση των πλαισίων που έχουν διαμορφώσει ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (ΟΟΣΑ), η Ευρωπαϊκή Επιτροπή και ο Οργανισμός των Ηνωμένων Εθνών για την Εκπαίδευση, την Επιστήμη και τον Πολιτισμό (UNESCO). Η επιλογή των τριών αυτών πλαισίων βασίζεται στην υψηλή θεσμική αξιοπιστία των φορέων που τα εισήγαγαν, στο αναλυτικό περιεχόμενό τους, καθώς και στο ευρύ πεδίο διεθνούς αποδοχής τους.

Παρότι τα επιμέρους κείμενα δεν ταυτίζονται, συγκλίνουν σε θεμελιώδεις αρχές, όπως η δικαιοσύνη, η διαφάνεια και η λογοδοσία. Οι αρχές αυτές αποτελούν θεμέλιο για τον περιορισμό των ηθικών κινδύνων στα συστήματα τεχνητής νοημοσύνης, καθώς στοχεύουν στη μη διάκριση, στη σαφήνεια των διαδικασιών λήψης αποφάσεων και στην ανάληψη ευθύνης από τους σχεδιαστές και τους χρήστες των συστημάτων (Valavan, 2023).

3.2.1 Η έννοια της ηθικής

Η ηθική, ή αλλιώς ηθική φιλοσοφία, ορίζεται γενικά ως ο κλάδος της φιλοσοφίας που εξετάζει τις έννοιες του σωστού και του λάθους, καθώς και τις αρχές που διέπουν την ανθρώπινη πράξη και συμπεριφορά. Αποτελεί έναν από τους αρχαιότερους τομείς της

φιλοσοφικής σκέψης, με τις ιστορικές ρίζες της να ανάγονται στην αρχαιότητα. Ο πολιτισμός της αρχαίας Ελλάδας, με εξέχοντες στοχαστές όπως ο Σωκράτης, ο Πλάτωνας και ο Αριστοτέλης (Chaddha & Agrawal, 2023; Greene, 2023), καθώς και οι πρώιμοι πολιτισμοί της Μεσοποταμίας, συνέβαλαν καθοριστικά στη θεμελίωση των πρώτων εννοιών της ηθικής σκέψης (Chaddha & Agrawal, 2023).

Τα πρώτα γνωστά λογοτεχνικά έργα αυτών των πολιτισμών, όπως η Ιλιάδα του Ομήρου και το Έπος του Γκιλγκαμές, καθόρισαν ένα σύνολο αξιών που αφορούσαν τα κοινωνικά πρότυπα και τις αρετές των ηρώων, προβάλλοντας ιδανικά συμπεριφοράς που θεωρούνταν αξιοσημείωτα και επιθυμητά. Επιπλέον, ηθικές αρχές εντοπίζονται στις αρχαίες αιγυπτιακές και σουμερικές γραφές, οι οποίες ενθάρρυναν τη φροντίδα των λιγότερο ευνοημένων μελών της κοινωνίας ως τρόπο τιμής και εξευμενισμού των θεών. Στο πλαίσιο αυτών των παραδόσεων καταγράφεται και η πρώιμη μορφή του λεγόμενου «Χρυσού Κανόνα της Ηθικής», σύμφωνα με τον οποίο ο άνθρωπος οφείλει να συμπεριφέρεται στους άλλους όπως θα επιθυμούσε να του συμπεριφέρονται (Chaddha & Agrawal, 2023).

Η εγκυκλοπαίδεια Britannica ορίζει ότι η ηθική μπορεί να αναφέρεται στον κλάδο της επιστήμης της φιλοσοφίας που μελετά την έννοια του σωστού και του λάθους, καθώς και του καλού και του κακού. Μπορεί επίσης να περιγράφει οποιαδήποτε φιλοσοφική θεωρία που εξετάζει τι θεωρείται σωστό ή λανθασμένο και τι αποτελεί καλό ή κακό και μπορεί να χρησιμοποιείται για να δηλώσει ένα σύστημα ή έναν κώδικα ηθικών κανόνων, αρχών ή αξιών. Ένα τέτοιο σύστημα μπορεί να συνδέεται με συγκεκριμένες θρησκείες, πολιτισμούς, επαγγελματικές ομάδες ή άλλες κοινότητες που χαρακτηρίζονται, τουλάχιστον εν μέρει, από την ιδιαίτερη ηθική τους προοπτική (Singer, 2025). Αντίστοιχα, το πανεπιστήμιο Santa Clara των ΗΠΑ όρισε την ηθική δίνοντάς της δύο διακριτές διαστάσεις. Από τη μία η ηθική αναφέρεται στα θεμελιωμένα πρότυπα του σωστού και του λάθους που καθορίζουν τι πρέπει να κάνουν οι άνθρωποι, συνήθως όσον αφορά τα δικαιώματα, τις υποχρεώσεις, τα οφέλη για την κοινωνία, τη δικαιοσύνη ή συγκεκριμένες αρετές. Από την άλλη, η ηθική αναφέρεται στη συνεχή μελέτη και την ανάπτυξη των ηθικών προτύπων του ατόμου διασφαλίζοντας ότι τα άτομα και οι θεσμοί τους ανταποκρίνονται σε πρότυπα που είναι λογικά και έχουν στέρεα βάση (Velasquez κ.ά., 2010).

Η σύγχρονη δυτική ηθική φιλοσοφία διακρίνει τρεις βασικούς κλάδους: τη μεταηθική, την κανονιστική και την εφαρμοσμένη ηθική (Greene, 2023; Singer, 2025). Η μεταηθική εξετάζει

και αναθεωρεί την προέλευση, τη φύση και το νόημα των ηθικών μας αξιών, αρχών και του γλωσσικού πλαισίου μέσα στο οποίο αυτές εκφράζονται, αποτελώντας ουσιαστικά την ηθική της ηθικής. Η κανονιστική ηθική επιδιώκει να θέσει τα ηθικά πρότυπα που καθορίζουν τη διάκριση μεταξύ σωστής και εσφαλμένης πράξης, ενώ η εφαρμοσμένη ηθική επιδιώκει την εφαρμογή των κανονιστικών ηθικών θεωριών σε συγκεκριμένους κλάδους για την παροχή πρακτικών λύσεων σε ηθικά διλήμματα (Chaddha & Agrawal, 2023; Greene, 2023; Singer, 2025).

Στο πλαίσιο της εφαρμοσμένης ηθικής, η ανάπτυξη και χρήση συστημάτων τεχνητής νοημοσύνης αναδύεται ως ένα σημαντικό θέμα ενδιαφέροντος εξαιτίας των δυνατοτήτων μετασχηματισμού που αυτά προσφέρουν. Η ευρεία υιοθέτηση τους εγείρει κρίσιμα ηθικά ζητήματα, ιδίως σε σχέση με τις ανθρώπινες αντιλήψεις για το τι θεωρείται σωστό ή αποδεκτό και τι λάθος ή απαράδεκτο. Ερωτήματα όπως η αναγκαιότητα ανάπτυξης συστημάτων τεχνητής νοημοσύνης γενικότερα ή ο τρόπος χρήσης τους σε συγκεκριμένους ευαίσθητους τομείς, πχ. δικαστικό σύστημα, αποτελούν σημαντικά διλήμματα της σύγχρονης κοινωνίας τα οποία πρέπει να απαντηθούν.

3.2.2 Επισκόπηση των αρχών τεχνητής νοημοσύνης του ΟΟΣΑ

Οι αρχές του Οργανισμού Οικονομικής Συνεργασίας και Ανάπτυξης για την τεχνητή νοημοσύνη υιοθετήθηκαν αρχικά το 2019 και επικαιροποιήθηκαν το 2024, ώστε να ενσωματώσουν τις νέες τεχνολογικές και πολιτικές εξελίξεις, διασφαλίζοντας την εγκυρότητα και καταλληλόλητά τους. Αποτελούν κατευθυντήριες γραμμές τόσο για τους φορείς που ασχολούνται με την ανάπτυξη συστημάτων τεχνητής νοημοσύνης, όσο και για τους υπεύθυνους χάραξης πολιτικής. Οι αρχές ταξινομούνται σε δύο κατηγορίες:

- τις αρχές βασισμένες σε αξίες, οι οποίες αποτελούν κατευθυντήριες γραμμές για την ανάπτυξη στρατηγικών και πολιτικών για αξιόπιστη τεχνητή νοημοσύνη και μπορούν να χρησιμοποιηθούν τόσο από κυβερνήσεις όσο και από άλλους φορείς·
- τις συστάσεις προς τους υπεύθυνους χάραξης εθνικής πολιτικής με σκοπό τα οικοσυστήματα τεχνητής νοημοσύνης να ωφελήσουν τις κοινωνίες (OECD, 2024a).

Οι αρχές βασισμένες σε αξίες του ΟΟΣΑ είναι:

- 1.1 Συμπεριληπτική και βιώσιμη ανάπτυξη και ευημερία

Οι εμπλεκόμενοι φορείς οφείλουν να προωθούν ενεργά την υπεύθυνη και αξιόπιστη τεχνητή νοημοσύνη, προκειμένου να επιτύχουν θετικά αποτελέσματα για την ανθρωπότητα και τον πλανήτη. Τούτο περιλαμβάνει την ενίσχυση των ανθρώπινων ικανοτήτων, τη συμπερίληψη, την ισότητα, την προστασία του περιβάλλοντος και την προώθηση της βιώσιμης και χωρίς αποκλεισμούς ανάπτυξης.

- 1.2 Σεβασμός του κράτους δικαίου, των ανθρωπίνων δικαιωμάτων και των δημοκρατικών αξιών, συμπεριλαμβανομένης της δικαιοσύνης και της ιδιωτικής ζωής

Οι εμπλεκόμενοι φορείς πρέπει να σέβονται το κράτος δικαίου, τα ανθρώπινα δικαιώματα και τις δημοκρατικές αξίες σε ολόκληρο τον κύκλο ζωής ενός συστήματος τεχνητής νοημοσύνης. Πρέπει να καταπολεμούν την παραπληροφόρηση που ενισχύεται από την τεχνητή νοημοσύνη, προστατεύοντας παράλληλα την ελευθερία της έκφρασης και άλλες ελευθερίες που κατοχυρώνονται από το διεθνές δίκαιο, και να εφαρμόζουν κατάλληλους μηχανισμούς ελέγχου, όπως η ανθρώπινη εποπτεία, για τον μετριασμό των κινδύνων από την κατάχρηση ή τις ακούσιες εφαρμογές.

- 1.3 Διαφάνεια και εξηγησιμότητα

Οι εμπλεκόμενοι παράγοντες οφείλουν να δεσμεύονται ρητά στη διαφάνεια και την υπεύθυνη αποκάλυψη χρήσης των συστημάτων τεχνητής νοημοσύνης, παρέχοντας κατάλληλες πληροφορίες προς το ζητούμενο πλαίσιο. Η εν λόγω δέσμευση καλλιεργεί την κατανόηση των δυνατοτήτων και περιορισμών των εν λόγω συστημάτων, ενημερώνει τους ενδιαφερόμενους φορείς περί των αλληλεπιδράσεών τους με την τεχνητή νοημοσύνη, διευκρινίζει τις πηγές δεδομένων, τις διεργασίες και τη λογική πίσω από τα παραγόμενα αποτελέσματα, όπου είναι εφικτό, και διευκολύνει την αμφισβήτηση δυσμενών αποφάσεων.

- 1.4 Ανθεκτικότητα, ασφάλεια και προστασία

Τα συστήματα τεχνητής νοημοσύνης οφείλουν να επιδεικνύουν ανθεκτικότητα, ασφάλεια και προστασία καθ' όλη τη διάρκεια του κύκλου ζωής τους, εξασφαλίζοντας την ορθή λειτουργία τους υπό οποιεσδήποτε συνθήκες. Θα πρέπει να υπάρχουν διαθέσιμοι μηχανισμοί παράκαμψης, επισκευής ή απόσυρσης των συστημάτων σε

περιπτώσεις πρόκλησης βλάβης ή ανεπιθύμητης συμπεριφοράς, ενώ, όπου είναι εφικτό, θα πρέπει να ενισχύεται η ακεραιότητα των πληροφοριών με σεβασμό στην ελευθερία έκφρασης.

- 1.5 Λογοδοσία

Οι εμπλεκόμενοι φορείς φέρουν ευθύνη για την ορθή λειτουργία των συστημάτων τεχνητής νοημοσύνης και την τήρηση των σχετικών αρχών, σύμφωνα με τους ρόλους τους, το πλαίσιο και την τεχνολογική πρόοδο. Πρέπει να εξασφαλίζουν την ιχνηλασιμότητα των δεδομένων, των διεργασιών και των αποφάσεων καθ' όλη τη διάρκεια του κύκλου ζωής, διευκολύνοντας την ανάλυση πιθανών ερωτημάτων που μπορεί να προκύψουν. Τέλος, καλούνται να εφαρμόσουν μια συστηματική μεθοδολογία διαχείρισης κινδύνων, υιοθετώντας υπεύθυνη επιχειρηματική συμπεριφορά για την αντιμετώπιση προκαταλήψεων, παραβιάσεων ανθρωπίνων δικαιωμάτων, ασφάλειας, ιδιωτικότητας, εργασιακών και πνευματικών δικαιωμάτων (OECD, 2024a).

Οι συστάσεις προς τους υπεύθυνους χάραξης εθνικής πολιτικής είναι:

- 2.1 Επένδυση στην έρευνα και ανάπτυξη της τεχνητής νοημοσύνης

Οι κυβερνήσεις οφείλουν να εξετάσουν μακροπρόθεσμες δημόσιες επενδύσεις και να ενθαρρύνουν ιδιωτικές στον τομέα της έρευνας και ανάπτυξης αξιόπιστης τεχνητής νοημοσύνης, εστιάζοντας στις πιθανές τεχνικές, κοινωνικές, νομικές και ηθικές επιπτώσεις. Επιπλέον, πρέπει να εξετάσουν την χρηματοδότηση εργαλείων ανοικτού κώδικα, καθώς και ανοικτών, αντιπροσωπευτικών συνόλων δεδομένων για την καλύτερη υποστήριξη της έρευνας στον τομέα της τεχνητής νοημοσύνης.

- 2.2 Προώθηση ενός οικοσυστήματος τεχνητής νοημοσύνης

Οι κυβερνήσεις οφείλουν να προωθήσουν την ανάπτυξη και την πρόσβαση σε ένα περιεκτικό, βιώσιμο και διαλειτουργικό ψηφιακό οικοσύστημα για την τεχνητή νοημοσύνη, το οποίο, μεταξύ άλλων, θα περιλαμβάνει δεδομένα, τεχνολογίες τεχνητής νοημοσύνης, υπολογιστική υποδομή κ.ά., καθώς και μηχανισμούς ανταλλαγής.

- 2.3 Διαμόρφωση και ενεργοποίηση ενός διαλειτουργικού περιβάλλοντος διακυβέρνησης και πολιτικής για την τεχνητή νοημοσύνη

Οι κυβερνήσεις οφείλουν να προωθήσουν ένα ευέλικτο πλαίσιο διακυβέρνησης, διευκολύνοντας τη μετάβαση συστημάτων τεχνητής νοημοσύνης από το στάδιο της έρευνας στο στάδιο ανάπτυξης και λειτουργίας. Για αυτό τον σκοπό, οφείλουν να εξετάζουν και να προσαρμόζουν, κατά περίπτωση, τα κανονιστικά πλαίσια και τους μηχανισμούς αξιολόγησης που εφαρμόζουν στα συστήματα τεχνητής νοημοσύνης, προκειμένου να ενθαρρύνουν τη διασυννοριακή συνεργασία, την καινοτομία και τον ανταγωνισμό για αξιόπιστη τεχνητή νοημοσύνη.

- 2.4 Ανάπτυξη των ανθρώπινων ικανοτήτων και προετοιμασία για τον μετασχηματισμό της αγοράς εργασίας

Οι κυβερνήσεις οφείλουν να συνεργαστούν στενά με τους εμπλεκόμενους φορείς και να εξοπλίσουν τους πολίτες με τις κατάλληλες δεξιότητες ώστε να τους προετοιμάσουν για τον επερχόμενο κοινωνικό και εργασιακό μετασχηματισμό. Οφείλουν επίσης, να προωθήσουν την υπεύθυνη χρήση της τεχνητής νοημοσύνης στον εργασιακό τομέα, ενισχύοντας την ασφάλεια, την ποιότητα των θέσεων εργασίας, την επιχειρηματικότητα, την παραγωγικότητα και τη δίκαιη κατανομή των ωφελειών.

- 2.5 Διεθνής συνεργασία για αξιόπιστη τεχνητή νοημοσύνη

Οι κυβερνήσεις και οι εμπλεκόμενοι φορείς, οφείλουν να συνεργάζονται ενεργά για την προώθηση αρχών υπεύθυνης διαχείρισης της τεχνητής νοημοσύνης. Οφείλουν να ενθαρρύνουν την ανταλλαγή γνώσεων, τις διεπιστημονικές πρωτοβουλίες, την ανάπτυξη παγκόσμιων τεχνικών προτύπων διαλειτουργικότητας και συγκρίσιμων δεικτών μέτρησης έρευνας και ανάπτυξης της τεχνητής νοημοσύνης.

Οι αρχές του ΟΟΣΑ αποτελούν ένα ολοκληρωμένο πλαίσιο που προωθεί τη δημιουργία αξιόπιστων συστημάτων τεχνητής νοημοσύνης και τη θέσπιση στοχευμένων πολιτικών που θα ωφελήσουν το ευρύτερο κοινωνικό σύνολο. Με αυτόν τον τρόπο, ισορροπεί την ανάγκη για τεχνολογική πρόοδο με την υποχρέωση προστασίας των ανθρώπινων δικαιωμάτων και του γενικότερου καλού.

3.2.3 Κατευθυντήριες γραμμές της Ευρωπαϊκής Επιτροπής

Η Ευρωπαϊκή Επιτροπή δημιούργησε το 2018 μία ανεξάρτητη ομάδα εμπειρογνομόνων για την τεχνητή νοημοσύνη (High-Level Expert Group on Artificial Intelligence – AI HLEG), ζητώντας ένα από τα βασικά παραδοτέα να είναι οι κατευθυντήριες γραμμές για την ηθική της. Βασική επιδίωξή ήταν να καταστήσει την ηθική βασικό πυλώνα των συστημάτων τεχνητής νοημοσύνης, με σκοπό την ενδυνάμωση και προστασία τόσο της ατομικής όσο και της κοινωνικής ευημερίας.

Κατά την ομάδα των εμπειρογνομόνων τα αξιόπιστα συστήματα τεχνητής νοημοσύνης συνίστανται από τρία θεμελιώδη στοιχεία, τα οποία απαιτείται να εκπληρώνονται καθ' όλη τη διάρκεια του κύκλου ζωής τους. Συγκεκριμένα, ένα σύστημα οφείλει να είναι νόμιμο, ηθικό, και ανθεκτικό, τόσο από τεχνικής όσο και από κοινωνικής σκοπιάς, αφού ακόμη και με αγαθές προθέσεις, τα συστήματα τεχνητής νοημοσύνης ενδέχεται να προκαλέσουν ακούσιες βλάβες. Η συνεχής εκπλήρωση των στοιχείων αυτών παρέχει τη βάση πάνω στην οποία όλοι όσοι επηρεάζονται από τα συστήματα αυτά μπορούν να εμπιστευτούν ότι ο σχεδιασμός, η ανάπτυξη και η χρήση τους είναι νόμιμη, ηθική και αξιόπιστη. Η εξασφάλιση της αξιοπιστίας αποτελεί κρίσιμο παράγοντα ώστε οι πολίτες να αποκομίσουν πλήρως τις ωφέλειες των συστημάτων τεχνητής νοημοσύνης, με την ταυτόχρονη ελαχιστοποίηση των πιθανών κινδύνων (High-Level Expert Group on Artificial Intelligence, 2019).

Η ομάδα των εμπειρογνομόνων λαμβάνοντας υπόψη τα θεμελιώδη ανθρώπινα δικαιώματα που ορίζονται από το διεθνές δίκαιο και το νομοθετικό πλαίσιο της Ευρωπαϊκής Ένωσης, πρότεινε τέσσερις ηθικές αρχές, τηρουμένων των οποίων, διασφαλίζεται ότι τα συστήματα τεχνητής νοημοσύνης αναπτύσσονται, χρησιμοποιούνται και εφαρμόζονται με αξιόπιστο τρόπο. Οι αρχές αυτές είναι:

- Σεβασμός στην ανθρώπινη αυτονομία

Τα θεμελιώδη δικαιώματα της Ευρωπαϊκής Ένωσης διασφαλίζουν την ελευθερία και αυτονομία των πολιτών. Κατά την αλληλεπίδραση με συστήματα τεχνητής νοημοσύνης, οι πολίτες πρέπει να διατηρούν πλήρως το δικαίωμα στην αυτοδιάθεση, ενώ τα συστήματα τεχνητής νοημοσύνης δεν πρέπει να εξαναγκάζουν, εξαπατούν, χειραγωγούν, επηρεάζουν ή να κατευθύνουν τους πολίτες χωρίς λόγο. Ο σχεδιασμός των συστημάτων πρέπει να είναι ανθρωποκεντρικός με σκοπό την ενίσχυση των δυνατοτήτων των πολιτών και την δημιουργία αξίας.

- Πρόληψη βλάβης

Τα συστήματα τεχνητής νοημοσύνης πρέπει να διασφαλίζουν την προστασία της ανθρώπινης αξιοπρέπειας, της ψυχικής και σωματικής ακεραιότητας, αποφεύγοντας την πρόκληση ή επιδείνωση βλάβης. Πρέπει να είναι ασφαλή, ανθεκτικά και προστατευμένα από κακόβουλη χρήση. Ιδιαίτερη φροντίδα απαιτείται για ευάλωτες ομάδες, ειδικά σε περιπτώσεις ανισορροπίας δύναμης και πληροφόρησης, όπως στην σχέση εργοδότη και εργαζόμενου, καθώς και σε περιπτώσεις προστασίας του φυσικού περιβάλλοντος και άλλων ζωντανών οργανισμών.

- Αρχή της δικαιοσύνης

Η ανάπτυξη, εφαρμογή και χρήση συστημάτων τεχνητής νοημοσύνης οφείλει να διακρίνεται από δικαιοσύνη, τόσο σε ουσιαστική όσο και σε διαδικαστική διάσταση. Η ουσιαστική διάσταση απαιτεί τη δίκαιη κατανομή ωφελειών και βαρών, την αποφυγή διακρίσεων, εξαπάτησης ή περιορισμού της ελευθερίας των πολιτών, την παροχή ίσων ευκαιριών πρόσβασης στην εκπαίδευση, σε αγαθά, σε υπηρεσίες και τεχνολογίες, καθώς και την τήρηση της αναλογικότητας των μέσων για την επίτευξη των σκοπών. Η διαδικαστική διάσταση συνεπάγεται την ικανότητα αμφισβήτησης και αναζήτησης αποτελεσματικής αποκατάστασης έναντι αποφάσεων που λαμβάνονται από συστήματα τεχνητής νοημοσύνης και από τους ανθρώπους που τα χειρίζονται.

- Αρχή της εξηγησιμότητας

Η εξηγησιμότητα αποτελεί κρίσιμο παράγοντα εδραίωσης και διατήρησης της εμπιστοσύνης των χρηστών στα συστήματα τεχνητής νοημοσύνης, απαιτώντας τη διαφάνεια των διεργασιών, τη σαφή επικοινωνία των δυνατοτήτων και του σκοπού των συστημάτων, καθώς και εξηγήσιμες αποφάσεις για τους άμεσα και έμμεσα ενδιαφερόμενους. Σε περιπτώσεις όπου δεν είναι δυνατή η εξήγηση της λειτουργίας ενός συστήματος, φαινόμενο γνωστό και ως «μαύρο κουτί» (black box), οφείλουν να εφαρμοστούν εναλλακτικά μέτρα ικανά να ενισχύσουν την εξηγησιμότητα του συστήματος (High-Level Expert Group on Artificial Intelligence, 2019).

Οι αρχές αυτές θέτουν τα θεμέλια για την ανάπτυξη και χρήση αξιόπιστων συστημάτων τεχνητής νοημοσύνης. Για την ολοκλήρωση των συστημάτων, οι αρχές αυτές οφείλουν να μετατραπούν σε συγκεκριμένες απαιτήσεις που πρέπει να ισχύουν για όλους τους εμπλεκόμενους φορείς, από τους δημιουργούς μέχρι τους τελικούς χρήστες και την ευρύτερη

κοινωνία. Η λίστα των απαιτήσεων δεν είναι εξαντλητική και αναφέρεται σε συστημικές, ατομικές και κοινωνικές πτυχές:

- Ανθρώπινη δράση και εποπτεία

Η υποστήριξη της ανθρώπινης αυτονομίας και της δυνατότητας λήψης αποφάσεων αποτελεί θεμελιώδη αρχή για την ανάπτυξη αξιόπιστων συστημάτων τεχνητής νοημοσύνης. Τα συστήματα οφείλουν να ενισχύουν τη δημοκρατική συμμετοχή, την ισότητα και τα θεμελιώδη δικαιώματα, διασφαλίζοντας παράλληλα ουσιαστικό ανθρώπινο έλεγχο. Παρότι η τεχνητή νοημοσύνη μπορεί να προάγει δικαιώματα, όπως την πρόσβαση στην εκπαίδευση ή τη διαχείριση προσωπικών δεδομένων, ενδέχεται επίσης να τα απειλήσει: συνεπώς, απαιτείται η εκτίμηση των επιπτώσεων στα θεμελιώδη δικαιώματα ήδη από το στάδιο του σχεδιασμού. Οι χρήστες πρέπει να διαθέτουν επαρκή γνώση και εργαλεία ώστε να κατανοούν, να αξιολογούν ή να αμφισβητούν αποφάσεις συστημάτων, διατηρώντας την ικανότητα αυτόνομης επιλογής. Τέλος, η ανθρώπινη εποπτεία εξασφαλίζει ότι τα συστήματα δεν υπονομεύουν την ανθρώπινη ευθύνη και αυτονομία, αλλά παραμένουν υπό διαρκή θεσμικό και ηθικό έλεγχο.

- Τεχνική ανθεκτικότητα και ασφάλεια

Η τεχνική ανθεκτικότητα και ασφάλεια συνδέεται στενά με την αρχή αποφυγής βλάβης. Τα συστήματα τεχνητής νοημοσύνης πρέπει να σχεδιάζονται εξασφαλίζοντας την ασφαλή και συνεπή λειτουργία από τις μεταβολές του περιβάλλοντος ή εχθρικές παρεμβάσεις, περιορίζοντας ταυτόχρονα τις ακούσιες ή αναπάντεχες βλάβες. Η ανθεκτικότητα απέναντι σε επιθέσεις απαιτεί την προστασία των δεδομένων, των μοντέλων και της υποδομής, ενώ η ύπαρξη μηχανισμών ασφαλείας και εναλλακτικών σχεδίων δράσης διασφαλίζει τη συνέχιση της λειτουργίας χωρίς πρόκληση βλάβης σε ανθρώπους ή στο περιβάλλον. Επιπλέον, η ακρίβεια, η αξιοπιστία και η δυνατότητα αναπαραγωγής των αποτελεσμάτων αποτελούν κρίσιμα κριτήρια, ιδιαίτερα όταν το σύστημα επιδρά άμεσα στη ζωή των ανθρώπων, καθώς ενισχύουν τη δυνατότητα ελέγχου, επαλήθευσης και έγκαιρης διόρθωσης σφαλμάτων.

- Προστασία της ιδιωτικότητας και διακυβέρνηση των δεδομένων

Η προστασία της ιδιωτικότητας και των δεδομένων αποτελεί θεμελιώδες δικαίωμα που επηρεάζεται άμεσα από τη λειτουργία των συστημάτων τεχνητής νοημοσύνης και συνδέεται με την αρχή της πρόληψης βλάβης. Η πρόληψη των κινδύνων για την ιδιωτικότητα προϋποθέτει αποτελεσματική διακυβέρνηση δεδομένων, διασφαλίζοντας την ποιότητα, την

ακεραιότητα και τη συνάφειά τους, Τα συστήματα τεχνητής νοημοσύνης οφείλουν να σέβονται την ιδιωτικότητα και να προστατεύουν τα δεδομένα σε ολόκληρο τον κύκλο ζωής τους. Η ύπαρξη προκαταλήψεων ή σφαλμάτων στα σύνολα δεδομένων πρέπει να εντοπίζεται και να διορθώνεται εκ των προτέρων, ενώ η πρόσβαση σε αυτά οφείλει να επιτρέπεται αποκλειστικά σε κατάλληλα εξουσιοδοτημένο και καταρτισμένο προσωπικό.

- Διαφάνεια

Η απαίτηση για διαφάνεια στα συστήματα τεχνητής νοημοσύνης συνδέεται άμεσα με την αρχή της εξηγησιμότητας και αφορά τρία κρίσιμα πεδία: τα δεδομένα, το ίδιο το σύστημα και τα επιχειρηματικά μοντέλα που το υποστηρίζουν. Η δυνατότητα τεκμηρίωσης και επαλήθευσης είναι σημαντική τόσο για τις διαδικασίες συλλογής και ανάλυσης δεδομένων όσο και για τους αλγορίθμους και τις εξόδους του. Όταν ένα σύστημα τεχνητής νοημοσύνης έχει σημαντική επίδραση στη ζωή των ανθρώπων, θα πρέπει να είναι δυνατή η εξήγηση της διαδικασίας λήψης μίας απόφασης. Παράλληλα, είναι σκόπιμο να εξασφαλίζεται η σαφής ενημέρωση των ανθρώπων όταν αλληλοεπιδρούν με ένα σύστημα τεχνητής νοημοσύνης, παρέχοντας ταυτόχρονα τη δυνατότητα εναλλακτικής, ανθρώπινης επιλογής.

- Διαφορετικότητα, αποφυγή διακρίσεων και δικαιοσύνη

Η απαίτηση για διαφορετικότητα, αποφυγή διακρίσεων και δικαιοσύνη συνδέεται με την αρχή της δικαιοσύνης. Η αποφυγή των προκαταλήψεων απαιτεί τη διόρθωση ιστορικών ή δομικών διακρίσεων στα σύνολα δεδομένων και στους αλγορίθμους, την υιοθέτηση μηχανισμών εποπτείας και την ευρεία συμμετοχή διαφορετικών ομάδων καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων. Παράλληλα, η εφαρμογή αρχών καθολικού σχεδιασμού (Universal Design) διασφαλίζει την προσβασιμότητα και τη χρήση των συστημάτων από όλους, ανεξαρτήτως φύλου, ηλικίας ή ικανοτήτων, με ιδιαίτερη μέριμνα για τα άτομα με αναπηρίες. Τέλος, η συμμετοχή όλων των εμπλεκόμενων φορέων σε όλα τα στάδια ανάπτυξης, υλοποίησης και αξιολόγησης των συστημάτων τεχνητής νοημοσύνης ενισχύει τη διαφάνεια, τη λογοδοσία και την κοινωνική αποδοχή των τεχνολογιών τεχνητής νοημοσύνης.

- Κοινωνική και περιβαλλοντική ευημερία

Η απαίτηση για κοινωνική και περιβαλλοντική ευημερία συνδέεται με τις αρχές της δικαιοσύνης και της πρόληψης βλάβης. Η ανάπτυξη και χρήση συστημάτων τεχνητής νοημοσύνης οφείλει να λαμβάνει υπόψη το ευρύτερο κοινωνικό σύνολο και το περιβάλλον

και να προωθεί την ενίσχυση της βιωσιμότητας και της οικολογικής ευθύνης σε όλα τα στάδια του κύκλου ζωής των συστημάτων. Η ευρεία παρουσία των συστημάτων τεχνητής νοημοσύνης επιδρά στις κοινωνικές σχέσεις και την ψυχική ευεξία των ανθρώπων, γεγονός που απαιτεί συστηματική παρακολούθηση τους. Τέλος, η χρήση τους οφείλει να αξιολογείται και ως προς τον αντίκτυπό στη δημοκρατία και στους θεσμούς, ιδίως σε διαδικασίες πολιτικής ή εκλογικής φύσης.

- Λογοδοσία

Η απαίτηση για λογοδοσία συμπληρώνει τις παραπάνω απαιτήσεις, συνδέεται με την αρχή της δικαιοσύνης και επιβάλλει τη δημιουργία μηχανισμών που διασφαλίζουν την υπευθυνότητα σε όλα τα στάδια ανάπτυξης, εφαρμογής και χρήσης των συστημάτων τεχνητής νοημοσύνης. Στο πλαίσιο της λογοδοσίας είναι αναγκαία η ελεύθερη δυνατότητα ελέγχου και αξιολόγησης των συστημάτων ενισχύοντας τη διαφάνεια και την αξιοπιστία, ιδίως όταν διακυβεύονται τα θεμελιώδη δικαιώματα των ανθρώπων. Εξίσου κρίσιμη είναι η τεκμηρίωση, η αξιολόγηση και η ελαχιστοποίηση των αρνητικών επιπτώσεων από τα συστήματα, με την ταυτόχρονη προστασία όσων αναδεικνύουν ζητήματα ηθικής ή ασφάλειας. Τέλος, αναγκαία προϋπόθεση αποτελεί η ύπαρξη αποτελεσματικών μηχανισμών αποκατάστασης για όσους υφίστανται αδικαιολόγητες αρνητικές συνέπειες από τα συστήματα τεχνητής νοημοσύνης (High-Level Expert Group on Artificial Intelligence, 2019).

Οι κατευθυντήριες γραμμές της Ευρωπαϊκής Επιτροπής εδραιώνουν ένα πλαίσιο όπου η νομιμότητα, η ηθική, η αποφυγή βλάβης και ο σεβασμός των θεμελιωδών δικαιωμάτων αποτελούν βασικές προϋποθέσεις για τον σχεδιασμό και τη χρήση των συστημάτων. Μέσα από αρχές όπως η ανθρώπινη αυτονομία, η διαφάνεια, η λογοδοσία και η προστασία της ιδιωτικότητας, επιδιώκουν να διασφαλίσουν ότι η τεχνητή νοημοσύνη ενδυναμώνει τους πολίτες, προστατεύει τις ευάλωτες ομάδες και στηρίζει μια δημοκρατική και κοινωνικά δίκαιη ψηφιακή μετάβαση (High-Level Expert Group on Artificial Intelligence, 2019).

3.2.4 Σύσταση για την ηθική της τεχνητής νοημοσύνης από την UNESCO

Η UNESCO υιοθέτησε τον Νοέμβριο του 2021 τη Σύσταση για την ηθική της τεχνητής νοημοσύνης⁴, εγκαινιάζοντας μια συστηματική και ολοκληρωμένη προσέγγιση η οποία

⁴ UNESCO - Recommendation on the Ethics of Artificial Intelligence

εδράζεται σε ένα ολιστικό, συμπεριληπτικό και εξελισσόμενο πλαίσιο αλληλένδετων αξιών, αρχών και δράσεων (UNESCO, 2022).

Η Σύσταση περιλαμβάνει αξίες και αρχές, οι οποίες οφείλουν να γίνονται σεβαστές από όλους τους εμπλεκόμενους φορείς καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων τεχνητής νοημοσύνης. Οι αξίες διαδραματίζουν καίριο ρόλο λειτουργώντας ως κίνητρα για τη διαμόρφωση κανονιστικών πλαισίων, με τις αρχές να στοχεύουν στην ευχερέστερη λειτουργική ενσωμάτωσή τους μέσω πλαισίων πολιτικής και δράσης. Λαμβάνοντας υπόψη τις αξίες και αρχές της σύστασης, η UNESCO προτείνει την προώθηση τροποποιήσεων στο υφιστάμενο κανονιστικό πλαίσιο και τις επιχειρηματικές κατευθυντήριες γραμμές, πάντα σε συμφωνία με το διεθνές δίκαιο, τις υποχρεώσεις των κρατών στον τομέα των ανθρωπίνων δικαιωμάτων, και άλλων εν ενεργεία διεθνών συμφωνιών. Τυχόν περιορισμοί των ανθρωπίνων δικαιωμάτων και των θεμελιωδών ελευθεριών από τα συστήματα τεχνητής νοημοσύνης πρέπει να έχουν νόμιμη βάση, να είναι εύλογοι, αναγκαίοι, αναλογικοί, και να συνάδουν με τις υποχρεώσεις των κρατών βάσει του διεθνούς δικαίου. Η εμπιστοσύνη και ακεραιότητα του κύκλου ζωής των εν λόγω συστημάτων συνιστούν προϋπόθεση για τη διασφάλιση της λειτουργίας τους προς όφελος των πολιτών, των κοινωνιών, του περιβάλλοντος και των οικοσυστημάτων, με ταυτόχρονη λήψη μέτρων μετριασμού των κινδύνων για την προαγωγή των ατομικών και συλλογικών ωφελειών (UNESCO, 2022).

Οι αξίες που περιλαμβάνονται στη σύσταση είναι:

- Σεβασμός, προστασία και προώθηση των ανθρωπίνων δικαιωμάτων, των θεμελιωδών ελευθεριών και της ανθρώπινης αξιοπρέπειας

Ο σεβασμός, η προστασία και η προαγωγή της ανθρώπινης αξιοπρέπειας απαιτούνται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων τεχνητής νοημοσύνης, ανεξαρτήτως φυλής, φύλου, ηλικίας ή άλλων χαρακτηριστικών. Όλοι οι εμπλεκόμενοι φορείς οφείλουν να συμμορφώνονται με τα διεθνή πρότυπα ανθρωπίνων δικαιωμάτων, προάγοντας νέα μέσα άσκησής τους μέσω της τεχνητής νοημοσύνης.

- Ευημερία του περιβάλλοντος και του οικοσυστήματος

Η ευημερία του περιβάλλοντος και του οικοσυστήματος πρέπει να αναγνωρίζεται, να προστατεύεται και να προωθείται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων τεχνητής νοημοσύνης. Όλοι οι εμπλεκόμενοι φορείς οφείλουν να μειώσουν το περιβαλλοντικό αποτύπωμα των συστημάτων τεχνητής νοημοσύνης,

προκειμένου να εξασφαλίσουν την ελαχιστοποίηση των παραγόντων κινδύνου για το κλίμα και το περιβάλλον.

- Διασφάλιση της διαφορετικότητας και συμπεριληπτικότητας

Η προστασία της διαφορετικότητας και η διασφάλιση της συμπεριληπτικότητας οφείλει να εξασφαλίζεται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων τεχνητής νοημοσύνης, μέσω της προώθησης της ενεργού συμμετοχής όλων των ατόμων ή ομάδων άνευ διακρίσεων. Στο πλαίσιο αυτό είναι αναγκαίο να καταβληθούν προσπάθειες για να ξεπεραστεί η έλλειψη της απαραίτητης τεχνολογικής υποδομής, εκπαίδευσης, δεξιοτήτων, και νομικών πλαισίων στις λιγότερο ανεπτυγμένες ή απομονωμένες χώρες.

- Διαβίωση σε ειρηνικές, δίκαιες και αλληλένδετες κοινωνίες

Η αξία αυτή απαιτεί την προώθηση της ειρήνης, της συμμετοχικότητας, της δικαιοσύνης, της ισότητας και της αλληλεξάρτησης σε όλο τον κύκλο ζωής των συστημάτων τεχνητής νοημοσύνης. Οι εμπλεκόμενοι φορείς οφείλουν την διασφάλιση ειρηνικών και δίκαιων κοινωνιών, οι οποίες στοχεύουν σε ένα μέλλον προς όφελος όλων, σύμφωνα με τα ανθρώπινα δικαιώματα και τις θεμελιώδεις αξίες (UNESCO, 2022).

Οι αρχές που περιλαμβάνονται στη σύσταση της UNESCO είναι:

- Αναλογικότητα και αποφυγή βλάβης

Η επιλογή και εφαρμογή διαφορετικών μεθόδων της τεχνητής νοημοσύνης οφείλουν να συμμορφώνονται με τις αρχές της νομιμότητας και της αναλογικότητας, διασφαλίζοντας παράλληλα την επίτευξη σκοπών με θεμιτά μέσα, την πλήρη προστασία των ανθρωπίνων δικαιωμάτων και την τήρηση υψηλών προδιαγραφών επιστημονικής εγκυρότητας. Αποφάσεις με μη αναστρέψιμες συνέπειες ή συνέπειες κρίσιμες για την ανθρώπινη ζωή πρέπει να υπάγονται σε τελική ανθρώπινη κρίση, ενώ η χρήση συστημάτων τεχνητής νοημοσύνης για σκοπούς κοινωνικής βαθμολόγησης ή μαζικής επιτήρησης απαγορεύεται ρητά.

- Ασφάλεια και προστασία

Οι ανεπιθύμητες βλάβες, καθώς και η ευπάθεια σε περιστατικά ασφαλείας οφείλουν να αποφεύγονται, αντιμετωπίζονται, προλαμβάνονται και εξαλείφονται καθ' όλη τη

διάρκεια του κύκλου ζωής των συστημάτων της τεχνητής νοημοσύνης, προκειμένου να διασφαλίζεται η ασφάλεια και η προστασία των ανθρώπων, του περιβάλλοντος και του οικοσυστήματος.

- Δικαιοσύνη και αποφυγή διακρίσεων

Οι εμπλεκόμενοι φορείς οφείλουν την προώθηση της κοινωνικής δικαιοσύνης, την αποφυγή διακρίσεων καθ' όλη τη διάρκεια ζωής των συστημάτων τεχνητής νοημοσύνης και κατά συνέπεια την διάχυση των ωφελειών της τεχνητής νοημοσύνης σε όλους.

- Βιωσιμότητα

Η έλευση των συστημάτων τεχνητής νοημοσύνης μπορεί να ωφελήσει ή και να εμποδίσει τους στόχους βιωσιμότητας και την υλοποίησή τους. Κατά συνέπεια κρίνεται αναγκαία η συνεχής αξιολόγηση των επιπτώσεων των συστημάτων αυτών στους συμφωνημένους στόχους βιωσιμότητας.

- Δικαίωμα στην ιδιωτικότητα και προστασία δεδομένων

Η προστασία της ιδιωτικότητας οφείλει να τηρείται αυστηρά, να διασφαλίζεται και να προάγεται καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων τεχνητής νοημοσύνης. Παράλληλα, η προστασία δεδομένων πρέπει να συνιστά άμεση προτεραιότητα για όλους τους εμπλεκόμενους φορείς, με τη διαμόρφωσή πλαισίων και μηχανισμών διακυβέρνησης σε εθνικό ή διεθνές επίπεδο, τα οποία υποστηρίζονται από τα δικαστικά συστήματα και εφαρμόζονται πλήρως καθ' όλη τη διάρκεια ζωής των εν λόγω συστημάτων.

- Ανθρώπινη εποπτεία και λήψη αποφάσεων

Τα κράτη μέλη οφείλουν να διασφαλίζουν την ατομική και συλλογική ανθρώπινη εποπτεία επί των συστημάτων τεχνητής νοημοσύνης, καθώς και την ρητή δυνατότητα λήψης απόφασης περί παραχώρησης ελέγχου στα συστήματα αυτά.

- Διαφάνεια και εξηγησιμότητα

Η διαφάνεια και η εξηγησιμότητα των συστημάτων τεχνητής νοημοσύνης συνιστούν θεμελιώδεις προϋποθέσεις για την εγγύηση του σεβασμού, της προστασίας, της προώθησης των ανθρωπίνων δικαιωμάτων, των θεμελιωδών ελευθεριών και των ηθικών αρχών, διευκολύνοντας την αποτελεσματική εφαρμογή εθνικών και διεθνών πλαισίων λογοδοσίας. Η έλλειψη διαφάνειας υπονομεύει την δυνατότητα

αμφισβήτησης των αποφάσεων που βασίζονται σε τέτοια συστήματα, παραβιάζοντας το δικαίωμα σε δίκαιη δίκη και αποτελεσματική αποκατάσταση της νομιμότητας.

- Ευθύνη και λογοδοσία

Οι εμπλεκόμενοι φορείς και τα κράτη μέλη οφείλουν να προασπίζουν τα ανθρώπινα δικαιώματα, τις θεμελιώδεις ελευθερίες, την προστασία του περιβάλλοντος και να αναλαμβάνουν την ευθύνη για τις αποφάσεις και τις ενέργειες που βασίζονται στην τεχνητή νοημοσύνη καθ' όλη τη διάρκεια του κύκλου ζωής των συστημάτων αυτών. Παράλληλα, θα πρέπει να αναπτυχθούν οι κατάλληλοι μηχανισμοί εποπτείας και ελέγχου που θα διασφαλίζουν τη λογοδοσία για τα συστήματα τεχνητής νοημοσύνης και τον αντίκτυπό τους καθ' όλη τη διάρκεια του κύκλου ζωής τους.

- Ευαισθητοποίηση και εκπαίδευση

Όλοι οι εμπλεκόμενοι φορείς οφείλουν να ευαισθητοποιήσουν και να βοηθήσουν τους πολίτες και το κοινωνικό σύνολο στην κατανόηση της τεχνητής νοημοσύνης και της αξίας των δεδομένων. Η ενημέρωση οφείλει να εστιάζει στις επιπτώσεις των συστημάτων στα ανθρώπινα δικαιώματα, την πρόσβαση σε αυτά, καθώς και στην προστασία του περιβάλλοντος και των οικοσυστημάτων.

- Πολυμερής και προσαρμοστική διακυβέρνηση και συνεργασία

Τα κράτη μέλη μπορούν να θεσπίζουν την κατάλληλη νομοθεσία για τα δεδομένα που παράγονται ή διέρχονται από το έδαφός τους, στο πλαίσιο της τήρησης του διεθνούς δικαίου και της εθνικής κυριαρχίας. Η εξασφάλιση της διαφορετικότητας και συμπεριληπτικότητας στη διακυβέρνηση της τεχνητής νοημοσύνης είναι επιτακτική για την ισότιμη κατανομή των ωφελειών και την προώθηση της βιώσιμης ανάπτυξης. Στο πλαίσιο αυτό, προτείνεται να υιοθετηθούν ανοικτά πρότυπα, να προωθηθεί η διαλειτουργικότητα για τη διευκόλυνση της συνεργασίας, να λαμβάνονται υπόψη η εξέλιξη της τεχνολογίας, η εμφάνιση νέων ομάδων ενδιαφερομένων και να επιτραπεί η ουσιαστική συμμετοχή περιθωριοποιημένων ομάδων, κοινοτήτων και ατόμων.

Συνολικά, η σύσταση της UNESCO αποτελεί ένα περιεκτικό και δυναμικό πλαίσιο που εμπεριέχει τέσσερις θεμελιώδεις αξίες και εννέα λειτουργικές αρχές, δημιουργώντας μια ολιστική προσέγγιση για τη διακυβέρνηση της τεχνητής νοημοσύνης που αποσκοπεί στην

καθοδήγηση των κοινωνιών προς μια υπεύθυνη διαχείριση τόσο των προβλέψιμων όσο και των απρόβλεπτων συνεπειών της.

4. Η τεχνητή νοημοσύνη στον τραπεζικό τομέα

Ο τραπεζικός τομέας αποτελεί κομβικό κομμάτι της παγκόσμιας οικονομίας, καθώς παρέχει μια σειρά από χρηματοοικονομικές υπηρεσίες, όπως η αποδοχή καταθέσεων, η παροχή δανείων, η διευκόλυνση χρηματοοικονομικών συναλλαγών και η διαχείριση περιουσιακών στοιχείων. Ταυτόχρονα, συγκαταλέγεται μεταξύ των πρώτων κλάδων που υιοθέτησαν την τεχνητή νοημοσύνη, με στόχο την αποκόμιση σημαντικών ωφελειών από την εφαρμογή της. Η τεχνητή νοημοσύνη προβλέπεται να μετασχηματίσει τις αλληλεπιδράσεις των πελατών με τα τραπεζικά ιδρύματα, οδηγώντας στη δημιουργία νέων προϊόντων και υπηρεσιών, στην ενίσχυση της ασφάλειας των συναλλαγών, στη βελτίωση της προστασίας των δεδομένων και στην πρόληψη της οικονομικής απάτης. Μέσω της τεχνητής νοημοσύνης, αναμένεται η παροχή οικονομικότερων υπηρεσιών και προϊόντων, η περαιτέρω βελτιστοποίηση των επιχειρησιακών λειτουργιών καθώς και η εξοικονόμηση πόρων (Ridzuan κ.ά., 2024).

4.1 Ψηφιακός μετασχηματισμός στον τραπεζικό τομέα

Η παραδοσιακή τραπεζική βασίζεται σε χειροκίνητες ή αναλογικές διαδικασίες για κρίσιμες επιχειρησιακές λειτουργίες όπως η αξιολόγηση πιστοληπτικής ικανότητας, η ανίχνευση απάτης, η διαχείριση κινδύνου και η εξυπηρέτηση πελατών. Το μεγαλύτερο μέρος αυτών των λειτουργιών λάμβανε χώρα στα τραπεζικά υποκαταστήματα και γραφεία, με την εξάρτηση στην ανθρώπινη αλληλεπίδραση και την φυσική τεκμηρίωση να περιορίζει την επιχειρησιακή αποτελεσματικότητα και να αποτελεί εμπόδιο στην επεκτασιμότητα των τραπεζικών υπηρεσιών (Galazona & Magomaeva, 2019; Munira, 2025).

Ο ψηφιακός μετασχηματισμός του τραπεζικού τομέα και η έναρξη της μετάβασης από τα παραδοσιακά τραπεζικά συστήματα ξεκίνησε με τα μηχανήματα ανάληψης μετρητών (Automated Teller Machines – ATMs) τη δεκαετία του 1960. Τα ATM έφεραν επανάσταση στις συναλλαγές των πελατών επιτρέποντας πρόσβαση στις τραπεζικές υπηρεσίες καθ' όλο το 24ωρο, μειώνοντας ταυτόχρονα την εξάρτηση από τα τραπεζικά υποκαταστήματα. Η ανάπτυξη μεγάλων υπολογιστικών συστημάτων (mainframes) διευκόλυνε την αυτοματοποίηση των υποστηρικτικών τραπεζικών λειτουργιών, ενισχύοντας τις δυνατότητες επεξεργασίας μεγάλου όγκου δεδομένων, ενώ στις αρχές της δεκαετίας του 2000 η ηλεκτρονική τραπεζική και η εξέλιξη του ίντερνετ επέτρεψε στους πελάτες να πραγματοποιούν συναλλαγές από τον προσωπικό τους χώρο. Στη διάρκεια της δεκαετίας του

2010 η εισαγωγή της τεχνολογίας blockchain παρείχε μια ασφαλή και διαφανή πλατφόρμα για ψηφιακές συναλλαγές. Παράλληλα, η ενσωμάτωση της τεχνητής νοημοσύνης και της μηχανικής μάθησης προς τα τέλη της δεκαετίας του 2010 έφερε επανάσταση στον τομέα, επιτρέποντας τη χρήση προγνωστικής ανάλυσης για την αξιολόγηση πιστοληπτικού κινδύνου και την ανίχνευση απάτης. Τέλος, η ανάπτυξη του υπολογιστικού νέφους (cloud computing), με τις δυνητικά απεριόριστες δυνατότητες κλιμάκωσης σε ισχύ και αποθήκευση, μείωσε τα κόστη υποδομής (Munira, 2025).

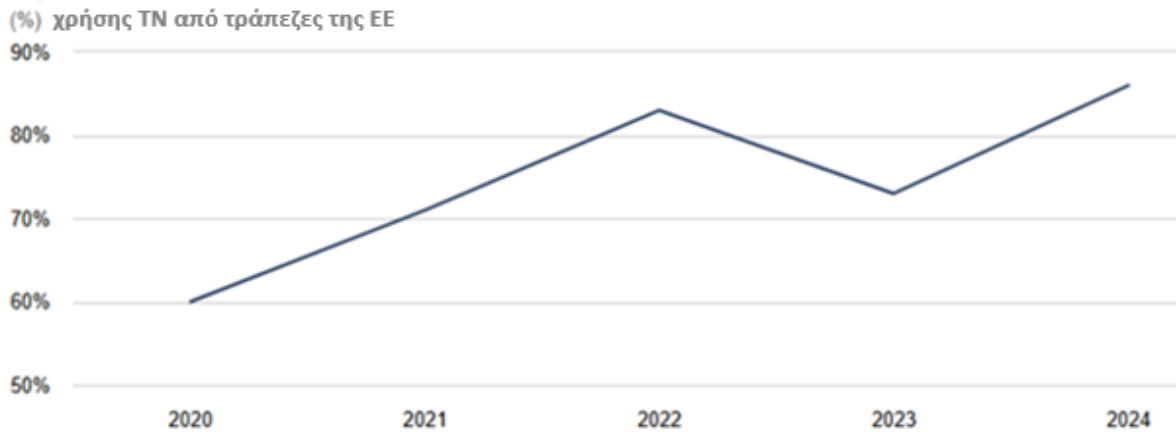
Ο ψηφιακός μετασχηματισμός αποτελεί τον ακρογωνιαίο λίθο της επιτυχίας για τα τραπεζικά ιδρύματα που επιδιώκουν να διατηρήσουν το ανταγωνιστικό τους πλεονέκτημα εν μέσω της διαρκούς εξέλιξης των προσδοκιών των πελατών και της ενίσχυσης των πιέσεων της αγοράς. Η τεχνολογία λειτουργεί ως παράγοντας ενεργοποίησης, επιτρέποντας στα τραπεζικά ιδρύματα να συνδυάζουν την πληροφορία με καινοτόμα προϊόντα και υπηρεσίες αποτελεσματικά. Ο τραπεζικός κλάδος έχει υιοθετήσει ένα ευρύ φάσμα τεχνολογιών, όπως η τεχνητή νοημοσύνη, η τεχνολογία blockchain και το υπολογιστικό νέφος με αποτέλεσμα την βελτιωμένη λειτουργική αποδοτικότητα, την αυξημένη ικανοποίηση των πελατών και τη διατήρηση της ανταγωνιστικότητας στην αγορά. Συνεπώς, οι σύγχρονες τραπεζικές υπηρεσίες παρέχονται αδιαλείπτως χωρίς την ανάγκη φυσικών υποκαταστημάτων ή διαδικασιών βασισμένων σε χαρτί (Lavanja & Rajkumar, 2024; Munira, 2025).

Στο πλαίσιο αυτό, η ενσωμάτωση αλλά και η καινοτομία στον τομέα της τεχνητής νοημοσύνης έχει αναδειχθεί ως ένας παράγοντας που θα αναδιαμορφώσει τη δυναμική του κλάδου και θα ενισχύσει την οικονομική απόδοση των τραπεζών. Οι δαπάνες του τραπεζικού τομέα αποκλειστικά για την παραγωγική τεχνητή νοημοσύνη αναμένεται να αυξηθούν από 3,86 δισεκατομμύρια δολάρια ΗΠΑ το 2023 σε σχεδόν 85 δισεκατομμύρια δολάρια ΗΠΑ το 2030. Οι χρήσεις της τεχνητής νοημοσύνης, οι οποίες εκτείνονται από την προγνωστική ανάλυση και τη διαχείριση κινδύνου έως την εξυπηρέτηση πελατών και την ανίχνευση απάτης, έχει εγκαινιάσει μια νέα εποχή ευκαιριών και ανταγωνισμού για τα τραπεζικά ιδρύματα παγκοσμίως (Baffour Gyau κ.ά., 2024; Crisanto κ.ά., 2024).

4.2 Χρήσεις της τεχνητής νοημοσύνης στον τραπεζικό τομέα

Η τεχνητή νοημοσύνη αναδεικνύεται ως μία μετασχηματιστική τεχνολογία για τον τραπεζικό τομέα, σηματοδοτώντας μια στρατηγική και συστημική στροφή προς την αυτοματοποίηση και τη βελτιστοποίηση των χρηματοπιστωτικών λειτουργιών. Τα τραπεζικά ιδρύματα

βασίζονται ολοένα και περισσότερο σε αυτή για την υποστήριξη λήψης τεκμηριωμένων αποφάσεων, την αξιολόγηση πιστωτικού κινδύνου και την ανίχνευση απάτης. Χαρακτηριστικά είναι τα στοιχεία της Ευρωπαϊκής Αρχής Τραπεζών στην εικόνα 3, που δείχνουν ότι το ποσοστό των τραπεζών από χώρες της Ευρωπαϊκής Ένωσης που χρησιμοποιούν τεχνολογίες τεχνητής νοημοσύνης έχει αυξηθεί σε 86% το 2024, όταν το 2020 ήταν στο 60% (European Banking Authority, 2024).



Χρονολογία	ΤΝ
2020	60
2021	71
2022	83
2023	73
2024	86

Εικόνα 3 Ποσοστό τραπεζών της ΕΕ που χρησιμοποιούν ΤΝ

Οι χρήσεις της τεχνητής νοημοσύνης στον τραπεζικό τομέα είναι πολυδιάστατες, συνεχώς αναπτυσσόμενες και σηματοδοτούν την απαρχή μιας νέας εποχής στην τραπεζική, η οποία χαρακτηρίζεται από την ενίσχυση της προστιθέμενης αξίας προς τους πελάτες και τη δημιουργία ενός μακροπρόθεσμου ανταγωνιστικού πλεονεκτήματος για τα τραπεζικά ιδρύματα (A. Kaur, 2025; Lawrence Damilare Oyeniyi κ.ά., 2024; Rashid & Kausik, 2024; Ridzuan κ.ά., 2024).

Στη συνέχεια του υποκεφαλαίου παρουσιάζονται κάποιες από τις πλέον συχνές χρήσεις της τεχνητής νοημοσύνης στον τραπεζικό τομέα, χωρίς αυτό να αποκλείει την ύπαρξη επιπρόσθετων εφαρμογών ή τη μη υιοθέτηση των αναφερόμενων πρακτικών από όλα τα τραπεζικά ιδρύματα.

4.2.1 Εξυπηρέτηση πελατών και εξατομίκευση υπηρεσιών

Στο πλαίσιο των χρήσεων της τεχνητής νοημοσύνης στον τραπεζικό τομέα, η εξυπηρέτηση πελατών και εξατομίκευση των υπηρεσιών ξεχωρίζει ως μία από τις πιο σημαντικές και επιδραστικές επιχειρησιακές λειτουργίες.

Η εισαγωγή εικονικών βοηθών με ενσωματωμένη τεχνητή νοημοσύνη συνιστά καίριο ορόσημο στην εξυπηρέτηση των πελατών του τραπεζικού τομέα, καθιστώντας τους αναπόσπαστο τμήμα των σύγχρονων τραπεζικών λειτουργιών και συμβάλλοντας στη βελτίωση της επιχειρησιακής αποδοτικότητας και εξατομίκευσης των υπηρεσιών. Οι συγκεκριμένοι ευφυείς βοηθοί μπορούν να απαντούν σε ερωτήματα πελατών, να παρέχουν στατιστικά και συναλλακτικά δεδομένα, να υποστηρίζουν σύνθετες λειτουργικές ανάγκες και να προσφέρουν εξατομικευμένη χρηματοοικονομική καθοδήγηση. Αποτελούν καθοριστικό παράγοντα για την αναβάθμιση της εμπειρίας του πελάτη και τη βελτιστοποίηση της τραπεζικής εξυπηρέτησης, προσφέροντας αδιάλειπτη διαθεσιμότητα και εξασφαλίζοντας την ταυτόχρονη διαχείριση πολυάριθμων αλληλεπιδράσεων. Συνολικά, μέσω της προσαρμοστικής λειτουργίας και της λήψης αποφάσεων βάσει δεδομένων, οι εικονικοί βοηθοί ενισχύουν την ανάπτυξη ευέλικτων, κλιμακώσιμων και πελατοκεντρικών τραπεζικών υπηρεσιών (Naik, 2024; Ridzuan κ.ά., 2024; World Economic Forum, 2025).

Μια επιπλέον εφαρμογή της τεχνητής νοημοσύνης στην εξυπηρέτηση πελατών εντοπίζεται στη συλλογή και ανάλυση μεγάλου όγκου δεδομένων για σκοπούς μάρκετινγκ. Μέσω της ανάλυσης δεδομένων που αφορούν τις καταναλωτικές προτιμήσεις, την οικονομική κατάσταση, τα πρότυπα συμπεριφοράς και άλλες μεταβλητές, τα τραπεζικά ιδρύματα μπορούν να δημιουργούν με ακρίβεια το προφίλ ενός πελάτη, βάσει του οποίου σχεδιάζονται και προωθούνται προσαρμοσμένα προϊόντα. Οι επαγγελματίες του μάρκετινγκ, αξιοποιώντας τις δυνατότητες της τεχνητής νοημοσύνης, τμηματοποιούν την αγορά, προβλέπουν τις προτιμήσεις των πελατών, προβάλλουν στοχευμένα ψηφιακές διαφημίσεις σε πραγματικό χρόνο και παρέχουν βέλτιστες στρατηγικές προώθησης και τιμολόγησης προς τα τραπεζικά ιδρύματα. Οι εξατομικευμένες τραπεζικές υπηρεσίες που βασίζονται στην

τεχνητή νοημοσύνη ενισχύουν τις σχέσεις μεταξύ των τραπεζικών ιδρυμάτων και των πελατών τους (Munira, 2025; Ridzuan κ.ά., 2024).

4.2.2 Πρόληψη απάτης

Η ενσωμάτωση της τεχνητής νοημοσύνης και της μηχανικής μάθησης στον τραπεζικό τομέα έχει επιφέρει σημαντικές αλλαγές στην ενίσχυση της ασφάλειας και της αποτελεσματικότητας των συστημάτων ανίχνευσης και πρόληψης απάτης. Η αυξανόμενη πολυπλοκότητα των συναλλακτικών διαδικασιών, σε συνδυασμό με τη ραγδαία εξέλιξη της ψηφιακής τεχνολογίας, έχει καταστήσει τα παραδοσιακά συστήματα ανεπαρκή για την αντιμετώπιση των σύγχρονων μορφών χρηματοοικονομικής απάτης. Η τεχνητή νοημοσύνη, καθιστά δυνατή την ανάλυση εκτεταμένων συνόλων δεδομένων σε πραγματικό χρόνο, επιτρέποντας τον έγκαιρο εντοπισμό ανωμαλιών και συναλλακτικών προτύπων που αποκλίνουν από τη συνήθη συμπεριφορά (Crisanto κ.ά., 2024; Munira, 2025; Naik, 2024; Rashid & Kausik, 2024).

Παρά το γεγονός ότι η αξιοποίηση της τεχνητής νοημοσύνης για την πρόληψη της απάτης δεν αποτελεί καινοφανή πρακτική, η τρέχουσα περίοδος χαρακτηρίζεται από τη ραγδαία διεύρυνση των εφαρμογών της και τη συνεχή βελτίωση της ακρίβειας και αποδοτικότητάς της (Crisanto κ.ά., 2024). Τα σύγχρονα μοντέλα έχουν τη δυνατότητα ανίχνευσης και ανάλυσης περίπλοκων συναλλακτικών μοτίβων, εντοπίζοντας ύποπτες ενέργειες όπως η ταχεία μετακίνηση χρηματικών ποσών μεταξύ λογαριασμών ή συναλλαγές που αποκλίνουν αισθητά από τα στατιστικά αναμενόμενα όρια. Επίσης, τα μοντέλα αυτά συμβάλλουν στον εντοπισμό ύποπτων προσώπων, λογαριασμών αχυρανθρώπων και οργανωμένων δικτύων που επιδιώκουν να παρακάμψουν τους ελέγχους ασφαλείας των τραπεζικών συστημάτων. Στο πλαίσιο της καταπολέμησης νομιμοποίησης εσόδων από παράνομες δραστηριότητες και της χρηματοδότησης της τρομοκρατίας, οι εφαρμογές της τεχνητής νοημοσύνης αποκτούν ακόμη μεγαλύτερη σημασία. Τα τραπεζικά ιδρύματα μπορούν να αξιολογούν εάν θα εγκρίνουν ή θα απορρίψουν πληρωμές που φαίνονται ακανόνιστες, ενισχύοντας την πρόληψη εγκληματικών δραστηριοτήτων (Crisanto κ.ά., 2024; Munira, 2025; Naik, 2024; Rashid & Kausik, 2024).

Ένα ιδιαίτερα σημαντικό χαρακτηριστικό των συστημάτων τεχνητής νοημοσύνης είναι η ικανότητά τους να προσαρμόζονται δυναμικά, βελτιώνοντας συνεχώς την απόδοσή τους μέσω της μάθησης από νέα δεδομένα και μεταβαλλόμενες συνθήκες. Παράλληλα, ο αριθμός

των ψευδώς θετικών ειδοποιήσεων μειώνεται αισθητά, επιτρέποντας στα χρηματοπιστωτικά ιδρύματα να επικεντρώνονται σε πραγματικά περιστατικά. Κατά συνέπεια, γίνεται δυνατή η καλύτερη αξιοποίηση των ανθρώπινων πόρων για την αύξηση της επιχειρησιακής αποδοτικότητας και τη συνολική βελτίωση της διαχείρισης κινδύνου (Crisanto κ.ά., 2024; Naik, 2024).

4.2.3 Πίστωση και αξιολόγηση κινδύνου

Ο δανεισμός κεφαλαίων αποτελεί μία από τις σημαντικότερες δραστηριότητες των τραπεζικών ιδρυμάτων. Στο πλαίσιο αυτό καλούνται να αξιολογήσουν και να διαχειριστούν τον πιστωτικό κίνδυνο, με τον ακριβή προσδιορισμό της πιστοληπτικής ικανότητας ενός ατόμου ή μιας επιχείρησης να είναι απαραίτητος.

Η χρήση μοντέλων τεχνητής νοημοσύνης έχει βελτιώσει την ακρίβεια και την αποδοτικότητα των διαδικασιών αξιολόγησης πιστοληπτικής ικανότητας, αφού λαμβάνουν υπόψη ένα ευρύτερο φάσμα παραμέτρων. Τα τραπεζικά ιδρύματα μπορούν να επεξεργάζονται σύνθετα σύνολα δεδομένων, τα οποία περιλαμβάνουν ιστορικά στοιχεία πιστωτικής συμπεριφοράς, οικονομικές εκθέσεις και τάσεις της αγοράς προκειμένου να εκτιμήσουν με ακρίβεια τον πιστωτικό κίνδυνο των δανειοληπτών. Τα συστήματα τεχνητής νοημοσύνης διευκολύνουν τη λήψη τεκμηριωμένων αποφάσεων, μειώνοντας ταυτόχρονα τις πιθανότητες οικονομικών απωλειών και αποτελώντας ένα αποτελεσματικό εργαλείο ελέγχου (Jain, 2023; Munira, 2025; Naik, 2024; Ridzuan κ.ά., 2024). Επιπλέον, οι αλγόριθμοι μηχανικής μάθησης μπορούν να υποστηρίξουν την εκτίμηση των δανειακών εξασφαλίσεων και τον υπολογισμό των επιτοκίων, επιτρέποντας τη δημιουργία και προώθηση εξατομικευμένων δανειακών προϊόντων (Crisanto κ.ά., 2024).

Μέσω αυτών των τεχνολογιών, τα τραπεζικά ιδρύματα προβλέπουν πιθανά προβλήματα εξυπηρέτησης δανείων, αξιολογούν την πιστοληπτική ικανότητα και βελτιστοποιούν τα επενδυτικά τους χαρτοφυλάκια. Η εφαρμογή τέτοιων καινοτομιών αυξάνει την ταχύτητα και την αποτελεσματικότητα των χρηματοοικονομικών προβλέψεων, συνεισφέροντας στη μείωση των κινδύνων και στην ενίσχυση της λειτουργικής αποδοτικότητας των τραπεζικών οργανισμών (World Economic Forum, 2025).

4.2.4 Ανάλυση αγοράς και διαχείριση επενδύσεων

Η ανάλυση των τάσεων της αγοράς αποτελεί μια σύνθετη και χρονοβόρα λειτουργία που απαιτεί ακρίβειά και προϋποθέτει την εξέταση μεγάλου όγκου ετερογενών δεδομένων. Τα μοντέλα τεχνητής νοημοσύνης υπερτερούν στην επεξεργασία μεγάλου όγκου δεδομένων, που συμπεριλαμβάνουν αλλά δεν περιορίζονται, στις τάσεις της αγοράς, διάφορους οικονομικούς δείκτες, αξιολογήσεις και ειδησεογραφικά άρθρα. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να ανακαλύψουν τάσεις, συσχετίσεις και ανωμαλίες σε αυτά, παρέχοντας πληροφορίες που οι παραδοσιακές τεχνικές αξιολόγησης μπορεί να παραβλέψουν. Αξιοποιώντας τα μοντέλα τεχνητής νοημοσύνης, τα τραπεζικά ιδρύματα αποκτούν βαθύτερη γνώση της δυναμικής της αγοράς που τους επιτρέπει να ενημερώνουν τις επενδυτικές τους στρατηγικές (Naik, 2024).

Επιπλέον, η τεχνητή νοημοσύνη αποτελεί πλέον ουσιώδες εργαλείο στη διαχείριση επενδύσεων και την παροχή χρηματοοικονομικών συμβουλών. Τα τραπεζικά ιδρύματα αξιοποιούν προηγμένα μοντέλα για τον σχεδιασμό επενδυτικών χαρτοφυλακίων, την παροχή εξειδικευμένης χρηματοοικονομικής καθοδήγησης και τη διασφάλιση πρόσβασης των πελατών σε πληροφορίες και συστάσεις συναλλαγών σε πραγματικό χρόνο (World Economic Forum, 2025). Στο πλαίσιο αυτό, τα τραπεζικά ιδρύματα έχουν δημιουργήσει τους ψηφιακούς συμβούλους επενδύσεων (robo-advisors), οι οποίοι αξιοποιούν την τεχνητή νοημοσύνη για την παροχή εξατομικευμένων επενδυτικών προτάσεων, προσαρμοσμένων στο προφίλ κάθε επενδυτή. Η έννοια του συμβούλου επενδύσεων περιλαμβάνει την εφαρμογή θεωρητικών μοντέλων βελτιστοποίησης χαρτοφυλακίου και ευφύων υπολογιστικών τεχνικών για την χρηματοοικονομική καθοδήγηση βασισμένη σε δεδομένα. Οι σύμβουλοι λαμβάνουν υπόψη παραμέτρους όπως η ανοχή στον κίνδυνο, οι χρηματοοικονομικές ανάγκες και το επενδυτικό ύψος του κάθε επενδυτή, ενώ παρακολουθούν και προσαρμόζονται διαρκώς στις διακυμάνσεις των αγορών μέσω δυναμικής κατανομής στοιχείων ενεργητικού. Οι προηγμένοι σύμβουλοι επενδύσεων επιτρέπουν στις τράπεζες να προσφέρουν αποδοτικές και κλιμακούμενες υπηρεσίες διαχείρισης πλούτου, οι οποίες χαρακτηρίζονται από αυξημένη αποτελεσματικότητα στη λήψη επενδυτικών αποφάσεων και στην εκτέλεση συναλλαγών, συμβάλλοντας ταυτόχρονα στον περιορισμό των συναισθηματικών προκαταλήψεων που τείνουν να επηρεάζουν τους ανθρώπους (Jain, 2023; Naik, 2024; Ridzuan κ.ά., 2024).

4.2.5 Ανάπτυξη και ασφάλεια πληροφοριακών συστημάτων

Τα τραπεζικά ιδρύματα αξιοποιούν σε ευρεία κλίμακα την τεχνητή νοημοσύνη για την ανάπτυξη τεχνολογικών λύσεων, μετασχηματίζοντας τόσο τις εσωτερικές τους λειτουργίες όσο και τις υπηρεσίες προς τους πελάτες. Η υιοθέτηση της τεχνητής νοημοσύνης παρέχει στα τραπεζικά ιδρύματα τη δυνατότητα αυτοματοποίησης και βελτιστοποίησης ενός μεγάλου φάσματος διαδικασιών σε ολόκληρο τον κύκλο ζωής ανάπτυξης πληροφοριακών συστημάτων, συμπεριλαμβανομένης της ανάπτυξης κώδικα, της αυτοματοποίησης των διαδικασιών δοκιμών και της κατανόησης και παύσης λειτουργίας των παλαιότερων συστημάτων. Παράλληλα, διευκολύνει τη δημιουργία καινοτόμων ψηφιακών προϊόντων που ενισχύουν την επιχειρησιακή αποδοτικότητα και την ανταγωνιστική τους θέση (European Banking Authority, 2024; Giovine κ.ά., 2024; World Economic Forum, 2025).

Οι προηγμένες τεχνολογικές δυνατότητες της τεχνητής νοημοσύνης ενισχύουν ουσιαστικά την αντιμετώπιση των κυβερνοαπειλών, καθιστώντας την έναν από τους βασικούς πυλώνες σύγχρονων στρατηγικών ασφάλειας των πληροφοριακών συστημάτων. Η αξιοποίηση της για την ανίχνευση επερχόμενων απειλών μέσω ανάλυσης μεγάλων δεδομένων σε πραγματικό χρόνο, επιτρέπει την έγκαιρη αποτροπή επιθέσεων και τη σημαντική μείωση των δυνητικών επιπτώσεων τους. Επιπλέον, η αυτοματοποίηση της διαχείρισης περιστατικών απελευθερώνει κρίσιμους ανθρώπινους πόρους, δίνοντας τη δυνατότητα εστίασης σε σύνθετες απειλές με μεγαλύτερη αποτελεσματικότητα. Τέλος, η δυναμική της συνεχούς μάθησης και προσαρμογής που χαρακτηρίζει τα συστήματα τεχνητής νοημοσύνης εξασφαλίζει ότι τα μέτρα προστασίας εξελίσσονται παράλληλα με το μεταβαλλόμενο τοπίο απειλών, αξιοποιώντας την ανατροφοδοτούμενη μάθηση για την ενίσχυση της μακροπρόθεσμης αποτελεσματικότητάς τους (Chlouverakis, 2024; Google, 2025).

4.3 Οφέλη από την υιοθέτηση της τεχνητής νοημοσύνης για τον τραπεζικό τομέα

Τα τραπεζικά ιδρύματα επενδύουν ολοένα και μεγαλύτερους πόρους στην υιοθέτηση και λειτουργική ενσωμάτωση των συστημάτων τεχνητής νοημοσύνης, επιδιώκοντας την ενίσχυση της αποδοτικότητας και της ανταγωνιστικότητάς τους. Σύμφωνα με στοιχεία της Statista, οι δαπάνες του παγκόσμιου τραπεζικού κλάδου που αφορούν αποκλειστικά την παραγωγική τεχνητή νοημοσύνη εκτιμάται ότι θα αυξηθούν από 3,86 δισεκατομμύρια

δολάρια το 2023 σε περίπου 85 δισεκατομμύρια δολάρια το 2030. Σημαντικό μέρος των εν λόγω επενδύσεων κατευθύνεται αφενός προς την ανάπτυξη εξειδικευμένου ανθρώπινου δυναμικού, ικανού να διαμορφώσει και να εξελίξει τις σχετικές τεχνολογίες, και αφετέρου προς την ενίσχυση της απαιτούμενης υποδομής πληροφορικής. Η έντονη επενδυτική δραστηριότητα υποδηλώνει ότι τα τραπεζικά ιδρύματα προσδοκούν ουσιαστικές αποδόσεις από την ενσωμάτωση της τεχνητής νοημοσύνης, γεγονός που επαναπροσδιορίζει τις βασικές τραπεζικές λειτουργίες, βελτιώνοντας την επιχειρησιακή αποτελεσματικότητα, ενισχύοντας την εμπειρία των πελατών και προάγοντας την χρηματοοικονομική καινοτομία (Crisanto κ.ά., 2024). Η καινοτομία στον τομέα της τεχνητής νοημοσύνης ασκεί θετική επίδραση στην απόδοση των περιουσιακών στοιχείων των τραπεζικών ιδρυμάτων, επιβεβαιώνοντας ότι αποτελεί κρίσιμο παράγοντα βελτίωσης των χρηματοοικονομικών επιδόσεών τους (Baffour Gyau κ.ά., 2024).

4.3.1 Λειτουργική και οικονομική αποδοτικότητα

Η τεχνητή νοημοσύνη μπορεί να βελτιώσει σημαντικά την παραγωγικότητα στον τραπεζικό τομέα, επιταχύνοντας και αυτοματοποιώντας διαδικασίες, μειώνοντας την ανάγκη για ανθρώπινο δυναμικό και, κατά συνέπεια, περιορίζοντας τα ανθρώπινα λάθη. Η αυτοματοποίηση κρίσιμων εργασιών, όπως η ανίχνευση απάτης, η εκτίμηση κινδύνου και η ανάλυση δεδομένων, εκτός από τη βελτίωση της παραγωγικότητας, συμβάλει και στη μείωση του κόστους λειτουργίας (European Central Bank, 2024; Naik, 2024; Ridzuan κ.ά., 2024).

Η βιβλιογραφία επιβεβαιώνει συστηματικά ότι πρωταρχικός στόχος της υιοθέτησης της τεχνητής νοημοσύνης από τα τραπεζικά ιδρύματα είναι η επίτευξη λειτουργικής και οικονομικής αποδοτικότητας, καθώς οι πρωτοβουλίες μείωσης του κόστους συνήθως αποφέρουν ταχύτερα και πιο μετρήσιμα αποτελέσματα (World Economic Forum, 2025). Αντανακλώντας αυτή την εστίαση, τα μεγάλα τραπεζικά ιδρύματα επεκτείνουν το ανθρώπινο δυναμικό τους στον τομέα της τεχνητής νοημοσύνης, ενώ μειώνουν το προσωπικό σε άλλους τομείς, προβλέποντας ότι η αύξηση της παραγωγικότητας που επιτυγχάνεται χάρη στην τεχνητή νοημοσύνη θα αντισταθμίσει τη μείωση του ανθρώπινου δυναμικού (Crisanto κ.ά., 2024).

Αξίζει να τονιστεί ότι η τεχνητή νοημοσύνη υπερβαίνει τον απλό ρόλο της αυτοματοποίησης διαδικασιών. Τα τραπεζικά ιδρύματα που αξιοποιούν την τεχνητή νοημοσύνη την αντιμετωπίζουν ως μια μετασχηματιστική τεχνολογία για την υλοποίηση βασικών

στρατηγικών προτεραιοτήτων, όπως η αύξηση των εσόδων, η απόκτηση ανταγωνιστικού πλεονεκτήματος και η αύξηση της ικανοποίησης των πελατών και των εργαζομένων. Τα κορυφαία τραπεζικά ιδρύματα οφείλουν να διασφαλίσουν ότι οι πρωτοβουλίες της τεχνητής νοημοσύνης καθοδηγούνται από την επιχειρησιακή στρατηγική τους και όχι από την τεχνολογική εξέλιξη. Αυτό προϋποθέτει ότι τα ανώτατα επιχειρησιακά στελέχη αναλαμβάνουν από κοινού με τα υψηλόβαθμα τεχνολογικά στελέχη την ευθύνη διαμόρφωσης των αναγκαίων παρεμβάσεων, διασφαλίζοντας ότι τα παραδοτέα ευθυγραμμίζονται άμεσα με τις στρατηγικές ανάγκες της επιχείρησης (Giovine κ.ά., 2024).

Η McKinsey σε έρευνά της τονίζει ότι η παραγωγική τεχνητή νοημοσύνη θα μπορούσε να αυτοματοποιήσει έως και το 70% των επιχειρησιακών δραστηριοτήτων σε σχεδόν όλα τα επαγγέλματα έως το 2030, προσθέτοντας τρισεκατομμύρια δολάρια σε αξία στην παγκόσμια οικονομία (Durth κ.ά., 2023). Τα στοιχεία αυτά υπογραμμίζουν τον κρίσιμο ρόλο της τεχνητής νοημοσύνης στην ενίσχυση της αποδοτικότητας και τη μείωση του κόστους στον τραπεζικό τομέα, τοποθετώντας την ως μια τεχνολογία ζωτικής σημασίας για την αναδιαμόρφωση των λειτουργιών του κλάδου και τη δημιουργία προστιθέμενης αξίας.

4.3.2 Λήψη αποφάσεων

Οι αλγόριθμοι τεχνητής νοημοσύνης είναι σε θέση να επεξεργάζονται τεράστιους όγκους δεδομένων, παράγοντας αξιοποιήσιμα αποτελέσματα. Οι αλγόριθμοι διευκολύνουν την πιο ακριβή εκτίμηση πιστωτικών ή άλλων κινδύνων, την εφαρμογή ισχυρών μέτρων πρόληψης της απάτης και τη βελτιστοποίηση των επενδυτικών στρατηγικών. Η ικανότητα της τεχνητής νοημοσύνης να μετατρέπει μεγάλα σύνολα δεδομένων σε εξατομικευμένες πληροφορίες δίνει τη δυνατότητα στους υπεύθυνους λήψης αποφάσεων να βελτιώσουν τις επενδυτικές επιλογές, να εξειδικεύσουν τις αξιολογήσεις κινδύνου και να βελτιώσουν τις καθημερινές επιχειρησιακές αποφάσεις (Ridzuan κ.ά., 2024; World Economic Forum, 2025).

Αναλύοντας τόσο ιστορικά όσο και τρέχοντα δεδομένα, η τεχνητή νοημοσύνη μπορεί να συμβάλλει στη δημιουργία προβλέψεων σχετικά με τις ροές εσόδων, τις βέλτιστες στρατηγικές τιμολόγησης και το στρατηγικό χρονοδιάγραμμα για την απόκτηση ή την πώληση μετοχών. Αυτή η ταχεία σύνθεση πληροφοριών σε αξιόπιστες προβλέψεις μπορεί να αποκαλύψει μοτίβα που ενδέχεται να διαφύγουν από τους ανθρώπινους αναλυτές, διευκολύνοντας έτσι λεπτομερείς συγκρίσεις και υπογραμμίζοντας υποσχόμενες ευκαιρίες ανάπτυξης (Ridzuan κ.ά., 2024).

4.3.3 Εμπειρία πελάτη

Οι καινοτομίες που βασίζονται στην τεχνητή νοημοσύνη συμβάλλουν στη βελτίωση της εμπειρίας του πελάτη μέσω της παροχής στοχευμένων υπηρεσιών σε πραγματικό χρόνο, ενισχύοντας με αυτόν τον τρόπο τη διατήρηση της πελατειακής βάσης (Munira, 2025; Ridzuan κ.ά., 2024). Η τεχνητή νοημοσύνη αξιοποιεί την ανάλυση μεγάλων όγκων δεδομένων πελατών με σκοπό τον εντοπισμό των καταναλωτικών συνηθειών και προτιμήσεών τους, επιτρέποντας έτσι τη δημιουργία και προώθηση εξατομικευμένων χρηματοοικονομικών προϊόντων, την παροχή κατευθύνσεων για επενδύσεις, καθώς και την εφαρμογή προληπτικών μέτρων διαχείρισης κινδύνου. Οι εξατομικευμένες αυτές αλληλεπιδράσεις ενισχύουν την αφοσίωση και την ικανοποίηση των πελατών, ξεδιπλώνοντας ταυτόχρονα νέες εμπορικές ευκαιρίες (Ridzuan κ.ά., 2024).

Επιπλέον, η τεχνητή νοημοσύνη βελτιώνει τις διαδικασίες πιστοληπτικής αξιολόγησης και έγκρισης δανείων, αφού τα μοντέλα πιστωτικής βαθμολόγησης εξετάζουν ένα ευρύτερο φάσμα πληροφοριών σε σύγκριση με τις παραδοσιακές μεθόδους, οδηγώντας σε πιο δίκαιες και χωρίς αποκλεισμούς διαδικασίες χρηματοδότησης. Το χαρακτηριστικό αυτό ενισχύει τη χρηματοοικονομική συμπερίληψη, αφού διευκολύνει την πρόσβαση σε πιστώσεις για πελάτες που στο παρελθόν ενδεχομένως να ήταν αποκλεισμένοι (Naik, 2024).

Είναι χαρακτηριστικό, ότι το 70% των στελεχών στον χρηματοπιστωτικό τομέα εκτιμά ότι η τεχνητή νοημοσύνη θα συσχετιστεί άμεσα με την αύξηση των εσόδων τα επόμενα χρόνια, μέσω της βελτίωσης της εμπειρίας των πελατών, της παροχής πιο εξατομικευμένων και συναφών προϊόντων και υπηρεσιών, της υποστήριξης καινοτόμων νέων προσφορών, της ενίσχυσης των διασταυρούμενων και συμπληρωματικών πωλήσεων και της ενδυνάμωσης της ασφάλειας έναντι πιθανών απειλών (World Economic Forum, 2025).

4.4 Προκλήσεις και περιορισμοί από την υιοθέτηση της τεχνητής νοημοσύνης για τον τραπεζικό τομέα

Η τεχνητή νοημοσύνη μπορεί να παρουσιάσει σημαντικές προκλήσεις και περιορισμούς, ειδικότερα όσο η υιοθέτηση της γίνεται πιο διαδεδομένη. Παρά τις σημαντικές ωφέλειες που αναμένονται, είναι αναγκαίο να αναγνωριστούν οι εγγενείς περιορισμοί και προκλήσεις που πηγάζουν από αυτήν για τον τραπεζικό τομέα.

Τα τραπεζικά ιδρύματα συχνά στερούνται επαρκών και αξιόπιστων δεδομένων, περιορίζοντας την ικανότητά τους να εκπαιδεύουν και να αξιοποιούν αποτελεσματικά μοντέλα τεχνητής νοημοσύνης. Για την αντιμετώπιση του συγκεκριμένου περιορισμού, δύναται να στραφούν σε εξωτερικές πηγές δεδομένων ή σε συνεργασία με τρίτους φορείς. Ωστόσο, τέτοιες πρακτικές εγείρουν σημαντικά ζητήματα απορρήτου και ασφάλειας, ιδίως όσον αφορά τη διαχείριση και την προστασία ευαίσθητων δεδομένων πελατών (Dhashanamoorthi, 2023; European Banking Authority, 2024).

Οι μεγάλες τεχνολογικές επιχειρήσεις, οι οποίες φαντάζουν ως οι ιδανικοί συνεργάτες, κατέχουν δεσπόζουσα θέση στον τομέα της τεχνητής νοημοσύνης, επηρεάζοντας καθοριστικά τόσο την πορεία της έρευνας όσο και την ανάπτυξη των σχετικών εφαρμογών. Η έντονη συγκέντρωση της παροχής των συγκεκριμένων υπηρεσιών σε περιορισμένο αριθμό παρόχων εγείρει ανησυχίες τόσο για τον τραπεζικό τομέα όσο και για τις ρυθμιστικές αρχές, καθώς η τάση αυτή ενδέχεται να εκθέσει τα τραπεζικά ιδρύματα σε αυξημένους κινδύνους συγκέντρωσης και εξάρτησης (vendor lock-in). Φορείς και οργανισμοί, συμπεριλαμβανομένης της Ευρωπαϊκής Κεντρικής Τράπεζας, έχουν επισημάνει την αυξανόμενη εξάρτηση των τραπεζών από τρίτους παρόχους τεχνολογικών λύσεων, υπογραμμίζοντας την ανάγκη τακτικής αξιολόγησης του βαθμού εξάρτησης, ιδίως όσον αφορά τις υπηρεσίες τεχνητής νοημοσύνης. Η περαιτέρω τεχνολογική διεξόδωση και η αυξανόμενη συγκέντρωση των παρόχων δύναται να δημιουργήσει συστημικούς κινδύνους, αλλά και να εντείνει την έκθεση σε κυβερνοαπειλές (Crisanto κ.ά., 2024; World Economic Forum, 2025).

Στο μέτωπο των κυβερνοαπειλών, η χρήση εργαλείων τεχνητής νοημοσύνης ενισχύει την πολυπλοκότητα και την αποτελεσματικότητα των κυβερνοεπιθέσεων, αφού μπορούν να αξιοποιηθούν οι ευπάθειες των μοντέλων ή να παραχθεί ψεύτικο, εξαιρετικά ρεαλιστικό, περιεχόμενο (deepfake content). Οι επιθέσεις αυτού του τύπου παρουσιάζουν αυξημένη δυσκολία εντοπισμού, καθώς χαρακτηρίζονται από τη δυνατότητα προσαρμογής σε πραγματικό χρόνο, με περαιτέρω δυνατότητες αυτοματοποίησης και κλιμάκωσης (Crisanto κ.ά., 2024; World Economic Forum, 2025). Η ενσωμάτωση εφαρμογών τεχνητής νοημοσύνης στο τεχνολογικό οικοσύστημα των τραπεζικών ιδρυμάτων θα απαιτήσει την επένδυση πόρων για την ευθυγράμμιση με την υπάρχουσα τεχνολογική υποδομή. Τα τραπεζικά ιδρύματα ενδέχεται να χρησιμοποιούν παλαιά και πολύπλοκα υπολογιστικά συστήματα (legacy IT

systems), για τα οποία θα απαιτηθεί μερική ή πλήρης αναβάθμιση για την αποφυγή ασυμβατότητας και τον μετριασμό των απειλών ασφαλείας. (Dhashanamoorthi, 2023).

Παράλληλα, τα τραπεζικά ιδρύματα θα πρέπει να μετασχηματίσουν τις οργανωτικές τους δομές, τις διαδικασίες και την εταιρική τους κουλτούρα ώστε να υιοθετήσουν την τεχνητή νοημοσύνη και να ενισχύσουν την καινοτομία. Για την επίτευξη των σκοπών αυτών μπορεί να απαιτηθεί επένδυση για την ανάπτυξη των ικανοτήτων του ανθρώπινου δυναμικού, στη διαχείριση της αλλαγής και σε ευέλικτες μεθόδους εργασίας (Dhashanamoorthi, 2023). Η υιοθέτηση της τεχνητής νοημοσύνης προβλέπεται να οδηγήσει στη μείωση της αναγκαιότητας ορισμένων ρόλων ή ακόμη και στην πλήρη κατάργησή τους, δεδομένου ότι οι διάφορες ροές εργασίας αυτοματοποιούνται, δημιουργώντας ανησυχίες για την ανεργία και την οικονομική ανισότητα. Τα τραπεζικά ιδρύματα μπορεί να βρεθούν αντιμέτωπα με ένα δίλημμα, όπου θα πρέπει να ισορροπήσουν μεταξύ της κοινωνικής και οικονομικής αστάθειας και της στρατηγικής τους επιλογής για βελτιστοποίηση των λειτουργιών και των κερδών τους (Chlouverakis, 2024; Naik, 2024).

Η ενσωμάτωση της τεχνητής νοημοσύνης στον τραπεζικό τομέα προϋποθέτει αυστηρή συμμόρφωση με ένα σύνθετο πλαίσιο κανονιστικών προτύπων και κανόνων δεοντολογίας, στο οποίο δίνεται ιδιαίτερη έμφαση στην αμεροληψία, τη διαφάνεια, την εξηγησιμότητα και τη λογοδοσία. Το πλαίσιο αυτό αποσκοπεί στη διασφάλιση της συμμόρφωσης των εφαρμογών της τεχνητής νοημοσύνης τόσο με τις νομικές υποχρεώσεις, όσο και με τις προσδοκίες των πελατών, των ρυθμιστικών αρχών και της κοινωνίας (Dhashanamoorthi, 2023). Η βιώσιμη και υπεύθυνη ανάπτυξη της τεχνητής νοημοσύνης στον τραπεζικό τομέα βασίζεται στην ευθυγράμμισή της με τις ανθρωποκεντρικές και κοινωνικές αξίες, διασφαλίζοντας ότι η τεχνολογική πρόοδος υπηρετεί τη δικαιοσύνη, την εμπιστοσύνη και την ανθεκτικότητα του χρηματοοικονομικού οικοσυστήματος (A. Kaur, 2025).

Η αναγνώριση των περιορισμών της τεχνητής νοημοσύνης αποτελεί ένα κρίσιμο κομμάτι στο ευρύτερο πλαίσιο προκλήσεων που αντιμετωπίζουν τα τραπεζικά ιδρύματα. Η αξιοποίηση της τεχνητής νοημοσύνης συνεπάγεται τόσο την υιοθέτηση των σχετικών τεχνολογιών αλλά και την προσαρμογή στις ηθικές, νομικές και κοινωνικές διαστάσεις της χρήσης της. Το ενδιαφέρον των τραπεζικών ιδρυμάτων υπερβαίνει την απλή τεχνολογική εφαρμογή και εστιάζει στην προώθηση ενός οικοσυστήματος βασισμένου στην τεχνητή νοημοσύνη, το

οποίο χαρακτηρίζεται από ηθική, υπευθυνότητα, διαφάνεια και άρση των αποκλεισμών (Chlouverakis, 2024).

5. Ανάλυση ηθικών ζητημάτων από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα

Τα συστήματα τεχνητής νοημοσύνης εγείρουν νέες ηθικές προκλήσεις που αφορούν, μεταξύ άλλων, τη λήψη αποφάσεων, την απασχόληση, την προστασία των προσωπικών δεδομένων, την ιδιωτικότητα και την αποτροπή διακρίσεων. Η ικανότητά τους να επιτελούν εργασίες που άλλοτε θεωρούνταν αποκλειστικά ανθρώπινες, αλλά και η δυνατότητά τους να ξεπερνούν την απόδοση του ανθρώπου σε κάποιες από αυτές, αναδεικνύει τη δυνητική επίδρασή τους στις ανθρώπινες κοινωνίες (UNESCO, 2022).

Η ηθική της τεχνητής νοημοσύνης, ως κλάδος της εφαρμοσμένης ηθικής, εξετάζει τις ηθικές προεκτάσεις της ανάπτυξης, εφαρμογής και χρήσης των συστημάτων τεχνητής νοημοσύνης, επιδιώκοντας να προσδιορίσει τον τρόπο με τον οποίο αυτά μπορούν να προάγουν ή να υπονομεύσουν την ποιότητα ζωής, την ανθρώπινη αυτονομία και θεμελιώδεις αξίες μιας δημοκρατικής κοινωνίας (High-Level Expert Group on Artificial Intelligence, 2019). Η αυξανόμενη εξάρτηση από συστήματα λήψης αποφάσεων βασισμένα στην τεχνητή νοημοσύνη εντείνει την ανάγκη για ηθική και υπεύθυνη τεχνητή νοημοσύνη, με την ανθρώπινη εποπτεία να λειτουργεί ως κρίσιμη προϋπόθεση απέναντι στους κινδύνους της υπερβολικής αυτοματοποίησης. Αν και τα συστήματα αυτά μπορούν να βελτιώσουν την αποδοτικότητα σε σχέση με τις παραδοσιακές διαδικασίες, η απουσία επαρκούς ανθρώπινης παρέμβασης ενδέχεται να οδηγήσει σε ανεπιθύμητες συνέπειες, ιδίως σε πολύπλοκες ή υψηλού κινδύνου καταστάσεις, καθιστώντας απαραίτητη τη διαρκή ισορροπία μεταξύ τεχνολογικής αποτελεσματικότητας και ηθικής υπευθυνότητας (Lee κ.ά., 2025). Οι ηθικές προεκτάσεις αυτές αποτελούν σημαντικές προκλήσεις που πρέπει να αντιμετωπιστούν από τα τραπεζικά ιδρύματα προκειμένου να διατηρήσουν την εμπιστοσύνη των πελατών τους και να συμμορφωθούν με τα αυστηρά κανονιστικά πλαίσια του τομέα.

5.1 Ιδιωτικότητα και προστασία δεδομένων

Ο σεβασμός στην ιδιωτικότητα και η προστασία των δεδομένων εκτός από βασικές αρχές των πλαισίων ηθικής που αναλύθηκαν κατά την παρούσα εργασία, αναγνωρίζονται και ως βασικά δικαιώματα του ανθρώπου από διεθνείς οργανισμούς. Η Οικουμενική Διακήρυξη των Δικαιωμάτων του Ανθρώπου του ΟΗΕ, στο άρθρο 12, ορίζει ότι κανείς δεν πρέπει να

υφίσταται αυθαίρετη ή παράνομη επέμβαση στην ιδιωτική του ζωή (Universal Declaration of Human Rights, 1948), ενώ ο Χάρτης των Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης στα άρθρα 7 και 8 ορίζει τον σεβασμό στην προσωπική ζωή και την προστασία των δεδομένων (Chart of Fundamental Rights of the European Union, 2012). Η προστασία των δεδομένων, παρότι δεν αναφέρεται ρητά σε ορισμένες βασικές συνθήκες προστασίας ανθρωπίνων δικαιωμάτων, αναγνωρίζεται ως θεμελιώδης έκφανση του δικαιώματος στην ιδιωτικότητα (Danish Institute for Human Rights, 2025).

Τα δεδομένα συνιστούν το θεμελιώδες υπόβαθρο των συστημάτων τεχνητής νοημοσύνης, με πολλά από τα εν λόγω συστήματα να εκπαιδεύονται και να λειτουργούν βάσει δεδομένων προερχομένων από ανθρώπινες δραστηριότητες. Τα δεδομένα αυτά ενδέχεται να ενσωματώνουν και ευαίσθητες πληροφορίες, γεγονός που εγείρει εύλογες ανησυχίες σχετικά με την προστασία, την ασφάλεια και την διακυβέρνηση τους σε ολόκληρη τη διάρκεια ζωής των συστημάτων τεχνητής νοημοσύνης. Η απουσία των κατάλληλων μηχανισμών δύναται να υπονομεύσει την εμπιστοσύνη των ανθρώπων και να οδηγήσει σε μη συμμόρφωση με το ισχύον νομικό πλαίσιο για την προστασία των δεδομένων, όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) της Ευρωπαϊκής Ένωσης. Οι πολίτες θα πρέπει να διατηρούν τον πλήρη έλεγχο επί των προσωπικών τους δεδομένων, διασφαλίζοντας ότι αυτά δεν θα χρησιμοποιηθούν με τρόπο παράνομο ή άδικο, που ενδεχομένως τους προκαλέσει βλάβη (Díaz-Rodríguez κ.ά., 2023).

Η ηθική αναγκαιότητα της προστασίας των δεδομένων αποκτά ιδιαίτερη σημασία στον τραπεζικό τομέα, όπου η διαχείριση προσωπικών και οικονομικών πληροφοριών αποτελεί αναπόσπαστο μέρος των καθημερινών λειτουργιών. Τα τραπεζικά ιδρύματα συλλέγουν, αναλύουν και αποθηκεύουν ένα τεράστιο όγκο ευαίσθητων δεδομένων πελατών, όπως ιστορικά συναλλαγών, πιστωτικά προφίλ και οικονομικές καταστάσεις (Naik, 2024; Valavan, 2023). Η υπεύθυνη διαχείριση αυτών των πληροφοριών δεν αποτελεί μόνο νομική υποχρέωση, αλλά και θεμελιώδη ηθική δέσμευση, η οποία συνδέεται άμεσα με τη διαφύλαξη της εμπιστοσύνης και της αξιοπιστίας του χρηματοπιστωτικού συστήματος (Valavan, 2023).

Στο πλαίσιο της τραπεζικής λειτουργίας, τα συστήματα τεχνητής νοημοσύνης ενδέχεται να έχουν πρόσβαση σε ευαίσθητα προσωπικά, οικονομικά και εμπορικά δεδομένα (Rashid & Kausik, 2024). Οι κύριες προκλήσεις που ανακύπτουν σχετίζονται με την προστασία αυτών

των δεδομένων, τη διασφάλιση της συγκατάθεσης των χρηστών για τη χρήση τους, καθώς και τον κίνδυνο μη εξουσιοδοτημένης πρόσβασης ή διαρροής πληροφοριών (Naik, 2024; Ridzuan κ.ά., 2024). Η πρόσβαση σε τέτοιου είδους δεδομένα μπορεί να αποτελέσει ελκυστικό στόχο για τους κυβερνοεγκληματίες, είτε με σκοπό την πρόκληση οικονομικών ή λειτουργικών ζημιών στα τραπεζικά ιδρύματα και τους πελάτες τους είτε για ίδιο οικονομικό όφελος. Τα τραπεζικά ιδρύματα οφείλουν να διασφαλίζουν όχι μόνο την προστασία των δεδομένων, αλλά και την ασφάλεια των ίδιων των συστημάτων τεχνητής νοημοσύνης, προκειμένου να αποτρέψουν τυχόν παραβιάσεις ή λανθασμένα και δυνητικά επιζήμια αποτελέσματα (Rashid & Kausik, 2024).

Η αντιμετώπιση των προκλήσεων στο πλαίσιο της υιοθέτησης των συστημάτων τεχνητής νοημοσύνης απαιτεί μια πολυδιάστατη προσέγγιση. Τα τραπεζικά ιδρύματα οφείλουν να εφαρμόσουν ισχυρές μεθόδους κρυπτογράφησης που διασφαλίζουν την ακεραιότητα και την εμπιστευτικότητα των δεδομένων, τη θέσπιση αυστηρών μηχανισμών ελέγχου πρόσβασης και τη διενέργεια τακτικών ελέγχων για την αποτροπή μη εξουσιοδοτημένων ενεργειών. Τέλος, η συμμόρφωση με την εκάστοτε εθνική και διεθνή νομοθεσία περί προστασίας προσωπικών δεδομένων αποτελεί απαραίτητη προϋπόθεση για τη διασφάλιση της νομιμότητας και της εταιρικής υπευθυνότητας (Ridzuan κ.ά., 2024)

5.2 Μεροληψία των συστημάτων τεχνητής νοημοσύνης

Τα συστήματα τεχνητής νοημοσύνης οφείλουν να σχεδιάζονται και να αναπτύσσονται με τρόπο που να εξασφαλίζει την χωρίς αποκλεισμούς πρόσβαση, την ισότιμη μεταχείριση και την αποφυγή άδικων διακρίσεων εις βάρος ατόμων, κοινοτήτων ή κοινωνικών ομάδων. Μια τέτοια προσέγγιση ευθυγραμμίζεται με τις αρχές της υπεύθυνης και αξιόπιστης τεχνητής νοημοσύνης, καθώς προωθεί τον δίκαιο σχεδιασμό, την ανάπτυξη και τη χρήση των συστημάτων, περιορίζοντας τις προκαταλήψεις και ενισχύοντας την κοινωνική ένταξη και δικαιοσύνη. Ωστόσο, οι προκαταλήψεις μπορούν να αναδυθούν σε όλα τα στάδια του κύκλου ζωής των συστημάτων τεχνητής νοημοσύνης· από τη συλλογή και επεξεργασία των δεδομένων, μέχρι τη μοντελοποίηση, την εκπαίδευση και την παραγωγή των εξόδων, γεγονός που καθιστά αναγκαία τη συστηματική ενσωμάτωση κριτηρίων δικαιοσύνης σε κάθε στάδιο (Lee κ.ά., 2025). Τα συστήματα τεχνητής νοημοσύνης τείνουν να κληρονομούν και, σε ορισμένες περιπτώσεις, να μεγεθύνουν τις κοινωνικές προκαταλήψεις που είναι ενσωματωμένες στα σύνολα δεδομένων στα οποία εκπαιδεύονται, με αποτέλεσμα η

μεροληψία στα δεδομένα εκπαίδευσης να αποτελεί κρίσιμη πρόκληση για την ηθική και αξιόπιστη χρήση της τεχνητής νοημοσύνης. Η αντιμετώπιση των φαινομένων αυτών προϋποθέτει τη χρήση αλγορίθμων με επίγνωση της αξίας της δικαιοσύνης, την ανάπτυξη ποικίλων και αντιπροσωπευτικών συνόλων δεδομένων, καθώς και τη θέσπιση διαδικασιών συνεχούς παρακολούθησης, ελέγχου και αξιολόγησης των συστημάτων μετά την ανάπτυξή τους, ώστε να εντοπίζονται και να μετριάζονται εγκαίρως τυχόν διακρίσεις (A. Kaur, 2025).

5.2.1 Τύποι μεροληψίας

Το Αμερικανικό Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας ταξινομεί τη μεροληψία στα συστήματα τεχνητής νοημοσύνης σε τρεις βασικές κατηγορίες: τη στατιστική και υπολογιστική μεροληψία, τη συστημική μεροληψία και την ανθρώπινη μεροληψία. Η λίστα δεν είναι εξαντλητική, ωστόσο αντικατοπτρίζει σημαντικούς κινδύνους που πρέπει να λαμβάνονται υπόψη καθ' όλη τη διάρκεια ζωής των συστημάτων τεχνητής νοημοσύνης (Schwartz κ.ά., 2022).

Η στατιστική και υπολογιστική μεροληψία προκύπτει όταν το χρησιμοποιούμενο δείγμα δεδομένων δεν είναι αντιπροσωπευτικό του πληθυσμού αναφοράς. Στο πλαίσιο των συστημάτων τεχνητής νοημοσύνης αυτού του είδους η μεροληψία ενσωματώνεται στα σύνολα δεδομένων και στις αλγοριθμικές διαδικασίες που χρησιμοποιούνται κατά την ανάπτυξη τους και γίνονται ορατές όταν οι αλγόριθμοι εκπαιδεύονται σε δεδομένα ενός συγκεκριμένου τύπου και αδυνατούν να συνάγουν συμπεράσματα εκτός του πλαισίου αυτού. Τα σφάλματα μπορεί να οφείλονται στην ετερογένεια των δεδομένων, στη μαθηματική απλούστευση σύνθετων πληροφοριών, στη χρήση εσφαλμένων δεδομένων ή σε αλγοριθμικές μεροληψίες που σχετίζονται με φαινόμενα όπως η υπερπροσαρμογή ή η υπό προσαρμογή, η διαχείριση ακραίων τιμών, καθώς και οι διαδικασίες καθαρισμού ή συμπλήρωσης δεδομένων (Schwartz κ.ά., 2022).

Η συστημική μεροληψία προέρχεται από τον τρόπο λειτουργίας και δομής των οργανισμών και θεσμών που οδηγούν στην ευνοϊκή μεταχείριση ορισμένων κοινωνικών ομάδων και στην ταυτόχρονη περιθωριοποίηση άλλων. Το φαινόμενο αυτό δεν εδράζεται κατ' ανάγκην σε συνειδητές πρακτικές προκατάληψης ή διάκρισης, αντιθέτως, προκύπτει κυρίως από την αναπαραγωγή εδραιωμένων θεσμικών προτύπων, κανόνων και πρακτικών που συντηρούν τις υφιστάμενες κοινωνικές ανισότητες. Παράλληλα, νέες μορφές συστημικής μεροληψίας αναδύονται όταν οι υλικές και ψηφιακές υποδομές δεν σχεδιάζονται βάσει καθολικών αρχών

προσβασιμότητας, γεγονός που εξαιρεί ή παρεμποδίζει τη συμμετοχή κοινωνικών ομάδων που αντιμετωπίζουν μειονεκτήματα, όπως άτομα με αναπηρίες ή κάτοικοι απομακρυσμένων περιοχών (Schwartz κ.ά., 2022). Στο πλαίσιο της τεχνητής νοημοσύνης, τα σύνολα δεδομένων που χρησιμοποιούνται όσο και οι κανόνες, πρακτικές και διαδικασίες που διέπουν τον κύκλο ζωής των συστημάτων, ενδέχεται να αντανakλούν υπάρχουσες μορφές συστημικής μεροληψίας όπως ο ρατσισμός, οι κοινωνικοοικονομικές προκαταλήψεις και οι ακούσιες προτιμήσεις στα δεδομένα εισόδου (Cheng κ.ά., 2021).

Η ανθρώπινη μεροληψία αντικατοπτρίζει συστηματικά σφάλματα στη διαδικασία της ανθρώπινης σκέψης. Αυτού του είδους η μεροληψία είναι, συνήθως, μη συνειδητή και σχετίζεται με τον τρόπο με τον οποίο ένα άτομο ή μια ομάδα ατόμων αντιλαμβάνεται, ερμηνεύει και χρησιμοποιεί πληροφορίες προκειμένου να λάβει αποφάσεις. Στο πλαίσιο της τεχνητής νοημοσύνης η παρουσία της είναι καθολική, επηρεάζοντας ολόκληρο τον κύκλο ζωής των συστημάτων τεχνητής νοημοσύνης. Η απλή αύξηση της επίγνωσής ύπαρξης του συγκεκριμένου είδους μεροληψίας δεν αρκεί για την πλήρη αποτροπή ή τον έλεγχο της (Schwartz κ.ά., 2022). Για παράδειγμα η έλλειψη ποικιλομορφίας μεταξύ των ομάδων ανάπτυξης ενός συστήματος τεχνητής νοημοσύνης, ενδέχεται να συμβάλει στην ενσωμάτωσή μεροληψίας προερχόμενη από τα μέλη της συγκεκριμένης ομάδας.

5.2.2 Μεροληψία στα συστήματα τεχνητής νοημοσύνης του τραπεζικού τομέα

Μία από τις βασικές ωφέλειες από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα είναι η βελτιστοποίηση της διαδικασίας λήψης αποφάσεων. Ωστόσο, τα συστήματα τεχνητής νοημοσύνης ενδέχεται να ενσωματώνουν και τα τρία είδη μεροληψίας εγκυμονώντας τον κίνδυνο διακρίσεων εις βάρος συγκεκριμένων ομάδων, ιδίως όταν μεροληπτικά δεδομένα οδηγούν σε άδικες αποφάσεις δανειοδότησης και πιστωτικής βαθμολόγησης (Naik, 2024; Ridzuan κ.ά., 2024).

Στο πλαίσιο της δανειοδότησης και πιστωτικής βαθμολόγησης, η στατιστική και συστημική μεροληψία εδράζεται στα δεδομένα εκπαίδευσης των μοντέλων που μπορεί να ενσωματώνουν ανισοροπίες ή κοινωνικές προκαταλήψεις, όπως ιστορικές διακρίσεις σε δανειοδοτήσεις, δημογραφική ή φυλετική διάκριση κ.ά., με αποτέλεσμα μη αντιπροσωπευτικά σύνολα δεδομένων, διακρίσεις σε εγκρίσεις δανείων και πιστωτικές βαθμολογήσεις. Η υπολογιστική μεροληψία, από την άλλη, πηγάζει από τον σχεδιασμό, την επιλογή χαρακτηριστικών και τους στόχους βελτιστοποίησης των

αλγορίθμων, όπως η προτεραιότητα μείωσης κόστους ή μεγιστοποίησης εσόδων, με αποτέλεσμα άδικες εκτιμήσεις κινδύνου ή κόστους δανειοδότησης. Τέλος, η ανθρώπινη μεροληψία διαπερνά ολόκληρο τον κύκλο ζωής των συστημάτων, επηρεάζοντας τη συλλογή δεδομένων, την ανάπτυξη, την εκπαίδευση και συντήρηση τους (Valavan, 2023). Τα μεροληπτικά αποτελέσματα των μοντέλων της τεχνητής νοημοσύνης ενισχύουν περαιτέρω τις υπάρχουσες ανισότητες· για παράδειγμα, η συστηματική άρνηση δανείων σε συγκεκριμένες δημογραφικές ομάδες μπορεί να τις περιθωριοποιήσει οικονομικά, δημιουργώντας ένα συνεχή κύκλο διακρίσεων (Naik, 2024).

5.2.3 Τρόποι αντιμετώπισης

Τα τραπεζικά ιδρύματα οφείλουν να διασφαλίζουν την ίση μεταχείριση από τα συστήματα της τεχνητής νοημοσύνης. Για την αντιμετώπιση των προκλήσεων της μεροληψίας απαιτείται η αξιολόγηση των μοντέλων ως προς τις αρχές της δικαιοσύνης και ίσης μεταχείρισης, η χρήση αντιπροσωπευτικών συνόλων δεδομένων εκπαίδευσης και ο συνεχής έλεγχος για τον εντοπισμό και τον μετριασμό μεροληπτικών αποφάσεων (Faheem, 2024; Naik, 2024; Ridzuan κ.ά., 2024).

Η τακτική αξιολόγηση των αλγορίθμων τεχνητής νοημοσύνης ως προς τη μεροληψία και τον βαθμό δικαιοσύνης τους, είναι δυνατή μέσα από τη χρήση μετρικών δικαιοσύνης. Η επιλογή των κατάλληλων μετρικών εξαρτάται από το εκάστοτε πεδίο εφαρμογής, τα χαρακτηριστικά των δεδομένων και τους ειδικούς στόχους του συστήματος. Η συστηματική ενσωμάτωση των μετρικών δικαιοσύνης στο συνολικό πλαίσιο αξιολόγησης συνιστά βασική προϋπόθεση για τον έγκαιρο εντοπισμό και τον περιορισμό της μεροληψίας στα μοντέλα τεχνητής νοημοσύνης.

Επιπλέον, η χρήση αντιπροσωπευτικών συνόλων δεδομένων εκπαίδευσης είναι ζωτικής σημασίας για την αποφυγή της μεροληψίας. Ο εντοπισμός και μετριασμός της στα δεδομένα εκπαίδευσης, απαιτεί τον σχολαστικό έλεγχο και εμπλουτισμό των δεδομένων, καθώς και την εξασφάλιση αντιπροσωπευτικής εκπροσώπησης των ομάδων αναφοράς για την αποφυγή υπέρ ή υπό εκπροσώπηση συγκεκριμένων ομάδων.

Τέλος, τα συστήματα τεχνητής νοημοσύνης είναι απαραίτητο να παρακολουθούνται συνεχώς για τον εντοπισμό και τη διόρθωση διακρίσεων που μπορεί να αναδυθούν με την πάροδο του χρόνου. Οι πραγματικές συνθήκες λειτουργίας και οι αλληλεπιδράσεις με τους χρήστες επηρεάζουν την απόδοση των μοντέλων, καθιστώντας την τακτική παρακολούθηση βασικό

στοιχείο της υπεύθυνης ανάπτυξής τους. Η εγκαθίδρυση βρόγχων ανατροφοδότησης, με περιοδικές αξιολογήσεις, σχόλια χρηστών και εξωτερικούς ελέγχους, εξασφαλίζει την προσαρμογή των συστημάτων σε μεταβαλλόμενα περιβάλλοντα και τη διατήρηση της δικαιοσύνης καθ' όλη τη διάρκεια του λειτουργικού τους κύκλου (Akinrinola κ.ά., 2024).

5.3 Λογοδοσία και εποπτεία της τεχνητής νοημοσύνης

Η λογοδοσία και η εποπτεία των συστημάτων τεχνητής νοημοσύνης αποτελούν μια σημαντική πρόκληση που απαιτεί τη θέσπιση κατάλληλων μηχανισμών σε όλα τα στάδια του κύκλου ζωής τους. Η δημιουργία αυτών των μηχανισμών καθιστά τόσο τα συστήματα όσο και τους δημιουργούς τους υπεύθυνους και διασφαλίζει ότι όσοι επηρεάζονται έχουν το δικαίωμα προσφυγής σε περιπτώσεις σφαλμάτων, προκαταλήψεων ή δυσμενών συνεπειών (Valavan, 2023). Η λογοδοσία και εποπτεία δεν περιορίζεται στον καθορισμό ρόλων και ευθυνών των φορέων της τεχνητής νοημοσύνης, αλλά περιλαμβάνει τη διασφάλιση της ιχνηλασιμότητας για τα δεδομένα, τις διαδικασίες και τις αποφάσεις, ενώ εξασφαλίζει και την εφαρμογή πλαισίων διακυβέρνησης (Lee κ.ά., 2025).

Στον τραπεζικό τομέα, η απουσία τέτοιων μηχανισμών αναδεικνύεται ως κρίσιμο ζήτημα, ιδίως όταν τα συστήματα αυτά χρησιμοποιούνται για τη λήψη κρίσιμων αποφάσεων που επηρεάζουν τους πελάτες. Η πιθανή μεροληψία που μπορεί να ενσωματώνεται στα συστήματα τεχνητής νοημοσύνης λόγω της χρήσης μη αντιπροσωπευτικών συνόλων δεδομένων καθιστά τον προσδιορισμό της ευθύνης ακόμα πιο σύνθετο, εντείνοντας τις προκλήσεις γύρω από τη διαφάνεια και τη δικαιοσύνη των αποφάσεων (Marple κ.ά., 2023).

Τα τραπεζικά ιδρύματα οφείλουν να διασφαλίζουν ότι οι εφαρμογές τεχνητής νοημοσύνης δίνουν προτεραιότητα στην ευημερία των πελατών τους (Valavan, 2023). Για τη βελτίωση της λογοδοσίας και εποπτείας απαιτείται ο ορισμός των ρόλων και των ευθυνών των εμπλεκόμενων μερών, η θέσπιση νομικών και ηθικών προτύπων που ορίζουν την ευθύνη στο πλαίσιο της τεχνητής νοημοσύνης (Akinrinola κ.ά., 2024; Valavan, 2023) και η ενίσχυση της ανθρώπινης παρέμβασης, ώστε να περιοριστεί ο κίνδυνος λανθασμένων ή επιζήμιων αποφάσεων (Crisanto κ.ά., 2024). Στο πλαίσιο επαύξησης της ανθρώπινης παρέμβασης, δίνεται έμφαση σε προσεγγίσεις όπως το «human-in-the-loop» που εξασφαλίζει την ανθρώπινη εμπλοκή στον κύκλο λήψης αποφάσεων, το «human-on-the-loop» που εξασφαλίζει την ανθρώπινη εποπτεία κατά τον σχεδιασμό και τη συνεχή αξιολόγηση των

συστημάτων και το «human-in-control» το οποίο διασφαλίζει την υπεροχή της ανθρώπινης κρίσης για κρίσιμες αποφάσεις (Crisanto κ.ά., 2024).

Η αποτελεσματική εφαρμογή μηχανισμών διακυβέρνησης προϋποθέτει τεχνογνωσία σε όλα τα επίπεδα διοίκησης και προσωπικού των τραπεζικών ιδρυμάτων, καθώς τα στελέχη οφείλουν να κατανοούν και να διαχειρίζονται τους κινδύνους που απορρέουν από τα συστήματα τεχνητής νοημοσύνης (Crisanto κ.ά., 2024). Η εμπιστοσύνη σε αυτά τα συστήματα μπορεί να διαταραχθεί εύκολα, δεδομένου του εύρους χρήσης τους και της ταχύτητας λήψης αποφάσεων ή παραγωγής αποτελεσμάτων, γεγονός που ενισχύει την ανάγκη για διαφάνεια και εξηγησιμότητα στην εφαρμογή τους (Maple κ.ά., 2023).

5.4 Διαφάνεια και εξηγησιμότητα της τεχνητής νοημοσύνης

Ο τρόπος λειτουργίας των συστημάτων τεχνητής νοημοσύνης συχνά παραμένει αδιαφανής και δύσκολα εξηγήσιμος, γεγονός που έχει οδηγήσει πολλούς ερευνητές να τα παρομοιάζουν με «μαύρα κουτιά» (Akinrinola κ.ά., 2024; A. Kaur, 2025; Maple κ.ά., 2023; Naik, 2024). Η έλλειψη διαφάνειας δυσχεραίνει τον εντοπισμό και τη διόρθωση των διακρίσεων, περιορίζει τις δυνατότητες βελτίωσης των ίδιων των συστημάτων και υπονομεύει την εμπιστοσύνη των ανθρώπων προς την τεχνητή νοημοσύνη (Akinrinola κ.ά., 2024; Faheem, 2024; Maple κ.ά., 2023). Η διαφάνεια στα συστήματα τεχνητής νοημοσύνης μπορεί προσεγγιστεί ανάλογα με την οπτική του ενδιαφερόμενου μέρους. Για τους προγραμματιστές, στόχος είναι η κατανόηση του τρόπου λειτουργίας και των αιτιών πίσω από τις αποφάσεις των συστημάτων, για τους οργανισμούς που τα διαθέτουν, η διαφάνεια αποσκοπεί στην ενίσχυση της ασφάλειας και της εμπιστοσύνης των χρηστών, ενώ για τους τελικούς χρήστες η διαφάνεια συνδέεται με την κατανόηση του τι κάνει το σύστημα και γιατί. Τέλος, η διαφάνεια μπορεί να ταξινομηθεί με βάση το εύρος εστίασης, όπου η ολική διαφάνεια σκοπεύει στη συνολική ερμηνεία λειτουργίας του συστήματος, με την τοπική διαφάνεια να εστιάζει στην ερμηνεία εξόδων εντός συγκεκριμένου πλαισίου (Cheng κ.ά., 2021).

Στον τραπεζικό τομέα, η διαφάνεια αποτελεί θεμελιώδη παράγοντα για την ενίσχυση της εμπιστοσύνης των πελατών και τη διασφάλιση της κατανόησης των αποφάσεων που λαμβάνονται μέσω των συστημάτων τεχνητής νοημοσύνης. Είναι ουσιώδες τα άτομα που επηρεάζονται από αυτές τις αποφάσεις να μπορούν να κατανοήσουν τους παράγοντες και τις παραμέτρους που τις διαμόρφωσαν. Οι πελάτες των τραπεζικών ιδρυμάτων διαθέτουν δικαίωμα ενημέρωσης σχετικά με τον τρόπο που τα συστήματα τεχνητής νοημοσύνης

αξιολογούν την πιστοληπτική τους ικανότητα, λαμβάνουν αποφάσεις δανειοδότησης και επηρεάζουν τις οικονομικές τους συναλλαγές (Naik, 2024; Valavan, 2023). Αντίθετα, η έλλειψη διαφάνειας στα συστήματα αυτά υπονομεύει την εμπιστοσύνη προς ολόκληρο τον τραπεζικό τομέα και ενδέχεται να οδηγήσει σε αυξημένη δυσaréσκεια και απώλεια πελατών (A. Kaur, 2025; Naik, 2024).

Η στρατηγική για την αντιμετώπιση του προβλήματος του «μαύρου κουτιού» επικεντρώνονται στην ανάπτυξη τεχνικών εξηγήσιμης τεχνητής νοημοσύνης (Explainable AI – XAI), με στόχο τη διασφάλιση ότι τα συστήματα τεχνητής νοημοσύνης είναι ερμηνεύσιμα και συνοδεύονται από σαφή τεκμηρίωση των διαδικασιών λήψης αποφάσεων (Valavan, 2023). Η εξηγησιμότητα αποτελεί παράγοντα ενεργοποίησης και εξασφαλίζει την κατανόηση και τον έλεγχο των συστημάτων τεχνητής νοημοσύνης από χρήστες, ρυθμιστικές αρχές και εμπλεκόμενους φορείς. Οι διαφανείς διαδικασίες λήψης αποφάσεων επιτρέπουν στους εμπλεκόμενους φορείς να αξιολογούν τις ηθικές προεκτάσεις των αποτελεσμάτων της τεχνητής νοημοσύνης, να επαληθεύουν την απουσία μεροληψίας και να κατανοούν τη συλλογιστική πίσω από συγκεκριμένες προτάσεις ή ενέργειες (Akinrinola κ.ά., 2024).

Ωστόσο, είναι σημαντικό να επισημανθεί ότι η ενίσχυση της εξηγησιμότητας των μοντέλων μπορεί να δημιουργήσει νέες προκλήσεις, όπως η μείωση της απόδοσης και η αύξηση του κόστους εκπαίδευσης και χρήσης τους (Marple κ.ά., 2023). Παράλληλα, η δημοσιοποίηση πληροφοριών σχετικά με τον τρόπο λειτουργίας των μοντέλων ενδέχεται να δημιουργήσει κινδύνους ασφαλείας, καθώς μπορεί να αξιοποιηθούν από κυβερνοεγκληματίες για την χειραγώγηση των μοντέλων. Είναι σαφές ότι η διαφάνεια, αν και συχνά ευεργετική, δεν είναι καθολικά θετική, φαινόμενο το οποίο συναντάται ως το «παράδοξο της διαφάνειας στην τεχνητή νοημοσύνη». Συνεπώς, η επίτευξη της διαφάνειας προϋποθέτει τον σεβασμό και την προστασία της ιδιωτικότητας και των προσωπικών δεδομένων (Cheng κ.ά., 2021).

5.5 Επιπτώσεις στην απασχόληση από την τεχνητή νοημοσύνη

Η ενσωμάτωση της τεχνητής νοημοσύνης ενισχύει σημαντικά την παραγωγικότητα των οργανισμών, ωστόσο ενέχει τον κίνδυνο απώλειας θέσεων εργασίας, ιδίως σε τομείς που χαρακτηρίζονται από χαμηλό βαθμό εξειδίκευσης και επαναλαμβανόμενες δραστηριότητες, λόγω της αυξανόμενης αυτοματοποίησης (Naik, 2024; Rashid & Kausik, 2024). Οι οικονομίες που στηρίζονται σε ανειδίκευτο εργατικό δυναμικό ενδέχεται να υποστούν σοβαρές κοινωνικοοικονομικές αναταράξεις, γεγονός που καθιστά αναγκαία την υιοθέτηση

στοχευμένων πολιτικών παρεμβάσεων για τον μετριασμό των ανισοτήτων που προκαλεί αυτή η τεχνολογική μετάβαση (Rashid & Kausik, 2024).

Τις ανησυχίες αυτές συμμερίζεται και η Ευρωπαϊκή Κεντρική Τράπεζα. Όπως επεσήμανε ο Piero Cipollone, μέλος του Εκτελεστικού Συμβουλίου της ΕΚΤ, σε ομιλία του στο Εθνικό Συνέδριο Στατιστικής, η τεχνητή νοημοσύνη διαθέτει τη δυναμική τόσο να υποκαταστήσει όσο και να συμπληρώσει την ανθρώπινη εργασία. Μπορεί να αντικαταστήσει εργασίες ρουτίνας, ενώ παράλληλα ενισχύει την παραγωγικότητα και δημιουργεί νέους επαγγελματικούς ρόλους. Σύμφωνα με εκτιμήσεις της κεντρικής τράπεζας, περίπου το ένα τέταρτο των θέσεων εργασίας στην Ευρώπη θεωρείται ιδιαίτερα εκτεθειμένο στην αυτοματοποίηση που βασίζεται στην τεχνητή νοημοσύνη, ενώ ένα επιπλέον σημαντικό ποσοστό παρουσιάζει μέτριο βαθμό έκθεσης, κυρίως στους τομείς έντασης γνώσης, όπως ο χρηματοοικονομικός. Τα τελικά αποτελέσματα στην απασχόληση παραμένουν αβέβαια και, σύμφωνα με την κεντρική τράπεζα, θα εξαρτηθούν σε μεγάλο βαθμό από την ικανότητα των εργαζομένων να αποκτήσουν συμπληρωματικές δεξιότητες και να προσαρμοστούν στις νέες τεχνολογικές απαιτήσεις (European Central Bank, 2024).

Το McKinsey Global Institute, σε έρευνά του το 2024, παρουσιάζει μια διαφοροποιημένη προοπτική, επισημαίνοντας ότι η υιοθέτηση της παραγωγικής τεχνητής νοημοσύνης αναμένεται να οδηγήσει σε μείωση των διαθέσιμων θέσεων εργασίας στον ευρωπαϊκό χρηματοπιστωτικό τομέα έως το 2030. Οι σημαντικότερες απώλειες εκτιμάται ότι θα σημειωθούν σε υποστηρικτικές διοικητικές θέσεις γραφείου και σε ρόλους εξυπηρέτησης πελατών. Η Εικόνα 4 αποτυπώνει τις προβλεπόμενες μεταβολές, θετικές και αρνητικές, στην απασχόληση του χρηματοπιστωτικού τομέα έως το 2030. Σύμφωνα με τα ευρήματα της μελέτης, η αύξηση της ζήτησης σε αναδυόμενους ρόλους υπολείπεται κατά 276.000 θέσεων σε σχέση με τη μείωση που προβλέπεται στους παραδοσιακούς τομείς (McKinsey Global Institute, 2024). Αξίζει, ωστόσο, να σημειωθεί ότι η έρευνα δεν αποσαφηνίζει αν οι απώλειες αυτές αφορούν αποκλειστικά άμεσες καταργήσεις θέσεων ή και μη αντικαταστάσεις εργαζομένων λόγω συνταξιοδότησης, γεγονός που θα μπορούσε να μετριάσει τον αντίκτυπο στους ήδη απασχολούμενους.

	Ρόλος	Μεταβολή ζήτησης έως 2030, σε χιλιάδες	Ποσοστιαία μεταβολή	Απασχόληση 2022, σε χιλιάδες
TOP 5	Επαγγελματίες επιχειρήσεων και χρηματοοικονομικών	91	7	1,356
	Στελέχη και διευθυντές	60	8	703
	Μηχανικοί και ειδικοί πληροφορικής	51	15	330
	Υποστηρικτικό προσωπικό πληροφορικής	9	11	82
	Δικηγόροι και επαγγελματίες νομικής	5	7	67
BOTTOM 5	Επαγγελματίες πωλήσεων	-16	(3)	608
	Βοηθοί διοίκησης	-64	(21)	305
	Υπάλληλοι καταχώρισης πληροφοριών	-72	(12)	599
	Υπάλληλοι υποστήριξης γραφείου	-98	(22)	436
	Υπάλληλοι ταμείου και λογιστηρίου	-242	(27)	900

Εικόνα 4 Κατάταξη ρόλων που θα επηρεαστούν από την τεχνητή νοημοσύνη στον χρηματοπιστωτικό τομέα

Η επιτάχυνση της ψηφιακής μετάβασης του τραπεζικού κλάδου θα ενισχύσει τη ζήτηση για επαγγελματίες στους τομείς των θετικών και τεχνολογικών επιστημών, εξαιτίας της στρατηγικής εστίασης στην αξιοποίηση των δεδομένων για τη βελτίωση της εμπειρίας των πελατών. Η δυναμική αυτή φέρνει σημαντική αύξηση των αναγκών για εξειδικευμένο προσωπικό, καθώς τα χρηματοπιστωτικά ιδρύματα επενδύουν συστηματικά στις ψηφιακές υπηρεσίες και στα αναγκαία πληροφοριακά συστήματα. Η εκτίμηση είναι ότι έως το 2030 περίπου 600.000 εργαζόμενοι στον τραπεζικό τομέα ενδέχεται να χρειαστεί να αλλάξουν ρόλο, καθώς η ζήτηση για αποφοίτους τριτοβάθμιας εκπαίδευσης και ιδίως για άτομα με ανεπτυγμένες τεχνολογικές, κοινωνικές και συναισθηματικές δεξιότητες θα αυξηθεί, αντανακλώντας την ολοένα μεγαλύτερη ανάγκη για στελέχη σε διευθυντικές θέσεις και ρόλους άμεσης επικοινωνίας με τους πελάτες (McKinsey Global Institute, 2024).

Τα χρηματοπιστωτικά ιδρύματα οφείλουν να διασφαλίσουν ότι διαθέτουν την αναγκαία τεχνογνωσία για την ανάπτυξη και τη συντήρηση των συστημάτων τεχνητής νοημοσύνης (Crisanto κ.ά., 2024). Η αποτελεσματική αξιοποίησή τους προϋποθέτει την καλλιέργεια ενός πολυδιάστατου συνόλου δεξιοτήτων, το οποίο καλύπτει τεχνικούς, αναλυτικούς, διοικητικούς και στρατηγικούς τομείς. Ωστόσο, η υφιστάμενη έλλειψη επαγγελματιών με επαρκή κατάρτιση στην τεχνητή νοημοσύνη έχει οδηγήσει σε μια αυξανόμενη ζήτηση για εξειδικευμένο ανθρώπινο δυναμικό (Ridzuan κ.ά., 2024). Η πρόκληση αυτή μπορεί ταυτόχρονα να αποτελέσει ευκαιρία για την υλοποίηση προληπτικών πρωτοβουλιών

επανεκπαίδευσης και αναβάθμισης δεξιοτήτων του εργατικού δυναμικού, συμβάλλοντας στον μετριασμό των κοινωνικών και οικονομικών επιπτώσεων της τεχνολογικής αυτής μετάβασης (Naik, 2024). Στο πλαίσιο αυτό, τα τραπεζικά ιδρύματα μπορούν να επιδιώξουν συνεργασίες με εκπαιδευτικά ιδρύματα, να αξιοποιήσουν διαδικτυακά προγράμματα κατάρτισης και πιστοποίησης, καθώς και να αναπτύξουν εσωτερικές εκπαιδευτικές πρωτοβουλίες προσαρμοσμένες στις ανάγκες τους (Ridzuan κ.ά., 2024).

6. Μελέτη περίπτωσης της Apple Card

Μία ενδεικτική περίπτωση που αναδεικνύει τις προκλήσεις εφαρμογής της τεχνητής νοημοσύνης στον τραπεζικό τομέα είναι οι αναφορές για διακρίσεις κατά τον καθορισμό του πιστωτικού ορίου της Apple Card. Η υπόθεση αυτή συνδέθηκε με διαφοροποιήσεις στα πιστωτικά όρια μεταξύ συζύγων με παρόμοια οικονομικά χαρακτηριστικά και ανέδειξε με σαφήνεια τη σημασία των ζητημάτων αλγοριθμικής μεροληψίας και δικαιοσύνης στα συστήματα τεχνητής νοημοσύνης στον τραπεζικό τομέα.

6.1 Χαρακτηριστικά της Apple Card

Στις 25 Μαρτίου 2019, η Apple ανακοίνωσε την κυκλοφορία της πιστωτικής κάρτας Apple Card, η οποία δημιουργήθηκε με στόχο την απλότητα, τη διαφάνεια, την προστασία της ιδιωτικότητας και εν τέλει την καλύτερη οικονομική ζωή των πελατών. Στόχος της Apple, ήταν να αναδιαμορφώσει την εμπειρία χρήσης των πιστωτικών καρτών, απλοποιώντας τη διαδικασία απόκτησης, εξαλείφοντας όλες τις χρεώσεις και ενθαρρύνοντας την ελαχιστοποίηση των τόκων (Apple, 2019).

Η πιστωτική κάρτα μπορεί να αποκτηθεί άμεσα από τους χρήστες συσκευών iPhone, αποκλειστικά στις ΗΠΑ, μέσω της εφαρμογής Wallet και να χρησιμοποιηθεί σε φυσικά και ηλεκτρονικά καταστήματα με άμεση ενημέρωση για τις συναλλαγές και το υπόλοιπό τους. Η κάρτα αξιοποιεί αλγορίθμους μηχανικής μάθησης για την ορθή ταξινόμηση δαπανών σε κατηγορίες και παρέχει αναλυτικές εβδομαδιαίες και μηνιαίες αναφορές. Ταυτόχρονα, υπάρχει η δυνατότητα επιστροφής χρημάτων ανάλογα με τις συναλλαγές (cashback), δεν επιβάλλονται τέλη χρήσης (ετήσια, καθυστερήσεων, διεθνών συναλλαγών ή υπερανάληψης) και παρέχονται χαμηλά επιτόκια, με διαφανής μηχανισμούς για τον υπολογισμό τόκων και πληρωμών (Apple, χ.χ.).

Σε επίπεδο ασφάλειας και ιδιωτικότητας, τα στοιχεία της κάρτας είναι αποθηκευμένα στη συσκευή iPhone του χρήστη, η φυσική κάρτα δεν αναγράφει κανένα στοιχείο, ενώ κάθε συναλλαγή εξουσιοδοτείται μέσω Face ή Touch ID και δυναμικού κωδικού ασφαλείας. Η Apple δεν συλλέγει ούτε κοινοποιεί δεδομένα αγορών χρηστών (Apple, χ.χ.), ενώ όπως σχολίασε ο διευθύνων σύμβουλος της Apple κατά την παρουσίαση της Apple Card *«σχεδιάστηκε θέτοντας την ιδιωτικότητα και την ασφάλεια σε προτεραιότητα»* (New York State Department of Financial Services, 2021).

Η Apple Card εκδίδεται σε συνεργασία με τις Goldman Sachs, και Mastercard, που προσφέρουν πρόσβαση σε τραπεζική υποστήριξη και στο διεθνές δίκτυο πληρωμών. Η Goldman Sachs συγκεκριμένα, είναι υπεύθυνη για την πιστωτική πολιτική, τις αποφάσεις αναδοχής κινδύνου της Apple Card, καθώς και για τη διαχείριση εξυπηρέτησης πελατών (New York State Department of Financial Services, 2021).

6.2 Αναφορές για μεροληψία

Τον Νοέμβριο του 2019, λίγους μόλις μήνες μετά την κυκλοφορία της πιστωτικής κάρτας τον Αύγουστο του ίδιου έτους, ένας χρήστης του κοινωνικού δικτύου X (πρώην Twitter) κατήγγειλε δημόσια ότι η Apple Card του παρείχε πιστωτικό όριο είκοσι φορές υψηλότερο από εκείνο της συζύγου του, παρότι το ζευγάρι υπέβαλε κοινή φορολογική δήλωση και η σύζυγος του διέθετε καλύτερη βαθμολογία πιστωτικής αξιολόγησης. Ο ίδιος χρήστης υποστήριξε ότι η τράπεζα βασιζόταν σε αλγορίθμους και τεχνολογίες μηχανικής μάθησης για τη λήψη των σχετικών αποφάσεων, ενώ εξέφρασε τη δυσαρέσκειά του για το γεγονός ότι η εξυπηρέτηση πελατών της Apple Card δεν ήταν σε θέση να εξηγήσει τον τρόπο λειτουργίας των αλγορίθμων ή τα κριτήρια που οδήγησαν στη σημαντική διαφορά των πιστωτικών ορίων. Με αφορμή τη δημοσιότητα που έλαβε η υπόθεση, ένας εκ των συνιδρυτών της Apple ανέφερε ότι και σε αυτόν προσφέρθηκε πιστωτικό όριο περίπου δέκα φορές υψηλότερο από εκείνο της συζύγου του, παρά το γεγονός ότι το ζευγάρι διατηρούσε κοινές καταθέσεις και συνδιαχειριζόταν τα περιουσιακά του στοιχεία (BBC, 2019; New York State Department of Financial Services, 2021; Reuters, 2019).

Λίγες ημέρες αργότερα και κατόπιν επανεξέτασης των δύο υποθέσεων, η Goldman Sachs προέβη σε αναπροσαρμογή των πιστωτικών ορίων των δύο γυναικών, εξισώνοντάς τα με εκείνα που είχαν αρχικά χορηγηθεί στους συζύγους τους. Παρά τη μεταγενέστερη διόρθωση, οι καταγγελίες αποδείχθηκαν επαρκείς ώστε το Τμήμα Χρηματοοικονομικών Υπηρεσιών (Department of Financial Services - DFS) της Πολιτείας της Νέας Υόρκης να κινήσει επίσημη διαδικασία διερεύνησης για πιθανή διάκριση λόγω φύλου στον τρόπο καθορισμού των πιστωτικών ορίων της Apple Card (BBC, 2019; New York State Department of Financial Services, 2021; Reuters, 2019).

6.3 Ευρήματα από την έρευνα της ρυθμιστικής αρχής

Το DFS ερεύνησε το κατά πόσο η Τράπεζα προέβη σε παράνομη διάκριση εις βάρος των γυναικών κατά την έκδοση της Apple Card, χωρίς να βρει στοιχεία παράνομης δραστηριότητας, αφού γυναίκες και άνδρες με ισοδύναμα πιστωτικά χαρακτηριστικά είχαν παρόμοια αποτελέσματα στις αιτήσεις τους. Συγκεκριμένα, το DFS υπογράμμισε τα παρακάτω:

i. Δίκαιη αξιολόγηση δανεισμού

Η ισχύουσα νομοθεσία της Νέας Υόρκης διέκρινε δύο βασικές μορφές διακρίσεων στον τομέα του δανεισμού: την άμεση διάκριση (disparate treatment) και τη διάκριση λόγω άνισων επιπτώσεων (disparate impact). Η πρώτη αφορά την εκούσια χρήση απαγορευμένων κριτηρίων, όπως το φύλο του υποψηφίου, κατά τη διαδικασία λήψης πιστωτικών αποφάσεων, ενώ η δεύτερη σχετίζεται με πολιτικές ή πρακτικές που φαίνονται ουδέτερες, αλλά στην πράξη οδηγούν σε δυσμενή αποτελέσματα για μέλη προστατευόμενων ομάδων. Η έρευνα δεν εντόπισε κανένα από τα δύο είδη διακρίσεων.

Αναφορικά με την άμεση διάκριση, δεν διαπιστώθηκαν πολιτικές που να προβλέπουν χαμηλότερα πιστωτικά όρια για τις γυναίκες, ενώ όσον αφορά τις άνισες επιπτώσεις, η εξειδικευμένη στατιστική ανάλυση των δεδομένων αξιολόγησης πιστοληπτικής ικανότητας δεν κατέδειξε παραβίαση των σχετικών νομοθετικών διατάξεων. Κατά τη διάρκεια της έρευνας διαπιστώθηκε ότι η Goldman Sachs είχε υιοθετήσει και εφαρμόσει ένα πρόγραμμα δίκαιου δανεισμού, διασφαλίζοντας ότι η πιστωτική της πολιτική, καθώς και το χρησιμοποιούμενο μοντέλο, δεν ενσωμάτωναν απαγορευμένα χαρακτηριστικά των αιτούντων και δεν παρήγαγαν άνισες επιπτώσεις.

Τέλος, η Goldman Sachs ήταν σε θέση να εξηγήσει τους παράγοντες που συνέβαλαν στη λήψη των πιστωτικών αποφάσεων του αλγορίθμου της. Οι εν λόγω παράγοντες εναρμονίζονταν με την υφιστάμενη πιστωτική της πολιτική και κανένας από αυτούς δεν αποτελούσε παράνομη βάση για τη διαμόρφωση της τελικής απόφασης.

ii. Διαφάνεια

Οι καταναλωτές υπέβαλαν προσφυγές καταγγέλλοντας την έλλειψη διαφάνειας στον καθορισμό των όρων πίστωσης της Apple Card, με έναν εξ αυτών να επισημαίνει ότι ο αλγόριθμος προσδιορισμού του πιστωτικού ορίου λειτουργούσε ως «μαύρο κουτί». Βάσει

της ισχύουσας νομοθεσίας, ο πιστωτής δεν υποχρεούται να αιτιολογεί τους όρους της χορηγηθείσας πίστωσης, παρά μόνον σε περίπτωση απόρριψης της αίτησης. Ωστόσο, το DFS τόνισε ότι η ενισχυμένη διαφάνεια και η δυνατότητα επαρκούς αιτιολόγησης των αποφάσεων ενδυναμώνουν την εμπιστοσύνη των πελατών προς το τραπεζικό σύστημα.

Η Goldman Sachs ήταν σε θέση να αιτιολογήσει τις αποφάσεις του αλγοριθμικού μοντέλου ενώπιον του ρυθμιστή, εντούτοις η έλλειψη αντίστοιχης διαφάνειας έναντι των ίδιων των καταγγελλόντων οδήγησε σε σύγχυση, η οποία θα μπορούσε να είχε αποφευχθεί. Μια περισσότερο προσανατολισμένη στον πελάτη πολιτική διαχείρισης αιτημάτων και προσφυγών θα είχε τη δυνατότητα να μετριάσει τη σύγχυση αυτή, ιδίως λαμβανομένης υπόψη της περιορισμένης ενημέρωσης των καταναλωτών ως προς τη βάση προσδιορισμού των όρων πίστωσης που τους χορηγήθηκαν.

Επιπλέον, η Goldman Sachs εμφανίστηκε απροετοίμαστη για την πιθανότητα το σύνθετο μοντέλο ανάληψης κινδύνου να παράγει αποτελέσματα ικανά να εκπλήξουν τους ενδιαφερόμενους. Τα προβλήματα αυτά εντάθηκαν περαιτέρω λόγω των εσωτερικών προθεσμιών και της πίεσης στην οποία υπεβλήθη για την έγκαιρη κυκλοφορία του προϊόντος. Τέλος, το DFS σημείωσε θετικά τις πρωτοβουλίες ενίσχυσης της διαφάνειας, όπως το «Path to Apple Card», από την Goldman Sachs και την Apple, το οποίο παρέχει οδηγίες βελτίωσης της πιστοληπτικής ικανότητας σε απορριφθέντες αιτούντες (π.χ. έγκαιρες πληρωμές, μείωση χρεών). Ο ιστότοπος της Apple Card πλέον αναφέρει τους παράγοντες καθορισμού όρων και προτείνει μέτρα βελτίωσης εγκρίσεων.

iii. Διαφορετικά πιστωτικά όρια μεταξύ συζύγων

Οι καταναλωτές φαίνεται να θεωρούν ότι οι σύζυγοι δικαιούνται αυτόματα ίδιους όρους πίστωσης εφόσον μοιράζονται τα οικονομικά τους. Παρόλο που τα εν λόγω ζευγάρια διαμαρτυρήθηκαν για διαφορετική μεταχείριση από την Τράπεζα, συνήθως παρουσίαζαν διαφορετικές πιστωτικές βαθμολογίες, ενώ η λεπτομερής εξέταση των αιτήσεων τους αποκάλυψε διαφορές στα πιστωτικά τους προφίλ. Για παράδειγμα, σε ορισμένες περιπτώσεις ο ένας σύζυγος εμφανιζόταν ως συνυπογράφων σε στεγαστικό δάνειο, ενώ ο άλλος όχι, στοιχεία τα οποία μπορούν να δικαιολογήσουν διαφοροποιήσεις στα πιστωτικά όρια. Αποτελεί, λοιπόν, μια κοινή παρανόηση ότι οι σύζυγοι έχουν δικαίωμα σε ίδιους όρους πίστωσης εάν απλά μοιράζονται τα οικονομικά τους (New York State Department of Financial Services, 2021).

6.4 Γενικές παρατηρήσεις από τη ρυθμιστική αρχή

Το DFS ολοκληρώνοντας την έρευνα του εστίασε σε τρία θέματα γενικότερου ενδιαφέροντος τα οποία έχουν εκτεταμένες επιπτώσεις στις πιστωτικές αποφάσεις από τα τραπεζικά ιδρύματα.

Το πρώτο αφορούσε τις πιστωτικές βαθμολογίες των καταναλωτών που χρησιμοποιούνται από τα τραπεζικά ιδρύματα ως βασικό μέσω προσδιορισμού της πιστοληπτικής ικανότητας τους. Παρά την δημοφιλία τους, οι πιστωτικές βαθμολογίες συχνά δεν είναι εύκολα κατανοητές αφού ο τρόπος αλλά και οι χρησιμοποιούμενες πηγές για τον καθορισμό τους δεν είναι εξηγήσιμες ή/και διαφανής. Όπως χαρακτηριστικά σημειώνει η ρυθμιστική αρχή, υπάρχουν διαφορετικά είδη πιστωτικής βαθμολογίας μεταξύ διαφορετικών τραπεζών, τα οποία χρησιμοποιούν διαφορετικά δεδομένα για τον καθορισμό τους (New York State Department of Financial Services, 2021).

Παράλληλα, το DFS τόνισε την ανάγκη οι καταναλωτές να ελέγχουν τακτικά τις πιστωτικές τους εκθέσεις για τυχόν σφάλματα, τα οποία μπορεί να υπονομεύσουν την πιστοληπτική τους ικανότητα. Μια ακριβής πιστωτική έκθεση διασφαλίζει ότι οι καταναλωτές έχουν πρόσβαση σε προϊόντα σύμφωνα με το πραγματικό πιστωτικό προφίλ τους, ιδίως σε περιπτώσεις σημαντικών δανειακών δεσμεύσεων όπως αγορές ακινήτων. Τα σοβαρά σφάλματα στις πιστωτικές εκθέσεις μπορούν να αυξήσουν το κόστος δανεισμού των καταναλωτών, υπογραμμίζοντας την κρισιμότητα των τακτικών ελέγχων (New York State Department of Financial Services, 2021).

Το δεύτερο ζήτημα αφορούσε τον κίνδυνο διακρίσεων από τις πιστωτικές βαθμολογίες, ακόμη και όταν αυτές διαμορφώνονται νομίμως. Τα ιστορικά δεδομένα δανεισμού, που χρησιμοποιούνται για τον υπολογισμό των πιστωτικών βαθμολογιών, ενδέχεται να ενσωματώνουν μακροχρόνιες φυλετικές και έμφυλες διακρίσεις, περιορίζοντας την πρόσβαση σε πιστωτικά προϊόντα. Επιπλέον, οι δανειστές βασίζονται σε αυτά τα δεδομένα για την ανάπτυξη και δοκιμή μοντέλων ανάληψης κινδύνου, διαιωρίζοντας έτσι προκαταλήψεις από προηγούμενες πρακτικές. Αυτό το φαινόμενο οδηγεί ακούσια σε μεροληπτικά αποτελέσματα, επιβάλλοντας υψηλότερο κόστος δανεισμού ή πλήρη αποκλεισμό ορισμένων καταναλωτών, εντείνοντας τις κοινωνικές ανισότητες (New York State Department of Financial Services, 2021).

Το τελευταίο ζήτημα αφορά τη χρήση εναλλακτικών μεθόδων πιστοληπτικής αξιολόγησης, οι οποίες αξιοποιούν μη παραδοσιακές πηγές δεδομένων με στόχο τον μετριασμό ιστορικών προκαταλήψεων. Ενδεικτικά, ως εναλλακτικά δεδομένα μπορούν να χρησιμοποιηθούν οι πληρωμές λογαριασμών κοινής ωφέλειας, οι ταμειακές ροές τραπεζικών λογαριασμών, καθώς και στοιχεία που σχετίζονται με τη διαδικτυακή δραστηριότητα των καταναλωτών. Η αξιοποίηση τέτοιων πηγών δύναται να βελτιώσει την πρόσβαση σε πιστωτικά προϊόντα συνολικά ή να επιτρέψει τη χορήγηση προϊόντων υψηλότερης ποιότητας σε καταναλωτές με περιορισμένο παραδοσιακό πιστωτικό ιστορικό. Η ενσωμάτωση εναλλακτικών δεδομένων μπορεί να καταστήσει ορατούς, καταναλωτές οι οποίοι μέχρι σήμερα αποκλείονται από προϊόντα πίστωσης, καθώς δεν πληρούν τα συμβατικά κριτήρια αξιολόγησης. Με τον τρόπο αυτό, διευρύνεται δυνητικά η βάση των επιλέξιμων δανειοληπτών και ενισχύεται η πιστωτική ένταξη κοινωνικών ομάδων που παραδοσιακά υποεκπροσωπούνται. Ωστόσο, ακόμη και οι εναλλακτικές μέθοδοι πιστοληπτικής αξιολόγησης οφείλουν να συμμορφώνονται με το ισχύον κανονιστικό πλαίσιο. Η χρήση τους ενέχει κινδύνους, όπως η αξιοποίηση ανακριβών ή αδιαφανών πληροφοριών και η ενσωμάτωση νέων μορφών προκατάληψης, γεγονός που καθιστά αναγκαία την ύπαρξη ισχυρών μηχανισμών εποπτείας και διαφάνειας (New York State Department of Financial Services, 2021).

7. Συζήτηση

Η υπόθεση της Apple Card ανέδειξε με ιδιαίτερη έμφαση το ζήτημα της μεροληψίας στα συστήματα τεχνητής νοημοσύνης, με την παρέμβαση του DFS να έρχεται έπειτα από καταγγελίες καταναλωτών σχετικά με την άνιση μεταχείριση μεταξύ συζύγων στον καθορισμό των πιστωτικών ορίων. Η έρευνα δεν διαπίστωσε παραβάσεις της νομοθεσίας, ωστόσο η περίπτωση αυτή υποδηλώνει ότι οι χρήστες αντιλαμβάνονται τις αποφάσεις των αλγορίθμων ως δυνητικά άδικες, ακόμη και όταν αυτές συμμορφώνονται με το νομικό πλαίσιο.

Οι παρατηρήσεις από το DFS ανέδειξαν κρίσιμα σημεία που είναι αναγκαίο να ληφθούν υπόψη για την αποφυγή διακρίσεων κατά την λήψη πιστωτικών αποφάσεων. Αφενός, υπογραμμίστηκε η ανάγκη ελέγχου από τους καταναλωτές των πιστωτικών τους εκθέσεων, καθώς πιθανά σφάλματα στα υποκείμενα δεδομένα μπορούν να υπονομεύσουν την ακεραιότητα του συνολικού συστήματος αξιολόγησης πιστοληπτικής ικανότητας. Αφετέρου, επισημάνθηκε ο κίνδυνος οι πιστωτικές βαθμολογίες, ακόμη και όταν παράγονται σύμφωνα με το ισχύον ρυθμιστικό πλαίσιο, να ενσωματώνουν και να διαιωνίζουν ιστορικές φυλετικές και έμφυλες διακρίσεις, οδηγώντας σε περιορισμένη πιστωτική πρόσβαση ή σε αυξημένο κόστος δανεισμού για ορισμένες κατηγορίες καταναλωτών (New York State Department of Financial Services, 2021). Και οι δύο παρατηρήσεις υπογραμμίζουν τη σημαντικότητα των δεδομένων που χρησιμοποιούνται σε ολόκληρη τη διάρκεια ζωής των συστημάτων τεχνητής νοημοσύνης και την αναγκαιότητα αξιοπιστίας και ακεραιότητας τους.

Οι παρατηρήσεις του DFS ευθυγραμμίζονται με το θεωρητικό πλαίσιο της εργασίας, αφού τα συστήματα τεχνητής νοημοσύνης για την πιστωτική βαθμολόγηση και δανειοδότηση είναι ευάλωτα σε πολλαπλές μορφές μεροληψίας, οι οποίες μπορεί να λειτουργούν ταυτόχρονα και επικουρικά. Τα δεδομένα εκπαίδευσης είναι πιθανό να ενσωματώνουν διαχρονικές και δομικές ανισότητες, ο σχεδιασμός των μοντέλων να ευνοεί στόχους όπως η μείωση του κόστους ή του κινδύνου εις βάρος της δίκαιης μεταχείρισης, ενώ η ανθρώπινη παρέμβαση σε όλα τα στάδια του κύκλου ζωής των συστημάτων μπορεί να εισάγει πρόσθετες στρεβλώσεις και διακρίσεις (Valavan, 2023).

Παράλληλα, η περίπτωση της Apple Card αποκαλύπτει ελλείψεις στα ζητήματα λογοδοσίας και ανθρώπινης εποπτείας των αλγοριθμικών συστημάτων τεχνητής νοημοσύνης, παρά την

τυπική συμμόρφωση με τα υφιστάμενα ρυθμιστικά πλαίσια. Η Apple και η Goldman Sachs ανταποκρίθηκαν στις απαιτήσεις του DFS, παρέχοντας την απαραίτητη τεκμηρίωση και δεδομένα που τους ζητήθηκαν, ενώ ήταν ξεκάθαρη και η κατανομή των ευθυνών, με την Goldman Sachs να είναι υπεύθυνη για την πιστωτική πολιτική, τις αποφάσεις ανάληψης κινδύνου και την εξυπηρέτηση πελατών (New York State Department of Financial Services, 2021). Ωστόσο, ο τρόπος εποπτείας του αλγορίθμου που αποδέχεται ή απορρίπτει αιτήσεις πίστωσης και καθορίζει τα πιστωτικά όρια βάσει των βαθμολογιών πιστωτικής αξιολόγησης, παρέμεινε ασαφής, αποκαλύπτοντας ένα κενό μεταξύ της τυπικής και πραγματικής λογοδοσίας έναντι των χρηστών.

Η έλλειψη διαφάνειας και εξηγησιμότητας συνιστά ακόμα ένα ελάττωμα που απορρέει από τη μελέτη περίπτωσης, υπονομεύοντας θεμελιωδώς την εμπιστοσύνη των καταναλωτών στις αλγοριθμικές αποφάσεις. Η μη αιτιολόγηση των πιστωτικών ορίων προς τους πελάτες, αν και αποδεκτή βάσει της νομοθεσίας της Νέας Υόρκης, δημιούργησε σύγχυση και δυσπιστία, ιδίως όταν ακόμα και οι αντιπρόσωποι της εξυπηρέτησης πελατών αδυνατούσαν να αιτιολογήσουν τις αποφάσεις του αλγορίθμου (Kureishy, 2019; New York State Department of Financial Services, 2021). Παρά το γεγονός ότι η Goldman Sachs κατόρθωσε να εξηγήσει επαρκώς τις αποφάσεις του μοντέλου ενώπιον του ρυθμιστή, η αδυναμία αντίστοιχης εξήγησης προς τους καταγγέλλοντες ενέτεινε την κρίση εμπιστοσύνης (Kureishy, 2019; New York State Department of Financial Services, 2021), αποτυγχάνοντας να μείνει πιστή τις αρχές της διαφάνειας και εξηγησιμότητας που τονίζονται στο θεωρητικό πλαίσιο της εργασίας.

Η μελέτη περίπτωσης επιτρέπει επίσης την έμμεση εξέταση των ζητημάτων ιδιωτικότητας. Αν και η Apple Card, δεν κατηγορήθηκε για παραβιάσεις ιδιωτικότητας και προσωπικών δεδομένων, είχε φροντίσει μέσα από το υλικό μάρκετινγκ να τονίσει την ιδιαίτερη προσοχή που δόθηκε σε θέματα ασφαλείας και προστασίας δεδομένων, τονίζοντας ότι δεν συλλέγει ούτε κοινοποιεί δεδομένα αγορών χρηστών (New York State Department of Financial Services, 2021). Επιπλέον, η ενσωμάτωση της Apple Card στο λογισμικό iOS αποτέλεσε ακόμα ένα εχέγγυο ασφαλείας, αφού προσφέρει πολυεπίπεδη προστασία (υλικό, λογισμικό, κρυπτογράφηση κ.ά.) (Apple, 2024), ενώ το ίδιο το λογισμικό είναι κλειστού κώδικα, ή όπως χαρακτηριστικά είχε δηλώσει ο ιδρυτής της αποτελεί παράδειγμα ολοκληρωμένου (integrated) λογισμικού (Jobs, 2010), ενισχύοντας έτσι την περίμετρο ασφαλείας.

Τέλος, αν και οι δυνητικές επιπτώσεις στην απασχόληση δεν αποτελούν το αντικείμενο της συγκεκριμένης μελέτης περίπτωσης, η φύση της Apple Card τις κάνει εμφανείς. Η μετάβαση από τις παραδοσιακές τραπεζικές διαδικασίες, που απαιτούσαν τη φυσική παρουσία σε τραπεζικά καταστήματα και την ανθρώπινη αξιολόγηση, στις ηλεκτρονικές, αυτοματοποιημένες «γραμμές παραγωγής» αναμένεται να επηρεάσει ποικιλοτρόπως τις θέσεις εργασίας σε ολόκληρο τον κύκλο ζωής των υπηρεσιών πίστωσης και τραπεζικής.

Ολοκληρώνοντας το κεφάλαιο της συζήτησης, είναι αναγκαίο να αναφερθούν οι περιορισμοί που παρατηρήθηκαν κατά τη διάρκεια εκπόνησης της εργασίας. Η μελέτη περίπτωσης της Apple Card βασίζεται κυρίως σε δευτερογενή δεδομένα (έκθεση DFS, δημοσιογραφικά άρθρα), αφού πρωτογενή τεχνικά στοιχεία, ιδιαίτερα σχετικά με την λειτουργία του πιστωτικού αλγορίθμου παραμένουν εμπιστευτικά και μη δημοσίως προσβάσιμα. Επιπλέον, μέχρι και τη συγγραφή της παρούσας, η Apple Card διατίθεται αποκλειστικά σε χρήστες συσκευών iPhone στις ΗΠΑ, δημιουργώντας ένα συγκεκριμένο δείγμα αναφοράς το οποίο είναι πολύ πιθανόν να μην είναι αντιπροσωπευτικό. Τέλος, η απουσία ενός παγκοσμίως αποδεκτού ορισμού της τεχνητής νοημοσύνης δυσχεραίνει την κατανόηση των επιμέρους περιπτώσεων χρήσης της στον διεθνή χρηματοπιστωτικό τομέα και περιορίζει τη περαιτέρω δυνατότητα εντοπισμού τομέων αυξημένου ρίσκου.

8. Συμπεράσματα

Τα ερευνητικά ερωτήματα που τέθηκαν στο αρχικό στάδιο της εργασίας απαντήθηκαν μέσα από την συστηματική τους ανάλυση, αφού παρουσιάστηκαν οι πρακτικές εφαρμογές των συστημάτων τεχνητής νοημοσύνης στον τραπεζικό τομέα, τα οφέλη τους, οι προκλήσεις και οι ηθικές προεκτάσεις που ανακύπτουν από την υιοθέτησή τους.

Η μελέτη περίπτωσης της Apple Card δεν αποτελεί μόνο ένα χαρακτηριστικό παράδειγμα της δυναμικής της τεχνητής νοημοσύνης, όντας ένα πιστωτικό προϊόν που αναπτύχθηκε από μια εταιρεία τεχνολογίας σε συνεργασία με μια επενδυτική τράπεζα, αλλά και μία ενδεικτική περίπτωση που συμπυκνώνει τα βασικότερα ηθικά ζητήματα από την ενσωμάτωση της τεχνητής νοημοσύνης στον τραπεζικό κλάδο. Η έλλειψη διαφάνειας και εξηγησιμότητας στις αποφάσεις του πιστωτικού αλγορίθμου οδήγησε σε σημαντική διάβρωση της εμπιστοσύνης των καταναλωτών, τόσο απέναντι στο συγκεκριμένο προϊόν όσο και ευρύτερα απέναντι στους αλγορίθμους τεχνητής νοημοσύνης και στο τραπεζικό σύστημα συνολικά. Ο παράγοντας αυτός λειτούργησε καταλυτικά για την εμφάνιση υποψιών μεροληψίας εκ μέρους του πιστωτικού αλγορίθμου, ανεξάρτητα από το εάν στοιχειοθετήθηκε ή όχι παραβίαση της κείμενης νομοθεσίας.

Τα ηθικά ζητήματα που απορρέουν από την υιοθέτηση της τεχνητής νοημοσύνης δεν πρέπει να αντιμετωπίζονται αποσπασματικά ή εκ των υστέρων από τα τραπεζικά ιδρύματα. Αντιθέτως, οφείλουν να αναπτύξουν ολοκληρωμένες και μακροπρόθεσμες στρατηγικές για την προληπτική και υπεύθυνη διαχείριση των ηθικών προεκτάσεων της τεχνητής νοημοσύνης, εφόσον επιδιώκουν πράγματι τη διατήρηση και ενίσχυση της εμπιστοσύνης των πελατών. Στο πλαίσιο αυτό, η θεσμοθέτηση πελατοκεντρικών μηχανισμών διακυβέρνησης και ελέγχου της τεχνητής νοημοσύνης από τα τραπεζικά ιδρύματα κρίνεται αναγκαία, προκειμένου να διασφαλίζεται ο επαρκής έλεγχος και η αποτελεσματική εποπτεία των σχετικών συστημάτων καθ' όλη τη διάρκεια ζωής τους.

Επιπλέον, η περίπτωση της Apple Card καταδεικνύει ότι η τυπική συμμόρφωση των τραπεζικών οργανισμών με τις εκάστοτε ρυθμιστικές απαιτήσεις δεν επαρκεί για την ουσιαστική αντιμετώπιση των ηθικών προκλήσεων που εγείρει η χρήση της τεχνητής νοημοσύνης. Παρότι οι ΗΠΑ έχουν υιοθετήσει μία απορρυθμιστική προσέγγιση ως προς το κανονιστικό πλαίσιο, δεν πρέπει να παραγνωρίζεται ότι ο τραπεζικός κλάδος αποτελεί έναν

από τους πλέον αυστηρά ρυθμιζόμενους τομείς της παγκόσμιας οικονομίας, με την εμπειρία να δείχνει ότι η αυστηρή κανονιστική ρύθμιση, από μόνη της, δεν είναι πάντοτε ικανή να απαντήσει στις αναδυόμενες προκλήσεις. Οι δύο κυρίαρχες προσεγγίσεις διαμόρφωσης του κανονιστικού πλαισίου για την τεχνητή νοημοσύνη, είτε μέσω μη δεσμευτικών αρχών είτε μέσω δεσμευτικών κανόνων (Crisanto κ.ά., 2024; KPMG, 2024), φαίνεται στην πράξη, να μην επαρκούν για την αποτελεσματική εποπτεία της, συμπέρασμα που ευθυγραμμίζεται με τη διαθέσιμη βιβλιογραφία (Vukoníc κ.ά., 2025).

Υβριδικές ρυθμιστικές προσεγγίσεις, οι οποίες συνδυάζουν πτυχές αρχών και κανόνων και μπορούν να προσαρμόζονται στις ταχείες τεχνολογικές εξελίξεις, διατηρώντας παράλληλα ρυθμιστική συνέπεια, θα μπορούσαν να αποτελέσουν μία βιώσιμη λύση. Βασική προϋπόθεση για την επιτυχή εφαρμογή τους συνιστά ένα υψηλό επίπεδο διεθνούς συντονισμού, προκειμένου να διασφαλιστεί η απαιτούμενη συνοχή μεταξύ διαφορετικών κρατών και θεσμικών οργάνων (Vukoníc κ.ά., 2025). Δεδομένου ότι η παραπάνω προϋπόθεση ενέχει υψηλό βαθμό πολυπλοκότητας και δυσκολίας υλοποίησης, μία εναλλακτική προσέγγιση θα μπορούσε να είναι η ενισχυμένη αυτορρύθμιση του τραπεζικού τομέα. Αυτή θα μπορούσε να υλοποιηθεί μέσω της από κοινού διαμόρφωσης και υιοθέτησης πλαισίων ηθικής για την τεχνητή νοημοσύνη, στο πλαίσιο συνεργασίας αξιόπιστων φορέων, ρυθμιστικών αρχών και μεγάλων τραπεζικών ιδρυμάτων. Ένα τέτοιο πλαίσιο, εφόσον εξασφαλίζει την απαραίτητη ευελιξία και την προσαρμογή σε γεωγραφικές, οικονομικές και οργανωτικές ιδιαιτερότητες (είδος και μέγεθος ιδρυμάτων), θα μπορούσε να αποτελέσει σημείο αναφοράς για την υπεύθυνη υιοθέτηση και χρήση της τεχνητής νοημοσύνης στον τραπεζικό τομέα.

Στο πλαίσιο της αναγκαίας μελλοντικής έρευνας, και με στόχο τόσο τον περαιτέρω εμπλουτισμό της ακαδημαϊκής γνώσης όσο και την αποτελεσματικότερη διαχείριση των ηθικών ζητημάτων που προκύπτουν από την υιοθέτηση συστημάτων τεχνητής νοημοσύνης στον τραπεζικό τομέα, ο συγγραφέας της παρούσας εργασίας προτείνει τη συνέχιση της ερευνητικής δραστηριότητας σε τρεις βασικούς άξονες:

- Στη διερεύνηση των επιπτώσεων στην απασχόληση από την υιοθέτηση της τεχνητής νοημοσύνης στον τραπεζικό τομέα, με έμφαση στις μεταβολές στα επαγγελματικά προφίλ, στις απαιτούμενες δεξιότητες και στις στρατηγικές προσαρμογής των εργαζομένων.

- Στη συγκριτική ανάλυση των διαφορετικών κανονιστικών και ηθικών πλαισίων που διέπουν τη χρήση της τεχνητής νοημοσύνης στον τραπεζικό τομέα, με απώτερο στόχο τη διαμόρφωση ενός ενιαίου, ευρέως αποδεκτού πλαισίου αρχών και κανόνων.
- Στη μελέτη και τεκμηρίωση των βέλτιστων πρακτικών σχετικά με την ανάπτυξη πελατοκεντρικών μηχανισμών διακυβέρνησης και ελέγχου των συστημάτων τεχνητής νοημοσύνης, οι οποίοι θα μπορούν να αποτελέσουν σημείο αναφοράς για τα τραπεζικά ιδρύματα διεθνώς.

Η συστηματική διερεύνηση των παραπάνω αναμένεται να συμβάλει ουσιαστικά στην εμβάθυνση της κατανόησης της τεχνητής νοημοσύνης στον τραπεζικό τομέα και στη διαμόρφωση υπεύθυνων, ανθεκτικών και κοινωνικά δίκαιων πρακτικών υιοθέτησής της και αντιμετώπισης των ηθικών ζητημάτων από προκύπτουν από αυτή.

Βιβλιογραφία

1. Ahmed, F. (2022). *Ethical aspects of Artificial Intelligence in Banking*. <https://www.researchgate.net/publication/373492572>
2. Akinrinola, O., Okoye, C. C., Ofodile, O. C., & Ugochukwu, C. E. (2024). Navigating and reviewing ethical dilemmas in AI development: Strategies for transparency, fairness, and accountability. *GSC Advanced Research and Reviews*, 18(3), 050–058. <https://doi.org/10.30574/gscarr.2024.18.3.0088>
3. American Psychological Association. (2018, Απρίλιος 19). *Definition of Intelligence*. American Psychological Association
4. Apple. (χ.χ.). *Apple Card*. Ανακτήθηκε 18 Δεκέμβριος 2025, από <https://www.apple.com/apple-card/features/>
5. Apple. (2019, Μάρτιος 25). *Introducing Apple Card, a new kind of credit card created by Apple*. <https://www.apple.com/newsroom/2019/03/introducing-apple-card-a-new-kind-of-credit-card-created-by-apple/>
6. Apple. (2024, Δεκέμβριος). *Apple Platform Security*. Apple. <https://support.apple.com/guide/security/welcome/web>
7. Baffour Gyau, E., Appiah, M., Gyamfi, B. A., Achie, T., & Naeem, M. A. (2024). Transforming banking: Examining the role of AI technology innovation in boosting banks financial performance. *International Review of Financial Analysis*, 96. <https://doi.org/10.1016/j.irfa.2024.103700>
8. BBC. (2019, Νοέμβριος 11). *Apple's 'sexist' credit card investigated by US regulator*. <https://www.bbc.com/news/business-50365609>
9. Belcic, I., & Stryker, C. (2025). *What is supervised learning?* IBM Think. <https://www.ibm.com/think/topics/supervised-learning#1509394340>
10. Bergmann, D. (2025). *What is machine learning?* <https://www.ibm.com/think/topics/machine-learning>
11. Chaddha, R., & Agrawal, G. (2023). Ethics and Morality. *Indian Journal of Orthopaedics*, 57(11), 1707–1713. <https://doi.org/10.1007/s43465-023-01004-3>
12. Chart of Fundamental Rights of the European Union (2012). https://eur-lex.europa.eu/eli/treaty/char_2012/oj/eng

13. Cheng, L., Varshney, K. R., & Liu, H. (2021). Socially Responsible AI Algorithms: Issues, Purposes, and Challenges. *Journal of Artificial Intelligence Research*, 71, 1137–1181. <https://doi.org/10.48550/arXiv.2101.02032>
14. Chlouverakis, K. (2024, Απρίλιος 26). *How artificial intelligence is reshaping the financial services industry*. EY. https://www.ey.com/en_gr/insights/financial-services/how-artificial-intelligence-is-reshaping-the-financial-services-industry
15. Condoleezza Rice, Taylor, J. B., Widom, J., Zegart, A., Lin, H. S., & Giles, M. (2025). *THE STANFORD EMERGING TECHNOLOGY REVIEW 2025 A Report on Ten Key Technologies and Their Policy Implications*. <https://setr.stanford.edu/technology-areas-2025>
16. Crisanto, J. C., Leuterio, C. B., Prenio, J., & Yong, J. (2024). *Regulating AI in the financial sector: recent developments and main challenges*. Bank for International Settlements.
17. Danish Institute for Human Rights. (2025). *Data protection & privacy*. <https://globalnaps.org/issue/data-protection-and-privacy/>
18. Delipetrev, B., Tsinaraki, C., & Kostić, U. (2020). *AI Watch Historical Evolution of Artificial Intelligence*. <https://doi.org/10.2760/801580>
19. Dhashanamoorthi, B. (2023). Opportunities and challenges of artificial intelligence in banking and financial services. *International Journal of Science and Research Archive*, 10(2), 272–279. <https://doi.org/10.30574/ijrsra.2023.10.2.0947>
20. Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., López de Prado, M., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, 99. <https://doi.org/10.1016/j.inffus.2023.101896>
21. Durth, S., Hancock, B., Maor, D., & Sukharevsky, A. (2023). *The organization of the future: Enabled by gen AI, driven by people Generative AI can empower people-but only if leaders take a broad view of its capabilities and deeply consider its implications for the organization*. <https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/the-organization-of-the-future-enabled-by-gen-ai-driven-by-people#/>

22. European Banking Authority. (2024). *Artificial intelligence*.
<https://www.eba.europa.eu/publications-and-media/publications/special-topic-artificial-intelligence>
23. European Central Bank (2024). *Artificial intelligence: a central bank's view* (P. Cipollone,)
https://www.ecb.europa.eu/press/key/date/2024/html/ecb.sp240704_1~e348c05894.en.html
24. European Commission, 8 (2025).
<https://digital-strategy.ec.europa.eu/en/library/commission-publishes-guidelines-ai-system-definition-facilitate-first-ai-acts-rules-application>
25. European Parliament. (2025, Φεβρουάριος 19). *EU AI Act: first regulation on artificial intelligence*. European Parliament.
<https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence>
26. Executive Order 14110, 88 (2023).
<https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence>
27. Executive Order 14119 (2025).
<https://www.federalregister.gov/documents/2025/01/31/2025-02172/removing-barriers-to-american-leadership-in-artificial-intelligence>
28. Faheem, M. A. (2024). Ethical AI: Addressing bias, fairness, and accountability in autonomous decision-making systems. *World Journal of Advanced Research and Reviews*, 23(2), 1703–1711. <https://doi.org/10.30574/wjarr.2024.23.2.2510>
29. Galazova, S. S., & Magomaeva, L. R. (2019). The Transformation of Traditional Banking Activity in Digital. *International Journal of Economics and Business Administration*, VII(2), 41–51. <https://doi.org/10.35808/ijeba/369>
30. Gignac, G. E., & Szodorai, E. T. (2024). Defining intelligence: Bridging the gap between human and artificial perspectives. *Intelligence*, 104.
<https://doi.org/10.1016/j.intell.2024.101832>

31. Giovine, C., Lerner, L., Thomas, R., Singh, S., Kakulavarapu, S., Chung, V., & Motwani, Y. (2024). *Extracting value from AI in banking: Rewiring the enterprise*. <https://www.mckinsey.com/industries/financial-services/our-insights/extracting-value-from-ai-in-banking-rewiring-the-enterprise>
32. Google. (2025). *What is artificial intelligence (AI) in banking?* <https://cloud.google.com/discover/ai-in-banking#how-do-banks-use-ai>
33. Google Cloud. (χ.χ.). *Artificial intelligence (AI): a simple-to-understand guide*. Ανακτήθηκε 3 Νοέμβριος 2025, από <https://cloud.google.com/learn/what-is-artificial-intelligence#what-is-artificial-intelligence-ai>
34. Google Cloud. (2025a). *What is Deep Learning?* <https://cloud.google.com/discover/what-is-deep-learning?hl=en>
35. Google Cloud. (2025b). *What is machine learning (ML)?* <https://cloud.google.com/learn/what-is-machine-learning#types-of-machine-learning>
36. Greene, J. (2023). *Ethics: Research Starters Topic.Ethics*. EBSCO. <https://www.ebsco.com/research-starters/religion-and-philosophy/ethics-research-starters-topicethics#full-article>
37. High-Level Expert Group on Artificial Intelligence. (2019). *HIGH-LEVEL EXPERT GROUP ON ARTIFICIAL INTELLIGENCE SET UP BY THE EUROPEAN COMMISSION ETHICS GUIDELINES FOR TRUSTWORTHY AI*. <https://digital-strategy.ec.europa.eu/en/policies/expert-group-ai>
38. Hirz, J., Kivisaari, E., Miehe, P., Senatore, C., Tautan, B., & Toraldo, F. (2024). *What should an actuary know about Artificial Intelligence?* <https://actuary.eu/wp-content/uploads/2024/01/What-should-an-actuary-know-about-Artificial-Intelligence.pdf>
39. IBM. (2025). *What is deep learning?* <https://www.ibm.com/think/topics/deep-learning#763338456>
40. IBM Data and AI Team. (2025). *Understanding the different types of artificial intelligence*. IBM Think. <https://www.ibm.com/think/topics/artificial-intelligence-types>
41. Jain, R. (2023). Role of artificial intelligence in banking and finance. *Journal of Management and Science*, 13(3), 1–4. <https://jmseleyon.com/index.php/jms/article/view/678>

42. Jobs, S. (2010, Οκτώβριος 18). *Steve Jobs's Epic 5-Minute Anti-Google Rant*. Business Insider. <https://www.businessinsider.com/steve-jobs-epic-5-minute-anti-google-rant-2010-10?IR=T>
43. Kaur, A. (2025). A Comprehensive Analysis of Types of Artificial Intelligence: Classification, Applications, and Future Directions. *International Journal of Advanced Research in Computer and Communication Engineering Impact Factor*, 14(2). <https://doi.org/10.17148/IJARCCCE.2025.14221>
44. Kaur, M. (2024). A Comprehensive Overview of Artificial Intelligence-Based Classification Techniques. *International Journal of Science and Research Archive*, 11(2), 125–129. <https://doi.org/10.30574/ijstra>
45. Kaur, N., Sahdev, S. L., Sharma, M., & Siddiqui, L. (2020). Banking 4.0: The influence of Artificial Intelligence on the banking industry & how AI is changing the face of modern day banks. *International Journal of Management*, 11(6). <https://doi.org/10.34218/ijm.11.6.2020.049>
46. KPMG. (2024). *Effective model risk management framework for AI/ML based models*. <https://assets.kpmg.com/content/dam/kpmg/in/pdf/2024/10/effective-model-risk-management-framework-for-ai-ml-based-models.pdf>
47. Kureishy, A. (2019, Νοέμβριος 12). *What the Apple Card Controversy Says About our AI Future*. Teradata. <https://www.teradata.com/blogs/what-the-apple-card-controversy-says-about-our-ai-future>
48. Lavanya, B., & Rajkumar, A. D. (2024). Paradigm shift in the digital transformation of the banking sector: A bibliometric analysis. *International Journal of Advanced and Applied Sciences*, 11(3), 115–126. <https://doi.org/10.21833/ijaas.2024.03.013>
49. Lawrence Damilare Oyeniya, Chinonye Esther Ugochukwu, & Noluthando Zamanjomane Mhlongo. (2024). Implementing AI in banking customer service: A review of current trends and future applications. *International Journal of Science and Research Archive*, 11(2), 1492–1509. <https://doi.org/10.30574/ijstra.2024.11.2.0639>
50. Lee, S. U., Perera, H., Liu, Y., Xia, B., Lu, Q., Zhu, L., Salvado, O., & Whittle, J. (2025). *Responsible AI Question Bank: A Comprehensive Tool for AI Risk Assessment*. <http://arxiv.org/abs/2408.11820>

51. Maple, C., Szpruch, L., Epiphaniou, G., Staykova, K., Singh, S., Penwarden, W., Wen, Y., Wang, Z., Hariharan, J., & Avramovic, P. (2023). *The AI Revolution: Opportunities and Challenges for the Finance Sector*.
<https://www.turing.ac.uk/news/publications/ai-revolution-opportunities-and-challenges-finance-sector>
52. McKinsey Global Institute. (2024). *A new future of work: The race to deploy AI and raise skills in Europe and beyond*.
https://www.mckinsey.de/~ /media/mckinsey/locations/europe%20and%20middle%20east/deutschland/news/presse/2024/2024%20-%2005%20-%2023%20mgi%20genai%20future%20of%20work/mgi%20report_a-new-future-of-work-the-race-to-deploy-ai.pdf
53. Morandín-Ahuerma, F. (2022). What is Artificial Intelligence? *International Journal of Research Publication and Reviews*, 3, 1947-1951.
54. Munira, M. S. K. (2025). Digital Transformation in Banking: A Systematic Review Of Trends, Technologies, And Challenges. *Strategic Data Management and Innovation*, 2(01), 78–95. <https://doi.org/10.71292/sdmi.v2i01.12>
55. Naik, A. (2024). Influence of AI in Banking: Ethical and Compliance Implications. *Journal of Global Research in Computer Sciences*, 15(1).
<https://doi.org/10.4172/2229-371X.15.2.002>
56. New York State Department of Financial Services. (2021). *New York State Department of Financial Services Report on Apple Card Investigation*.
https://www.dfs.ny.gov/system/files/documents/2021/03/rpt_202103_apple_card_investigation.pdf
57. OECD. (2024a). *AI, Data governance and privacy: Synergies and areas of co-operation*. [https://doi.org/DSTI/CDEP/AIGO/DGP\(2023\)1/FINAL](https://doi.org/DSTI/CDEP/AIGO/DGP(2023)1/FINAL)
58. OECD. (2024b). *Explanatory memorandum on the updated OECD definition of an AI system*. https://www.oecd.org/en/publications/explanatory-memorandum-on-the-updated-oecd-definition-of-an-ai-system_623da898-en.html
59. Rashid, A. Bin, & Kausik, M. A. K. (2024). AI revolutionizing industries worldwide: A comprehensive overview of its diverse applications. *Hybrid Advances*, 7.
<https://doi.org/10.1016/j.hybadv.2024.100277>

60. Regulation (EU) 2024/1689 (2024).
<http://data.europa.eu/eli/reg/2024/1689/oj>
61. Reuters. (2019, Νοέμβριος 10). *Goldman faces probe after entrepreneur slams Apple Card algorithm in tweets*. <https://www.reuters.com/article/us-goldman-sachs-probe/goldman-faces-probe-after-entrepreneur-slams-apple-card-algorithm-in-tweets-idUSKBN1XK00L/>
62. Ridzuan, N. N., Masri, M., Anshari, M., Fitriyani, N. L., & Syafrudin, M. (2024). AI in the Financial Sector: The Line between Innovation, Regulation and Ethical Responsibility. *Information (Switzerland)*, 15(8).
<https://doi.org/10.3390/info15080432>
63. Schettenhelm, M. (2025, Αύγουστος 8). US regulation of artificial intelligence. *Bloomberg Intelligence*. <https://www.bloomberg.com/professional/insights/artificial-intelligence/us-regulation-of-artificial-intelligence/>
64. Schwartz, R., Vassilev, A., Greene, K., Perine, L., Burt, A., & Hall, P. (2022). *Towards a standard for identifying and managing bias in artificial intelligence*.
<https://doi.org/10.6028/NIST.SP.1270>
65. Singer, P. (2025). Ethics philosophy. *Britannica*.
<https://www.britannica.com/topic/ethics-philosophy>
66. Stryker, C., & Holdsworth, J. (2025). *What is NLP (natural language processing)?* IBM Think. <https://www.ibm.com/think/topics/natural-language-processing>
67. Stryker, C., & Kavlakoglu, E. (2024). *What is artificial intelligence (AI)?* IBM Think. <https://www.ibm.com/think/topics/artificial-intelligence>
68. Stryker, C., & Scapicchio, M. (2025). *What is generative AI?* IBM Think. <https://www.ibm.com/think/topics/generative-ai#257779831>
69. Turing, A. M. (1950). Computing Machinery and Intelligence. Στο *New Series* (τ. 59, Τεύχος 236).
70. UN. Advisory Body on Artificial Intelligence. (2024). *Governing AI for Humanity*. <https://digitallibrary.un.org/record/4062495>
71. UNESCO. (2022). *Recommendation on the Ethics of Artificial Intelligence*.
<https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence>

72. Universal Declaration of Human Rights (1948). <https://www.un.org/en/about-us/universal-declaration-of-human-rights>
73. Valavan, T. A. (2023). AI Ethics and Bias: Exploratory study on the ethical considerations and potential biases in ai and data-driven decision-making in banking, with a focus on fairness, transparency, and accountability. *World Journal of Advanced Research and Reviews*, 20(4), 197–206.
<https://doi.org/10.30574/wjarr.2023.20.2.2245>
74. Velasquez, M., Andre, C., Shanks, T., S.J., & Meyer, M. J. (2010, Ιανουάριος 1). *What is Ethics?* Issues in Ethics IIE V1 N1 (Fall 1987). Revised in 2010.
<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/what-is-ethics/>
75. Vuković, D. B., Dekpo-Adza, S., & Matović, S. (2025). AI integration in financial services: a systematic review of trends and regulatory challenges. *Humanities and Social Sciences Communications* (τ. 12, Τεύχος 1). Springer Nature.
<https://doi.org/10.1057/s41599-025-04850-8>
76. Wechsler, D. (1944). *The Measurement of Adult Intelligence* (Third Edition). THE WILLIAMS & WILKINS COMPANY.
77. Weitzel, P. (2025). *Defining Artificial Intelligence*.
<https://dx.doi.org/10.2139/ssrn.5154389>
78. World Economic Forum. (2025). *Artificial Intelligence in Financial Services*.