



ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

**Η Ηθική Διάσταση στη Σχεδίαση Εφαρμογών και
Συστημάτων Τεχνητής Νοημοσύνης**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Κατσίγρα Αβραάμ

Επιβλέπουσα Καθηγήτρια : Δρ. Διαμαντοπούλου Βασιλική, Επίκουρη Καθηγήτρια

Μέλη εξεταστικής επιτροπής: Κοκολάκης Σπυρίδων, Καθηγητής
Μαλιάτσος Κωνσταντίνος, Επίκουρος Καθηγητής

Σάμος, Μάρτιος 2025

Η σελίδα αυτή είναι σκόπιμα λευκή.

Πρόλογος και ευχαριστίες

Η παρούσα διπλωματική εργασία αποτελεί το αποτέλεσμα μιας ακαδημαϊκής πορείας που απαιτούσε αφοσίωση και συνεχή προσπάθεια. Θα ήθελα να εκφράσω τις ευχαριστίες μου στην επιβλέπουσα καθηγήτριά μου, Δρ. Διαμαντοπούλου Βασιλική, για την καθοδήγηση και τις πολύτιμες συμβουλές της κατά τη διάρκεια της εκπόνησης αυτής της μελέτης.

Επίσης, ευχαριστώ θερμά τους γονείς μου για τη συνεχή στήριξη που μου παρείχαν σε όλη τη διάρκεια των σπουδών μου. Τέλος, ένα ιδιαίτερο ευχαριστώ στη Δέσποινα Ζωγραφίδου για την υποστήριξη και την κατανόησή της καθ' όλη τη διάρκεια αυτής της διαδικασίας.

© 2025

του

Κατσίγρα Αβραάμ

Τμήμα Μηχανικών Πληροφοριακών και Επικοινωνιακών Συστημάτων

ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΙΓΑΙΟΥ

Η σελίδα αυτή είναι σκόπιμα λευκή

Πίνακας περιεχομένων

Περίληψη

Abstract

1 Εισαγωγή.....	1
1.1 Πλαίσιο Διπλωματικής Εργασίας.....	1
1.2 Σκοπός Διπλωματικής Εργασίας.....	3
1.3 Μεθοδολογία της Διπλωματικής Εργασίας.....	4
1.4 Βιβλιογραφική Ανασκόπηση.....	5
1.4.1 Θεωρητικές Προσεγγίσεις στην Ηθική της Τεχνητής Νοημοσύνης.....	6
1.4.2 Κύριες Ηθικές Προκλήσεις στα Συστήματα ΤΝ.....	6
1.4.3 Ρυθμιστικά Πλαίσια και Κατευθυντήριες Γραμμές.....	7
1.4.4 Συμπερασματική Ανασκόπηση.....	7
1.5 Δομή Διπλωματικής Εργασίας.....	7
2 Θεωρητικό Υπόβαθρο.....	10
2.1 Βασικές Έννοιες και Ορισμοί.....	10
2.2 Βασικές Τεχνολογίες και Εφαρμογές.....	12
2.3 Πρόσφατες Τάσεις ΤΝ και Αυτοματισμού.....	16
2.4 ΤΝ Εφαρμογές και Επιπτώσεις.....	18
3 Τεχνητή Νοημοσύνη και Ηθική.....	20
3.1 Προκλήσεις και ανησυχίες της ΤΝ.....	20
3.2 Εισαγωγή στην Ηθική.....	22
3.3 Ηθική Διάσταση στη Σχεδίαση ΤΝ.....	23
3.4 Ηθικές Αρχές και Απαιτήσεις.....	23
3.5 Η Ανάγκη για Κατευθυντήριες Γραμμές.....	25
4 Πλαίσια, Κατευθυντήριες Γραμμές και η Εφαρμογή τους.....	28
4.1 Διεθνή Πλαίσια και Κατευθυντήριες Γραμμές.....	28
4.1.1 «Δεοντολογικές Οδηγίες για αξιόπιστη Τεχνητή Νοημοσύνη» της Ευρωπαϊκής Επιτροπής (European Commission - EU).....	30
4.1.2 «Ηθικά Ευθυγραμμισμένος Σχεδιασμός» του Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (Institute of Electrical and Electronic Engineers - IEEE).....	32
4.1.3 «Ηθικές αρχές για την Τεχνητή Νοημοσύνη και την Ανάλυση Δεδομένων» του Συνδέσμου Βιομηχανίας Λογισμικού και Πληροφορικής (Software and Information Industry - SIIA).....	33

4.2	Κανονισμοί και Νομικά Πλαίσια.....	35
4.2.1	Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR).....	35
4.2.2	Νομοθεσίες για τη Λογοδοσία και την Ασφάλεια.....	37
4.2.3	Παγκόσμια Συγκριτική Ανάλυση.....	38
4.2.4	Ιδιωτικότητα και Υγεία.....	40
4.2.5	Διασυννοριακή Συμμόρφωση.....	41
4.3	Ηθικές Κατευθυντήριες Γραμμές και Βέλτιστες Πρακτικές.....	42
4.3.1	Μετάφραση Ηθικών Αρχών σε Καθημερινές Πρακτικές.....	42
4.3.2	Παραδείγματα Εφαρμογής.....	42
4.3.3	Βέλτιστες Πρακτικές.....	43
4.3.4	Δικαιοσύνη στην Τεχνητή Νοημοσύνη (AI Fairness).....	43
4.3.5	Τα Πρότυπα IEEE για Ηθική ΤΝ.....	44
4.4	Προκλήσεις στην Εφαρμογή των Κατευθυντήριων Γραμμών.....	45
4.4.1	Τεχνικές Προκλήσεις.....	46
4.4.2	Νομικές Προκλήσεις.....	46
4.4.3	Κοινωνικές και Πολιτισμικές Προκλήσεις.....	47
4.4.4	Επιχειρηματικές Προκλήσεις.....	47
4.5	Προς ένα Παγκόσμιο Ρυθμιστικό Πλαίσιο.....	48
4.5.1	Διαφορετικότητα Προσεγγίσεων ανάμεσα σε Χώρες και Ηπείρους.....	48
4.5.2	Πρωτοβουλίες Διεθνών Οργανισμών.....	49
4.5.3	Προκλήσεις και Προοπτικές Εναρμόνισης.....	50
5	Ηθική Σχεδίαση Αλγορίθμων ΤΝ.....	51
5.1	Πρακτικές σε διάφορα στρώματα.....	51
5.2	Προβλήματα στην εφαρμογή των Πρακτικών.....	53
5.3	Υπόδειγμα Εφαρμογής.....	55
5.3.1	Παρουσίαση της Υποθετικής Εταιρείας "EthicAI Systems".....	55
5.3.2	Στάδια Εφαρμογής των Ηθικών Κατευθυντήριων Γραμμών.....	55
5.3.3	Ανάλυση Επιτυχημένων Case Studies στην Πραγματική Ζωή.....	56
6	Παραδείγματα και Επιπτώσεις.....	58
6.1	Καλά παραδείγματα.....	58
6.2	Προβληματικά παραδείγματα.....	59
6.3	Οι επιπτώσεις.....	60
7	Συμπεράσματα και Προτάσεις.....	61
7.1	Βασικά Συμπεράσματα.....	61
7.2	Προτάσεις για το Μέλλον.....	62

7.3	Τελική Ανασκόπηση.....	63
8	Βιβλιογραφία.....	64

Λίστα Σχημάτων

Σχήμα 1: Κατευθυντήριες γραμμές της ΕΕ ως ένα πλαίσιο για αξιολόγηση ΤΝ

Λίστα Πινάκων

Πίνακας 1: Σύγκριση Τεχνητής Νοημοσύνης και Αυτοματοποίησης.

Πίνακας 2: Σύγκριση Μηχανική Μάθησης και Βαθιά Μάθησης.

Πίνακας 3: Σύγκριση ΕΦΓ και Μεγάλων Γλωσσικών Μοντέλων

Πίνακας 4: Κατευθυντήριες γραμμές ως ένα πλαίσιο για αξιολόγηση ΤΝ

Πίνακας 5: Συγκριτική Ανάλυση των Κανονισμών GDPR, CCPA και PIPL

Λίστα Εικόνων

Εικόνα 1: Τρόπος λειτουργίας ενός συστήματος ΤΝ.

Εικόνα 2: Επίπεδα Κινδύνου Εφαρμογών ΤΝ.

Ακρωνύμια

TN	Τεχνητή Νοημοσύνη	AI	Artificial Intelligence
MM	Μηχανική Μάθηση	ML	Machine Learning
BM	Βαθιά Μάθηση	DL	Deep Learning
ΕΦΓ	Επεξεργασία Φυσικής Γλώσσας	NLP	Natural Language Processing
MGM	Μεγάλα Γλωσσικά Μοντέλα	LLM	Large Language Models
KY	Κβαντική Υπολογιστική	QC	Quantum Computing
YA	Υπολογιστική Αιχμής	EC	Edge Computing
ΕΕ	Ευρωπαϊκή Ένωση	EU	European Union
ΓΚΠΑ	Γενικό Κανονισμό Προστασίας Δεδομένων	GDPR	General Data Protection Regulation
ΙΗΗΜ	Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών	IEEE	Institute of Electrical and Electronic Engineers
ΣΒΑΠ	Σύνδεσμος Βιομηχανίας Λογισμικού και Πληροφορικής	SIIA	Software and Information Industry Association
ΑΜΜ	Αυτόματη Μηχανική Μάθηση	AutoML	Automated Machine Learning
ΠΥ	Προηγμένη Υπολογιστική	AC	Advanced Computing
ΠΤΝ	Πράξη για την Τεχνητή Νοημοσύνη (Κανονισμός της ΕΕ)	AI Act	Artificial Intelligence Act (EU Regulation)
ΟΟΣΑ	Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης	OECD	Organisation for Economic Co-operation and Development
ΟΥΝΕΣΚΟ	Εκπαιδευτικός, Επιστημονικός και Πολιτιστικός Οργανισμός του ΟΗΕ	UNESCO	United Nations Educational, Scientific and Cultural Organization
WEF	Παγκόσμιο Οικονομικό Φόρουμ	WEF	World Economic Forum
G20	Ομάδα των 20 Ισχυρότερων Οικονομιών	G20	Group of Twenty
G7	Ομάδα των 7 Ανεπτυγμένων Οικονομιών	G7	Group of Seven
ISO	Διεθνής Οργανισμός Τυποποίησης	ISO	International Organization for Standardization
FDA	Υπηρεσία Τροφίμων και Φαρμάκων (ΗΠΑ)	FDA	Food and Drug Administration
NIST	Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (ΗΠΑ)	NIST	National Institute of Standards and Technology
IEC	Διεθνής Ηλεκτροτεχνική Επιτροπή	IEC	International Electrotechnical Commission

ECJ	Δικαστήριο της Ευρωπαϊκής Ένωσης	ECJ	European Court of Justice
AI4EU	Πλατφόρμα Τεχνητής Νοημοσύνης της ΕΕ	AI4EU	Artificial Intelligence for the European Union
HLEG AI	Ομάδα Υψηλού Επιπέδου για την ΤΝ	HLEG AI	High-Level Expert Group on AI
ETSI	Ευρωπαϊκό Ινστιτούτο Τηλεπικοινωνιακών Προτύπων	ETSI	European Telecommunications Standards Institute
CEPEJ	Ευρωπαϊκή Επιτροπή για την Αποτελεσματικότητα της Δικαιοσύνης	CEPEJ	European Commission for the Efficiency of Justice
HIPAA	Νόμος για τη Φορητότητα και Λογοδοσία Ασφάλισης Υγείας	HIPAA	Health Insurance Portability and Accountability Act
XAI	Εξηγήσιμη Τεχνητή Νοημοσύνη	XAI	Explainable Artificial Intelligence
GPT	Γενικά Προεκπαιδευμένο Μετασχηματιστικό Μοντέλο	GPT	Generative Pre-trained Transformer
BERT	Διπλής Κατεύθυνσης Κωδικοποιητής Αντιπροσώπευσης από Μετασχηματιστές	BERT	Bidirectional Encoder Representations from Transformers
LIME	Τοπικά Διαφωτισμένη Μοντέλο-εξήγηση	LIME	Local Interpretable Model-Agnostic Explanations
SHAP	Ερμηνεύσιμη Μέθοδος Κατανομής Συνεισφορών	SHAP	SHapley Additive exPlanations
PROV	Στοιχεία Ιχνηλασιμότητας και Διαφάνειας	PROV	Provenance Information
CHI	Διεπαφή Ανθρώπου-Υπολογιστή	CHI	Computer-Human Interaction
CCPA	Κανονισμός Ιδιωτικότητας Καταναλωτών Καλιφόρνιας	CCPA	California Consumer Privacy Act
COMPAS	Σύστημα Πρόβλεψης Εγκληματικής Συμπεριφοράς	COMPAS	Correctional Offender Management Profiling for Alternative Sanctions
CSR	Εταιρική Κοινωνική Ευθύνη	CSR	Corporate Social Responsibility
UNDP	Πρόγραμμα Ανάπτυξης των Ηνωμένων Εθνών	UNDP	United Nations Development Programme
ITU	Διεθνής Ένωση Τηλεπικοινωνιών	ITU	International Telecommunication Union
A/IS	Αυτόνομα και Ευφυή Συστήματα	A/IS	Autonomous and Intelligent Systems
ACM	Ένωση Μηχανικών Υπολογιστών	ACM	Association for Computing Machinery
EAD	Δεοντολογικός Ευθυγραμμισμένος Σχεδιασμός	EAD	Ethically Aligned Design
PHI	Προστασία Υγειονομικών Πληροφοριών	PHI	Protected Health Information

GPT-3	Γενικά Προεκπαιδευμένο Μετασχηματιστικό Μοντέλο 3	GPT-3	Generative Pre-trained Transformer 3
P9_TA- PROV	Αναφορά Πρότασης Νομοθεσίας ΕΕ	P9_TA- PROV	EU Legislative Proposal Report

Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει τις ηθικές και νομικές προκλήσεις που εγείρονται από την ανάπτυξη και εφαρμογή της τεχνητής νοημοσύνης (TN), εστιάζοντας στην επίδρασή της σε κρίσιμους τομείς όπως η ιατρική, η αυτοκινητοβιομηχανία, η δικαιοσύνη και τα χρηματοοικονομικά. Ιδιαίτερη έμφαση δίνεται σε ηθικά διλήμματα όπως η διαφάνεια των αλγορίθμων, η προκατάληψη στα δεδομένα, η ασφάλεια και προστασία της ιδιωτικότητας, καθώς και η αυτονομία της TN στη λήψη αποφάσεων.

Επιπλέον, η εργασία αναλύει τις υπάρχουσες νομικές ρυθμίσεις που στοχεύουν στη διαχείριση αυτών των ζητημάτων, εξετάζοντας διεθνή και εθνικά νομικά πλαίσια, καθώς και προτάσεις για τη βελτίωση της σχετικής νομοθεσίας. Στόχος είναι η ανάδειξη του ρόλου της ηθικής και νομικής ρύθμισης στη βιώσιμη ανάπτυξη της TN, καθώς και η παρουσίαση προτάσεων για την ενίσχυση της διαφάνειας, της λογοδοσίας και της προστασίας των δικαιωμάτων των πολιτών.

Μέσα από αυτήν την ανάλυση, η εργασία φιλοδοξεί να συμβάλει στον ακαδημαϊκό διάλογο για την υπεύθυνη χρήση της TN και να προτείνει στρατηγικές για την καλύτερη διαχείριση των κινδύνων που αυτή συνεπάγεται.

Λέξεις Κλειδιά: *Τεχνητή Νοημοσύνη, Μηχανική Μάθηση, Δικαιοσύνη, Ευθύνη, Διαφάνεια, Ασφάλεια, Απόρρητο, Αλγοριθμική Προκατάληψη, Νομοθετική Ρύθμιση.*

Abstract

This thesis examines the ethical and legal challenges arising from the development and implementation of artificial intelligence (AI), focusing on its impact across critical sectors such as healthcare, the automotive industry, justice, and finance. Special attention is given to ethical dilemmas such as algorithmic transparency, data bias, security and privacy protection, as well as AI's autonomy in decision-making.

Furthermore, the thesis analyzes existing legal frameworks designed to address these issues, exploring both international and national regulatory approaches, while also proposing improvements to current legislation. The objective is to highlight the role of ethical and legal governance in the sustainable development of AI and to present recommendations for enhancing transparency, accountability, and the protection of citizens' rights.

Through this analysis, the research aims to contribute to the academic discussion on the responsible use of AI and suggest strategies for better managing the risks associated with its deployment.

Keywords: *Artificial Intelligence, Machine Learning, Justice, Accountability, Transparency, Security, Privacy, Algorithmic Bias, Legal Regulation*

1 Εισαγωγή

1.1 Πλαίσιο Διπλωματικής Εργασίας

Η Τεχνητή Νοημοσύνη (TN) αποτελεί ένα από τα πλέον καινοτόμα και ραγδαία αναπτυσσόμενα πεδία της σύγχρονης επιστήμης και τεχνολογίας. Επηρεάζει ένα ευρύ φάσμα τομέων, από την υγειονομική περίθαλψη και τη βιομηχανία μέχρι τη χρηματοοικονομική ανάλυση και τη νομική επιστήμη^[24]. Σύμφωνα με τον Marvin Minsky, έναν από τους πρωτοπόρους της έρευνας στην TN, «η Τεχνητή Νοημοσύνη είναι η επιστήμη του να κάνεις τις μηχανές να εκτελούν λειτουργίες που, όταν επιτελούνται από ανθρώπους, απαιτούν ευφυΐα» (Minsky, 1986).

Οι σύγχρονες εφαρμογές της TN περιλαμβάνουν την αναγνώριση προσώπων, την αυτοματοποίηση σύνθετων διαδικασιών, τη φυσική γλωσσική επεξεργασία (Natural Language Processing – NLP) και την ανάλυση μεγάλων δεδομένων (Big Data Analytics). Ωστόσο, καθώς η ανάπτυξη των συστημάτων TN εξελίσσεται, προκύπτουν πολυάριθμα ζητήματα που σχετίζονται με την ηθική, τη νομική ρύθμιση και τη διαφάνεια των αλγορίθμων^{[8][14][9]}.

Κατηγοριοποίηση της Τεχνητής Νοημοσύνης

Η TN διακρίνεται σε τρεις βασικές κατηγορίες, καθεμία από τις οποίες αντιπροσωπεύει διαφορετικά επίπεδα γνωστικής ικανότητας και αυτονομίας ^{[1][5]}:

- 1. Στενή TN (Narrow AI):** Περιλαμβάνει εξειδικευμένα συστήματα που εκτελούν συγκεκριμένες, προδιαγεγραμμένες λειτουργίες μέσω αλγοριθμικής επεξεργασίας. Παραδείγματα περιλαμβάνουν την αναγνώριση αντικειμένων σε εικόνες, τα συστήματα αυτόματης μετάφρασης και τους αλγορίθμους προτάσεων περιεχομένου (Netflix, YouTube)^{[6][17]}. Αυτά τα συστήματα είναι εξαιρετικά αποδοτικά στις προκαθορισμένες λειτουργίες τους, αλλά στερούνται κατανόησης και προσαρμοστικότητας.
- 2. Γενική TN (General AI):** Αυτή η μορφή TN, που βρίσκεται ακόμη σε πειραματικό στάδιο, επιδιώκει να μιμηθεί την ανθρώπινη σκέψη και μάθηση. Περιλαμβάνει αλγορίθμους που μπορούν να αναγνωρίζουν συναισθήματα, να διαχειρίζονται πολύπλοκες καταστάσεις και να προσαρμόζονται σε νέα δεδομένα χωρίς προγραμματισμένη καθοδήγηση^{[4][16]}.
- 3. Υπερ-TN (Super AI):** Πρόκειται για ένα υποθετικό στάδιο ανάπτυξης, όπου τα συστήματα TN θα ξεπερνούσαν την ανθρώπινη νοημοσύνη σε όλους τους τομείς. Αυτή η ιδέα έχει προκαλέσει έντονες συζητήσεις σχετικά με τις επιπτώσεις και τους κινδύνους, καθώς η αυτονομία της TN ενδέχεται να θέσει προκλήσεις για τον ανθρώπινο έλεγχο^{[9][18]}.

Παρόλο που η Γενική και η Υπερ-TN παραμένουν θεωρητικές έννοιες, η Στενή TN εφαρμόζεται ευρέως σε πολλούς τομείς της καθημερινής ζωής, συμπεριλαμβανομένης της βιομηχανίας, της ιατρικής διάγνωσης, των αυτόνομων οχημάτων και των συστημάτων κυβερνοασφάλειας^{[10][12][15]}.

Ηθικές και Νομικές Προκλήσεις της Τεχνητής Νοημοσύνης

Η εξέλιξη και η ολοένα αυξανόμενη χρήση της TN γεννούν σημαντικά ηθικά και νομικά ζητήματα, τα οποία αποτελούν αντικείμενο διεθνούς προβληματισμού^{[8][14][18]}. Τα πιο καίρια ζητήματα περιλαμβάνουν:

- **Αυτοματοποίηση και ανεργία:** Η αντικατάσταση εργαζομένων από αυτοματοποιημένα συστήματα TN έχει εγείρει ανησυχίες για την απώλεια θέσεων εργασίας, ειδικά σε τομείς που βασίζονται σε επαναλαμβανόμενες εργασίες^[5].
- **Προσωπικό απόρρητο και προστασία δεδομένων:** Τα συστήματα TN απαιτούν μεγάλες ποσότητες δεδομένων για να λειτουργήσουν αποτελεσματικά, γεγονός που εγείρει ερωτήματα σχετικά με την προστασία της ιδιωτικότητας και την ασφάλεια των δεδομένων των χρηστών^{[2][19]}.
- **Αλγοριθμική προκατάληψη (Algorithmic Bias):** Οι αλγόριθμοι μηχανικής μάθησης (Machine Learning) συχνά εκπαιδεύονται με δεδομένα που περιέχουν εγγενείς προκαταλήψεις, οδηγώντας σε διακρίσεις, π.χ., στις προσλήψεις ή στη χορήγηση δανείων^{[6][11]}.
- **Νομική ευθύνη και ηθική λογοδοσία:** Σε περιπτώσεις όπου τα συστήματα TN προκαλούν βλάβες (π.χ., ατυχήματα αυτόνομων οχημάτων), τίθεται το ζήτημα της ευθύνης. Πρέπει να θεωρείται υπεύθυνος ο προγραμματιστής, ο κατασκευαστής, ή το ίδιο το σύστημα^{[15][17]}.
- **Πνευματική ιδιοκτησία και πνευματικά δικαιώματα:** Όταν ένας αλγόριθμος δημιουργεί ένα έργο (μουσική, κείμενο, εικόνες), τίθεται το ερώτημα της ιδιοκτησίας και της προστασίας πνευματικών δικαιωμάτων. Το νομικό πλαίσιο δεν έχει ακόμη πλήρως προσαρμοστεί σε τέτοιες περιπτώσεις^[20].

Ρυθμιστικά Πλαίσια και Διεθνείς Προσεγγίσεις

Λόγω της πολυπλοκότητας των ηθικών και νομικών ζητημάτων που σχετίζονται με την TN, έχουν διαμορφωθεί διάφορες στρατηγικές ρύθμισης σε διεθνές επίπεδο. Μερικές από τις πιο σημαντικές περιλαμβάνουν:

- **Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) της Ευρωπαϊκής Ένωσης:** Επιβάλλει αυστηρούς κανόνες στη συλλογή και επεξεργασία δεδομένων από συστήματα TN, προστατεύοντας την ιδιωτικότητα των πολιτών^[19].
- **Η Πρωτοβουλία για Ηθική TN της UNESCO (2021):** Στοχεύει στη διαμόρφωση παγκόσμιων ηθικών κατευθυντήριων γραμμών για τη χρήση της TN^[12].
- **Οι Κανονισμοί για την TN της Ευρωπαϊκής Επιτροπής (2023):** Περιλαμβάνουν προτάσεις για την κατηγοριοποίηση της TN ανάλογα με το επίπεδο κινδύνου και τη θέσπιση υποχρεωτικών αξιολογήσεων ασφαλείας^[25].

Η ανάπτυξη της Τεχνητής Νοημοσύνης συνοδεύεται από σημαντικές προκλήσεις που απαιτούν διεπιστημονική προσέγγιση. Η διερεύνηση των ηθικών και νομικών ζητημάτων είναι κρίσιμη για

την αποτελεσματική ρύθμιση της ΤΝ, ώστε να διασφαλιστεί η διαφάνεια, η ασφάλεια και η υπεύθυνη χρήση της τεχνολογίας^{[3][4][7]}.

1.2 Σκοπός Διπλωματικής Εργασίας

Στην προσπάθεια για ανάπτυξη ηθικών συστημάτων ΤΝ δημιουργούνται εντάσεις μεταξύ των θεωρητικών πλαισίων που σκοπεύουν να καλύψουν όλο το φάσμα της Τεχνητής Νοημοσύνης με την πρακτική εφαρμογή τους που απαιτεί σαφείς οδηγίες, στοχευμένες στο συγκεκριμένο σύστημα που αναπτύσσεται^{[3][4]}. Οι εντάσεις αυτές είναι αναμενόμενες, εφόσον δεν παράγουν όλα τα συστήματα το ίδιο αποτέλεσμα, δεν αναλύουν τα ίδια δεδομένα και δε φέρουν το ίδιο ρίσκο σε όλο το φάσμα των ηθικών απαιτήσεων^[9]. Για παράδειγμα, η ανθρώπινη αυτενέργεια είναι ιδιαίτερα σημαντική σε ένα αυτόνομο όχημα το οποίο για τον οποιονδήποτε λόγο μπορεί να σφάλλει και χρειάζεται την άμεση επέμβαση του οδηγού για την αποφυγή ατυχήματος^[15]. Ένα σύστημα προτάσεων προϊόντων ή υπηρεσιών σχετικά με τα ενδιαφέροντα του καταναλωτή θέτει περισσότερο βάρος στην ιδιωτικότητα του χρήστη παρά στην αυτενέργεια του, αν και πρέπει να σημειωθεί ότι και εδώ είναι σημαντική^[2].

Η διαδικασία ανάπτυξης της ΤΝ απαιτεί εξαιρετική κατανόηση των προβλημάτων που ενδέχεται να εμφανιστούν και να ληφθούν οι απαραίτητες ενέργειες για την αποφυγή σφάλματος ή κακοχρησίας^[12]. Τα θεωρητικά πλαίσια που έχουν προταθεί χρησιμεύουν στην αναγνώριση των δυνητικών προβλημάτων κατά τη διάρκεια του αρχικού σχεδιασμού του συστήματος^[5]. Τα πρακτικά βήματα αντλούνται από τα θεωρητικά πλαίσια και η χρησιμότητά τους φαίνεται στην ανάπτυξη του συστήματος και στην τελική υλοποίησή του^[7]. Βέβαια, οι προτάσεις για ηθική δεν αρκούν, αν και είναι απαραίτητες. Καθίσταται λοιπόν υποχρεωτική η θέση του νομικού πλαισίου για την εγγύηση της υπευθυνότητας που φέρει ο κατασκευαστής του συστήματος απέναντι στον τελικό χρήστη^{[19][20]}. Εκεί προκύπτει μία ακόμη ένταση: η ένταση μεταξύ του νομικού πλαισίου που μέσω ελέγχου εγγυάται για την ασφάλεια του πολίτη και του οργανισμού που επιθυμεί να αναπτύξει και να εξάγει ένα ανταγωνιστικό προϊόν στην αγορά χωρίς εμπόδια^[25].

Η κατανόηση και ανάλυση των ηθικών ζητημάτων της ΤΝ και η επίλυση των εντάσεων που αναφέρθηκαν παραπάνω θα επιτρέψουν στην τεχνολογία αυτή να συνεχίσει την ήδη ραγδαία εξέλιξή της με κύριο σκοπό τη βελτίωση της ανθρώπινης ζωής^[8]. Η εργασία αυτή αποσκοπεί:

1. **Εμβάθυνση στην Κατανόηση της ΤΝ:** Παρουσίαση και ανάλυση των διαφορετικών τύπων ΤΝ και των εφαρμογών τους, με ιδιαίτερη έμφαση στη στενή ΤΝ που χρησιμοποιείται ευρέως στη βιομηχανία, την υγεία, την αυτοκινητοβιομηχανία και άλλους τομείς^{[1][6]}.
2. **Ανάλυση Ηθικών Διλημάτων:** Εξέταση των ηθικών ζητημάτων που ανακύπτουν από τη χρήση της ΤΝ, όπως η προστασία της ιδιωτικότητας, η αυτονομία των μηχανών, η δικαιοσύνη στις αποφάσεις των αλγορίθμων και η ασφάλεια των συστημάτων ΤΝ^{[14][16]}.
3. **Εξέταση Νομικών Πλαισίων:** Διερεύνηση των υφιστάμενων και προτεινόμενων νομικών ρυθμίσεων που αφορούν την ΤΝ σε εθνικό και διεθνές επίπεδο. Ανάλυση των κενών στη νομοθεσία και των προκλήσεων που αντιμετωπίζουν οι νομοθέτες στη ρύθμιση της χρήσης της ΤΝ^{[19][23]}.

4. **Προτάσεις για τη Διαμόρφωση Πολιτικών:** Ανάπτυξη προτάσεων για τη βελτίωση των νομικών και κανονιστικών πλαισίων, με στόχο την προαγωγή μιας ασφαλούς, δίκαιης και ηθικά αποδεκτής χρήσης της ΤΝ. Προτάσεις για τη θέσπιση κατευθυντήριων γραμμών και προτύπων που θα διευκολύνουν την ενσωμάτωση των ηθικών αρχών στη σχεδίαση και την ανάπτυξη των συστημάτων ΤΝ^{[11][22]}.
5. **Ευαισθητοποίηση και Δημόσια Ενημέρωση:** Συμβολή στην ενημέρωση και ευαισθητοποίηση του κοινού, των επαγγελματιών του κλάδου και των νομοθετών σχετικά με τις ηθικές και νομικές διαστάσεις της ΤΝ. Παρουσίαση των πιθανών κινδύνων και προκλήσεων που προκύπτουν από τη χρήση της ΤΝ και προτάσεις για την αντιμετώπισή τους^{[13][24]}.

Αυτή η διπλωματική εργασία αποσκοπεί να συμβάλλει στην κρίσιμη συζήτηση γύρω από την τεχνητή νοημοσύνη, προσδιορίζοντας τρόπους για τη διαχείριση των τεχνολογικών καινοτομιών με έναν τρόπο που σέβεται τις ηθικές αρχές και προωθεί την κοινωνική ευημερία^[17]. Η εργασία αυτή ελπίζει να συνεισφέρει στη διαμόρφωση μιας πιο υπεύθυνης μελλοντικής προοπτικής στον τομέα της ΤΝ^[18].

1.3 Μεθοδολογία της Διπλωματικής Εργασίας

Η μεθοδολογία της παρούσας διπλωματικής εργασίας βασίζεται σε μια συνδυαστική προσέγγιση που περιλαμβάνει βιβλιογραφική έρευνα, συγκριτική ανάλυση ρυθμιστικών και ηθικών πλαισίων και τη μελέτη περιπτώσεων (case studies) για την κατανόηση της εφαρμογής των αρχών της ηθικής στην Τεχνητή Νοημοσύνη (ΤΝ). Η προσέγγιση αυτή επιτρέπει μια ολοκληρωμένη ανάλυση των ηθικών, νομικών και κοινωνικών προκλήσεων που σχετίζονται με την ανάπτυξη και εφαρμογή των συστημάτων ΤΝ.

Βιβλιογραφική Ανασκόπηση

Η αρχική φάση της έρευνας περιλάμβανε μια συστηματική ανασκόπηση της βιβλιογραφίας, με στόχο τη χαρτογράφηση των θεωρητικών προσεγγίσεων και των ηθικών αρχών που διέπουν την ανάπτυξη και εφαρμογή της ΤΝ. Για την ανάλυση αυτή χρησιμοποιήθηκαν:

- **Ακαδημαϊκές δημοσιεύσεις και άρθρα** σχετικά με την ηθική της ΤΝ και τις προκλήσεις της εφαρμογής της.
- **Διεθνή ρυθμιστικά πλαίσια και κατευθυντήριες γραμμές** (π.χ. EU AI Act, GDPR, IEEE Ethically Aligned Design).
- **Μελέτες περιπτώσεων** από τομείς όπως η υγεία, η δικαιοσύνη και η απασχόληση, όπου η ΤΝ έχει ήδη επηρεάσει την καθημερινή ζωή.

Συγκριτική Ανάλυση Ρυθμιστικών και Ηθικών Πλαισίων

Για την αξιολόγηση των υπαρχόντων κατευθυντήριων γραμμών, πραγματοποιήθηκε συγκριτική ανάλυση των ηθικών και νομικών πλαισίων που έχουν προταθεί από διάφορους διεθνείς οργανισμούς και κυβερνήσεις. Η ανάλυση αυτή περιλαμβάνει:

- Σύγκριση των κανονισμών της Ευρωπαϊκής Ένωσης (EU AI Act) με αντίστοιχες πρωτοβουλίες στις Ηνωμένες Πολιτείες και την Ασία.
- Αξιολόγηση της αποτελεσματικότητας των ηθικών κατευθυντήριων γραμμών μέσω της εφαρμογής τους σε πραγματικά παραδείγματα.
- Ανάλυση πιθανών νομικών και ηθικών κενών, που μπορεί να οδηγήσουν σε καταχρήσεις ή ανεπαρκή ρύθμιση των συστημάτων ΤΝ.

Μελέτες Περίπτωσης (Case Studies)

Η έρευνα συμπεριέλαβε μελέτες περίπτωσης, στις οποίες αναλύθηκαν εφαρμογές ΤΝ που προκάλεσαν ηθικά ζητήματα ή πέτυχαν τη βέλτιστη ενσωμάτωση ηθικών αρχών. Ενδεικτικά παραδείγματα:

- **IBM Watson for Oncology:** Θέματα διαφάνειας και ανακρίβειας στις διαγνώσεις.
- **DeepMind & NHS:** Νομικά ζητήματα σχετικά με τη χρήση δεδομένων ασθενών.
- **Amazon Rekognition:** Ζητήματα αλγοριθμικής προκατάληψης στην αναγνώριση προσώπου.

Ποιοτική Ανάλυση και Συμπεράσματα

Μετά τη συλλογή και ανάλυση των δεδομένων, ακολούθησε ποιοτική αξιολόγηση των ευρημάτων, με στόχο:

- Την αναγνώριση των βασικών ηθικών ζητημάτων που προκύπτουν από τη χρήση της ΤΝ.
- Την αξιολόγηση της αποτελεσματικότητας των υπαρχόντων κανονισμών και ηθικών πλαισίων.
- Την πρόταση μέτρων και κατευθυντήριων γραμμών που θα μπορούσαν να βελτιώσουν τη ρύθμιση και την ηθική εφαρμογή της ΤΝ.

Η μεθοδολογική αυτή προσέγγιση επιτρέπει τη δημιουργία μιας ολοκληρωμένης και τεκμηριωμένης ανάλυσης, παρέχοντας τόσο θεωρητικά όσο και πρακτικά εργαλεία για τη μελέτη της ηθικής στη σχεδίαση εφαρμογών και συστημάτων Τεχνητής Νοημοσύνης.

1.4 Βιβλιογραφική Ανασκόπηση

Η ηθική στη σχεδίαση εφαρμογών και συστημάτων Τεχνητής Νοημοσύνης (ΤΝ) αποτελεί ένα ζήτημα που απασχολεί έντονα την επιστημονική κοινότητα τα τελευταία χρόνια^{[8][9]}. Η ραγδαία εξέλιξη των αλγορίθμων μηχανικής μάθησης και η αυξανόμενη διείσδυση της ΤΝ σε κρίσιμους τομείς, όπως η υγεία, η δικαιοσύνη και η χρηματοοικονομική ανάλυση, έχουν οδηγήσει στην ανάγκη καθορισμού ηθικών αρχών και κατευθυντήριων γραμμών για τη διασφάλιση της δικαιοσύνης, της διαφάνειας και της υπευθυνότητας των συστημάτων αυτών^{[12][14]}. Στην παρούσα ενότητα παρουσιάζονται οι βασικότερες θεωρητικές προσεγγίσεις, οι προκλήσεις που έχουν

εντοπιστεί στη βιβλιογραφία και οι υφιστάμενες κατευθυντήριες γραμμές που διαμορφώνουν το πεδίο της ηθικής στην ΤΝ.

1.4.1 Θεωρητικές Προσεγγίσεις στην Ηθική της Τεχνητής Νοημοσύνης

Οι ερευνητές έχουν αναπτύξει διάφορες θεωρητικές προσεγγίσεις για την κατανόηση των ηθικών προκλήσεων στην ΤΝ. Οι κυριότερες από αυτές περιλαμβάνουν:

- **Δεοντολογικές και Ωφελιμιστικές Προσεγγίσεις:** Σύμφωνα με τη δεοντολογική ηθική (Kant, 1785), οι πράξεις πρέπει να βασίζονται σε ηθικούς κανόνες και αρχές ανεξάρτητα από τις συνέπειές τους. Στον αντίποδα, η ωφελιμιστική ηθική (Bentham & Mill, 1863) αξιολογεί τις πράξεις με βάση το συνολικό όφελος που παράγουν. Στην ΤΝ, η δεοντολογική προσέγγιση υποστηρίζει ότι οι αλγόριθμοι πρέπει να σχεδιάζονται με βάση ηθικά πλαίσια, ενώ η ωφελιμιστική αντίληψη εστιάζει στην αξιολόγηση των αποτελεσμάτων τους^{[4][10]}.
- **Ηθική της Αρετής:** Η προσέγγιση αυτή (Αριστοτέλης, 350 π.Χ.) δίνει έμφαση στις ηθικές αρετές των σχεδιαστών και των χρηστών της ΤΝ, προτείνοντας ότι οι αλγόριθμοι πρέπει να αναπτύσσονται από άτομα με υψηλές ηθικές αξίες^[15].
- **Ηθική των Δικαιωμάτων και της Ιδιωτικότητας:** Δεδομένης της ευρείας χρήσης προσωπικών δεδομένων από τα συστήματα ΤΝ, πολλοί ερευνητές τονίζουν τη σημασία της προστασίας των δικαιωμάτων των χρηστών και της διασφάλισης της ιδιωτικότητας τους^{[2][17]}.

1.4.2 Κύριες Ηθικές Προκλήσεις στα Συστήματα ΤΝ

Η βιβλιογραφία αναγνωρίζει διάφορες ηθικές προκλήσεις που ανακύπτουν κατά τη σχεδίαση και χρήση συστημάτων ΤΝ. Οι σημαντικότερες από αυτές περιλαμβάνουν:

- **Αλγοριθμική Προκατάληψη (Algorithmic Bias):** Πολλές μελέτες έχουν δείξει ότι οι αλγόριθμοι μηχανικής μάθησης ενδέχεται να αναπαράγουν ή και να ενισχύουν κοινωνικές ανισότητες^{[6][11]}. Για παράδειγμα, τα συστήματα πρόβλεψης εγκληματικότητας έχουν κατηγορηθεί για φυλετική προκατάληψη^[2].
- **Διαφάνεια και Επεξηγησιμότητα (Explainability & Transparency):** Οι αλγόριθμοι ΤΝ, ιδιαίτερα οι βαθιές νευρωνικές δομές, συχνά λειτουργούν ως «μαύρα κουτιά», δυσχεραίνοντας την ερμηνεία των αποφάσεών τους^[17]. Αυτό αποτελεί σημαντικό ζήτημα σε εφαρμογές όπου η λήψη αποφάσεων πρέπει να είναι κατανοητή από τους ανθρώπους^[6].
- **Ιδιωτικότητα και Ασφάλεια Δεδομένων:** Η χρήση μεγάλων όγκων προσωπικών δεδομένων από εφαρμογές ΤΝ εγείρει σοβαρά ζητήματα ιδιωτικότητας^{[19][20]}. Οι τεχνολογίες ανωνυμοποίησης και κρυπτογράφησης έχουν προταθεί ως λύσεις για την προστασία των δεδομένων^[9].
- **Αυτόνομη Λήψη Αποφάσεων και Ευθύνη (Accountability & Responsibility):** Ποιος είναι υπεύθυνος όταν ένα σύστημα ΤΝ λαμβάνει λανθασμένες ή ανήθικες αποφάσεις; Η ευθύνη μπορεί να αποδοθεί στον προγραμματιστή, στον οργανισμό που ανέπτυξε την ΤΝ ή στο ίδιο το σύστημα, εάν λειτουργεί αυτόνομα^{[5][13]}.

1.4.3 Ρυθμιστικά Πλαίσια και Κατευθυντήριες Γραμμές

Για την αντιμετώπιση των παραπάνω προκλήσεων, έχουν προταθεί διάφορα ρυθμιστικά πλαίσια και ηθικές κατευθυντήριες γραμμές σε διεθνές επίπεδο:

- **Κατευθυντήριες Αρχές της Ευρωπαϊκής Ένωσης για την Αξιόπιστη ΤΝ:** Η Ευρωπαϊκή Επιτροπή (2019) δημοσίευσε έναν οδηγό που προτείνει βασικές αρχές για την ανάπτυξη αξιόπιστης ΤΝ, όπως η ανθρώπινη επίβλεψη, η διαφάνεια, η ασφάλεια και η ιδιωτικότητα^[19].
- **IEEE Ethically Aligned Design:** Το Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE) έχει προτείνει ένα σύνολο ηθικών κατευθυντήριων αρχών που επικεντρώνονται στη δικαιοσύνη, την υπευθυνότητα και τη βιωσιμότητα στη σχεδίαση της ΤΝ^[7].
- **OECD Principles on AI:** Ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (OECD, 2019) έχει προτείνει πέντε αρχές για υπεύθυνη ανάπτυξη της ΤΝ, συμπεριλαμβανομένης της διαφάνειας και της δικαιοσύνης^[23].
- **AI Act της Ευρωπαϊκής Ένωσης:** Πρόκειται για ένα προτεινόμενο νομικό πλαίσιο που θα ρυθμίζει τη χρήση της ΤΝ στην Ευρώπη, δίνοντας ιδιαίτερη έμφαση σε ζητήματα ασφάλειας, διαφάνειας και ηθικής ευθύνης^[25].

1.4.4 Συμπερασματική Ανασκόπηση

Από τη βιβλιογραφική ανασκόπηση προκύπτει ότι η ηθική στη σχεδίαση και χρήση της Τεχνητής Νοημοσύνης είναι ένα πολυδιάστατο ζήτημα που απαιτεί διεπιστημονική προσέγγιση^{[3][10]}. Οι ερευνητές και οι φορείς χάραξης πολιτικής έχουν διαμορφώσει θεωρητικά και πρακτικά πλαίσια για την ενίσχυση της ηθικής διάστασης των εφαρμογών ΤΝ, ωστόσο, υπάρχουν ακόμη σημαντικές προκλήσεις που απαιτούν περαιτέρω διερεύνηση^{[8][14]}. Η ανάπτυξη διαφανών και υπεύθυνων συστημάτων ΤΝ αποτελεί έναν από τους μεγαλύτερους στόχους της σύγχρονης έρευνας στον τομέα, με σκοπό τη δημιουργία τεχνολογιών που θα ωφελούν την κοινωνία, ενώ παράλληλα θα σέβονται τις θεμελιώδεις ηθικές αξίες^[18].

1.5 Δομή Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία είναι οργανωμένη σε έξι κύρια κεφάλαια, προκειμένου να παρέχει μια ολοκληρωμένη ανάλυση της ηθικής διάστασης στη σχεδίαση εφαρμογών και συστημάτων Τεχνητής Νοημοσύνης (ΤΝ). Κάθε κεφάλαιο εστιάζει σε διαφορετικές πτυχές του θέματος, συνδέοντας θεωρητικά πλαίσια, εφαρμογές, κανονιστικά πρότυπα και ηθικά ζητήματα που προκύπτουν από τη χρήση της ΤΝ^{[3][8][14]}. Ακολουθεί μια σύντομη περιγραφή της δομής της εργασίας:

Θεωρητικό Υπόβαθρο

Στο κεφάλαιο αυτό αναλύονται οι βασικές έννοιες που σχετίζονται με την Τεχνητή Νοημοσύνη, η εξέλιξή της και οι βασικές τεχνολογίες που χρησιμοποιούνται στη σχεδίαση των συστημάτων ΤΝ^{[4][10]}. Εξετάζονται επίσης οι πρόσφατες τάσεις στον τομέα της ΤΝ και η αλληλεπίδρασή της με τις σύγχρονες τεχνολογίες αυτοματισμού^[20].

Τεχνητή Νοημοσύνη και Ηθική

Το τρίτο κεφάλαιο επικεντρώνεται στις ηθικές προκλήσεις που σχετίζονται με τη χρήση της ΤΝ. Αναλύονται οι βασικοί κίνδυνοι, όπως η αλγοριθμική προκατάληψη, η ιδιωτικότητα των δεδομένων, η ασφάλεια και η ανάγκη για διαφάνεια στη λήψη αποφάσεων^{[2][6]}. Εξετάζεται επίσης η έννοια της ηθικής στον σχεδιασμό συστημάτων ΤΝ και οι βασικές αρχές που πρέπει να ακολουθούνται^{[9][11]}.

Πλαίσια, Κατευθυντήριες Γραμμές και η Εφαρμογή τους

Στο κεφάλαιο αυτό παρουσιάζονται τα υφιστάμενα ηθικά και νομικά πλαίσια που ρυθμίζουν τη χρήση της ΤΝ σε διεθνές επίπεδο^{[15][23]}. Συγκεκριμένα, εξετάζονται οι κατευθυντήριες γραμμές της Ευρωπαϊκής Ένωσης, του Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE) και άλλων διεθνών οργανισμών^{[7][19]}. Παράλληλα, γίνεται σύγκριση διαφορετικών προσεγγίσεων που έχουν υιοθετηθεί από κυβερνήσεις και οργανισμούς^[25].

Ηθική Σχεδίαση Αλγορίθμων ΤΝ

Το πέμπτο κεφάλαιο εστιάζει στις πρακτικές που εφαρμόζονται στη σχεδίαση ηθικών συστημάτων ΤΝ. Εξετάζονται μεθοδολογίες για τη μείωση των αλγοριθμικών προκαταλήψεων, η διαφάνεια στη λήψη αποφάσεων και οι βέλτιστες πρακτικές για την ενσωμάτωση ηθικών αρχών στον σχεδιασμό των αλγορίθμων^{[6][12]}. Οι προσεγγίσεις αυτές είναι κρίσιμες για τη βελτίωση της αξιοπιστίας και της κοινωνικής αποδοχής των εφαρμογών ΤΝ^[13].

Παραδείγματα και Επιπτώσεις

Στο κεφάλαιο αυτό παρουσιάζονται πρακτικές εφαρμογές της ΤΝ και οι επιπτώσεις τους στην κοινωνία^{[5][18]}. Αναλύονται θετικά παραδείγματα όπου η ηθική σχεδίαση έχει συμβάλει στη βελτίωση της αξιοπιστίας και της αποδοτικότητας των συστημάτων, καθώς και περιπτώσεις όπου η έλλειψη ηθικών κανόνων είχε αρνητικές συνέπειες^[22].

Συμπεράσματα και Προτάσεις

Το τελευταίο κεφάλαιο συνοψίζει τα βασικά ευρήματα της έρευνας και διατυπώνει προτάσεις για την ανάπτυξη ηθικών κατευθυντήριων γραμμών που θα μπορούσαν να συμβάλουν στη βελτίωση της σχεδίασης συστημάτων ΤΝ^[16]. Παράλληλα, επισημαίνονται οι προκλήσεις που παραμένουν ανοιχτές και προτείνονται κατευθύνσεις για μελλοντική έρευνα^[21].

Η συνολική δομή της διπλωματικής εργασίας ακολουθεί μια λογική και συνεκτική προσέγγιση που συνδέει τη θεωρία με την πράξη, προκειμένου να προσφέρει μια ολοκληρωμένη

ανάλυση της ηθικής διάστασης στη σχεδίαση εφαρμογών και συστημάτων Τεχνητής Νοημοσύνης^[8].

2 Θεωρητικό Υπόβαθρο

2.1 Βασικές Έννοιες και Ορισμοί

Εισαγόμενοι βαθύτερα στην ΤΝ καθώς και στα ηθικά ζητήματα που προκύπτουν, είναι αναγκαίο να αναλύσουμε κάποιες βασικές έννοιες και ορισμούς για την ολοκληρωμένη κατανόηση τους^{[1][3][8]}.

Τεχνητή Νοημοσύνη

Η Τεχνητή Νοημοσύνη (Artificial Intelligence - AI) είναι η επιστήμη και η μηχανική που αφορούν τη δημιουργία ευφύων μηχανών, κυρίως μέσω ευφύων προγραμμάτων υπολογιστών^[5]. Η ΤΝ περιλαμβάνει την ανάπτυξη συστημάτων ικανών να εκτελούν καθήκοντα που, εάν εκτελούνταν από ανθρώπους, θα απαιτούσαν νοημοσύνη, όπως η μάθηση, η κατανόηση φυσικής γλώσσας, η αναγνώριση προτύπων και η λήψη αποφάσεων^{[7][10]}. Επομένως, η ΤΝ μιμείται την ανθρώπινη σκέψη αλλά επεξεργάζεται τις πληροφορίες ταχύτερα και ακριβέστερα.

Ο πιο πρόσφατος ορισμός από την αρμόδια επιτροπή του Ευρωπαϊκού Κοινοβουλίου το 2021 περιγράφει τα συστήματα ΤΝ ως:

«Συστήματα που βασίζονται σε λογισμικό ή είναι ενσωματωμένα σε συσκευές υλισμικού και εμφανίζουν συμπεριφορά που προσομοιώνει νοημοσύνη, συλλέγοντας και επεξεργάζοντας δεδομένα, αναλύοντας και ερμηνεύοντας το περιβάλλον τους και αναλαμβάνοντας δράση με κάποιο βαθμό αυτονομίας για την επίτευξη συγκεκριμένων στόχων.»^[19]

Ο όρος "Τεχνητή Νοημοσύνη" πρωτοχρησιμοποιήθηκε το 1956 από τον John McCarthy, έναν από τους θεμελιωτές της ΤΝ, ο οποίος την όρισε ως:

«Την επιστήμη και μηχανική της κατασκευής ευφύων μηχανών, πιο συγκεκριμένα ευφύων προγραμμάτων υπολογιστών.»

Το 1968, ο Marvin Minsky περιέγραψε την ΤΝ ως:

«Την επιστήμη που επιτρέπει στις μηχανές να κάνουν πράγματα τα οποία, αν γίνονταν από ανθρώπους, θα απαιτούσαν νοημοσύνη.»

Μηχανική Μάθηση και Υποπεδία της ΤΝ

Μια σημαντική υποενότητα της ΤΝ είναι η Μηχανική Μάθηση (Machine Learning - ML), που αναφέρεται σε προγράμματα που βελτιώνονται από μόνα τους μέσω εμπειρίας^[6]. Ο καθηγητής Tom Mitchell από το Carnegie Mellon University ορίζει τη ΜΜ ως:

«Ένα πρόγραμμα μαθαίνει από εμπειρία E σε σχέση με κάποια τάξη καθηκόντων T και κάποια απόδοση P , αν η απόδοσή του στα καθήκοντα της τάξης T , όπως μετράται από P , βελτιώνεται με την εμπειρία E .»

Η ΤΝ περιλαμβάνει επίσης την Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP), που επιτρέπει την αναγνώριση και κατανόηση της ανθρώπινης γλώσσας^[9], και την Υπολογιστική Όραση (Computer Vision), που δίνει τη δυνατότητα στις μηχανές να αναγνωρίζουν και να επεξεργάζονται εικόνες^[12].

Τόσο η ΤΝ όσο και η ΜΜ έχουν εξελιχθεί σε θεμελιώδη εργαλεία στον κόσμο της τεχνολογίας, επιτρέποντας ταχεία προσαρμογή και εξέλιξη σε διάφορες καταστάσεις, βελτιώνοντας την απόδοση με την πάροδο του χρόνου^[14].

Ωστόσο, αυτή η ταχεία εξέλιξη και προσαρμοστικότητα ενέχει και κινδύνους, καθώς τα συστήματα ΤΝ μπορεί να γίνουν στόχος κακόβουλων επιθέσεων. Είναι κρίσιμο οι ερευνητές και οι προγραμματιστές να λαμβάνουν υπόψη τους αυτούς τους κινδύνους και να εργάζονται για την ανάπτυξη ασφαλών και αξιόπιστων συστημάτων ΤΝ^[18].

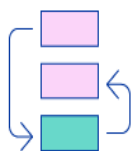
Η ΤΝ αντιπροσωπεύει μια δυναμική και συνεχώς εξελισσόμενη επιστημονική περιοχή με τη δυνατότητα να μεταμορφώσει πολλούς τομείς της ανθρώπινης δραστηριότητας, από την ιατρική και την εκπαίδευση μέχρι την αυτοματοποίηση και την ανάλυση δεδομένων^[22].



1. αναλύει μεγάλο όγκο εισαγόμενων δεδομένων



2. εντοπίζει μοτίβα μέσω αλγορίθμων



3. εκτελεί την εργασία



4. μαθαίνει και βελτιώνει τις επιδόσεις

Εικόνα 1: Τρόπος λειτουργία ενός συστήματος ΤΝ^[20]

Αυτοματοποίηση

Ο όρος αυτοματοποίηση αναφέρεται στη χρήση τεχνολογιών για την εκτέλεση εργασιών χωρίς ανθρώπινη παρέμβαση, στοχεύοντας στην αύξηση της αποδοτικότητας, της ακρίβειας και της ταχύτητας σε διάφορες διαδικασίες^[23].

Ενσωματώνεται σε διάφορους τομείς, όπως:

- **Βιομηχανία** → Ρομποτικά συστήματα και αυτοματοποιημένα μηχανήματα σε γραμμές παραγωγής για βελτίωση της αποδοτικότητας και μείωση του κόστους.

- **Πληροφορική** → Αυτοματοποιημένα συστήματα για διαχείριση δεδομένων, ανάλυση και λήψη αποφάσεων ^[16].

Σύμφωνα με την επιτροπή του Ευρωπαϊκού Κοινοβουλίου το 2021, ένα αυτόνομο σύστημα ΤΝ λειτουργεί:

«Ερμηνεύοντας ορισμένες εισροές και χρησιμοποιώντας μια δέσμη προκαθορισμένων οδηγιών, χωρίς να περιορίζεται σε αυτές, επικεντρώνοντας τη συμπεριφορά του στην επίτευξη του στόχου που έχει οριστεί από τον προγραμματιστή του.» ^[19]

Σύγκριση ΤΝ & Αυτοματοποίηση

Χαρακτηριστικό	Τεχνητή Νοημοσύνη (ΤΝ)	Αυτοματοποίηση
Κύριος στόχος	Δημιουργία ευφυών μηχανών	Εκτέλεση εργασιών χωρίς ανθρώπινη παρέμβαση
Ικανότητες	Μάθηση, λήψη αποφάσεων, ανάλυση δεδομένων	Προκαθορισμένες λειτουργίες
Παραδείγματα	Chatbots, αυτόνομα οχήματα, ανάλυση εικόνων	Ρομποτικά συστήματα, αυτοματοποιημένα scripts
Προσαρμογή	Μαθαίνει και βελτιώνεται με την εμπειρία	Δε μαθαίνει, εκτελεί προκαθορισμένες ενέργειες

Πίνακας 1: Σύγκριση Τεχνητής Νοημοσύνης και Αυτοματοποίησης

Η ΤΝ εστιάζει στη δημιουργία ευφυών μηχανών που έχουν τη δυνατότητα να επιλύουν προβλήματα και να λαμβάνουν αποφάσεις, ενώ η αυτοματοποίηση προσηλώνεται στην αυτόματη διεκπεραίωση καθηκόντων από φυσικές μηχανές και συστήματα, εξοικονομώντας χρόνο και πόρους ^[25].

2.2 Βασικές Τεχνολογίες και Εφαρμογές

Μηχανική Μάθηση και Βαθιά Μάθηση

Η Μηχανική Μάθηση (Machine Learning - ML) είναι ένας κλάδος της ΤΝ που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να βελτιώνουν την απόδοσή τους σε μια εργασία μέσω εμπειρίας^{[6][8]}. Ουσιαστικά, οι αλγόριθμοι μηχανικής μάθησης χρησιμοποιούν δεδομένα για να εκπαιδευτούν και να κάνουν προβλέψεις ή να λάβουν αποφάσεις χωρίς να έχουν προγραμματιστεί ρητά για τη συγκεκριμένη εργασία^[14].

Βασικές Κατηγορίες ΜΜ:

1. **Εποπτευόμενη Μάθηση (Supervised Learning):** Τα μοντέλα εκπαιδεύονται με δεδομένα που περιέχουν εισόδους και τις αντίστοιχες εξόδους.
 - Παράδειγμα: Ανάλυση εικόνας για την αναγνώριση γραμμάτων.^[7]
2. **Μη Εποπτευόμενη Μάθηση (Unsupervised Learning):** Τα μοντέλα εκπαιδεύονται με δεδομένα που περιέχουν μόνο εισόδους χωρίς τις αντίστοιχες εξόδους.

- Παράδειγμα: Ομαδοποίηση (clustering) δεδομένων για τον εντοπισμό κρυφών προτύπων.^[12]

3. Ενισχυτική Μάθηση (Reinforcement Learning): Τα μοντέλα μαθαίνουν μέσω αλληλεπίδρασης με το περιβάλλον τους και προσαρμόζονται με βάση τις ανταμοιβές ή τις ποινές που λαμβάνουν.

- Παράδειγμα: Ένα ρομπότ που μαθαίνει να περπατάει.^[10]

Σημαντικό είναι να δοθεί ιδιαίτερη προσοχή στην υπερεκπαίδευση, ώστε τα μοντέλα να μη μάθουν υπερβολικά τα δεδομένα εκπαίδευσης, καθώς αυτό μπορεί να οδηγήσει σε κακές προβλέψεις για νέα δεδομένα.^[13]

Η Βαθιά Μάθηση (Deep Learning - DL) είναι ένα υποσύνολο της MM που χρησιμοποιεί πολυεπίπεδες τεχνητές νευρωνικές δομές (νευρωνικά δίκτυα) για την ανάλυση και την εκμάθηση από μεγάλα σύνολα δεδομένων^[9]. Το "βάθος" στη BM αναφέρεται στον αριθμό των επιπέδων (layers) στο νευρωνικό δίκτυο^[13].

Τα βασικά χαρακτηριστικά της BM είναι τα εξής:

- 1. Νευρωνικά Δίκτυα (Neural Networks):** Δομές εμπνευσμένες από τον ανθρώπινο εγκέφαλο που αποτελούνται από νευρώνες οργανωμένους σε επίπεδα^[11].
- 2. Εκμάθηση από Μεγάλα Δεδομένα:** Τα νευρωνικά δίκτυα είναι ιδιαίτερα αποτελεσματικά στη μάθηση από μεγάλα σύνολα δεδομένων ανιχνεύοντας αυτόματα χαρακτηριστικά του χωρίς την ανάγκη χειροκίνητης προεπεξεργασίας^[15].
- 3. Εφαρμογές σε Πολύπλοκα Προβλήματα:** Χρησιμοποιείται σε εφαρμογές όπως η αναγνώριση ομιλίας, η επεξεργασία φυσικής γλώσσας, η ανάλυση εικόνας και η αυτόνομη οδήγηση^[16].

Η MM και η BM είναι δύο συναφή πεδία της Τεχνητής Νοημοσύνης (TN) που ασχολούνται με την ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να μαθαίνουν από δεδομένα και να κάνουν προβλέψεις^[13].

- Η BM απαιτεί μεγαλύτερα σύνολα δεδομένων και αυξανόμενους υπολογιστικούς πόρους σε σύγκριση με τη MM, κάτι που προσδίδει μεγαλύτερη πολυπλοκότητα στις διαδικασίες της.^[18]
- Τα μοντέλα βαθιάς μάθησης τείνουν να είναι πιο ακριβή και ευέλικτα, αλλά η ερμηνεία τους συχνά παρουσιάζει μεγαλύτερες δυσκολίες (black box problem).^[20]

Χαρακτηριστικό	Μηχανική Μάθηση (ML)	Βαθιά Μάθηση (DL)
Αλγόριθμοι	Απλούστεροι, όπως δέντρα απόφασης, SVM, κ.ά.	Πολυεπίπεδα νευρωνικά δίκτυα
Απαιτήσεις δεδομένων	Μικρότερα σύνολα δεδομένων	Πολύ μεγάλα σύνολα δεδομένων
Εφαρμογές	Πρόβλεψη καιρού, ανάλυση κειμένου, συστάσεις προϊόντων	Αυτόματη μετάφραση, ανάλυση ιατρικών εικόνων, αυτόνομα οχήματα
Ερμηνευσιμότητα	Ευκολότερη ερμηνεία αποφάσεων	Πολύπλοκες αποφάσεις, δύσκολη εξήγηση

Πίνακας 2: Σύγκριση Μηχανική Μάθησης και Βαθιά Μάθησης

Συμπερασματικά, η BM χρησιμοποιείται κυρίως στην αντιμετώπιση πολύπλοκων προβλημάτων ενώ η MM παραμένει χρήσιμη για πολλές εφαρμογές όπου η απλότητα είναι προτιμότερη^[21].

Παραδείγματα εφαρμογών:

- **Μηχανική Μάθηση:** Πρόβλεψη καιρού, ανάλυση κειμένου, συστάσεις προϊόντων^[17].
- **Βαθιά Μάθηση:** Αυτόματη μετάφραση, ανάλυση ιατρικών εικόνων, αυτόνομα οχήματα^[19].

Η συνεχής πρόοδος στην υπολογιστική ισχύ και τα δεδομένα έχει επιτρέψει τη ραγδαία ανάπτυξη της BM, καθιστώντας την ένα ισχυρό εργαλείο για την ανάλυση μεγάλων δεδομένων και την ανάπτυξη ευφυών εφαρμογών ^[23].

Ρομποτική και Αυτοματισμός

Η χρήση της ρομποτικής και του αυτοματισμού έχει συνεισφέρει σημαντικά στο τρόπο με τον οποίο διεκπεραιώνονται οι εργασίες σε διάφορους τομείς. Οι τεχνολογίες αυτές επιτρέπουν επαναλαμβανόμενες εργασίες χωρίς ανθρώπινη παρέμβαση προσφέροντας αυξημένη απόδοση, παραγωγικότητα και ακρίβεια^{[6][12]}.

Επιπροσθέτως, η χρήση τους έχει μειώσει σημαντικά τους κίνδυνους ανθρώπινων λαθών καθώς και τους κίνδυνους ασφάλειας σε επικίνδυνα περιβάλλοντα εργασίας, όπως εργοστάσια, κατασκευές και χημικά εργαστήρια ^[8].

Η χρήση της ρομποτικής και του αυτοματισμού έχει κατ' επέκταση αποδειχθεί οικονομικά αποδοτική για πολλές επιχειρήσεις, καθώς τους επιτρέπει να παραμείνουν ανταγωνιστικές, μειώνοντας το κόστος παραγωγής και βελτιώνοντας την ποιότητα των προϊόντων^[5].

Παρακάτω, παραθέτουμε μερικά παραδείγματα εφαρμογής των προαναφερόμενων τεχνολογιών, σε διάφορους τομείς:

1. **Βιομηχανικός Τομέας:** Χρησιμοποιούνται ρομπότ στα εργοστάσια και στις γραμμές παραγωγής, για τη συναρμολόγηση προϊόντων, τη συγκόλληση, τη βαφή και την επεξεργασία υλικών. Οι αυτοματοποιημένες γραμμές παραγωγής βελτιώνουν την αποδοτικότητα και μειώνουν το κόστος παραγωγής^{[9][16]}.
2. **Υγειονομικός Τομέας:** Η ρομποτική χειρουργική χρησιμοποιείται σε χειρουργικές επεμβάσεις για μεγαλύτερη ακρίβεια και για λιγότερο επεμβατικές διαδικασίες.
 - **Παράδειγμα:** Το Da Vinci Surgical System που επιτρέπει ακριβείς ελάχιστα επεμβατικές χειρουργικές επεμβάσεις^{[7][18]}.
3. **Γεωργικός Τομέας:** Χρησιμοποιούνται ρομποτικοί συλλέκτες για τη συγκομιδή καρπών και λαχανικών μειώνοντας την ανάγκη για ανθρώπινο δυναμικό και αυξάνοντας την αποδοτικότητα^[10].
 - **Παράδειγμα:** Ρομποτικοί βραχίονες που επιλέγουν φρούτα με ακρίβεια χωρίς να τα καταστρέφουν.
4. **Τομέας Λιανεμπορίου:** Χρησιμοποιούνται αυτοματοποιημένα ταμεία για την ταχύτερη εξυπηρέτηση πελατών.
 - **Παράδειγμα:** Τα αυτόματα ταμεία σε καταστήματα όπως τα Amazon Go, που χρησιμοποιούν αισθητήρες και κάμερες αντί για ταμίες.

5. **Στον τομέα Μεταφορών:** Χρήση Αυτόματων Οχημάτων χωρίς οδηγό για τη μεταφοράς εμπορευμάτων με σκοπό τη βελτίωση της ασφάλειας και τη μείωση του κόστους μεταφοράς^[11].
 - **Παράδειγμα:** Τα αυτόνομα φορτηγά της Tesla και της Waymo, που μειώνουν το κόστος logistics.
6. **Κατασκευαστικός Τομέας:** Χρήση ρομπότ για την εξόρυξη υλικών και την κατασκευή δομών μειώνοντας τον κίνδυνο για τους εργάτες^[15].
7. **Στην Καθημερινή Ζωή:** Υπάρχουν οικιακά ρομπότ όπως οι σκούπες ρομπότ (π.χ. Roomba) που βοηθάνε στις οικιακές εργασίες, καθώς και προσωπικοί βοηθοί, δηλαδή ρομπότ ή ψηφιακοί βοηθοί όπως της Amazon Echo και Google Home, που προσφέρουν βοήθεια στην καθημερινή διαχείριση πληροφοριών και εργασιών^[17].

Επεξεργασία Φυσικής Γλώσσας και Μεγάλα Γλωσσικά Μοντέλα

Ένα από τα σημαντικότερα πεδία της Τεχνητής Νοημοσύνης (TN) αποτελεί η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP). Η ΕΦΓ εστιάζει στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να αλληλεπιδρούν με τους ανθρώπους, αναλύοντας, επεξεργάζοντας και παράγοντας γλωσσικά δεδομένα με τρόπο που προσομοιάζει την ανθρώπινη κατανόηση ^[20].

Η ΕΦΓ έχει σημειώσει σημαντική πρόοδο τα τελευταία χρόνια, ιδιαίτερα με την ανάπτυξη Μεγάλων Γλωσσικών Μοντέλων (LLM - Large Language Models). Αυτά τα μοντέλα, όπως το GPT-4 της OpenAI και το BERT της Google, έχουν φέρει επανάσταση στον τρόπο με τον οποίο οι μηχανές κατανοούν και παράγουν την ανθρώπινη γλώσσα, βελτιώνοντας σημαντικά την ποιότητα της αυτοματοποιημένης επικοινωνίας και των εφαρμογών που βασίζονται στη γλώσσα ^[22].

Τα Μεγάλα Γλωσσικά Μοντέλα είναι προηγμένα συστήματα TN που αξιοποιούν τεράστιες ποσότητες δεδομένων και εξελιγμένους αλγορίθμους για την κατανόηση, την ερμηνεία και τη δημιουργία ανθρώπινης γλώσσας. Χρησιμοποιούν τεχνικές βαθιάς μάθησης (Deep Learning), όπως τα νευρωνικά δίκτυα, για να αναλύουν και να μαθαίνουν από τεράστιους όγκους δεδομένων κειμένου, βελτιώνοντας έτσι την ικανότητά τους να ανταποκρίνονται σε περίπλοκα γλωσσικά ερωτήματα και να δημιουργούν φυσικές και κατανοητές απαντήσεις ^[20].

Χαρακτηριστικό	Επεξεργασία Φυσικής Γλώσσας (NLP)	Μεγάλα Γλωσσικά Μοντέλα (LLMs)
Σκοπός	Κατανόηση και επεξεργασία κειμένου	Δημιουργία ανθρώπινης γλώσσας με μεγαλύτερη ακρίβεια
Μέθοδος	Συμβατικοί NLP αλγόριθμοι	Νευρωνικά δίκτυα μεγάλης κλίμακας
Απαιτήσεις δεδομένων	Μικρότερα σύνολα δεδομένων	Τεράστιες ποσότητες δεδομένων
Εφαρμογές	Ανάλυση κειμένου, αναγνώριση ομιλίας, μετάφραση	Chatbots, σύνοψη κειμένου, δημιουργική γραφή
Ερμηνευσιμότητα	Ευκολότερη ερμηνεία	Δύσκολη εξήγηση αποφάσεων (black box problem)

Πίνακας 3: Σύγκριση ΕΦΓ και Μεγάλων Γλωσσικών Μοντέλων

Οι εφαρμογές της ΕΦΓ είναι ποικίλες και έχουν αντίκτυπο σε διάφορους κλάδους και η ανάπτυξη των ΜΓΜ οδήγησε στη σημαντική βελτίωση της απόδοσης τους. Παρακάτω θα αναφέρουμε κάποια παραδείγματα εφαρμογής τους.

1. **Ανάλυση Συναισθημάτων:** Είναι μία διαδικασία της εξαγωγής συναισθηματικών πληροφοριών από κείμενο. Χρησιμοποιείται ευρέως για την ανάλυση σχολίων χρηστών σε μέσα κοινωνικής δικτύωσης, κριτικών προϊόντων και υπηρεσιών, προσφέροντας πολύτιμες πληροφορίες στις επιχειρήσεις σχετικά με τις προτιμήσεις και τις αντιδράσεις των καταναλωτών [23].
2. **Αυτόματη Μετάφραση:** Οι υπηρεσίες αυτόματης μετάφρασης, όπως το Google Translate, βασίζονται σε τεχνικές ΕΦΓ για τη μετάφραση κειμένων από μία γλώσσα σε άλλη. Οι βελτιώσεις στις τεχνολογίες αυτές έχουν συμβάλει στη διευκόλυνση της επικοινωνίας μεταξύ ατόμων διαφορετικών γλωσσικών υποβάθρων και στην ενίσχυση της διεθνούς συνεργασίας [20].
3. **Αναγνώριση Ομιλίας:** Η αναγνώριση ομιλίας επιτρέπει τη μετατροπή προφορικού λόγου σε γραπτό κείμενο. Εφαρμογές όπως οι ψηφιακοί βοηθοί (π.χ. Siri, Alexa) και τα συστήματα υπαγόρευσης αξιοποιούν την ΕΦΓ για να κατανοούν και να αναλύουν την ομιλία των χρηστών, διευκολύνοντας την αλληλεπίδραση μεταξύ ανθρώπου και μηχανής [20].
4. **Ανάκτηση Πληροφοριών και Αναζήτηση:** Η ΕΦΓ βελτιώνει σημαντικά τα συστήματα ανάκτησης πληροφοριών και τις μηχανές αναζήτησης, επιτρέποντας την πιο ακριβή και αποτελεσματική εξαγωγή σχετικών δεδομένων από μεγάλες βάσεις δεδομένων και το διαδίκτυο [24].
5. **Διάλογος και Chatbots:** Η σύνοψη κειμένου βασίζεται στην ΕΦΓ για την ανάλυση εκτενών εγγράφων και τη δημιουργία περιλήψεων που περιέχουν τις πιο σημαντικές πληροφορίες. Αυτή η τεχνική είναι ιδιαίτερα χρήσιμη στην επεξεργασία ειδήσεων, στην ανασκόπηση επιστημονικών άρθρων και στη διαχείριση μεγάλων όγκων πληροφοριών [22].
6. **Σύνοψη Κειμένου:** Η σύνοψη κειμένου βασίζεται στην ΕΦΓ για την ανάλυση εκτενών εγγράφων και τη δημιουργία περιλήψεων που περιέχουν τις πιο σημαντικές πληροφορίες. Αυτή η τεχνική είναι ιδιαίτερα χρήσιμη στην επεξεργασία ειδήσεων, στην ανασκόπηση επιστημονικών άρθρων και στη διαχείριση μεγάλων όγκων πληροφοριών [22].

Η συνεχής πρόοδος στον τομέα της ΕΦΓ και των Μεγάλων Γλωσσικών Μοντέλων αναμένεται να επιφέρει ακόμη πιο εξελιγμένες εφαρμογές στο μέλλον, βελτιώνοντας την ανθρώπινη-μηχανική επικοινωνία και επεκτείνοντας τις δυνατότητες των αυτόνομων συστημάτων [25].

2.3 Πρόσφατες Τάσεις ΤΝ και Αυτοματισμού

Η Τεχνητή Νοημοσύνη (ΤΝ) και ο αυτοματισμός συγκαταλέγονται μεταξύ των πιο ραγδαία αναπτυσσόμενων τεχνολογικών τάσεων του 21ου αιώνα. Η ταχεία πρόοδός τους έχει μεταμορφώσει ένα ευρύ φάσμα κλάδων, προσφέροντας νέες δυνατότητες, αλλά ταυτόχρονα δημιουργώντας προκλήσεις και ζητήματα που απαιτούν ιδιαίτερη προσοχή. Σε αυτή την ενότητα, αναλύονται

ορισμένες από τις πλέον σημαντικές τάσεις της ΤΝ και του αυτοματισμού, οι εφαρμογές τους και οι επιπτώσεις που επιφέρουν στη σύγχρονη κοινωνία και βιομηχανία.

1. Η **Μηχανική Μάθηση** (Machine Learning - ML) συνεχίζει να εξελίσσεται με ταχύτατους ρυθμούς, με νέους αλγορίθμους και τεχνικές που βελτιώνουν την απόδοση και την ευκολία εφαρμογής των μοντέλων ΤΝ. Μία από τις πιο καινοτόμες εξελίξεις είναι η Αυτόματη Μηχανική Μάθηση (AutoML), η οποία επιτρέπει την αυτοματοποίηση της δημιουργίας, της ρύθμισης και της εκπαίδευσης μοντέλων ML χωρίς την ανάγκη εξειδικευμένης ανθρώπινης παρέμβασης^{[7][17]}. Αυτό καθιστά την ανάπτυξη και χρήση της ΤΝ πιο προσιτή σε οργανισμούς και επαγγελματίες χωρίς προηγμένη γνώση στον τομέα της επιστήμης δεδομένων, διευρύνοντας σημαντικά τη χρήση της σε τομείς όπως η ιατρική διάγνωση, η χρηματοοικονομική ανάλυση και οι έξυπνες υποδομές^[14].
2. Η **Κβαντική Υπολογιστική** (Quantum Computing - QC) εμφανίζεται ως μία από τις πιο επαναστατικές τεχνολογίες του μέλλοντος, προσφέροντας υπολογιστική ισχύ που υπερβαίνει κατά πολύ τις δυνατότητες των παραδοσιακών (κλασικών) υπολογιστών^{[16][20]}. Στο πλαίσιο της ΤΝ, η ΚΥ έχει τη δυνατότητα να αλλάξει ριζικά τον τρόπο εκπαίδευσης και βελτιστοποίησης των αλγορίθμων, επιτρέποντας την ταχύτερη επίλυση πολύπλοκων προβλημάτων, όπως η ανάλυση μεγάλων όγκων δεδομένων και η ανάπτυξη προηγμένων αλγορίθμων μηχανικής μάθησης^[13]. Εφαρμογές της ΚΥ εντοπίζονται ήδη στην κρυπτογραφία, την προσομοίωση μοριακών δομών για φαρμακευτική έρευνα, αλλά και στη βελτιστοποίηση πολύπλοκων δικτύων, όπως τα δίκτυα μεταφορών και ενέργειας.
3. Η **Υπολογιστική Αιχμής** (Edge Computing) είναι μια ακόμη σημαντική τεχνολογική εξέλιξη που επιτρέπει την επεξεργασία δεδομένων σε πραγματικό χρόνο κοντά στην πηγή δημιουργίας τους, αντί να βασίζεται αποκλειστικά σε κεντρικά υπολογιστικά νέφη (cloud computing)^{[12][19]}. Η τεχνολογία αυτή είναι ιδιαίτερα κρίσιμη για εφαρμογές που απαιτούν χαμηλό χρόνο απόκρισης και άμεση λήψη αποφάσεων, όπως η αυτόνομη οδήγηση, τα έξυπνα συστήματα ελέγχου σε εργοστάσια και τα δίκτυα του διαδικτύου των πραγμάτων (IoT). Στον τομέα της αυτόνομης οδήγησης, για παράδειγμα, οι αισθητήρες των οχημάτων πρέπει να επεξεργάζονται δεδομένα τοπικά και σε πραγματικό χρόνο, ώστε να διασφαλίζεται η ασφάλεια των επιβατών και των πεζών^[11]. Επιπλέον, η ΥΑ συμβάλλει στη μείωση του φόρτου δικτύου, καθώς και στη μείωση του ενεργειακού κόστους μεταφοράς δεδομένων.

Ο συνδυασμός των παραπάνω τεχνολογιών, ανοίγει νέους ορίζοντες για δημιουργία και ανάπτυξη καινοτόμων εφαρμογών. Η δυνατότητα προηγμένης ανάλυσης δεδομένων σε πραγματικό χρόνο επιτρέπει τη δημιουργία προβλέψεων που μπορούν να βελτιώσουν τις λήψεις αποφάσεων σε κρίσιμους τομείς, όπως στον υγειονομικό, ενεργειακό και χρηματοοικονομικό τομέα. Οι δυνατότητες που παρέχονται είναι πράγματι απεριόριστες αλλά δε σημαίνει ότι μας προσφέρονται χωρίς περιορισμούς. Ο αυξανόμενος αυτοματισμός και η ολοένα και πιο διαδεδομένη χρήση της ΤΝ εγείρουν σημαντικά ερωτήματα σχετικά με τη δικαιοσύνη, διαφάνεια και την ηθική χρήση αυτών των τεχνολογιών.

2.4 ΤΝ Εφαρμογές και Επιπτώσεις

Η Τεχνητή Νοημοσύνη (ΤΝ) αποτελεί βασικό μοχλό ψηφιακού μετασχηματισμού και καινοτομίας, επιφέροντας σημαντικές αλλαγές στην κοινωνία, τη βιομηχανία και τις επιχειρήσεις^[1]. Με τη βελτίωση της αποδοτικότητας, την αυτοματοποίηση διαδικασιών και την ανάπτυξη έξυπνων λύσεων, η ΤΝ οδηγεί στη δημιουργία νέων δυνατοτήτων, αλλά και προκλήσεων^{[3][6]}. Παρακάτω αναλύονται βασικοί τομείς εφαρμογής της και οι αντίστοιχες επιπτώσεις της.

1. Βιομηχανία και Μεταποίηση

Η ΤΝ έχει μεταμορφώσει τον βιομηχανικό τομέα μέσω της έξυπνης αυτοματοποίησης, της προγνωστικής συντήρησης και της βελτιστοποίησης της παραγωγής^[7]. Στις σύγχρονες έξυπνες γραμμές παραγωγής, αισθητήρες και αλγόριθμοι μηχανικής μάθησης επιτρέπουν την αυτορρύθμιση μηχανημάτων, μειώνοντας τις απώλειες και αυξάνοντας την αποδοτικότητα. Παράλληλα, η χρήση ρομποτικών συστημάτων μειώνει την ανάγκη ανθρώπινης παρέμβασης, αυξάνοντας τη συνέπεια και την παραγωγικότητα^[12].

Επιπτώσεις:

- Μείωση κόστους παραγωγής μέσω αυτοματοποίησης
- Αύξηση ποιότητας προϊόντων μέσω ελέγχου σε πραγματικό χρόνο
- Ενίσχυση της βιωσιμότητας μέσω ενεργειακά αποδοτικών διαδικασιών

2. Ασφάλεια

Η ΤΝ έχει συμβάλλει στη βελτίωση της ασφάλειας μέσω έξυπνων συστημάτων παρακολούθησης, αναγνώρισης προτύπων και προγνωστικής ανάλυσης κινδύνων^[10]. Ειδικότερα στην κυβερνοασφάλεια, τα αλγοριθμικά μοντέλα μπορούν να ανιχνεύσουν κακόβουλες δραστηριότητες και κυβερνοεπιθέσεις σε πραγματικό χρόνο ^[19].

Επιπτώσεις:

- Προστασία προσωπικών δεδομένων και μείωση κινδύνου παραβιάσεων
- Ανίχνευση απειλών πριν την εκδήλωσή τους
- Μείωση του κινδύνου φυσικών ατυχημάτων μέσω ανάλυσης δεδομένων σε βιομηχανικά περιβάλλοντα

3. Εκπαίδευση

Η ΤΝ επιτρέπει την εξατομικευμένη μάθηση και την αναπροσαρμογή εκπαιδευτικών προγραμμάτων βάσει των αναγκών κάθε μαθητή^[14]. Πλατφόρμες όπως το Khan Academy, το Coursera και το Duolingo αξιοποιούν αλγορίθμους μηχανικής μάθησης για να προσαρμόζουν το υλικό μάθησης στις δυνατότητες κάθε μαθητή. Επιπλέον, η αυτοματοποιημένη βαθμολόγηση εξοικονομεί χρόνο από τους εκπαιδευτικούς, επιτρέποντάς τους να επικεντρωθούν περισσότερο στη διδασκαλία.

Επιπτώσεις:

- Βελτίωση της εκπαιδευτικής εμπειρίας μέσω εξατομικευμένων προγραμμάτων
- Μείωση ανισοτήτων στην πρόσβαση στη γνώση μέσω ψηφιακών εργαλείων
- Αποφόρτιση των εκπαιδευτικών μέσω αυτοματοποίησης διοικητικών εργασιών

4. Υγειονομική Περίθαλψη

Η ΤΝ χρησιμοποιείται ευρέως στην ανάλυση ιατρικών δεδομένων, την προγνωστική διάγνωση και την εξατομίκευση θεραπειών^[24]. Μοντέλα βαθιάς μάθησης (Deep Learning) έχουν αποδειχθεί εξαιρετικά αποτελεσματικά στην ανίχνευση ανωμαλιών σε απεικονιστικές εξετάσεις (MRI, CT scans)^[5]. Επίσης, εικονικοί ιατρικοί βοηθοί μπορούν να παρέχουν διαγνωστικές προτάσεις βασισμένες σε τεράστια σύνολα δεδομένων.

Επιπτώσεις:

- Βελτίωση της ακρίβειας διαγνώσεων
- Εξατομίκευση θεραπειών
- Μείωση χρόνου αναμονής για ιατρικές υπηρεσίες μέσω αυτοματοποίησης

5. Ενέργεια και Περιβάλλον

Η ΤΝ συμβάλλει στην πράσινη ενέργεια, βελτιώνοντας την απόδοση των ανανεώσιμων πηγών ενέργειας και τη διαχείριση δικτύων ηλεκτροδότησης^[21]. Οι αλγόριθμοι πρόβλεψης ζήτησης ενέργειας βελτιστοποιούν την παραγωγή ηλεκτρικής ενέργειας, μειώνοντας τις εκπομπές CO₂ και βελτιώνοντας την ενεργειακή απόδοση των πόλεων.

Επιπτώσεις:

- Βελτίωση της ενεργειακής αποδοτικότητας
- Μείωση εκπομπών CO₂ μέσω έξυπνων συστημάτων διαχείρισης
- Βελτιστοποίηση της κατανάλωσης ενέργειας στα κτίρια και στις πόλεις

6. Μεταφορές

Η ΤΝ έχει φέρει επανάσταση στον κλάδο των μεταφορών μέσω των αυτόνομων οχημάτων, των έξυπνων σηματοδοτών και των δικτύων πρόβλεψης κυκλοφορίας^[22]. Οι εταιρείες Tesla, Waymo και Uber αξιοποιούν αλγορίθμους βαθιάς μάθησης για την πλοήγηση και την αποφυγή συγκρούσεων.

Επιπτώσεις:

- Μείωση τροχαίων ατυχημάτων μέσω αυτόνομων οχημάτων
- Βελτιστοποίηση δρομολογίων μέσω ανάλυσης δεδομένων
- Μείωση περιβαλλοντικού αποτυπώματος μέσω έξυπνης διαχείρισης κυκλοφορίας

7. Καθημερινή Ζωή

Στην καθημερινή ζωή, η ΤΝ ενσωματώνεται σε οικιακές συσκευές, ψηφιακούς βοηθούς (Siri, Alexa, Google Assistant) και έξυπνα σπίτια^[20]. Οι συστάσεις περιεχομένου σε πλατφόρμες όπως το Netflix και το Spotify βασίζονται σε αλγορίθμους μηχανικής μάθησης.

Επιπτώσεις:

- Διευκόλυνση καθημερινών εργασιών μέσω έξυπνων οικιακών συσκευών
- Εξατομίκευση εμπειρίας χρήστη μέσω προσωποποιημένων συστάσεων
- Βελτίωση ποιότητας ζωής μέσω αυτοματοποιημένων βοηθών

3 Τεχνητή Νοημοσύνη και Ηθική

3.1 Προκλήσεις και ανησυχίες της TN

Η Τεχνητή Νοημοσύνη (TN) ενσωματώνεται όλο και περισσότερο στην καθημερινότητά μας, προσφέροντας τεράστιες δυνατότητες, αλλά ταυτόχρονα εγείρει πολλές προκλήσεις και ανησυχίες που πρέπει να αντιμετωπιστούν προκειμένου να αξιοποιηθεί με ασφαλή και δίκαιο τρόπο [1][3][8].

Τα συστήματα TN στηρίζονται κυρίως σε τρεις βασικούς παράγοντες:

1. Τη **συλλογή δεδομένων**, που αποτελεί τη βάση της εκπαίδευσης των αλγορίθμων.
2. Την **εκπαίδευση των συστημάτων TN**, όπου τα μοντέλα μαθαίνουν μέσω αλληλεπίδρασης με τους χρήστες.
3. Την **αυτοματοποίηση διαδικασιών**, η οποία αντικαθιστά τις ανθρώπινες παρεμβάσεις σε διάφορους τομείς [7][11].

Από αυτούς τους παράγοντες προκύπτει ένα ευρύ φάσμα προκλήσεων και ανησυχιών που σχετίζονται με ζητήματα ιδιωτικότητας, ηθικής, διαφάνειας, προκατάληψης και κοινωνικών επιπτώσεων [6][9].

Ζητήματα Ιδιωτικότητας και Προστασίας Δεδομένων

Τα συστήματα TN συλλέγουν και αναλύουν τεράστιες ποσότητες δεδομένων, με σκοπό την εξαγωγή προβλέψεων και συμπερασμάτων. Όταν τα δεδομένα περιλαμβάνουν προσωπικές πληροφορίες, προκύπτουν σοβαρά ζητήματα ιδιωτικότητας και διαφάνειας [14].

Η ανησυχία έγκειται στο πώς χρησιμοποιούνται αυτά τα δεδομένα, καθώς και στον κίνδυνο διαρροής τους. Για παράδειγμα, οι μεγάλες τεχνολογικές εταιρείες χρησιμοποιούν αλγορίθμους TN για στοχευμένες διαφημίσεις, αξιοποιώντας προσωπικές πληροφορίες των χρηστών χωρίς τη ρητή συγκατάθεσή τους [2]. Επιπλέον, η χρήση καμερών αναγνώρισης προσώπου και μικροφώνων δημιουργεί ανησυχίες παρακολούθησης και περιορισμού της ατομικής ελευθερίας [10][17].

- **Παράδειγμα:** Η ανάλυση μεγάλων δεδομένων από κοινωνικά δίκτυα θα μπορούσε να δημιουργήσει εξαιρετικά ακριβή προφίλ χρηστών, χωρίς να έχουν δώσει τη συγκατάθεσή τους [20], παραβιάζοντας την ιδιωτικότητά τους και θέτοντας σε κίνδυνο τα θεμελιώδη δικαιώματά τους, όπως η αυτονομία, η αξιοπρέπεια και η ελευθερία.

Προκαταλήψεις και Αλγοριθμική Αδικία

Η εκπαίδευση των συστημάτων της TN επιτυγχάνεται, με τη συλλογή δεδομένων του παρελθόντος και μέσω αλληλεπίδρασης με τους χρήστες, τα οποία μπορούν να περιέχουν μεροληπτικές πληροφορίες οδηγώντας στην ενσωμάτωση προκαταλήψεων και αδικιών στα μοντέλα της TN. Οι προκαταλήψεις μπορεί να προέλθουν από ιστορικές, κοινωνικές, πολιτισμικές

ή προσωπικές απόψεις των χρηστών. Αυτό μπορεί να οδηγήσει σε αλγοριθμικές αποφάσεις που ενισχύουν ή αναπαράγουν αυτές τις προκαταλήψεις, δημιουργώντας αθέμιτες ή διακριτικές πρακτικές^{[5][12]}.

Παράδειγμα:

- Ένα σύστημα πρόσληψης προσωπικού που βασίζεται στην ΤΝ ενδέχεται να προτιμά άνδρες υποψηφίους αν έχει εκπαιδευτεί με δεδομένα που προέρχονται από έναν ανδροκρατούμενο τομέα^[4].
- Η χρήση ΤΝ για δικαστικές αποφάσεις έχει δείξει ότι οι αλγόριθμοι τείνουν να επιβάλλουν αυστηρότερες ποινές σε μειονότητες, επειδή έχουν εκπαιδευτεί σε ιστορικά δεδομένα που ενσωματώνουν ρατσιστικές προκαταλήψεις ^[15].

Η αλγοριθμική διαφάνεια και η εξηγησιμότητα των αποφάσεων είναι κρίσιμες για τη διασφάλιση της δικαιοσύνης και της αντικειμενικότητας των συστημάτων ΤΝ^[18].

Αυτοματοποιημένη Λήψη Αποφάσεων και Λογοδοσία

Η χρήση και η εφαρμογή συστημάτων ΤΝ, έχει ως αποτέλεσμα την αυτοματοποίηση πολλών διαδικασιών. Η αυτοματοποίηση διαδικασιών όμως, μπορεί να οδηγήσει και στην αντικατάσταση ανθρώπινου δυναμικού, δημιουργώντας ανησυχίες για την αύξηση της ανεργίας και την κοινωνική ανισότητα^{[21][23]}.

Η ανεργία λόγω αυτοματοποίησης αποτελεί μία από τις πιο σημαντικές προκλήσεις καθώς η επανακατάρτιση των εργαζομένων και απόκτηση νέων δεξιοτήτων είναι μία διαδικασία χρονοβόρα και δαπανηρή. Πρέπει να ληφθεί υπόψιν λοιπόν, ότι ενώ η αυτοματοποίηση μπορεί να αυξήσει την αποδοτικότητα και παραγωγικότητα, μπορεί παράλληλα να προκαλέσει αρνητικές επιπτώσεις στην ποιότητα ζωής των ανθρώπων, συμπεριλαμβανομένων των όρων εργασίας και των αποδοχών.

Κατά επέκταση όλων των παραπάνω, σημαντική ανησυχία αποτελεί η λογοδοσία και η απόδοση της ευθύνης σε περίπτωση λάθους ή βλάβης που μπορεί να προκληθούν από τα συστήματα ΤΝ. Η πολυπλοκότητα των αλγορίθμων ΤΝ και η δυσκολία κατανόησης των εσωτερικών τους διαδικασιών καθιστούν τον προσδιορισμό της ευθύνης για λανθασμένες αποφάσεις ιδιαίτερα προβληματικό ζήτημα. Αυτή η δυσκολία έχει προκαλέσει έντονες συζητήσεις γύρω από το θέμα, ειδικά όταν οι αποφάσεις που λαμβάνονται από τα συστήματα αυτά έχουν σοβαρές επιπτώσεις, όπως στην περίπτωση της υγειονομικής περίθαλψης ή της δικαιοσύνης. Όταν ένας αλγόριθμος πάρει μια λανθασμένη απόφαση, ποιος ευθύνεται; Ο προγραμματιστής, ο φορέας που τον χρησιμοποιεί ή ο ίδιος ο αλγόριθμος;

- **Παράδειγμα:** Σε περιπτώσεις αυτόνομων οχημάτων, ποιος έχει την ευθύνη σε περίπτωση ατυχήματος; Ο κατασκευαστής του οχήματος, ο δημιουργός του λογισμικού ή ο χρήστης που το ενεργοποίησε; Αυτές οι ερωτήσεις δείχνουν την ανάγκη για σαφή νομικά πλαίσια και κατευθυντήριες γραμμές ^[19].

Τέλος, οι προοπτικές για το μέλλον της ΤΝ εξαρτώνται σε μεγάλο βαθμό από την ικανότητα των σχεδιαστών και των πολιτικών να αντιμετωπίσουν τις προκλήσεις αυτές με τρόπο που να προάγει την ηθική, τη διαφάνεια και τη δικαιοσύνη. Η ανάπτυξη κατευθυντήριων γραμμών και η ρύθμιση της χρήσης της ΤΝ αποτελούν σημαντικά βήματα προς αυτή την κατεύθυνση. Με την

κατάλληλη προσοχή στις ηθικές προκλήσεις, η ΤΝ μπορεί να προσφέρει σημαντικά οφέλη στην κοινωνία και να συμβάλει στη βελτίωση της ανθρώπινης ευημερίας.

3.2 Εισαγωγή στην Ηθική

Τα ηθικά ζητήματα που σχετίζονται με την ΤΝ δεν είναι απλώς θεωρητικές ανησυχίες, αλλά πρακτικά προβλήματα που διαμορφώνουν την καθημερινή ζωή, την κοινωνία και την οικονομία. Η αλληλεπίδραση αλγορίθμων και δεδομένων εγείρει ζητήματα διαφάνειας, προκατάληψης και ιδιωτικότητας, ενώ η αυτοματοποίηση δημιουργεί νέες προκλήσεις για την εργασία και την κοινωνική δικαιοσύνη.^{[9][14][16]}

Η διαφάνεια των αλγορίθμων αποτελεί κεντρική ηθική πρόκληση, καθώς χωρίς εξηγήσιμες αποφάσεις, οι χρήστες και οι υπεύθυνοι χάραξης πολιτικής δεν μπορούν να αξιολογήσουν την αξιοπιστία των συστημάτων ΤΝ. Πώς μπορούμε να είμαστε σίγουροι ότι οι αποφάσεις που λαμβάνονται από ένα σύστημα ΤΝ είναι δίκαιες και αντικειμενικές, όταν δεν κατανοούμε πλήρως πώς λειτουργεί ο αλγόριθμος; Η διαφάνεια είναι κρίσιμη για την οικοδόμηση εμπιστοσύνης μεταξύ των χρηστών και των συστημάτων ΤΝ, και απαιτεί την ανάπτυξη τεχνολογιών που να επιτρέπουν την εξήγηση και την κατανόηση των διαδικασιών λήψης αποφάσεων^{[6][7]}.

Επιπλέον, η αλγοριθμική δικαιοσύνη αποτελεί κεντρικό ζήτημα. Οι αλγόριθμοι που χρησιμοποιούνται για την ανάλυση δεδομένων μπορούν να ενισχύσουν ή να αναπαράγουν υπάρχουσες κοινωνικές ανισότητες και προκαταλήψεις. Αυτό δημιουργεί την ανάγκη για προσεκτικό σχεδιασμό και συνεχή αξιολόγηση των συστημάτων ΤΝ, ώστε να διασφαλιστεί ότι λειτουργούν με αμεροληψία και δεν εισάγουν ή ενισχύουν διακρίσεις^{[2][11][15]}.

Η προστασία της ιδιωτικότητας και η ασφάλεια των δεδομένων είναι εξίσου σημαντικά ζητήματα. Οι τεχνολογίες ΤΝ συχνά επεξεργάζονται μεγάλες ποσότητες προσωπικών δεδομένων, καθιστώντας την προστασία αυτών των δεδομένων κρίσιμη για την προστασία των ατομικών δικαιωμάτων. Οι σχεδιαστές και οι προγραμματιστές συστημάτων ΤΝ πρέπει να διασφαλίζουν ότι εφαρμόζονται αυστηρά πρωτόκολλα για την προστασία της ιδιωτικότητας και την αποφυγή καταχρήσεων^{[19][25]}.

Τέλος, οι κοινωνικές και οικονομικές επιπτώσεις της αυτοματοποίησης που επιφέρει η ΤΝ είναι ένα ακόμη ζήτημα που απαιτεί προσοχή. Η αντικατάσταση των ανθρώπινων εργαζομένων από μηχανές σε διάφορους τομείς μπορεί να οδηγήσει σε σημαντικές αλλαγές στην αγορά εργασίας, με πιθανές συνέπειες την ανεργία και την κοινωνική ανισότητα^{[5][8]}. Επομένως, είναι απαραίτητο να εξετάσουμε στρατηγικές για την επανεκπαίδευση και την υποστήριξη των εργαζομένων που επηρεάζονται από αυτές τις αλλαγές^[23].

Η ηθική στη σχεδίαση και εφαρμογή της ΤΝ δεν αποτελεί πολυτέλεια, αλλά αναγκαιότητα για τη δημιουργία ενός δικαίου και βιώσιμου μέλλοντος. Η ευθύνη για την αντιμετώπιση αυτών των ηθικών ζητημάτων ανήκει σε όλους τους ενδιαφερόμενους φορείς, από τους επιστήμονες και τους προγραμματιστές έως τις κυβερνήσεις και την κοινωνία στο σύνολό της^{[9][13]}.

3.3 Ηθική Διάσταση στη Σχεδίαση ΤΝ

Η ηθική σχεδίαση της ΤΝ αφορά την ενσωμάτωση αξιών όπως η διαφάνεια, η αμεροληψία και η λογοδοσία στη διαδικασία ανάπτυξης αλγορίθμων και έξυπνων συστημάτων. Αυτό είναι ιδιαίτερα κρίσιμο σε τομείς όπως η υγεία, η δικαιοσύνη και η αγορά εργασίας, όπου οι αποφάσεις των αλγορίθμων μπορούν να έχουν άμεσο κοινωνικό και οικονομικό αντίκτυπο^{[11][3]}.

Καθώς η ΤΝ επηρεάζει όλο και περισσότερες πτυχές της καθημερινότητας, η ανάγκη για σαφή ηθικά πλαίσια και ρυθμίσεις γίνεται επιτακτική^{[14][16]}. Επιστημονικές έρευνες δείχνουν ότι η απουσία ηθικού σχεδιασμού μπορεί να οδηγήσει σε μεροληπτικά μοντέλα, αδιαφανείς αποφάσεις και κοινωνικές ανισότητες^{[7][18]}.

Για τον λόγο αυτό, εταιρείες, ακαδημαϊκά ιδρύματα και κυβερνήσεις αναπτύσσουν πολιτικές που στοχεύουν στη διαμόρφωση αξιόπιστων και δίκαιων συστημάτων ΤΝ. Παραδείγματα περιλαμβάνουν την ενσωμάτωση μηχανισμών εξηγήσιμης ΤΝ (Explainable AI), που επιτρέπουν στους χρήστες να κατανοούν πώς λαμβάνονται οι αποφάσεις, καθώς και τη χρήση αντιμεροληπτικών μεθόδων εκπαίδευσης δεδομένων^[10].

3.4 Ηθικές Αρχές και Απαιτήσεις

Διαφάνεια και Εξηγησιμότητα

Η διαφάνεια είναι ένα ουσιαστικό στοιχείο της ηθικής διάστασης στη σχεδίαση ΤΝ, καθώς αποτελεί θεμελιώδη αρχή για την ανάπτυξη εμπιστοσύνης στα συστήματα ΤΝ. Οι χρήστες πρέπει να κατανοούν πώς λειτουργούν τα συστήματα αυτά και πώς λαμβάνονται οι αποφάσεις που μπορεί να επηρεάζουν σημαντικές πτυχές της ζωής τους, όπως η υγειονομική περίθαλψη ή η πρόσβαση σε οικονομικούς πόρους^{[6][12]}. Επομένως, οι αλγόριθμοι πρέπει να είναι εξηγήσιμοι και οι αποφάσεις τους να μπορούν να αναλυθούν και να αιτιολογηθούν^[10]. Η διαφάνεια όμως δεν αφορά μόνο την εξήγηση των αποφάσεων, αλλά και την ανοικτή επικοινωνία σχετικά με τις δυνατότητες και τους περιορισμούς των συστημάτων ΤΝ.

Για παράδειγμα, σε ένα σύστημα που χρησιμοποιείται για την ανίχνευση απάτης, είναι σημαντικό οι χρήστες να γνωρίζουν πώς λειτουργεί το σύστημα, ποια δεδομένα χρησιμοποιούνται και ποια είναι τα ποσοστά σφάλματος. Αυτό επιτρέπει στους χρήστες να λαμβάνουν ενημερωμένες αποφάσεις σχετικά με τη χρήση του συστήματος και να αναγνωρίζουν πιθανά προβλήματα που μπορεί να προκύψουν^[17].

Ένα χαρακτηριστικό παράδειγμα έλλειψης διαφάνειας είναι τα αυτοματοποιημένα συστήματα αξιολόγησης αιτήσεων δανείων, όπου έχουν αναφερθεί περιπτώσεις απόρριψης αιτήσεων χωρίς σαφή αιτιολόγηση, δημιουργώντας ανησυχίες σχετικά με την αλγοριθμική προκατάληψη και την έλλειψη εξήγησης των αποφάσεων^{[2][11]}.

Ανθρωποκεντρικός Σχεδιασμός

Ο ανθρωποκεντρικός σχεδιασμός είναι μια ακόμα σημαντική πτυχή της ηθικής στη σχεδίαση ΤΝ. Η τεχνολογία πρέπει να αναπτύσσεται με επίκεντρο τον άνθρωπο, λαμβάνοντας υπόψη τις ανάγκες, τις αξίες και τις προτεραιότητες των χρηστών^{[7][11]}. Αυτό σημαίνει ότι οι

σχεδιαστές πρέπει να διασφαλίζουν ότι τα συστήματα ΤΝ βελτιώνουν την ανθρώπινη ευημερία και δεν προκαλούν βλάβη, είτε άμεσα είτε έμμεσα.

Για παράδειγμα, ένα σύστημα ΤΝ που χρησιμοποιείται σε ιατρικές διαγνώσεις πρέπει να λαμβάνει υπόψη τις ανθρώπινες ανησυχίες για την ακρίβεια και την ηθική χρήση των δεδομένων υγείας^[24].

Έρευνες δείχνουν ότι οι χρήστες εμπιστεύονται περισσότερο τα συστήματα ΤΝ όταν αισθάνονται ότι έχουν κάποιο επίπεδο ελέγχου ή δυνατότητα παρέμβασης στις αποφάσεις τους. Αυτό ενισχύει την αποδοχή και τη χρήση των τεχνολογιών ΤΝ, ιδίως σε κρίσιμους τομείς όπως η υγειονομική περίθαλψη και η εκπαίδευση ^[7].

Σεβασμός Ανθρώπινης Αυτονομίας και Δράσης

Μια από τις πιο σημαντικές ηθικές αρχές στη σχεδίαση ΤΝ είναι ο σεβασμός για την ανθρώπινη αυτονομία και δράση. Ο σεβασμός για την ανθρώπινη δράση απαιτεί από τα συστήματα ΤΝ να σχεδιάζονται με τρόπο που να ενισχύει την αυτονομία των χρηστών, επιτρέποντάς τους να λαμβάνουν ελεύθερα, ενημερωμένες αποφάσεις^{[5][14]}.

Αυτό σημαίνει ότι τα συστήματα δεν πρέπει να περιορίζουν την ελευθερία των χρηστών μέσω κρυφών ή παραπλανητικών διαδικασιών, αλλά να τους παρέχουν τα απαραίτητα εργαλεία και πληροφορίες για να ασκήσουν την αυτονομία τους^[19]. Ο σεβασμός για την ανθρώπινη αυτονομία είναι θεμελιώδης για τη διατήρηση της αξιοπρέπειας και της ελευθερίας των ατόμων σε μια εποχή όπου οι τεχνολογίες ΤΝ διαδραματίζουν όλο και πιο κεντρικό ρόλο στις αποφάσεις που επηρεάζουν τη ζωή τους^[16].

Ένα παράδειγμα παραβίασης της αυτονομίας των χρηστών είναι τα "σκοτεινά μοτίβα" (dark patterns) που χρησιμοποιούνται στις διεπαφές των συστημάτων, τα οποία καθοδηγούν σιωπηρά τους χρήστες προς συγκεκριμένες επιλογές χωρίς σαφή ενημέρωση, όπως η αποδοχή των όρων χρήσης δεδομένων χωρίς ξεκάθαρη παρουσίαση εναλλακτικών^[3].

Δικαιοσύνη και Μη Διάκριση

Η αμεροληψία και η αντιμετώπιση τις προκατάληψης είναι κρίσιμη για την ανάπτυξη ηθικών συστημάτων ΤΝ. Οι αλγόριθμοι πρέπει να σχεδιάζονται με τρόπο που να αποφεύγει την ενίσχυση των κοινωνικών ανισοτήτων και των διακρίσεων^{[2][9]}.

Αυτό απαιτεί συνεχή αξιολόγηση και βελτίωση των μοντέλων ΤΝ, καθώς και τη χρήση αντιπροσωπευτικών και αμερόληπτων δεδομένων εκπαίδευσης^{[3][15]}. Η αλγοριθμική δικαιοσύνη απαιτεί τη διασφάλιση ότι τα δεδομένα εκπαίδευσης είναι αντιπροσωπευτικά και απαλλαγμένα από προκαταλήψεις^[8].

Απόρρητο και Ασφάλεια Δεδομένων

Τα συστήματα ΤΝ συχνά επεξεργάζονται μεγάλες ποσότητες προσωπικών δεδομένων. Η προστασία του απορρήτου των χρηστών και η ασφάλεια των δεδομένων τους είναι θεμελιώδεις^{[19][25]}.

Οι σχεδιαστές πρέπει να εφαρμόζουν πρακτικές όπως η ανωνυμοποίηση δεδομένων και η ισχυρή κρυπτογράφηση^[20]. Ο σχεδιασμός διεπαφών χρήστη που δίνουν προτεραιότητα στο απόρρητο των χρηστών και την ασφάλεια των δεδομένων μπορεί να οικοδομήσει εμπιστοσύνη και αξιοπιστία μεταξύ των χρηστών^[22].

Ευθύνη και Λογοδοσία

Στην ηθική ανάπτυξη της ΤΝ περιλαμβάνεται και η λογοδοσία. Οι δημιουργοί και οι χειριστές των συστημάτων ΤΝ πρέπει να είναι υπεύθυνοι για τις ενέργειες και τις αποφάσεις των συστημάτων τους^{[4][9]}.

Αυτό περιλαμβάνει την ανάπτυξη μηχανισμών για την παρακολούθηση και την αξιολόγηση της απόδοσης των συστημάτων ΤΝ, καθώς και τη δυνατότητα διόρθωσης λαθών ή αδικιών που μπορεί να προκύψουν^[13]. Η ενσωμάτωση της λογοδοσίας είναι απαραίτητη για την ανάπτυξη υπεύθυνων οικοσυστημάτων ΤΝ που θα λειτουργούν με διαφάνεια και δικαιοσύνη^[18].

Κοινωνικές και Εργασιακές Επιπτώσεις

Η αυτοματοποίηση και η ΤΝ μπορεί να επηρεάσουν σημαντικά την εργασία και την κοινωνία. Οι σχεδιαστές πρέπει να λαμβάνουν υπόψη τις επιπτώσεις στις θέσεις εργασίας και να προωθούν την εκπαίδευση και την αναβάθμιση των δεξιοτήτων για να βοηθήσουν τους ανθρώπους να προσαρμοστούν στις νέες συνθήκες^{[5][12]}.

Συμμετοχή της Κοινωνίας

Η συμμετοχή της κοινωνίας στη διαδικασία σχεδιασμού της ΤΝ είναι απαραίτητη για να διασφαλιστεί ότι οι τεχνολογίες αντανακλούν τις αξίες και τις ανάγκες της κοινωνίας^{[15][23]}.

Η διαβούλευση με τους χρήστες και τους ενδιαφερόμενους φορείς μπορεί να βοηθήσει στην ανάπτυξη πιο ηθικών και αποδεκτών συστημάτων^[9].

Η ηθική στη σχεδίαση της ΤΝ δεν είναι απλώς θέμα τεχνολογίας, αλλά αφορά και τις αξίες και τις προτεραιότητες που ενσωματώνονται στην ανάπτυξή της. Οι σχεδιαστές, οι προγραμματιστές και οι πολιτικοί πρέπει να συνεργαστούν για να εξασφαλίσουν ότι οι τεχνολογίες ΤΝ θα χρησιμοποιηθούν προς όφελος όλων των μελών της κοινωνίας^[17].

3.5 Η Ανάγκη για Κατευθυντήριες Γραμμές

Οι κατευθυντήριες γραμμές για την ΤΝ αποτελούν ένα κρίσιμο εργαλείο για την εξασφάλιση της αξιοπιστίας, ασφαλούς και δεοντολογικής ανάπτυξης και χρήσης των συστημάτων αυτών. Στην εποχή της ψηφιακής επανάστασης, όπου η ΤΝ εισέρχεται δυναμικά σε όλους τους τομείς της κοινωνίας, η ανάγκη για συγκεκριμένες και σαφείς κατευθύνσεις γίνεται πιο επιτακτική από ποτέ. Με την ταχεία πρόοδο της ΤΝ από τα μέσα του 20ού αιώνα, έγινε εμφανής η ανάγκη για θεσμοθέτηση πλαισίων που θα διασφάλιζαν τη χρήση της με τρόπο που προάγει το κοινό καλό^{[9][14][16]}.

Αρχικά, οι συζητήσεις για την ηθική της ΤΝ περιορίζονταν σε ακαδημαϊκούς κύκλους, αλλά καθώς οι εφαρμογές της ΤΝ άρχισαν να επηρεάζουν ευρύτερους τομείς της κοινωνίας, η ανάγκη για συγκεκριμένες κατευθυντήριες γραμμές έγινε πιο επιτακτική^{[18][17]}. Οι κατευθυντήριες γραμμές λειτουργούν ως ένα πλαίσιο που καθοδηγεί τις πρακτικές των φορέων που αναπτύσσουν, υλοποιούν και χρησιμοποιούν την ΤΝ, προκειμένου να διασφαλιστεί ότι οι νέες τεχνολογίες υπηρετούν το κοινό καλό και σέβονται τα ανθρώπινα δικαιώματα^{[3][7]}.

Σε τομείς όπως η υγεία, όπου οι αποφάσεις που λαμβάνονται από συστήματα ΤΝ μπορεί να επηρεάσουν άμεσα την ανθρώπινη ζωή, οι κατευθυντήριες γραμμές παρέχουν ένα απαραίτητο πλαίσιο για τη διασφάλιση της ασφάλειας και της δικαιοσύνης^{[10][15]}. Για παράδειγμα, η εφαρμογή αλγορίθμων σε διαγνωστικές διαδικασίες απαιτεί αυστηρή τήρηση δεοντολογικών αρχών για την αποφυγή προκαταλήψεων και τη διασφάλιση αμεροληψίας^[5].

Οι κατευθυντήριες γραμμές απευθύνονται σε ένα ευρύ φάσμα ενδιαφερόμενων μερών. Οι κυβερνητικοί φορείς μπορούν να χρησιμοποιήσουν αυτές τις οδηγίες για τη διαμόρφωση κανονιστικών πλαισίων που προστατεύουν τα δικαιώματα των πολιτών και διασφαλίζουν τη διαφάνεια και τη λογοδοσία^{[19][25]}. Οι επιχειρήσεις, από την πλευρά τους, μπορούν να ενσωματώσουν αυτές τις κατευθυντήριες γραμμές στις πολιτικές τους, διασφαλίζοντας ότι οι τεχνολογίες ΤΝ που αναπτύσσουν είναι ασφαλείς, ηθικές και κοινωνικά αποδεκτές^[12].

Επίσης, οι ερευνητές και οι προγραμματιστές μπορούν να χρησιμοποιήσουν τις κατευθυντήριες γραμμές για να καθοδηγούν τον σχεδιασμό και την ανάπτυξη των τεχνολογιών ΤΝ με τρόπο που σέβεται τα ανθρώπινα δικαιώματα και προάγει την κοινωνική ευημερία^[6]. Υπάρχουν διάφοροι τύποι κατευθυντήριων γραμμών στην ΤΝ, καθένας από τους οποίους εξυπηρετεί διαφορετικούς σκοπούς και αποδέκτες.

Οι κανονιστικές ρυθμίσεις, όπως αυτές που προβλέπονται από τον Γενικό Κανονισμό Προστασίας Δεδομένων (General Data Protection Regulation - GDPR) στην Ευρωπαϊκή Ένωση, αποτελούν νομικά δεσμευτικά πρότυπα που προστατεύουν την ιδιωτικότητα και τα προσωπικά δεδομένα των πολιτών^[19]. Παράλληλα, οι ηθικές οδηγίες, όπως αυτές που αναπτύσσονται από διεθνείς οργανισμούς και ακαδημαϊκούς, καθοδηγούν τις βέλτιστες πρακτικές στην ανάπτυξη της ΤΝ, διασφαλίζοντας ότι οι τεχνολογίες αυτές θα χρησιμοποιούνται με τρόπο που είναι συμβατός με τις ηθικές αξίες και τα ανθρώπινα δικαιώματα^{[4][13]}.

Επιπλέον, βιομηχανικά πρότυπα που αναπτύσσονται από τον ιδιωτικό τομέα στοχεύουν στη διασφάλιση της ποιότητας και της ασφάλειας των προϊόντων και των υπηρεσιών που βασίζονται στην ΤΝ. Μεγάλες εταιρείες τεχνολογίας, όπως η Google και η Microsoft, έχουν δημοσιεύσει εσωτερικές ηθικές οδηγίες για την ΤΝ, αν και η υλοποίησή τους δεν είναι πάντα εγγυημένη, προκαλώντας διαφωνίες σχετικά με την αποτελεσματικότητα της αυτορρύθμισης^[18].

Οι κατευθυντήριες γραμμές για την ΤΝ δεν είναι απλώς μια συλλογή κανόνων και προτύπων. Αποτελούν ένα «ζωντανό» έγγραφο, το οποίο εξελίσσεται παράλληλα με την τεχνολογία και τις ανάγκες της κοινωνίας^{[9][23]}. Καθώς η ΤΝ συνεχίζει να προοδεύει, οι κατευθυντήριες γραμμές πρέπει να αναθεωρούνται και αναπροσαρμόζονται, ώστε να καλύπτουν τις νέες προκλήσεις που ανακύπτουν και να διασφαλίζουν ότι η χρήση της ΤΝ παραμένει πάντα προς όφελος της κοινωνίας^[22].

Η ανάπτυξη κατευθυντήριων γραμμών για την ΤΝ δεν αποτελεί μόνο εθνική υπόθεση, αλλά απαιτεί διεθνή συνεργασία. Οι διαφορετικές πολιτισμικές προσεγγίσεις και οι νομικές διατάξεις

μεταξύ των χωρών καθιστούν αναγκαία τη δημιουργία διεθνών πλαισίων που θα προωθούν την ασφαλή και ηθική χρήση της ΤΝ σε παγκόσμιο επίπεδο^[21].

Η εφαρμογή των κατευθυντήριων γραμμών απαιτεί την ενεργή συμμετοχή όλων των ενδιαφερόμενων μερών, συμπεριλαμβανομένων των κυβερνήσεων, των επιχειρήσεων, των ερευνητικών ιδρυμάτων και των οργανώσεων της κοινωνίας των πολιτών^[24]. Μόνο μέσω μιας συλλογικής προσπάθειας μπορεί να επιτευχθεί η ανάπτυξη μιας ΤΝ που θα είναι ταυτόχρονα καινοτόμος και ηθικά αποδεκτή, συμβάλλοντας έτσι στη δημιουργία ενός βιώσιμου και δίκαιου μέλλοντος^[20].

4 *Πλαίσια, Κατευθυντήριες Γραμμές και η Εφαρμογή τους*

Το Κεφάλαιο 4 εξετάζει τη σημασία των πλαισίων και των κατευθυντήριων γραμμών στην ανάπτυξη και εφαρμογή της ΤΝ. Αρχικά, εισάγονται οι γενικές αρχές που διέπουν τη δημιουργία κατευθυντήριων γραμμών, ενώ στη συνέχεια αναλύονται τα διεθνή πλαίσια, οι νομικές ρυθμίσεις, οι ηθικές αρχές και οι προκλήσεις που συναντώνται κατά την εφαρμογή τους. Τέλος, συζητείται η ανάγκη για ένα παγκόσμιο ρυθμιστικό πλαίσιο που θα εναρμονίζει τις διεθνείς πρακτικές και θα εξασφαλίζει την ασφαλή και υπεύθυνη χρήση της ΤΝ^{[4][9][15]}.

4.1 *Διεθνή Πλαίσια και Κατευθυντήριες Γραμμές*

Υπάρχουν περισσότερα από 80 δημόσια και ιδιωτικά έγγραφα δεοντολογικών οδηγιών ΤΝ. Πρόσφατα δημοσιευμένες κατευθυντήριες γραμμές δεοντολογίας είναι:

1. **«Δεοντολογικές Οδηγίες για αξιόπιστη Τεχνητή Νοημοσύνη» της Ευρωπαϊκής Επιτροπής (European Commission - EU).** Αυτές οι οδηγίες υπογραμμίζουν την ανάγκη για αξιοπιστία και διαφάνεια στα συστήματα ΤΝ, προτείνοντας αρχές όπως η εξηγησιμότητα, η προστασία της ιδιωτικότητας και η αποφυγή διακρίσεων. Στόχος είναι η δημιουργία συστημάτων ΤΝ που είναι όχι μόνο τεχνολογικά προηγμένα, αλλά και ηθικά ευθυγραμμισμένα με τις αξίες της κοινωνίας^{[19][25]}.
2. **«Ηθικά Ευθυγραμμισμένος Σχεδιασμός» («Ethically Aligned Design») του Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (Institute of Electrical and Electronic Engineers - IEEE).** Το οποίο είναι ένα πλαίσιο το οποίο επικεντρώνεται στην ενσωμάτωση ηθικών αρχών σε όλα τα στάδια της ανάπτυξης της ΤΝ. Το πλαίσιο αυτό προάγει την ευθύνη, τη διαφάνεια και την ασφάλεια ως κρίσιμες διαστάσεις για τη δημιουργία συστημάτων ΤΝ που σέβονται την ανθρώπινη αξιοπρέπεια και δικαιώματα^{[7][14]}.
3. **«Ηθικές αρχές για την Τεχνητή Νοημοσύνη και την Ανάλυση Δεδομένων» του Συνδέσμου Βιομηχανίας Λογισμικού και Πληροφορικής (Software and Information Industry Association - SIIA).** Οι οποίες δίνουν έμφαση στη δίκαιη και υπεύθυνη χρήση της ΤΝ και των δεδομένων. Αυτές οι αρχές προωθούν την αποφυγή διακρίσεων, τη διασφάλιση της προστασίας των δεδομένων και την προαγωγή της εμπιστοσύνης μέσω της διαφάνειας^{[10][17]}.

Παρόμοιες πρωτοβουλίες έχουν αναπτυχθεί από μεγάλες τεχνολογικές εταιρείες, όπως η Google και η Microsoft, που έχουν δημοσιεύσει ηθικές οδηγίες για την ανάπτυξη της ΤΝ. Ωστόσο, η πραγματική υιοθέτηση και εφαρμογή αυτών των πλαισίων παραμένει αμφιλεγόμενη, καθώς η αυτορρύθμιση δε διασφαλίζει πάντα ότι οι επιχειρήσεις θα ακολουθήσουν αυτές τις ηθικές αρχές στην πράξη [3][18].

Οι παραπάνω κατευθυντήριες γραμμές έχουν σημαντική επίδραση στην παγκόσμια ρύθμιση της ΤΝ, καθώς προσφέρουν ένα κοινό πλαίσιο αναφοράς για κυβερνήσεις, εταιρείες και οργανισμούς. Αυτό διευκολύνει τη συνεργασία και την εναρμόνιση των κανόνων σε διεθνές επίπεδο, κάτι που είναι απαραίτητο για την ανάπτυξη ασφαλών και υπεύθυνων συστημάτων ΤΝ.

Η ανάγκη για διεθνή συνεργασία και συντονισμό σε θέματα ρύθμισης της ΤΝ είναι κρίσιμη, καθώς οι τεχνολογίες αυτές επηρεάζουν παγκοσμίως διάφορες πτυχές της ζωής μας και απαιτούν συνεπή και ηθική διαχείριση [11][23].

Συνολικά, τα διεθνή πλαίσια και οι κατευθυντήριες γραμμές παρέχουν τη βάση για μια ηθική προσέγγιση στην ανάπτυξη της ΤΝ, ενισχύοντας την εμπιστοσύνη του κοινού και διασφαλίζοντας ότι η ΤΝ θα λειτουργεί προς όφελος όλων των ανθρώπων [9][13].

<i>Κατευθυντήριες Γραμμές</i>	<i>Ευρωπαϊκή Ένωση</i>	<i>Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών</i>	<i>Σύνδεσμος Βιομηχανίας Λογισμικού και Πληροφορικής</i>
<i>Διαφάνεια</i>	Ορίζεται	Ορίζεται	Ορίζεται
<i>Ενημερία</i>	Ορίζεται	Ορίζεται	Ορίζεται
<i>Επίγνωση της κακής χρήσης και βλάβης</i>	Ορίζεται	Ορίζεται	Ορίζεται
<i>Επεξηγησιμότητα</i>	Ορίζεται	Αναφορά	Ορίζεται
<i>Ευθύνη</i>	Ορίζεται	Ορίζεται	--
<i>Αυτονομία</i>	Ορίζεται	Ορίζεται	--
<i>Δικαιοσύνη</i>	Ορίζεται	Αναφορά	Αναφορά
<i>Ασφάλεια</i>	Ορίζεται	Αναφορά	--
<i>Μυστικότητα</i>	Ορίζεται	Αναφορά	--
<i>Σκοπός Συστημάτων ΤΝ</i>	Αναφορά	Ορίζεται	--
<i>Επάρκεια</i>	Αναφορά	Ορίζεται	--

Πίνακας 4: Κατευθυντήριες γραμμές ως ένα πλαίσιο για αξιολόγηση ΤΝ

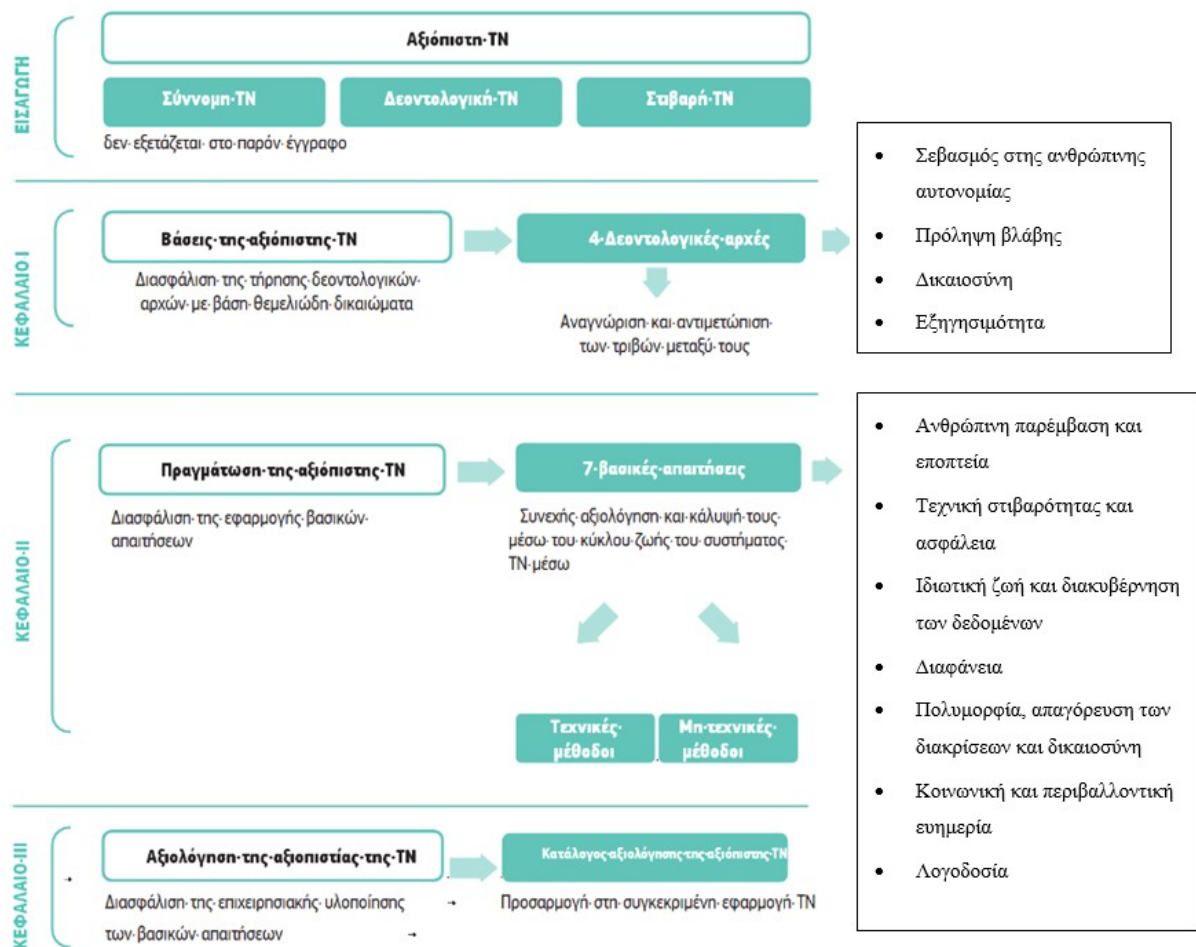
✶ Ο παραπάνω πίνακας συνοψίζει τις κύριες αρχές των πιο γνωστών κατευθυντήριων γραμμών που σχετίζονται με την ανάπτυξη και εφαρμογή της ΤΝ. Η ανάλυση δείχνει ότι η Ευρωπαϊκή Ένωση έχει την πιο ολοκληρωμένη προσέγγιση, ενώ το IEEE και η SIIA δίνουν έμφαση σε πιο συγκεκριμένες πτυχές, όπως η αυτονομία και η υπευθυνότητα [19][21].

4.1.1 «Δεοντολογικές Οδηγίες για αξιόπιστη Τεχνητή Νοημοσύνη» της Ευρωπαϊκής Επιτροπής (European Commission - EU)

Οι κατευθυντήριες γραμμές της ΕΕ εστιάζουν σε τρεις βασικές συνιστώσες για την προαγωγή της αξιόπιστης ΤΝ, οι οποίες πρέπει να πληρούνται σε ολόκληρο τον κύκλο ζωής του συστήματος^{[19][25]}:

1. Θα πρέπει να είναι **σύννομη**: Η ΤΝ πρέπει να λειτουργεί εντός του πλαισίου των υφιστάμενων νομοθετικών και κανονιστικών διατάξεων, διασφαλίζοντας τη συμμόρφωση με όλους τους ισχύοντες νόμους και κανονισμούς^[4].
2. Θα πρέπει να είναι **δεοντολογική**: Η ΤΝ πρέπει να συμμορφώνεται με δεοντολογικές αρχές και αξίες, με έμφαση στον σεβασμό της ανθρώπινης αξιοπρέπειας και την προστασία των ατομικών δικαιωμάτων^{[9][14]}.
3. Θα πρέπει να είναι **στιβαρή**: Η ΤΝ πρέπει να είναι στιβαρή από τεχνικής και κοινωνικής άποψης, εξασφαλίζοντας τη λειτουργική ασφάλεια και αξιοπιστία των συστημάτων της, με τρόπο που προλαμβάνει τις ακούσιες βλάβες^{[15][17]}.

Κάθε μία από αυτές τις συνιστώσες είναι αναγκαία, ωστόσο δεν επαρκεί από μόνη της για να διασφαλίσει την αξιοπιστία της ΤΝ. Ιδανικά, οι τρεις συνιστώσες πρέπει να συνεργάζονται αρμονικά, ενώ η κοινωνία έχει την ευθύνη να διασφαλίσει ότι δεν υπάρχουν συγκρούσεις ή παραβιάσεις στις αρχές αυτές^{[13][18]}.



Σχήμα 1: Κατευθυντήριες γραμμές της ΕΕ ως ένα πλαίσιο για αξιολόγηση ΤΝ

Οι κατευθυντήριες γραμμές δεν περιορίζονται σε έναν κατάλογο δεοντολογικών αρχών, αλλά παρέχουν κατευθύνσεις σχετικά με τους τρόπους με τους οποίους οι αρχές αυτές είναι δυνατό να υλοποιηθούν επιχειρησιακά στο πλαίσιο κοινωνικοτεχνικών συστημάτων. Οι κατευθύνσεις παρέχονται σε τρία επίπεδα, από τα πιο αφηρημένα στο κεφάλαιο I έως τα πιο συγκεκριμένα στο κεφάλαιο III, και ολοκληρώνονται με παραδείγματα ευκαιριών και καίριων προβληματισμών που εγείρουν τα συστήματα TN.

Κεφάλαιο I: Βάσεις της αξιόπιστης TN

Το πρώτο κεφάλαιο των κατευθυντήριων γραμμών εστιάζει στη θεμελίωση της αξιόπιστης TN. Αυτό περιλαμβάνει την αναγνώριση και την ενσωμάτωση θεμελιωδών δεοντολογικών αρχών, όπως ο σεβασμός της ανθρώπινης αξιοπρέπειας, η προστασία των ατομικών δικαιωμάτων και η προώθηση της κοινωνικής ευημερίας. Αυτές οι αρχές αποτελούν τη βάση για την ανάπτυξη και τη χρήση της TN με τρόπο που να είναι σύμφωνος με τα ανθρώπινα δικαιώματα και τις ηθικές αξίες της κοινωνίας^{[5][6][11]}.

Κεφάλαιο II: Πραγμάτωση της αξιόπιστης TN

Το δεύτερο κεφάλαιο ασχολείται με την εφαρμογή των δεοντολογικών αρχών μέσα από επτά συγκεκριμένες απαιτήσεις, που διασφαλίζουν την ασφάλεια, την υπευθυνότητα και τη διαφάνεια των συστημάτων TN:

1. Ανθρώπινη παρέμβαση και εποπτεία
2. Τεχνική στιβαρότητα και ασφάλεια
3. Ιδιωτική ζωή και διακυβέρνηση των δεδομένων
4. Διαφάνεια
5. Πολυμορφία, απαγόρευση των διακρίσεων και δικαιοσύνη
6. Κοινωνική και περιβαλλοντική ευημερία
7. Λογοδοσία

Οι απαιτήσεις αυτές συνδέουν τις δεοντολογικές αρχές με συγκεκριμένες τεχνικές και κοινωνικές πρακτικές, διασφαλίζοντας ότι η TN παραμένει ασφαλής και αξιόπιστη καθ' όλη τη διάρκεια του κύκλου ζωής της ^{[7][10][23]}. Μεταξύ των απαιτήσεων αυτών είναι η διαφάνεια, η λογοδοσία και η ασφάλεια, που θεωρούνται κρίσιμες για την ενίσχυση της εμπιστοσύνης του κοινού.

Κεφάλαιο III: Αξιολόγηση της αξιοπιστίας της TN

Το τρίτο κεφάλαιο επικεντρώνεται στην αξιολόγηση της αξιοπιστίας της TN. Παρέχεται ένας κατάλογος αξιολόγησης που βοηθά τους επαγγελματίες της TN να εφαρμόζουν τις απαιτήσεις και να διασφαλίζουν ότι τα συστήματά τους είναι σύμφωνες με τις κατευθυντήριες γραμμές. Αυτός ο κατάλογος αξιολόγησης είναι σχεδιασμένος να προσαρμόζεται ανάλογα με τις συγκεκριμένες ανάγκες και τις συνθήκες εφαρμογής, υπογραμμίζοντας ότι η διασφάλιση της αξιοπιστίας της TN είναι μια συνεχής και δυναμική διαδικασία^{[16][21]}.

4.1.2 «Ηθικά Ευθυγραμμισμένος Σχεδιασμός» του Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (Institute of Electrical and Electronic Engineers - IEEE)

Ο «Ηθικά Ευθυγραμμισμένος Σχεδιασμός» (Ethically Aligned Design - EAD) του Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE) αποτελεί μία παγκόσμια πρωτοβουλία που στοχεύει στην ενσωμάτωση ηθικών αρχών στην ανάπτυξη αυτόνομων και ευφυών συστημάτων (A/IS). Το IEEE συγκεντρώνει ακαδημαϊκούς, ερευνητές, επαγγελματίες του κλάδου, κυβερνήσεις και κοινωνικούς φορείς με στόχο την ανάπτυξη προτύπων και ρυθμιστικών πολιτικών που διασφαλίζουν τη συμμόρφωση των συστημάτων A/IS με τις ανθρώπινες αξίες και τα διεθνή νομικά πλαίσια ^{[8][14]}.

I. Στόχοι του «Ethically Aligned Design - EAD»

Το EAD του IEEE επιδιώκει να καθοδηγήσει τους σχεδιαστές και τους φορείς ανάπτυξης συστημάτων TN και A/IS, ώστε αυτά να είναι ηθικά ευθυγραμμισμένα και κοινωνικά υπεύθυνα. Οι βασικοί στόχοι του περιλαμβάνουν:

1. **Προτεραιότητα στην ανθρώπινη ευημερία:** Οι τεχνολογίες A/IS πρέπει να αναπτύσσονται με επίκεντρο τον άνθρωπο, λαμβάνοντας υπόψη δείκτες ευημερίας που σχετίζονται με τη σωματική, ψυχολογική και κοινωνική ευημερία των χρηστών και των κοινωνιών όπου εφαρμόζονται ^[11].
2. **Ευθύνη και Λογοδοσία:** Οι φορείς ανάπτυξης και οι σχεδιαστές συστημάτων A/IS πρέπει να αναλαμβάνουν ρητά την ευθύνη για τις αποφάσεις που λαμβάνουν τα συστήματά τους. Αυτό περιλαμβάνει μηχανισμούς λογοδοσίας, ώστε τα λάθη ή οι αρνητικές επιπτώσεις να είναι επιλύσιμες και αναστρέψιμες ^{[9][17]}.
3. **Διαφάνεια και Εξηγησιμότητα:** Οι αποφάσεις που λαμβάνονται από τις TN και τα αυτόνομα συστήματα πρέπει να είναι κατανοητές από τους ανθρώπους. Η διαφάνεια αφορά τόσο τη δομή και τη λειτουργία των αλγορίθμων, όσο και τη σαφή επικοινωνία των περιορισμών και των δυνατοτήτων τους ^{[5][7]}.
4. **Αποφυγή Κατάχρησης και Βλάβης:** Το IEEE επισημαίνει την ανάγκη πρόληψης καταχρηστικών χρήσεων των A/IS, καθώς και την ελαχιστοποίηση των κινδύνων που σχετίζονται με παραπληροφόρηση, παραβίαση της ιδιωτικότητας και καταπάτηση θεμελιωδών δικαιωμάτων ^[10].

II. Θεμελιώδεις Αρχές

Το EAD του IEEE στηρίζεται σε θεμελιώδεις αρχές, οι οποίες καθοδηγούν την ανάπτυξη υπεύθυνων και ηθικών συστημάτων A/IS.

1. **Σεβασμός των Ανθρωπίνων Δικαιωμάτων:** Οι τεχνολογίες A/IS πρέπει να σέβονται τα διεθνώς αναγνωρισμένα ανθρώπινα δικαιώματα, διασφαλίζοντας ότι δε θα παραβιάζουν την ανθρώπινη αξιοπρέπεια, τα ατομικά δικαιώματα και την πολιτιστική πολυμορφία ^[6].

2. **Προαγωγή της Ευημερίας:** Οι τεχνολογίες A/IS πρέπει να σχεδιάζονται με στόχο την προώθηση της ευημερίας της κοινωνίας, λαμβάνοντας υπόψη κοινωνικές, περιβαλλοντικές και οικονομικές διαστάσεις^[12].
3. **Υπευθυνότητα και Διαφάνεια:** Οι σχεδιαστές πρέπει να διασφαλίζουν τη λογοδοσία και την αναστρεψιμότητα των αποφάσεων που λαμβάνονται από τα συστήματα τους. Επιπλέον, πρέπει να παρέχουν εξηγήσιμες αποφάσεις, ώστε να αποφεύγονται φαινόμενα αλγοριθμικής προκατάληψης ^{[13][15]}.

III. Εφαρμογές και Κατευθυντήριες Γραμμές

Το IEEE, μέσω του EAD, προτείνει συγκεκριμένες κατευθυντήριες γραμμές για την ενσωμάτωση ηθικών κριτηρίων στη διακυβέρνηση των A/IS.

1. **Πλαίσια Διακυβέρνησης:** Η προτεινόμενη πολιτική του IEEE περιλαμβάνει τη δημιουργία ρυθμιστικών πλαισίων που θα επιβλέπουν τα συστήματα TN και A/IS, διασφαλίζοντας τη συμμόρφωση με ηθικές αρχές και διεθνείς συνθήκες ^{[16][21]}.
2. **Πολιτικές για την Εκπαίδευση και Ευαισθητοποίηση:** Το EAD τονίζει τη σημασία της εκπαίδευσης των σχεδιαστών, των προγραμματιστών αλλά και του κοινού σχετικά με τις κοινωνικές και ηθικές επιπτώσεις των A/IS. Αυτό περιλαμβάνει προγράμματα ηθικής εκπαίδευσης για επιστήμονες δεδομένων και μηχανικούς, ώστε να ενσωματώνουν τις αρχές αυτές στον σχεδιασμό τους ^[22].

4.1.3 «Ηθικές αρχές για την Τεχνητή Νοημοσύνη και την Ανάλυση Δεδομένων» του Σύνδεσμου Βιομηχανίας Λογισμικού και Πληροφορικής (Software and Information Industry Association - SIIA)

Ο Σύνδεσμος Βιομηχανίας Λογισμικού και Πληροφορικής (Software and Information Industry Association - SIIA) έχει θεσπίσει ένα σύνολο ηθικών αρχών που καθοδηγούν τη χρήση της Τεχνητής Νοημοσύνης (TN) και της Ανάλυσης Δεδομένων (Data Analytics). Οι κατευθυντήριες γραμμές του SIIA επιδιώκουν να διασφαλίσουν ότι οι τεχνολογίες αυτές χρησιμοποιούνται υπεύθυνα, διαφανώς και με σεβασμό στα ανθρώπινα δικαιώματα, παρέχοντας ένα ρυθμιστικό και ηθικό πλαίσιο για οργανισμούς, κυβερνήσεις και ερευνητές ^{[7][15]}.

I. Στόχοι των Αρχών

Οι αρχές του SIIA αποσκοπούν στην αντιμετώπιση κρίσιμων ζητημάτων που προκύπτουν από την ευρεία χρήση της TN και της ανάλυσης δεδομένων. Οι βασικοί στόχοι του πλαισίου περιλαμβάνουν:

1. Υπευθυνότητα στη χρήση και ανάπτυξη της TN

Επιδιώκεται η ανάπτυξη τεχνολογιών που ανταποκρίνονται στις ανάγκες της κοινωνίας, προάγουν το δημόσιο συμφέρον και μειώνουν τους κινδύνους κακής χρήσης. Η υπευθυνότητα περιλαμβάνει τη λήψη μέτρων για την πρόληψη των αρνητικών συνεπειών και τη διασφάλιση ότι οι αποφάσεις που λαμβάνονται από τέτοιες τεχνολογίες είναι δίκαιες και επεξηγήσιμες^{[3][9]}.

2. **Διαφάνεια στην επεξεργασία δεδομένων και στη λειτουργία αλγορίθμων**
Η διαφάνεια αναφέρεται στη δυνατότητα κατανόησης του τρόπου λειτουργίας των αλγορίθμων, των δεδομένων που χρησιμοποιούνται και των παραμέτρων που επηρεάζουν την ανάλυση και τις αποφάσεις. Με τον τρόπο αυτό, ενισχύεται η εμπιστοσύνη των χρηστών και των φορέων ανάπτυξης^[11].
3. **Δικαιοσύνη και Αμεροληψία**
Οι τεχνολογίες ΤΝ και ανάλυσης δεδομένων πρέπει να διασφαλίζουν ότι δεν υπάρχουν προκαταλήψεις και διακρίσεις που να ενισχύουν τις ανισότητες. Οι αποφάσεις πρέπει να βασίζονται σε αντικειμενικά και αμερόληπτα δεδομένα^{[5][10]}.
4. **Αξιοπιστία και Ασφάλεια**
Οι τεχνολογίες αυτές πρέπει να σχεδιάζονται έτσι ώστε να ελαχιστοποιείται ο κίνδυνος σφαλμάτων και παραβιάσεων. Εξασφαλίζεται ότι είναι αξιόπιστες, ασφαλείς και λειτουργούν όπως προβλέπεται, προστατεύοντας τα δεδομένα και την ιδιωτικότητα των χρηστών^{[14][16]}.

II. Θεμελιώδεις Αρχές

Οι ηθικές αρχές του ΣΙΑ βασίζονται στις εξής θεμελιώδεις κατευθύνσεις:

1. **Προστασία των Ανθρωπίνων Δικαιωμάτων**
Οι τεχνολογίες ΤΝ και ανάλυσης δεδομένων πρέπει να σέβονται τα διεθνώς αναγνωρισμένα ανθρώπινα δικαιώματα, όπως η προστασία της ιδιωτικότητας, η διατήρηση της αξιοπρέπειας και η ισότιμη μεταχείριση όλων των ανθρώπων. Επιπλέον, απαιτείται η διασφάλιση ότι τα δεδομένα δε χρησιμοποιούνται με τρόπο που παραβιάζει την πολιτιστική πολυμορφία και τις ατομικές ελευθερίες^[8].
2. **Ευθύνη των Οργανισμών και Ανθρώπινη Εποπτεία**
Οι οργανισμοί που σχεδιάζουν ή χρησιμοποιούν τέτοιες τεχνολογίες πρέπει να είναι υπεύθυνοι για τις επιπτώσεις τους. Παράλληλα, οι τεχνολογίες πρέπει να συνοδεύονται από ανθρώπινη εποπτεία, ώστε να προλαμβάνονται περιπτώσεις κατάχρησης ή σφαλμάτων^{[4][12]}.
3. **Ασφάλεια και Προστασία Ιδιωτικότητας**
Οι διαδικασίες ανάλυσης δεδομένων πρέπει να είναι σύμφωνες με τα υψηλότερα πρότυπα προστασίας δεδομένων, διασφαλίζοντας την ιδιωτικότητα των χρηστών και την ασφάλεια των πληροφοριών^{[6][13]}.
4. **Διαφάνεια και Επεξηγησιμότητα**
Οι αλγόριθμοι πρέπει να είναι κατανοητοί και επεξηγήσιμοι, ώστε οι χρήστες και τα ενδιαφερόμενα μέρη να μπορούν να αντιληφθούν τη διαδικασία λήψης αποφάσεων. Αυτό βοηθά στην οικοδόμηση εμπιστοσύνης και στη λογοδοσία^[2].

III. Εφαρμογές και Κατευθυντήριες Γραμμές

Το ΣΙΑ παρέχει ένα σύνολο πρακτικών κατευθυντήριων γραμμών για την εφαρμογή των αρχών:

1. Διακυβέρνηση και Κανονιστικά Πλαίσια

Ενθαρρύνεται η θέσπιση κανονιστικών πλαισίων που να διέπουν τη λειτουργία των τεχνολογιών ΤΝ και ανάλυσης δεδομένων. Τα πλαίσια αυτά πρέπει να είναι ευέλικτα και να προσαρμόζονται στις τεχνολογικές εξελίξεις, εξασφαλίζοντας τη συμμόρφωση με τις ηθικές αρχές^[19].

2. Προώθηση της Εκπαίδευσης

Η εκπαίδευση επαγγελματιών, οργανισμών και κοινού σχετικά με τις κοινωνικές και ηθικές επιπτώσεις της ΤΝ και της ανάλυσης δεδομένων αποτελεί βασική προτεραιότητα. Η εκπαίδευση μπορεί να περιλαμβάνει κατανόηση της διαφάνειας, των προκαταλήψεων και των κινδύνων των δεδομένων^[18].

3. Αξιολόγηση Αντίκτυπου

Εφαρμόζονται διαδικασίες αξιολόγησης αντίκτυπου για την ανάλυση των επιπτώσεων των τεχνολογιών, τόσο κατά τη φάση του σχεδιασμού όσο και κατά τη χρήση τους^[20].

4.2 Κανονισμοί και Νομικά Πλαίσια

Η αυξανόμενη ενσωμάτωση της Τεχνητής Νοημοσύνης (ΤΝ) σε ποικίλους τομείς της κοινωνίας έχει καταστήσει επιτακτική τη ρύθμιση της ανάπτυξης και χρήσης της μέσω νομικών κανονισμών και πλαισίων. Αυτά τα ρυθμιστικά πλαίσια στοχεύουν όχι μόνο στη διασφάλιση της ασφάλειας, της διαφάνειας και της δικαιοσύνης, αλλά και στην αντιμετώπιση σύνθετων ηθικών ζητημάτων που προκύπτουν από τη χρήση τεχνολογιών ΤΝ^{[11][2]}. Σε αυτήν την ενότητα, αναλύονται σημαντικά νομοθετικά παραδείγματα και οι σχετικές προκλήσεις που προκύπτουν.

4.2.1 Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR)

Ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR), που τέθηκε σε εφαρμογή στην Ευρωπαϊκή Ένωση το 2018, αποτελεί πρότυπο για τη ρύθμιση της προστασίας των προσωπικών δεδομένων^[3]. Ο GDPR διασφαλίζει ότι τα δεδομένα που χρησιμοποιούνται από τα συστήματα ΤΝ επεξεργάζονται με διαφάνεια και σύμφωνα με τα δικαιώματα των πολιτών. Απαιτεί τη συγκατάθεση των χρηστών για τη συλλογή και επεξεργασία δεδομένων, καθώς και την παροχή δυνατότητας διαγραφής των δεδομένων τους ("δικαίωμα στη λήθη")^[4].

Μία από τις μεγαλύτερες προκλήσεις στην εφαρμογή του GDPR είναι η διαχείριση δεδομένων που χρησιμοποιούνται για την εκπαίδευση μοντέλων μηχανικής μάθησης. Τα μοντέλα ΤΝ συχνά απαιτούν τεράστιες ποσότητες δεδομένων, κάτι που έρχεται σε σύγκρουση με τις αυστηρές ρυθμίσεις για τη συλλογή και αποθήκευση δεδομένων προσωπικού χαρακτήρα^[5].

Για παράδειγμα, στην ανάπτυξη συστημάτων πρόβλεψης ιατρικών διαγνώσεων, οι ερευνητές χρειάζονται μεγάλα σύνολα δεδομένων. Ωστόσο, ο GDPR περιορίζει τη χρήση προσωπικών δεδομένων, επιβάλλοντας ανωνυμοποίηση και αυστηρούς ελέγχους πρόσβασης^[6].

Επίσης, η απαίτηση για εξηγήσιμες αποφάσεις αποτελεί σημείο σύγκρουσης με τις δυσκολίες διαφάνειας που παρουσιάζουν τα βαθιά νευρωνικά δίκτυα. Ένα σύστημα ΤΝ μπορεί να

δώσει διαφορετικά αποτελέσματα με τρόπους που οι προγραμματιστές δεν μπορούν εύκολα να εξηγήσουν^[7].

Οι βασικές αρχές του περιλαμβάνουν:

- **Διαφάνεια:** Οι χρήστες πρέπει να γνωρίζουν πώς χρησιμοποιούνται τα δεδομένα τους.
- **Λογοδοσία:** Οι οργανισμοί είναι υπεύθυνοι για τη συμμόρφωση με τον κανονισμό.
- **Προστασία "by design":** Τα συστήματα πρέπει να σχεδιάζονται με ενσωματωμένα μέτρα προστασίας της ιδιωτικότητας.

Η εφαρμογή του GDPR έχει αναγκάσει τους παρόχους ΤΝ να προσαρμοστούν μέσω συγκεκριμένων μηχανισμών και διαδικασιών:

1. Μηχανισμοί για τη μείωση προκαταλήψεων:

- **Ανάλυση δεδομένων εκπαίδευσης:** Οι πάροχοι εξετάζουν προσεκτικά τα σύνολα δεδομένων που χρησιμοποιούν για την εκπαίδευση αλγορίθμων, εξασφαλίζοντας ότι δεν περιλαμβάνουν συστηματικές προκαταλήψεις^[8].
- **Αλγοριθμικοί έλεγχοι:** Χρήση τεχνικών όπως τα αντίστροφα audits για την αναγνώριση προκαταλήψεων σε δεδομένα.
- **Διαφοροποίηση δεδομένων:** Προστίθενται περισσότερα δεδομένα από υποεκπροσωπούμενες ομάδες για τη βελτίωση της ισορροπίας^[9].

2. Διαφάνεια στους αλγορίθμους:

- **Επεξήγηση αποτελεσμάτων:** Οι πάροχοι υποχρεούνται να παρέχουν κατανοητές εξηγήσεις για τον τρόπο λήψης αποφάσεων από τους αλγορίθμους.
- **Αποκαλύψεις πολιτικής:** Οι οργανισμοί δημοσιεύουν λεπτομέρειες για τον τρόπο με τον οποίο επεξεργάζονται δεδομένα.
- **Εργαλεία αξιολόγησης:** Δημιουργούνται εργαλεία για να μπορούν οι χρήστες να ελέγχουν τη λειτουργία των συστημάτων^[10].

3. Μέτρα για την αποτροπή διαρροών δεδομένων:

1. **Κρυπτογράφηση:** Όλα τα δεδομένα προστατεύονται με ισχυρούς αλγορίθμους κρυπτογράφησης τόσο κατά την αποθήκευση όσο και κατά τη μεταφορά^[11].
2. **Περιορισμένη πρόσβαση:** Η πρόσβαση στα δεδομένα δίνεται μόνο σε εξουσιοδοτημένα άτομα, μειώνοντας τον κίνδυνο διαρροών.
3. **Τακτικοί έλεγχοι ασφαλείας:** Διεξάγονται συστηματικοί έλεγχοι για την αξιολόγηση της ασφάλειας των συστημάτων και της προστασίας των δεδομένων^[12].

Η επίδραση του GDPR στην ΤΝ είναι πολυδιάστατη. Από τη μία, διασφαλίζει ότι τα συστήματα είναι σχεδιασμένα για να προστατεύουν τους χρήστες, ενώ από την άλλη επιβάλλει αυστηρούς περιορισμούς που μπορεί να επιβραδύνουν την καινοτομία^[13].

Θετικά αποτελέσματα:

- Ενίσχυση εμπιστοσύνης των χρηστών προς τις τεχνολογίες ΤΝ.
- Καλύτερη διαχείριση προσωπικών δεδομένων από τις εταιρείες.

Προκλήσεις:

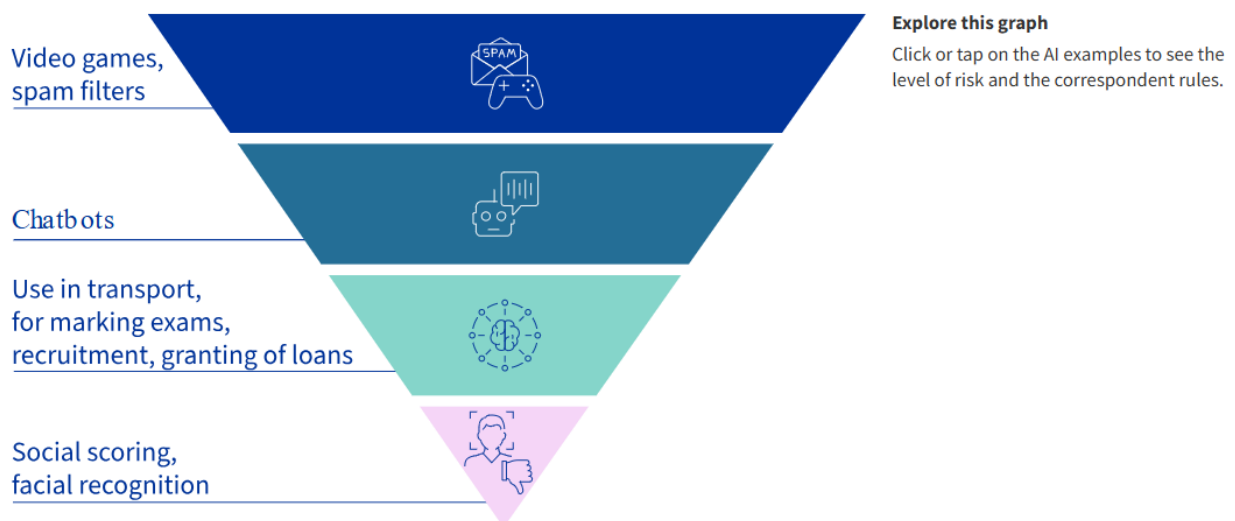
- Δυσκολία στη διαφάνεια των πολύπλοκων αλγορίθμων ΤΝ.
- Υψηλό κόστος συμμόρφωσης για τις εταιρείες που αναπτύσσουν ΤΝ^[14].

Έρευνες δείχνουν ότι ο GDPR συμβάλλει στην ενίσχυση της εμπιστοσύνης των χρηστών και στη δημιουργία ενός ασφαλέστερου περιβάλλοντος για τη χρήση δεδομένων στην ΤΝ^[15].

4.2.2 Νομοθεσίες για τη Λογοδοσία και την Ασφάλεια

Το AI Act της Ευρωπαϊκής Ένωσης είναι μια φιλόδοξη νομοθετική πρωτοβουλία που στοχεύει στη δημιουργία ενός συνεκτικού πλαισίου για τη ρύθμιση της ΤΝ, με έμφαση στη λογοδοσία και την ασφάλεια^[10]. Βασίζεται στη διάκριση των εφαρμογών ΤΝ σε τέσσερα επίπεδα κινδύνου^[19]:

- **Εφαρμογές ελάχιστου κινδύνου:** Όπως τα συστήματα φίλτρων ανεπιθύμητης αλληλογραφίας (spam filters) και οι αλγόριθμοι που χρησιμοποιούνται σε βιντεοπαιχνίδια. Αυτές οι εφαρμογές δεν απαιτούν πρόσθετους κανονισμούς.
- **Εφαρμογές περιορισμένου κινδύνου:** Όπως τα chatbots, που απαιτούν διαφάνεια για να ενημερώνουν τους χρήστες ότι αλληλεπιδρούν με μηχανές.
- **Εφαρμογές υψηλού κινδύνου:** Όπως οι αλγόριθμοι που χρησιμοποιούνται για ιατρικές διαγνώσεις, στη μεταφορά, στην αυτόματη αξιολόγηση εξετάσεων, στις προσλήψεις ή στην έγκριση δανείων. Αυτές υπόκεινται σε αυστηρούς ελέγχους συμμόρφωσης, συμπεριλαμβανομένων αξιολογήσεων κινδύνου, διαφάνειας και ασφάλειας^[14].
- **Απαγορευμένες εφαρμογές:** Όπως συστήματα ΤΝ για κοινωνική βαθμολόγηση (social scoring) ή τεχνολογίες αναγνώρισης προσώπου για μαζική παρακολούθηση. Αυτές θεωρούνται απαράδεκτες λόγω των ηθικών και κοινωνικών κινδύνων που ενέχουν^[25].



Εικόνα 2: Επίπεδα Κινδύνου Εφαρμογών ΤΝ

Το AI Act επιβάλλει στις εταιρείες και τους παρόχους ΤΝ τις παρακάτω υποχρεώσεις^[17]:

- **Αξιολόγηση Κινδύνου:** Ενδελεχή ανάλυση των πιθανών επιπτώσεων από τη χρήση των συστημάτων τους.
- **Διαφάνεια:** Υποχρέωση ενημέρωσης των χρηστών για τον τρόπο λειτουργίας και τα δεδομένα που χρησιμοποιούνται.
- **Εξωτερικούς Ελέγχους:** Διενέργεια ελέγχων από ανεξάρτητους φορείς για τη διασφάλιση της συμμόρφωσης.

- **Δημιουργία Μηχανισμών Αναφοράς:** Παροχή εργαλείων στους χρήστες για την αναφορά προβλημάτων ή παραβιάσεων.

Η συμπληρωματικότητα του **AI Act** με τον **GDPR** είναι εμφανής^[15].

- Ο **GDPR** επικεντρώνεται στην προστασία των προσωπικών δεδομένων και των δικαιωμάτων των πολιτών.
- Το **AI Act** διευρύνει το πεδίο ρύθμισης, καλύπτοντας ηθικές, κοινωνικές και τεχνικές πτυχές της ανάπτυξης ΤΝ^[11].

Μαζί, οι δύο κανονισμοί δημιουργούν ένα ισχυρό πλαίσιο για την προστασία των πολιτών και την ενίσχυση της εμπιστοσύνης στην τεχνολογία.

Σε διεθνές επίπεδο, διάφορες χώρες έχουν εισαγάγει νομικά πλαίσια που αποσκοπούν στη ρύθμιση της λογοδοσίας και της ασφάλειας των συστημάτων ΤΝ^[16].

- **Algorithmic Accountability Act (ΗΠΑ):** Υποχρεώνει τις εταιρείες να αξιολογούν και να διορθώνουν τις μεροληπτικές επιδράσεις των αλγορίθμων τους.
- **Ευρωπαϊκό AI Act:** Καθορίζει επίπεδα κινδύνου και απαιτεί αυστηρότερα κριτήρια για τις υψηλού κινδύνου εφαρμογές, όπως η ιατρική διάγνωση και οι αυτοματοποιημένες αποφάσεις προσλήψεων ^[13].
- **Νόμοι στην Κίνα για την παρακολούθηση ΤΝ:** Εστιάζουν κυρίως στον κρατικό έλεγχο και τη διασφάλιση εθνικών συμφερόντων ^[23].

Ωστόσο, οι νόμοι αυτοί αντιμετωπίζουν δυσκολίες λόγω της ταχείας εξέλιξης της ΤΝ, καθώς και των προκλήσεων στον εντοπισμό και τη διόρθωση της αλγοριθμικής προκατάληψης^[12].

4.2.3 Παγκόσμια Συγκριτική Ανάλυση

Η σύγκριση νομοθεσιών, όπως ο **GDPR** και ο **CCPA**, αναδεικνύει τις διαφορές στις προσεγγίσεις της Ευρωπαϊκής Ένωσης και των Ηνωμένων Πολιτειών όσον αφορά την προστασία προσωπικών δεδομένων^[10].

Γενική Φιλοσοφία των Νομοθεσιών

- **GDPR (ΕΕ):** Προστασία δικαιωμάτων χρηστών, διαφάνεια, λογοδοσία και σαφής ρύθμιση των αυτοματοποιημένων αποφάσεων.
- **CCPA (ΗΠΑ):** Ενημέρωση και έλεγχος καταναλωτών στα δεδομένα τους, με έμφαση στην αποτροπή πώλησης προσωπικών πληροφοριών.
- **PIPL (Κίνα):** Κρατικός έλεγχος και αυστηρή ρύθμιση των εταιρειών που επεξεργάζονται δεδομένα, με ιδιαίτερη έμφαση στην εθνική ασφάλεια και την κυβερνητική εποπτεία.

California Consumer Privacy Act (CCPA)

Ο **California Consumer Privacy Act (CCPA)** είναι ένας νόμος προστασίας προσωπικών δεδομένων που εφαρμόζεται στην Καλιφόρνια των Ηνωμένων Πολιτειών και επικεντρώνεται στην προστασία των δικαιωμάτων των καταναλωτών. Συγκεκριμένα, προσφέρει:

- **Δικαίωμα πρόσβασης:** Οι καταναλωτές μπορούν να γνωρίζουν ποια προσωπικά δεδομένα συλλέγονται για αυτούς.
- **Δικαίωμα διαγραφής:** Δίνεται η δυνατότητα στους καταναλωτές να ζητήσουν τη διαγραφή των δεδομένων τους.
- **Δικαίωμα εξαίρεσης από την πώληση δεδομένων:** Οι επιχειρήσεις υποχρεούνται να παρέχουν επιλογή «Do Not Sell My Personal Information.», επιτρέποντας στους χρήστες να αποτρέπουν τη μεταπώληση των δεδομένων τους [15].

Ωστόσο, παρά τις δυνατότητες που προσφέρει, ο CCPA δε διαθέτει εξίσου αυστηρά μέτρα για τη διαχείριση αλγορίθμων ή τη λογοδοσία των παρόχων ΤΝ.

- **Παράδειγμα:** Η εταιρεία Sephora το 2022 κλήθηκε να πληρώσει 1,2 εκατομμύρια δολάρια πρόστιμο επειδή δεν ενημέρωνε επαρκώς τους πελάτες της για την πώληση των προσωπικών τους δεδομένων, παραβιάζοντας έτσι τον CCPA (California Attorney General, 2022).

General Data Protection Regulation (GDPR)

Ο GDPR της Ευρωπαϊκής Ένωσης προσφέρει ένα πιο αυστηρό και συνεκτικό πλαίσιο, το οποίο εστιάζει στη διασφάλιση της ιδιωτικότητας και στη λογοδοσία των οργανισμών. Ο GDPR επιβάλλει:

- **Δικαίωμα στη λήθη:** Οι χρήστες μπορούν να ζητήσουν τη διαγραφή των δεδομένων τους.
- **Διαφάνεια στις αυτοματοποιημένες αποφάσεις:** Οι πολίτες έχουν το δικαίωμα να γνωρίζουν πώς οι αλγόριθμοι επεξεργάζονται τα δεδομένα τους.
- **Βαρύτατες κυρώσεις:** Οι επιχειρήσεις που δε συμμορφώνονται μπορούν να αντιμετωπίσουν πρόστιμα έως 20 εκατομμύρια ευρώ ή το 4% του ετήσιου κύκλου εργασιών.
- **Παράδειγμα:** Η Meta (Facebook) το 2023 δέχτηκε πρόστιμο 1,2 δισ. ευρώ από την ΕΕ επειδή μετέφερε δεδομένα χρηστών στις ΗΠΑ χωρίς να πληροί τα πρότυπα προστασίας προσωπικών δεδομένων του GDPR [24].

Personal Information Protection Law (PIPL)

Ο Personal Information Protection Law (PIPL) της Κίνας, ο οποίος τέθηκε σε ισχύ τον Νοέμβριο του 2021, αποτελεί το πρώτο ολοκληρωμένο πλαίσιο προστασίας προσωπικών δεδομένων στη χώρα. Το PIPL έχει σημαντικές διαφορές από το GDPR και τον CCPA, καθώς ενσωματώνει κρατικό έλεγχο και απαιτεί αυστηρότερες ρυθμίσεις για πολυεθνικές εταιρείες που δραστηριοποιούνται στην Κίνα [20].

- **Συναίνεση για συλλογή δεδομένων:** Οι εταιρείες πρέπει να λάβουν ρητή συγκατάθεση των χρηστών πριν συλλέξουν τα προσωπικά τους δεδομένα.
- **Κρατική εποπτεία:** Οι αρχές έχουν την εξουσία να ελέγχουν και να διαγράφουν δεδομένα που θεωρούνται επικίνδυνα για την εθνική ασφάλεια.
- **Απαγόρευση διασυνοριακής μεταφοράς δεδομένων:** Οι επιχειρήσεις πρέπει να αποθηκεύουν τα δεδομένα των Κινέζων χρηστών εντός της Κίνας, εκτός εάν λάβουν ειδική κυβερνητική έγκριση.

- **Παράδειγμα:** Η Didi Global, η μεγαλύτερη εταιρεία διαμοιρασμού οχημάτων της Κίνας, δέχτηκε πρόστιμο 1,2 δισ. δολάρια το 2022 επειδή παραβίασε τις διατάξεις του PIPL για διασυννοριακή μεταφορά δεδομένων, μη αποθηκεύοντας τις πληροφορίες των χρηστών της εντός της Κίνας^[21].

Χαρακτηριστικό	GDPR (ΕΕ)	CCPA (Καλιφόρνια, ΗΠΑ)	PIPL (Κίνα)
Ποιοι προστατεύονται	Όλοι οι πολίτες της ΕΕ	Μόνο κάτοικοι Καλιφόρνιας	Όλοι οι πολίτες της Κίνας
Κάλυψη	Όλες οι εταιρείες που διαχειρίζονται προσωπικά δεδομένα	Μόνο επιχειρήσεις με έσοδα > 25 εκατ. \$ ή 50.000+ χρήστες	Όλες οι εταιρείες που συλλέγουν δεδομένα στην Κίνα
Δικαίωμα διαγραφής δεδομένων	✓ (Δικαίωμα στη λήθη)	✓ (Κατόπιν αιτήματος)	✓ (Απαίτηση κρατικής έγκρισης)
Δικαίωμα ενημέρωσης για χρήση δεδομένων	✓	✓	✓ (Με κρατικό έλεγχο)
Διαφάνεια αλγορίθμων & αυτοματοποιημένων αποφάσεων	✓ (Απαιτεί εξηγήσεις)	✗ (Δεν απαιτείται)	✗ (Υπόκειται σε κρατική ρύθμιση)
Πρόστιμα για μη συμμόρφωση	Έως 20 εκατ. € ή 4% του κύκλου εργασιών	Έως 7.500 \$ ανά παραβίαση ανά χρήστη	Χρηματικά πρόστιμα + Κυβερνητικές κυρώσεις

Πίνακας 5: Συγκριτική Ανάλυση των Κανονισμών GDPR, CCPA και PIPL

Ο GDPR προσφέρει το πιο ισχυρό νομικό πλαίσιο προστασίας δεδομένων, δίνοντας έμφαση στη λογοδοσία και τη διαφάνεια, διασφαλίζοντας ότι οι χρήστες έχουν τον έλεγχο των προσωπικών τους πληροφοριών και ότι οι εταιρείες τηρούν αυστηρούς κανόνες συμμόρφωσης. Αντίθετα, ο CCPA ακολουθεί μια πιο ελαστική προσέγγιση, επικεντρώνοντας στην ενδυνάμωση των καταναλωτών μέσω του δικαιώματος πρόσβασης, διαγραφής και εξαίρεσης από την πώληση των δεδομένων τους, χωρίς όμως να επιβάλλει αυστηρές ρυθμίσεις σχετικά με τη διαφάνεια των αλγορίθμων. Από την άλλη πλευρά, ο PIPL της Κίνας δίνει κυρίαρχο ρόλο στο κράτος, επιβάλλοντας αυστηρούς ελέγχους στις εταιρείες που διαχειρίζονται δεδομένα και απαιτώντας τη διατήρησή τους εντός Κίνας, με έμφαση στην εθνική ασφάλεια και κυβερνητική εποπτεία. Συνολικά, η βασική διαφορά μεταξύ των τριών νομοθεσιών έγκειται στον προσανατολισμό τους: ο GDPR εστιάζει στα δικαιώματα των χρηστών, ο CCPA στην επιλογή των καταναλωτών, ενώ ο PIPL εξυπηρετεί κυρίως κρατικά συμφέροντα ^[22].

4.2.4 Ιδιωτικότητα και Υγεία

Στον τομέα της υγείας, η χρήση ΤΝ υπόκειται σε αυστηρότερες ρυθμίσεις λόγω της ευαισθησίας των δεδομένων. Στις ΗΠΑ, ο Health Insurance Portability and Accountability Act (HIPAA) αποτελεί έναν ομοσπονδιακό νόμο που θεσπίστηκε το 1996, με σκοπό να προστατεύσει την ιδιωτικότητα και την ασφάλεια των ιατρικών πληροφοριών, ενώ παράλληλα διευκολύνει την

ηλεκτρονική μεταφορά και χρήση δεδομένων υγείας για τη βελτίωση της αποδοτικότητας στον τομέα της υγείας^[11]. Ο HIPAA ρυθμίζει:

- **Προστασία Προσωπικών Δεδομένων Υγείας (Protected Health Information - PHI):** Εξασφαλίζει ότι ευαίσθητα δεδομένα υγείας δεν αποκαλύπτονται χωρίς τη συγκατάθεση του ασθενούς, προστατεύοντας την ταυτότητά του^[11].
- **Ασφάλεια Δεδομένων:** Επιβάλλει τεχνικά και φυσικά μέτρα για την προστασία δεδομένων υγείας από μη εξουσιοδοτημένη πρόσβαση, παραβιάσεις ή απώλειες^[12].
- **Δικαιώματα Ασθενών:** Παρέχει στους ασθενείς τη δυνατότητα πρόσβασης στα ιατρικά τους αρχεία και δικαίωμα διόρθωσης τυχόν ανακριβών πληροφοριών.^[24]
- **Κανόνες Μεταφοράς Δεδομένων:** Καθορίζει πρότυπα για τη μετάδοση ηλεκτρονικών δεδομένων υγείας, εξασφαλίζοντας τη διαλειτουργικότητα και τη συμμόρφωση με τις αρχές απορρήτου^[25].

Η συμμόρφωση με τον HIPAA είναι υποχρεωτική για νοσοκομεία, γιατρούς, ασφαλιστικές εταιρείες και άλλους φορείς που διαχειρίζονται δεδομένα υγείας. Στον τομέα της ΤΝ, αυτό σημαίνει ότι τα συστήματα που χρησιμοποιούν δεδομένα ασθενών πρέπει να είναι ανωνυμοποιημένα, ασφαλή και σχεδιασμένα ώστε να συμμορφώνονται πλήρως με τις ρυθμίσεις του νόμου^[13].

- **Παράδειγμα:** Το 2015, η Anthem Inc., μία από τις μεγαλύτερες ασφαλιστικές εταιρείες υγείας των ΗΠΑ, παραβίασε τον HIPAA όταν δέχτηκε κυβερνοεπίθεση, με αποτέλεσμα την έκθεση προσωπικών και ιατρικών δεδομένων 79 εκατομμυρίων ασφαλισμένων. Το περιστατικό αποκάλυψε σοβαρές αδυναμίες στην ασφάλεια δεδομένων της εταιρείας, με την Anthem να καταλήγει σε διακανονισμό 16 εκατομμυρίων δολαρίων με το Υπουργείο Υγείας και Ανθρωπίνων Υπηρεσιών των ΗΠΑ ^[17].

Ο GDPR και ο HIPAA διαφέρουν σημαντικά ως προς την εφαρμογή τους. Ενώ ο GDPR εστιάζει στη διαφάνεια και την προστασία των δικαιωμάτων των πολιτών της ΕΕ, ο HIPAA επικεντρώνεται στη συμμόρφωση των οργανισμών υγείας στις Ηνωμένες Πολιτείες με ειδικά πρότυπα απορρήτου, καθιστώντας τους δύο κανονισμούς συμπληρωματικούς σε διεθνές επίπεδο^[17].

Στον τομέα της υγείας, η χρήση ΤΝ υπόκειται σε αυστηρότερες ρυθμίσεις λόγω της ευαισθησίας των δεδομένων. Στις ΗΠΑ, ο Health Insurance Portability and Accountability Act (HIPAA) διασφαλίζει ότι η χρήση δεδομένων ασθενών από συστήματα ΤΝ γίνεται με απόλυτη προστασία και ασφάλεια.

4.2.5 Διασυννοριακή Συμμόρφωση

Η έλλειψη εναρμόνισης μεταξύ νομοθεσιών, όπως ο GDPR και ο CCPA, προκαλεί σημαντικές προκλήσεις για τις πολυεθνικές εταιρείες. Αυτές καλούνται να προσαρμόσουν τις πρακτικές τους ώστε να ανταποκρίνονται σε διαφορετικές ρυθμιστικές απαιτήσεις, γεγονός που αυξάνει σημαντικά το κόστος συμμόρφωσης ^[10].

Η ανάγκη για παγκόσμια συνεργασία και συντονισμό είναι ιδιαίτερα κρίσιμη, καθώς οι τεχνολογίες ΤΝ έχουν παγκόσμιο αντίκτυπο και δεν περιορίζονται σε εθνικά σύνορα ^[12]. Η έλλειψη μιας ενιαίας νομικής προσέγγισης οδηγεί σε ρυθμιστικά κενά και αλληλοσυγκρουόμενες

απαιτήσεις, δημιουργώντας δυσκολίες για τις εταιρείες που δραστηριοποιούνται σε πολλαπλές δικαιοδοσίες ^[14].

Για την αντιμετώπιση αυτών των ζητημάτων, διεθνείς οργανισμοί όπως ο Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (OECD) και ο Διεθνής Οργανισμός Τυποποίησης (ISO) προωθούν πρωτοβουλίες για την εναρμόνιση των ρυθμιστικών πλαισίων, δίνοντας έμφαση στην ηθική και ασφαλή ανάπτυξη της ΤΝ ^[15].

- **Χαρακτηριστικό Παράδειγμα:**

Ένα από τα πιο πρόσφατα βήματα προς αυτή την κατεύθυνση είναι η ανάπτυξη κοινών αρχών για την ηθική χρήση της ΤΝ, που στοχεύουν στην ενίσχυση της διαλειτουργικότητας μεταξύ διαφορετικών νομικών πλαισίων και την προώθηση παγκόσμιων προτύπων συμμόρφωσης ^[13].

4.3 Ηθικές Κατευθυντήριες Γραμμές και Βέλτιστες Πρακτικές

Η εφαρμογή ηθικών αρχών στην Τεχνητή Νοημοσύνη (ΤΝ) δεν περιορίζεται στη διατύπωση θεωρητικών εννοιών, αλλά απαιτεί συγκεκριμένα πρακτικά βήματα για την υλοποίησή τους. Στην ενότητα αυτή, παρουσιάζονται πρακτικές μέθοδοι που επιτρέπουν την ενσωμάτωση ηθικών κατευθυντήριων γραμμών στις διαδικασίες ανάπτυξης και χρήσης της ΤΝ^{[12][14]}.

4.3.1 Μετάφραση Ηθικών Αρχών σε Καθημερινές Πρακτικές

Η εφαρμογή αρχών όπως η διαφάνεια, η εξηγησιμότητα, η αμεροληψία και η προστασία της ιδιωτικότητας, μπορεί να επιτευχθεί μέσω συγκεκριμένων πρακτικών:

- **Διαφάνεια μέσω τεκμηρίωσης:** Η ανάπτυξη συστημάτων ΤΝ πρέπει να συνοδεύεται από αναλυτική τεκμηρίωση που περιγράφει τις επιλογές σχεδιασμού, τις πηγές δεδομένων και τους στόχους του αλγορίθμου^[13].
- **Εξηγησιμότητα με εργαλεία ερμηνείας αλγορίθμων:** Εφαρμογές όπως η LIME (Local Interpretable Model-Agnostic Explanations) βοηθούν στη δημιουργία εύκολα κατανοητών εξηγήσεων για τις αποφάσεις που λαμβάνονται από τα συστήματα^[15].
- **Αμεροληψία μέσω ελέγχων δεδομένων:** Οι προγραμματιστές πρέπει να εφαρμόζουν τεχνικές όπως ο έλεγχος δεδομένων εισόδου για εντοπισμό προκαταλήψεων και η χρήση αντισταθμιστικών αλγορίθμων για τη μείωση της μεροληψίας^[16].
- **Προστασία ιδιωτικότητας μέσω ανωνυμοποίησης:** Η ενσωμάτωση τεχνικών ανωνυμοποίησης, όπως differential privacy, διασφαλίζει την προστασία των προσωπικών δεδομένων^[18].

4.3.2 Παραδείγματα Εφαρμογής

Η μετάφραση ηθικών αρχών σε πρακτική δράση γίνεται εμφανής σε πολλές επιτυχημένες εφαρμογές:

- Στον τομέα της υγείας, εργαλεία όπως το DeepMind της Google ελέγχονται συστηματικά για την εξηγησιμότητα και τη συμμόρφωσή τους με τις κατευθυντήριες γραμμές προστασίας προσωπικών δεδομένων. Το NHS στο Ηνωμένο Βασίλειο το χρησιμοποίησε για τη διάγνωση οφθαλμικών παθήσεων, εγείροντας όμως ζητήματα ιδιωτικότητας και ασφάλειας δεδομένων [20].
- Στον χρηματοπιστωτικό τομέα, η πλατφόρμα Zest AI εφαρμόζει διαφανείς αλγορίθμους πιστοδοτήσεων, ελαχιστοποιώντας τη μεροληψία μέσω της χρήσης αμερόληπτων δεδομένων και αλγοριθμικών ελέγχων [19].
- Οι μηχανές αναζήτησης, όπως η Bing AI, ενσωματώνουν εργαλεία διαχείρισης προκαταλήψεων για τη βελτίωση της εξηγησιμότητας και την πρόληψη διαστρεβλωμένων αποτελεσμάτων [21].

Στην ιατρική, το Watson της IBM χρησιμοποιεί την TN για τη διάγνωση και τη θεραπεία ασθενειών, παρέχοντας στους γιατρούς τεκμηριωμένες προτάσεις βασισμένες σε μεγάλες βάσεις ιατρικών δεδομένων [22].

4.3.3 Βέλτιστες Πρακτικές

Η εφαρμογή αυτών των αρχών προϋποθέτει τη δέσμευση οργανισμών και επιχειρήσεων σε στρατηγικές που εξασφαλίζουν την ηθική χρήση της TN. Οι βασικές βέλτιστες πρακτικές περιλαμβάνουν:

- **Εκπαίδευση και Ευαισθητοποίηση:** Οι οργανισμοί πρέπει να επενδύουν στην εκπαίδευση των εργαζομένων και των χρηστών για την κατανόηση των ηθικών διαστάσεων της TN [15].
- **Αξιολόγηση Αντίκτυπου:** Η ενσωμάτωση διαδικασιών για την αξιολόγηση των ηθικών και κοινωνικών επιπτώσεων της TN αποτελεί προαπαιτούμενο για την υπεύνητη χρήση της [17].
- **Δημιουργία ομάδων ηθικής TN:** Εξειδικευμένες ομάδες μπορούν να παρέχουν συνεχή παρακολούθηση των αλγορίθμων και να διασφαλίζουν τη συμμόρφωση με τα ηθικά πρότυπα [18].
- **Εξωτερικός Έλεγχος:** Οι οργανισμοί πρέπει να ενσωματώνουν μηχανισμούς εξωτερικού ελέγχου, ώστε να διασφαλίζεται η συμμόρφωση με τις ηθικές κατευθυντήριες γραμμές [19].
- **Πολυφωνία και Συμμετοχικότητα:** Σχεδιασμός συστημάτων TN που ενσωματώνουν τη συμμετοχή διαφορετικών κοινωνικών ομάδων για τη μείωση του κινδύνου μεροληψίας [20].

4.3.4 Δικαιοσύνη στην Τεχνητή Νοημοσύνη (AI Fairness)

Η δικαιοσύνη στην Τεχνητή Νοημοσύνη (AI Fairness) αφορά τη δημιουργία αλγορίθμων που λαμβάνουν αμερόληπτες, αντικειμενικές και δίκαιες αποφάσεις, ανεξάρτητα από παράγοντες όπως το φύλο, η φυλή ή το κοινωνικοοικονομικό υπόβαθρο. Ωστόσο, η επίτευξη δίκαιης TN αποτελεί πρόκληση, καθώς πολλοί αλγόριθμοι αντικατοπτρίζουν προκαταλήψεις που προκύπτουν από τα ίδια τα δεδομένα εκπαίδευσής τους [8][9].

Προβλήματα στην AI Fairness

1. Μεροληψία στα δεδομένα εκπαίδευσης

Οι αλγόριθμοι βασίζονται σε ιστορικά δεδομένα, τα οποία συχνά περιέχουν προϋπάρχουσες κοινωνικές προκαταλήψεις. Αν αυτά τα δεδομένα δεν ελεγχθούν και διορθωθούν, οι αποφάσεις που λαμβάνει το σύστημα μπορεί να είναι άδικες και να ενισχύσουν διακρίσεις.

- **Παράδειγμα:** Το 2018, η Amazon ανέπτυξε ένα σύστημα προσλήψεων που απορρίφθηκε, καθώς είχε μεροληψία κατά των γυναικών. Ο αλγόριθμος είχε εκπαιδευτεί σε βιογραφικά ανδρών, με αποτέλεσμα να βαθμολογεί χαμηλότερα τα γυναικεία βιογραφικά ^[12].

2. Έλλειψη διαφάνειας και επεξηγησιμότητας

Πολλά αλγοριθμικά συστήματα λειτουργούν ως "μαύρα κουτιά", όπου οι τελικοί χρήστες δεν μπορούν να κατανοήσουν πώς λαμβάνονται οι αποφάσεις. Αυτό καθιστά δύσκολη τη διάγνωση και τη διόρθωση μεροληψιών.

- **Παράδειγμα:** Οι αλγόριθμοι πιστοληπτικής αξιολόγησης συχνά απορρίπτουν δανειολήπτες χωρίς να παρέχουν αιτιολόγηση, γεγονός που δημιουργεί αθέμιτες οικονομικές διακρίσεις ^[13].

3. Ανισορροπία στην εκπροσώπηση δεδομένων

Σε πολλές περιπτώσεις, τα δεδομένα εκπαίδευσης δεν αντιπροσωπεύουν ολόκληρο τον πληθυσμό. Αυτό μπορεί να οδηγήσει σε μειωμένη ακρίβεια σε ομάδες που υποεκπροσωπούνται, ενισχύοντας τις κοινωνικές ανισότητες.

- **Παράδειγμα:** Ένα μοντέλο αναγνώρισης προσώπων της IBM βρέθηκε να έχει πολύ χαμηλότερη ακρίβεια στον εντοπισμό ατόμων με σκούρο δέρμα σε σύγκριση με άτομα με ανοιχτόχρωμο δέρμα ^[9].

Μέθοδοι αντιμετώπισης των προβλημάτων AI Fairness

- **Έλεγχος και καθαρισμός δεδομένων:** Διεξαγωγή αναλύσεων για την ανίχνευση προκαταλήψεων και χρήση τεχνικών όπως το reweighting για τη διόρθωση τους ^[11].
- **Ανάπτυξη εξηγήσιμων μοντέλων (Explainable AI - XAI):** Τα συστήματα πρέπει να παρέχουν κατανοητές εξηγήσεις για τις αποφάσεις τους ^[14].
- **Ενίσχυση εκπροσώπησης μειονοτήτων στα δεδομένα:** Συλλογή περισσότερων δεδομένων από υποεκπροσωπούμενες ομάδες για βελτίωση της ακρίβειας των μοντέλων ^[15].
- **Διεθνή πρότυπα και κατευθυντήριες γραμμές:** Χρήση ρυθμιστικών πλαισίων, όπως τα IEEE πρότυπα ^[16].

4.3.5 Τα Πρότυπα IEEE για Ηθική ΤΝ

Ο Διεθνής Οργανισμός Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών (IEEE) έχει αναπτύξει ένα σύνολο προτύπων που στοχεύουν στην ανάπτυξη ηθικών και κοινωνικά υπεύθυνων συστημάτων ΤΝ. Τα πρότυπα αυτά παρέχουν κατευθυντήριες γραμμές για την εξηγήσιμη, διαφανή και δίκαιη χρήση των αλγορίθμων, διασφαλίζοντας τη συμμόρφωση των τεχνολογικών λύσεων με τις ηθικές απαιτήσεις ^{[17][18]}.

Βασικά IEEE Πρότυπα για Ηθική ΤΝ

- **IEEE 7000 – Μοντέλο Ηθικού Σχεδιασμού (Ethical System Design)**
 - Εστιάζει στην ενσωμάτωση ηθικών αξιών από το στάδιο της ανάπτυξης ενός συστήματος ΤΝ.
 - Διασφαλίζει τη λογοδοσία και την ηθική συμμόρφωση καθ' όλη τη διάρκεια του κύκλου ζωής του συστήματος.
- **IEEE 7010 – Δείκτες Κοινωνικής Ευημερίας (Well-being Metrics for AI)**
 - Παρέχει ένα πλαίσιο αξιολόγησης του κοινωνικού αντίκτυπου της ΤΝ.
 - Εξετάζει ζητήματα όπως δικαιοσύνη, βιωσιμότητα, κοινωνική ένταξη και ιδιωτικότητα.
- **IEEE 7001 – Εξηγήσιμη ΤΝ (Explainable AI - XAI)**
 - Καθορίζει αρχές για τη δημιουργία αλγορίθμων που προσφέρουν κατανοητές εξηγήσεις στους χρήστες.
 - Στόχος του είναι η αύξηση της διαφάνειας και της εμπιστοσύνης στα αλγοριθμικά συστήματα.
- **IEEE 7002 – Πρότυπο για την Εμπιστοσύνη στην ΤΝ (AI Trustworthiness Metrics)**
 - Ορίζει κριτήρια που καθορίζουν πόσο αξιόπιστο είναι ένα σύστημα ΤΝ.
 - Αξιολογεί παράγοντες όπως ακρίβεια, ασφάλεια, ανθεκτικότητα και απουσία προκαταλήψεων.

Σημασία των IEEE Προτύπων για την ΤΝ

- **Βελτίωση της ηθικής συμμόρφωσης:** Τα IEEE πρότυπα παρέχουν ένα διεθνές σημείο αναφοράς για την ανάπτυξη υπεύθυνων αλγορίθμων ΤΝ^[19].
- **Αύξηση της εμπιστοσύνης του κοινού:** Η ενσωμάτωση ηθικών προτύπων ενισχύει την αποδοχή των συστημάτων ΤΝ από την κοινωνία^[20].
- **Συμμόρφωση με ρυθμιστικά πλαίσια:** Τα πρότυπα μπορούν να βοηθήσουν τις εταιρείες να συμμορφωθούν με νόμους όπως ο GDPR και το AI Act της ΕΕ^[21].

4.4 Προκλήσεις στην Εφαρμογή των Κατευθυντήριων Γραμμών

Η εφαρμογή των κατευθυντήριων γραμμών για την ηθική χρήση της ΤΝ συναντά σημαντικές προκλήσεις σε διάφορα επίπεδα. Παρότι οι κατευθυντήριες αρχές, όπως η διαφάνεια, η εξηγησιμότητα, η αμεροληψία και η προστασία της ιδιωτικότητας, παρέχουν ένα θεωρητικό πλαίσιο, η εφαρμογή τους στην πράξη αποδεικνύεται σύνθετη και απαιτεί μια πολυεπίπεδη προσέγγιση. Στην ενότητα αυτή εξετάζονται τα τεχνικά, νομικά, κοινωνικά και επιχειρηματικά εμπόδια, καθώς και οι στρατηγικές για την αντιμετώπισή τους^{[8][9]}.

4.4.1 Τεχνικές Προκλήσεις

Η πολυπλοκότητα των συστημάτων ΤΝ αποτελεί έναν από τους σημαντικότερους φραγμούς για την εφαρμογή ηθικών αρχών. Τα μεγάλα γλωσσικά μοντέλα (LLMs) και άλλες προηγμένες τεχνολογίες συχνά λειτουργούν ως "μαύρα κουτιά", στα οποία η διαδικασία λήψης αποφάσεων είναι δύσκολο να παρακολουθηθεί και να κατανοηθεί, ακόμα και από τους ίδιους τους δημιουργούς τους. Αυτή η έλλειψη εξηγησιμότητας περιορίζει τη διαφάνεια και εγείρει ανησυχίες για την υπευθυνότητα και την αξιοπιστία των συστημάτων^[10].

Παρόλο που αυτά τα εργαλεία έχουν συμβάλει στη βελτίωση της εξηγησιμότητας, απαιτείται περαιτέρω έρευνα για την ενίσχυση της αποτελεσματικότητάς τους, ειδικά σε σύνθετα συστήματα ΤΝ^[12]. Η υιοθέτηση εργαλείων ανοιχτού κώδικα που προωθούν τη διαφάνεια αποτελεί ένα σημαντικό βήμα για την επίλυση των τεχνικών προκλήσεων. Αυτά τα εργαλεία επιτρέπουν στους προγραμματιστές να κατανοούν καλύτερα τις εσωτερικές διεργασίες των αλγορίθμων, συμβάλλοντας στην προώθηση της ηθικής χρήσης της ΤΝ^[13].

Ωστόσο, μια πρόσφατη μελέτη από τους Sanderson et al. το 2023 υποδεικνύει ότι πολλοί σχεδιαστές και προγραμματιστές δυσκολεύονται να ενσωματώσουν τις κατευθυντήριες γραμμές στην ανάπτυξη των συστημάτων τους, λόγω της περιορισμένης κατανόησης των διαθέσιμων εργαλείων^[15]. Αυτό αναδεικνύει την ανάγκη για εκπαίδευση και κατάρτιση των επαγγελματιών της ΤΝ σχετικά με τη χρήση αυτών των εργαλείων.

Η δημιουργία εκπαιδευτικών προγραμμάτων, η ανταλλαγή βέλτιστων πρακτικών και η προώθηση συνεργασιών μεταξύ ακαδημαϊκών και βιομηχανικών φορέων μπορούν να συμβάλουν στη γεφύρωση του χάσματος μεταξύ θεωρίας και πράξης^[16].

4.4.2 Νομικές Προκλήσεις

Η απουσία ενοποιημένου νομικού πλαισίου σε διεθνές επίπεδο αποτελεί ένα από τα μεγαλύτερα εμπόδια για την εφαρμογή των κατευθυντήριων γραμμών στην ηθική χρήση της ΤΝ. Παρότι υπάρχουν περιφερειακές πρωτοβουλίες, όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων (GDPR) στην Ευρωπαϊκή Ένωση, που εστιάζουν στην προστασία της ιδιωτικότητας και της διαφάνειας, η έλλειψη παγκόσμιας συναίνεσης περιπλέκει τη συμμόρφωση για πολυεθνικές εταιρείες που δραστηριοποιούνται σε πολλές διαφορετικές ρυθμιστικές ζώνες^[19].

Αυτό δημιουργεί νομικές και επιχειρησιακές προκλήσεις, καθώς οι επιχειρήσεις καλούνται να προσαρμόζονται ταυτόχρονα σε πολυάριθμα, και συχνά αντικρουόμενα, ρυθμιστικά καθεστώτα^[20].

Επιπλέον, οι νομοθεσίες συχνά αδυνατούν να παρακολουθήσουν τον ταχύ ρυθμό της τεχνολογικής εξέλιξης, αφήνοντας σημαντικά κενά και ασάφειες. Για παράδειγμα, ζητήματα όπως η εξηγησιμότητα των αλγορίθμων, η λογοδοσία για τις επιπτώσεις των αποφάσεων ΤΝ και η χρήση δεδομένων σε παγκόσμιο επίπεδο παραμένουν σε μεγάλο βαθμό αδιευκρίνιστα^[21].

- **Παράδειγμα:** Ένα χαρακτηριστικό παράδειγμα είναι οι αποκλίσεις ανάμεσα σε χώρες όσον αφορά την αναγνώριση της ΤΝ ως νομικής οντότητας, κάτι που περιπλέκει περαιτέρω την ευθύνη και την τήρηση της νομοθεσίας^[22].

Η ανάγκη για τη δημιουργία ενός ενοποιημένου διεθνούς ρυθμιστικού πλαισίου είναι κρίσιμη. Προτάσεις όπως η δημιουργία ενός παγκόσμιου συμβουλίου για την ΤΝ, παρόμοιο με τα διεθνή φόρουμ για το εμπόριο και το περιβάλλον, έχουν αρχίσει να κερδίζουν έδαφος^[23]. Ένα τέτοιο συμβούλιο θα μπορούσε να λειτουργήσει ως πλατφόρμα συντονισμού, επιτρέποντας τη διαμόρφωση κοινών προτύπων που θα εγγυώνται την ασφάλεια, τη διαφάνεια και τη δίκαιη χρήση της ΤΝ^[24].

Παράλληλα, οι πολυεθνικές εταιρείες μπορούν να αναλάβουν πρωτοβουλίες συνεργασίας με νομικούς ειδικούς και ακαδημαϊκούς φορείς για την ενίσχυση της συμμόρφωσης στις υπάρχουσες νομοθεσίες. Επιπλέον, οι βιομηχανικοί φορείς μπορούν να προωθήσουν τη χρήση εργαλείων ανοιχτού κώδικα που ενσωματώνουν ηθικές κατευθυντήριες γραμμές και διευκολύνουν τη συμμόρφωση. Ο John Danaher (2022) προτείνει την εισαγωγή "αλγοριθμικών εγγυητών", δηλαδή ανεξάρτητων φορέων που θα επαληθεύουν τη συμμόρφωση των αλγορίθμων με νομικές και ηθικές απαιτήσεις. Παράλληλα, η έγκαιρη συνεργασία με νομικούς και τεχνολογικούς εμπειρογνώμονες μπορεί να ελαχιστοποιήσει τον κίνδυνο μη συμμόρφωσης και να ενισχύσει τη διαφάνεια^[25].

4.4.3 Κοινωνικές και Πολιτισμικές Προκλήσεις

Οι κοινωνικές προκαταλήψεις και οι πολιτισμικές διαφορές αποτελούν κρίσιμες προκλήσεις για την ανάπτυξη και χρήση της ΤΝ. Οι αλγόριθμοι, όταν εκπαιδεύονται σε δεδομένα που δεν αντιπροσωπεύουν ολόκληρο το φάσμα της ανθρώπινης ποικιλομορφίας, ενδέχεται να παράγουν προκατειλημμένα αποτελέσματα, επηρεάζοντας τη δικαιοσύνη και την αξιοπιστία τους^[18].

Ένα χαρακτηριστικό παράδειγμα είναι η τεχνολογία αναγνώρισης προσώπου, η οποία έχει παρουσιάσει υψηλά ποσοστά σφαλμάτων στις φυλετικές και εθνοτικές ομάδες που δεν εκπροσωπούνται επαρκώς στα δεδομένα εκπαίδευσης^[14]. Αυτή η ανισορροπία μπορεί να οδηγήσει σε άνιση μεταχείριση, ιδιαίτερα σε τομείς όπως η ασφάλεια, η απασχόληση και η υγειονομική περίθαλψη.

4.4.4 Επιχειρηματικές Προκλήσεις

Οι εταιρείες αντιμετωπίζουν σημαντικές προκλήσεις κατά την ενσωμάτωση των κατευθυντήριων γραμμών στην πρακτική τους λειτουργία. Συχνά προκύπτει μια σύγκρουση μεταξύ της επιδίωξης οικονομικών στόχων και της τήρησης ηθικών αρχών^[20].

Για παράδειγμα, η επιτάχυνση της καινοτομίας ή η μείωση του κόστους μπορεί να οδηγήσουν σε παραμέληση βασικών ηθικών ζητημάτων, όπως η διαφάνεια και η εξηγησιμότητα των αλγορίθμων. Έρευνες έχουν δείξει ότι ο χρόνος και το κόστος που απαιτούνται για την ενσωμάτωση ηθικών πρακτικών συχνά αποτρέπουν τις επιχειρήσεις από τη συμμόρφωση, ιδιαίτερα όταν υπάρχει πίεση για ταχεία ανάπτυξη και λανσάρισμα προϊόντων^[11].

Κρίσιμα Εμπόδια

- **Έλλειψη εξειδικευμένων πόρων:** Πολλές εταιρείες δε διαθέτουν τις απαραίτητες ομάδες ή εργαλεία για την εφαρμογή ηθικών πρακτικών, γεγονός που επιτείνει τις δυσκολίες ενσωμάτωσης ηθικών αρχών στις διαδικασίες τους^[21].
- **Πίεση χρόνου και ανταγωνισμού:** Η ταχεία εξέλιξη της αγοράς ΤΝ δημιουργεί έντονη πίεση για ταχεία ανάπτυξη, συχνά εις βάρος της αξιολόγησης της ηθικής διάστασης των προϊόντων^[22].
- **Δυσκολίες μέτρησης ηθικής απόδοσης:** Οι επιχειρήσεις αντιμετωπίζουν προκλήσεις στην ανάπτυξη και εφαρμογή δεικτών που μπορούν να αξιολογήσουν την ηθική απόδοση των συστημάτων ΤΝ^[23].
- **Απουσία κανονιστικών υποχρεώσεων:** Η έλλειψη αυστηρών ρυθμιστικών πλαισίων για την ΤΝ σε πολλές περιοχές σημαίνει ότι οι εταιρείες δεν έχουν πάντα εξωτερική πίεση να τηρούν ηθικές κατευθυντήριες γραμμές^[24].

4.5 Προς ένα Παγκόσμιο Ρυθμιστικό Πλαίσιο

Η ανάγκη για ένα ενοποιημένο παγκόσμιο ρυθμιστικό πλαίσιο για την ΤΝ αποτελεί πλέον ένα από τα πλέον κρίσιμα ζητήματα της εποχής μας. Παρόλο που η ανάπτυξη της ΤΝ έχει επιφέρει σημαντικά οφέλη σε διάφορους τομείς, η έλλειψη συντονισμού και εναρμόνισης στις ρυθμίσεις μεταξύ των χωρών δημιουργεί ανησυχίες για τον τρόπο με τον οποίο η τεχνολογία επηρεάζει τα θεμελιώδη δικαιώματα, την ασφάλεια και τη διαφάνεια. Η ενότητα αυτή εξετάζει τις διαφορετικές προσεγγίσεις ρύθμισης σε επίπεδο χωρών, τις πρωτοβουλίες διεθνών οργανισμών και τις προκλήσεις και προοπτικές εναρμόνισης σε παγκόσμιο επίπεδο^{[9][14]}.

4.5.1 Διαφορετικότητα Προσεγγίσεων ανάμεσα σε Χώρες και Ηπείρους

Οι διαφορετικές πολιτικές και κοινωνικοοικονομικές συνθήκες οδηγούν σε αποκλίνουσες προσεγγίσεις ρύθμισης της ΤΝ ανάμεσα σε χώρες και ηπείρους.

Ευρωπαϊκή Ένωση (ΕΕ)

Η ΕΕ έχει υιοθετήσει μια ανθρωποκεντρική προσέγγιση για την ΤΝ, με έμφαση στα θεμελιώδη δικαιώματα και τη διαφάνεια. Ο Κανονισμός για την Τεχνητή Νοημοσύνη (AI Act), ο οποίος βρίσκεται σε τελικά στάδια επεξεργασίας, αποτελεί ένα πρότυπο πλαίσιο για τη ρύθμιση της ΤΝ. Ο AI Act κατατάσσει τις εφαρμογές ΤΝ σε διαφορετικές κατηγορίες κινδύνου (απαγορευμένες, υψηλού, μεσαίου και χαμηλού κινδύνου) και προβλέπει αυστηρούς κανόνες για συστήματα υψηλού κινδύνου, όπως η αναγνώριση προσώπου σε δημόσιους χώρους^[19].

Ηνωμένες Πολιτείες (ΗΠΑ)

Η ρύθμιση της ΤΝ στις ΗΠΑ είναι πιο αποσπασματική και εστιάζει κυρίως σε συγκεκριμένους τομείς, όπως η υγεία, οι μεταφορές και η προστασία της ιδιωτικότητας. Η πολιτική

της κυβέρνησης βασίζεται κυρίως σε κατευθυντήριες γραμμές από οργανισμούς όπως το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας (NIST), αφήνοντας μεγάλο περιθώριο στις επιχειρήσεις να αυτορρυθμιστούν. Η προσέγγιση αυτή βασίζεται στην αρχή της καινοτομίας με περιορισμένη κρατική παρέμβαση^[22].

Κίνα

Η Κίνα έχει επιλέξει μια πιο κεντρική και αυταρχική προσέγγιση, με έμφαση στον έλεγχο και την επιτήρηση. Η ρύθμιση της ΤΝ στοχεύει τόσο στην ενίσχυση της εθνικής ασφάλειας όσο και στη δημιουργία ανταγωνιστικού πλεονεκτήματος στις παγκόσμιες αγορές. Πρόσφατα, εισήχθησαν κανονισμοί για τους αλγόριθμους συστάσεων, απαιτώντας διαφάνεια και τη δυνατότητα χρηστών να απενεργοποιούν συγκεκριμένες λειτουργίες. Ωστόσο, η ρύθμιση συχνά επικρίνεται για την περιορισμένη προστασία της ιδιωτικότητας των χρηστών^[24].

Αυτές οι διαφορετικές προσεγγίσεις δείχνουν ότι η εναρμόνιση των ρυθμίσεων είναι δύσκολη, αλλά ταυτόχρονα αναδεικνύουν την ανάγκη συνεργασίας για τη διαμόρφωση κοινών κανόνων.

4.5.2 Πρωτοβουλίες Διεθνών Οργανισμών

Διάφοροι διεθνείς οργανισμοί έχουν αναλάβει δράσεις για την προώθηση ενός παγκόσμιου ρυθμιστικού πλαισίου που θα διασφαλίζει τη συνεπή και ηθική χρήση της ΤΝ:

UNESCO

Η UNESCO το 2021 ενέκρινε τη Σύσταση για την Ηθική της Τεχνητής Νοημοσύνης, η οποία περιλαμβάνει αρχές για τη διαφάνεια, τη λογοδοσία, την προστασία των ανθρωπίνων δικαιωμάτων και την προώθηση της βιώσιμης ανάπτυξης. Το κείμενο της Σύστασης αποτελεί σημείο αναφοράς για την ανάπτυξη εθνικών πολιτικών και την ενίσχυση της διεθνούς συνεργασίας^[20].

Οργανισμός Οικονομικής Συνεργασίας και Ανάπτυξης (OECD)

Ο ΟΟΣΑ έχει αναπτύξει το Πλαίσιο για την Υπεύθυνη Τεχνητή Νοημοσύνη, το οποίο περιλαμβάνει πέντε βασικές αρχές, όπως η ανθρώπινη ευημερία, η διαφάνεια και η ασφάλεια. Το πλαίσιο έχει υιοθετηθεί από περισσότερες από 40 χώρες, συμπεριλαμβανομένων μελών του G20, και λειτουργεί ως οδηγός για την ανάπτυξη εθνικών ρυθμίσεων^[21].

G20 και G7

Οι κορυφαίες οικονομίες του κόσμου έχουν αρχίσει να συζητούν την ανάγκη για παγκόσμιες κατευθυντήριες γραμμές. Στο πλαίσιο αυτό, οι χώρες του G20 έχουν προτείνει τη δημιουργία ενός διεθνούς οργανισμού για την ΤΝ, ο οποίος θα λειτουργεί ως φόρουμ συνεργασίας και συντονισμού^[25].

Παγκόσμιο Οικονομικό Φόρουμ (WEF)

Το WEF έχει αναλάβει πρωτοβουλίες για την προώθηση της υπεύθυνης ΤΝ μέσω της Πρωτοβουλίας για την Τεχνητή Νοημοσύνη. Το πρόγραμμα αυτό εστιάζει στη διαμόρφωση κανονισμών και στην ανάπτυξη εργαλείων για την ενσωμάτωση των ηθικών κατευθυντήριων γραμμών^[25].

4.5.3 Προκλήσεις και Προοπτικές Εναρμόνισης

Η διαμόρφωση ενός παγκόσμιου ρυθμιστικού πλαισίου συνοδεύεται από προκλήσεις, όπως:

- **Πολιτισμικές διαφορές:** Οι πολιτισμικές αξίες και προσεγγίσεις στα ανθρώπινα δικαιώματα διαφέρουν σημαντικά μεταξύ χωρών, γεγονός που καθιστά δύσκολη την εναρμόνιση^[16].
- **Τεχνολογικές ανισότητες:** Οι χώρες με λιγότερη πρόσβαση σε τεχνολογικές υποδομές και κεφάλαια ενδέχεται να μην μπορούν να εφαρμόσουν αυστηρές ρυθμίσεις^[18].
- **Εθνική κυριαρχία:** Ορισμένες χώρες είναι απρόθυμες να δεχτούν διεθνείς κανόνες που θα μπορούσαν να περιορίσουν την ανεξαρτησία τους^[15].

Ωστόσο, η δημιουργία ενός διεθνούς ρυθμιστικού πλαισίου είναι εφικτή μέσω:

- **Διακρατικών συνεργασιών:** Η ενίσχυση της συνεργασίας μεταξύ κυβερνήσεων, εταιρειών και διεθνών οργανισμών μπορεί να συμβάλει στη γεφύρωση των διαφορών^[14].
- **Εναρμόνιση προτύπων:** Η δημιουργία κοινών προτύπων που βασίζονται σε παγκοσμίως αποδεκτές αρχές, όπως αυτές της UNESCO και του ΟΟΣΑ, μπορεί να λειτουργήσει ως κοινός παρονομαστής^[22].
- **Τεχνολογική εκπαίδευση:** Η εκπαίδευση και η ευαισθητοποίηση σχετικά με τις ηθικές πτυχές της ΤΝ μπορούν να ενισχύσουν τη στήριξη για την υιοθέτηση διεθνών κανόνων^[24].

5 Ηθική Σχεδίαση Αλγορίθμων ΤΝ

Η ηθική σχεδίαση αλγορίθμων ΤΝ αποτελεί κρίσιμο βήμα για τη διασφάλιση της υπευθυνότητας, της διαφάνειας και της αμεροληψίας των συστημάτων. Σε αυτή την ενότητα, εξετάζονται πρακτικές και εργαλεία που εφαρμόζονται σε διάφορα επίπεδα, από την οργάνωση επιχειρήσεων έως την τεχνική σχεδίαση αλγορίθμων, καθώς και τα προβλήματα που προκύπτουν κατά την εφαρμογή τους^{[8][12]}.

5.1 Πρακτικές σε διάφορα στρώματα

Η ηθική σχεδίαση απαιτεί τη συνεργασία πολλαπλών φορέων, καθώς και την υιοθέτηση κατάλληλων στρατηγικών σε όλα τα επίπεδα λειτουργίας.

Οργάνωση Επιχειρήσεων

Οι επιχειρήσεις διαδραματίζουν καθοριστικό ρόλο στη διαμόρφωση μιας κουλτούρας που προάγει την ηθική. Ορισμένες βασικές πρακτικές περιλαμβάνουν:

- **Καθιέρωση Εξειδικευμένων Ομάδων Ηθικής:** Η συγκρότηση ομάδων που θα εστιάζουν αποκλειστικά στην ηθική της ΤΝ μπορεί να βοηθήσει στην αντιμετώπιση προβλημάτων και στη διασφάλιση της συμμόρφωσης με κατευθυντήριες γραμμές. Αυτές οι ομάδες μπορούν να λειτουργούν παράλληλα με άλλες επιχειρηματικές μονάδες, διασφαλίζοντας την εναρμόνιση της ηθικής με τους οικονομικούς στόχους^[15].
- **Εταιρική Κοινωνική Ευθύνη (CSR):** Οι επιχειρήσεις που ενσωματώνουν ηθικές κατευθυντήριες γραμμές στις στρατηγικές CSR προβάλλουν την υπευθυνότητά τους, ενισχύοντας τη φήμη και τη διαφάνεια τους, καθώς και το ανταγωνιστικό τους πλεονέκτημα, σύμφωνα με το Forbes(2021). Για παράδειγμα, η Google έχει αναπτύξει εσωτερικές διαδικασίες αξιολόγησης για να διασφαλίσει τη συμμόρφωση των προϊόντων της με ηθικές αρχές^[21].
- **Υιοθέτηση Ανοιχτού Κώδικα:** Η χρήση εργαλείων και πλατφορμών ανοιχτού κώδικα που προάγουν τη διαφάνεια και την εξηγησιμότητα, όπως το AI Explainability 360, μπορεί να συμβάλει στη μείωση του κόστους και στην προώθηση της υπευθυνότητας^[16].

Οργάνωση Ομάδων Ανάπτυξης

Η ηθική σχεδίαση ξεκινά από τις ίδιες τις ομάδες που αναπτύσσουν τα συστήματα ΤΝ:

- **Κώδικας και Ανασκόπηση Κώδικα:** Πρακτικές όπως τα code reviews για ανίχνευση προκαταλήψεων μπορούν να ενσωματώσουν τις ηθικές κατευθυντήριες γραμμές σε πρώιμο στάδιο ανάπτυξης^[18].
- **Δημιουργία Εργαλείων Αξιολόγησης:** Εργαλεία όπως το FairLearn και το AI Fairness 360 μπορούν να εντοπίσουν και να μετριάσουν τις μεροληψίες στα μοντέλα^[19].
- **Εκπαίδευση Ομάδων:** Η εκπαίδευση των προγραμματιστών και σχεδιαστών σχετικά με την εξηγησιμότητα, την αμεροληψία και τη διαφάνεια είναι απαραίτητη για την ανάπτυξη ηθικών συστημάτων^[23].

Τεχνικές Πρακτικές και Εργαλεία

Σε τεχνικό επίπεδο, η εφαρμογή ηθικών αρχών μπορεί να επιτευχθεί μέσω εξειδικευμένων εργαλείων και τεχνικών. Ορισμένα από τα πιο διαδεδομένα είναι:

- **LIME (Local Interpretable Model-Agnostic Explanations):** Δημιουργεί τοπικά μοντέλα εξηγησιμότητας για συγκεκριμένες προβλέψεις, παρέχοντας κατανόηση των αποφάσεων ενός μοντέλου μέσω της ανάλυσης των χαρακτηριστικών που επηρεάζουν κάθε πρόβλεψη. Χρησιμοποιείται ευρέως σε μοντέλα μηχανικής μάθησης για τη δημιουργία ερμηνειών χωρίς να επηρεάζεται το μοντέλο που χρησιμοποιείται^[10].
 - **Πλεονεκτήματα:** Ευκολία χρήσης και δυνατότητα εφαρμογής σε διαφορετικά είδη μοντέλων.
 - **Περιορισμοί:** Η τοπική φύση της εξήγησης μπορεί να μην αντικατοπτρίζει τις συνολικές ιδιότητες του μοντέλου.
- **SHAP (SHapley Additive exPlanations):** Βασισμένο στη θεωρία παιγνίων, το SHAP υπολογίζει τη συμβολή κάθε χαρακτηριστικού σε μια πρόβλεψη, προσφέροντας μια σαφή και ποσοτικοποιημένη εικόνα της σημασίας των δεδομένων εισόδου^[12].
 - **Πλεονεκτήματα:** Ιδανικό για ανάλυση σημαντικότητας χαρακτηριστικών σε πολύπλοκα μοντέλα (π.χ., δέντρα απόφασης και νευρωνικά δίκτυα).
 - **Περιορισμοί:** Υπολογιστικά απαιτητικό εργαλείο για μεγάλα σύνολα δεδομένων ή πολύπλοκα μοντέλα.
- **Integrated Gradients:** Είναι ένα ειδικό εργαλείο για νευρωνικά δίκτυα που χρησιμοποιεί υπολογισμό ολοκληρωμένων παραγώγων για να αξιολογήσει τη συμβολή κάθε χαρακτηριστικού στην πρόβλεψη^[15].
 - **Πλεονεκτήματα:** Είναι ιδανικό για βαθιά νευρωνικά δίκτυα που εφαρμόζονται σε δεδομένα εικόνας ή κειμένου. Είναι απλό στη χρήση και αποτελεσματικό για μεγάλα μοντέλα.
 - **Περιορισμοί:** Περιορίζεται η χρήση του σε νευρωνικά δίκτυα.
- **What-If Tool (Google AI):** Παρέχει έναν διαδραστικό πίνακα εργαλείων για την κατανόηση του τρόπου με τον οποίο διαφορετικές παράμετροι επηρεάζουν τις προβλέψεις^[17]. Είναι πολύ χρήσιμο σε εταιρικές εφαρμογές όπου απαιτείται εξηγησιμότητα για λήψη αποφάσεων. Αποτελεί ένα διαισθητικό και χρήσιμο εργαλείο για επαγγελματίες που δεν είναι εξοικειωμένοι με τον προγραμματισμό.

- **Πλεονεκτήματα:** Εύκολο στη χρήση για επαγγελματίες χωρίς τεχνικές γνώσεις.
- **Περιορισμοί:** Δεν προσφέρει λεπτομερή ανάλυση σε περίπλοκα μοντέλα.
- **AI Explainability 360 (IBM):** Ένα ολοκληρωμένο εργαλείο που προσφέρει πολλές μεθόδους εξηγησιμότητας και διευκολύνει τη σύγκριση διαφορετικών τεχνικών^[19].
 - **Πλεονεκτήματα:** Κατάλληλο για όλα τα είδη μοντέλων και ενσωματώνει πολλαπλές στρατηγικές εξηγησιμότητας.

5.2 Προβλήματα στην εφαρμογή των Πρακτικών

Αυτή η ενότητα είναι πολύ χρήσιμη, καθώς δίνει τη δυνατότητα να αναγνωρίσεις τα εμπόδια (τεχνικά, οργανωσιακά, ακόμα και ψυχολογικά) και να προτείνεις λύσεις. Πρόσθεσε μερικές σύντομες μελέτες περιπτώσεων (π.χ., προβλήματα που συνάντησαν εταιρείες όπως το Facebook όταν προσπάθησαν να διορθώσουν τη μεροληψία στα συστήματά τους).

Παρά την ύπαρξη εργαλείων και πρακτικών, η ενσωμάτωσή τους στις διαδικασίες ανάπτυξης ΤΝ συναντά σημαντικά εμπόδια:

Τεχνικά Εμπόδια

- **Πολύπλοκότητα Ενσωμάτωσης:** Η ενσωμάτωση εργαλείων όπως το SHAP ή το LIME σε μεγάλα γλωσσικά μοντέλα είναι υπολογιστικά απαιτητική, ιδιαίτερα για πολύπλοκα σύνολα δεδομένων^[14].
- **Έλλειψη Επαρκών Εργαλείων:** Παρότι υπάρχουν εξελιγμένα εργαλεία, η εφαρμογή τους σε πραγματικό χρόνο ή σε διαφορετικές εφαρμογές παραμένει περιορισμένη^[13].

Οργανωτικά Εμπόδια

- **Αντίσταση στην Αλλαγή:** Πολλές εταιρείες διστάζουν να επενδύσουν πόρους σε νέες ηθικές πρακτικές λόγω κόστους ή χρονικών περιορισμών^[11].
- **Ελλιπής Κατανόηση:** Οι ομάδες ανάπτυξης συχνά δεν κατανοούν πλήρως πώς να χρησιμοποιήσουν τα διαθέσιμα εργαλεία^[16].

Προτάσεις για Αντιμετώπιση

- **Διασύνδεση Ακαδημαϊκών και Επαγγελματικών Φορέων:** Η συνεργασία μεταξύ πανεπιστημίων και εταιρειών μπορεί να οδηγήσει στη δημιουργία νέων εργαλείων και τεχνολογιών^[20].
- **Εκπαίδευση:** Επενδύσεις σε εκπαιδευτικά προγράμματα και πιστοποιήσεις σχετικά με την ηθική σχεδίαση αλγορίθμων, καθώς και για τα διαθέσιμα εργαλεία, μπορεί να ενισχύσει την ικανότητα των εταιρειών να συμμορφώνονται με τις κατευθυντήριες γραμμές^[18].
- **Ανατροφοδότηση:** Συνεχής αξιολόγηση των μοντέλων με χρήση εργαλείων όπως το FairLearn^[21].

Επέκταση Προτάσεων για την Αντιμετώπιση της Μεροληψίας και της Έλλειψης Διαφάνειας

Για να επιλυθούν αυτά τα ζητήματα, απαιτείται μια συστηματική και πολύπλευρη προσέγγιση. Ορισμένες βασικές προτάσεις περιλαμβάνουν:

- **Διασφάλιση ποικιλομορφίας στα δεδομένα εκπαίδευσης:** Η συλλογή και επεξεργασία δεδομένων που περιλαμβάνουν διαφορετικές δημογραφικές ομάδες, πολιτισμικά πλαίσια και κοινωνικοοικονομικά στρώματα είναι ζωτικής σημασίας^[9].
 - **Παράδειγμα:** Έρευνα του Baeza-Yates (2018) έδειξε ότι τα συστήματα αναζήτησης στο διαδίκτυο επηρεάζονται από τη γλώσσα και την προέλευση των δεδομένων, με αποτέλεσμα μεροληπτικά αποτελέσματα που ενισχύουν στερεότυπα.
- **Συνεχής αξιολόγηση για μεροληψίες:** Η υιοθέτηση εργαλείων αξιολόγησης όπως το FairLearn και το AI Fairness 360 μπορεί να βοηθήσει στον εντοπισμό και τη διόρθωση των προκαταλήψεων στα συστήματα ΤΝ^[13]. Αυτά τα εργαλεία παρέχουν μεθόδους ανάλυσης που βοηθούν στην αξιολόγηση της δικαιοσύνης και στη διασφάλιση ίσης μεταχείρισης για όλες τις ομάδες.
 - **Παράδειγμα:** Το 2018, η Amazon ανέπτυξε ένα αλγοριθμικό σύστημα προσλήψεων, το οποίο απορρίφθηκε λόγω μεροληψίας κατά των γυναικών. Ο αλγόριθμος είχε εκπαιδευτεί με βάση ιστορικά δεδομένα ανδρών, γεγονός που τον έκανε να υποτιμά τις υποψηφιότητες γυναικών.
- **Συμμετοχή τοπικών κοινοτήτων και ενδιαφερόμενων μερών:** Η συμμετοχή τοπικών φορέων, πολιτισμικών ομάδων και ομάδων συμφερόντων στον σχεδιασμό και την ανάπτυξη συστημάτων ΤΝ διασφαλίζει ότι λαμβάνονται υπόψη οι τοπικές αξίες και ανάγκες^[15]. Η ενσωμάτωση αυτών των ομάδων στις διαδικασίες λήψης αποφάσεων μπορεί να μειώσει τις πολιτισμικές ανισότητες και να προάγει την κοινωνική αποδοχή.
 - **Παράδειγμα:** Σε πολλές χώρες της Αφρικής, οι εφαρμογές αναγνώρισης προσώπου έχουν αποτύχει λόγω εκπαίδευσης των αλγορίθμων σε δυτικά δεδομένα, με αποτέλεσμα υψηλά ποσοστά σφαλμάτων στην ταυτοποίηση ατόμων με σκούρο δέρμα.
- **Προσαρμογή των αλγορίθμων στις πολιτισμικές ιδιαιτερότητες:** Οι αλγόριθμοι πρέπει να προσαρμόζονται για να λαμβάνουν υπόψη τις πολιτισμικές διαφορές, όπως η γλώσσα, τα κοινωνικά πρότυπα και οι αξίες^[17].
 - **Παράδειγμα:** Οι τεχνολογίες επεξεργασίας φυσικής γλώσσας (NLP) που αναπτύσσονται σε αγγλικά δεδομένα έχουν χαμηλή απόδοση όταν εφαρμόζονται σε λιγότερο διαδεδομένες γλώσσες.
- **Δημόσια ευαισθητοποίηση και εκπαίδευση:** Η ευαισθητοποίηση του κοινού και η εκπαίδευση σχετικά με τις δυνατότητες και τους κινδύνους της ΤΝ μπορεί να ενισχύσει την κοινωνική εμπιστοσύνη^[23]. Επιπλέον, εκπαιδευτικά προγράμματα που απευθύνονται σε μη ειδικούς χρήστες μπορούν να τους βοηθήσουν να κατανοήσουν καλύτερα τη λειτουργία και τις επιπτώσεις των αλγορίθμων.
 - **Παράδειγμα:** Το Ευρωπαϊκό Κοινοβούλιο έχει προτείνει την ένταξη μαθημάτων ΤΝ και ηθικής αλγορίθμων στα σχολεία, ώστε οι πολίτες να κατανοούν τις επιπτώσεις των αλγοριθμικών αποφάσεων.

5.3 Υπόδειγμα Εφαρμογής

Στην ενότητα αυτή, θα παρουσιαστεί ένα υπόδειγμα εφαρμογής που περιγράφει πώς μια υποθετική εταιρεία μπορεί να εφαρμόσει τις ηθικές κατευθυντήριες γραμμές από τον σχεδιασμό έως την υλοποίηση ενός αυτόνομου συστήματος. Η ανάλυση βασίζεται σε υπάρχουσες ηθικές πρακτικές και πραγματικές περιπτώσεις επιτυχημένης εφαρμογής.

5.3.1 Παρουσίαση της Υποθετικής Εταιρείας "EthicAI Systems"

Η EthicAI Systems είναι μια υποθετική εταιρεία που αναπτύσσει αυτόνομα συστήματα υγειονομικής περίθαλψης, εστιάζοντας στη διαφάνεια, την αξιοπιστία και την ηθική συμμόρφωση. Το κύριο προϊόν της είναι ένας αυτόνομος βοηθός υγείας (AI Health Assistant) που βοηθά στη διάγνωση ασθενειών και στη λήψη ιατρικών αποφάσεων, σε συνεργασία με γιατρούς.

Η εταιρεία εφαρμόζει τις ηθικές κατευθυντήριες γραμμές από το αρχικό στάδιο του σχεδιασμού έως την εμπορική διάθεση του προϊόντος^[12].

5.3.2 Στάδια Εφαρμογής των Ηθικών Κατευθυντήριων Γραμμών

Στάδιο 1: Δεοντολογικός Σχεδιασμός και Προδιαγραφές

Πριν την ανάπτυξη του συστήματος, η EthicAI Systems συγκροτεί μια Διεπιστημονική Επιτροπή Δεοντολογίας, αποτελούμενη από:

- Ειδικούς στην τεχνητή νοημοσύνη,
- Ιατρούς και βιοηθικολόγους,
- Νομικούς συμβούλους για την προστασία δεδομένων,
- Εκπροσώπους χρηστών (ασθενείς και γιατρούς).

Η επιτροπή αυτή χαράσσει τις βασικές ηθικές αρχές, που περιλαμβάνουν:

- Δικαιοσύνη & αμεροληψία στον αλγόριθμο διάγνωσης^[11],
- Διαφάνεια & εξηγησιμότητα των αποφάσεων του AI^[4],
- Προστασία προσωπικών δεδομένων (σύμφωνα με τον GDPR και διεθνείς κανονισμούς)^[9].

Στάδιο 2: Ανάπτυξη και Εκπαίδευση του Συστήματος

Κατά τη διάρκεια της ανάπτυξης, η εταιρεία ακολουθεί την προσέγγιση "Ηθική από τον Σχεδιασμό" (Ethics by Design):

- Χρησιμοποιεί ελεγχόμενα και πιστοποιημένα ιατρικά δεδομένα για την εκπαίδευση του AI^[10].
- Ενσωματώνει εξηγήσιμους αλγόριθμους, που επιτρέπουν στους γιατρούς να κατανοούν πώς και γιατί προτείνονται συγκεκριμένες διαγνώσεις^[6].
- Διεξάγει δοκιμές με διαφορετικές πληθυσμιακές ομάδες για να ελαχιστοποιήσει μεροληψίες και να διασφαλίσει την ισότιμη πρόσβαση στη διάγνωση^[12].

Στάδιο 3: Δοκιμές και Βελτιστοποίηση μέσω Case Studies

Για να διασφαλίσει την πρακτική εφαρμογή των ηθικών προτύπων, η EthicAI Systems συνεργάζεται με τρία μεγάλα νοσοκομεία για πιλοτικές δοκιμές.

- **Case Study 1: Πανεπιστημιακό Νοσοκομείο Αθηνών**

- Δοκιμές του συστήματος ΑΙ σε επαγγελματίες υγείας, ώστε να μετρηθεί η ακρίβεια και η διαφάνεια των αποφάσεων.
- Ανατροφοδότηση: Οι γιατροί ζήτησαν πιο εξηγήσιμα διαγνωστικά αποτελέσματα.

- **Case Study 2: Ιδιωτική Κλινική στη Γερμανία**

- Έλεγχος συμμόρφωσης με το GDPR για την ασφαλή αποθήκευση προσωπικών δεδομένων ασθενών^[19].
- Ανατροφοδότηση: Η προσθήκη δυνατότητας ελέγχου των δεδομένων από τους ίδιους τους ασθενείς αύξησε την εμπιστοσύνη στη χρήση του ΑΙ.

- **Case Study 3: Δημόσιο Νοσοκομείο στις ΗΠΑ**

- Ανάλυση των αποτελεσμάτων σε ασθενείς διαφορετικών εθνοτικών και κοινωνικοοικονομικών ομάδων.
- Ανατροφοδότηση: Εντοπίστηκαν μεροληψίες, π.χ., μεγαλύτερη πιθανότητα λανθασμένων διαγνώσεων σε μειονότητες. Έγιναν διορθώσεις στον αλγόριθμο.
- Βελτίωση: Έγιναν διορθώσεις στον αλγόριθμο με βάση εργαλεία όπως FairLearn & AI Fairness 360 ^[15].

Στάδιο 4: Τελική Πιστοποίηση και Διάθεση στην Αγορά

Η EthicAI Systems λαμβάνει πιστοποιήσεις από διεθνείς οργανισμούς, όπως:

- **EU AI Act** (Κανονισμός της Ε.Ε. για την ΤΝ)^[25]
- **FDA Approval** (για χρήση σε ιατρικά περιβάλλοντα στις ΗΠΑ)^[24]
- **ISO 13485** (πιστοποίηση ιατρικών συσκευών)

Στάδιο 5: Μετά-εφαρμοστική Επιτήρηση και Συνεχής Βελτίωση

Μετά την κυκλοφορία του συστήματος, η EthicAI Systems:

- Παρακολουθεί συνεχώς τις αποφάσεις του ΑΙ για να διασφαλίσει ότι παραμένουν ηθικές και αξιόπιστες.
- Παρέχει τακτικές ενημερώσεις και εκπαίδευση στους χρήστες (ιατρούς & ασθενείς).
- Δημιουργεί μηχανισμό αναφορών προβλημάτων, επιτρέποντας στους ασθενείς να καταγγέλλουν λανθασμένες ή μη δίκαιες αποφάσεις του ΑΙ^[23].

5.3.3 Ανάλυση Επιτυχημένων Case Studies στην Πραγματική Ζωή

Παρόμοιες πρακτικές έχουν εφαρμοστεί από πραγματικές εταιρείες και οργανισμούς:

IBM Watson for Oncology

Η IBM ανέπτυξε ένα σύστημα τεχνητής νοημοσύνης για τη διάγνωση καρκίνου. Ωστόσο, υπήρξαν ζητήματα διαφάνειας και υπερβολικής εξάρτησης από μη αντιπροσωπευτικά δεδομένα, κάτι που προκάλεσε ανακρίβειες σε ορισμένες διαγνώσεις^{[26][27]}.

DeepMind & NHS (Ηνωμένο Βασίλειο)

Η DeepMind συνεργάστηκε με το Εθνικό Σύστημα Υγείας (NHS) για την ανάλυση δεδομένων ασθενών, αλλά αντιμετώπισε νομικά ζητήματα σχετικά με τη μη συγκατάθεση των ασθενών στη χρήση των προσωπικών τους δεδομένων^{[28][29]}.

Ada Health (Γερμανία)

Επιτυχημένο παράδειγμα που ενσωμάτωσε ηθικές αρχές στον σχεδιασμό του, επιτρέποντας στους χρήστες πλήρη έλεγχο των δεδομένων τους.

6 Παραδείγματα και Επιπτώσεις

Η Τεχνητή Νοημοσύνη (TN) έχει ήδη διαμορφώσει σημαντικούς τομείς της κοινωνίας, από την ιατρική και την εκπαίδευση έως την οικονομία και την πολιτική. Το παρόν κεφάλαιο αναλύει συγκεκριμένα παραδείγματα, τόσο θετικά όσο και αρνητικά, και εξετάζει τις επιπτώσεις τους στη σύγχρονη κοινωνία.

6.1 Καλά παραδείγματα

Παρά την ανησυχία γύρω από την TN, υπάρχουν παραδείγματα που δείχνουν πώς μπορεί να χρησιμοποιηθεί για το κοινό καλό. Εξετάζουμε τρεις βασικές περιοχές εφαρμογής.

A. Τεχνητή Νοημοσύνη στην Ιατρική

Η TN έχει φέρει επαναστατικές αλλαγές στην ιατρική, με σημαντικά οφέλη στη διάγνωση ασθενειών και στη βελτίωση της φροντίδας των ασθενών. Παραδείγματα περιλαμβάνουν:

- **Ανίχνευση καρκίνου μέσω μηχανικής μάθησης:** Συστήματα όπως το Google DeepMind έχουν αναπτύξει μοντέλα ανίχνευσης καρκίνου του μαστού που είναι πιο ακριβή από τους γιατρούς σε ορισμένες περιπτώσεις^[28].
- **ΑΙ στην πρόβλεψη επιδημιών:** Το σύστημα BlueDot κατάφερε να προβλέψει την εξάπλωση του COVID-19 πριν από τον Παγκόσμιο Οργανισμό Υγείας^[12].

B. Τεχνητή Νοημοσύνη για τη διαφάνεια και την ανάλυση δεδομένων

Η χρήση της TN για την ανάλυση δεδομένων και την προώθηση της διαφάνειας είναι ένα άλλο θετικό παράδειγμα.

- Ο **ΟΗΕ (UNDP)** έχει αναπτύξει εργαλεία ΑΙ που επιτρέπουν την ανάλυση οικονομικών δεδομένων για την ανίχνευση διαφθοράς και την ενίσχυση της λογοδοσίας σε κυβερνήσεις^[16].
- Το σύστημα "**AI for Good**" του ITU βοηθά στη συλλογή και ανάλυση δεδομένων για την αντιμετώπιση της κλιματικής αλλαγής^[14].

Γ. Τεχνητή Νοημοσύνη στην εκπαίδευση και τη δικαιοσύνη

Η TN χρησιμοποιείται ήδη για να βελτιώσει τη μάθηση και τη δικαστική αμεροληψία.

- **Προσωποποιημένη μάθηση:** Πλατφόρμες όπως το Khan Academy AI tutor προσαρμόζουν τη διδασκαλία στις ανάγκες κάθε μαθητή^[9].

- **ΑΙ στη δικαιοσύνη:** Στις ΗΠΑ, συστήματα όπως το COMPAS επιχειρούν να αξιολογήσουν την επικινδυνότητα κρατουμένων, αν και εγείρονται ηθικά ερωτήματα^[11].

6.2 Προβληματικά παραδείγματα

Παρά τις θετικές εφαρμογές, υπάρχουν επίσης περιπτώσεις όπου η ΤΝ έχει χρησιμοποιηθεί καταχρηστικά ή έχει οδηγήσει σε αρνητικές συνέπειες.

A. Το σκάνδαλο Cambridge Analytica

Το 2018, η Cambridge Analytica χρησιμοποίησε την ΤΝ για να αναλύσει δεδομένα από το Facebook και να δημιουργήσει ψυχολογικά προφίλ χρηστών, τα οποία χρησιμοποιήθηκαν για στοχευμένη πολιτική διαφήμιση στις προεδρικές εκλογές των ΗΠΑ (2016) και στο δημοψήφισμα για το Brexit.

- Αυτή η υπόθεση έθεσε σοβαρά ερωτήματα περί ιδιωτικότητας και δεοντολογίας στην ανάλυση μεγάλων δεδομένων^{[19][9]}.

B. Αναγνώριση προσώπου και ρατσιστική προκατάληψη

Τα συστήματα αναγνώρισης προσώπου έχουν αποδειχθεί ανακριβή, ειδικά για άτομα από εθνοτικές μειονότητες, γεγονός που έχει οδηγήσει σε λανθασμένες συλλήψεις.

- **Χαρακτηριστικά παραδείγματα:**
 - Τα συστήματα της Amazon Rekognition απέτυχαν να αναγνωρίσουν σωστά άτομα με σκούρο δέρμα, ενώ ταυτοποίησαν λανθασμένα μέλη του Κογκρέσου ως εγκληματίες^{[2][17]}.
 - Στην Κίνα, η τεχνολογία αναγνώρισης προσώπου χρησιμοποιείται για τον έλεγχο πληθυσμών, όπως οι Ουιγούροι μουσουλμάνοι, μέσω κρατικών προγραμμάτων μαζικής επιτήρησης^[18].

Γ. Προβλήματα με τα Μεγάλα Γλωσσικά Μοντέλα (LLMs)

Τα LLMs (π.χ. ChatGPT, Bard) μπορεί να παράγουν παραπληροφόρηση και προκαταλήψεις.

- **Παραδείγματα**
 - Τη δημιουργία ψευδών ειδήσεων μέσω αυτοματοποιημένων bots, που χειραγωγούν τη δημόσια γνώμη^[5].
 - Την προώθηση στερεοτύπων που υπάρχουν στα δεδομένα εκπαίδευσής τους, καθώς βασίζονται σε μεγάλες βάσεις δεδομένων που περιέχουν ιστορικά μεροληπτικές πληροφορίες^[14].

6.3 Οι επιπτώσεις

Οι επιπτώσεις της ΤΝ είναι πολυδιάστατες, επηρεάζοντας την απασχόληση, την ιδιωτικότητα και τις κοινωνικές ανισότητες.

Α. Επιπτώσεις στην απασχόληση

Η ΤΝ αυτοματοποιεί πολλές εργασίες, οδηγώντας σε:

- Απώλεια θέσεων εργασίας σε τομείς όπως η εξυπηρέτηση πελατών, η βιομηχανική παραγωγή και η λογιστική^[8].
- Δημιουργία νέων επαγγελμάτων, όπως η ανάπτυξη αλγορίθμων ΑΙ και η ηθική εποπτεία της ΤΝ^[16].

Β. Απειλή για την ιδιωτικότητα

Η ΤΝ επιτρέπει τη μαζική επιτήρηση και τη συλλογή δεδομένων χωρίς συγκατάθεση^[19].

- **Παραδείγματα:**

- Τη χρήση "έξυπνων" καμερών για την αναγνώριση προσώπων.
- Τη δημιουργία προφίλ καταναλωτών από μεγάλες εταιρείες για διαφημιστικούς σκοπούς.

Γ. Αύξηση κοινωνικής ανισότητας

Τα μεροληπτικά δεδομένα που χρησιμοποιούνται στην εκπαίδευση της ΤΝ μπορεί να ενισχύσουν τις κοινωνικές ανισότητες^[11].

- **Παραδείγματα:**

- Οι αλγόριθμοι προσλήψεων (Amazon AI Hiring) είχαν αρνητικές προκαταλήψεις απέναντι στις γυναίκες, καθώς είχαν εκπαιδευτεί σε ανδρικά βιογραφικά.
- Οι ασφαλιστικές εταιρείες χρησιμοποιούν ΤΝ για να απορρίπτουν αιτήσεις πελατών με βάση προκατειλημμένα δεδομένα.

7 *Συμπεράσματα και Προτάσεις*

Η ραγδαία εξέλιξη της Τεχνητής Νοημοσύνης (TN) και η ευρεία της εφαρμογή σε ποικίλους τομείς της κοινωνίας έχει αναδείξει πληθώρα ηθικών και νομικών ζητημάτων. Η παρούσα διπλωματική εργασία ανέλυσε διεξοδικά τις προκλήσεις που προκύπτουν από την ανάπτυξη και εφαρμογή της TN, εστιάζοντας στην ανάγκη για ηθική σχεδίαση και ρύθμιση των συστημάτων αυτών.

Τα κυριότερα συμπεράσματα που προκύπτουν από την έρευνα συνοψίζονται ως εξής:

7.1 Βασικά Συμπεράσματα

Η ανάγκη για διαφάνεια και εξηγήσιμη TN

Τα συστήματα TN επηρεάζουν κρίσιμες αποφάσεις στην υγεία, τη δικαιοσύνη, τις χρηματοοικονομικές υπηρεσίες και πολλούς άλλους τομείς^[12]. Ωστόσο, η έλλειψη διαφάνειας και η δυσκολία εξήγησης των αποφάσεών τους δημιουργεί σημαντικά ζητήματα εμπιστοσύνης. Είναι απαραίτητο να αναπτυχθούν τεχνικές εξηγησιμότητας (XAI) που επιτρέπουν την κατανόηση της λογικής πίσω από τις αποφάσεις των αλγορίθμων^[10].

Η αντιμετώπιση των αλγοριθμικών προκαταλήψεων

Οι αλγόριθμοι TN μπορούν να ενισχύσουν ή να αναπαράγουν κοινωνικές προκαταλήψεις, εάν δε σχεδιαστούν σωστά. Αυτό οφείλεται στο γεγονός ότι εκπαιδεύονται σε υπάρχοντα δεδομένα, τα οποία συχνά αντικατοπτρίζουν τις ανισότητες που υπάρχουν στον πραγματικό κόσμο^{[2][17]}. Είναι απαραίτητη η διαρκής αξιολόγηση και προσαρμογή των μοντέλων ώστε να διασφαλίζεται η δίκαιη και αμερόληπτη λειτουργία τους.

Η διασφάλιση της προστασίας των δεδομένων και της ιδιωτικότητας

Οι εφαρμογές TN βασίζονται στη συλλογή και ανάλυση τεράστιων ποσοτήτων δεδομένων. Η χρήση αυτών των δεδομένων χωρίς την κατάλληλη προστασία μπορεί να οδηγήσει σε καταπάτηση της ιδιωτικότητας και των δικαιωμάτων των πολιτών^[19]. Η συμμόρφωση με κανονισμούς όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) είναι κρίσιμη για την ασφαλή χρήση της TN.

Η ισορροπία μεταξύ ρύθμισης και καινοτομίας

Οι νομικές και δεοντολογικές κατευθυντήριες γραμμές είναι απαραίτητες για την προστασία των ατόμων από τυχόν καταχρήσεις της ΤΝ. Ωστόσο, οι υπερβολικά αυστηροί κανονισμοί ενδέχεται να περιορίσουν την καινοτομία και την ανταγωνιστικότητα^[5]. Είναι σημαντικό να επιτευχθεί μια ισορροπία μεταξύ της νομοθετικής ρύθμισης και της ενίσχυσης της έρευνας και ανάπτυξης.

Η κοινωνική και οικονομική επίδραση της αυτοματοποίησης

Η αυτοματοποίηση μέσω ΤΝ επηρεάζει την αγορά εργασίας, οδηγώντας σε απώλεια θέσεων εργασίας σε συγκεκριμένους κλάδους, αλλά και στη δημιουργία νέων ευκαιριών σε άλλους^[8]. Οι κυβερνήσεις και οι επιχειρήσεις πρέπει να υιοθετήσουν στρατηγικές επανεκπαίδευσης και ενσωμάτωσης των εργαζομένων στις νέες τεχνολογικές συνθήκες^[16].

7.2 Προτάσεις για το Μέλλον

Με βάση τα παραπάνω συμπεράσματα, προτείνονται τα ακόλουθα μέτρα και πολιτικές για τη διασφάλιση μιας υπεύθυνης και ηθικής χρήσης της ΤΝ:

Ενίσχυση των μηχανισμών εξηγησιμότητας και διαφάνειας

- Ανάπτυξη τεχνικών για την εξήγηση των αποφάσεων των αλγορίθμων (Explainable AI - ΧΑΙ)^[11].
- Εφαρμογή συστημάτων εποπτείας που θα επιτρέπουν την αξιολόγηση της λειτουργίας των αλγορίθμων από ειδικούς.
- Διασφάλιση της δυνατότητας ανθρώπινης παρέμβασης σε αυτοματοποιημένες διαδικασίες.

Ανάπτυξη αλγοριθμικής δικαιοσύνης και μείωση των προκαταλήψεων

- Καθιέρωση αυστηρών κριτηρίων για τη συλλογή και την προ επεξεργασία των δεδομένων εκπαίδευσης^[9].
- Συστηματικός έλεγχος των μοντέλων για πιθανές προκαταλήψεις και αδικίες^[2].
- Δημιουργία εργαλείων που θα επιτρέπουν τη συνεχή παρακολούθηση και αξιολόγηση των αποφάσεων της ΤΝ.

Ενίσχυση της προστασίας δεδομένων και της ιδιωτικότητας

- Βελτίωση των τεχνικών ανωνυμοποίησης και ασφαλούς αποθήκευσης δεδομένων^[19].
- Θέσπιση αυστηρότερων κανονισμών για τη χρήση προσωπικών δεδομένων από συστήματα ΤΝ^[18].
- Ενίσχυση της διαφάνειας στις πολιτικές συλλογής και διαχείρισης δεδομένων.

Ανάπτυξη ηθικών και νομικών πλαισίων για τη ρύθμιση της TN

- Υιοθέτηση διεθνών κατευθυντήριων γραμμών για την ηθική ανάπτυξη TN, όπως αυτές που προτείνονται από την Ευρωπαϊκή Επιτροπή και τον ΟΗΕ^[23].
- Δημιουργία μηχανισμών συμμόρφωσης που θα διασφαλίζουν ότι οι εταιρείες που χρησιμοποιούν TN ακολουθούν ηθικά πρότυπα^[25].
- Διαμόρφωση νομοθεσίας που θα ορίζει την ευθύνη για αποφάσεις που λαμβάνονται από συστήματα TN^[9].

Πρόωθηση της κοινωνικής και οικονομικής βιωσιμότητας της TN

- Επενδύσεις στην εκπαίδευση και την επανεκπαίδευση των εργαζομένων που επηρεάζονται από την αυτοματοποίηση^[8].
- Ανάπτυξη νέων επαγγελματικών ρόλων που θα επικεντρώνονται στη διαχείριση και την εποπτεία των συστημάτων TN^[16].
- Θέσπιση πολιτικών που διασφαλίζουν ότι η TN χρησιμοποιείται για την ενίσχυση της κοινωνικής δικαιοσύνης και της συμπερίληψης^[11].

7.3 Τελική Ανασκόπηση

Η Τεχνητή Νοημοσύνη αποτελεί μια από τις πιο ισχυρές τεχνολογίες του 21ου αιώνα, με τεράστιες δυνατότητες αλλά και σημαντικές προκλήσεις. Η παρούσα εργασία ανέδειξε την ανάγκη για έναν υπεύθυνο και ηθικό σχεδιασμό των συστημάτων TN, προκειμένου να διασφαλιστεί ότι η ανάπτυξή τους θα συμβάλει στη βελτίωση της ανθρώπινης ζωής χωρίς να δημιουργήσει νέες ανισότητες και κινδύνους.

Η ισορροπία μεταξύ τεχνολογικής προόδου και ηθικής ευθύνης απαιτεί τη συνεργασία κυβερνήσεων, ερευνητών, επιχειρήσεων και της κοινωνίας των πολιτών. Με τη σωστή προσέγγιση, η TN μπορεί να μετατραπεί σε ένα εργαλείο που όχι μόνο θα ενισχύσει την καινοτομία και την παραγωγικότητα αλλά και θα εξασφαλίσει μια πιο δίκαιη και βιώσιμη κοινωνία.

Η συνέχιση της έρευνας και της διαμόρφωσης πολιτικών για την ηθική ανάπτυξη της TN είναι επιτακτική, προκειμένου να αξιοποιηθούν στο έπακρο τα οφέλη της τεχνολογίας αυτής και να διασφαλιστεί ότι θα υπηρετεί το κοινωνικό καλό.

8 Βιβλιογραφία

- [1] Almánzar, Arlette D Román, Laura M Edinger-Schons, and David J Grüning. «**The Impact of Social Responsibility and Social Dominance Attitudes on AI Programmers**» n.d.
- [2] Baeza-Yates, Ricardo. «**Bias on the Web.**» *Communications of the ACM* 61, no. 6 (May 23, 2018): 54–61. <https://doi.org/10.1145/3209581>.
- [3] Balasubramaniam, Nagadivya, Marjo Kauppinen, Sari Kujala, and Kari Hiekkanen. «**Ethical Guidelines for Solving Ethical Issues and Developing AI Systems.**» In *Product-Focused Software Process Improvement*, edited by Maurizio Morisio, Marco Torchiano, and Andreas Jedlitschka, 12562:331–46. *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-64148-1_21.
- [4] Brey, Philip, and Brandt Dainow. «**Ethics by Design for Artificial Intelligence.**» *AI and Ethics*, September 21, 2023. <https://doi.org/10.1007/s43681-023-00330-4>.
- [5] Danaher, John. «**Freedom in an Age of Algocracy.**» In *The Oxford Handbook of Philosophy of Technology*, by John Danaher, 249–72. edited by Shannon Vallor. Oxford University Press, 2022. <https://doi.org/10.1093/oxfordhb/9780190851187.013.16>.
- [6] Du, Fan, Catherine Plaisant, Neil Spring, Kenyon Crowley, and Ben Shneiderman. «**EventAction: A Visual Analytics Approach to Explainable Recommendation for Event Sequences.**» *ACM Transactions on Interactive Intelligent Systems* 9, no. 4 (December 31, 2019): 1–31. <https://doi.org/10.1145/3301402>.
- [7] Figalist, Iris, Christoph Elsner, Jan Bosch, and Helena Holmström Olsson. «**An End-to-End Framework for Productive Use of Machine Learning in Software Analytics and Business Intelligence Solutions.**» In *Product-Focused Software Process Improvement*, edited by Maurizio Morisio, Marco Torchiano, and Andreas Jedlitschka, 12562:217–33. *Lecture Notes in Computer Science*. Cham: Springer International Publishing, 2020. https://doi.org/10.1007/978-3-030-64148-1_14.
- [8] Forbes, Kelly. «**Opening the Path to Ethics in Artificial Intelligence.**» *AI and Ethics* 1, no. 3 (August 2021): 297–300. <https://doi.org/10.1007/s43681-020-00031-2>.
- [9] Hagendorff, Thilo. «**The Ethics of AI Ethics: An Evaluation of Guidelines.**» *Minds and Machines* 30, no. 1 (March 2020): 99–120. <https://doi.org/10.1007/s11023-020-09517-8>.

- [10] Heer, Jeffrey. «**Agency plus Automation: Designing Artificial Intelligence into Interactive Systems.**» *Proceedings of the National Academy of Sciences* 116, no. 6 (February 5, 2019): 1844–50. <https://doi.org/10.1073/pnas.1807184115>.
- [11] Holstein, Kenneth, Jennifer Wortman Vaughan, Hal Daumé, Miro Dudik, and Hanna Wallach. «**Improving Fairness in Machine Learning Systems: What Do Industry Practitioners Need?**» In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–16. Glasgow Scotland UK: ACM, 2019. <https://doi.org/10.1145/3290605.3300830>.
- [12] Morley, Jessica, Luciano Floridi, Libby Kinsey, and Anat Elhalal. «**From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices.**» *Science and Engineering Ethics* 26, no. 4 (August 2020): 2141–68. <https://doi.org/10.1007/s11948-019-00165-5>.
- [13] Noul, Deni. «**Ethical Horizons in AI Development: Interdisciplinary Perspectives for Responsible Progress.**» n.d.
- [14] Prem, Erich. «**From Ethical AI Frameworks to Tools: A Review of Approaches.**» *AI and Ethics* 3, no. 3 (August 2023): 699–716. <https://doi.org/10.1007/s43681-023-00258-9>.
- [15] Sanderson, Conrad, David Douglas, Qinghua Lu, Emma Schleiger, Jon Whittle, Justine Lacey, Glenn Newnham, Stefan Hajkowicz, Cathy Robinson, and David Hansen. «**AI Ethics Principles in Practice: Perspectives of Designers and Developers.**» *IEEE Transactions on Technology and Society* 4, no. 2 (June 2023): 171–87. <https://doi.org/10.1109/TTS.2023.3257303>.
- [16] Seppälä, Akseli, Teemu Birkstedt, and Matti Mäntymäki. «**From Ethical AI Principles to Governed AI**» 2021.
- [17] Shneiderman, Ben. «**Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems.**» *ACM Transactions on Interactive Intelligent Systems* 10, no. 4 (December 31, 2020): 1–31. <https://doi.org/10.1145/3419764>.
- [18] Stahl, Bernd Carsten. «**Embedding Responsibility in Intelligent Systems: From AI Ethics to Responsible AI Ecosystems.**» *Scientific Reports* 13, no. 1 (May 18, 2023): 7586. <https://doi.org/10.1038/s41598-023-34622-w>.
- [19] Ευρωπαϊκό Κοινοβούλιο 2019- 2024, «**Τεχνητή Νοημοσύνη: ζητήματα ερμηνείας και εφαρμογής του διεθνούς δικαίου**», Κείμενα που εγκρίθηκαν, Προσωρινή έκδοση, P9_TA-PROV (2021) 0009 σ.5 https://www.europarl.europa.eu/doceo/document/TA-9-2021-0009_EL.pdf
- [20] IBM «**What is Deep Learning?** » IBM <https://www.ibm.com/topics/deep-learning>

- [21] «Τρεις τάσεις θα καθορίσουν τις εξελίξεις στην τεχνητή νοημοσύνη» <https://www.euro2day.gr/specials/opinions/article/2231246/treis-taseis-tha-kathorisoun-tis-exelixeis-sthn-te.html>
- [22] Αντουάν Ταρντίφ «Αποκαλύπτοντας τη δύναμη των μεγάλων γλωσσικών μοντέλων (LLM)»(2024) www.unite.ai/el/large-language-models/
- [23] Spandis 2023, «*The Ethics of Artificial Intelligence and Automation*» <https://dspace.lib.ntua.gr/xmlui/bitstream/handle/123456789/59650/The%20Ethics%20of%20Artificial%20Intelligence%20and%20Automation%20-spandis.pdf?isAllowed=y&sequence=1>
- [24] MIT Press 2023, «*AI in Healthcare: Challenges and Ethical Considerations*» <https://www.libre.gr/2024/03/19/fakelos-ai-sta-vathia-tis-technitis-noimosynis-apo-to-a-eos-to-o/>
- [25] Harvard Law Review 2021, «**Regulating AI: EU AI Act and Ethical Considerations**» από τον ΣΕΒ https://www.sev.org.gr/wp-content/uploads/2024/05/ODIGOS-SEV_AI.pdf
- [26] Ross, C., & Swetlitz, I. (2018). «IBM's Watson recommended 'unsafe and incorrect' cancer treatments, internal documents show.» Stat News. <https://www.statnews.com/wp-content/uploads/2018/09/IBMs-Watson-recommended-unsafe-and-incorrect-cancer-treatments-STAT.pdf>
- [27] Strickland, E. (2019). «**IBM Watson, heal thyself: How IBM overpromised and underdelivered on AI health care.**» IEEE Spectrum. <https://ieeexplore.ieee.org/abstract/document/8678513>
- [28] Powles, J., & Hodson, H. (2017). «**Google DeepMind and healthcare in an age of algorithms.**» Health and Technology Journal. <https://link.springer.com/article/10.1007/s12553-017-0179-1>
- [29] Hern, A. (2017). «**Google DeepMind 1.6m patient records deal 'inappropriate,'**» says UK watchdog. The Guardian. <https://www.theguardian.com/technology/2017/may/16/google-deepmind-16m-patient-record-deal-inappropriate-data-guardian-royal-free>