



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ
ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ
ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

ΣΥΝΘΕΣΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΜΕ ΣΚΟΠΟ ΤΗΝ ΑΝΑΠΤΥΞΗ ΗΘΙΚΩΝ
ΛΥΣΕΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Βικέντιος Βιτάλης

Επιβλέπων : Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος/2024



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ
ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

ΣΥΝΘΕΣΗ ΜΕΘΟΔΟΛΟΓΙΑΣ ΜΕ ΣΚΟΠΟ ΤΗΝ ΑΝΑΠΤΥΞΗ ΗΘΙΚΩΝ
ΛΥΣΕΩΝ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Βικέντιος Βιτάλης

Επιβλέπων : Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 12^η Ιουλίου 2024.

.....
Δημήτριος Ασκούνης
Καθηγητής Ε.Μ.Π.

.....
Ιωάννης Ψαρράς
Καθηγητής Ε.Μ.Π.

.....
Χρυσόστομος Δούκας
Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος/2024

.....

Βικέντιος Βιτάλης

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Copyright © Βικέντιος Βιτάλης, 2024.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, η αποθήκευση και η διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, η αποθήκευση και η διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Η παρούσα μελέτη διεισδύει στην Τεχνητή Νοημοσύνη – ΤΝ και την κρίσιμης σημασίας συζήτηση γύρω από την Ηθική ΤΝ. Αναπτύσσεται το θεωρητικό υπόβαθρο, τα θεωρητικά μοντέλα και οι αλγόριθμοι ΤΝ, τα είδη ΤΝ και πως η Μηχανική Μάθηση - ΜΜ αποτελεί την βάση της. Εν συνεχεία, διερευνούνται οι γνωστικές ικανότητες - λειτουργίες της ΤΝ και οι εφαρμογές της σε όλους τους κλάδους της Οικονομίας και τις Επιχειρήσεις. Αναπτύσσεται η ταξονομία της Ομάδας Εμπειρογνομόνων Υψηλού Επιπέδου – ΟΕΥΕ Τεχνητής Νοημοσύνης και οι Ευρωπαϊκές κατευθυντήριες γραμμές για Αξιόπιστη ΤΝ, το ALTAI Ερωτηματολόγιο και το ΑΙ Νομικό Πλαίσιο. Ερευνούνται πρακτικές εφαρμογές συστημάτων ΤΝ στην βιομηχανία της Ναυτιλίας, στην Γενετική Τεχνητή Νοημοσύνη και στην Αυτοκινητοβιομηχανία. Εν κατακλείδι, τα συστήματα ΤΝ και η ηθικότητά τους αξιολογούνται από πρακτική μεθοδολογία αξιολόγησης.

Λέξεις Κλειδιά: Τεχνητή Νοημοσύνη, Ηθική Τεχνητή Νοημοσύνη, Αλγόριθμοι Τεχνητής Νοημοσύνης, ALTAI Ερωτηματολόγιο, ΑΙ Νομικό Πλαίσιο, Μεθοδολογία Ηθικότητας Συστημάτων ΤΝ, Ναυτιλία, Γενετική Τεχνητή Νοημοσύνη, Αυτοκινητοβιομηχανία.

Abstract

This study delves into Artificial Intelligence – AI and the crucial debate around Ethical AI. The theoretical background, theoretical models and AI algorithms, the types of AI and how Machine Learning – ML consists its technical basis. Subsequently, the cognitive abilities – functions of AI and its applications in all branches of Economy and Businesses are explored. The taxonomy of the High-Level Expert Group – HLEG on AI and the European Guidelines for Reliable AI, the ALTAI Questionnaire and the AI Act are being developed. Practical applications of AI systems in the Maritime industry, the Generative AI and the Automotive industry are being discussed. Last but not least, AI systems and their ethicality are evaluated by practical evaluation methodology.

Key Words: Artificial Intelligence, Ethical Artificial Intelligence, Artificial Intelligence Algorithms, ALTAI Questionnaire, AI Act, Methodology for AI Systems Ethics, Maritime Industry, Generative AI, Automotive Industry.

Ευχαριστίες

Θέλω να ευχαριστήσω όλους εκείνους με τους οποίους άνοιξε και έκλεισε ο κύκλος των σπουδών μου στους Ηλεκτρολόγους Μηχανικούς και Μηχανικούς Υπολογιστών στο Εθνικό Μετσόβιο Πολυτεχνείο. Θέλω να ευχαριστήσω εκείνους που συνεχίζουν σήμερα να είναι δίπλα μου και τους περαστικούς καθώς συνετέλεσαν στο να είμαι αυτός που είμαι σήμερα.

Τον επιβλέποντα καθηγητή κο. Δημήτριο Ασκούνη, αξιότιμο μέλος της ΣΗΜΜΥ και του ΕΜΠ, καθώς ο ίδιος και τα μαθήματά του στην ροή Διοίκησης και Απόφασης είναι ο λόγος που ανακάλυψα το πάθος μου για την διοίκηση και τις επιχειρήσεις.

Θέλω να ευχαριστήσω τον υποψήφιο διδάκτορα κο. Χρήστο Κοντζίνο για την συνεχή καθοδήγησή του κατά την διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Οι θερμότερες ευχαριστίες ανήκουν στην οικογένειά μου για την ανατροφή μου και την αγάπη τους.

Κλείνοντας, ευχαριστώ ιδιαίτερα όλους εκείνους που είναι σήμερα δίπλα μου στην ζωή και μοιραζόμαστε τις νίκες μας και τις στιγμές μας.

«Δεν υπάρχει τίποτε αδύνατον γι' αυτόν που θα προσπαθήσει.»

~ Μέγας Αλέξανδρος ~

Περιεχόμενα

.....	2
Περίληψη.....	3
Abstract.....	5
Ευχαριστίες.....	7
Λίστα Πινάκων.....	13
Λίστα Εικόνων.....	15
Ακρωνύμια.....	17
Κεφάλαιο 1: Εισαγωγή.....	21
1.1 Αντικείμενο και Σκοπός.....	21
1.2 Μεθοδολογία Διπλωματικής.....	23
1.3 Οργάνωση Κειμένου.....	23
Κεφάλαιο 2: Θεωρητικό Υπόβαθρο Τεχνητής Νοημοσύνης.....	24
2.1 Τεχνητή Νοημοσύνη.....	24
2.1.1 Ορισμός Συστήματος Τεχνητής Νοημοσύνης.....	24
2.1.2 Θεωρητικά Μοντέλα και Αλγόριθμοι.....	24
2.1.3 Συμβολική Τεχνητή Νοημοσύνη.....	26
2.1.4 Συνδυαστική Τεχνητή Νοημοσύνη.....	27
2.1.5 Εξελικτική Τεχνητή Νοημοσύνη.....	28
2.1.6 Μπεϋζιανή Τεχνητή Νοημοσύνη.....	29
2.1.7 Αναλογική Τεχνητή Νοημοσύνη.....	29
2.1.8 Μηχανική Μάθηση: Πρακτική Εφαρμογή της Τεχνητής Νοημοσύνης.....	29
2.1.9 Επιβλεπόμενη Μάθηση.....	30
2.1.10 Ανεπίβλεπτη Μάθηση.....	31
2.1.11 Ημι-Επιβλεπόμενη Μάθηση.....	32

2.1.12 Ενισχυτική Μάθηση	32
2.1.13 Αντιθετική Μάθηση.....	32
2.1.14 Μεταφορική Μάθηση	33
2.1.15 Γνωστικές Ικανότητες που Αναπαράγονται από την Τεχνητή Νοημοσύνη...	33
2.1.16 Λειτουργίες που εκτελούνται από την Τεχνητή Νοημοσύνη	35
2.1.17 Τεχνητή Νοημοσύνη στις Επιχειρήσεις.....	37
2.1.18 Τεχνητή Νοημοσύνη ανά Κλάδο της Οικονομίας	39
2.1.19 Ταξονομία Επίβλεψης Ομάδας Εμπειρογνομόνων Υψηλού Επιπέδου Τεχνητής Νοημοσύνης.....	40
2.2 Ηθικό και Νομικό Πλαίσιο Τεχνητής Νοημοσύνης.....	41
2.2.1 Ευρωπαϊκές Κατευθυντήριες Γραμμές για Αξιόπιστη Τεχνητή Νοημοσύνη ...	41
2.2.2 Αξιόπιστη Τεχνητή Νοημοσύνη.....	41
2.2.3 Ακροατήριο και Πεδίο Εφαρμογής	43
2.2.4 Πλαίσιο Αξιόπιστης Τεχνητής Νοημοσύνης	43
2.2.5 Νομικά Συμμορφούμενη Τεχνητή Νοημοσύνη	44
2.2.6 Ηθική Τεχνητή Νοημοσύνη	44
2.2.7 Εύρωστη Τεχνητή Νοημοσύνη.....	45
2.2.8 Τα θεμέλια για Αξιόπιστη Τεχνητή Νοημοσύνη	45
2.2.9 Τα Θεμελιώδη Δικαιώματα ως Ηθικά και Νομικά Δικαιώματα	46
2.2.10 Τα Θεμελιώδη Δικαιώματα ως Βάση για Αξιόπιστη Τεχνητή Νοημοσύνη	47
2.2.11 Σεβασμός Ανθρώπινης Αξιοπρέπειας.....	48
2.2.12 Ελευθερία Ατόμου.....	48
2.2.13 Σεβασμός στην Δημοκρατία, την Δικαιοσύνη και το Κράτος Δικαίου.....	49
2.2.14 Ισότητα, Μη-διάκριση και Αλληλεγγύη	49
2.2.15 Δικαιώματα των Πολιτών	49

2.2.16 Ηθικές Αρχές στο Πλαίσιο Συστημάτων Τεχνητής Νοημοσύνης	50
2.2.17 Αρχή Σεβασμού Ανθρώπινης Αυτονομίας	50
2.2.18 Αρχή Πρόληψης Βλάβης.....	51
2.2.19 Αρχή Δικαιοσύνης.....	51
2.2.20 Αρχή Εξηγησιμότητας.....	52
2.2.21 Εντάσεις Μεταξύ των Αρχών	53
2.2.22 Προϋποθέσεις για Αξιόπιστη Τεχνητή Νοημοσύνη	53
2.2.23 Ανθρώπινη Παρέμβαση και Εποπτεία	55
2.2.24 Τεχνική Ευρωστία και Ασφάλεια	57
2.2.25 Ιδιωτικότητα και Επίβλεψη Δεδομένων	59
2.2.26 Διαφάνεια.....	60
2.2.27 Ποικιλομορφία, κατά Διακρίσεων, Δικαιοσύνη	61
2.2.28 Κοινωνική και Περιβαλλοντολογική Ευεξία.....	62
2.2.29 Λογοδοσία	63
2.2.30 AI Act – AI Νομικό Πλαίσιο Τεχνητής Νοημοσύνης.....	65
Κεφάλαιο 3: Ερευνητικό Πλαίσιο Νομικών και Ηθικών Λύσεων Τεχνητής Νοημοσύνης	67
3.1 Ερευνητικές Δημοσιεύσεις Ηθικών Λύσεων Τεχνητής Νοημοσύνης στην Ναυτιλία	67
3.1.1 Ψηφιακοί Πλοηγοί.....	67
3.1.2 Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης	71
3.1.3 Θαλάσσια Αυτόνομα Πλοία Επιφάνειας.....	76
3.2 Ερευνητικές Δημοσιεύσεις Ηθικών Λύσεων Τεχνητής Νοημοσύνης στην Γενετική Τεχνητή Νοημοσύνη	80
3.2.1 Διαχείριση Προϊόντων Λογισμικού	80
3.2.2 Μηχανική Λογισμικού και Κυβερνο-Φυσικά Συστήματα	85
3.2.3 Διαδίκτυο των Πραγμάτων	88

3.3 Ερευνητικές Δημοσιεύσεις Ηθικών Λύσεων Τεχνητής Νοημοσύνης στην Αυτοκινητοβιομηχανία.....	94
3.3.1 Ταυτόχρονος Εντοπισμός και Χαρτογράφηση.....	97
3.3.2 Έλεγχος και Δημιουργία Πορείας.....	100
3.3.3 Βιομηχανική και Βιβλιογραφική Ανασκόπηση Αυτοκινητοβιομηχανίας	107
Κεφάλαιο 4: Σύνθεση Μεθοδολογίας για την Ανάπτυξη Ηθικών Λύσεων Τεχνητής Νοημοσύνης.....	113
4.1 Μεθοδολογικό Πλαίσιο Αξιολόγησης και Ανάπτυξης Ηθικών Λύσεων Τεχνητής Νοημοσύνης.....	113
4.2 Εφαρμογή Μεθοδολογίας.....	115
4.2.1 Περίπτωση Χρήσης στην Ναυτιλία.....	115
4.2.1.1 Εφαρμογή Μεθοδολογίας στους Ψηφιακούς Πλοηγούς.....	115
4.2.1.2 Εφαρμογή Μεθοδολογίας στην Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης.....	118
4.2.1.3 Εφαρμογή Μεθοδολογίας στα Θαλάσσια Αυτόνομα Πλοία Επιφάνειας....	121
4.2.2 Περίπτωση Χρήσης στην Γενετική Τεχνητή Νοημοσύνη	124
4.2.2.1 Εφαρμογή Μεθοδολογίας στην Διαχείριση Προϊόντων Λογισμικού.....	124
4.2.2.2 Εφαρμογή Μεθοδολογίας στην Μηχανική Λογισμικού και Κυβερνο-Φυσικά Συστήματα	127
4.2.2.3 Εφαρμογή Μεθοδολογίας στο Διαδίκτυο των Πραγμάτων	131
4.2.3 Περίπτωση Χρήσης στην Αυτοκινητοβιομηχανία	135
4.2.3.1 Εφαρμογή Μεθοδολογίας στον Ταυτόχρονο Εντοπισμό και Χαρτογράφηση	135
4.2.3.2 Εφαρμογή Μεθοδολογίας στον Έλεγχο και Δημιουργία Πορείας	140
Κεφάλαιο 5: Συμπεράσματα και Μελλοντικές Προοπτικές.....	146
5.1 Συμπεράσματα.....	146

5.1.1 Ανάπτυξη ALTAI Ερωτηματολογίου.....	146
5.1.2 Τομέας Ερευνητικής Ανάπτυξης: Ναυτιλιακά	147
5.1.3 Τομέας Ερευνητικής Ανάπτυξης: Γενετική Τεχνητή Νοημοσύνη	150
5.1.4 Τομέας Ερευνητικής Ανάπτυξης: Αυτοκινητοβιομηχανία	153
5.2 Προτάσεις για περαιτέρω έρευνα	156
5.2.1 Θεσμοθέτηση Ηθικής Τεχνητής Νοημοσύνης	156
5.2.2 Όραμα.....	156
Βιβλιογραφία	158

Λίστα Πινάκων

Πίνακας 1: Περιοχές Προτεραιότητας Ευρωπαϊκού Κοινοβουλίου και Θαλάσσιας Ολοκληρωμένης Επίγνωσης της Επιτήρησης – ALTAI Ερωτηματολόγιο: Ιδιωτικότητα και Διακυβέρνηση δεδομένων	72
Πίνακας 2: Συνοψισμός νομικού και δεοντολογικού πλαισίου της Νοημοσύνης Ανοιχτού Κώδικα – ALTAI Ερωτηματολόγιο: Κοινωνική και Περιβαλλοντολογική Ευεξία.....	74
Πίνακας 3: Δεοντολογικές και νομικές προκλήσεις της Νοημοσύνης Κοινωνικών Μέσων – ALTAI Ερωτηματολόγιο: Διαφάνεια, Ασφάλεια.....	74
Πίνακας 4: Πλαίσιο για μεγάλα δεδομένα – ALTAI Ερωτηματολόγιο: Ιδιωτικότητα και Διακυβέρνηση δεδομένων	76
Πίνακας 5: Σχεδιαστικές Αρχές για Θαλάσσια Αυτόνομα Πλοία Επιφάνειας και οδοί για περαιτέρω έρευνα – ALTAI Ερωτηματολόγιο: Αξιοπιστία, Διαφάνεια και Λογοδοσία....	77
Πίνακας 6: Εφαρμογές της Γενετικής Τεχνητής Νοημοσύνης στην Διαχείριση Προϊόντων Λογισμικού.	81
Πίνακας 7: ALTAI Ερωτηματολόγιο: Δηλώσεις δεοντολογικών αρχών.....	86
Πίνακας 8: Αυτόνομη Οδήγηση: Εξέλιξη Τεχνολογιών – Κλειδιών [31]	104
Πίνακας 9: ALTAI Προτάσεις για Ψηφιακούς Πλοηγούς - ΨΠ.	117
Πίνακας 10: ALTAI Προτάσεις για Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ.....	120
Πίνακας 11: ALTAI Προτάσεις για Θαλάσσια Αυτόνομα Πλοία Επιφάνειας - ΘΑΠΕ. ..	123
Πίνακας 12: ALTAI Προτάσεις για Διαχείριση Προϊόντων Λογισμικού - ΔΠΛ	126
Πίνακας 13: ALTAI Προτάσεις για Μηχανική Λογισμικού – ΜΛ και Κυβερνο-Φυσικά Συστήματα - ΚΦΣ.....	129
Πίνακας 14: ALTAI Προτάσεις για Διαδίκτυο των Πραγμάτων - ΔτΠ.....	133
Πίνακας 15: ALTAI Προτάσεις για Ταυτόχρονο Εντοπισμό και Χαρτογράφηση - TEX (0 – 3).....	137

Πίνακας 16: ALTAI Προτάσεις για Έλεγχο και Δημιουργία Πορείας (4 – 5) 142

Λίστα Εικόνων

Εικόνα 1: ΟΕΥΕ, Ταξονομία Τεχνητής Νοημοσύνης ^[1]	41
Εικόνα 2: ΟΕΥΕ, Αλληλεπίδραση των επτά Απαιτήσεων Αξιόπιστης ΤΝ ^[4]	54
Εικόνα 3: AI Act, το Πρώτο Νομικό Πλαίσιο ΤΝ ^[15]	66
Εικόνα 4: Ηθικά ζητήματα στην βιομηχανία ΑΟ ^[46]	108
Εικόνα 5: Το πλαίσιο αξιόπιστης Τεχνητής Νοημοσύνης όπως καθορίζεται από την ΟΕΥΕ	114
Εικόνα 6: Ιεραρχία AI Νομικού Πλαισίου: Ψηφιακοί Πλοηγοί – ΨΠ	116
Εικόνα 7: ALTAI Αξιολόγηση: Ψηφιακών Πλοηγών - ΨΠ	116
Εικόνα 8: Ιεραρχία AI Νομικού Πλαισίου: Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης – ΘΟΕΕ	119
Εικόνα 9: ALTAI Αξιολόγηση Θαλάσσιας Ολοκληρωμένης Επίγνωσης της Επιτήρησης - ΘΟΕΕ	119
Εικόνα 10: Ιεραρχία AI Νομικού Πλαισίου: Θαλάσσια Αυτόνομα Πλοία Επιφάνειας – ΘΑΠΕ	122
Εικόνα 11: ALTAI Αξιολόγηση: Θαλάσσια Αυτόνομα Πλοία Επιφάνειας - ΘΑΠΕ	122
Εικόνα 12: Ιεραρχία AI Νομικού Πλαισίου: Διαχείριση Προϊόντων Λογισμικού – ΔΠΛ. 125	
Εικόνα 13: ALTAI Αξιολόγηση: Διαχείριση Προϊόντων Λογισμικού – ΔΠΛ	125
Εικόνα 14: Ιεραρχία AI Νομικού Πλαισίου: Μηχανική Λογισμικού – ΜΛ και Κυβερνο-Φυσικά Συστήματα – ΚΦΣ.....	128
Εικόνα 15: ALTAI Αξιολόγηση: Μηχανική Λογισμικού και Κυβερνο-Φυσικά Συστήματα	129
Εικόνα 16: Ιεραρχία AI Νομικού Πλαισίου: Διαδίκτυο των Πραγμάτων - ΔΤΠ	132
Εικόνα 17: ALTAI Αξιολόγηση: Διαδίκτυο των Πραγμάτων – ΔΤΠ	132

Εικόνα 18: Ιεραρχία ΑΙ Νομικού Πλαισίου: Ταυτόχρονος Εντοπισμός και Χαρτογράφηση (0 – 3).....	136
Εικόνα 19: ALTAI Αξιολόγηση: Ταυτόχρονος Εντοπισμός και Χαρτογράφηση (0 – 3)	136
Εικόνα 20: Ιεραρχία ΑΙ Νομικού Πλαισίου: Έλεγχος και Δημιουργία Πορείας - ΕΔΠ (4 – 5).....	141
Εικόνα 21: ALTAI Αξιολόγηση: Έλεγχος και Δημιουργία Πορείας – ΕΔΠ (4 – 5).....	141

Ακρωνύμια

Ελληνικό Ακρωνύμιο	English Acronym
Τεχνητή Νοημοσύνη - ΤΝ	Artificial Intelligence - AI
Μηχανική Μάθηση - ΜΜ	Machine Learning - ML
Ομάδα Εμπειρογνομόνων Υψηλού Επιπέδου - ΟΕΥΕ	High-Level Expert Group - HLEG
Κατάλογος Αξιολόγησης για Αξιόπιστη Τεχνητή Νοημοσύνη - ΚΑΑΤΝ	Assessment List on Trustworthy Artificial Intelligence - ALTAI
Γενετική Τεχνητή Νοημοσύνη - ΓΤΝ	Generative Artificial Intelligence - GAI
Στενή Τεχνητή Νοημοσύνη - ΣΤΝ	Artificial Narrow Intelligence - ANI
Γενική Τεχνητή Νοημοσύνη - ΓΤΝ	Artificial General Intelligence - AGI
Τεχνητά Νευρωνικά Δίκτυα - ΤΝΔ	Artificial Neural Networks - ANN
Βαθιά Μάθηση - ΒΜ	Deep Learning - DL
Συλλογισμός Βάσει Κανόνων - ΣΒΚ	Rule-Based Reasoning - RBR
Συλλογιστική Βάσει Περιπτώσεων - ΣΒΠ	Case-Based Reasoning - CBR
Νοημοσύνη Σμήνους - ΝΣ	Swarm Intelligence - SI
Βελτιστοποίηση Σμήνους Σωματιδίων - ΒΣΣ	Particle Swarm Optimization - PSO
Αλγόριθμοι Τεχνητής Άλγης - ΑΤΑ	Artificial Algae Algorithms - AAA
Τεχνητό Σμήνος Ψαριών - ΑΣΨ	Artificial Fish Swarm - AFS
Μοντελοποίηση Βάσει Πρακτόρων - ΜΒΠ	Agent-Based Modelling - ABM
Καταργητικός Λογικός Προγραμματισμός - ΚΛΠ	Destructive Logic Programming - DeLP
Γενετικοί Αλγόριθμοι - ΓΑ	Generative Algorithms - GA
Διαφορική Εξέλιξη - ΔΕ	Differential Evolvment - DE
Αλγόριθμοι Αναζήτησης Αρμονίας - ΑΑΑ	Harmony Search Algorithms - HSA
Αλγόριθμοι Βελτιστοποίησης Κοινωνικής Μηχανικής - ΑΒΚΜ	Social Engineering Optimization Algorithms - SEOA
Κρυφά Μοντέλα Μαρκόβ - ΚΜΜ	Hidden Markov Models - HMM
Μηχανές Διανυσμάτων Υποστήριξης - ΜΔΥ	Support Vector Machines - SVM
k-Κοντινότερων Γειτόνων - k-ΚΓ	k-Nearest Neighbors - k-NN
Επιβλεπόμενη Μάθηση - ΕΠ	Supervised Learning - SL
Παλινδρόμηση Διαδικασίας Γκάους - ΠΔΓ	Gauss Process Regression - GPR
Πρόμπιτ Παλινδρόμηση - ΠΠ	Probit Regression - PR
Πιθανοτικά Γραφικά Μοντέλα - ΠΓΜ	Probalistic Graphical Models - PLM
Περιορισμένη Μηχανή Μπόλτζμαν - ΠΜΜ	Restricted Boltzmann Machine - RBM
Βαθιά Μηχανή Μπόλτζμαν - ΒΜΜ	Deep Boltzmann Machine - DBM
Βαθείς Αυτο-Κωδικοποιητές - ΒΑΚ	Deep Auto-Decoders - DAE
Βαθιά Επαναληπτικά Νευρωνικά Δίκτυα - ΒΕΝΔ	Deep Recursive Neural Networks - DRNN
Βαθιά Δίκτυα Πίστης - ΒΔΠ	Deep Belief Networks - DBN
Επαναληπτικά Νευρωνικά Δίκτυα - ΕΝΔ	Recursive Neural Networks - RNN

Συνελικτικά Νευρωνικά Δίκτυα - ΣΝΔ	Convolutional Neural Networks - CNN
Ανεπίβλεπτη Μάθηση - AM	Unsupervised Learning – UL
Ιεραρχική Συσσωρευτική Ομαδοποίηση - ΙΣΟ	Hierarchical Accumulative Clustering - HAC
Μεγιστοποίηση Προσδοκίας - EM	Expectation Maximization - EM
Χωρική Ομαδοποίηση Βάσει την Πυκνότητα Εφαρμογών με Θόρυβο - ΧΟΒΠΕΘ	Density Based Spatial Clustering on Applications with Noise - DBSCAN
Ανάλυση Κύριων Συνιστωσών - ΑΚΣ	Principal Components Analysis - PCA
Γενετικά Αντιφατικά Δίκτυα - ΓΑΔ	Generative Adversarial Networks - GAN
Αυτο-Κωδικοποιητές - ΑΕ	Auto-Decoders - AD
Ημι-Επιβλεπόμενη Μάθηση - ΗΕΜ	Semi-Unsupervised Learning – SUL
Ενισχυτική Μάθηση - EM	Reinforcement Learning – RL
Αντιθετική Μάθηση - AM	Adversarial Learning – AL
Μεταφορική Μάθηση - MM	Transfer Learning – TL
Διαδικασίες Μνήμης - ΔΜ	Memory Processes - MP
Αισθητικοκινητική Αλληλεπίδραση – ΑΑ	Sensorimotor Interaction - SI
Οπτική Επεξεργασία - ΟΕ	Visual Processing - VP
Ακουστική Επεξεργασία - ΑΕ	Auditory Processing - AP
Προσοχή και Αναζήτηση - ΠΑ	Attention and Search - AS
Σχεδιασμός και Δράση - ΣΔ	Planning and Action - PA
Κατανόηση και Έκφραση - ΚΕ	Comprehension and Expression - CE
Συναισθήματα και Αυτο-Έλεγχος - ΣΑΕ	Emotions and Self-Control - ESC
Εννοιολογική Μάθηση και Αφαίρεση - ΕΜΑ	Conceptualization Learning and Abstraction - CLS
Ποσοτικός και Λογικός Συλλογισμός – ΠΛΣ	Quantitative and Logical Reasoning - QLR
Μοντελοποίηση Νου και Κοινωνική Αλληλεπίδραση - ΜΝΚΑ	Mind Modelling and Social Interaction - MMSI
Μηχανή-σε-Μηχανή - ΜσΜ	Machine-to-Machine – M2M
Μηχανή-σε-Σύννεφο – ΜσΣ	Machine-to-Cloud - M2C
Φυσική Γλώσσα Επεξεργασίας - ΦΓΕ	Natural Language Processing - NLP
Τεχνολογία Πληροφοριών και Επικοινωνιών – ΤΠΕ	Information and Communication Technology - ICT
Γενικός Κανονισμός για την Προστασία Δεδομένων - ΓΚΠΔ	General Data Protection Regulation - GDPR
Οδηγία Ευθύνης Προϊόντων - ΟΕΠ	Product Liability Directive - PLD
Ελεύθερη Ροή Μη-προσωπικών Δεδομένων - ΕΡΜΔ	Free Flow of Non-personal Data - FFND
Ευρωπαϊκή Ένωση - ΕΕ	European Union - EU
Ευρωπαϊκή Ομάδα Δεοντολογίας – ΕΟΔ	European Group on Ethics - EGE
Άνθρωπος Εντός Βρόχου - ΑΕΒ	Human-In-the-Loop - HIL
Άνθρωπος Πάνω στον Βρόχο - ΑΠΒ	Human-On-the-Loop - HOTL
Άνθρωπος στην Απόφαση - ΑΣΑ	Human-In-Command - HIC
Διοξείδιο του Άνθρακα – CO ₂	Carbon Dioxide - CO ₂

Βασικοί Δείκτες Επιδόσεων – ΒΔΕ	Key Performance Indicators - KPIs
Πολυ-Επίπεδα Νευρωνικά Δίκτυα - ΠΕΝΔ	Multi-Layer Neural Networks – MLNN
Συνελικτικά Νευρωνικά Δίκτυα - ΣΝΔ	Convolutional Neural Networks – CNN
Επαναληπτικά Νευρωνικά Δίκτυα - ΕΝΔ	Recursive Convolutional Networks – RCN
Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ	Maritime Integrated Surveillance Awareness – MARISA
Νοημοσύνη Ανοιχτού Κώδικα – ΝΑΚ	Open-Source Intelligence – OSINT
Νοημοσύνη Κοινωνικών Μέσων – ΝΚΜ	Social Media Intelligence – SOCMINT
Θαλάσσια Αυτόνομα Πλοία Επιφάνειας – ΘΑΠΕ	Maritime Autonomous Surface Ships – MASS
Απομακρυσμένα Κέντρα Ελέγχου - ΑΚΕ	Remote Control Centers - RCC
Ανάλυση Γνωστικών Εργασιών – ΑΓΕ	Cognitive Task Analysis – CTA
Ιεραρχική Ανάλυση Εργασίας – ΙΕΑ	Hierarchical Task Analysis – HTA
Συστήματα Υποστήριξης Απόφασης – ΣΥΑ	Decision Support Systems - DSS
Ομάδες Ανθρώπου Μηχανής – ΟΑΜ	Human-Machine Team – HMT
Επίγνωση της Κατάστασης – ΕΚ	Situation Awareness - SA
Διαχείριση Προϊόντων Λογισμικού - ΔΠΛ	Software Product Management - SPM
Προδιαγραφές Απαιτήσεων Λογισμικού – ΠΑΛ	Software Requirement Specifications – SRS
Διεπαφή Χρήστη/Εμπειρία Χρήστη - ΔΧ/ΕΧ	User Interface/User Experience - UI/UX
Μεγάλα Γλωσσικά Μοντέλα – ΜΓΜ	Large Language Models – LLMs
Μηχανική Λογισμικού – ΜΛ	Software Engineering – SE
Κυβερνο-Φυσικά Συστήματα – ΚΦΣ	Cyber-Physical Systems – CPS
Εκπαιδευτική Επιστημονική και Πολιτιστική Οργάνωση των Ηνωμένων Εθνών - ΕΕΠΟΗΕ	United Nations Educational Scientific and Cultural Organization – UNESCO
Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών - ΙΗΗΜ	Institute of Electrical and Electronics Engineers – IEEE
Διεθνείς Μηχανές Επιχειρήσεων - ΔΜΕ	International Business Machines – IBM
Διαδίκτυο των Πραγμάτων – ΔτΠ	Internet of Things – IoT
Βιομηχανικό Διαδίκτυο των Πραγμάτων – ΒΔτΠ	Industrial Internet of Things – IIoT
Αυτόνομα Οχήματα – ΑΟ	Autonomous Vehicles – AV
Προηγμένα Συστήματα Υποβοήθησης Οδηγού – ΠΣΥΟ	Advanced Driver Assistance Systems - ADAS
Κοινωνία Μηχανικών Αυτοκινήτων – ΚΜΑ	Society of Automotive Engineers – SAE
Ανίχνευση Φωτός και Εμβέλεια – ΑΦΕ	Light Detection And Ranging - LiDAR
Ταυτόχρονος Εντοπισμός και Χαρτογράφηση – ΤΕΧ	Simultaneous Localization and Mapping – SLM
Λογισμικό Εντός Βρόχου - ΛΕΒ	Software-in-the-Loop – SIL
Υλικό Εντός Βρόχου – ΥΕΒ	Hardware-in-the-Loop – HIL
Ανθρώπινος Ουσιαστικός Έλεγχος – ΑΟΕ	Human Meaningful Control – HMC

Αρχή Προστασίας Δεδομένων – ΑΠΔ	Data Protection Authority – DPA
Διεθνής Ένωση Τηλεπικοινωνιών – ΔΕΤ	International Telecommunication Union – ITU

Κεφάλαιο 1: Εισαγωγή

1.1 Αντικείμενο και Σκοπός

Η Τεχνητή Νοημοσύνη – TN ή Artificial Intelligence – AI έχει γίνει αναπόσπαστο μέρος της καθημερινής μας ζωής, είναι συνυφασμένη με διάφορους τομείς, ενώ παράλληλα ενισχύει τις δυνατότητες στις επιχειρήσεις, στην υγειονομική περίθαλψη, τη γεωργία και τις δημόσιες υπηρεσίες. Ωστόσο, καθώς η TN γίνεται όλο και πιο διαδεδομένη, η συζήτηση γύρω από τις ηθικές της επιπτώσεις αποκτά μεγαλύτερη δυναμική. Γίνονται πολλές συζητήσεις σχετικά με την εφαρμογή της ηθικής στην TN και πολλοί εμπειρογνώμονες προσπαθούν να δημιουργήσουν έναν ολοκληρωμένο χάρτη των ηθικών προβληματισμών, των στρατηγικών εφαρμογών και των αναμενόμενων κατευθύνσεων του τοπίου της TN.

Αποκαλύπτοντας την Ηθική Τεχνητή Νοημοσύνη

Η ηθική TN, το ηθικό πλαίσιο των τεχνολογιών TN, διασφαλίζει ότι η αδιάκοπη πορεία των τεχνολογικών δυνατοτήτων ευθυγραμμίζεται με τις θεμελιώδεις κοινωνικές αξίες και την ανθρώπινη αξιοπρέπεια. Αυτό το πλαίσιο δεν είναι ένα στατικό σύνολο εντολών, αλλά μια ζωντανή, αναπνέουσα κατανόηση του τρόπου με τον οποίο πρέπει να χρησιμοποιούνται τεχνολογίες όπως η TN. Η κατανόηση αυτή εξελίσσεται με κάθε άλμα προόδου στην TN, προκαλώντας τόσο τους προγραμματιστές όσο και τους χρήστες να εξετάσουν τις επιπτώσεις κάθε καινοτομίας. Υπηρετεί τον σκοπό της συνείδησης της TN, διερευνώντας τις προθέσεις, εξετάζοντας τις μεθόδους και αξιολογώντας τις επιπτώσεις των συστημάτων TN, διασφαλίζοντας έτσι ότι αυτά δεν είναι απλώς αποτελεσματικά αλλά και δίκαια, όχι απλά ισχυρά αλλά και με αρχές.

Τα συστήματα TN από την φύση τους δεν είναι αδρανείς αλγόριθμοι που «κάθονται» σε ψηφιακά ράφια – είναι δυναμικές οντότητες, που πάλλονται με την δυνατότητα να αναλύουν τεράστια σύνολα δεδομένων, να λαμβάνουν αποφάσεις και το πιο κρίσιμο, να επηρεάζουν τις ανθρώπινες ενέργειες και τους κοινωνικούς κανόνες. Με εφαρμογές που εκτείνονται σε όλο το φάσμα της υγειονομικής περίθαλψης έως την προγνωστική αστυνόμευση, καθίσταται ολοφάνερη η επιτακτική ανάγκη για ηθική προσέγγιση της TN.

Καθώς αυτά τα συστήματα γίνονται όλο και πιο αυτόνομα και οι αποφάσεις τους όλο και πιο σημαντικές, η ηθική διακυβέρνηση της ΤΝ μετατρέπεται από θεωρητική άσκηση σε πρακτική αναγκαιότητα, ένα προπύργιο ενάντια στην διάβρωση της δράσης και της ισότητας μπροστά στην συνεχή τεχνολογική πρόοδο.

Δεοντολογικές Προκλήσεις και Στρατηγικές Απαντήσεις

Οι ηθικές προκλήσεις που θέτει η ΤΝ, ιδίως στο πεδίο της Γενετικής Τεχνητής Νοημοσύνης – ΓΤΝ ή Generative Artificial Intelligence – GAI, απαιτούν σαφείς και στρατηγικές απαντήσεις. Οι οργανισμοί που δεσμεύονται να διατυπώνουν με σαφήνεια τις προθέσεις και τους περιορισμούς των μοντέλων μπορούν να παρέχουν παραδείγματα για το πώς οι εταιρείες μπορούν να διαχειριστούν αυτές τις προκλήσεις. Παρέχοντας σαφή επικοινωνία και αυστηρή αξιολόγηση των αποτελεσμάτων της ΤΝ, οι ηγέτες του κλάδου μπορούν να διαχειριστούν τις ηθικές προκλήσεις όχι μόνο αναγνωρίζοντάς τες αλλά και αντιμετωπίζοντάς τες με υπευθυνότητα.

Δέσμευση για Χωρίς Αποκλεισμούς Τεχνητή Νοημοσύνη

Η επιδίωξη της ΤΝ χωρίς αποκλεισμούς είναι μια ισχυρή έκκληση για ποικιλομορφία και ισότητα, υπερασπιζόμενη την υπόθεση της εκπροσώπησης όλων των δημογραφικών ομάδων στην ανάπτυξη ΤΝ. Η δέσμευση αυτή αποτελεί απόδειξη της πεποίθησης ότι η ΤΝ πρέπει να είναι εργαλείο ενδυνάμωσης και όχι αποκλεισμού, βελτιώνοντας την ανθρώπινη εμπειρία για όλους, ανεξαρτήτως φυλής, φύλου ή γεωγραφίας.

Θεματικές Ανάπτυξης Ηθικού Λογισμικού ΤΝ

Στην παρούσα εργασία το ερευνητικό έργο έχει ως επίκεντρο εφαρμογές συστημάτων ΤΝ στην Ναυτιλία, στην Γενετική Τεχνητή Νοημοσύνη και στην Αυτοκινητοβιομηχανία. Αναλύονται τεχνικά χαρακτηριστικά, η ηθικότητα των εφαρμογών, ο τρόπος λειτουργίας τους, το πώς συνδέονται με τα Ευρωπαϊκά Νομικά Πλαίσια εύρυθμης λειτουργίας και ο βαθμός ρίσκου χρήσης τους.

1.2 Μεθοδολογία Διπλωματικής

Η εκπόνηση της εργασίας στηρίζεται σε μεγάλο βαθμό στην μεθοδολογία της βιβλιογραφικής έρευνας. Για την επίτευξη του σκοπού της χρειάζεται να μελετηθούν τρία κύρια ερευνητικά ζητήματα:

1. Το θεωρητικό υπόβαθρο της Τεχνητής Νοημοσύνης
2. Οι εφαρμογές ηθικών συστημάτων Τεχνητής Νοημοσύνης στην Ναυτιλία, στην Γενετική Τεχνητή Νοημοσύνη και στην Αυτοκινητοβιομηχανία
3. Η κατηγοριοποίηση τους με βάση τους κανονισμούς της Ευρωπαϊκής Ένωσης

Για τα πρώτα δύο ζητήματα μελετάται η υπάρχουσα βιβλιογραφία και για το τρίτο πραγματοποιείται συνδυαστική ανάλυση ακαδημαϊκών αναφορών και ψηφιακών εργαλείων αξιολόγησης συστημάτων ΤΝ.

1.3 Οργάνωση Κειμένου

Η παρούσα διπλωματική αναπτύσσεται σε πέντε κεφάλαια. Στο πρώτο κεφάλαιο αναλύεται η χρησιμότητα της παρούσας διπλωματικής. Το πώς η συζήτηση για την Ηθική Τεχνητή Νοημοσύνη την τελευταία περίοδο που διανύει η ανθρωπότητα έχει γίνει τόσο επίκαιρη, και αναπτύσσονται τα αίτια τα οποία καθιστούν την έρευνα όσον αφορά στην ηθικότητα των συστημάτων Τεχνητής Νοημοσύνης απαραίτητη. Εν συνεχεία, στο δεύτερο κεφάλαιο, περιγράφεται το Θεωρητικό υπόβαθρο της Τεχνητής Νοημοσύνης, οι βασικοί αλγόριθμοι, τα προαπαιτούμενα για ηθικά και νομικά συμμορφούμενα συστήματα Τεχνητής Νοημοσύνης καθώς και οι βασικοί τομείς εφαρμογής και οι μηχανισμοί της. Στο τρίτο κεφάλαιο αναπτύσσονται οι τρεις βασικές ερευνητικές περιοχές της παρούσας εργασίας για ηθικό λογισμικό ΤΝ, στην Ναυτιλία, στην Γενετική Τεχνητή Νοημοσύνη και στην Αυτοκινητοβιομηχανία. Στο τέταρτο κεφάλαιο κατηγοριοποιούνται με κριτήριο το ρίσκο τους οι αναφερθείσες εφαρμογές ΤΝ σύμφωνα με τα πλαίσια της Ευρωπαϊκής Ένωσης, μέσω ψηφιακών εργαλείων αξιολόγησης. Τέλος, στο πέμπτο κεφάλαιο αναπτύσσονται οι ορίζοντες περαιτέρω ερευνητικής ανάπτυξης και το όραμα για ΤΝ.

Κεφάλαιο 2: Θεωρητικό Υπόβαθρο Τεχνητής Νοημοσύνης

2.1 Τεχνητή Νοημοσύνη

2.1.1 Ορισμός Συστήματος Τεχνητής Νοημοσύνης

Η Ομάδα Εμπειρογνομόνων Υψηλού Επιπέδου Τεχνητής Νοημοσύνης – ΟΕΥΕ ή High Level Expert Group on Artificial Intelligence - HLEG ορίζει τα συστήματα Τεχνητής Νοημοσύνης ως «Συστήματα λογισμικού και πιθανώς συστήματα υλικού, σχεδιασμένα από τους ανθρώπους, στα οποία έχει ανατεθεί ένας σύνθετος στόχος. Δρουν στην φυσική ή στην ψηφιακή διάσταση, λαμβάνουν τις παραμέτρους του περιβάλλοντος μέσω δεδομένων, τα οποία ερμηνεύουν σε συλλογές δομημένων και μη-δομημένων δεδομένων και συλλογίζονται με βάση την γνώση ή επεξεργάζονται τις πληροφορίες που προέρχονται από τα δεδομένα ώστε να αποφασίζουν ποιες δράσεις είναι κατάλληλες να πραγματοποιηθούν προκειμένου να πετύχουν τον δοσμένο στόχο. Τα συστήματα Τεχνητής Νοημοσύνης χρησιμοποιούν είτε συμβολικούς κανόνες είτε μαθαίνουν μέσω αριθμητικών μοντέλων. Μπορούν επίσης να προσαρμόσουν την συμπεριφορά τους αναλύοντας το περιβάλλον τους με βάση το πώς επηρεάζεται από τις προηγούμενες δράσεις τους».

2.1.2 Θεωρητικά Μοντέλα και Αλγόριθμοι

Η Τεχνητή Νοημοσύνη - TN ή Artificial Intelligence – AI ταξινομείται σύμφωνα με τις θεωρητικές έννοιες που διέπουν την ομαδοποίηση των αλγορίθμων που χρησιμοποιούνται για την ανάπτυξη της τεχνολογίας TN. Η ταξινόμηση της TN είναι μια ταξινόμηση της ίδιας της τεχνολογίας με έννοιες που προέρχονται από τα μαθηματικά, τη λογική, τη φιλοσοφία και τη θεωρία της πληροφορίας. Η παρακάτω ταξινόμηση συνίσταται από διάφορες μελέτες διαφορετικών οπτικών γωνιών και διαχωρίζει την Τεχνητή Νοημοσύνη σε Στενή Τεχνητή Νοημοσύνη - ΣΤΝ ή Artificial Narrow

Intelligence - ANI και Γενική Τεχνητή Νοημοσύνη - ΓΤΝ ή Artificial General Intelligence - AGI.

Η ΣΤΝ είναι ο τύπος Τεχνητής Νοημοσύνης που υπάρχει σήμερα. Τα συστήματα ΣΤΝ μπορούν να εκτελούν ένα συγκεκριμένο έργο και να λειτουργούν εντός ενός προκαθορισμένου περιβάλλοντος. Η ΣΤΝ μπορεί να επεξεργάζεται δεδομένα με μεγάλη ταχύτητα και να ενισχύει την παραγωγικότητα και την αποδοτικότητα σε πολλές πρακτικές εφαρμογές. Ενώ η ΣΤΝ είναι ανώτερη σε εξειδικευμένους τομείς, δεν είναι ικανή για γενίκευση, δηλαδή, δεν είναι σε θέση να επαναχρησιμοποιεί τη γνώση που απέκτησε σε άλλους τομείς.

Η Γενική ΤΝ αναφέρεται σε μηχανές που παρουσιάζουν ανθρώπινη νοημοσύνη. Με άλλα λόγια, η Γενική ΤΝ έχει ως στόχο να εκτελεί οποιαδήποτε διανοητική εργασία που μπορεί να εκτελέσει ένας άνθρωπος. Η τρέχουσα κατάσταση της ανάπτυξης της Τεχνητής Νοημοσύνης απέχει πολύ από το να φτάσει στην Γενική ΤΝ. Υπό αυτό το πρίσμα, μια καλή κατηγοριοποίηση είναι αυτή της ταξινόμησης της ΤΝ σε Συμβολική, Συνδετική, Εξελικτική, Μπεϋζιανή και Αναλογική.

Η Συμβολική και η Συνδετική ΤΝ είναι τα πιο κυρίαρχα πεδία στην ιστορία της ΤΝ. Η συμβολική ήταν το κυρίαρχο πεδίο της ΤΝ μεταξύ των δεκαετιών του 1970 και του 1990, μια περίοδο με σημαντικές επενδύσεις και διαφημιστική εκστρατεία στα μέσα ενημέρωσης. Με έμπνευση τον ανθρώπινο εγκέφαλο, εφευρέθηκε ο «περσέπτρον» - perceptron που θεωρείται η γέννηση της Συνδετικής ΤΝ, το θεμέλιο των Τεχνητών Νευρωνικών Δικτύων - ΤΝΔ ή Artificial Neural Networks - ANN και της Βαθιάς Μάθησης - ΒΜ ή Deep Learning - DL. Από το 2010, η Βαθιά Μάθηση είναι ο κυρίαρχος τομέας της ΤΝ και ο λόγος πίσω από την τρέχουσα έξαρση της ΤΝ.

Παρακάτω παρατίθεται αναλυτικότερη εξήγηση των πεδίων και ένας κατάλογος αλγορίθμων για το εκάστοτε πεδίο. Τα πεδία συνδυάζονται συχνά στην πράξη και διαπλέκονται με άλλα πεδία από την επιστήμη των υπολογιστών, τη στατιστική και την επεξεργασία σήματος, τα οποία παραδοσιακά δεν περιλαμβάνονταν στην ΤΝ^[1,2,3,4].

2.1.3 Συμβολική Τεχνητή Νοημοσύνη

Διαχωρίζεται περαιτέρω η Συμβολική Τεχνητή Νοημοσύνη βασισμένη σε κανόνες, βασισμένη στην λογική και βασισμένη στην γνώση και αντιπροσωπεύει τις ορθολογικές διαδικασίες σκέψης και συλλογισμού. Ονομάζεται Συμβολική, διότι οι έννοιες και οι διαδικασίες εκφράζονται από σύνολα συμβόλων, τα οποία ακολουθούν ένα περιορισμένο σύνολο κανόνων που βασίζονται στη λογική και χρησιμοποιούν παρελθοντικές γνώσεις και περιορισμούς για τον καθορισμό του χώρου αναζήτησης. Ο συλλογισμός επιτυγχάνεται μέσω του συνδυασμού διαφορετικών συμβόλων που ακολουθούν λογική και τυπικούς κανόνες για την επίλυση ενός προβλήματος. Η Συμβολική ΤΝ υλοποιείται αρχικά ως "συστήματα εμπειρογνωμόνων" ή "συστήματα βασισμένα στη γνώση". Η κύρια ιδέα είναι να αποκτηθεί ανθρώπινη εξειδικευμένη γνώση σε έναν υπολογιστή και να διαμοιραστεί σαν πρόγραμμα σε πολλούς προσωπικούς υπολογιστές. Η Συμβολική ΤΝ έχει δύο συνιστώσες:

- I) Την βάση γνώσης - μια συλλογή γεγονότων, κανόνων και σχέσεων σε έναν συγκεκριμένο τομέα.
- II) Την μηχανή εξαγωγής συμπερασμάτων που περιγράφει τον τρόπο χειρισμού και συνδυασμού αυτών των συμβόλων.

Τα γεγονότα και οι κανόνες έχουν ρητή εκπροσώπηση και είναι τροποποιήσιμα. Τα συμβολικά εργαλεία και πλαίσια εργασίας δεν έχουν επαρκή εκφραστική δύναμη για να αποτυπώσουν το εύρος της γνώσης και της συμπεριφοράς των εμπειρογνωμόνων που απαιτούνται για την επίτευξη ικανοποιητικής απόδοσης. Η απόδοση των μεταγλωττισμένων γλωσσών, όπως η C, είναι καλύτερη από το συμβολικό περιβάλλον. Στη δεκαετία του '90, η συμβολική έρευνα και η υιοθέτηση της ΣΤΝ επιβραδύνεται. Οι ακόλουθες κατηγορίες υπάγονται στο πλαίσιο της Συμβολικής ΤΝ:

I) Συστήματα Εμπειρογνωμόνων

- ✓ Συλλογισμός Βάσει Κανόνων – ΣΒΚ ή Rule-Based Reasoning - RBR
- ✓ Ασαφή Συστήματα Εμπειρογνωμόνων
- ✓ Συλλογιστική Βάσει Περιπτώσεων – ΣΒΠ ή Case-Based Reasoning - CBR
- ✓ Πολυγλωσσικά Συστήματα βασισμένα στη Γνώση
- ✓ Γράφοι Γνώσης

II) Νοημοσύνη Σμήνους – ΝΣ ή Swarm Intelligence - SI

- ✓ Στοχαστική Αναζήτηση Διάχυσης
- ✓ Αλγόριθμοι Βελτιστοποίησης Σμήνους Σωματιδίων - ΒΣΣ ή Particle Swarm Optimization – PSO, οι οποίοι είναι Επιταχυνόμενοι & Υβριδικοί
- ✓ Αλγόριθμος Τεχνητής Άλγης – ΑΤΑ ή Artificial Algae Algorithms - AAA
- ✓ Τεχνητό Σμήνος Ψαριών – ΑΣΨ ή Artificial Fish Swarm - AFS
- ✓ Αλγόριθμος Βελτιστοποίησης Βροχόπτωσης
- ✓ Άλλοι Ζωικοί Αλγόριθμοι

III) Μοντελοποίηση Βάσει Πρακτόρων – ΜΒΠ ή Agent-Based Modelling - ABM

- ✓ Συστήματα Πολλαπλών Πρακτόρων και Προσομοιώσεων
- ✓ Εικονικοί Πράκτορες και Ψηφιακά Δίδυμα

IV) Λογικός Προγραμματισμός - Λογική Συλλογιστική

- ✓ Συμπερασματικός
- ✓ Επαγωγικός
- ✓ Απαγωγικός
- ✓ Πιθανοτικός ή Στατιστικός
- ✓ Καταρρητικός Λογικός Προγραμματισμός – ΚΛΠ ή Destructive Logic Programming - DeLP

2.1.4 Συνδεδετική Τεχνητή Νοημοσύνη ή Τεχνητά Νευρωνικά Δίκτυα

Πρόκειται για προσεγγίσεις βασισμένες σε πιθανότητες, δεδομένα και στατιστική. Η ιδέα πίσω από την Συνδεδετική ΤΝ, που συχνά αναφέρεται ως Τεχνητά Νευρωνικά Δίκτυα, βασίζεται στο δίκτυο διασυνδεδεμένων νευρώνων του εγκεφάλου. Στην προσέγγιση ΤΝΔ, οι έννοιες σχηματίζονται με απλά μεμονωμένα στοιχεία επεξεργασίας, μονάδες ή κόμβους, διασυνδεδεμένα σε ένα δίκτυο. Το μέγεθος του νευρωνικού δικτύου μπορεί να είναι μεγάλο με πολλαπλά επίπεδα τα οποία συχνά αναφέρονται ως Βαθιά Μάθηση, η οποία επιτρέπει την παράλληλη επεξεργασία πληροφοριών μέσω των διασυνδεδεμένων κόμβων.

Η επεξεργασία πραγματοποιείται με τη διάδοση των ενεργοποιήσεων μέσω των διασυνδεδεμένων κόμβων. Η ικανότητα επεξεργασίας του δικτύου αποθηκεύεται στις διακομβικές δυνάμεις σύνδεσης, ή βάρη, που λαμβάνονται από μια διαδικασία προσαρμογής ή μάθησης από ένα σύνολο προτύπων εκπαίδευσης. Αυτά τα βάρη μεταφέρουν τις αριθμητικές αγωγιμότητες των επιδράσεων που έχει ένας κόμβος σε έναν

άλλον και επιτρέπουν τη διάδοση των ενεργοποιήσεων, με παρόμοιο τρόπο που οι συνδεδεμένοι νευρώνες επηρεάζουν ο ένας τον άλλον μέσω των συνάψεων. Η τιμή ενεργοποίησης για κάθε κόμβο μεταβάλλεται ανάλογα με τη δραστηριότητα των κόμβων με τους οποίους συνδέεται. Το τμήμα μάθησης, το οποίο είναι αποτέλεσμα ενός σήματος, δίνει μια συνεχή έξοδο, υποδεικνύοντας το βαθμό διαφοράς μεταξύ του αποτελέσματος από το δίκτυο και του επιδιωκόμενου αποτελέσματος, ή μια διακριτή έξοδο, υποδεικνύοντας επιτυχία - αποτυχία.

Κατά την τελευταία δεκαετία, οι προσεγγίσεις της Βαθιάς Μάθησης έχουν γίνει η προτιμητέα λύση για την όραση, την ομιλία, τη γλώσσα και άλλα προβλήματα. Είναι εγγενώς ευέλικτες, καθώς περιλαμβάνουν προσαρμοστικούς μηχανισμούς μέσω της μάθησης, με αποτέλεσμα να πετυχαίνουν καλύτερες επιδόσεις από τις συμβολικές. Ωστόσο, το μειονέκτημα είναι ότι είναι δύσκολο να κατανοήσουμε πώς επιτυγχάνεται η λύση, δηλαδή πώς λαμβάνονται οι αποφάσεις, γεγονός που μειώνει την εξηγησιμότητα και τη διαφάνεια των μοντέλων που χρησιμοποιούν αυτή την προσέγγιση.

Ανάλογα με την αρχιτεκτονική του νευρωνικού δικτύου και τον τομέα του προβλήματος, η Βαθιά Μάθηση μπορεί να έχει διάφορες κατηγορίες συμπεριλαμβανομένων της Επιβλεπόμενης, της Μη-επιβλεπόμενης και της Ημι-επιβλεπόμενης Μάθησης.

2.1.5 Εξελικτική Τεχνητή Νοημοσύνη

Οι εξελικτικοί αλγόριθμοι είναι εμπνευσμένοι από τη βιολογική εξέλιξη και εφαρμόζονται σε μια ποικιλία εργασιών μάθησης και άλλων προβλημάτων βελτιστοποίησης. Η λύση του προβλήματος αναζητείται ξεκινώντας από έναν πληθυσμό που εξελίσσεται με την εφαρμογή μηχανισμών που σχετίζονται με την εξέλιξη, την αναπαραγωγή, την μετάλλαξη, και την επιλογή. Σε κάθε επανάληψη του αλγορίθμου δημιουργείται ένας νέος πληθυσμός που αξιολογείται σε σχέση με το καθορισμένο μέτρο καταλληλότητας. Μερικοί από τους αλγορίθμους που ανήκουν σε αυτή την κατηγορία είναι:

- ✓ Γενετικοί Αλγόριθμοι – ΓΑ ή Generative Algorithms - GA
- ✓ Διαφορική Εξέλιξη – ΔΕ ή Differential Evolvment - DE
- ✓ Αλγόριθμος Αναζήτησης Αρμονίας – AAA ή Harmony Search Algorithms - HSA
- ✓ Αλγόριθμος Βελτιστοποίησης Κοινωνικής Μηχανικής – ABKM ή Social Engineering Optimization Algorithms - SEOA

2.1.6 Μπεϋζιανή Τεχνητή Νοημοσύνη

Χρησιμοποιεί Μπεϋζιανές μεθόδους συμπερασματολογίας για την ανάπτυξη μιας αρχιτεκτονικής λογισμικού TN. Γίνεται η εκ των προτέρων υπόθεση ότι κάθε εισερχόμενη πληροφορία περιέχει μια τιμή αβεβαιότητας, οπότε ο υπολογισμός των πιθανοτήτων προκειμένου να είναι σωστή ή να συμβεί και η ενημέρωσή της με βάση το θεώρημα Μπεϋζ και τα πρόσθετα δεδομένα, οδηγεί στο πιο πιθανό αποτέλεσμα. Μέθοδοι που χρησιμοποιούνται για την επίλυση προβλημάτων πιθανολογικής συμπερασματολογίας περιλαμβάνουν τα Κρυφά Μοντέλα Μαρκόβ - KMM ή Hidden Markov Models - HMM και τα Γραφικά Μοντέλα.

2.1.7 Αναλογική Τεχνητή Νοημοσύνη

Οι μέθοδοι επικεντρώνονται στην ανάλυση μοτίβων, με την αντιστοίχιση σχετικών δεδομένων και τον εντοπισμό ομοιοτήτων. Οι νέες ή παλιές είσοδοι στο σύστημα αντιστοιχίζονται εκ των προτέρων σε μία από την κλάση-κατηγορία οντοτήτων που δημιουργούνται με βάση τις ομοιότητες μεταξύ των δεδομένων. Οι έξοδοι προβλέπονται με βάση γενικευμένες ομοιότητες με τα μέλη μιας κλάσης-κατηγορίας. Τα πιο γνωστά αναλογικά μοντέλα είναι, οι Μηχανές Διανυσμάτων Υποστήριξης - ΜΔΥ ή Support Vector Machines - SVM και οι αλγόριθμοι k-Κοντινότερων Γειτόνων - k-ΚΓ ή k-Nearest Neighbors - k-NN. Χρησιμοποιούνται ευρέως στα συστήματα προτάσεων.

2.1.8 Μηχανική Μάθηση: Πρακτική Εφαρμογή της Τεχνητής Νοημοσύνης

Μηχανική Μάθηση είναι η επιστημονική μελέτη των αλγορίθμων υπολογιστών που βελτιώνονται αυτόματα μέσω εμπειρίας. Η MM μπορεί να θεωρηθεί ως η πρακτική εφαρμογή της TN που συνδέεται με ορισμένα από τα θεωρητικά μοντέλα τα οποία παρουσιάζονται παρακάτω. Οι αλγόριθμοι MM δημιουργούν ένα μοντέλο με βάση δεδομένα εκπαίδευσης, προκειμένου να κάνουν προβλέψεις ή να λάβουν αποφάσεις χωρίς να είναι ρητά προγραμματισμένοι να το κάνουν.

Πιο συγκεκριμένα, οι ακόλουθες οικογένειες αλγορίθμων μπορούν να θεωρηθούν ως προσεγγίσεις MM. Η παρουσίαση δεν ακολουθεί μια ιεραρχική προσέγγιση αλλά μια

απαρίθμηση υφιστάμενων ομαδοποιήσεων που μπορεί να παρουσιάζουν αλληλοεπικαλύψεις μεταξύ τους^[1,2,3,4].

2.1.9 Επιβλεπόμενη Μάθηση

Επιβλεπόμενη Μάθηση - ΕΠ ή Supervised Learning – SL: Αντιστοιχίζει τις τιμές εισόδου στις τιμές εξόδου με βάση παραδείγματα-ζεύγη με ετικέτες εισόδου-εξόδου. Η Επιβλεπόμενη Μάθηση χρειάζεται σημαντικές ποσότητες δεδομένων με ετικέτες, μια διαδικασία που συχνά γίνεται από ανθρώπους.

I) Παλινδρόμηση

- ✓ Γραμμική
- ✓ Μη-γραμμική
- ✓ Παλινδρόμηση Διαδικασίας Γκάους - ΠΔΓ ή Gauss Process Regression - GPR
- ✓ Λογιστική
- ✓ Πρόμπτια Παλινδρόμηση – ΠΠ ή Probit Regression - PR

II) Πιθανοτικά Γραφικά Μοντέλα – ΠΓΜ ή Probabilistic Graphical Models - PLM

- ✓ Κρυφά Μοντέλα Μαρκόβ - KMM
- ✓ Μπεϋζιανά Δίκτυα ή Δίκτυα Πίστης ή Αιτιώδη Δίκτυα
 - Αφελής Μπέϋζ
 - Ημι-αφελής Μπέϋζ
 - Καθαρός Μπέϋζ
- ✓ Περιορισμένη Μηχανή Μπόλτζμαν - ΠΜΜ ή Restricted Boltzmann Machine - RBM και Βαθιά Μηχανή Μπόλτζμαν - ΒΜΜ ή Deep Boltzmann Machine - DBM

III) Τεχνητά Νευρωνικά Δίκτυα - ΤΝΔ και Βαθιά Μάθηση - ΒΜ

- ✓ Νευρωνικά Δίκτυα με έως και τέσσερα επίπεδα
- ✓ Βαθιά Νευρωνικά Δίκτυα - ΒΝΔ
- ✓ Βαθείς Αυτο-Κωδικοποιητές – ΒΑΚ ή Deep Auto-Decoders - DAE
- ✓ Βαθιά Μηχανή Μπόλτζμαν
- ✓ Βαθιά Επαναληπτικά Νευρωνικά Δίκτυα – ΒΕΝΔ ή Deep Recursive Neural Networks - DRNN
- ✓ Βαθιά Δίκτυα Πίστης – ΒΔΠ ή Deep Belief Networks - DBN
- ✓ Επαναλαμβανόμενα Νευρωνικά Δίκτυα – ΕΝΔ ή Recursive Neural Networks - RNN
- ✓ Συνελκτικά Νευρωνικά Δίκτυα – ΣΝΝ ή Convolutional Neural Networks - CNN

IV) Δέντρα Αποφάσεων

- ✓ Ταξινόμηση ή Δυαδικά Δέντρα - Διακριτά

- ✓ Δέντρα Παλινδρόμησης - Συνεχή
- ✓ Τυχαία Δέντρα Αποφάσεων
- ✓ Τυχαία Δάση
- ✓ Δέντρα Απόφασης Σάκων
- ✓ Ενισχυμένα Δέντρα Απόφασης
- ✓ Συμπεριφορικά Δέντρα

V) Ταξινομητές

- ✓ Ανάλυση Διακριτότητας - Γενίκευση της γραμμικής διακριτότητας του Φίσερ - Fisher
- ✓ Μηχανή Διανυσμάτων Υποστήριξης - ΜΔΥ
- ✓ κ-Κοντινότερος Γείτονας – κ-ΚΓ

VI) Εκμάθηση Συνόλου

- ✓ Τόνωση - Boosting

2.1.10 Ανεπίβλεπτη Μάθηση

Ανεπίβλεπτη Μάθηση - AM ή Unsupervised Learning – UL: Βοηθά στην εύρεση προηγούμενων άγνωστων μοτίβων σε σύνολα δεδομένων χωρίς προϋπάρχουσες ετικέτες. Ο στόχος είναι η ανακάλυψη της υποκείμενης δομής των δεδομένων, για παράδειγμα, με την ομαδοποίηση παρόμοιων στοιχείων για τον σχηματισμό συστάδων.

I) Ομαδοποίηση

- ✓ κ-Μέσοι – k-Means
- ✓ Ιεραρχική Συσσωρευτική Ομαδοποίηση – ΙΣΟ ή Hierarchical Accumulative Clustering - HAC
- ✓ Μεγιστοποίηση Προσδοκίας – EM ή Expectation Maximization - EM
- ✓ Χωρική Ομαδοποίηση Βάσει την Πυκνότητα Εφαρμογών με Θόρυβο – ΧΟΒΠΕΘ ή Density Based Spatial Clustering on Applications with Noise - DBSCAN
- ✓ Παραγοντοποίηση Πινάκων π.χ. Ομαδοποίηση, Μείωση Διαστάσεων

II) Μείωση Διαστάσεων - Συμπύεση Δεδομένων

- ✓ Ανάλυση Κύριων Συνιστωσών – ΑΚΣ ή Principal Components Analysis - PCA
- ✓ Γραμμική Διακριτική Ανάλυση
- ✓ Παραγοντοποίηση Πινάκων π.χ. Ομαδοποίηση, Μείωση Διαστάσεων

III) Γενετικά Αντιφατικά Δίκτυα – ΓΑΔ ή Generative Adversarial Networks - GAN

IV) Αυτο-Κωδικοποιητές – ΑΚ ή Auto-Decoders - AD

2.1.11 Ημι-Επιβλεπόμενη Μάθηση

Ημι-Επιβλεπόμενη Μάθηση - HEM ή Semi-Unsupervised Learning – SUL: Θεωρείται μια κατηγορία μεταξύ της επιβλεπόμενης και της μη επιβλεπόμενης μάθησης στην οποία τα δεδομένα χωρίζονται σε δεδομένα με ετικέτες και χωρίς.

2.1.12 Ενισχυτική Μάθηση

Ενισχυτική Μάθηση - EM ή Reinforcement Learning – RL: Διερευνά τον τρόπο με τον οποίο οι πράκτορες¹ αναλαμβάνουν δράσεις σε ένα περιβάλλον προκειμένου να μεγιστοποιήσουν μια ανταμοιβή.

I) Βαθιά Ενισχυτική Μάθηση

II) Αντίστροφη Ενισχυτική Μάθηση

III) Κιου-Αλγοριθμική μάθηση – Q-learning Algorithm²

2.1.13 Αντιθετική Μάθηση

Αντιθετική Μάθηση - AM ή Adversarial Learning – AL: Είναι μια από τις πιο δημοφιλείς μεθόδους για τη βελτίωση της αντιθετικής ευρωστίας των μοντέλων νευρωνικών δικτύων. Ο στόχος της αντιθετικής άμυνας είναι η εκπαίδευση ενός μοντέλου για την ελαχιστοποίηση του εύρωστου σφάλματος έναντι συγχύσεων χειρότερης περίπτωσης εντός μιας ορισμένης απόστασης. Αυτός ο στόχος εκπαίδευσης μπορεί να γραφεί φυσικά ως μια βελτιστοποίηση δύο επιπέδων, όπου ο επιτιθέμενος στοχεύει να λάβει τη χειρότερη σύγκυση στον εσωτερικό βρόχο και ο αμυνόμενος στοχεύει να εκπαιδεύσει το μοντέλο που ελαχιστοποιεί το σφάλμα με το συγχυσμένο παράδειγμα^[5].

¹ TN Πράκτορας - AI Agent: είναι ένα πρόγραμμα λογισμικού το οποίο αλληλεπιδρά με το περιβάλλον του, συλλέγει δεδομένα και τα χρησιμοποιεί για να εκτελέσει αυτο-καθορισμένες ενέργειες προκειμένου να επιτύχει προκαθορισμένους στόχους.

² Κιου-Αλγοριθμική Μάθηση - Q-Learning Algorithm: Είναι ένας πολύ αποδοτικός τρόπος για έναν πράκτορα να μάθει πως λειτουργεί το περιβάλλον.

2.1.14 Μεταφορική Μάθηση

Μεταφορική Μάθηση - MM ή Transfer Learning – TL: Επαναχρησιμοποιεί προϋπάρχοντα μοντέλα για την επίλυση τρεχουσών προκλήσεων. Η Μάθηση Μεταφοράς είναι μια τεχνική που χρησιμοποιείται για την από κοινού εκπαίδευση μοντέλων. Οι έννοιες των προϋπαρχόντων δεδομένων εκπαίδευσης αξιοποιούνται για την ενίσχυση της απόδοσης της τρέχουσας πρόκλησης, ώστε η λύση να μην χρειάζεται να αναπτυχθεί από το μηδέν^{6]}.

2.1.15 Γνωστικές Ικανότητες που Αναπαράγονται από την Τεχνητή Νοημοσύνη

Η παρακάτω προσέγγιση για την ταξινόμηση της TN ομαδοποιεί τις κύριες γνωστικές ικανότητες που η TN αυτοματοποιεί ή αναπαράγει.

I) Διαδικασίες Μνήμης - ΔΜ ή Memory Processes - MP: ικανότητα δημιουργίας, επεξεργασίας, αποθήκευσης και ανάκτησης πληροφοριών.

II) Αισθητικοκινητική Αλληλεπίδραση – AA ή Sensorimotor Interaction - SI³: αντίληψη των πραγμάτων, αναγνώριση προτύπων, χειρισμός των πραγμάτων σε φυσικά ή εικονικά περιβάλλοντα.

III) Οπτική Επεξεργασία - ΟΕ ή Visual Processing - VP: επεξεργασία και αναγνώριση οπτικών πληροφοριών, τόσο αντικειμένων όσο και συμβόλων, σε εικόνες και βίντεο σε διαφορετικά επίπεδα θορύβου, γωνιών και μετασχηματισμών.

IV) Ακουστική Επεξεργασία - ΑΕ ή Auditory Processing - AP: επεξεργασία ακουστικών πληροφοριών, δηλαδή της ομιλίας, εξαιρουμένης της πλήρους κατανόησης των προτάσεων, και της μουσικής, και της υποκειμενικής αντίληψης της αρμονίας, σε διαφορετικές συχνότητες και επίπεδα θορύβου.

V) Προσοχή και Αναζήτηση - ΠΑ ή Attention and Search - AS: προσοχή είναι η ικανότητα εστίασης σε μόνο ένα μέρος της πληροφορίας εισόδου που πληροί ορισμένα κριτήρια και η αδιαφορία για άσχετα αντικείμενα, μοτίβα κ.λπ. Αναζήτηση είναι η ικανότητα εύρεσης των μερών της πληροφορίας που πληρούν ορισμένα κριτήρια.

³ Στην παρούσα σελίδα η συντομογραφία SI σημαίνει Sensorimotor Interaction

VI) Σχεδιασμός και Δράση - ΣΔ ή Planning and Action - PA: ικανότητα αναμονής των συνεπειών μιας σειράς ενεργειών, με χρονική ή διαφορετική εξάρτηση, κατανόηση του λόγου της πραγματοποίησής τους, και προσδιορισμός του καλύτερου σχεδίου δράσης σε μια δεδομένη περίπτωση.

VII) Κατανόηση και Έκφραση - ΚΕ ή Comprehension and Expression - CE: επεξεργασία, κατανόηση και παραγωγή φυσικής γλώσσας και σημασιολογικών αναπαραστάσεων, μέσω διαφόρων μορφών: κείμενο, ήχος, σχέδια κ.λπ.

VIII) Επικοινωνία ή Communication: ικανότητα ανταλλαγής και κατανόησης πληροφοριών μεταξύ ομότιμων, σε επίσημα και ανεπίσημα πλαίσια.

IX) Συναισθήματα και Αυτο-Έλεγχος - ΣΑΕ ή Emotions and Self-Control - ESC: κατανόηση της συναισθηματικής κατάστασης άλλων πρακτόρων, του αντίκτυπου αυτών των συναισθημάτων στη συμπεριφορά τους, καθώς και την αναγνώριση και τον έλεγχο της δικής τους συναισθηματικής κατάστασης, εξαιρούμενης της ικανότητας μοντελοποίησης του νου και της κοινωνικής αλληλεπίδρασης.

X) Πλοήγηση ή Navigation: ικανότητα μετακίνησης του εαυτού ή αντικειμένων μεταξύ θέσεων, μέσω κατάλληλων και ασφαλών μονοπατιών, λαμβάνοντας υπόψιν άλλα αντικείμενα ή πράκτορες που βρίσκονται σε σχετική εγγύτητα. Όπως επίσης η ικανότητα προσαρμογής των μονοπατιών για την επίτευξη μιας θέσης.

XI) Εννοιολογική Μάθηση και Αφαίρεση - ΕΜΑ ή Conceptualization Learning and Abstraction - CLS: η ικανότητα σχηματισμού μιας ιδέας ή μιας λειτουργίας με βάση γενικεύσεις προηγούμενων παραδειγμάτων, η λήψη οδηγιών και η λειτουργία με βάση επιδείξεις και τέλος, η συλλογή γνώσεων σε διάφορα επίπεδα αφαίρεσης. Τα προαναφερθέντα μέρη της ικανότητας στοχεύουν στην ολοκλήρωση μιας εργασίας, χωρίς η εργασία να περιορίζεται από εκμάθηση παρελθοντικών ιδεών, λειτουργιών ή αφαιρέσεων.

XII) Ποσοτικός και Λογικός Συλλογισμός - ΠΛΣ ή Quantitative and Logical Reasoning - QLR: η ικανότητα αναπαράστασης ποσοτικών ή/και λογικών πληροφοριών που είναι απαραίτητες για το έργο που έχει τεθεί και η εξαγωγή νέων πληροφοριών που μπορούν να επιλύσουν το έργο π.χ. πιθανότητες, αντιπαραδείγματα, άλλοι τύποι αναλυτικού

συλλογισμού. Η ικανότητα αυτή εκτείνεται πέρα από την παρακολούθηση ενός συνδυασμού απλών κανόνων ή οδηγιών με την επεξεργασία εσωτερικών συμβόλων ή αριθμούς, όπως οι γλώσσες προγραμματισμού, οι νευρώνες ενός δικτύου κ.λπ.

XIII) Μοντελοποίηση του Νου και Κοινωνική Αλληλεπίδραση - MNKA ή Mind Modelling and Social Interaction - MMSI: δημιουργία μοντέλων άλλων πρακτόρων που κατανοούν τις προθέσεις, τις επιθυμίες και τις πεποιθήσεις τους, με στόχο την πρόβλεψη των ενεργειών και των συμφερόντων τους.

XIV) Μεταγνώση και Αξιολόγηση Εμπιστοσύνης - MAE ή Metacognition and Confidence Assessment - MCA: ικανότητα αυτοαξιολόγησης της πιθανότητας και της εμπιστοσύνης των ενεργειών του για την επιτυχία, την προσπάθεια και τον κίνδυνο, που υπερβαίνει τα καθήκοντα σχεδιασμού ή την απλή επιλογή της δράσης με τη μεγαλύτερη πιθανότητα^[1,2,3,4].

2.1.16 Λειτουργίες που εκτελούνται από την Τεχνητή Νοημοσύνη

Η Τεχνητή Νοημοσύνη μπορεί επίσης να κατηγοριοποιηθεί με βάση την λειτουργία που εκτελούν οι αλγόριθμοι. Οι εφαρμογές ΤΝ χρησιμοποιούνται για την επίλυση και την βελτιστοποίηση, την αυτοματοποίηση, την αντίληψη και την επικοινωνία, τον πειραματισμό και την δημιουργία.

I) Επίλυση και Βελτιστοποίηση

- ✓ Βελτιστοποίηση Απόφασης: πρόβλεψη ζήτησης, κατανομή εργασιών, συστήματα ειδικών, αποθήκες, εφοδιαστική αλυσίδα, εργατικό δυναμικό κ.λπ.
- ✓ Διαδικασίες ή Βελτιστοποίηση Εξοπλισμού: πχ. Πρόβλεψη ανάγκης συντήρησης
- ✓ Προτάσεις και Μηχανές Εξατομίκευσης: μέσω αλγορίθμων αντιστοίχισης και ανάκτηση πληροφορίας, παρέχεται εξατομικευμένη προώθηση, προτάσεις δράσης κ.λπ.

II) Αυτοματισμός

- ✓ Αυτοματισμός Διαδικασιών: αυτοματοποίηση αποθηκών, αυτόματες συναλλαγές, αυτοματοποίηση διαδικασιών ρομπότ κ.λπ.

- ✓ Αυτόματα Συστήματα: αυτόνομα ρομπότ, αυτόνομα οχήματα, μη επανδρωμένα αεροσκάφη κ.λπ.

III) Αντίληψη και Επικοινωνία: συμπεριλαμβανομένου κειμένου, ήχου, βίντεο και εικόνας για επικοινωνία μεταξύ ανθρώπου και μηχανής, ή Μηχανή-σε-Μηχανή – ΜσΜ ή Machine-to-Machine - M2M, ή Μηχανή-σε-Σύννεφο – ΜσΣ ή Machine-to-Cloud - M2C, μερικά εκ των οποίων χρησιμοποιούνται στην αυτοματοποίηση εφαρμογών.

- ✓ Φυσική Γλώσσα Επεξεργασίας – ΦΓΕ ή Natural Language Processing - NLP: αναγνώριση ομιλίας και παραγωγής, αυτοματοποιημένη συνομιλία, φωνητικοί βοηθοί, αυτοματισμοί σύνοψης κειμένου, εξαγωγή όρων όπως μηχανική μετάφραση.
- ✓ Όραση Υπολογιστών: απόκτηση εικόνας, προεπεξεργασία, ανίχνευση, τμηματοποίηση, αναγνώριση εικόνας, επεξεργασία υψηλού επιπέδου, εξαγωγή χαρακτηριστικών, αναγνώριση προσώπου, οπτική διάγνωση κ.λπ.
- ✓ Ανάλυση Συναισθήματος: εντοπισμός συναισθημάτων και αναγνώριση συμπεριφορών, εξαγωγή, ανάλυση συναισθημάτων και συμπεριφορών, κατηγοριοποίηση από κείμενο - συμπεριλαμβανομένης της δραστηριότητας στα μέσα κοινωνικής δικτύωσης, ήχο, βίντεο ή βιομετρικούς αισθητήρες κ.λπ.
- ✓ Ανίχνευση Ανωμαλιών: ανίχνευση απάτης, ανίχνευση επιθέσεων στον κυβερνοχώρο όπως δηλητηρίαση, ανίχνευση ελαττωμάτων κ.λπ.

IV) Πειραματισμός και Δημιουργικές Δραστηριότητες

- ✓ Δημιουργία Δεδομένων: αυτόματη σύνθεση εικόνων, αυτόματη δημιουργία σκηνών κ.λπ.
- ✓ Τεχνητός Πειραματισμός και Ανακάλυψη: εικονικός πράκτορας, ψηφιακός δίδυμος, εικονικό πρωτότυπο, φαρμακευτική, χημική, βιολογική ανακάλυψη, συστήματα πολλαπλών πρακτόρων, προσομοιώσεις κ.λπ.
- ✓ Καλλιτεχνική Δημιουργία και Σχεδιασμός: τεχνητή σύνθεση μουσικής, ζωγραφικής, κινούμενα σχέδια χαρακτήρων κ.λπ.

2.1.17 Τεχνητή Νοημοσύνη στις Επιχειρήσεις

Στον πραγματικό κόσμο, μπορούμε να βρούμε πολλαπλές εφαρμογές ΤΝ που χρησιμοποιούνται από εταιρείες, κυβερνήσεις και την κοινωνία σε γενικότερο πλαίσιο. Επειδή δεν είναι δυνατό να παραχθεί μια πλήρως επικαιροποιημένη λίστα όλων των εφαρμογών ΤΝ που υπάρχουν στο εμπόριο, καθώς η τεχνολογία συνεχώς εξελίσσεται, παρέχεται μια ενδεικτική λίστα εφαρμογών ΤΝ, ανά λειτουργία της επιχείρησης.

I) Πωλήσεις, Προώθηση Προϊόντων

- ✓ Κατηγοριοποίηση πελατών, εξατομικευμένες προσφορές προώθησης και προτάσεις, εκπτώσεις με βάση τον πελάτη, στοχευμένη ανάλυση αγοράς, πρόληψη απώλειας πελατείας. Πορίσματα τα οποία έχουν εξαχθεί βασισμένα στην Μηχανική Μάθηση
- ✓ Αυτόματοι συνομιλητές βασισμένοι στην Φυσική Γλώσσα Επεξεργασίας για εξυπηρέτηση πελατών

II) Διαδικασία Παραγωγής

- ✓ Πρόβλεψη συντήρησης
- ✓ Ταξινόμηση προϊόντων ή ανίχνευση ελαττωμάτων με βάση την όραση υπολογιστή
- ✓ Εργασίες συναρμολόγησης από αυτόνομα ρομπότ
- ✓ Αυτόνομα μη-επανδρωμένα αεροσκάφη για την επιτήρηση της παραγωγής, την ασφάλεια ή την επιθεώρηση εργασιών

III) Οργάνωση των Διαδικασιών Διοίκησης Επιχειρήσεων

- ✓ Επιχειρηματικοί εικονικοί βοηθοί
- ✓ Μετατροπή φωνής σε κείμενο με βάση την αναγνώριση ομιλίας
- ✓ Αυτοματοποιημένος σχεδιασμός ή προγραμματισμός
- ✓ Μηχανική μετάφραση

IV) Διοίκηση Εταιρειών

- ✓ Αναλύσεις δεδομένων για τη λήψη επενδυτικών ή άλλων αποφάσεων
- ✓ Προβλέψεις πωλήσεων ή επιχειρήσεων
- ✓ Εκτίμηση κινδύνου

V) Εφοδιαστική Αλυσίδα

- ✓ Πρόβλεψη ζήτησης
- ✓ Αυτόνομα ρομπότ για λύσεις «πικ-και-πακ» - pick-and-pack - επιλογή των παραγγελιών του πελάτη και τοποθέτησή τους απευθείας σε κουτί αποστολής, αποστολή δεμάτων και εντοπισμός
- ✓ Βελτιστοποίηση διαδρομών
- ✓ Μη επανδρωμένα σκάφη για την παράδοση δεμάτων

VI) Τεχνολογία Πληροφοριών και Επικοινωνιών – ΤΠΕ ή Information and Communication Technology - ICT

- ✓ Πιστοποίηση ταυτότητας χρηστών ΤΠΕ με αναγνώριση προσώπου με βάση την όραση υπολογιστή
- ✓ Ανίχνευση και πρόληψη επιθέσεων στον κυβερνοχώρο

VII) Διαχείριση Ανθρώπινου Δυναμικού και Προσλήψεις

- ✓ Προεπιλογή υποψηφίων, αυτοματοποίηση προσλήψεων
- ✓ Υποστήριξη διαχείρισης ανθρώπινου δυναμικού ή προσλήψεων
- ✓ Προφίλ εργαζομένων, ανάλυση επιδόσεων
- ✓ Αυτοματοποίηση εργασιών με ρομπότ ανθρώπινου δυναμικού και ηλεκτρονικό γραφείο υποστήριξης εργαζομένων
- ✓ Ανάλυση συναισθήματος των εργαζομένων και της ομαδικής συμπεριφοράς για την πρόβλεψη του κύκλου εργασιών των εργαζομένων

VIII) Διαδικασίες Έρευνας και Ανάπτυξης

- ✓ Εικονική πρωτοτυποποίηση προϊόντων
- ✓ Βελτιστοποίηση των χαρακτηριστικών του προϊόντος

IX) Υποδομές και Διαχείριση Τεχνολογίας

- ✓ Έξυπνα κτίρια
- ✓ Αναλυτικά εργαλεία για βελτιστοποίηση και προγραμματισμό επενδύσεων

2.1.18 Τεχνητή Νοημοσύνη ανά Κλάδο της Οικονομίας

I) Γεωργία, Δασοκομία και Αλιεία

- ✓ Γεωργικά ρομπότ, παρακολούθηση καλλιεργειών και εδάφους

II) Πετρέλαιο και Φυσικό Αέριο

- ✓ Γεωτρήσεις ακριβείας, προληπτική συντήρηση, πρόληψη και ανίχνευση βλαβών

III) Παραγωγή

- ✓ Προληπτική συντήρηση, πρόληψη και ανίχνευση αποτυχίας, αυτοματοποιημένα και συνδεδεμένα οχήματα, αυτοματοποίηση διαδικασιών, βελτιστοποίηση διαδικασιών

IV) Παροχή Ηλεκτρικής Ενέργειας και Φυσικού Αερίου

- ✓ Έξυπνη αποθήκευση ενέργειας, αυτόνομη διαχείριση δικτύου, πρόληψη και ανίχνευση βλαβών, έξυπνοι μετρητές, αυτοματοποίηση διαδικασιών

V) Παροχή Νερού, Διαχείριση αποβλήτων

- ✓ Έξυπνοι μετρητές, αυτοματοποίηση διαδικασιών, βελτιστοποίηση διαδικασιών

VI) Εμπόριο, Λιανικό Εμπόριο

- ✓ Εξατομικευμένη προώθηση, σύσταση προϊόντων, βελτιστοποίηση τιμών, αυτόματη διαχείριση αποθεμάτων, ανάλυση αγοράς

VII) Μεταφορά και Αποθήκευση

- ✓ Βελτιστοποίηση αποθήκευσης, βελτιστοποίηση διαδρομής, αυτόνομη λειτουργία τρένων, αυτόνομη κινητικότητα κατά παραγγελία

VIII) Διανομή και Τρόφιμα

- ✓ Βελτιστοποίηση επιχειρηματικών διαδικασιών μέσω διαδικτυακών υπηρεσιών

IX) Πληροφορίες και Επικοινωνία

- ✓ Ανίχνευση ανωμαλιών, πρόβλεψη, βελτιστοποίηση συστημάτων, αποδοτικότητα δικτύων, εφαρμογές κυβερνοασφάλειας, όπως φιλτράρισμα ανεπιθύμητης αλληλογραφίας και προστασία δεδομένων

X) Χρηματοοικονομικά και Ασφάλειες

- ✓ Προβλέψεις, Αυτοματοποίηση Ρομποτικών Διαδικασιών, Συναλλαγές Υψηλής Συχνότητας

XI) Ακίνητα

- ✓ Ανάλυση συναισθήματος, μηχανές προτάσεων, προβλέψεις

XII) Εκπαίδευση

- ✓ Αυτόματη δημιουργία εξετάσεων, ηλεκτρονική μάθηση, έγκαιρη ανίχνευση δυσλεξίας και ελλειμματικής προσοχής

XIII) Ανθρώπινη Υγεία, Κοινωνική Εργασία

- ✓ Ηλεκτρονικά ραντεβού, προσωπικοί σύντροφοι/ανθρωποειδή ρομπότ, διάγνωση με βάση την εικόνα, πρόβλεψη, συσκευές μέτρησης, ιατρικά ρομπότ για χειρουργικές επεμβάσεις

XIV) Δραστηριότητες Αναψυχής

- ✓ Δημιουργικές δραστηριότητες, εικονική και επαυξημένη πραγματικότητα, αθλητικά εξαρτήματα

2.1.19 Ταξονομία Επίβλεψης Ομάδας Εμπειρογνομόνων Υψηλού Επιπέδου Τεχνητής Νοημοσύνης

Η παρακάτω ταξονομία Επίβλεψης TN δεν προορίζεται να αποτελέσει μια αυστηρή ή εξαντλητική ταξινόμηση, αλλά μια ολοκληρωμένη συλλογή των περιοχών που αντιπροσωπεύουν την TN από τις τρεις προοπτικές: την πολιτική, την έρευνα και την βιομηχανία. Περιλαμβάνει την πλειονότητα των προαναφερθέντων περιπτώσεων από εφαρμογές TN και αλγορίθμων, σε μια μορφή που επιτρέπει στους υποτομείς να είναι συναφή και όχι διαχωρισμένα υποσύνολα της TN. Επιτρέπει στον αναγνώστη να περιηγηθεί σε κάθε υποτομέα και να εντοπίσει μια τεχνολογία ή έναν επιχειρηματικό σκοπό, με βάση τον τρόπο λειτουργίας της εκάστοτε περίπτωσης^[1,2,3,4].

		Ταξονομία TN	
		Κατηγορία TN	Υποκατηγορία TN
Κύρια	Συλλογιστική		Αναπαράσταση Γνώσης
			Αυτοματοποιημένη Συλλογιστική
			Συλλογιστική Κοινής Λογικής
	Σχεδίαση		Σχεδίαση και Δρομολόγηση
			Αναζήτηση
			Βελτιστοποίηση
	Μάθηση		Μηχανική Μάθηση
Επικοινωνία		Φυσική Γλώσσα Επεξεργασίας	
Αντίληψη		Όραση Υπολογιστών	
		Επεξεργασία Ήχου	
Εγκάρσια	Ενσωμάτωση και Αλληλεπίδραση		Συστήματα Πολλαπλών Παραγόντων
			Ρομποτική κι Αυτοματοποίηση
			Συνδεδεμένα και Αυτόνομα Οχήματα
	Υπηρεσίες		Υπηρεσίες TN
Ηθική και Φιλοσοφία		Ηθικότητα TN	
		Φιλοσοφία TN	

Εικόνα 1: ΟΕΥΕ, Ταξονομία Τεχνητής Νοημοσύνης^[1]

2.2 Ηθικό και Νομικό Πλαίσιο Τεχνητής Νοημοσύνης

2.2.1 Ευρωπαϊκές Κατευθυντήριες Γραμμές για Αξιόπιστη Τεχνητή Νοημοσύνη

Σημαντικό ρόλο στην διαμόρφωση ενός ηθικού πλαισίου εφαρμογής της TN έχει η Ομάδα Εμπειρογνομώνων Υψηλού Επιπέδου για την Τεχνητή Νοημοσύνη της Ευρωπαϊκής Ένωσης - ΟΕΥΕ.

2.2.2 Αξιόπιστη Τεχνητή Νοημοσύνη

Η αξιοπιστία είναι προαπαιτούμενο για τους ανθρώπους και τις κοινωνίες για να αναπτύσσουν, να εφαρμόζουν και να χρησιμοποιούν συστήματα TN. Χωρίς τα συστήματα TN να είναι αποδεδειγμένα άξια εμπιστοσύνης και τους ανθρώπους να βρίσκονται από πίσω τους, οι ανεπιθύμητες συνέπειες ελλοχεύουν και η εφαρμογή των συστημάτων TN παρεμποδίζεται με αποτέλεσμα την επιβράδυνση της καινοτομίας και

κατ' επέκταση την απώλεια των οικονομικών και κοινωνικών οφελών από την χρήση τους. Στόχος είναι να επιτευχθεί η αξιόπιστη ΤΝ.

Η εμπιστοσύνη στην ανάπτυξη, στην εφαρμογή και στην χρήση συστημάτων ΤΝ δεν αφορά μόνο τις εγγενείς ιδιότητες της τεχνολογίας αυτής, αλλά και τα χαρακτηριστικά των κοινωνικοτεχνικών συστημάτων που περιέχουν εφαρμογές ΤΝ. Όπως και με τους προβληματισμούς απώλειας αξιοπιστίας στις αερομεταφορές, την πυρηνική ενέργεια ή την ασφάλεια τροφίμων, έτσι και τα συστήματα ΤΝ χρειάζεται να δημιουργούν αξιοπιστία στο ευρύτερο πλαίσιο λειτουργίας. Η επιδίωξη Αξιόπιστης ΤΝ δεν αφορά λοιπόν μόνο στην αξιοπιστία του ίδιου του συστήματος ΤΝ αλλά απαιτεί ολιστική και συστηματική προσέγγιση, η οποία περιλαμβάνει την αξιοπιστία όλων των φορέων και των διαδικασιών που αποτελούν μέρος του κοινωνικοτεχνικού πλαισίου του συστήματος καθ' όλη την διάρκεια του κύκλου ζωής του.

Η Αξιόπιστη ΤΝ αποτελείται από τρία στοιχεία τα οποία πρέπει να πληρούνται καθ' όλη την διάρκεια του κύκλου ζωής του συστήματος:

- I) Πρέπει να είναι νόμιμη, συμμορφούμενη με τους ισχύοντες νόμους και κανονισμούς.
- II) Πρέπει να είναι ηθική, εξασφαλίζοντας την τήρηση των ηθικών αρχών και αξιών.
- III) Πρέπει να είναι εύρωστη, τόσο από τεχνική όσο και από κοινωνική άποψη, καθώς ακόμα και συστήματα ΤΝ με καλές προθέσεις μπορεί να προκαλέσουν ακούσια βλάβη.

Κάθε ένα από αυτά τα συστατικά είναι απαραίτητο αλλά όχι επαρκές από μόνο του για την επίτευξη αξιόπιστης ΤΝ. Ιδανικά, και τα τρία λειτουργούν αρμονικά και επικαλύπτονται στην λειτουργία τους. Επί του πρακτέου, μπορεί να υπάρχουν τριβές μεταξύ αυτών των στοιχείων. Για παράδειγμα, κατά καιρούς το πεδίο εφαρμογής και το περιεχόμενο της υφιστάμενης νομοθεσίας μπορεί να μην συμβαδίζει με τους ηθικούς κανόνες. Είναι ατομική και συλλογική ευθύνη ως κοινωνία να εργαστούμε για την διασφάλιση και των τριών αυτών στοιχείων τα οποία συμβάλλουν στην αξιόπιστη Τεχνητή Νοημοσύνη.

Μια αξιόπιστη προσέγγιση είναι το κλειδί για την επίτευξη «υπεύθυνης ανταγωνιστικότητας», παρέχοντας τα θεμέλια πάνω στα οποία όλοι όσοι επηρεάζονται από τα συστήματα ΤΝ μπορούν να είναι σίγουροι ότι ο σχεδιασμός, η ανάπτυξη και η χρήση τους γίνεται με νόμιμο, ηθικό και αξιόπιστο τρόπο. Οι κατευθυντήριες γραμμές

αποσκοπούν στην προώθηση της υπεύθυνης και βιώσιμης καινοτομίας ΤΝ στην Ευρώπη. Επιδιώξή τους είναι να καταστεί η ηθική ως βασικός πυλώνας για την ανάπτυξης ΤΝ, η οποία αποσκοπεί στο όφελος, την ενδυνάμωση και την προστασία τόσο της ατομικής ανθρώπινης ευημερίας όσο και του κοινού καλού της κοινωνίας. Έτσι, οι πολίτες επωφελούνται πλήρως από τα οφέλη των συστημάτων ΤΝ, γνωρίζοντας ότι τα μέτρα είναι μέτρα για την προστασία τους από πιθανούς κινδύνους.

2.2.3 Ακροατήριο και Πεδίο Εφαρμογής

Οι παρούσες κατευθυντήριες γραμμές απευθύνονται σε όλους τους ενδιαφερομένους φορείς ΤΝ που σχεδιάζουν, αναπτύσσουν, εφαρμόζουν, χρησιμοποιούν πληροφοριακά συστήματα ή επηρεάζονται από την ΤΝ. Μερικοί τέτοιοι φορείς είναι εταιρείες, οργανισμοί, ερευνητές, δημόσιες υπηρεσίες, κυβερνήσεις, ιδρύματα, κοινωνικές οργανώσεις πολιτών, εργαζόμενοι και καταναλωτές. Οι φορείς που είναι αφοσιωμένοι στο να πετύχουν Αξιόπιστη ΤΝ, μπορούν να επιλέξουν εθελοντικά να εφαρμόσουν αυτές τις οδηγίες ώστε να μετουσιώσουν στην πράξη την αφοσίωσή τους. Εν συνεχεία, ακολουθεί η πλήρης ανάλυση της λίστας αξιολόγησης η οποία μπορεί να συμπληρώσει ή να ενσωματωθεί σε υπάρχουσες διαδικασίες αξιολόγησης.

Οι κατευθυντήριες γραμμές στοχεύουν να παρέχουν καθοδήγηση για εφαρμογές ΤΝ γενικά, χτίζοντας οριζόντια βάση για την αξιόπιστη ΤΝ. Διαφορετικές καταστάσεις, προκαλούν διαφορετικές προκλήσεις. Η αναγκαιότητα για επιπρόσθετη τμηματική προσέγγιση και συμπλήρωση του οριζόντιου πλαισίου μπορεί να διερευνηθεί.

2.2.4 Πλαίσιο Αξιόπιστης Τεχνητής Νοημοσύνης

Οι οδηγίες αυτές διαμορφώνουν ένα πλαίσιο αξιόπιστης ΤΝ με βάση τα θεμελιώδη δικαιώματα που κατοχυρώνονται στο Καταστατικό των Θεμελιωδών Δικαιωμάτων της Ευρωπαϊκής Ένωσης - Καταστατικό της ΕΕ και στο Διεθνές Δίκαιο περί Ανθρωπίνων Δικαιωμάτων. Παρακάτω εισάγονται οι τρεις βασικές έννοιες Αξιόπιστης Τεχνητής Νοημοσύνης.

2.2.5 Νομικά Συμμορφούμενη Τεχνητή Νοημοσύνη

Τα συστήματα ΤΝ λειτουργούν σε έναν κόσμο με νόμους. Ένας αριθμός νομικά δεσμευτικών κανόνων σε Ευρωπαϊκό, Εθνικό και Διεθνές επίπεδο ήδη εφαρμόζεται ή είναι συναφής με την ανάπτυξη, την εγκατάσταση και την χρήση συστημάτων ΤΝ σήμερα. Νομικές πηγές περιλαμβάνουν, μεταξύ άλλων, τις εξής:

I) Το πρωτογενές δίκαιο της ΕΕ σχετικά με τις Συνθήκες Ευρωπαϊκής Ένωσης και το Καταστατικό περί Θεμελιωδών Δικαιωμάτων^[7].

II) Το δευτερογενές δίκαιο της ΕΕ, όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων – ΓΚΠΔ ή General Data Protection Regulation - GDPR^[8], ο κανονισμός για την Οδηγία Ευθύνης Προϊόντων – ΟΕΠ ή Product Liability Directive - PLD^[9], ο κανονισμός για την Ελεύθερη Ροή Μη-προσωπικών Δεδομένων – ΕΡΜΔ ή Free Flow of Non-personal Data - FFND^[10], οι Οδηγίες κατά των Διακρίσεων^[11], το Δίκαιο των Καταναλωτών^[12] και οι Οδηγίες για την Ασφάλεια και την Υγεία στην Εργασία^[13].

III) Οι Συνθήκες των Ηνωμένων Εθνών για τα ανθρώπινα δικαιώματα και οι συμβάσεις του Συμβουλίου της Ευρώπης, όπως η Ευρωπαϊκή Σύμβαση για τα Ανθρώπινα Δικαιώματα καθώς και πολυάριθμοι νόμοι των κρατών μελών της ΕΕ.

IV) Ποικίλοι κανονισμοί εξειδικευμένων τομέων υφίστανται οι οποίοι εφαρμόζονται σε συγκεκριμένες εφαρμογές ΤΝ, εκτός από τους οριζοντίως ισχύοντες κανονισμούς. Ο νόμος προβλέπει θετικές και αρνητικές υποχρεώσεις, γεγονός που σημαίνει ότι δεν πρέπει να ερμηνεύεται μόνο σε σχέση του τι δε μπορεί να γίνει αλλά και σε σχέση με το τι μπορεί και πρέπει να γίνει. Ο νόμος δεν απαγορεύει μόνο ορισμένες πράξεις, αλλά επιτρέπει και άλλες. Αξίζει να σημειωθεί ότι το Ευρωπαϊκό Καταστατικό εμπεριέχει άρθρα σχετικά με την «επιχειρηματική ελευθερία» και την «ελευθερία των τεχνών και των επιστημών». Παράλληλα, περιέχει και άρθρα όπως η προστασία δεδομένων και η απαγόρευση διακρίσεων.

2.2.6 Ηθική Τεχνητή Νοημοσύνη

Το να πετύχει κανείς Αξιόπιστη ΤΝ δεν απαιτεί μόνο συμμόρφωση με τον νόμο, η οποία είναι ένα από τα τρία συστατικά της. Η νομοθετική ταχύτητα δεν συμβαδίζει αναγκαστικά

με τις τεχνολογικές εξελίξεις. Μπορεί να είναι εκτός συγχρονισμού με θέματα ηθικότητας ή απλά να μην μπορεί να επιλύσει τα εκάστοτε τρέχοντα ζητήματα. Προκειμένου να είναι αξιόπιστα τα συστήματα ΤΝ πρέπει να εξασφαλίζουν την ευθυγράμμιση με τους ηθικούς κανόνες.

2.2.7 Εύρωστη Τεχνητή Νοημοσύνη

Ακόμα και αν διασφαλιστεί μια ηθική στόχευση, τα άτομα και η κοινωνία πρέπει να είναι επίσης βέβαια πως τα συστήματα ΤΝ δεν προκαλούν καμία ακούσια βλάβη. Τέτοια συστήματα χρειάζεται να λειτουργούν με ασφάλεια, προστασία, αξιοπιστία και η περιφρούρησή τους πρέπει να έχει προσχεδιαστεί προκειμένου να αποτραπούν τυχόν ακούσιες αρνητικές επιπτώσεις. Είναι κατ' επέκταση σημαντικό να διασφαλιστεί πως τα συστήματα ΤΝ είναι εύρωστα. Γεγονός το οποίο είναι απαραίτητο τόσο από τεχνική άποψη, την εξασφάλιση δηλαδή της τεχνικής τους ευρωστίας είτε στον τομέα εφαρμογής, είτε στην φάση του κύκλου ζωής τους όσο και από κοινωνική άποψη, λαμβάνοντας υπόψιν δηλαδή το πλαίσιο και το περιβάλλον στα οποία το σύστημα λειτουργεί.

Η Ηθική και η Εύρωστη ΤΝ είναι στενά συνυφασμένες έννοιες, και η μία συμπληρώνει την άλλη. Οι αρχές και οι απαιτήσεις που διατυπώνονται στην συνέχεια αφορούν και τις δύο συνιστώσες^[1,2,3,4].

2.2.8 Τα θεμέλια για Αξιόπιστη Τεχνητή Νοημοσύνη

Η ηθικότητα της ΤΝ είναι μια υπο-περιοχή εφαρμοσμένης ηθικής και επικεντρώνεται στα ηθικά ζητήματα τα οποία εγείρονται από την ανάπτυξη, την εφαρμογή και την χρήση της ΤΝ. Κεντρικό της μέλημα είναι να προσδιορίσει πως η ΤΝ μπορεί να καθησυχάσει ή να αυξήσει τις ανησυχίες για την ευζωία των ατόμων, είτε μιλάμε για την ποιότητα ζωής, είτε για την ανθρώπινη αυτονομία και ελευθερία που είναι απαραίτητες για μια δημοκρατική κοινωνία.

Ο ηθικός στοχασμός πάνω στην τεχνολογία ΤΝ εξυπηρετεί πολλαπλούς σκοπούς. Αρχικά, μπορεί να οξύνει τον προβληματισμό σχετικά με την ανάγκη προστασίας ατόμων και ομάδων. Δεύτερον, μπορεί να ενισχύσει νέα είδη καινοτομιών που επιδιώκουν να προάγουν ηθικές αξίες, όπως αυτές που βοηθούν στην επίτευξη των στόχων Βιώσιμης

Ανάπτυξης του ΟΗΕ, οι οποίες είναι σταθερά ενσωματωμένες στο πρόγραμμα δράσης της ΕΕ μέχρι το 2030. Η αξιόπιστη ΤΝ μπορεί να βελτιώσει την ατομική άνθηση και την συλλογική ευημερία μέσω της δημιουργίας αξίας και την μεγιστοποίηση πλούτου, να συμβάλλει στην επίτευξη μιας δίκαιης κοινωνίας, συμμετέχοντας στην αύξηση της υγείας και της ευημερίας των πολιτών με τρόπους που προάγουν την ισότητα στην κατανομή πλούτου, στην κοινωνική και πολιτική ευμάρεια.

Η χρήση των συστημάτων ΤΝ στην κοινωνία μας εγείρει αρκετές ηθικές προκλήσεις, για παράδειγμα όσον αφορά τον αντίκτυπό τους στους ανθρώπους και την κοινωνία, τις ικανότητες λήψης αποφάσεων και την ασφάλεια. Αν πρόκειται να χρησιμοποιούμε όλο και περισσότερο τη βοήθεια των συστημάτων ΤΝ ή να αναθέτουμε τη λήψη αποφάσεων σε αυτά, πρέπει να διασφαλίσουμε ότι τα συστήματα αυτά έχουν δίκαιο αντίκτυπο στη ζωή των ανθρώπων, πως είναι σύμφωνα με αξίες που δεν πρέπει να παρακμάζουν, και πως οι κατάλληλες διαδικασίες λογοδοσίας μπορούν να διασφαλίσουν αυτές τις απαιτήσεις.

Ένας κώδικας δεοντολογίας συγκεκριμένου τομέα - όσο συνεπής, ανεπτυγμένος και λεπτομερής να είναι στην τρέχουσα μορφή αλλά και στις μελλοντικές εκδόσεις του - δεν μπορεί ποτέ να λειτουργήσει ως υποκατάστατο του ίδιου του ηθικού συλλογισμού, ο οποίος πρέπει πάντα να παραμένει ευαίσθητος στις λεπτομέρειες του πλαισίου που δεν μπορούν να αποτυπωθούν σε γενικές κατευθυντήριες γραμμές. Πέρα από την ανάπτυξη ενός συνόλου κανόνων, η εξασφάλιση αξιόπιστης ΤΝ απαιτεί από εμάς να οικοδομήσουμε και να διατηρήσουμε μια ηθική κουλτούρα και νοοτροπία μέσω δημόσιου διαλόγου, εκπαίδευσης και πρακτικής μάθησης.

2.2.9 Τα Θεμελιώδη Δικαιώματα ως Ηθικά και Νομικά Δικαιώματα

Η προσέγγιση ηθικής ΤΝ πρέπει να βασίζεται στα θεμελιώδη δικαιώματα που κατοχυρώνονται στις Συνθήκες της ΕΕ, στο Ευρωπαϊκό και στο Διεθνές Δίκαιο των ανθρωπίνων δικαιωμάτων. Ο σεβασμός των θεμελιωδών δικαιωμάτων, στο πλαίσιο της δημοκρατίας και του κράτους δικαίου, παρέχει τις πιο ελπιδοφόρες βάσεις για τον προσδιορισμό αφηρημένων ηθικών αρχών και αξιών, οι οποίες μπορούν να λειτουργήσουν στο πλαίσιο της ΤΝ.

Οι Συνθήκες της ΕΕ και το Ευρωπαϊκό Καταστατικό, προβλέπουν μια σειρά θεμελιωδών δικαιωμάτων που τα κράτη μέλη της ΕΕ και τα θεσμικά όργανα υποχρεούνται από το νόμο να σέβονται στο πλαίσιο εφαρμογής δικαίου της ΕΕ. Τα δικαιώματα αυτά περιγράφονται στο Ευρωπαϊκό Καταστατικό και περιέχουν την αξιοπρέπεια, τις ελευθερίες, την ισότητα, την αλληλεγγύη, τα δικαιώματα των πολιτών και την δικαιοσύνη. Το κοινό θεμέλιο των δικαιωμάτων αυτών, έχει τις ρίζες του στο σεβασμό της ανθρωπίνης αξιοπρέπειας, παραπέμποντας σε αυτό που περιγράφουμε ως "ανθρωποκεντρική προσέγγιση" στην οποία ο άνθρωπος απολαμβάνει ένα μοναδικό και αναφαίρετο ηθικό καθεστώς υπεροχής στον αστικό, πολιτικό, οικονομικό και κοινωνικό τομέα.

Ενώ τα δικαιώματα που ορίζονται στο Ευρωπαϊκό Καταστατικό είναι νομικώς δεσμευτικά, είναι σημαντικό να αναφερθεί ότι τα θεμελιώδη δικαιώματα δεν παρέχουν ολοκληρωμένη νομική προστασία σε κάθε περίπτωση καθώς το πεδίο εφαρμογής τους περιορίζεται σε τομείς δικαίου της ΕΕ. Το Διεθνές Δίκαιο των ανθρωπίνων δικαιωμάτων και ιδίως η Ευρωπαϊκή Σύμβαση ανθρωπίνων δικαιωμάτων είναι δεσμευτικά για τα κράτη μέλη της ΕΕ, μεταξύ άλλων και σε τομείς που εμπίπτουν εκτός του πεδίου εφαρμογής του δικαίου της ΕΕ. Τα θεμελιώδη δικαιώματα απευθύνονται σε άτομα και σε ομάδες λόγω της υπόστασής τους ως ανθρώπινα όντα, ανεξαρτήτως από τη νομική τους ισχύ. Αντιληπτά ως νομικά δικαιώματα, τα θεμελιώδη δικαιώματα εμπίπτουν επομένως στο πρώτο στοιχείο της αξιόπιστης ΤΝ, Νομικά Συμμορφούμενη ΤΝ, η οποία διασφαλίζει τη συμμόρφωση με το νόμο. Επίσης αντιληπτά ως τα δικαιώματα όλων, που έχουν τις ρίζες τους στην εγγενή ηθική υπόσταση των ανθρώπων, στηρίζουν επίσης το δεύτερο συστατικό της Αξιόπιστης ΤΝ, Ηθική ΤΝ, η οποία αφορά ηθικούς κανόνες που δεν είναι απαραίτητα νομικώς δεσμευτικοί, αλλά είναι κρίσιμοι για τη διασφάλιση της αξιοπιστίας.

2.2.10 Τα Θεμελιώδη Δικαιώματα ως Βάση για Αξιόπιστη Τεχνητή Νοημοσύνη

Μεταξύ του ολοκληρωμένου συνόλου αδιαίρετων δικαιωμάτων, τα οποία είναι ορισμένα στο Διεθνές Δίκαιο των Ανθρωπίνων Δικαιωμάτων, στις Συνθήκες της ΕΕ και στο Ευρωπαϊκό Καταστατικό οι παρακάτω συστάδες θεμελιωδών δικαιωμάτων είναι ιδιαίτερα κατάλληλες να καλύψουν τα συστήματα ΤΝ. Πλήθος εξ αυτών των δικαιωμάτων είναι, σε συγκεκριμένες περιπτώσεις, νομικά επιβαλλόμενα στην ΕΕ, έτσι ώστε η συμμόρφωση με

τους όρους τους να είναι νομικά υποχρεωτική. Όμως ακόμη και μετά την επίτευξη της συμμόρφωσης με τα νομικά κατοχυρωμένα θεμελιώδη δικαιώματα, ο ηθικός προβληματισμός μπορεί να βοηθήσει στην κατανόηση του πώς η ανάπτυξη και η χρήση συστημάτων ΤΝ μπορεί να εμπλέξει θεμελιώδη δικαιώματα και τις υποκείμενες αξίες τους, και μπορεί να βοηθήσει στην παροχή πιο λεπτομερούς καθοδήγησης όταν επιδιώκουμε να προσδιορίσουμε τί πρέπει να κάνουμε και όχι τί μπορούμε επί του παρόντος να κάνουμε με την τεχνολογία.

2.2.11 Σεβασμός Ανθρώπινης Αξιοπρέπειας

Η ανθρώπινη αξιοπρέπεια σημαίνει ότι κάθε ανθρώπινο ον διαθέτει μια "εγγενή αξία", η οποία δεν πρέπει ποτέ να μειώνεται, να διακυβεύεται ή να καταστέλλεται από τις νέες τεχνολογίες όπως η ΤΝ. Στο πλαίσιο αυτό, ο σεβασμός της ανθρώπινης αξιοπρέπειας συνεπάγεται ότι όλοι οι άνθρωποι αντιμετωπίζονται με τον σεβασμό που τους αρμόζει ως υποκείμενα με ηθική αξία και όχι απλώς ως αντικείμενα που πρέπει να κοσκινίζονται, να ταξινομούνται, να βαθμολογούνται, να εκτρέφονται, να εξαρτώνται ή να χειραγωγούνται. Τα συστήματα ΤΝ πρέπει συνεπώς να αναπτύσσονται με τρόπο που να σέβεται, να εξυπηρετεί και να προστατεύει την σωματική και την ψυχική υγεία του ανθρώπου. Όπως επίσης την ακεραιότητα, την προσωπική και την πολιτισμική αίσθηση ταυτότητας και την ικανοποίηση των βασικών αναγκών του ατόμου.

2.2.12 Ελευθερία Ατόμου

Τα ανθρώπινα όντα πρέπει να παραμένουν ελεύθερα να λαμβάνουν τα ίδια αποφάσεις για τους εαυτούς τους. Αυτό συνεπάγεται ελευθερία από απολυταρχικές επεμβάσεις, αλλά απαιτεί επίσης παρέμβαση από κυβερνητικές και μη οργανώσεις για να διασφαλιστεί ότι τα άτομα ή οι άνθρωποι που κινδυνεύουν να αποκλειστούν, να έχουν ισότιμη πρόσβαση στα οφέλη της ΤΝ και στις ευκαιρίες της. Στο πλαίσιο της ΤΝ, η ελευθερία του ατόμου, απαιτεί μετριασμό των παράνομων εξαναγκασμών, των απειλών κατά της ψυχικής αυτονομίας και της ψυχικής υγείας, της αδικαιολόγητης παρακολούθησης, της εξαπάτησης και της αθέμιτης χειραγώγησης.

Στην πραγματικότητα, η ελευθερία του ατόμου σημαίνει δέσμευση να επιτρέπεται στα άτομα να ασκούν ακόμη υψηλότερο έλεγχο στην ζωή τους, συμπεριλαμβανομένου της

προστασίας της ελεύθερης άσκησης επιχειρηματικής δραστηριότητας, της ελευθερίας των τεχνών και της επιστήμης, την ελευθερία της έκφρασης, το δικαίωμα στην ιδιωτική ζωή και την ιδιωτικότητα και την ελευθερία του συνέρχεσθαι και του συνεταιρίζεσθαι.

2.2.13 Σεβασμός στην Δημοκρατία, την Δικαιοσύνη και το Κράτος Δικαίου

Η κυβερνητική εξουσία στις συνταγματικές δημοκρατίες, οφείλει να είναι νομικά εξουσιοδοτημένη και περιορισμένη από το νόμο. Τα συστήματα ΤΝ πρέπει να υπηρετούν την διατήρηση και την προώθηση των δημοκρατικών διαδικασιών και να σέβονται τον πλουραλισμό των αξιών και των επιλογών ζωής των ατόμων. Πρέπει να μην υπονομεύουν τις δημοκρατικές διαδικασίες, την ανθρώπινη διαβούλευση ή τα δημοκρατικά συστήματα ψηφοφορίας. Πρέπει επίσης να ενσωματώνουν τη δέσμευση διασφάλισης ότι δεν λειτουργούν με τρόπους που υπονομεύουν τις θεμελιώδεις αρχές στις οποίες βασίζεται το κράτος δικαίου, τους υποχρεωτικούς νόμους και κανονισμούς, την τήρηση της δέουσας διαδικασίας και την ισότητα ενώπιον του νόμου.

2.2.14 Ισότητα, Μη-διάκριση και Αλληλεγγύη

Ο ισότιμος σεβασμός της αξιοπρέπειας όλων των ανθρώπων πρέπει να διασφαλίζεται, συμπεριλαμβανομένων των δικαιωμάτων των ατόμων που κινδυνεύουν με αποκλεισμό. Αυτό υπερβαίνει την απαγόρευση των διακρίσεων, η οποία ανέχεται τη διάκριση μεταξύ ανόμοιων καταστάσεων βάσει αντικειμενικών κριτηρίων. Στο πλαίσιο της ΤΝ, η ισότητα συνεπάγεται ότι οι λειτουργίες του συστήματος δεν μπορούν να παράγουν αθέμιτα προκατειλημμένες εξόδους. Για παράδειγμα, τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση των συστημάτων ΤΝ πρέπει να είναι όσο το δυνατόν πιο περιεκτικά, αντιπροσωπεύοντας διαφορετικές ομάδες πληθυσμού. Κάτι τέτοιο απαιτεί επαρκή σεβασμό για δυνητικά ευάλωτα άτομα και ομάδες, όπως οι εργαζόμενοι, οι γυναίκες, τα άτομα με αναπηρία, οι εθνικές μειονότητες, παιδιά, καταναλωτές ή άλλα άτομα που διατρέχουν τον κίνδυνο του αποκλεισμού.

2.2.15 Δικαιώματα των Πολιτών

Οι πολίτες έχουν πρόσβαση σε ένα δημοκρατικό φάσμα δικαιωμάτων, όπως το δικαίωμα ψήφου, το δικαίωμα στην καλή διοίκηση ή την πρόσβαση σε δημόσια έγγραφα, όπως και

το δικαίωμα υποβολής αιτήσεων προς τη διοίκηση. Τα συστήματα ΤΝ προσφέρουν σημαντικές δυνατότητες βελτίωσης της κλιμακωσιμότητας και της αποτελεσματικότητας της κυβέρνησης στην παροχή δημόσιων αγαθών και υπηρεσιών στην κοινωνία. Όμως, τα δικαιώματα των πολιτών μπορούν να επηρεαστούν αρνητικά από τα συστήματα ΤΝ και πρέπει να περιφρουρούνται.

2.2.16 Ηθικές Αρχές στο Πλαίσιο Συστημάτων Τεχνητής Νοημοσύνης

Πολλές δημόσιες, ιδιωτικές και αστικές οργανώσεις έχουν εμπνευστεί από τα θεμελιώδη δικαιώματα για να παράγουν ηθικά πλαίσια εργασίας για συστήματα ΤΝ. Στην ΕΕ, η Ευρωπαϊκή Ομάδα Δεοντολογίας – ΕΟΔ στην Επιστήμη και τις Νέες Τεχνολογίες ή European Group on Ethics - EGE πρότεινε ένα σύνολο 9 θεμελιωδών βασικών αρχών, με βάση τις θεμελιώδεις αξίες που ορίζονται στις Συνθήκες και στο Ευρωπαϊκό Καταστατικό. Η ανάπτυξη των ηθικών αρχών βασίζεται σε αυτό το έργο, εν συνεχεία αναγνωρίζει τις περισσότερες από τις αρχές που έχουν διατυπωθεί μέχρι σήμερα από διάφορες ομάδες και αποσαφηνίζει τους σκοπούς που όλες οι αρχές επιδιώκουν να καλλιεργήσουν και να υποστηρίξουν. Οι ηθικές αρχές αυτές μπορούν να εμπνεύσουν νέα και συγκεκριμένα ρυθμιστικά όργανα, μπορούν να βοηθήσουν στην ερμηνεία των θεμελιωδών δικαιωμάτων καθώς το κοινωνικο-τεχνικό περιβάλλον εξελίσσεται με την πάροδο του χρόνου, και μπορούν να καθοδηγήσουν τη λογική για την ανάπτυξη, την εφαρμογή και τη χρήση των συστημάτων ΤΝ.

Τα συστήματα ΤΝ χρειάζεται να βελτιώνουν την ατομική και συλλογική ευημερία. Παραθέτονται, χωρίς ιεραρχία, τέσσερις ηθικές αρχές, που έχουν τις ρίζες τους στα θεμελιώδη δικαιώματα, οι οποίες πρέπει να γίνονται σεβαστές προκειμένου να διασφαλιστεί ότι τα συστήματα ΤΝ αναπτύσσονται και χρησιμοποιούνται με αξιόπιστο τρόπο. Καθορίζονται ως ηθικές επιταγές, έτσι ώστε οι επαγγελματίες της ΤΝ να επιδιώκουν πάντα να τις τηρούν^[1,2,3,4].

2.2.17 Αρχή Σεβασμού Ανθρώπινης Αυτονομίας

Τα θεμελιώδη δικαιώματα στα οποία βασίζεται η ΕΕ αποσκοπούν στην εξασφάλιση του σεβασμού της ελευθερίας και της αυτονομίας του ανθρώπου. Οι άνθρωποι που αλληλεπιδρούν με συστήματα ΤΝ πρέπει να μπορούν να διατηρούν πλήρη και

αποτελεσματική αυτοδιάθεση των εαυτών τους και να μπορούν να συμμετέχουν στη δημοκρατική διαδικασία. Τα συστήματα ΤΝ δεν πρέπει να υποδαυλίζουν, να εξαναγκάζουν, να εξαπατούν ή να χειραγωγούν αδικαιολόγητα τους ανθρώπους. Αντίθετα, πρέπει να σχεδιάζονται με τέτοιο τρόπο ώστε να ενισχύουν, να συμπληρώνουν και να ενδυναμώνουν τις ανθρώπινες γνωστικές, κοινωνικές και πολιτιστικές δεξιότητες. Η κατανομή των λειτουργιών μεταξύ των ανθρώπων και των συστημάτων ΤΝ πρέπει να ακολουθεί ανθρωποκεντρικό σχεδιασμό και να αφήνει ουσιαστικές ευκαιρίες για ανθρώπινη επιλογή. Αυτό σημαίνει πως διασφαλίζεται η ανθρώπινη εποπτεία του κύκλου εργασίας συστημάτων ΤΝ, η υποστήριξη των ανθρώπων στο εργασιακό περιβάλλον και η στόχευση στην δημιουργία εργασίας με νόημα.

2.2.18 Αρχή Πρόληψης Βλάβης

Τα συστήματα ΤΝ δεν πρέπει να προκαλούν, να επιδεινώνουν βλάβες ή να επηρεάζουν με τρόπο αρνητικό τον άνθρωπο. Συνεπάγεται η προστασία της ανθρώπινης αξιοπρέπειας και της ψυχικής και σωματικής ακεραιότητας. Τα συστήματα ΤΝ και τα περιβάλλοντα στα οποία λειτουργούν οφείλουν να είναι ασφαλή και προστατευμένα. Πρέπει να είναι τεχνικά εύρωστα και πρέπει να διασφαλίζεται ότι δεν είναι ανοικτά σε κακόβουλη χρήση. Τα ευάλωτα άτομα πρέπει να λάβουν μεγαλύτερη προσοχή και να συμπεριληφθούν στην ανάπτυξη και την χρήση των συστημάτων ΤΝ. Ιδιαίτερη προσοχή πρέπει επίσης να δοθεί στις περιπτώσεις όπου τα συστήματα ΤΝ μπορούν να προκαλέσουν ή να επιδεινώσουν δυσμενείς επιπτώσεις λόγω κατώτερης εξουσίας ή πληροφόρησης, όπως μεταξύ εργοδοτών και εργαζομένων, επιχειρήσεων και καταναλωτών ή κυβερνήσεων και πολιτών. Η πρόληψη της βλάβης συνεπάγεται επίσης την προστασία του φυσικού περιβάλλοντος και όλων των έμβιων όντων.

2.2.19 Αρχή Δικαιοσύνης

Αναγνωρίζεται το γεγονός ότι υφίστανται διαφορετικές ερμηνείες της δικαιοσύνης, η δικαιοσύνη έχει τόσο μια ουσιαστική όσο και μια διαδικαστική διάσταση. Η ουσιαστική διάσταση συνεπάγεται δέσμευση για διασφάλιση της ίσης και δίκαιης κατανομής τόσο των οφελών όσο και του κόστους της ΤΝ. Όπως επίσης, τη διασφάλιση ότι τα άτομα και οι ομάδες είναι απαλλαγμένες από άδικες προκαταλήψεις, διακρίσεις και στιγματισμό.

Εάν οι άδικες προκαταλήψεις μπορούν να αποφευχθούν, τα συστήματα ΤΝ μπορούν ακόμη και να αυξήσουν την κοινωνική δικαιοσύνη προσφέροντας ίσες ευκαιρίες όσον αφορά την πρόσβαση στην εκπαίδευση, στα αγαθά, στις υπηρεσίες και στην τεχνολογία.

Η σχεδίαση συστημάτων ΤΝ δεν πρέπει ποτέ να οδηγεί στην εξαπάτηση ανθρώπων ή να περιοριστεί αδικαιολόγητα η ελευθερία επιλογών τους. Επιπλέον, η δικαιοσύνη συνεπάγεται ότι οι επαγγελματίες ΤΝ πρέπει να σέβονται την αρχή της αναλογικότητας μεταξύ μέσων και σκοπών και να εξετάζουν προσεκτικά τον τρόπο εξισορρόπησης ανταγωνιστικών συμφερόντων και στόχων. Η διαδικαστική διάσταση της δικαιοσύνης συνεπάγεται τη δυνατότητα αμφισβήτησης και της αποτελεσματικής προσφυγής κατά των αποφάσεων που λαμβάνονται από τα συστήματα ΤΝ και από τους ανθρώπους που τα χειρίζονται. Προκειμένου να γίνει αυτό, η οντότητα που είναι υπεύθυνη για την απόφαση πρέπει να είναι ταυτοποιήσιμη και οι διαδικασίες λήψης αποφάσεων να είναι εξηγήσιμες.

2.2.20 Αρχή Εξηγησιμότητας

Η εξηγησιμότητα είναι ζωτικής σημασίας για την δημιουργία σχέσεων εμπιστοσύνης στα συστήματα ΤΝ. Αυτό σημαίνει ότι οι διαδικασίες πρέπει να είναι διαφανείς, οι δυνατότητες και ο σκοπός των συστημάτων να κοινοποιούνται, και οι αποφάσεις, στο βαθμό που είναι δυνατόν, να εξηγούνται σε όσους επηρεάζονται άμεσα και έμμεσα. Χωρίς αυτές τις πληροφορίες, η απόφαση δεν μπορεί να αμφισβητηθεί δεόντως. Μια εξήγηση σχετικά με το γιατί ένα μοντέλο παράγει μια συγκεκριμένη έξοδο ή απόφαση και ποιος συνδυασμός παραγόντων εισόδου συνέβαλε σε αυτήν δεν είναι πάντα εφικτό. Οι περιπτώσεις αυτές αναφέρονται ως αλγόριθμοι "μαύρου κουτιού" και απαιτούν ιδιαίτερη προσοχή. Σε αυτές τις περιπτώσεις, επιπλέον μέτρα εξηγησιμότητας χρειάζονται, όπως η ιχνηλασιμότητα, η ελεγχιμότητα και η διαφανής επικοινωνία σχετικά με τις δυνατότητες του συστήματος, υπό την προϋπόθεση φυσικά ότι το σύστημα στο σύνολό του σέβεται τα θεμελιώδη δικαιώματα. Ο βαθμός στον οποίο απαιτείται εξηγησιμότητα εξαρτάται σε μεγάλο βαθμό από το πλαίσιο και τη σοβαρότητα των συνεπειών εάν η εξήγηση είναι λανθασμένη ή ανακριβής.

2.2.21 Εντάσεις Μεταξύ Αρχών

Μπορεί να προκύψουν εντάσεις μεταξύ των ανωτέρω αρχών, σε περιπτώσεις για τις οποίες δεν υπάρχει σταθερή λύση. Σύμφωνα με την Ευρωπαϊκή θεμελιώδη απαίτηση για δημοκρατία, οι δέουσες διαδικασίες, η ανοικτή συμμετοχή πολιτών και η υπεύθυνη διαβούλευση, πρέπει να καθιερωθούν για την αντιμετώπιση τέτοιων εντάσεων. Σε διάφορες εφαρμογές, η αρχή της πρόληψης της βλάβης και η αρχή της ανθρώπινης αυτονομίας μπορεί να συγκρούονται. Για παράδειγμα, η χρήση ΤΝ με στόχο την προγνωστική αστυνόμευση, μπορεί να συμβάλει στη μείωση της εγκληματικότητας, με τρόπους όμως οι οποίοι συνεπάγονται δραστηριότητες επιτήρησης θίγοντας έτσι την ατομική ελευθερία και την ιδιωτική ζωή.

Τα οφέλη των συστημάτων ΤΝ πρέπει να αντισταθμίζουν κατά πολύ τους φανερούς ατομικούς κινδύνους. Ενώ οι παραπάνω αρχές προσφέρουν σίγουρα καθοδήγηση προς την κατεύθυνση λύσεων, παραμένουν αφηρημένες ηθικές επιταγές. Οι επαγγελματίες ΤΝ δεν μπορεί συνεπώς να αναμένεται να βρουν την σωστή λύση με βάση τις παραπάνω αρχές, ωστόσο πρέπει να προσεγγίζουν τα ηθικά διλήμματα και τα κέρδη-απώλειες μέσω λογικού, τεκμηριωμένου προβληματισμού και όχι μέσω διαίσθησης ή τυχαίας διακριτικής ευχέρειας.

Ενδέχεται, ωστόσο, να υπάρξουν καταστάσεις στις οποίες δεν μπορούν να προσδιοριστούν ηθικά αποδεκτές σχέσεις κέρδους και απώλειας. Ορισμένα θεμελιώδη δικαιώματα και συσχετιζόμενες αρχές είναι απόλυτες και δεν μπορούν να αποτελέσουν αντικείμενο άσκησης εξισορρόπησης, όπως η ανθρώπινη αξιοπρέπεια.

2.2.22 Προϋποθέσεις για Αξιόπιστη Τεχνητή Νοημοσύνη

Οι παραπάνω αρχές πρέπει να ερμηνευθούν ως ισχυρές απαιτήσεις επίτευξης Αξιόπιστης ΤΝ. Οι απαιτήσεις εφαρμόζονται σε διάφορα ενδιαφερόμενα μέρη που συμμετέχουν στον κύκλο ζωής των συστημάτων ΤΝ. Προγραμματιστές, φορείς ανάπτυξης, τελικοί χρήστες καθώς και η ευρύτερη κοινωνία. Με τον όρο προγραμματιστές, αναφερόμαστε σε όσους ερευνούν, σχεδιάζουν ή/και αναπτύσσουν συστήματα ΤΝ. Με τον όρο φορείς ανάπτυξης, αναφερόμαστε σε δημόσιους ή ιδιωτικούς οργανισμούς που χρησιμοποιούν συστήματα ΤΝ στο πλαίσιο των επιχειρησιακών

διαδικασιών ώστε να προσφέρουν προϊόντα και υπηρεσίες σε άλλους. Οι τελικοί χρήστες είναι εκείνοι που εμπλέκονται με το σύστημα ΤΝ, άμεσα ή έμμεσα. Η ευρύτερη κοινωνία περιλαμβάνει όλους τους άλλους που επηρεάζονται άμεσα ή έμμεσα από τα συστήματα ΤΝ.

Διαφορετικές ομάδες ενδιαφερομένων μερών, διαδραματίζουν διαφορετικό ρόλο στη διασφάλιση της ικανοποίησης των απαιτήσεων. Οι προγραμματιστές πρέπει να εφαρμόζουν τις απαιτήσεις στις διαδικασίες σχεδιασμού και ανάπτυξης. Οι φορείς ανάπτυξης πρέπει να διασφαλίζουν ότι τα συστήματα που χρησιμοποιούν και τα προϊόντα και οι υπηρεσίες που προσφέρουν πληρούν τις απαιτήσεις. Οι τελικοί χρήστες και η ευρύτερη κοινωνία πρέπει να ενημερώνονται για τις απαιτήσεις αυτές και να απαιτούν την τήρησή τους.

Το παρακάτω σχήμα απαιτήσεων δεν είναι εξαντλητικό. Περιλαμβάνει συστημικές, ατομικές και κοινωνικές πτυχές:



Εικόνα 2: ΟΕΥΕ, Αλληλεπίδραση των επτά Απαιτήσεων Αξιοπιστίας ΤΝ^[4]

2.2.23 Ανθρώπινη Παρέμβαση και Εποπτεία

Τα συστήματα ΤΝ πρέπει να υποστηρίζουν την ανθρώπινη αυτονομία και τη λήψη αποφάσεων, όπως ορίζει η αρχή του σεβασμού της ανθρώπινης αυτονομίας. Αυτό προϋποθέτει ότι ενεργούν ως καταλύτες για μια δημοκρατική, ακμάζουσα και δίκαιη κοινωνία, υποστηρίζοντας τη δράση του χρήστη, προωθώντας τα θεμελιώδη δικαιώματα και επιτρέποντας την ανθρώπινη εποπτεία.

I) Θεμελιώδη Δικαιώματα

Τα συστήματα ΤΝ, όπως πολλές τεχνολογίες, μπορούν να σεβαστούν αλλά και να παραβιάσουν τα θεμελιώδη δικαιώματα. Μπορούν να ωφελήσουν τους ανθρώπους, για παράδειγμα αυξάνοντας την πρόσβαση στην εκπαίδευση υποστηρίζοντας έτσι το δικαίωμά τους στην παιδεία. Ωστόσο, δεδομένης της εμβέλειας και της ικανότητας τους, μπορούν επίσης να επηρεάσουν αρνητικά τα θεμελιώδη δικαιώματα. Σε περιπτώσεις όπου υπάρχουν τέτοιοι κίνδυνοι, πρέπει να αξιολογείται ο αντίκτυπος στα θεμελιώδη δικαιώματα. Η αξιολόγηση πρέπει να γίνει πριν από την ανάπτυξη του συστήματος και να περιλαμβάνει μια αποτίμηση του κατά πόσον οι κίνδυνοι αυτοί μπορούν να μειωθούν ή να δικαιολογηθούν όπως είναι απαραίτητο σε μια δημοκρατική κοινωνία, προκειμένου να τηρούνται τα δικαιώματα και οι ελευθερίες των άλλων. Επιπλέον, πρέπει να λειτουργήσουν μηχανισμοί για τη λήψη εξωτερικής ανατροφοδότησης σχετικά με τα συστήματα ΤΝ που ενδεχομένως παραβιάζουν τα θεμελιώδη δικαιώματα.

II) Ανθρώπινη Παρέμβαση

Οι χρήστες πρέπει να είναι σε θέση να λαμβάνουν πληροφορημένες αυτόνομες αποφάσεις σχετικά με τα συστήματα ΤΝ. Πρέπει να τους παρέχονται οι γνώσεις και τα εργαλεία για να κατανοούν και να αλληλεπιδρούν με τα συστήματα ΤΝ σε ικανοποιητικό βαθμό και όπου είναι δυνατόν, να τους δίνεται η δυνατότητα να αξιολογούν ή να αμφισβητούν το σύστημα. Τα συστήματα ΤΝ μπορούν μερικές φορές να αναπτυχθούν για να διαμορφώνουν και να επηρεάζουν την ανθρώπινη συμπεριφορά μέσω μηχανισμών που μπορεί να είναι δύσκολο να εντοπιστούν, καθώς μπορούν να αξιοποιήσουν υποσυνείδητες διεργασίες, συμπεριλαμβανομένων διαφόρων μορφών αθέμιτης χειραγώγησης, εξαπάτησης, αγελαίας συμπεριφοράς, μηχανισμοί εν δυνάμει απειλητικοί για την ατομική αυτονομία. Η γενική αρχή της αυτονομίας του χρήστη πρέπει

να βρίσκεται στο επίκεντρο της λειτουργίας του συστήματος TN. Βασικό στοιχείο είναι να έχει το δικαίωμα να μην υπόκειται σε αποφάσεις που βασίζονται αποκλειστικά σε αυτοματοποιημένη επεξεργασία, όταν αυτή παράγει νομικές συνέπειες ή τον επηρεάζει σημαντικά με παρεμφερή τρόπο.

III) Ανθρώπινη Εποπτεία

Βοηθάει να διασφαλιστεί ότι ένα σύστημα TN δεν υπονομεύει την ανθρώπινη αυτονομία ή προκαλεί άλλες δυσμενείς επιπτώσεις. Η επίβλεψη μπορεί να επιτευχθεί μέσω μηχανισμών διακυβέρνησης όπως η προσέγγιση Άνθρωπος Εντός Βρόχου – AEB ή Human-In-the-Loop - HIL, Άνθρωπος Πάνω στον Βρόχο – ΑΠΒ ή Human-On-the-Loop – HOTL, ή Άνθρωπος Στην Απόφαση – ΑΣΑ ή Human-In-Command - HIC. Ο Άνθρωπος Εντός Βρόχου αναφέρεται στην ικανότητα ανθρώπινης παρέμβασης σε κάθε κύκλο λήψης αποφάσεων του συστήματος, η οποία σε πολλές περιπτώσεις δεν είναι ούτε δυνατή, ούτε επιθυμητή. Ο Άνθρωπος Πάνω στον Βρόχο αναφέρεται στη δυνατότητα ανθρώπινης παρέμβασης κατά τη διάρκεια του κύκλου σχεδιασμού του συστήματος και την παρακολούθηση της λειτουργίας του συστήματος. Ο Άνθρωπος Στην Απόφαση αναφέρεται στην ικανότητα εποπτείας της συνολικής δραστηριότητας του συστήματος TN, συμπεριλαμβανομένου του ευρύτερου οικονομικού, κοινωνικού, νομικού και ηθικού αντίκτυπου και την ικανότητα απόφασης για την χρήση του συστήματος σε οποιαδήποτε κατάσταση. Αυτό μπορεί να περιλαμβάνει την απόφαση να μην χρησιμοποιηθεί ένα σύστημα TN σε μια συγκεκριμένη κατάσταση ή να εξασφαλιστεί η δυνατότητα παράκαμψης μιας απόφασης που έχει ληφθεί από ένα σύστημα.

Επιπλέον, πρέπει να διασφαλιστεί ότι οι δημόσιοι φορείς επιβολής νόμου έχουν τη δυνατότητα να εποπτεύουν σύμφωνα με εντολή τους. Οι μηχανισμοί εποπτείας μπορεί να απαιτούνται σε διαφορετικό βαθμό για την υποστήριξη άλλων μέτρων ασφάλειας και ελέγχου, ανάλογα με την περιοχή εφαρμογής του συστήματος TN και τον δυνητικό κίνδυνο. Όλοι οι άλλοι παράγοντες είναι ίσοι, όσο λιγότερη εποπτεία μπορεί να ασκήσει ένας άνθρωπος σε ένα σύστημα TN, τόσο πιο εκτεταμένες δοκιμές και αυστηρότερη διακυβέρνηση απαιτούνται^[1,2,3,4].

2.2.24 Τεχνική Ευρωστία και Ασφάλεια

Μια κρίσιμη συνιστώσα για την επίτευξη αξιόπιστης ΤΝ είναι η τεχνική ευρωστία, η οποία συνδέεται στενά με την αρχή πρόληψης της βλάβης. Η τεχνική ευρωστία προϋποθέτει ότι τα συστήματα ΤΝ, αναπτύσσονται με ενσωματωμένη πρόληψη κινδύνων και με τέτοιο τρόπο ώστε να συμπεριφέρονται αξιόπιστα, ελαχιστοποιώντας την ακούσια και απροσδόκητη βλάβη. Αυτό πρέπει να ισχύει και για πιθανές αλλαγές στο λειτουργικό τους περιβάλλον ή την παρουσία άλλων παραγόντων, ανθρώπινων και τεχνητών, που μπορεί να αλληλεπιδράσουν με το σύστημα με ανταγωνιστικό τρόπο. Επιπλέον, πρέπει να διασφαλίζεται η σωματική και ψυχική ακεραιότητα των ανθρώπων.

I) Ανθεκτικότητα σε Επιθέσεις και Ασφάλεια

Τα συστήματα ΤΝ, όπως όλα τα συστήματα λογισμικού, πρέπει να προστατεύονται από τρωτά σημεία που μπορούν να τους επιτρέψουν να γίνουν αντικείμενο εκμετάλλευσης από "αντιπάλους", όπως κακόβουλοι προγραμματιστές. Οι επιθέσεις μπορεί να στοχεύουν στην νοθεία δεδομένων, την διαρροή μοντέλου ή την υποκείμενη υποδομή, τόσο σε επίπεδο λογισμικού όσο και σε υλικού. Εάν ένα σύστημα ΤΝ δεχθεί επίθεση, μπορεί να το εξαναγκάσει να λάβει διαφορετικές αποφάσεις, ή να το αναγκάσει να κλείσει εντελώς. Τα συστήματα και τα δεδομένα μπορούν επίσης να διαφθαρθούν από κακόβουλες προθέσεις ή από την έκθεση σε απροσδόκητες καταστάσεις. Οι ανεπαρκείς διαδικασίες ασφάλειας μπορούν επίσης να οδηγήσουν σε εσφαλμένες αποφάσεις ή ακόμη και σε σωματικές βλάβες.

Για να θεωρηθούν τα συστήματα ΤΝ ασφαλή, οι πιθανές ακούσιες εφαρμογές των συστημάτων ΤΝ, όπως οι εφαρμογές διπλής χρήσης και η πιθανή κατάχρηση του συστήματος από κακόβουλες οντότητες, πρέπει να λαμβάνονται υπόψιν και μέτρα προληπτικού χαρακτήρα χρειάζεται να υιοθετούνται για τον μετριασμό τέτοιων συμβάντων.

II) Σχέδιο Εφεδρείας και Γενικής Ασφάλειας

Τα συστήματα ΤΝ πρέπει να έχουν ένα σχέδιο εφεδρείας σε περίπτωση που προκύψουν προβλήματα. Κάτι τέτοιο σημαίνει πως το σύστημα μεταβαίνει από μια στατιστική διαδικασία σε μία διαδικασία βασισμένη σε κανόνες, ή ότι ζητούν άδεια από έναν

άνθρωπο-χειριστή πριν συνεχίσουν την δράση τους. Πρέπει να διασφαλίζεται ότι το σύστημα εξακολουθεί να ολοκληρώνει τους στόχους του χωρίς να προκαλεί βλάβες σε ζωντανά όντα ή στο περιβάλλον. Περιλαμβάνεται δηλαδή η ελαχιστοποίηση ακούσιων συνεπειών και λαθών. Επιπροσθέτως, χρειάζεται να δημιουργηθούν διαδικασίες οι οποίες διασαφηνίζουν και αξιολογούν τους κινδύνους που είναι συνδεδεμένοι με την χρήση των συστημάτων TN σε όλο το φάσμα του πεδίου εφαρμογών τους. Το επίπεδο των μέτρων ασφαλείας που πρέπει να ληφθούν εξαρτάται από το εύρος του κινδύνου που θέτει το σύστημα TN, εξαρτάται δηλαδή από τις δυνατότητες του συστήματος αυτού καθ' εαυτού. Όπου μπορεί να προβλεφθεί ότι η διαδικασία ανάπτυξης ή ότι το ίδιο το σύστημα ενδέχεται να ενέχει υψηλούς κινδύνους, έχει μεγάλη σημασία τα μέτρα ασφαλείας να αναπτυχθούν και να δοκιμαστούν προληπτικά.

III) Ακρίβεια

Η ακρίβεια αφορά την ικανότητα ενός συστήματος TN να είναι σε θέση να διαμορφώσει ορθές κρίσεις, όπως το να ταξινομεί πληροφορίες στις κατάλληλες κατηγορίες, να διαθέτει την ικανότητα δημιουργίας ορθών προβλέψεων και προτάσεων, να λαμβάνει αποφάσεις με βάση δεδομένα και μοντέλα. Μια καλά διαμορφωμένη διαδικασία ανάπτυξης και αξιολόγησης μπορεί να μετριάσει και να διορθώσει ακούσιους κινδύνους από ανακριβείς προβλέψεις. Όταν οι ανακριβείς προβλέψεις είναι αναπόφευκτο να συμβούν, είναι σημαντικό το σύστημα να υποδεικνύει πόσο πιθανά είναι αυτά τα σφάλματα. Ένα υψηλό επίπεδο ακρίβειας, είναι απαραίτητο όταν το σύστημα επηρεάζει ανθρώπινες ζωές.

IV) Αξιοπιστία και Αναπαραξιμότητα

Αποτελεί ζωτικής σημασίας παράγοντα στα συστήματα TN τα αποτελέσματα να είναι αξιόπιστα και αναπαραξιμότητα. Αξιοπιστία στα συστήματα TN ορίζουμε την ορθή λειτουργία για μεγάλο εύρος εισόδων και καταστάσεων. Γεγονός που είναι απαραίτητο για τον έλεγχο του συστήματος και την αποφυγή ακούσιων βλαβών. Η αναπαραξιμότητα στα συστήματα TN ορίζεται ως η διατήρηση σταθερής συμπεριφοράς κατόπιν επανάληψης των πειραμάτων υπό τις ίδιες συνθήκες. Αυτό επιτρέπει στους επιστήμονες να περιγράψουν με ακρίβεια την λειτουργία συστημάτων TN. Τα αρχεία αντιγραφής είναι κατάλληλα για την φιλοξενία της διαδικασίας του ελέγχου και της αναπαραγωγής συμπεριφορών των συστημάτων TN^[1,2,3,4].

2.2.25 Ιδιωτικότητα και Επίβλεψη Δεδομένων

I) Ιδιωτικότητα και Προστασία Δεδομένων

Τα συστήματα ΤΝ πρέπει να εγγυώνται την προστασία της ιδιωτικής ζωής και των δεδομένων καθ' όλη την διάρκεια των κύκλων ζωών τους. Η εγγύηση περιλαμβάνει τις πληροφορίες που παρέχονται από τον χρήστη καθώς και τις πληροφορίες που παράγονται για τον χρήστη κατόπιν αλληλεπίδρασής του με το σύστημα. Για παράδειγμα, έξοδοι που το σύστημα ΤΝ παρήγαγε για συγκεκριμένους χρήστες ή πως ανταποκρίθηκαν οι χρήστες σε συγκεκριμένες προτάσεις.

Τα ψηφιακά αρχεία της ανθρώπινης συμπεριφοράς μπορούν να επιτρέψουν στα συστήματα ΤΝ να συμπεραίνουν τις προτιμήσεις των ατόμων, τον σεξουαλικό προσανατολισμό, την ηλικία, το φύλο, την θρησκευτική ή την πολιτική τους ταυτότητα. Προκειμένου η διαδικασία συλλογής δεδομένων να είναι έμπιστη χρειάζεται να διασφαλιστεί ότι τα δεδομένα που συλλέγονται για τα άτομα δεν χρησιμοποιούνται για παράνομες ή άδικες διακρίσεις εις βάρος τους.

II) Ποιότητα και Ακεραιότητα Δεδομένων

Η ποιότητα των υπό χρήση συνόλων δεδομένων είναι ύψιστης σημασίας για την απόδοση των συστημάτων ΤΝ. Κατά την συλλογή δεδομένων, υπάρχει το ενδεχόμενο να εμπεριέχονται κοινωνικά κατασκευασμένες προκαταλήψεις, ανακρίβειες και σφάλματα. Γεγονός που χρειάζεται να αντιμετωπιστεί πριν την εκπαίδευση ενός μοντέλου με οποιοδήποτε σύνολο δεδομένων. Πρέπει επίσης, να διασφαλίζεται η ακεραιότητα των δεδομένων. Η τροφοδότηση σε ένα σύστημα ΤΝ κακόβουλων δεδομένων μπορεί να αλλάξει την συμπεριφορά του, ειδικά σε συστήματα αυτο-εκμάθησης. Οι διαδικασίες και τα σύνολα των δεδομένων που χρησιμοποιούνται χρειάζεται να ελέγχονται και να τεκμηριώνονται σε κάθε βήμα. Το ίδιο πρέπει να ισχύει και για τα αποκτηθέντα συστήματα ΤΝ τα οποία δεν αναπτύχθηκαν εσωτερικά.

III) Πρόσβαση στα Δεδομένα

Κάθε οργανισμός ο οποίος διαχειρίζεται δεδομένα ατόμων, πρέπει να ελέγχεται μέσω της θέσπισης πρωτοκόλλων που επιτρέπουν την πρόσβαση στα δεδομένα. Τα εν λόγω πρωτόκολλα πρέπει να περιγράφουν ποιος μπορεί να έχει πρόσβαση στα δεδομένα και

υπό ποιες συνθήκες. Μόνο το εξειδικευμένο προσωπικό, με την κατάλληλη αρμοδιότητα, και την ανάγκη πρόσβασης στα δεδομένα ενός ατόμου επιτρέπεται να τα προσπελάζει.

2.2.26 Διαφάνεια

I) Ιχνηλασιμότητα

Η συλλογή δεδομένων, η επισήμανσή τους, οι αλγόριθμοι που χρησιμοποιούνται, τα σύνολα δεδομένων και οι διαδικασίες που οδηγούν στην απόφαση ενός συστήματος TN, πρέπει να τεκμηριώνονται με τα καλύτερα δυνατά πρότυπα ώστε να είναι δυνατή η ιχνηλασιμότητα και η αύξηση της διαφάνειας. Το ίδιο πρέπει να ισχύει και για τις αποφάσεις που λαμβάνει το σύστημα TN. Έτσι, καθίσταται δυνατό να εντοπιστούν οι λόγοι για τους οποίους μια απόφαση είναι εσφαλμένη, γεγονός που με την σειρά του μπορεί να συμβάλλει στην πρόληψη μελλοντικών σφαλμάτων. Διευκολύνεται μέσω της ιχνηλασιμότητας η δυνατότητα ελέγχου και η επεξήγηση των συστημάτων TN.

II) Εξηγησιμότητα

Αφορά την ικανότητα εξήγησης τόσο τεχνικών διαδικασιών ενός συστήματος TN όσο και των σχετικών ανθρώπινων αποφάσεων. Η τεχνική εξηγησιμότητα απαιτεί ότι οι αποφάσεις που λαμβάνονται από ένα σύστημα TN μπορούν να είναι κατανοητές από τον άνθρωπο. Επιπροσθέτως, ενδέχεται να γίνουν συμβιβασμοί μεταξύ της βελτίωσης της εξηγησιμότητας ενός συστήματος και της αύξησης της ακρίβειάς του, καθώς είναι ποσά αντιστρόφως ανάλογα. Όταν ένα σύστημα TN έχει αντίκτυπο στην ζωή των ανθρώπων πρέπει να είναι δυνατόν να ζητηθεί η διαδικασία λήψης αποφάσεών του. Η εξήγηση πρέπει να είναι έγκαιρη, προσαρμοσμένη στην εμπειρία του ενδιαφερόμενου ανάλογα με τις γνώσεις του και την εξειδίκευσή του στο αντικείμενο. Επιπλέον, οι εξηγήσεις του βαθμού στον οποίο επηρεάζει ένα σύστημα TN την διαδικασία λήψης απόφασης σε έναν οργανισμό, τις επιλογές σχεδιασμού του συστήματος και το σκεπτικό ανάπτυξής του, πρέπει να είναι διαθέσιμες. Έτσι, διασφαλίζεται η διαφάνεια του επιχειρηματικού μοντέλου.

III) Επικοινωνία

Τα συστήματα TN δεν πρέπει να παρουσιάζουν τον εαυτό τους σαν ανθρώπους στους χρήστες. Οι άνθρωποι έχουν το δικαίωμα να πληροφορηθούν πως αλληλεπιδρούν με ένα

τέτοιο σύστημα. Αυτό συνεπάγεται πως τα συστήματα TN πρέπει να ταυτοποιούνται για την πραγματική τους υπόσταση. Επιπροσθέτως, η επιλογή αλληλεπίδρασης με ένα σύστημα TN ή με άνθρωπο πρέπει να διασφαλίζεται προκειμένου να υπάρχει συμμόρφωση με τα θεμελιώδη ανθρώπινα δικαιώματα. Εκτός αυτού, οι δυνατότητες και τα όρια ενός συστήματος TN πρέπει να επικοινωνούνται στους ειδικούς TN και στους τελικούς χρήστες με ανάλογο τρόπο για την εκάστοτε περίπτωση. Κάτι τέτοιο, συμπεριλαμβάνει και την επικοινωνία της ακρίβειας μια λύσης TN.

2.2.27 Ποικιλομορφία, κατά Διακρίσεων, Δικαιοσύνη

I) Αποφυγή Αθέμιτης Προκατάληψης

Συχνά τα συστήματα TN χρησιμοποιούν σύνολα δεδομένων, τόσο για την εκπαίδευση, όσο και για την λειτουργία τους, τα οποία πάσχουν από ακούσια ιστορική μεροληψία, ατέλεια ή λανθασμένα μοντέλα διακυβέρνησης. Η διαιώνιση τέτοιων προκαταλήψεων μπορεί να οδηγήσει σε ακούσιες διακρίσεις ή μη, εις βάρος ορισμένων ομάδων και ατόμων ενισχύοντας έτσι την περιθωριοποίησή τους. Μπορεί επίσης να προκληθεί βλάβη από την επί τούτου εκμετάλλευση της καταναλωτικής μεροληψίας, όπως η ομογενοποίηση των τιμών συμπαιγνιών ή της μαύρης αγοράς. Οι μεροληπτικές διακρίσεις οι οποίες μπορούν να αναγνωριστούν πρέπει να εξαλείφονται στην φάση της συλλογής των δεδομένων όταν είναι δυνατό.

Επίσης, κατά την ανάπτυξη ενός συστήματος TN, για παράδειγμα στην φάση προγραμματισμού αλγορίθμων, είναι πιθανό να εμφανιστούν φαινόμενα προκατάληψης. Γεγονός που μπορεί να περιοριστεί με την θέσπιση διαδικασιών εποπτείας για την ανάλυση του τρόπου με τον οποίο λαμβάνονται οι αποφάσεις.

Επιπλέον, όταν προσλαμβάνονται άτομα από διαφορετικά υπόβαθρα και κλάδους, μπορεί να διασφαλιστεί η ποικιλομορφία των δεδομένων, κουλτούρα που πρέπει να ενθαρρυνθεί.

II) Προσβασιμότητα και Καθολικός Σχεδιασμός

Ειδικά σε τομείς όπως Επιχείρηση-προς-Καταναλωτή, τα συστήματα TN πρέπει να επικεντρώνονται στον χρήστη και να σχεδιάζονται με τέτοιο τρόπο ώστε οι άνθρωποι να έχουν πρόσβαση σε αυτά ανεξάρτητα από την ηλικία, το φύλο, τις ικανότητες και τα

χαρακτηριστικά τους. Η πρόσβαση στην τεχνολογία της ΤΝ για άτομα με αναπηρία, τα οποία υπάρχουν σε όλες τις κοινωνικές ομάδες, έχει ιδιαίτερη σημασία. Τα συστήματα ΤΝ δεν πρέπει να σχεδιάζονται με την προσέγγιση του να ταιριάζουν σε όλους, αλλά πρέπει να ληφθεί υπόψιν καθολικός σχεδιασμός ο οποίος απευθύνεται στο ευρύτερο δυνατό φάσμα χρηστών, ακολουθώντας τα σχετικά πρότυπα προσβασιμότητας. Γεγονός που επιτρέπει την ισότιμη πρόσβαση και την ενεργή συμμετοχή όλων των ανθρώπων στις υπάρχουσες και στις αναδυόμενες δραστηριότητες που σχετίζονται με την διαμεσολάβηση υπολογιστή για υποστηρικτικές ως προς τον άνθρωπο τεχνολογίες.

III) Συμμετοχή Ενδιαφερόμενων Μερών

Για την ανάπτυξη αξιόπιστων συστημάτων ΤΝ, συνίσταται η διαβούλευση των ενδιαφερόμενων μερών τα οποία ενδέχεται να επηρεαστούν άμεσα ή έμμεσα από το σύστημα καθ' όλη την διάρκεια του κύκλου ζωής του. Είναι ωφέλιμο να υπάρχει τακτική ανατροφοδότηση ακόμα και μετά το πέρας της ανάπτυξης ενός συστήματος ΤΝ και να δημιουργούνται μηχανισμοί συμμετοχής των ενδιαφερόμενων μερών, διασφάλισης της ενημέρωσης και της συμμετοχής των εργαζομένων σε όλη την διαδικασία εφαρμογής συστημάτων ΤΝ σε οργανισμούς.

2.2.28 Κοινωνική και Περιβαλλοντολογική Ευεξία

I) Βιώσιμη και Περιβαλλοντολογικά Φιλική ΤΝ

Τα συστήματα ΤΝ υπόσχονται να βοηθήσουν στην αντιμετώπιση ορισμένων εκ των πιο πιεστικών κοινωνικών προβλημάτων, ωστόσο πρέπει να διασφαλιστεί ότι αυτό γίνεται με τον πιο φιλικό προς το περιβάλλον τρόπο. Η διαδικασία ανάπτυξης, εγκατάστασης και χρήσης, όπως και όλη η αλυσίδα εφοδιασμού ενός συστήματος ΤΝ, πρέπει να αξιολογείται μέσω μιας κριτικής εξέτασης της χρήσης πόρων και της κατανάλωσης ενέργειας κατά την διάρκεια εκπαίδευσης, επιλέγοντας φυσικά τις λιγότερο επιβλαβείς επιλογές. Μέτρα διασφάλισης της φιλικότητας προς το περιβάλλον ολόκληρου του κύκλου εργασιών των συστημάτων ΤΝ πρέπει να ενθαρρύνονται.

II) Κοινωνικός Αντίκτυπος

Η πανταχού παρούσα έκθεση σε κοινωνικά συστήματα ΤΝ σε όλους τους τομείς της ζωής μας, είτε πρόκειται για την εκπαίδευση, την εργασία, την περίθαλψη ή την ψυχαγωγία,

μπορεί να αλλάξει την αντίληψή μας για την κοινωνική δράση και να επηρεάσει τις κοινωνικές μας σχέσεις. Ενώ τα συστήματα ΤΝ μπορούν να χρησιμοποιούνται με τέτοιο τρόπο ώστε να ενισχύονται οι κοινωνικές δεξιότητες των ανθρώπων, μπορούν εξίσου να συμβάλλουν στην υποβάθμισή τους. Γεγονός που μπορεί επίσης, να επηρεάσει την σωματική και την ψυχική ευεξία των ανθρώπων. Οι επιπτώσεις των συστημάτων ΤΝ χρειάζονται συνεχή παρακολούθηση.

III) Κοινωνία και Δημοκρατία

Πέρα από την αξιολόγηση του αντικτύπου της ανάπτυξης και της χρήσης ενός συστήματος ΤΝ, πρέπει να αξιολογείται και η κοινωνική χροιά του, λαμβάνοντας υπόψιν την επίδρασή του στους θεσμούς, την δημοκρατία και την κοινωνία στο σύνολό της. Η χρήση συστημάτων ΤΝ πρέπει να εξετάζεται προσεκτικά, ιδίως σε καταστάσεις που σχετίζονται με την δημοκρατική διαδικασία, συμπεριλαμβανομένων των πολιτικών αποφάσεων και των εκλογικών πλαισίων.

2.2.29 Λογοδοσία

I) Ελεγκσιμότητα

Η δυνατότητα του ελέγχου συνεπάγεται την αξιολόγηση αλγορίθμων, δεδομένων και διαδικασιών σχεδιασμού. Αυτό δεν σημαίνει πως οι πληροφορίες σχετικά με τα επιχειρηματικά μοντέλα και την πνευματική ιδιοκτησία που σχετίζονται με το σύστημα ΤΝ πρέπει να είναι ευρέως διαθέσιμες. Στην αξιοπιστία της τεχνολογίας μπορεί να συμβάλλει η αξιολόγηση από εσωτερικούς και εξωτερικούς ελεγκτές, ταυτόχρονα με την διάθεση εκθέσεων αξιολόγησης των εν λόγω επαγγελματιών. Σε εφαρμογές που επηρεάζουν θεμελιώδη δικαιώματα, τα συστήματα ΤΝ πρέπει να μπορούν να ελέγχονται ανεξάρτητα.

II) Ελαχιστοποίηση και Αναφορά Αρνητικών Επιπτώσεων

Πρέπει να διασφαλίζεται η δυνατότητα αναφοράς αποφάσεων που συμβάλλουν σε ένα ορισμένο αποτέλεσμα συστήματος. Ο εντοπισμός, η αξιολόγηση και η τεκμηρίωση των πιθανών αρνητικών επιπτώσεων των συστημάτων ΤΝ είναι ιδιαίτερα κρίσιμα συστατικά ελαχιστοποίησης αρνητικών επιπτώσεων για τους άμεσα εμπλεκόμενους. Η δέουσα προστασία χρειάζεται να υπάρχει για τους καταγγέλλοντες ή άλλες οντότητες όταν αναφέρουν νόμιμες ανησυχίες σχετικά με ένα σύστημα ΤΝ. Η χρήση αξιολογήσεων

επίπτωσης, όπως η «κόκκινη ομάδα⁴» ή άλλες μορφές αλγοριθμικής εκτίμησης επιπτώσεων, τόσο πριν όσο και κατά την διάρκεια της ανάπτυξης και της χρήσης των συστημάτων TN, είναι χρήσιμη για την ελαχιστοποίηση των αρνητικών επιπτώσεων. Οι αξιολογήσεις αυτές πρέπει να είναι ανάλογες με το κίνδυνο που ενέχει η εκάστοτε τεχνολογία.

III) Συμβιβασμοί

Κατά την εφαρμογή των παραπάνω απαιτήσεων, ενδέχεται να προκύψουν σημεία τριβής, τα οποία μπορεί να οδηγήσουν σε αναπόφευκτους συμβιβασμούς. Τέτοιοι συμβιβασμοί χρειάζεται να αντιμετωπίζονται με ορθολογικό τρόπο και μεθοδολογία στο πλαίσιο τεχνολογιών αιχμής. Κατ' επέκταση τα συμφέροντα, οι αξίες και οι συμβιβασμοί που τίθενται από ένα σύστημα TN πρέπει να αναγνωρίζονται ρητά και να αξιολογούνται υπό το πρίσμα του κινδύνου που θέτουν ως προς τις ηθικές αρχές και τα θεμελιώδη δικαιώματα. Σε καταστάσεις στις οποίες δεν μπορούν να εντοπιστούν ηθικά αποδεκτά μέτρα συμβιβασμού, η ανάπτυξη και η χρήση του συστήματος TN δεν πρέπει να προχωρήσει στην τρέχουσα-εκάστοτε μορφή. Οι αποφάσεις σχετικά με τον συμβιβασμό πρέπει να είναι αιτιολογημένες και τεκμηριωμένες κατάλληλα. Επίσης, ο υπεύθυνος λήψης αποφάσεων πρέπει να είναι υπόλογος για τον τρόπο με τον οποίο γίνεται ο κάθε συμβιβασμός και πρέπει να επανεξετάζει συνεχώς την καταλληλότητα της απόφασης που προκύπτει, ώστε να διασφαλίζεται ότι μπορούν να γίνουν οι απαραίτητες αλλαγές στο σύστημα.

IV) Αποκατάσταση

Όταν προκύπτουν δυσμενείς επιπτώσεις, πρέπει να προβλέπονται μηχανισμοί πρόσβασης που εξασφαλίζουν επαρκή αποκατάσταση. Η γνώση πως η αποκατάσταση είναι εφικτή, είναι το κλειδί για την διασφάλιση της εμπιστοσύνης προς το σύστημα TN. Ιδιαίτερη προσοχή πρέπει να δοθεί στα ευάλωτα άτομα ή ομάδες^[1,2,3,4].

⁴ Η κοινότητα της TN υιοθέτησε την «κόκκινη ομάδα» από την ασφάλεια στον κυβερνοχώρο, όπου οι επαγγελματίες της κόκκινης ομάδας προσπαθούν να επιτεθούν ή να αποκτήσουν πρόσβαση σε μια φυσική τοποθεσία ή ένα δίκτυο υπολογιστών, ενώ οι επαγγελματίες της μπλε ομάδας προσπαθούν να αμυνθούν ενάντια στην εισβολή^[14].

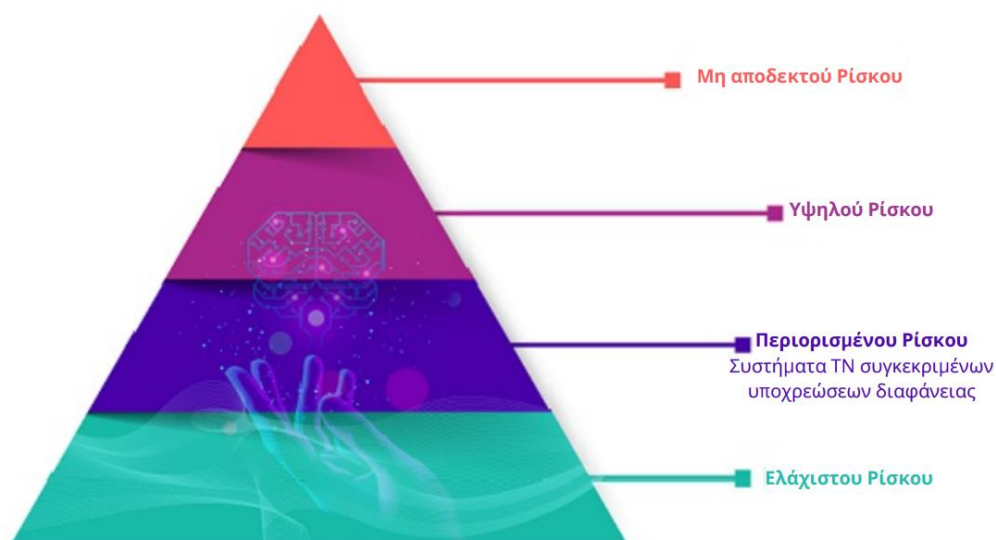
2.2.30 AI Act – AI Νομικό Πλαίσιο Τεχνητής Νοημοσύνης

Το AI Νομικό Πλαίσιο, είναι το πρώτο νομικό πλαίσιο όσον αφορά την Τεχνητή Νοημοσύνη, το οποίο διαχειρίζεται ρίσκα της TN και εισάγει τις θέσεις της Ευρώπης ώστε να διενεργήσει ηγετικό ρόλο παγκοσμίως.

Οι προτεινόμενοι κανόνες είναι οι εξής:

- I) Διαχείριση ρίσκων που δημιουργούνται από εφαρμογές TN
- II) Απαγόρευση TN πρακτικών οι οποίες θέτουν ανεπίτρεπτα ρίσκα
- III) Καθορισμός λίστας εφαρμογών υψηλού ρίσκου
- IV) Θέσπιση ξεκάθαρων απαιτήσεων για συστήματα TN υψηλού ρίσκου
- V) Καθορισμός συγκεκριμένων επιταγών για τους δημιουργούς και τους παρόχους TN εφαρμογών υψηλού ρίσκου
- VI) Απαίτηση αξιολόγησης συμβατότητας προτού ένα σύστημα TN τίθεται σε λειτουργία ή τοποθετείται στην αγορά
- VII) Επιβολή της νομοθεσίας μετά τη διάθεση ενός συγκεκριμένου συστήματος TN στην αγορά
- VIII) Εγκαθίδρυση δομής διακυβέρνησης σε Ευρωπαϊκό και σε Διεθνές επίπεδο

Προσέγγιση με βάση το ρίσκο - το AI Νομικό Ρυθμιστικό Πλαίσιο καθορίζει τα τέσσερα επίπεδα για τα συστήματα ΤΝ:



Εικόνα 3: AI Act, το Πρώτο Νομικό Πλαίσιο ΤΝ^[15]

Όλα τα συστήματα ΤΝ τα οποία θέτουν ξεκάθαρο κίνδυνο στις ζωές των ανθρώπων, στην ασφάλεια και στα δικαιώματά τους απαγορεύονται. Οι πτυχές των συστημάτων ΤΝ εξετάζονται και αξιολογούνται. Παραδείγματα τέτοιων πτυχών αποτελούν η κοινωνική βαθμολόγηση από τις κυβερνήσεις και οι βοηθοί φωνής που ενθαρρύνουν την επικίνδυνη συμπεριφορά^[15].

Κεφάλαιο 3: Ερευνητικό Πλαίσιο Νομικών και Ηθικών Λύσεων Τεχνητής Νοημοσύνης

3.1 Ερευνητικές Δημοσιεύσεις Ηθικών Λύσεων Τεχνητής Νοημοσύνης στην Ναυτιλία

Η εμπορική και οικονομική σημασία της ναυτιλιακής βιομηχανίας είναι αδιαμφισβήτητη, καθώς αποτελεί το μέσο για το 90% της μεταφοράς των αγαθών που ανταλλάσσονται παγκοσμίως. Με την συνεχή πρόοδο στο τομέα των αισθητήρων και των αλγορίθμων υποστήριξης απόφασης, νέες αυτόνομες εφαρμογές αντικαθιστούν τις παραδοσιακές ναυτικές τεχνολογίες καθώς προσφέρουν χαμηλότερα λειτουργικά έξοδα, μικρότερα κόστη συντήρησης με αποτέλεσμα την εκπομπή λιγότερων ρύπων στην ατμόσφαιρα και την μεγαλύτερη ασφάλεια των πλοίων και του κλάδου γενικότερα.

Παρά το γεγονός ότι οι άνθρωποι είναι σε θέση να αντιμετωπίσουν σύνθετες καταστάσεις λήψης απόφασης καλύτερα από τις μηχανές, τα ανθρώπινα λάθη στην λήψη απόφασης κατά την ναυσιπλοΐα είναι η πρώτη αιτία ατυχημάτων. Μπορεί να προκληθούν από κούραση ή υπερβολικό φόρτο εργασίας. Δεδομένου ότι η ανθρώπινη ικανότητα επεξεργασίας πολλών δεδομένων είναι περιορισμένη, ένας απαραίτητος παράγοντας αποτροπής ατυχημάτων είναι μια κατάλληλα διαμορφωμένη - ανθρώπινη και μηχανική - διεπαφή η οποία υποστηρίζει την αποτροπή συγκρούσεων, την επιλογή βέλτιστης διαδρομής, την βελτιστοποίηση καυσίμων και εκπομπών Διοξειδίου του Άνθρακα - CO₂ στην συμβατική ναυτιλία και συνολικά την υποστήριξη απόφασης. Η μετάβαση προς την αυτόνομη ναυτιλία αποτελεί ανατρεπτική καινοτομία στην εξέλιξη της ναυτιλίας και υπάρχει αυξανόμενο ενδιαφέρον στην ανάπτυξη αυτόνομων πλοίων. Πολλοί τεχνολόγοι ανά τον κόσμο προσπαθούν να αναπτύξουν, να υιοθετήσουν και να εφαρμόσουν στην πράξη νέες τεχνολογίες προκειμένου η αυτόνομη ναυτιλία να βελτιωθεί περισσότερο^[16].

3.1.1 Ψηφιακοί Πλοηγοί - ΨΠ

Η βιομηχανία της ναυτιλίας υιοθετεί τεχνολογία η οποία έχει δημιουργηθεί με βάση την ΤΝ, δηλαδή, αλγορίθμους Μηχανικής Μάθησης για αυτόνομα πλοία.

Παράδειγμα τεχνολογίας προς αυτή την κατεύθυνση είναι οι Ψηφιακοί Πλοηγοί - ΨΠ που αναπτύσσονται με χρήση Βαθών Νευρωνικών Δικτύων τα οποία είναι εκπαιδευμένα με σύνολα δεδομένων προηγούμενων ταξιδιών. Προς το παρόν δεν έχει δημιουργηθεί ένα συγκεκριμένο πλαίσιο αξιολόγησης της αξιοπιστίας τους. Υφίστανται ωστόσο Βασικοί Δείκτες Επιδόσεων – ΒΔΕ ή Key Performance Indicators - KPIs οι οποίοι αξιολογούν την αξιοπιστία ενός συστήματος ΤΝ όπως οι ψηφιακοί πλοηγοί^[16].

Η αξιοπιστία συστημάτων ΤΝ στην ναυτιλία, συμπεριλαμβανομένων των ψηφιακών πλοηγών, μπορεί να μελετηθεί σε επίπεδο λογισμικού και σε επίπεδο υλικού. Κάθε επίπεδο πρέπει να έχει συγκεκριμένα χαρακτηριστικά για να είναι αξιόπιστο. Με άλλα λόγια, ο κώδικας του λογισμικού πρέπει να είναι Διαυγής, Εξηγήσιμος, Δίκαιος και Υπεύθυνος. Επιπλέον, η αξιοπιστία σε επίπεδο υλικού μπορεί να ελεγχθεί υπό το πρίσμα της Ανθεκτικότητας και της Διαθεσιμότητας σχετικών συστημάτων και τεχνολογιών. Επιπρόσθετα, χαρακτηριστικά όπως η Αξιοπιστία, η Ιδιωτικότητα, η Ασφάλεια και η Προστασία πρέπει να μελετώνται.

ALTAI Ερωτηματολόγιο: Ανθρώπινη Εκπροσώπηση και Εποπτεία

Οι ψηφιακοί πλοηγοί αναπτύσσονται με την εξής λογική: μια σειρά γεγονότων τροφοδοτείται στο νευρωνικό δίκτυο και το δίκτυο μαθαίνει και προσομοιώνει την σωστή συμπεριφορά σε διαφορετικά σενάρια. Στο μέλλον αποφασίζει την δράση, ανάλογα με την ομοιότητα μια κατάστασης σε μεμονωμένο ή συνδυασμό παρελθοντικών σεναρίων. Έτσι, ένα πλαίσιο εξηγησιμότητας κατάλληλο για Βαθιά Νευρωνικά Δίκτυα πρέπει να επιλεγεί για τον ψηφιακό πλοηγό. Σε αυτό, προτείνεται να υπάρχουν οι μεταγενέστερες τοπικές εξηγήσεις και τα χαρακτηριστικά τεχνικής συνάφειας οι οποίες είναι οι πιο κατάλληλες και ευρέως υιοθετημένες μέθοδοι εξήγησης για ΒΝΔ. Το πλαίσιο αξιολόγησης κατηγοριοποιείται στα Πολυ-Επίπεδα Νευρωνικά Δίκτυα - ΠΕΝΔ ή Multi-Layer Neural Networks – MLNN, στα Συνελικτικά Νευρωνικά Δίκτυα - ΣΝΔ ή Convolutional Neural Networks – CNN, και στα Επαναληπτικά Νευρωνικά Δίκτυα - ΕΝΔ ή Recursive Convolutional Networks – RCN^[16].

Πολυ-Επίπεδα Νευρωνικά Δίκτυα - ΠΕΝΔ

Ένα ικανοποιητικό κριτήριο εξηγησιμότητας των ΠΕΝΔ που μπορεί να εφαρμοστεί είναι να αποσυντεθεί η ταξινόμηση δικτυακών αποφάσεων σε «κομμάτια» των εισόδων του.

Σε αυτή τη μέθοδο η μαθηματική λογική και οι αλγεβρικές εξισώσεις κάθε νευρώνα συνυπολογίζονται και κάθε νευρώνας είναι αντικείμενο που μπορεί να αποσυντεθεί και να επεκταθεί. Έτσι, οι αποσυνθέσεις διαδίδονται προς τα πίσω μέσω δικτύου από την έξοδο στην είσοδο. Με αυτό το τρόπο, εξάγεται η σχέση μεταξύ εισόδου-εξόδου.

Συνελικτικά Νευρωνικά Δίκτυα - ΣΝΔ

Τα ΣΝΔ χρησιμοποιούνται εκτενώς για εφαρμογές επεξεργασίας εικόνων οι οποίες έχουν κριτική σημασία στην ανάπτυξη ψηφιακής πλοήγησης. Αυτά τα δίκτυα αποτελούνται από μια ακολουθία συνελικτικών δικτύων τα οποία μπορούν να εκπαιδευτούν αυτόματα σε επίπεδο υψηλών προδιαγραφών. Η εξηγησιμότητα για τα ΣΝΔ είναι πιο ευθύς συγκριτικά με τους άλλους τύπους δικτύων. Βρίσκουν διαφορετικά χαρακτηριστικά ή αντικείμενα σε κάθε εικόνα και η λειτουργία τους βασίζεται στην μίμηση της ικανότητας του ανθρώπινου νου, με αποτέλεσμα να είναι ευκολότερο να εξηγηθεί η απόδοση του κάθε επιπέδου. Οι προσεγγίσεις κατανόησης των ΣΝΔ κατηγοριοποιούνται:

- I) Κατανόηση της διαδικασίας αντιστοίχισης εξόδου - εισόδου για να εντοπιστούν τα μέρη της εισόδου που συμβάλλουν στην έξοδο.
- II) Ερμηνεία για το πως τα επίπεδα λειτουργούν γενικότερα και όχι απαραίτητα συσχετιζόμενα με οποιαδήποτε συγκεκριμένη είσοδο.

Επαναληπτικά Νευρωνικά Δίκτυα - ΕΝΔ

Τα ΕΝΔ χρησιμοποιούνται ευρέως για προβλεπτικές εφαρμογές ακολουθιακών δεδομένων και ανάλυση χρονοσειρών. Σε έναν ψηφιακό πλοηγό, η πορεία ενός πλοίου αποτελείται από γεωγραφικά μήκη και πλάτη στη μορφή σειρών σημειακών δεδομένων. Μια σειρά γεωγραφικών συντεταγμένων καθορίζει το μονοπάτι που ένα επιλεγμένο πλοίο έχει πάρει και το μελλοντικό μονοπάτι μπορεί να προβλεφθεί χρησιμοποιώντας ιστορικά δεδομένα θέσης. Τέτοιες πληροφορίες είναι χρήσιμες για κοντινές συναντήσεις πλοίων και κινήσεις αποτροπής συγκρούσεων. Σύνολα δεδομένων χρονοσειρών επιδεικνύουν μακροπρόθεσμες αλληλεξαρτήσεις οι οποίες είναι πολύ σύνθετες να αναλυθούν από έναν αλγόριθμο MM. Από την άλλη, τα ΕΝΔ διαχειρίζονται αλληλεξαρτήσεις αποθηκεύοντας γνώση στους αντίστοιχους νευρώνες. Υπάρχουν δύο βασικά είδη ΕΝΔ μοντέλων:

I) Κατανόηση της γνώσης ενός ΕΝΔ μοντέλου από την χρονοσειρά εισόδου μέσω της ομοιότητας χαρακτηριστικών μεθόδων.

II) Τροποποίηση ΕΝΔ αρχιτεκτονικών - μια την φορά - προκειμένου να εξαχθούν συμπεράσματα για την διαδικασία λήψης απόφασης σαν μια προσεγγιστική τοπική εξήγηση.

Στην ανάπτυξη ΨΠ, όλα αυτά τα είδη Νευρωνικών Δικτύων χρησιμοποιούνται για διαφορετικούς σκοπούς, όπως η πρόβλεψη πορείας, η επεξεργασία εικόνων, η λήψη απόφασης, συνυπολογίζοντας βέβαια την σύνθετη δομή των πλοηγών. Έτσι, ο συνδυασμός όλων αυτών των μεθόδων οδηγεί σε μια καλή αξιολόγηση της εξηγησιμότητας ενός ψηφιακού πλοηγού.

ALTAI Ερωτηματολόγιο: Δικαιοσύνη

Είναι ιδιαίτερα σημαντικό ότι οι προτάσεις και οι αποφάσεις οι οποίες λαμβάνονται από εφαρμογές ή συστήματα δεν είναι θετικά προκατειλημμένες απέναντι σε κάποια κοινωνική ομάδα, και η διαδικασία επιλογής για την λήψη απόφασης δεν δρα διαχωριστικά κατά συγκεκριμένων χαρακτηριστικών.

Στους ψηφιακούς πλοηγούς, τα επίπεδα του κύκλου ζωής του συστήματος ΤΝ διαμορφώνονται ως το επίπεδο σκοπού, επίπεδο επιλογής δεδομένων, επίπεδο προεπεξεργασίας, επίπεδο αλγορίθμου, επίπεδο εκπαίδευσης, ανεξάρτητο επίπεδο ελέγχου και επίπεδο χρήσης.

Αφού εισαχθεί ένα σύστημα ΤΝ στον τελικό χρήστη – ναυτικό, είναι σημαντικό να ελεγχθεί το πόσο δίκαιο είναι και να εκπαιδευτεί με νέα δεδομένα προκειμένου να αποφεύγεται τυχόν προκατάληψη στα δεδομένα και κατ' επέκταση στο σύστημα ΤΝ. Επιπλέον, ο δείκτης επίδοσης του ανεπτυχθέντος συστήματος πρέπει να ελέγχεται περιοδικά προκειμένου να καθίσταται σαφές πότε το σύστημα χρειάζεται επανεκπαίδευση. Αυτή η προσέγγιση μπορεί να είναι πιο αποδοτική συγκριτικά με τις περισσότερες μελέτες οι οποίες επικεντρώνονται μόνο σε προκαταλήψεις συνόλων δεδομένων και αλγορίθμων, καθώς προτείνεται μια μέθοδος πλήρους κατανόησης των σταδίων ανάπτυξης ΤΝ, από την ιδέα μέχρι την υλοποίηση. Δεδομένου ότι τα στάδια που περιλαμβάνονται για την κατασκευή μοντέλων ΤΝ είναι συχνά παρόμοια, είναι ο ίδιος λόγος που η συγκεκριμένη

μέθοδος μπορεί να εφαρμοστεί στους ΨΠ. Διασφαλίζοντας έτσι, την δικαιοσύνη σε κάθε στάδιο της ανάπτυξης καθώς κάθε προκατάληψη μπορεί να αποφευχθεί σε ικανοποιητικό βαθμό.

ALTAI Ερωτηματολόγιο: Λογοδοσία

Μια διαδικασία λογοδοσίας με νόημα απαρτίζεται από τρεις φάσεις: πληροφορία, εξήγηση ή δικαιολόγηση και πιθανές συνέπειες. Υπάρχουν προτεινόμενες αρχές σχετικά με την κοινωνικο-τεχνική σκοπιά σχεδιασμού ενός υπεύθυνου συστήματος TN μέσω ρυθμιστικών κανονισμών και λειτουργικού συντονισμού. Ο έλεγχος απόκλισης, η οριακή τοποθεσία και η δύναμη της εξουσίας αποτελούν τις αρχές που σχετίζονται με την υπευθυνότητα μέσω ρυθμιστικών κανονισμών. Η υπευθυνότητα μέσω του λειτουργικού συντονισμού αποτελείται από τέσσερις αρχές: ενσωμάτωση του πυρήνα της διαδικασίας, κατανομή εργασίας σε ανθρώπους και μηχανές, συμφωνία και ανάληψη ευθύνης από τους διαχειριστές και τους χρήστες.

Οι Ψηφιακοί Πλοηγοί^[16] κατατάσσονται σε συστήματα TN με περιορισμένο ρίσκο καθώς με συγκεκριμένες επιταγές διαφάνειας μπορεί να μην θέτουν ξεκάθαρο κίνδυνο στις ζωές των ανθρώπων, στην ασφάλεια και στα δικαιώματά τους εφόσον φυσικά τηρούνται οι απαιτήσεις αξιοπιστίας.

3.1.2 Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ

Μια επιπλέον περιοχή ενδιαφέροντος για ζητήματα δεοντολογίας συστημάτων TN στην ναυτιλία αποτελεί η Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ ή Maritime Integrated Surveillance Awareness – MARISA το οποίο είναι ένα έργο που ασχολείται κατά κύριο λόγο με δεδομένα θαλάσσης, τα δεδομένα αποτελούν την βάση συστημάτων TN. Το έργο ΘΟΕΕ άπτεται θεμάτων επιτήρησης θαλάσσης, εφαρμογής του νόμου, ιδιωτικότητας δεδομένων και έχει τέσσερις βασικούς στόχους^[19]:

- I) Να δημιουργήσει βελτιωμένη επίγνωση και επίβλεψη της κατάστασης με βασικό στόχο να παρέχει ολοκληρωμένη και χρήσιμη εικόνα για την κατάσταση στην θάλασσα.
- II) Να υποστηρίξει τους επαγγελματίες καθ' όλο τον κύκλο ζωής καταστάσεων στην θάλασσα, από την παρακολούθηση στοιχείων στο περιβάλλον μέχρι τον εντοπισμό ανωμαλιών και την βοήθεια στην δημιουργία πλάνου διαχείρισης.

III) Να ενθαρρύνει την αποδοτική συνεργασία μεταξύ γειτονικών και διασυνοριακών αρχών που λειτουργούν στην σφαίρα της θαλάσσιας επιτήρησης. Παραδείγματα σωμάτων που χρειάζονται την εύρυθμη συνεργασία είναι το ναυτικό, η ακτοφυλακή, το τελωνείο και η συνοριακή αστυνομία προκειμένου να αντλούν κοινούς πόρους για κοινούς στόχους, οδηγώντας έτσι σε αποδοτική χρήση των ήδη υπάρχοντων πόρων.

IV) Να τροφοδοτεί την δημιουργία ενός δυναμικού οικοσυστήματος χρηστών και παρόχων, επιτρέποντας νέες υπηρεσίες συγχώνευσης δεδομένων, βασισμένες σε «αποσταγμένη» γνώση, να είναι προσβάσιμες σε διαφορετικές ναυτικές οντότητες στην θάλασσα μέσω της ενσωμάτωσης ενός ευρέος πεδίου δεδομένων και αισθητήρων.

Πίνακας 1: Περιοχές Προτεραιότητας Ευρωπαϊκού Κοινοβουλίου και Θαλάσσιας Ολοκληρωμένης Επίγνωσης της Επιτήρησης – ALTAI Ερωτηματολόγιο: Ιδιωτικότητα και Διακυβέρνηση δεδομένων

Περιοχή Προτεραιότητας	Περιεχόμενα	Πτυχές Θαλάσσιας Ολοκληρωμένης Επίγνωσης της Επιτήρησης
<p>Ελευθερία, Ασφάλεια και Δικαιοσύνη.</p> <p>«Το Ευρωπαϊκό Κοινοβούλιο δίνει έμφαση στην σημασία μιας καλής συνεργασίας σε θέματα ασφάλειας όπως η τρομοκρατία και η διαχείριση μεταναστευτικών ροών.»</p>	<p>Καλύτερη διαχείριση όλων των πτυχών μετανάστευσης, συμπεριλαμβανομένων ασύλου και διαχείριση συνόρων.</p> <p>Αποτροπή και πάταξη οργανωμένου εγκλήματος, διαφθοράς και τρομοκρατίας βελτιώνοντας την δικαστική συνεργασία μεταξύ Ευρωπαϊκών χωρών.</p>	<p>Η ιδιωτικότητα είναι αυστηρά συνδεδεμένη με την ελευθερία, και σε μια κοινωνία όπου κάθε κίνηση και δράση καταγράφεται θεωρείται αντιφατική πραγματικότητα στην ιδέα της ελευθερίας. Στο πλαίσιο της θαλάσσιας επιτήρησης είναι σημαντικό να προστατευθεί η αρχή της «ελεύθερης ναυσιπλοΐας».</p> <p>Αυξανόμενος έλεγχος και μέτρα ασφαλείας είναι δικαιολογημένος λόγω της ανάγκης προστασίας της Ευρώπης από το διασυνοριακό έγκλημα, όπως η εμπορία ανθρώπων και το λαθρεμπόριο. Το Ευρωπαϊκό θαλάσσιο σύνορο δεν είναι θέμα ασφάλειας μόνο για την Ευρώπη αλλά και για εκείνους που αναζητούν να εισέλθουν στην Ευρώπη μέσω θαλάσσης.</p>

<p>Η Ευρωπαϊκή Ένωση είναι ένας δυνατός παγκόσμιος παίκτης.</p> <p>«Το Ευρωπαϊκό Κοινοβούλιο καλεί την Ευρωπαϊκή Ένωση να διασφαλίσει την ισχυρή της δέσμευση σε παγκόσμια ζητήματα.»</p>	<p>Διασφαλίζοντας συνέπεια μεταξύ των Μελών Κρατών και των στόχων εξωτερικής πολιτικής της Ευρωπαϊκής Ένωσης.</p> <p>Πρωθώντας σταθερότητα, ευημερία και δημοκρατία στις γειτονικές χώρες της Ευρώπης.</p> <p>Παρακινώντας τους διεθνείς συμμάχους σε ένα εύρος ζητημάτων όπως το εμπόριο, η κυβερνοασφάλεια, τα ανθρώπινα δικαιώματα και η διαχείριση κρίσεων.</p>	<p>Στο πλαίσιο της θαλάσσιας επιτήρησης, η έλλειψη αξιοπιστίας, οι ξεκάθαρες γραμμές ευθύνης μεταξύ των Ευρωπαϊκών Κρατών Μελών και οι διαφορετικοί τους ρόλοι είναι ένα επίμονο πρόβλημα.</p> <p>Επιπλέον, οι αποκλίνουσες ερμηνείες διεθνών νομικών κανόνων υποδαυλίζουν την συνεργασία των Κρατών Μελών για την θαλάσσια επιτήρηση.</p> <p>Η Θαλάσσια επιτήρηση βασίζεται στον συντονισμό και τον διαμοιρασμό πληροφοριών μεταξύ των Κρατών Μελών. Επομένως, υπάρχει το δυναμικό να δημιουργηθεί ένας κοινός ελεγκτικός μηχανισμός μεταξύ των συμμετεχόντων πρακτόρων, σχετικά με τα θεμελιώδη ανθρώπινα δικαιώματα, τους μεταναστευτικούς νόμους και τις επιταγές διάσωσης.</p>
---	---	---

Η Νοημοσύνη Ανοιχτού Κώδικα – NAK ή Open-Source Intelligence – OSINT, είναι νοημοσύνη η οποία συλλέγεται από πηγές δημόσια διαθέσιμες, συμπεριλαμβανομένου του διαδικτύου, των εφημερίδων, την τηλεόραση, κυβερνητικές αναφορές και επαγγελματικές και ακαδημαϊκές έρευνες. Το τελευταίο διάστημα το διαδίκτυο και ειδικότερα τα μέσα κοινωνικής δικτύωσης έχουν φέρει επανάσταση στην NAK και στην πληροφορία που αναλύεται. Η NAK επίσης λέγεται και δεοντολογική πειρατεία και χρησιμοποιείται για καλοπροαίρετη ψυχολογική και φυσιολογική χειραγώγηση με στόχο την απόκτηση πληροφοριών.

Πίνακας 2: Σύνοψη νομικού και δεοντολογικού πλαισίου της Νοημοσύνης Ανοιχτού Κώδικα – ALTAI Ερωτηματολόγιο: Κοινωνική και Περιβαλλοντολογική Ευεξία

Κύριες ανησυχίες για NAK	Διεθνείς και Ευρωπαϊκοί Κανονισμοί
Προέλευση και πρόθεση πηγών Χωρίς κατηγοριοποίηση αλλά ευαίσθητες πληροφορίες Επίδραση Μωσαϊκού ⁵ Εξάρτηση σε αυτοματοποιημένη ανάλυση Δημοσιότητα και Ορατότητα	Θεμελιώδη Δικαιώματα Ευρωπαϊκή Σύμβαση για Ανθρώπινα Δικαιώματα Σύμβαση Κυβερνοεγκλήματος Δικαιώματα Πνευματικής Ιδιοκτησίας ⁶ Αξιοπιστία Κανονισμοί για πρακτορία ερευνών

Νοημοσύνη Κοινωνικών Μέσων – NKM ή Social Media Intelligence – SOCMINT μπορεί να οριστεί ως η αναλυτική εκμετάλλευση πληροφοριών που είναι διαθέσιμες στα μέσα κοινωνικής δικτύωσης. Προσδιορίζεται από το περιεχόμενο των μέσων κοινωνικής δικτύωσης ως ευκαιρία και πρόκληση για τις έρευνες ανοιχτού κώδικα. Η παρακολούθηση εκατομμυρίων λογαριασμών σε πραγματικό χρόνο, η ανάλυσή τους και η αποθήκευση τους αποτελεί κύριο ζητούμενο της NKM.

Πίνακας 3: Δεοντολογικές και νομικές προκλήσεις της Νοημοσύνης Κοινωνικών Μέσων – ALTAI Ερωτηματολόγιο: Διαφάνεια, Ασφάλεια

Αστυνομική Χρήση Κοινωνικών Δικτύων	Πολίτες Αστυνομικοί
Νομικά ζητούμενα I) Ο διπλός ρόλος των δημόσιων φορέων ασφάλειας οι οποίοι επιβάλλουν τον νόμο II) Θεμελιώδη δικαιώματα πολιτών III) Συμμετοχή των πολιτών στην πρόληψη της δημόσιας ασφάλειας	Δυσκολία διασφάλισης διαφάνειας, αξιοπιστίας και μη-διάκρισης. Οι πολίτες καθοδηγούνται από ίδιες ερμηνείες του νόμου και ηθικό κώδικα χωρίς δημοκρατικά νομιμοποιημένη εξουσία.

⁵ Επίδραση Μωσαϊκού – Mosaic Effect: Είναι η επίδραση διαφορετικών κομματιών δεδομένων ή πληροφορίας – παρά το γεγονός ότι από μόνες τους δεν έχουν ιδιαίτερη χρησιμότητα – γίνονται σημαντικές όταν συνδυάζονται με άλλους τύπους πληροφοριών^[17].

⁶ Δικαιώματα Πνευματικής Ιδιοκτησίας – IPR Legislation^[18].

<p>Δεοντολογικά ζητούμενα</p> <p>I) Δυσανάλογη παρέμβαση στην ιδιωτικότητα αθών ατόμων ή ομάδων</p> <p>II) Ρίσκο άδικων διακρίσεων</p> <p>III) Δικαιώματα αστυνομικών για ιδιωτική ζωή και ελευθερία έκφρασης.</p>	<p>Κύριες προκλήσεις σχετικά με το σκοτεινό διαδίκτυο⁷</p> <p>I) Πώς διαχωρίζονται παράνομες και προσβλητικές ή ανήθικες συμπεριφορές</p> <p>II) Πώς καθορίζεται η γραμμή μεταξύ δικαιολογημένων και συγκαλυμμένων αλληλεπιδράσεων μεταξύ εγκληματιών με στόχο την παγίδευσή τους.</p>
--	---

Σχετικά με τα δυνητικά μεγάλα ποσά δεδομένων που αποθηκεύονται στο έργο ΘΟΕΕ, προτείνεται να υιοθετηθεί μια αυθεντική προσέγγιση για «μεγάλα δεδομένα» τα οποία ορίζονται ως η συλλογή συνόλων δεδομένων τόσο μεγάλων και πολύπλοκων που γίνεται δύσκολη η επεξεργασία τους χρησιμοποιώντας χειρόγραφα διαχειριστικά εργαλεία δεδομένων.

Επί του πρακτέου, τα «μεγάλα δεδομένα» καθιστούν ένα σύνολο τεχνικών και εργαλείων τα οποία είναι κατάλληλα για την αποθήκευση και την επεξεργασία τέτοιων συνόλων δεδομένων συνήθως μέσω NoSQL⁸ προσέγγισης. Οι δυνατότητες συλλογής, ανάλυσης, διάδοσης και αποθήκευσης μεγάλων ποσοτήτων δεδομένων εγείρει ανησυχίες για την ιδιωτικότητα και τους τρόπους που η ατομική ελευθερία μπορεί να παραβιάζεται ή να προστατεύεται. Η ανωνυμία επικαλύπτεται με την ιδιωτικότητα αλλά δεν ταυτίζονται ως έννοιες. Ομοίως, η ικανότητα να λάβει κανείς ευαίσθητες προσωπικές αποφάσεις χωρίς κυβερνητική παρέμβαση πρέπει να θεωρείται ιδιωτικό δικαίωμα, όπως και η προστασία από διακρίσεις λόγω συγκεκριμένων ατομικών χαρακτηριστικών όπως η φυλή, το φύλο ή το γονιδίωμα.

⁷ Σκοτεινό Διαδίκτυο είναι το μέρος του παγκόσμιου δικτύου το οποίο είναι προσβάσιμο αποκλειστικά και μόνο από ειδικό λογισμικό και φυλλομετρητές οι οποίοι επιτρέπουν τους χρήστες να παραμείνουν ανώνυμοι και μη-εντοπίσιμοι.

⁸ Η NoSQL προσέγγιση στην διαχείριση βάσεων δεδομένων φιλοξενεί ποικιλία μοντέλων δεδομένων, συμπεριλαμβανομένων ζευγαριών κλειδιών-τιμών, κειμένου, στηλών και γραφημάτων. Μια βάση δεδομένων NoSQL σημαίνει πως είναι μη-σχεσιακή, καταναμημένη, ευέλικτη και κλιμακώσιμη.

Πίνακας 4: Πλαίσιο για μεγάλα δεδομένα – ALTAI Ερωτηματολόγιο: Ιδιωτικότητα και Διακυβέρνηση δεδομένων

Κανονισμός στην φάση της ανάλυσης	Νομικό και ρητό καθήκον φροντίδας Εξωτερικές αξιολογήσεις και έλεγχοι Ρήτρες λήξης ισχύος Αξιολόγηση των επιπτώσεων επιτήρησης
Κανονισμός για την χρήση των δεδομένων	Προφίλ οριοθέτησης Μη-αυτοματοποιημένη ή ημι-αυτοματοποιημένη διαδικασία λήψης απόφασης Κάτοχος δεδομένων, κάτοχος ευθύνης
Ενίσχυση εποπτείας	Αυξανόμενη διαφάνεια επιπέδων Αυξανόμενες πιθανότητες για δικαστικές αξιολογήσεις

Η Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ αποτελεί ένα έργο που είναι η βάση για συστήματα TN στην ναυτιλία, το οποίο κατατάσσεται στην κλάση περιορισμένου ρίσκου στην πυραμίδα του AI Νομικού Πλαισίου. Δεδομένου των θεμάτων διαφάνειας του έργου, το AI Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας οι οποίες διαβεβαιώνουν ότι οι άνθρωποι πληροφορούνται όταν είναι απαραίτητο, ενισχύοντας την εμπιστοσύνη του έργου και των δυνητικών αποκημάτων του.

3.1.3 Θαλάσσια Αυτόνομα Πλοία Επιφάνειας - ΘΑΠΕ

Μια ακόμη περιοχή ενδιαφέροντος με θέματα δεοντολογίας TN αποτελούν τα Θαλάσσια Αυτόνομα Πλοία Επιφάνειας – ΘΑΠΕ ή Maritime Autonomous Surface Ships – MASS τα οποία εισάγονται με υψηλά επίπεδα αυτονομίας. Δηλαδή, δεν αλλάζει μόνο ο ρόλος του ανθρώπινου λειτουργού αλλά αλλάζει και ο τρόπος που οι διαδικασίες λαμβάνουν χώρα. Τα ΘΑΠΕ είναι μη επανδρωμένες πλατφόρμες οι οποίες ελέγχονται από Απομακρυσμένα Κέντρα Ελέγχου – ΑΚΕ ή Remote Control Centers - RCC. Έτσι, οι Ομάδες Ανθρώπων και Μηχανών – ΟΑΜ ή Human-Machine Teams – HMT αποκτούν σημασία^[20].

Στην παρούσα μελέτη έχει πραγματοποιηθεί συστηματική ακαδημαϊκή ανασκόπηση η οποία επικεντρώνεται στην λήψη απόφασης, προκειμένου να ενισχυθούν οι γνώσεις στην λειτουργία μη-επανδρωμένων πλοίων και να γίνουν σχεδιαστικές προτάσεις ανάπτυξης ΘΑΠΕ. Αναπτυξιακές περιοχές της έρευνας αποτελούν τα συστήματα λήψης απόφασης, η εμπιστοσύνη, η διαφάνεια, οι ομάδες, η κατανομή εργασίας - ρόλου, η αξιοπιστία και η επίγνωση της κατάστασης.

Πίνακας 5: Σχεδιαστικές Αρχές για Θαλάσσια Αυτόνομα Πλοία Επιφάνειας και οδοί για περεταίρω έρευνα – ALTAI Ερωτηματολόγιο: Αξιοπιστία, Διαφάνεια και Λογοδοσία

Θεματική Περιοχή	Σχεδιαστικές Προτάσεις για Θαλάσσια Αυτόνομα Πλοία Επιφάνειας
Συστήματα Υποστήριξης Απόφασης	<p>Ανθρωποκεντρικές αρχές σχεδίασης όπως η Ανάλυση Γνωστικών Εργασιών – ΑΓΕ ή Cognitive Task Analysis – CTA και η Ιεραρχική Ανάλυση Εργασίας – ΙΕΑ ή Hierarchical Task Analysis – HTA χρειάζονται προκειμένου να διασφαλιστεί ότι τα Συστήματα Υποστήριξης Απόφασης – ΣΥΑ ή Decision Support Systems - DSS για ΘΑΠΕ είναι σχεδιασμένα αποδοτικά έτσι ώστε να υποστηρίξουν τους λειτουργούς στις εργασίες τους.</p> <p>Ο τρόπος με τον οποίο ο σχεδιασμός ΣΥΑ για ΘΑΠΕ επηρεάζει το φόρτο εργασίας του λειτουργού όταν διεξάγει τις εργασίες του και τότε είναι πιθανό να αποφευχθούν υψηλές διακυμάνσεις εργασίας και να διασφαλιστεί πως ο φόρτος εργασίας παραμένει σε σταθερό επίπεδο.</p> <p>Αλλαγή του σχεδιασμού των ΣΥΑ ώστε να ενισχύεται η επικοινωνία και η κατανόηση μεταξύ των αυτοματοποιημένων συστημάτων των ΘΑΠΕ και των λειτουργών τους.</p>
Εμπιστοσύνη	<p>Εξέταση του τρόπου με τον οποίο η εκπαίδευση του λειτουργού ΘΑΠΕ μπορεί να ενισχύεται μέσω ακριβών μοντέλων βιο-προσομοίωσης, δίνοντας πρακτική εμπειρία με αυτοματοποιημένα συστήματα και καλύτερη κατανόηση των δυνατοτήτων των ΘΑΠΕ, προκειμένου να προωθούνται τα κατάλληλα επίπεδα εμπιστοσύνης.</p> <p>Εξέταση του επιπέδου διαφάνειας των αυτοματοποιημένων συστημάτων καθώς επηρεάζει την ικανότητα του λειτουργού να</p>

	<p>κατανοεί τις διαδικασίες των συστημάτων και κατ' επέκταση την εμπιστοσύνη του στο αυτοματοποιημένο σύστημα.</p> <p>Η χρήση ευέλικτων μεθόδων όπως οι προσαρμόσιμοι αυτοματισμοί πρέπει να λαμβάνονται υπόψιν κατά τον σχεδιασμό ΘΑΠΕ όπου οι εργασίες είναι κατάλληλες για ανθρώπινο και μηχανικό έλεγχο.</p>
<p>Διαφάνεια</p>	<p>Για τον σχεδιασμό των ΘΑΠΕ συστημάτων, είναι σημαντικό να ληφθεί υπόψιν το επίπεδο διαφάνειας των αυτοματοποιημένων συστημάτων ώστε οι λειτουργοί να έχουν ικανοποιητικό ποσό πληροφορίας προκειμένου να λαμβάνουν πλήρως πληροφορημένες αποφάσεις.</p> <p>Προτείνεται η διαφάνεια του συστήματος να βελτιωθεί με το να συμπεριλαμβάνει πληροφορίες όπως η δημοσιοποίηση του στόχου του, του πλάνου δράσης επίτευξης του στόχου, η πρόοδος των ενδιάμεσων σταδίων και η διαδικασία λογισμού του.</p> <p>Είναι επίσης απαραίτητο να διερευνηθεί περαιτέρω ποιοι άλλοι τύποι πληροφορίας θα μπορούσαν να χρησιμοποιηθούν κατά την σχεδίαση ΘΑΠΕ και πως οι πληροφορίες θα ενσωματώνονται και θα προβάλλονται στον λειτουργό του συστήματος προκειμένου να βελτιωθεί η αποδοτικότητα κάθε εργασίας και να προωθηθεί η αξιοπιστία των αυτοματοποιημένων ΘΑΠΕ συστημάτων.</p>
<p>Ομάδες</p>	<p>Είναι σημαντικό για τα ΘΑΠΕ ότι ο ρόλος του λειτουργού έχει καθοριστεί πλήρως και ότι ο λειτουργός έχει κατανοήσει τον ρόλο του όπως και τον ρόλο του αυτοματοποιημένου συστήματος ως υποσύνολο του συνολικού στόχου του συστήματος.</p> <p>Κατά τον σχεδιασμό των ΘΑΠΕ πρέπει να λαμβάνεται υπόψιν πως το αυτοματοποιημένο σύστημα μπορεί να διαβεβαιώσει ότι ο λειτουργός έχει τις απαραίτητες πληροφορίες προκειμένου να λάβει πληροφορημένες αποφάσεις.</p> <p>Όταν γίνεται έρευνα για ΘΑΠΕ με Ομάδες Ανθρώπου Μηχανής – ΟΑΜ ή Human-Machine Teams – HMT, το επίπεδο πιστότητας πρέπει να λαμβάνεται σοβαρά υπόψιν προκειμένου η οικολογική ορθότητα των ευρημάτων να είναι έμπιστη.</p> <p>Περαιτέρω έρευνα χρειάζεται προκειμένου να βρεθεί πως διαφορετικές δομές ομάδων και μεθόδων επικοινωνίας θα</p>

	επηρεάσουν την συνεργασία μεταξύ των ΟΑΜ συγκεκριμένα στην λειτουργία των ΘΑΠΕ.
Κατανομή Εργασίας - Ρόλου	<p>Τα ρίσκα τα οποία σχετίζονται με την ανάπτυξη ΘΑΠΕ συστημάτων πρέπει να λαμβάνονται υπόψιν προκειμένου να διαβεβαιωθεί πως ένα ικανοποιητικό επίπεδο αυτοματισμού χρησιμοποιείται στο σύστημα. Εάν οι λειτουργοί εργάζονται με συστήματα κριτικής σημασίας για την ασφάλεια, όπως συστήματα αποφυγής συγκρούσεων πρέπει να διαβεβαιώνεται πως υπάρχει σταθμισμένο επίπεδο ελέγχου και πως συμμετέχουν στην διαδικασία λήψης απόφασης.</p> <p>Οι εργασίες του λειτουργού πρέπει να ληφθούν υπόψιν κατά τον σχεδιασμό ΘΑΠΕ συστημάτων καθώς το φόρτο εργασίας του λειτουργού επηρεάζεται. Προσαρμοστικοί αυτοματισμοί μπορούν να τεθούν σε λειτουργία προκειμένου να μειωθεί ο φόρτος εργασίας του λειτουργού μέσω ευελιξίας στις ΟΑΜ των ΘΑΠΕ.</p> <p>Ο τύπος της εργασίας πρέπει να λαμβάνεται υπόψιν και να μην δημιουργούνται υπερβολικά φορτία εργασίας κατά την παράδοση της εργασίας από τον λειτουργό στο αυτοματοποιημένο σύστημα και αντίστροφα.</p>
Λογοδοσία	<p>Υπάρχει ανάγκη περαιτέρω έρευνας σε θέματα λογοδοσίας των ΟΑΜ, καθώς όταν τα αυτοματοποιημένα συστήματα χρησιμοποιούν υψηλότερα επίπεδα αυτοματισμών υπάρχει λιγότερη άμεση ανθρώπινη εμπλοκή. Ωστόσο, υπάρχει η ανάγκη να αξιολογείται ένα σύστημα πέρα από την σχέση λειτουργού – ΘΑΠΕ σε θέματα λογοδοσίας και να γίνεται αξιολόγηση με βάση το εκάστοτε σύστημα. Επίσης, χρειάζεται να συνεκτιμάται η γενικότερη επιρροή του και μέσα στο πλαίσιο ενός οργανισμού.</p> <p>Το επίπεδο ελέγχου του λειτουργού στο σύστημα πρέπει να σταθμίζεται προσεκτικά ανάλογα με τα συσχετιζόμενα ρίσκα του συστήματος, προκειμένου να διαβεβαιώνεται πως ο λειτουργός διατηρεί ένα επαρκές επίπεδο ελέγχου και συνεχίζει να εμπλέκεται στην διαδικασία λήψης απόφασης της ομάδας ΟΑΜ.</p>
Επίγνωση της κατάστασης	<p>Περαιτέρω έρευνα είναι απαραίτητη με στόχευση το είδος της πληροφορίας που χρειάζεται προκειμένου οι λειτουργοί των ΘΑΠΕ να συντηρήσουν την Επίγνωση της Κατάστασης – EK ή Situation Awareness – SA και τον τρόπο που η πληροφορία πρέπει να εμφανίζεται στους λειτουργούς.</p>

	<p>Αφηγηματικοί τρόποι υποστήριξης της ΕΚ των λειτουργιών απαιτούνται καθώς δεν είναι δυνατόν να έχουν πρόσβαση στα ίδια στοιχεία του περιβάλλοντος που θα είχαν και επί του σκάφους. Μπορεί να έχουν έλλειψη «αίσθησης του πλοίου» και δυνητικά να απεμπλέκονται όσο μεγαλώνουν τα επίπεδα αυτοματισμού.</p> <p>Υπάρχει επίσης η ανάγκη να βρεθούν πρωτοποριακοί τρόποι να υποστηριχθεί ο λειτουργός των ΘΑΠΕ όταν επανέρχεται στον βρόχο και επανακτά τον έλεγχο σε περίπτωση έκτακτης ανάγκης.</p>
--	---

Τα ΘΑΠΕ κατηγοριοποιούνται στα συστήματα περιορισμένου ρίσκου του AI Νομικού Πλαισίου καθώς με συγκεκριμένες επιταγές διαφάνειας ενισχύεται η εμπιστοσύνη και η λήψη απόφασης δεν γίνεται παρακινδυνευμένα^[20].

3.2 Ερευνητικές Δημοσιεύσεις Ηθικών Λύσεων Τεχνητής Νοημοσύνης στην Γενετική Τεχνητή Νοημοσύνη

Η δεύτερη και ερευνητικά εξαιρετικά ενδιαφέρουσα περιοχή ανάπτυξης της παρούσας εργασίας είναι ο τομέας της Γενετικής Τεχνητής Νοημοσύνης – ΓΤΝ ή Generative Artificial Intelligence – GAI ο οποίος έχει προχωρήσει σημαντικά και οι επιπτώσεις του είναι ορατές σε πολλούς κλάδους.

3.2.1 Διαχείριση Προϊόντων Λογισμικού - ΔΠΛ

Ευρεία εφαρμογή υπάρχει στον κλάδο Διαχείρισης Προϊόντων Λογισμικού – ΔΠΛ ή Software Product Management - SPM. Η έρευνα αποδεικνύει ότι η συγκεκριμένη τεχνολογία βοηθά την ιδέα, την έρευνα αγοράς, την γνώση των καταναλωτών, την συγγραφή προδιαγραφών και την ανάπτυξη προϊόντων. Με την αυτοματοποίηση εργασιών όπως η παραγωγή κώδικα και η ανάλυση των εισροών των χρηστών, μπορεί να μειώσει την ακριβή και χρονοβόρα ανάπτυξη λογισμικού. Ωστόσο, εξακολουθούν να υπάρχουν ανησυχίες για την ακρίβεια και την ασφάλεια της τεχνολογίας καθώς και για τις ηθικές επιπτώσεις. Οι πρακτικές εφαρμογές της ΓΤΝ έχουν την δυνατότητα να

βελτιώσουν σημαντικά τις διαδικασίες διαχείρισης προϊόντων λογισμικού, οδηγώντας σε αποτελεσματική χρήση των πόρων με αποτέλεσμα την παραγωγή προϊόντων υψηλότερης ποιότητας και σε βελτιωμένες εμπειρίες των χρηστών^[21].

Πίνακας 6: Εφαρμογές της Γενετικής Τεχνητής Νοημοσύνης στην Διαχείριση Προϊόντων Λογισμικού.

Εφαρμογή	Σύνοψη
Έρευνα αγοράς	<p>Τα μοντέλα Γενετικής Τεχνητής Νοημοσύνης μπορούν να χρησιμοποιηθούν στην έρευνα αγοράς, δρώντας ως υποθετικοί πελάτες προκειμένου να κατανοήσουν τις προτιμήσεις. Τέτοια μοντέλα μπορούν να εμπνεύσουν νέα ερευνητικά παραδείγματα που δεν περιορίζονται από την έρευνα σε ανθρώπινο υποκείμενο.</p> <p>Η Γενετική Τεχνητή Νοημοσύνη μπορεί να παράγει ιδέες σε εξειδικευμένους κλάδους, με το μεγαλύτερο μοντέλο να παράγει πιο συνεκτικές και διασυνδεδεμένες ιδέες.</p>
Καθορισμός θέσης και προϊόντος	<p>Η ΓΤΝ που χρησιμοποιείται κυρίως για συγγραφή περιεχομένου, μπορεί να βοηθήσει στην τοποθέτηση και τον ορισμό προϊόντων, συμπεριλαμβανομένων προτάσεων ονομασίας και χαρακτηριστικών, με αποδεδειγμένη ικανότητα στην συγγραφή προϊόντων ηλεκτρονικού εμπορίου και στην δημιουργία τίτλων σε διάφορους κλάδους.</p>
Γνώσεις πελατών, υποστήριξη πελατών	<p>Μοντέλα ΓΤΝ όπως το ChatGPT και το Bard της Google, μπορούν να βελτιώσουν τις εμπειρίες πελατών ερμηνεύοντας και ενσωματώνοντας δεδομένα, μεταφράζοντας τις ανάγκες σε αιτήματα και να ανταποκρίνονται με εξατομικευμένες λύσεις. Μπορούν επίσης να αναλύουν μεγάλους όγκους δεδομένων ανατροφοδότησης πελατών για να εξάγουν πολύτιμες πληροφορίες. Επιπλέον η ΓΤΝ μπορεί να δημιουργήσει έναν θετικό βρόχο ανατροφοδότησης ενισχύοντας την κατανόηση του πελάτη για πιο εξατομικευμένες υπηρεσίες.</p>
Μηχανική απαιτήσεων προϊόντων	<p>Η ΓΤΝ μπορεί να αυτοματοποιήσει την ανάδειξη και την ανάλυση απαιτήσεων, παράγοντας δηλώσεις απαιτήσεων σε φυσική γλώσσα από δομημένα δεδομένα. Επίσης, μπορεί να βοηθήσει στον εντοπισμό συγκρούσεων σε έγγραφα Προδιαγραφών Απαιτήσεων Λογισμικού – ΠΑΛ ή Software Requirement Specifications – SRS.</p>

	<p>Η ΓΤΝ μπορεί να υποστηρίξει «σπριντ προγραμματισμό⁹», την διαχείριση των επόμενων εργασιών και την εκτίμησή τους, στην ευέλικτη ανάπτυξη λογισμικού.</p> <p>Μοντέλα ΓΤΝ όπως το GPT-4, μπορούν να δημιουργήσουν αυτόματα ιστορίες χρήστη, εξοικονομώντας χρόνο, μειώνοντας το ανθρώπινο λάθος και βελτιώνοντας την ποιότητα ιστοριών χρήστη.</p>
Εκτέλεση ανάπτυξης	<p>Η ΓΤΝ μπορεί να βοηθήσει στην αυτοματοποιημένη παραγωγή κώδικα, ενισχύοντας την παραγωγικότητα και την αποδοτικότητα. Εργαλεία όπως το Github Copilot έχουν επιδείξει σημαντική αύξηση της παραγωγικότητας. Επιπλέον, η ΓΤΝ μπορεί να χρησιμοποιηθεί για την αυτοματοποίηση της δημιουργίας τεκμηρίωσης και την βελτίωση της κατανόησης του πηγαίου κώδικα στην ανάπτυξη λογισμικού.</p> <p>Η ΓΤΝ μπορεί να δημιουργήσει οπτικά ελκυστικές και μοναδικές διεπαφές, απλοποιώντας την διαδικασία σχεδιασμού και επιτρέποντας την ταχεία δημιουργία πρωτοτύπων. Μπορεί να δημιουργήσει σχέδια ΔΧ/ΕΧ¹⁰ με βάση τις προτιμήσεις και τις απαιτήσεις των χρηστών, οδηγώντας σε πιο χρηστοκεντρικά σχέδια και βελτιωμένες εμπειρίες προϊόντων.</p>
Λήψη απόφασης	<p>Η ΓΤΝ μπορεί να βελτιώσει την λήψη απόφασης στις επιχειρήσεις αντιμετωπίζοντας τους περιορισμούς του μοντέλου περιορισμένου ορθολογισμού, με τους νεότερους χρήστες να είναι συνήθως πιο ανοιχτοί στην υιοθέτηση της τεχνολογίας.</p>

Ηθικές Επιπτώσεις

Η ΓΤΝ έχει ένα ευρύ φάσμα δυνητικών αρνητικών συνεπειών, γεγονός που αποδεικνύει την ανάγκη για αυστηρά πρότυπα και ρυθμιστικά μέτρα που εγγυώνται την υπεύθυνη χρήση της.

ALTAI Ερωτηματολόγιο: Δικαιοσύνη

Ο κίνδυνος μεροληψίας στα συστήματα ΓΤΝ είναι υπαρκτός. Η χρήση προκατειλημμένων δεδομένων εκπαίδευσης μπορεί να προκαλέσει το σύστημα ΤΝ να παρέχει

⁹ «Σπριντ προγραμματισμός» ή Sprint planning είναι ένας σχεδιασμός προϊόντος ορισμένου μήκους ενώ «Σκραμ προγραμματισμός» ή Scrum planning είναι ένα προσαρμοστικό πλαίσιο ανάπτυξης προϊόντος^[22].

¹⁰ User Interface/User Experience - UI/UX ή Διεπαφή Χρήστη/Εμπειρία Χρήστη - ΔΧ/ΕΧ.

προκατειλημμένα ή πολωμένα αποτελέσματα κατά λάθος. Προκειμένου να αποφευχθεί κάθε είδους προκατάληψη ή ανεπιθύμητη ζημία, είναι ζωτικής σημασίας η χρήση ποικίλων και δίκαιων δεδομένων εκπαίδευσης.

ALTAI Ερωτηματολόγιο: Ιδιωτικότητα Δεδομένων

Η εμπιστευτικότητα των προσωπικών δεδομένων είναι ένας άλλος σημαντικός δεοντολογικός παράγοντας. Δεδομένου ότι τα συστήματα ΓΤΝ χρειάζονται πολλά δεδομένα για να μάθουν, μπορούν να προστατευθούν με την τήρηση της νομοθεσίας για την προστασία της ιδιωτικής ζωής, όπως ο Γενικός Κανονισμός για την Προστασία Δεδομένων – ΓΚΠΔ ή General Data Protection Regulation – GDPR. Όταν πρόκειται για την δεοντολογική χρήση προσωπικών δεδομένων για την ΓΤΝ, οι επιχειρήσεις μπορούν να αναζητήσουν καθοδήγηση στις αρχές που περιγράφονται στον ΓΚΠΔ.

ALTAI Ερωτηματολόγιο: Λογοδοσία

Στο πλαίσιο της ΓΤΝ, η λογοδοσία είναι η αποδοχή της κυριότητας των αποτελεσμάτων που παράγει το σύστημα. Οι οργανισμοί είναι υπεύθυνοι για την διόρθωση των συστημάτων ΤΝ που παράγουν ακατάλληλο ή λανθασμένο υλικό και την διενέργεια των κατάλληλων κινήσεων με στόχο την αποτροπή παρόμοιων τέτοιων γεγονότων εάν αυτό συμβεί.

ALTAI Ερωτηματολόγιο: Διαφάνεια

Ένα ακόμη μείζον ζήτημα είναι η πιθανότητα ακατάλληλης χρήσης του υλικού το οποίο δημιουργήθηκε από την ΓΤΝ. Οι επιπτώσεις του πλαστού ή παραποιημένου υλικού το οποίο παρερμηνεύεται ως αληθινό σε ποικίλα μέρη της αγοράς, όπως οι ειδήσεις, η δημοσιογραφία, οι οικονομικές εκθέσεις, και τα νομικά έγγραφα είναι σημαντικές. Η αποτροπή τέτοιων καταχρήσεων απαιτεί αυστηρές τεχνικές επαλήθευσης. Τα πολύπλοκα μοντέλα είναι αποτέλεσμα της εξάπλωσης των βαθιών γενετικών μοντέλων ΤΝ. Ωστόσο, εξαρτώνται σε μεγάλο βαθμό από υψηλής ποιότητας δεδομένα εκπαίδευσης προκειμένου να λειτουργούν εύρυθμα. Μια αξιοσημείωτη δυσκολία με αυτά τα μοντέλα είναι ότι συχνά υποφέρουν από το πρόβλημα του «μαύρου κουτιού», το οποίο μειώνει την ερμηνευσιμότητα τους και μπορεί να οδηγήσει σε προβλήματα εμπιστοσύνης εάν δεν αντιμετωπιστεί.

ALTAI Ερωτηματολόγιο: Τεχνική Ευρωστία

Το θέμα των «ψευδαισθήσεων» και των ανακριβών πληροφοριών αφορά τα μεγάλα γλωσσικά μοντέλα. Για να διασφαλιστεί η ποιότητα των αποτελεσμάτων που παράγονται από την TN, πρέπει να πραγματοποιούνται σε τακτική βάση ανθρώπινες αξιολογήσεις.

Νομικοί Κίνδυνοι

Οι νομικοί κίνδυνοι που συνδέονται με την GTN περιλαμβάνουν την πιθανή παραβίαση των δικαιωμάτων πνευματικής ιδιοκτησίας και την έλλειψη σαφήνειας γύρω από τα ζητήματα όπως η ιδιοκτησία και η δυνατότητα εφαρμογής των νόμων περί πνευματικών δικαιωμάτων, δικαιωμάτων ευρεσιτεχνίας και εμπορικών σημάτων. Οι επιχειρήσεις πρέπει να γνωρίζουν τους κινδύνους αυτούς πριν χρησιμοποιήσουν την GTN στο πλήρες δυναμικό της. Οι επικαιροποιήσεις των συμβάσεων προμηθευτών και πελατών συμπεριλαμβανομένων των διατάξεων για την προστασία της πνευματικής ιδιοκτησίας και η σαφής γνωστοποίηση της χρήσης της GTN αποτελούν κρίσιμα βήματα. Για να αποτραπεί η αποκάλυψη ευαίσθητων πληροφοριών μέσω εργαλείων κειμένου, πρέπει να αυστηροποιηθεί η προστασία της ιδιωτικής ζωής.

Η χρήση λύσεων GTN στην διαχείριση προϊόντων λογισμικού δεν έρχεται χωρίς πιθανά όρια ή εμπόδια, τα οποία πρέπει να αναγνωριστούν και να αντιμετωπιστούν. Είναι σημαντικό να δοθεί ιδιαίτερη προσοχή στην τρέχουσα πρακτική όσο και στην μελλοντική μελέτη σε θέματα μεροληψίας, διαφάνειας, προστασίας της ιδιωτικής ζωής, λογοδοσίας, δεοντολογίας και προκλήσεων ενσωμάτωσης. Περιεχόμενο που δημιουργήθηκε από GTN εξακολουθεί να πρέπει να ελέγχεται από τους διαχειριστές προϊόντων για να επαληθεύεται η αυθεντικότητα και η ορθότητα του. Συμπερασματικά, κατηγοριοποιείται στην ιεραρχία του AI Νομικού Πλαισίου ως λύση περιορισμένου ρίσκου η οποία συσχετίζεται κατά κύριο λόγο με την έλλειψη διαφάνειας στην χρήση της GTN. Σε αυτήν την περίπτωση, το AI Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας ώστε να διασφαλιστεί πως οι άνθρωποι ενημερώνονται όταν χρειάζεται, ενισχύοντας έτσι την εμπιστοσύνη. Οι πάροχοι λύσεων GTN στην διαχείριση προϊόντων λογισμικού πρέπει να διασφαλίζουν πως το παραγόμενο περιεχόμενο από TN είναι ταυτοποιήσιμο.

Στις αρχές της δεκαετίας τους 2020, η πρόοδος των μετασχηματιστικών Βαθιών Νευρωνικών Δικτύων επιτρέπουν την ανάπτυξη ενός αυξανόμενου αριθμού συστημάτων

ΓΤΝ τα οποία δέχονται ως είσοδο υποδείξεις φυσικής γλώσσας. Ευρέως γνωστά παραδείγματα τέτοιων γλωσσικών μοντέλων είναι οι αυτόματοι συνομιλητές όπως το ChatGPT, το Bard και άλλα. Η ΓΤΝ έχει εφαρμογή σε ένα ευρύ φάσμα βιομηχανιών, όπως η τέχνη, η συγγραφή, η ανάπτυξη λογισμικού, ο σχεδιασμός προϊόντων, η υγειονομική περίθαλψη, τα χρηματοοικονομικά, τυχερά παιχνίδια κ.λπ. Στην παρούσα εργασία μελετούνται οι πρόσφατες εξελίξεις σε ιστορικό και κυβερνογνωσιακό πλαίσιο. Αναλύονται ηθικά ζητήματα στον τομέα της μηχανικής λογισμικού και των κυβερνοφυσικών συστημάτων. Επιπλέον διερευνώνται οι προκλήσεις που βασίζονται στην ΤΝ στην υγειονομική περίθαλψη. Η έρευνα δείχνει τη σημασία της αυστηρής ηθικής ανάλυσης και των συνακόλουθων εγγυήσεων για την αντιμετώπιση των αναδυόμενων ζητημάτων της ΤΝ^[23].

3.2.2 Μηχανική Λογισμικού – ΜΛ και Κυβερνο-Φυσικά Συστήματα - ΚΦΣ

Η αυξανόμενη εφαρμογή της ΤΝ, εργαλείων Μηχανικής Μάθησης, μοντέλων και αλγορίθμων μάθησης καθώς και η πρόσφατη ταχεία υιοθέτηση των Μεγάλων Γλωσσικών Μοντέλων – ΜΓΜ ή Large Language Models – LLMs και της ΓΤΝ έχουν πολύπλευρο αντίκτυπο στην Μηχανική Λογισμικού – ΜΛ ή Software Engineering – SE και στην ανάπτυξη Κυβερνο-Φυσικών Συστημάτων – ΚΦΣ ή Cyberphysical Systems – CPS. Η Μηχανική Λογισμικού αποτελεί κλάδο εφαρμογής αρχών και πρακτικών μηχανικής στον σχεδιασμό, την ανάπτυξη, τον έλεγχο, την συντήρηση και την εξέλιξη συστημάτων λογισμικού που εκτελούνται σε διάφορες συσκευές και πλατφόρμες. Τα Κυβερνο-Φυσικά Συστήματα αποτελούν υποσύνολο συστημάτων λογισμικού που ενσωματώνουν υπολογισμούς, επικοινωνία και φυσικές διαδικασίες για την επίτευξη μιας επιθυμητής λειτουργικότητας. Παραδείγματα ΚΦΣ είναι τα βιομηχανικά ρομπότ, ιατρικές συσκευές, έξυπνα δίκτυα, έξυπνα σπίτια και αυτόνομα οχήματα.

Πίνακας 7: ALTAI Ερωτηματολόγιο: Δηλώσεις δεοντολογικών αρχών

UNESCO ¹¹	IEEE ¹²	IBM ¹³
<p>Δεοντολογικές συστάσεις με βάση την ανθρώπινη αξιοπρέπεια, τα δικαιώματα, την προώθηση της ειρήνης και την φροντίδα για το περιβάλλον. Με βάση τις αξίες αυτές η UNESCO καθορίζει δέκα αρχές:</p> <ol style="list-style-type: none"> 1. Αναλογικότητα και Ακακία 2. Ασφάλεια και Προστασία 3. Δικαίωμα στην ιδιωτικότητα και προστασία δεδομένων 4. Πολυμερής, προσαρμοστική διακυβέρνηση και συνεργασία 5. Υπευθυνότητα και λογοδοσία 6. Διαφάνεια και Εξηγησιμότητα 7. Ανθρώπινη εποπτεία και αποφασιστικότητα 8. Βιωσιμότητα 9. Αντίληψη και Αλφαριθμητισμός 10. Δικαιοσύνη και κατά των διακρίσεων 	<p>Η IEEE Ένωση Προτύπων – ΕΠ ή Standards Association – SA έχει εγκαθιδρύσει μια παγκόσμια πρωτοβουλία για την δεοντολογία των αυτόνομων και ευφυών συστημάτων. Η προσέγγιση της IEEE βασίζεται σε οκτώ θεμελιώδεις αρχές:</p> <ol style="list-style-type: none"> 1. Ανθρώπινα δικαιώματα 2. Ευζήν 3. Εκπροσώπηση δεδομένων 4. Αποτελεσματικότητα 5. Διαφάνεια 6. Λογοδοσία 7. Εντοπισμός παράτυπης χρήσης 8. Ανταγωνιστικότητα 	<p>Η IBM προτείνει τρεις καθοδηγητικές αξίες για την TN:</p> <ol style="list-style-type: none"> 1. Ο στόχος της TN είναι να αυξήσει την ανθρώπινη ευφυΐα 2. Τα δεδομένα και η γνώση τους ανήκουν στον δημιουργό τους 3. Η τεχνολογία πρέπει να είναι διαφανής και εξηγησίμη <p>Μοχλεύοντας γνώση από το 1979 και την Αναφορά Belmont¹⁴, η IBM καθορίζει τρεις αρχές γενικής διεύθυνσης για την TN:</p> <ol style="list-style-type: none"> 1. Σεβασμός στους ανθρώπους 2. Ευεργεσία 3. Δικαιοσύνη, δηλαδή οφέλη και επιβαρύνσεις πρέπει να κατανέμονται με: <ol style="list-style-type: none"> 3.1 Ίσο μερίδιο 3.2 Ατομική ανάγκη 3.3 Ατομική προσπάθεια 3.4 Κοινωνική συμμετοχή 3.5 Αξιοκρατία

¹¹ United Nations Educational Scientific and Cultural Organization – UNESCO ή Εκπαιδευτική Επιστημονική και Πολιτιστική Οργάνωση των Ηνωμένων Εθνών - ΕΕΠΟΗΕ ^[24].

¹² Institute of Electrical and Electronics Engineers – IEEE ή Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών - ΙΗΗΜ ^[25].

¹³ International Business Machines – IBM ή Διεθνείς Μηχανές Επιχειρήσεων - ΔΜΕ ^[26].

¹⁴ Belmont Report - Αναφορά Belmont είναι μια έκθεση που δημιουργήθηκε από την Εθνική Επιτροπή για την Προστασία των Ανθρώπινων Υποκειμένων Βιοϊατρικής και Συμπεροφορικής Έρευνας ^[27].

Τα συστήματα GTN θέτουν νέους κινδύνους και προκλήσεις. Υφίστανται γνωστά προβλήματα ψευδαισθήσεων και προκαταλήψεων. Υπάρχει επίσης έλλειψη διαφάνειας και η εισαγωγή σφαλμάτων που είναι δύσκολο να διορθωθούν. Η χρήση των Μεγάλων Γλωσσικών Μοντέλων της GTN στην Μηχανική Λογισμικού και στην ανάπτυξη ΚΦΣ εγείρει πρόσθετες ηθικές ανησυχίες που πρέπει να αντιμετωπιστούν από όλα τα ενδιαφερόμενα μέρη, συμπεριλαμβανομένων των προγραμματιστών λογισμικού, τους εμπειρογνώμονες TN και τους επιστήμονες δεδομένων, τους μηχανικούς ΚΦΣ, τους χρήστες, τις ρυθμιστικές αρχές και την κοινωνία γενικότερα. Τα ΜΓΜ και η GTN δυνητικά υπονομεύουν ή διαβρώνουν την εμπιστοσύνη των ατόμων και των οργανισμών στο λογισμικό και τα ΚΦΣ παράγοντας ή προκαλώντας αναξιόπιστα, ανακριβή ή παραπλανητικά αποτελέσματα.

ALTAI Ερωτηματολόγιο: Ιδιωτικότητα δεδομένων, Αναπαραξιμότητα - Στρατηγικές για Πρακτικές Δεδομένων και Ελέγχου των Συστημάτων

Προκειμένου να αντιμετωπιστούν οι ηθικές προκλήσεις, κάποιες από τις στρατηγικές κλειδί και καλύτερες πρακτικές επικεντρώνονται στην ποιότητα των δεδομένων, τα μοντέλα, και τις εξόδους. Οι μηχανικοί λογισμικού και οι προγραμματιστές ΚΦΣ πρέπει να εκπαιδεύονται και να καταρτίζονται για την ορθή και ηθική αλληλεπίδραση με τα συστήματα TN. Πρέπει να διασφαλίζεται η ποιότητα των δεδομένων που χρησιμοποιούνται για την εκπαίδευση, τη δοκιμή ή την αξιολόγηση των ΜΓΜ και των συστημάτων και εφαρμογών GTN. Τα δεδομένα πρέπει να είναι συναφή, ακριβή, συνεπή και αντιπροσωπευτικά του προβλεπόμενου τομέα, της εργασίας και του πληθυσμού. Επιπλέον, τα δεδομένα πρέπει να λαμβάνονται, να υποβάλλονται σε επεξεργασία και να αποθηκεύονται με νόμιμο, δεοντολογικό και διαφανή τρόπο, με σεβασμό της ιδιωτικής ζωής, της συγκατάθεσης και της ιδιοκτησίας των παρόχων δεδομένων.

Οι μηχανικοί λογισμικού και οι προγραμματιστές ΚΦΣ πρέπει να διασφαλίζουν ότι τα μοντέλα που χρησιμοποιούνται για την υλοποίηση ΜΓΜ και GTN επικυρώνονται, επαληθεύονται και αξιολογούνται με τη χρήση κατάλληλων μεθόδων, μετρήσεων και κριτηρίων αναφοράς. Επίσης, τα αποτελέσματα να αναφέρονται και να τεκμηριώνονται με σαφή, ειλικρινή και αναπαράξιμο τρόπο. Περαιτέρω, τα αποτελέσματα πρέπει να παρακολουθούνται, να επανεξετάζονται και να διορθώνονται από ανθρώπινους

εμπειρογνώμονες και οι χρήστες πρέπει να ενημερώνονται και να εκπαιδεύονται σχετικά με την πηγή, τη φύση και τους περιορισμούς των αποτελεσμάτων.

Συμπερασματικά, η ΜΛ και τα ΚΦΣ αποτελούν συστήματα ΤΝ περιορισμένου ρίσκου τα οποία συνδέονται με την έλλειψη διαφάνειας κατά κύριο λόγο. Το ΑΙ Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας. Για παράδειγμα, όταν χρησιμοποιούνται αυτόματοι συνομιλητές ΤΝ, οι άνθρωποι πρέπει να ενημερώνονται πως αλληλεπιδρούν με ένα μηχάνημα ώστε να λάβουν πληροφορημένες αποφάσεις. Οι πάροχοι πρέπει επίσης να διασφαλίσουν ότι το περιεχόμενο που έχει δημιουργηθεί μέσω ΤΝ είναι ταυτοποιήσιμο. Κείμενα τα οποία έχουν δημιουργηθεί μέσω ΤΝ και είναι δημόσιου ενδιαφέροντος πρέπει να συνοδεύονται με ενημέρωση του δημόσιου καλού και της φύσης τους. Το ίδιο πρέπει να εφαρμόζεται για οπτικοακουστικό περιεχόμενο που αποτελεί ψευδές υλικό^[42].

3.2.3 Διαδίκτυο των Πραγμάτων - ΔΤΠ

Η Γενετική Τεχνητή Νοημοσύνη και το Διαδίκτυο των Πραγμάτων – ΔΤΠ ή Internet of Things – IoT είναι δύο ταχύτατα εξελισσόμενες τεχνολογίες που έχουν την δυνατότητα να καινοτομήσουν σε πολλές πτυχές της ζωής μας σήμερα. Η ΓΤΝ, η οποία αποτελεί υποσύνολο τεχνολογιών ΤΝ, έχει μια μοναδική ικανότητα να δημιουργεί και να χειρίζεται περιεχόμενο με βάση μοτίβα που μαθαίνουν από τεράστια σύνολα δεδομένων και όταν συνδυάζεται με την διασυνδεδεμένη φύση των συσκευών ΔΤΠ, υπόσχεται πολλά σε όλους τους κλάδους. Αξιοποιώντας τις δυνατότητες της Βαθιάς Μάθησης, η ΓΤΝ μπορεί να αναλύσει έξυπνα και να εφαρμόσει τις τεράστιες ποσότητες δεδομένων που παράγονται από το διαδίκτυο για την βελτιστοποίηση των λειτουργιών και την προώθηση της καινοτομίας. Ωστόσο, η πιθανή ενσωμάτωση της ΓΤΝ και των συστημάτων ΔΤΠ στο μέλλον μπορεί να εγείρει αξιοσημείωτες ηθικές και κοινωνικές ανησυχίες. Ζητήματα σχετικά με την προστασία της ιδιωτικής ζωής, τους κινδύνους της ασφάλειας, τις προκαταλήψεις, τις προκλήσεις δικαιοσύνης, την έλλειψη αυτονομίας, τον έλεγχο των ανθρώπων καθώς και αναδυόμενα ζητήματα όπως η έλλειψη ηθικής δημιουργίας περιεχομένου. Δεδομένης της σχετικά νέας φύσης της ΓΤΝ, στην παρούσα εργασία ερευνώνται οι ηθικές και κοινωνικές ανησυχίες που προκύπτουν όταν η ΓΤΝ και τα περιβάλλοντα ΔΤΠ ενσωματώνονται. Ρίχνοντας φως σε αυτούς τους ηθικούς και

κοινωνικούς προβληματισμούς επιδιώκεται η προώθηση μιας διαφοροποιημένης κατανόησης της κρίσιμης ισορροπίας μεταξύ τεχνολογικής προόδου και υπεύθυνης ανάπτυξης, καθοδηγώντας τα ενδιαφερόμενα μέρη προς μια ηθικά ορθή ενσωμάτωση της ΓΤΝ στα οικοσυστήματα ΔΤΠ^[28].

Ορισμός Διαδικτύου των Πραγμάτων - ΔΤΠ

Το ΔΤΠ είναι ένα δίκτυο φυσικών συσκευών, όπως οχήματα, συσκευές και άλλες τεχνολογικές εφευρέσεις, τα οποία είναι εξοπλισμένα με αισθητήρες, λογισμικό και συνδεσιμότητα δικτύου η οποία τους επιτρέπει να συλλέγουν και να ανταλλάσσουν δεδομένα. Το ΔΤΠ επιτρέπει σε αυτές τις έξυπνες συσκευές να επικοινωνούν μεταξύ τους και με άλλες συσκευές που είναι συνδεδεμένες στο διαδίκτυο, όπως τα έξυπνα τηλέφωνα και οι πύλες δικτύων¹⁵ για να δημιουργήσουν ένα μεγάλο δίκτυο διασυνδεδεμένων συσκευών που μπορούν να ανταλλάσσουν δεδομένα και να εκτελούν εργασίες ανεξάρτητα.

Ηθικές και Κοινωνικές Ανησυχίες σχετικά με τα Ρίσκα Παραγόμενου Περιεχόμενου

Μια από τις βασικές ηθικές και κοινωνικές ανησυχίες είναι η πιθανότητα παραπληροφόρησης και εξαπάτησης. Το περιεχόμενο που παράγεται με ΤΝ μπορεί να χρησιμοποιηθεί για την δημιουργία παραπλανητικών πληροφοριών που μπορούν να διαδοθούν μέσω συσκευών και δικτύων ΔΤΠ. Μια άλλη ηθική και κοινωνική ανησυχία είναι η έλλειψη προστασίας των δικαιωμάτων πνευματικής ιδιοκτησίας σε περίπτωση που το περιεχόμενο που παράγεται από ΤΝ βασίζεται στην χρήση υλικού το οποίο προστατεύεται από πνευματικά δικαιώματα, εμπορικά σήματα ή άλλες πληροφορίες ιδιοκτησίας. Στην βιομηχανία, χρησιμοποιείται η ΓΤΝ και το Βιομηχανικό Διαδίκτυο των Πραγμάτων – ΒΔΤΠ ή Industrial Internet of Things – IIoT για να βελτιστοποιήσει τις κατασκευαστικές διαδικασίες. Η χρήση υλικού που προστατεύεται από πνευματικά δικαιώματα για την εκπαίδευση μοντέλων ΓΤΝ μπορεί να οδηγήσει σε ισχυρισμούς παραβίασης δικαιωμάτων. Δεδομένης της ικανότητας των συστημάτων ΔΤΠ να αναλύουν δεδομένα που χρησιμοποιούνται για την παροχή πληροφοριών και την λήψη

¹⁵ Πύλες δικτύων ή Gateways είναι ένα δίκτυο κόμβων το οποίο χρησιμοποιείται στις τηλεπικοινωνίες και ενώνει δύο δίκτυα με διαφορετικά πρωτόκολλα μετάδοσης.

αποφάσεων, εάν ψευδή δεδομένα που παράγονται από ΤΝ διεισδύσουν στο σύνολο δεδομένων, μπορούν να οδηγήσουν σε ανακριβείς αναλύσεις, παραπλανητικές προβλέψεις που βασίζονται σε ψευδείς πληροφορίες, καθώς και σε ατοπήματα που μπορούν να προέλθουν από εσφαλμένη επεξεργασία δεδομένων.

ALTAI Ερωτηματολόγιο: Δικαιοσύνη

Τα μοντέλα ΓΤΝ που εκπαιδεύονται χρησιμοποιώντας μεροληπτικά σύνολα δεδομένων έχουν την δυνατότητα να ενισχύσουν τις υπάρχουσες κοινωνικές προκαταλήψεις. Για παράδειγμα όταν μεροληπτικά δεδομένα εκπαίδευσης χρησιμοποιούνται σε συστήματα ΔΤΠ για την υγειονομική περίθαλψη, μπορούν να οδηγήσουν σε ανισότητες στην διάγνωση καταστάσεων ή στη σύσταση θεραπειών για διάφορες δημογραφικές ομάδες. Μια μελέτη που διεξήχθη από το Πανεπιστήμιο της Ανατολικής Αγγλίας στο Ηνωμένο Βασίλειο υποδεικνύει ενδείξεις αριστερής προκατάληψης στο ChatGPT, ένα μοντέλο ΓΤΝ που αναπτύχθηκε από την OpenAI. Η προκατάληψη αυτή φαίνεται να εκδηλώνεται ως προτίμηση προς το Εργατικό Κόμμα του Ηνωμένου Βασιλείου και προς το Δημοκρατικό Κόμμα των ΗΠΑ. Η μελέτη υποδηλώνει ότι δεδομένης της αυξανόμενης δημόσιας χρήσης και εξάρτησης από πλατφόρμες ΤΝ όπως το ChatGPT, οι ανησυχίες σχετικά με την δικαιοσύνη και την μεροληψία ενδέχεται να αυξηθούν, ιδίως καθώς η χρήση του ChatGPT και η ενσωμάτωσή του σε εμπορικές και καταναλωτικές πλατφόρμες αυξάνεται.

Οι πολύπλοκοι αλγόριθμοι των μοντέλων ΓΤΝ και των συστημάτων ΔΤΠ μπορούν να εισάγουν ακούσιες προκαταλήψεις, οδηγώντας ενδεχομένως σε αποτελέσματα που εισάγουν διακρίσεις ή ενισχύοντας κοινωνικές προκαταλήψεις. Για παράδειγμα, η μεροληπτική τοποθέτηση αισθητήρων ΔΤΠ σε συγκεκριμένες γειτονίες μπορεί να οδηγήσει σε άνιση συλλογή δεδομένων, ενώ οι μεροληπτικοί αλγόριθμοι στα συστήματα ΔΤΠ μπορεί να οδηγήσουν σε άδικη κατανομή πόρων για συγκεκριμένες ομάδες μέσω της μαζικής επιτήρησης και της προληπτικής αστυνόμευσης. Σε μια υπόθεση που αφορά την ΤΝ και την μεροληψία του ΔΤΠ, η οποία πλήττει ιδιαίτερα τους έγχρωμους ανθρώπους, το σύστημα αναγνώρισης προσώπου με βάση τη ΤΝ αναγνώρισε λανθασμένα μια έγκυο γυναίκα ως εμπλεκόμενη σε μια βίαιη κλοπή αυτοκινήτου μετά την αντιστοίχιση μιας παλιότερης φωτογραφίας της με υλικό παρακολούθησης στο διαδίκτυο. Όταν εξετάζεται το πλαίσιο της πιθανής ενσωμάτωσης της ΓΤΝ και του ΔΤΠ, η παραπάνω

περίπτωση παρέχει πληροφορίες για πιθανές προκαταλήψεις και προκλήσεις που μπορούν να προκύψουν σε μοντέλα που αντικατοπτρίζουν προκαταλήψεις ενσωματωμένες στα δεδομένα εκπαίδευσής τους.

ALTAI Ερωτηματολόγιο: Ιδιωτικότητα και Διακυβέρνηση Δεδομένων, Τεχνική Ευρωστία και Ασφάλεια

Ένας άλλος κίνδυνος που συνδέεται με την χρήση ΓΤΝ είναι η πιθανότητα παραβίασης ή διαρροής δεδομένων και αυτό οφείλεται στο γεγονός ότι τα εργαλεία αυτά έχουν σχεδιαστεί για την αυτόματη παραγωγή περιεχομένου, συλλέγουν και αποθηκεύουν συχνά μεγάλες ποσότητες δεδομένων. Επιπλέον, η πολυπλοκότητα και η εκτεταμένη διασύνδεση των συσκευών ΔΤΠ παρουσιάζουν προκλήσεις για την προστασία της ιδιωτικής ζωής και την ασφάλεια, διότι κάθε συσκευή αποτελεί πιθανό σημείο εισόδου για τους επιτιθέμενους. Ως εκ τούτου, η ενσωμάτωση της ΓΤΝ σε περιβάλλοντα ΔΤΠ έχει την δυνατότητα να εισάγει ένα πρόσθετο επίπεδο πολυπλοκότητας, δεδομένης της εντατικής ανάγκης πόρων αυτών των μοντέλων και της δυνατότητας εισαγωγής νέων οδών για επιθέσεις. Παρά τις συνεχείς προσπάθειες για την ανάπτυξη ηθικών και ασφαλών συστημάτων ΤΝ, η ανάλυσή τους υπογραμμίζει ότι το ChatGPT εξακολουθεί να έχει την ικανότητα να παράγει ακατάλληλο περιεχόμενο, αναδεικνύοντας την εκτεταμένη φύση αυτών των κινδύνων.

Το απόρρητο και η ασφάλεια αποτελούν σημαντική ανησυχία για τα άτομα, τους οργανισμούς και την κοινωνία, αλλά οι κίνδυνοι οι οποίοι ελλοχεύουν δεν είναι εύκολα προφανείς στους περισσότερους. Η σύγκλιση της ΓΤΝ και του ΔΤΠ εγείρει ανησυχίες σχετικά με την ασφάλεια και την παραβίαση της ιδιωτικής ζωής σε περιπτώσεις όπου ευαίσθητες πληροφορίες μπορεί να εκτεθούν κατά λάθος ή να χρησιμοποιηθούν καταχρηστικά χωρίς την δέουσα συγκατάθεση. Για τον μετριασμό αυτών των προκλήσεων όσον αφορά την ιδιωτικότητα και την ασφάλεια, αποτελεί επιτακτική ανάγκη να υιοθετηθεί μια πολυεπίπεδη προσέγγιση ασφάλειας. Η ιδέα περιλαμβάνει ασφαλή σχεδιασμό μοντέλων ΓΤΝ, ισχυρούς μηχανισμούς ελέγχου ταυτότητας, πρόσβασης και τεχνικές κρυπτογράφησης για την προστασία της συλλογής, μετάδοσης και αποθήκευσης δεδομένων. Πρέπει να εφαρμόζονται τακτικές ενημερώσεις και επιδιορθώσεις ασφαλείας

για την αντιμετώπιση τυχόν ταυτοποιημένων τρωτών σημείων τόσο στα μοντέλα ΓΤΝ όσο και στην υποκείμενη υποδομή ΔΤΠ.

ALTAI Ερωτηματολόγιο: Περιβαλλοντολογικό και Κοινωνικό Ευ ζήν

Καθώς τα συστήματα ΓΤΝ και ΔΤΠ εξελίσσονται σε εκλεπτυσμένα συστήματα, προκύπτουν πιθανές επιπτώσεις σχετικά με την μείωση της ανθρώπινης δράσης και της εξουσίας λήψης αποφάσεων στο οικοσύστημα. Η πιθανή διασταύρωση της ΓΤΝ, του ΔΤΠ και της ιδιοκτησίας ψηφιακής ταυτότητας εγείρει ανησυχίες, ιδίως όσον αφορά την διαφάνεια και την αναζήτηση της κατάλληλης ισορροπίας μεταξύ της αυτονομίας της ΤΝ και του ανθρώπινου ελέγχου καθώς αποτελεί μείζον ζήτημα η διατήρηση των ανθρώπινων αξιών, προτιμήσεων, ηθικών και κοινωνικών προβληματισμών στον πυρήνα της ενσωμάτωσης της ΓΤΝ στα ΔΤΠ. Η πρόσβαση στους ίδιους αλγορίθμους και εργαλεία ΤΝ αυξάνει τον κίνδυνο παραγκωνισμού της πρωτοτυπίας και της δημιουργικότητας προς όφελος της αυτοματοποιημένης μίμησης. Αυτή η κατάσταση μπορεί ενδεχομένως να καταπνίξει την εμφάνιση νέων ειδών στον κινηματογράφο, την τέχνη ή την μουσική καθώς βρίσκονται οι βιομηχανίες παγιδευμένες σε έναν κύκλο ανακυκλωμένων ιδεών και αντιγραφής περιεχομένου.

Η αυτονομία κατέχει κεντρικό ρόλο στην ηθική και πολιτική φιλοσοφία εδώ και αιώνες, καθιερώνοντας τον εαυτό της ως κεντρικό στοιχείο τόσο της δικαιοσύνης όσο και της ευημερίας. Ως εκ τούτου, είναι σημαντικό να διασφαλιστεί η αυτονομία και ο ανθρώπινος έλεγχος από τον σχεδιασμό των συστημάτων μέχρι την συγκατάθεση για την συλλογή και την επεξεργασία δεδομένων. Οι χρήστες πρέπει να έχουν πλήρη κατανόηση του τρόπου με τον οποίο χρησιμοποιούνται τα δεδομένα τους και να έχουν την ελευθερία να εξαιρεθούν, εάν το επιθυμούν. Επιπλέον, το περιεχόμενο που απορρέει από αυτά τα μοντέλα πρέπει να επιδεικνύει ακρίβεια και σεβασμό προς τα άτομα που αντιπροσωπεύουν, καθώς και να αποφεύγει την παραποίηση ή την χειραγώγηση.

ALTAI Ερωτηματολόγιο: Διαφάνεια, Λογοδοσία

Το μέλλον των τεχνολογιών ΓΤΝ στο περιβάλλον του ΔΤΠ έχει την δυνατότητα να παράγει παραπλανητικό, επιζήμιο ή ακατάλληλο περιεχόμενο σε τομείς όπως η προώθηση προϊόντων και η εκπαίδευση. Στον τομέα της προώθησης προϊόντων η ΓΤΝ μπορεί να χρησιμοποιηθεί για ανήθικες επιχειρηματικές πρακτικές, όπως για την χειραγώγηση

διαδικτυακών κριτικών για λόγους κέρδους ή για την δημιουργία χιλιάδων λογαριασμών με πλαστές ταυτότητες με στόχο την αύξηση των αλληλεπιδράσεων. Η ενσωμάτωση της ΓΤΝ σε έξυπνες συσκευές ΔΤΠ μπορεί να επιτρέψει στις επιχειρήσεις να έχουν την δυνατότητα να αναλύουν εκτεταμένα δεδομένα και να εντοπίζουν μοτίβα στην συμπεριφορά των καταναλωτών, επιτρέποντας την δημιουργία καλά δομημένων στρατηγικών προώθησης, ωστόσο εάν το παραγόμενο περιεχόμενο δεν ελέγχεται και δεν επεξεργάζεται επαρκώς, μπορεί να οδηγήσει σε ανακρίβειες ή παραπλανητικές πληροφορίες πλήττοντας έτσι την εταιρεία ή την φήμη της ονομασίας της.

Στην εκπαίδευση, οι τεχνολογικές εξελίξεις και το διαδίκτυο έχουν το πιο βαθύ αντίκτυπο στον τρόπο διδασκαλίας και μάθησης. Το ΔΤΠ επιτρέπει στην εκπαιδευτική κοινότητα και το σχολείο να γίνουν ένα περιβάλλον τεχνολογικής μάθησης και κατάρτισης. Ωστόσο από την αρχική του κυκλοφορία, τα εργαλεία ΓΤΝ όπως το ChatGPT έχουν αποτελέσει αντικείμενο έντονης συζήτησης στην εκπαίδευση. Οι ανησυχίες για την ΓΤΝ όπως η κατάχρηση ψευδών ή παραπλανητικών πληροφοριών που παρουσιάζονται ως γεγονότα είναι επίκαιρες. Το ChatGPT έχει τοποθετηθεί ως εναλλακτική λύση στην θέση μηχανών αναζήτησης όπως η Google για την αναζήτηση πληροφοριών και οι μαθητές μπορεί να καταστούν ευάλωτοι στην παραπληροφόρηση, μειώνοντας έτσι τις δεξιότητες γραμματισμού τους. Επιπλέον, το ChatGPT μπορεί να χρησιμοποιηθεί για την δημιουργία υλικού που δεν είναι μόνο ανακριβές αλλά και ιδεολογικά προκατειλημμένο και μπορεί να διακινεί θεωρίες συνωμοσίας ή να παραθέτει ψευδείς επιστημονικές μελέτες^[28].

Η σύγκλιση της ΓΤΝ με τα συστήματα ΔΤΠ αποτελεί σημαντικό βήμα για την αξιοποίηση του τεράστιου δυναμικού των δεδομένων που παράγονται από έξυπνες συσκευές. Η σύγκλιση αυτή αποτελεί ένα σύστημα ΤΝ περιορισμένου ρίσκου καθώς σχετίζεται κατά βάση με ρίσκα που συνδέονται με την έλλειψη διαφάνειας στον τρόπο χρήσης του ή εξαγωγής των συμπερασμάτων του. Σε τέτοιες περιπτώσεις το ΑΙ Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας προκειμένου να διασφαλίσει πως οι άνθρωποι πληροφορούνται όταν χρειάζεται, ενισχύοντας την εμπιστοσύνη στο σύστημα ΤΝ και διασφαλίζοντας την εύρυθμη λειτουργία του.

3.3 Ερευνητικές Δημοσιεύσεις Ηθικών Λύσεων Τεχνητής Νοημοσύνης στην Αυτοκινητοβιομηχανία

Ο τρίτος τομέας που αναπτύσσεται στην παρούσα εργασία με ιδιαίτερο ερευνητικό ενδιαφέρον είναι τα Αυτόνομα Οχήματα – ΑΟ ή Autonomous Vehicles – AV τα οποία υποστηρίζονται από την ΤΝ και αντιπροσωπεύουν ένα μετασχηματιστικό άλμα στην τεχνολογία των μεταφορών, υποσχόμενα μειωμένη κυκλοφοριακή συμφόρηση και αυξημένη κινητικότητα. Ωστόσο, η ενσωμάτωση της ΤΝ στα ΑΟ εισάγει πληθώρα ηθικών προβληματισμών, ιδίως όσον αφορά τη λεπτή ισορροπία μεταξύ ασφάλειας και ιδιωτικότητας. Καθώς τα ΑΟ γίνονται όλο και πιο διαδεδομένα στους δρόμους μας, είναι επιτακτική ανάγκη να εξεταστούν τα ηθικά διλήμματα για να διασφαλιστεί ότι τα οφέλη αυτής της τεχνολογίας υλοποιούνται χωρίς να διακυβευθούν θεμελιώδεις ανθρώπινες αξίες. Επί του παρόντος, ο κλάδος των αυτόνομων οχημάτων βρίσκεται σε μια δυναμική φάση πειραματισμού και επέκτασης. Αρκετές εταιρείες, από γίγαντες της αυτοκινητοβιομηχανίας έως νεοφυείς επιχειρήσεις τεχνολογίας, επενδύουν σημαντικά στην έρευνα και την ανάπτυξη για να φέρουν αυτόνομα οχήματα στην αγορά.

Η εφαρμογή των ΑΟ με ΤΝ εγείρει σημαντικές ανησυχίες σχετικά με την προστασία ιδιωτικής ζωής των δεδομένων. Όπως όλα τα συστήματα ΤΝ, έτσι και τα ΑΟ, βασίζονται σε τεράστιες ποσότητες δεδομένων, συμπεριλαμβανομένων των δεδομένων αισθητήρων, των πληροφοριών θέσης, των προσωπικών προτιμήσεων προκειμένου να επιτευχθεί ασφαλής και αποτελεσματική πλοήγηση. Τα δεδομένα είναι απαραίτητα για την ενίσχυση των επιδόσεων των ΑΟ, παρουσιάζουν επίσης κινδύνους για την προστασία της ιδιωτικής ζωής, καθώς ενδέχεται να περιέχουν ευαίσθητες προσωπικές πληροφορίες. Επιπλέον, η συλλογή, η αποθήκευση και η κοινή χρήση αυτών των δεδομένων εγείρουν ερωτήματα σχετικά με τη συγκατάθεση, την διαφάνεια και την αυτονομία του ατόμου.

Εκτός από τις ανησυχίες για την προστασία της ιδιωτικής ζωής, οι ηθικές επιπτώσεις των ΑΟ με ΤΝ επεκτείνονται σε θέματα ασφάλειας και αξιοπιστίας. Τα ΑΟ πρέπει να λαμβάνουν αποφάσεις σε κλάσματα δευτερολέπτου που ενδεχομένως απειλούν την ζωή των ανθρώπων, εγείροντας ερωτήματα σχετικά με τις ηθικές αρχές που διέπουν αυτές τις αποφάσεις. Επιπλέον, ο προσδιορισμός της ευθύνης σε περίπτωση ατυχημάτων που

αφορούν τα ΑΟ είναι ένα πολύπλοκο ζήτημα, καθώς η ευθύνη μπορεί να ανήκει στους κατασκευαστές, στους προγραμματιστές, στους χειριστές ή ακόμη και στην ίδια την ΤΝ.

Ορισμός Αυτόνομων Αυτοκινήτων - ΑΟ

Η αυτόνομη οδήγηση αποτελεί μία από τις πιο μελετημένες τεχνολογίες στον τομέα των μεταφορών. Ενώ η κατανόηση των αυτόνομων οχημάτων μπορεί να είναι απλή σαν ιδέα – οχήματα ικανά να οδηγούν από μόνα τους – η εμβάθυνση αποκαλύπτει έναν διαφορετικό ορισμό. Τα αυτόνομα οχήματα είναι εξοπλισμένα με Προηγμένα Συστήματα Υποβοήθησης Οδηγού – ΠΣΥΟ ή Advanced Driver Assistance Systems - ADAS τα οποία είναι σχεδιασμένα να υποστηρίξουν τους οδηγούς τόσο σε συνηθισμένες διαδρομές όσο και σε κρίσιμες καταστάσεις όπως οι συγκρούσεις. Ο ορισμός αυτός περιλαμβάνει διάφορα επίπεδα αυτονομίας, γεγονός που οδηγεί στην καθιέρωση μιας αυτοκινητοβιομηχανικής τυποποίησης από την Κοινωνία Μηχανικών Αυτοκινήτων – ΚΜΑ ή Society of Automotive Engineers – SAE^[30]. Τα επίπεδα αυτά κυμαίνονται από το σημείο 0 έως 5 τα οποία αντιπροσωπεύουν μηδαμινή έως πλήρη αυτοματοποίηση χωρίς να απαιτείται ανθρώπινη παρέμβαση.

I) Επίπεδο 0: Μηδενική αυτονομία – Ο οδηγός διατηρεί τον πλήρη έλεγχο.

II) Επίπεδο 1: Υποβοήθεια οδηγού - Το όχημα βοηθά με το χειρισμό του τιμονιού και την πέδηση.

III) Επίπεδο 2: Μερική αυτονομία – Το όχημα ελέγχει τον χειρισμό του τιμονιού και την επιτάχυνση – επιβράδυνση, με τον οδηγό να την επιβλέπει.

IV) Επίπεδο 3: Υπο συνθήκη αυτονομία – Το όχημα μπορεί να διαχειριστεί τις περισσότερες οδηγικές απαιτήσεις χωρίς την ανθρώπινη παρέμβαση, απαιτώντας από τον οδηγό να είναι έτοιμος να επέμβει.

V) Επίπεδο 4: Υψηλή αυτονομία – Το όχημα είναι σε θέση να εκτελέσει τις περισσότερες οδηγικές εργασίες, περιορίζεται από συγκεκριμένες συνθήκες ή περιβάλλοντα.

VI) Επίπεδο 5: Πλήρης αυτονομία – Το όχημα μπορεί να διαχειριστεί όλες τις οδηγικές απαιτήσεις χωρίς καμία ανθρώπινη παρέμβαση.

Θέματα Δεοντολογίας για Αυτόνομα Αυτοκίνητα

Η ηθική, παραδοσιακά επικεντρώνεται σε λογικούς συλλογισμούς και ηθικές εκτιμήσεις, οριοθετεί τις σωστές και τις λανθασμένες ανθρώπινες ενέργειες. Ωστόσο, η μετάβαση από λήψη αποφάσεων με βάση τον άνθρωπο σε βάση με το σύστημα εισάγει ένα στρώμα απρόβλεπτων συμπεριφορών, διλημμάτων και πολυπλοκότητας. Οι πολιτισμικοί και κοινωνικοί παράγοντες επηρεάζουν την ηθική λήψης αποφάσεων στα αυτόνομα οχήματα, ιδίως σε σενάρια που απειλούν την ζωή. Οι πολιτισμικές διαφορές οδηγούν σε διαφορετικές ιεραρχήσεις, περιπλέκοντας την λήψη αποφάσεων.

Η αντιμετώπιση αυτών των ζητημάτων απαιτεί βαθιά κατανόηση των πολιτισμικών και κοινωνικών πλαισίων, καθώς η ηθική συμπεριφορά διαφέρει από πολιτισμό σε πολιτισμό και πλαίσιο. Η προσαρμοστικότητα των ΑΟ στις πολιτισμικές διαφοροποιήσεις είναι ύψιστης σημασίας, θέτοντας σύνθετα ερωτήματα σχετικά με την συμπεριφορά τους σε διάφορα σενάρια κυκλοφορίας.

ALTAI Ερωτηματολόγιο: Διαφάνεια, Τεχνική Ευρωστία και Ασφάλεια

Η διαφάνεια, η αξιοπιστία και η ασφάλεια αποτελούν θεμελιώδεις πυλώνες για έμπιστη ΤΝ, ενισχύοντας την εμπιστοσύνη μεταξύ των χρηστών, των ενδιαφερόμενων μερών και της κοινωνίας γενικότερα. Η διαφάνεια διευκολύνει την κατανόηση από τους χρήστες των διαδικασιών λήψης απόφασης της ΤΝ, ενώ η αξιοπιστία διασφαλίζει συνεπή απόδοση και προβλεψιμότητα. Η ασφάλεια είναι ζωτικής σημασίας για την αποτροπή ακούσιων βλαβών ή συνεπειών και υπογραμμίζει τη σημασία του σχεδιασμού και της υλοποίησης ενός συστήματος.

Ένα διαφανές σύστημα ΤΝ επιτρέπει στους χρήστες του να κατανοήσουν τις μηχανικές διαδικασίες, ενισχύοντας την εμπιστοσύνη και την αξιοπιστία του. Η αξιοπιστία απαιτεί τεχνικές μεθόδους για την ανθρώπινη ευθύνη, συμπεριλαμβανομένων των διαδρομών ελέγχου και των αναλυτικών τεχνικών, ώστε να διασφαλίζεται η συνεχής απόδοση.

Η ασφάλεια, αποτελεί αναπόσπαστο στοιχείο ανάπτυξης ΤΝ, απαιτεί αυστηρές στρατηγικές διαχείρισης και εποπτείας για μετριασμό των κινδύνων και τη διασφάλιση της ακεραιότητας του συστήματος. Σε τελική ανάλυση, ένα ασφαλές σύστημα δίνει προτεραιότητα στην πρόληψη ακούσιων βλαβών και απροσδόκητων συνεπειών^[29].

Μηχανική Μάθηση και Βαθιά Μάθηση στα Αυτόνομα Οχήματα

Η MM και συγκεκριμένα η BM αποτελεί τον ακρογωνιαίο λίθο των τεχνολογιών ΤΝ που χρησιμοποιούνται στα ΑΟ. Τα νευρωνικά δίκτυα, εμπνευσμένα από τη δομή του ανθρώπινου εγκεφάλου, έχουν επιδείξει εξαιρετικές ικανότητες σε εργασίες όπως η αντίληψη, η λήψη αποφάσεων και ο έλεγχος. Τα Συνελικτικά Νευρωνικά Δίκτυα – ΣΝΔ υπερέχουν στην όραση υπολογιστών επιτρέποντας στα οχήματα να κατανοούν το περιβάλλον τους μέσω καμερών. Τα Επαναληπτικά Νευρωνικά Δίκτυα – ΕΝΔ διευκολύνουν την λήψη διαδοχικών αποφάσεων σε πολύπλοκες διαστραυρώσεις και στον χειρισμό της κυκλοφορίας.

Σύντηξη Αισθητήρων για Αντίληψη

Τα ΑΟ βασίζονται σε μια σειρά αισθητήρων, Ανίχνευση Φωτός και Εμβέλειας – ΑΦΕ ή Light Detection And Ranging - LiDAR¹⁶, ραντάρ, κάμερες, υπερήχους και άλλους. Οι τεχνικές σύντηξης αισθητήρων συλλέγουν, επεξεργάζονται και ενσωματώνουν δεδομένα από αυτές τις διαφορετικές πηγές για να δημιουργήσουν μια ολοκληρωμένη κατανόηση περιβάλλοντος του οχήματος. Αυτή η σύντηξη είναι ζωτικής σημασίας για ισχυρή αντίληψη, ιδίως σε δυσμενείς καιρικές συνθήκες ή σε απαιτητικά περιβάλλοντα.

3.3.1 Ταυτόχρονος Εντοπισμός και Χαρτογράφηση – TEX (0 – 3)

Ο ακριβής εντοπισμός αποτελεί βάση της αυτόνομης οδήγησης. Οι τεχνικές Ταυτόχρονου Εντοπισμού και Χαρτογράφησης – TEX ή Simultaneous Localization and Mapping – SLM επιτρέπουν στα οχήματα να δημιουργούν και να ενημερώνουν λεπτομερείς χάρτες του περιβάλλοντος τους και ταυτόχρονα να προσδιορίζουν την ακριβή τους θέση μέσα σε αυτούς τους χάρτες. Αυτή η ικανότητα είναι θεμελιώδης για την ασφαλή πλοήγηση, ιδίως σε αστικά περιβάλλοντα με πολύπλοκες οδικές δομές.

¹⁶ LiDAR – Light Detection And Ranging ή Ανίχνευση Φωτός και Εμβέλειας, είναι τεχνολογία η οποία βασίζεται στην εκπομπή παλμικής ακτινοβολίας λέιζερ στην ατμόσφαιρα.

ALTAI Ερωτηματολόγιο: Διαφάνεια, Τεχνική Ευρωστία και Ασφάλεια

Η διασφάλιση της ασφάλειας των ΑΟ οχημάτων περιλαμβάνει μια πολύπλευρη προσέγγιση. Πρακτικά, περιλαμβάνει αυστηρές δοκιμές, πλεόνασμα σε κριτικά συστήματα και μηχανισμούς ασφαλείας. Χαρακτηριστικά ασφαλείας που βασίζονται στην ΤΝ όπως η αποφυγή συγκρούσεων και η ανίχνευση πεζών είναι ύψιστης σημασίας για την αποφυγή των ατυχημάτων.

Ηθικά Διλήμματα

Η ενσωματωμένη ΤΝ σε ΑΟ διακατέχεται από ποικίλα ηθικά διλήμματα, με επιτομή το «πρόβλημα του τρόλεϊ¹⁷». Αυτό το πείραμα σκέψης θέτει ερωτήματα σχετικά με τον ηθικό προγραμματισμό των οχημάτων όταν έρχονται αντιμέτωπα με αποφάσεις ζωής και θανάτου. Η αντιμετώπιση αυτών των διλημάτων απαιτεί μια λεπτή ισορροπία μεταξύ της διατήρησης της ανθρώπινης ζωής και της τήρησης των νομικών και ηθικών αρχών^[29].

ALTAI Ερωτηματολόγιο: Ιδιωτικότητα και Διακυβέρνηση δεδομένων - Ο Κύκλος Ζωής Ανάπτυξης Συστημάτων ΤΝ στα ΑΟ: Εκπαίδευση και Ανάπτυξη μοντέλων

I) Συλλογή και Προεπεξεργασία Δεδομένων: Συγκέντρωση τεράστιου όγκου δεδομένων από αισθητήρες του πραγματικού κόσμου, προϋπάρχοντα σύνολα δεδομένων και άλλες πηγές, όπως συνθετικά σύνολα δεδομένων. Καθαρισμός και προεπεξεργασία δεδομένων ώστε να καταστούν κατάλληλα για μοντέλα MM.

II) Εκπαίδευση Μοντέλου: Χρησιμοποιεί μοντέλα εκμάθησης, όπως Νευρωνικά Δίκτυα, Βαθιά Μάθηση και Φυσική Γλώσσα Επεξεργασίας για την κατανόηση μοτίβων και δομών, με βάση τα δεδομένα. Εκπαίδευση των μοντέλων μέχρι ένα επιθυμητό επίπεδο ακρίβειας με βάση κάθε σενάριο ή σε γενικές αφηρημένες περιπτώσεις η δυνατότητα να εξάγονται τα πρότυπα κατά την διάρκεια της ζωντανής λειτουργίας των οχημάτων.

III) Παραγωγή Μοντέλου: Εκπαίδευση μοντέλων για την εκτέλεση μιας συγκεκριμένης εργασίας λήψης απόφασης, λειτουργιών με βάση τα διδαχθέντα πρότυπα. Αυτά τα

¹⁷ Το πρόβλημα του τρόλεϊ: το τρόλεϊ διασχίζει τις ράγες. Στο πλάι της σιδηρογραμμικής γραμμής βρίσκεται ο σταθμάρχης. Μπροστά του είναι ένας μοχλός που αλλάζει την πορεία του τρόλεϊ. Μπορεί να σώσει τους πέντε, σκοτώνοντας τον έναν ή να σώσει τον έναν σκοτώνοντας τους πέντε. Ανάλογα με την επιλογή παρέμβασης στην πορεία του τρόλεϊ με χρήση του μοχλού επιλέγεται η μια εκ των δύο παραπάνω επιλογών. Πώς μετριέται η αξία της ανθρώπινης ζωής σε κάθε μια επιλογή και ποια η ηθικά σωστή επιλογή;^[32]

μοντέλα μπορούν να χρησιμοποιούν διάφορες αρχιτεκτονικές όπως δέντρα αποφάσεων, τυχαία δάση, δέντρα παλινδρόμησης, βαθιά στρώματα, μάθηση συνόλου κ.λπ.

IV) Βελτίωση και Βελτιστοποίηση Κώδικα: Βελτίωση της ποιότητας, της αναγνωσιμότητας και της λειτουργικότητάς του. Η επεξεργασία μετά την δημιουργία εξασφαλίζει ότι ο κώδικας τηρεί τα πρότυπα κωδικοποίησης, τις συμβάσεις και τις απαιτήσεις.

V) Αξιολόγηση Ποιότητας: Αξιολόγηση του παραγόμενου κώδικα ως προς την τήρηση των προβλεπόμενων λειτουργιών. Αυτό περιλαμβάνει διαδικασίες δοκιμής, αποσφαλμάτωσης και επικύρωσης.

VI) Ενσωμάτωση και Ανάπτυξη: Ενσωμάτωση του μοντέλου σε ένα ευρύτερο σύστημα υπό ανάπτυξη για αυτόνομη εφαρμογή. Εφαρμογή και έλεγχος της εφαρμογής λογισμικού εμπιρεύοντας το νέο μοντέλο με χρήση πολλαπλών μεθόδων όπως Λογισμικό Εντός Βρόχου - ΛΣΒ ή Software-in-the-Loop – SIL¹⁸, Υλικό Εντός Βρόχου – ΥΕΒ ή Hardware-in-the-Loop – HIL¹⁹, Άνθρωπος Στον Βρόχο – ΑΣΒ ή Human-in-the-Loop – HIL²⁰ κ.λπ., χρησιμοποιώντας προσομοίωση, κλειστή πορεία και περιορισμένα περιβάλλοντα δημόσιων δρόμων. Κάποια μοντέλα εκπαιδεύονται προκειμένου να βελτιώνουν την γνώση τους ακόμα και μετά την εφαρμογή τους σε λειτουργία, τα οποία χρειάζονται επανέλεγχο για μελλοντικές κατευθύνσεις γνώσης ώστε να διασφαλιστούν οι ηθικές επιταγές και άλλες απαιτήσεις.

Με την χρήση μιας συστηματικής διαδικασίας αυξάνεται η αυτοπεποίθηση σε κάθε μοντέλο που αναπτύσσεται και εφαρμόζεται σε διάφορα υποσυστήματα ΑΟ όπως στην αντίληψη, στην πορεία, στον έλεγχο και στις εφαρμογές Διεπαφής Ανθρώπου Μηχανής.

¹⁸ Η προσομοίωση Λογισμικού Στον Βρόχο αντιπροσωπεύει την ενσωμάτωση μεταγλωττισμένου κώδικα παραγωγής σε μαθηματικοποιημένο μοντέλο προσομοίωσης παρέχοντας στους μηχανικούς ένα ψηφιακό περιβάλλον για ανάπτυξη και έλεγχο σύνθετων συστημάτων^[33].

¹⁹ Ο έλεγχος Υλικό Στον Βρόχο είναι μια τεχνική όπου πραγματικά σήματα από έναν ελεγκτή είναι συνδεδεμένα σε ένα σύστημα το οποίο προσομοιώνει την πραγματικότητα παραπέμποντας τον ελεγκτή να λειτουργεί σαν να είναι συνδεδεμένος στο τελικό προϊόν^[34].

²⁰ Ο έλεγχος Άνθρωπος Στον Βρόχο αποτελεί μια μίξη Επιβλεπόμενης Μηχανικής Μάθησης και Ενεργούς Μάθησης όπου οι άνθρωποι συμμετέχουν και στην εκπαίδευση και στον έλεγχο ενός αναπτυσσόμενου αλγόριθμου^[35].

Τα αυτόνομα οχήματα είναι εξοπλισμένα με Προηγμένα Συστήματα Υποβοήθησης Οδηγού – ΠΣΥΟ. Τα συστήματα ΤΝ που σχετίζονται με τον Εντοπισμό και την Χαρτογράφηση αποτελούν συστήματα τα οποία χαρακτηρίζονται από μηδενική αυτονομία – επίπεδο 0 έως υπό συνθήκη αυτονομία – επίπεδο 3. Πρακτικά, ο βαθμός αυτονομίας είναι τέτοιος που χρειάζεται στις περισσότερες περιπτώσεις να αναλαμβάνουν τον έλεγχο οι οδηγοί του ΑΟ. Συμπερασματικά, τα συστήματα Εντοπισμού και Χαρτογράφησης (0 - 3), ταξινομούνται ως συστήματα ΤΝ περιορισμένου ρίσκου σύμφωνα με την διαβάθμιση του ΑΙ Νομικού Πλαισίου. Το περιορισμένο ρίσκο σχετίζεται με την έλλειψη διαφάνειας. Σε τέτοιες περιπτώσεις το ΑΙ Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας προκειμένου να διασφαλίσει πως οι άνθρωποι ενημερώνονται όταν χρειάζεται. Οι πάροχοι πρέπει επίσης να διασφαλίζουν πως τα αποτελέσματα τα οποία είναι ΤΝ-παραγόμενα είναι ταυτοποιήσιμα.

3.3.2 Έλεγχος και Δημιουργία Πορείας – ΕΔΠ (4 – 5)

Οι αλγόριθμοι ελέγχου καθορίζουν τον τρόπο με τον οποίο ένα αυτόνομο όχημα κάνει φυσικούς ελιγμούς, ενώ οι αλγόριθμοι σχεδιασμού διαδρομής καθορίζουν τη διαδρομή του μέσα στο περιβάλλον. Αυτές οι τεχνολογίες συνεργάζονται για να διασφαλίσουν ότι οι κινήσεις του οχήματος είναι όχι μόνο ασφαλείς αλλά και αποτελεσματικές. Από τον προσαρμοσμένο έλεγχο ταχύτητας έως τα συστήματα διατήρησης λωρίδας κυκλοφορίας, η ΤΝ παίζει καθοριστικό ρόλο στην εκτέλεση καθηκόντων οδήγησης.

ALTAI Ερωτηματολόγιο: Τεχνική Ευρωστία και Ασφάλεια - Διασφάλιση Ποιότητας Λογισμικού και Ασφάλειας

Στα ΑΟ η ενσωμάτωση ΤΝ σε ποικίλες πλευρές της ανάπτυξης και συντήρησης λογισμικού παίζει καθοριστικό ρόλο στην διασφάλιση ευρωστίας και ασφάλειας του συστήματος συνολικά. Ο αυτοματοποιημένος έλεγχος που πραγματοποιείται βασισμένος σε εργαλεία ΤΝ, αποτελεί σημείο κλειδί της διαδικασίας ελέγχου. Αυτά τα εργαλεία αποδοτικά ταυτοποιούν σφάλματα, αδυναμίες και διασφαλίζουν ότι οι λειτουργίες του λογισμικού πράττουν όπως σχεδιάστηκαν, συμβάλλοντας έτσι στην αξιοπιστία του λογισμικού των ΑΟ.

Επιπρόσθετα, οι ικανότητες της TN επεκτείνονται σε ανάλυση και επανέλεγχο κώδικα, παρέχοντας πλήρη έλεγχο της βάσης κώδικα για την διασφάλιση της ποιότητας και τον εντοπισμό πιθανών προβλημάτων ή αδυναμιών.

Η προβλεπτική συντήρηση είναι απαραίτητη για την πιθανή εμφάνιση αποτυχιών του λογισμικού, με αποτέλεσμα την μείωση του ανενεργού χρόνου και την ενίσχυση της συνολικής αποδοτικότητας των ΑΟ.

Επιπρόσθετα, η TN-καθοδηγούμενη εύρεση ανωμαλιών και η παρακολούθηση ασφάλειας συμμετέχουν σημαντικά στην ασφάλεια των ΑΟ. Μέσω της συνεχούς παρακολούθησης του περιβάλλοντος λογισμικού, τα συστήματα TN ταυτοποιούν παράτυπα μοτίβα ή συμπεριφορές και έτσι γρήγορα αντιμετωπίζονται απειλές ασφάλειας σε πραγματικό χρόνο.

Η αξιολόγηση τρωτότητας, η οποία αποτελεί μια ακόμη εφαρμογή των εργαλείων TN, διάγει εις βάθος ελέγχους προκειμένου να εντοπίσει τις αδυναμίες του λογισμικού συστημάτων, παρέχοντας πολύτιμες πληροφορίες για τον αποτελεσματικό μετριασμό κινδύνων. Η υποβοηθούμενη συμπεριφορική ανάλυση από TN αποδεικνύεται καθοριστική στην κατανόηση των αλληλεπιδράσεων των χρηστών εντός του λογισμικού. Αυτή η δυνατότητα βοηθά στην εύρεση και στην αποφυγή ύποπτων ή κακοπροαίρετων δραστηριοτήτων, καλλιεργώντας ένα ασφαλές και αξιόπιστο οικοσύστημα ΑΟ. Επίσης, ο ρόλος της TN στην εύρεση απάτης στις εφαρμογές λογισμικού προσθέτει ένα επιπλέον επίπεδο ασφάλειας, διασφαλίζοντας την ακεραιότητα των συστημάτων στα ΑΟ και θωρακίζοντάς τα από πιθανές παραβιάσεις ασφάλειας. Συνοψίζοντας, η ενσωμάτωση TN σε όλες τις προαναφερθείσες πτυχές ενισχύει σημαντικά την συνολική ασφάλεια και την αποδοτικότητα των ΑΟ.

Ηθικοί Προβληματισμοί και Προκατάληψη στην Ανάπτυξη Λογισμικού TN στα ΑΟ

Προκειμένου να αντιμετωπιστούν οι προκλήσεις συσχετιζόμενες με την προκατάληψη χρειάζεται πρώτα να γίνουν κατανοητές και μετά να επιλυθούν έτσι ώστε να χτιστεί υπεύθυνη και δίκαιη TN για ΑΟ. Παρατίθενται τα σημεία κλειδιά υποδεικνύοντας τους ηθικούς προβληματισμούς και την προκατάληψη στα συστήματα TN των ΑΟ.

I) ALTAI Ερωτηματολόγιο - Ιδιωτικότητα και Διακυβέρνηση δεδομένων

Πρόκληση: Τα μοντέλα TN μαθαίνουν από ιστορικά δεδομένα και αν τα δεδομένα εκπαίδευσης έχουν προκατάληψη, το μοντέλο μπορεί να διαιωνίσει και να οξύνει τις τρέχουσες προκαταλήψεις.

Οι αλγόριθμοι μπορεί να κωδικοποιούν κατά λάθος προκαταλήψεις που υπάρχουν στα δεδομένα εκπαίδευσης οδηγώντας σε αποτελέσματα που αμβλύνουν τις διακρίσεις.

Τα συστήματα TN συχνά επεξεργάζονται μεγάλα ποσά προσωπικών δεδομένων, εγείροντας ανησυχίες για την ιδιωτικότητα του χρήστη.

Αντιμετώπιση: Κρίνεται απαραίτητη η αυστηρή προ-επεξεργασία δεδομένων, ποικιλία σε δεδομένα εκπαίδευσης και συνεχής παρακολούθηση προκατάληψης. Ηθικές διαδικασίες συλλογής δεδομένων χρειάζεται να υποστηρίζονται.

Τακτικοί έλεγχοι των αλγορίθμων για προκατάληψη, διαφάνεια στην αλγοριθμική λήψη απόφασης και στην ενσωμάτωση μετρικών δικαιοσύνης κατά την αξιολογήση του μοντέλου.

Εφαρμογή τεχνολογιών διατήρησης ιδιωτικότητας, λαμβάνοντας την συγκατάθεση του χρήστη κατόπιν ενημέρωσής του και τήρηση των κανονισμών προστασίας δεδομένων για την διασφάλιση της ιδιωτικής ζωής των χρηστών.

II) ALTAI Ερωτηματολόγιο - Δικαιοσύνη και Λογοδοσία

Πρόκληση: Η διασφάλιση δίκαιων αποτελεσμάτων και η εγκαθίδρυση λογοδοσίας στις αποφάσεις TN είναι σύνθετη, ειδικά όταν τα μοντέλα είναι αδιαφανή.

Αντιμετώπιση: Η εφαρμογή εξηγήσιμων τεχνικών TN, ο καθορισμός ξεκάθαρων ορίων απόφασης και η εγκαθίδρυση πλαισίων λογοδοσίας για αποφάσεις που παράγονται μέσω TN.

III) ALTAI Ερωτηματολόγιο - Διαφάνεια

Πρόκληση: Πολλά μοντέλα TN λειτουργούν ως «μαύρα-κουτιά TN²¹» με αποτέλεσμα να μην είναι ορατό το πως λαμβάνονται οι αποφάσεις. Η ασφάλεια της TN είναι απαραίτητο να διασφαλίζεται καθώς είναι κριτικής σημασίας στα ΑΟ.

Αντιμετώπιση: Η προτεραιοποίηση της εξηγησιμότητας στα μοντέλα TN, χρησιμοποιώντας ερμηνεύσιμους αλγορίθμους και παρέχοντας καθαρή και πλήρη τεκμηρίωση της συμπεριφοράς του μοντέλου.

IV) ALTAI Ερωτηματολόγιο - Τεχνική Ευρωστία και Ασφάλεια

Πρόκληση: Τα μοντέλα TN μπορεί να είναι επιρρεπή σε επιθέσεις αντιπάλων και έτσι υφίστανται ρίσκα ασφάλειας.

Αντιμετώπιση: Εύρωστοι έλεγχοι με δυνητικά σενάρια επιθέσεων, ενσωματώνοντας μέτρα ασφαλείας, και συνεχόμενες αναβαθμίσεις για την αντιμετώπιση εκκολαπτόμενων και δυνητικών απειλών.

V) ALTAI Ερωτηματολόγιο - Συμμετοχή Ενδιαφερόμενων Μερών

Πρόκληση: Η προκατάληψη στην TN μπορεί να έχει ως αποτέλεσμα τον αποκλεισμό συγκεκριμένων δημογραφικών ομάδων, ενισχύοντας τις ψηφιακές διακρίσεις.

Αντιμετώπιση: Προτεραιοποίηση ποικιλίας στην ανάπτυξη ομάδων, ενεργή αναζήτηση ανατροφοδότησης χρήστη και διενέργεια αξιολογήσεων προσβασιμότητας προκειμένου να διασφαλίζεται η συμμετοχικότητα.

VI) ALTAI Ερωτηματολόγιο - Κοινωνική και Περιβαλλοντολογική Ευεξία

Πρόκληση: Η εφαρμογή συστημάτων TN με προκατάληψη μπορεί να έχει αρνητικές κοινωνικές επιπτώσεις, επηρεάζοντας απομακρυσμένες κοινότητες δυσανάλογα.

Αντιμετώπιση: Η διενέργεια λεπτομερών αξιολογήσεων επίδρασης με συμμετέχοντες ποικίλων ενδιαφερομένων μερών στην αναπτυξιακή διαδικασία και λαμβάνοντας υπόψιν κοινωνικές επιδράσεις κατά την ανάπτυξη TN.

²¹ «Μαύρο-Κουτί TN» - Black Box AI – είναι οποιοδήποτε σύστημα TN του οποίου οι εισδοί και οι λειτουργίες δεν είναι ορατές στον χρήστη ή σε κάποιο άλλο ενδιαφερόμενο μέλος.

VII) ALTAI Ερωτηματολόγιο - Ανθρώπινη Εκπροσώπηση και Εποπτεία

Πρόκληση: Τα μοντέλα TN μπορεί να αντιμετωπίσουν νέες προκαταλήψεις ή ηθικές προκλήσεις καθώς λειτουργούν σε δυναμικά περιβάλλοντα.

Αντιμετώπιση: Η εγκαθίδρυση μηχανισμών για ζωντανή παρακολούθηση, βρόχους ανατροφοδότησης και προσαρμοστικότητα μοντέλων για την αντιμετώπιση τρέχοντων ηθικών προβληματισμών.

Ηθικά Πλαίσια και Οδηγίες

Πρόκληση: Η απουσία κανονικοποιημένων ηθικών πλαισίων μπορεί να οδηγήσει σε ασυνεπείς πρακτικές στην ανάπτυξη TN.

Αντιμετώπιση: Τηρώντας τις καθιερωμένες ηθικές κατευθυντήριες γραμμές, όπως αυτές που παρέχονται από οργανισμούς όπως το Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών – IHME ή Institute of Electrical and Electronics Engineers – IEEE, την Κοινωνία Μηχανικών Αυτοκινήτων – KMA ή Society of Automotive Engineers – SAE, κυβερνητικές ρυθμιστικές επιτροπές, την Ευρωπαϊκή Ένωση κ.λπ. και συμμετέχοντας ενεργά στην ανάπτυξη προτύπων σε ολόκληρη την βιομηχανία.

Η αντιμετώπιση ηθικών προβληματισμών και προκαταλήψεων στην ανάπτυξη λογισμικού με βάση την TN στα ΑΟ απαιτεί ολιστική και προληπτική προσέγγιση. Εμπεριέχει αφοσίωση στην δικαιοσύνη, την διαφάνεια, την ιδιωτικότητα χρήστη και την κοινωνική ευθύνη καθ' όλο τον κύκλο ζωής ενός συστήματος TN. Καθώς ο τομέας εξελίσσεται συνεχώς, συνεχόμενες προσπάθειες χρειάζονται να καταβάλλονται για να καθοριστούν ηθικές πρακτικές και να προωθηθεί η υπεύθυνη ανάπτυξη TN.

Πίνακας 8: Αυτόνομη Οδήγηση: Εξέλιξη Τεχνολογιών – Κλειδιών [31]

Τεχνολογία	Ανάπτυξη στο πέρασμα του χρόνου	Μέλλον
Περιβαλλοντολογική αντίληψη	Βαθιά Μάθηση για τον εντοπισμό αντικειμένων	Μεγάλη πρόκληση, χρειάζεται περισσότερη έρευνα για τον καλύτερο εντοπισμό αντικειμένων σε ομιχλώδη και δυσμενή καιρική συνθήκη σε πραγματικό χρόνο

	(YOLOv3 ²² , K-means clustering ²³)	
Εντοπισμός πεζών	Μοντέλα για εντοπισμό αντικειμένων σε κακές καιρικές συνθήκες (RCNN ²⁴)	Για την επίλυση του προβλήματος σύγκλισης εφαρμογή OrientNet, Predictor-Net ²⁵ , RPN ²⁶
Σχεδιασμός πορείας	Αλγόριθμοι Βαθιάς Μάθησης βασισμένοι στα Συνελικτικά Νευρωνικά Δίκτυα	Πολυαισθητηριακά συστήματα σύντηξης σε συνδυασμό με INS ²⁷ , GNSS ²⁸ και LiDAR συστήματα σε συνδυαστική χρήση για να εφαρμοστεί το σύστημα 3D-SLAM ⁴³
Κυβερνοασφάλεια οχήματος	Έλεγχος ασφάλειας και TARA ⁴⁴	Απομακρυσμένος έλεγχος των ΑΟ χρησιμοποιώντας IoT ²⁹ αισθητήρες
Σχεδιασμός κίνησης	Αλγόριθμος Κιου-Μάθησης, Κρυφό Μοντέλο Μαρκόβ - KMM	Μοντέλο πρόβλεψης Grey ⁴⁵ με χρήση προηγμένου προβλεπτικού μοντέλου ελέγχου αλλαγής λωρίδας

²² YOLOv3 – You Only Look Once, Version 3 είναι ένας αλγόριθμος ανίχνευσης αντικειμένων σε πραγματικό χρόνο ο οποίος ταυτοποιεί αντικείμενα σε βίντεο ή εικόνες. Χρησιμοποιεί Βαθιά Συνελικτικά Νευρωνικά Δίκτυα³⁶.

²³ K-means clustering – Κ-μέσων ομαδοποίηση είναι μια τεχνική μη-επιβλεπόμενης μάθησης που ταξινομεί δεδομένα με βάση χαρακτηριστικά και όχι προκαθορισμένες κατηγορίες³⁷.

²⁴ RCNN – Region-based Convolutional Neural Network αποτελεί ένα Συνελικτικό Νευρωνικό Δίκτυο που χρησιμοποιεί προτάσεις με βάση την περιοχή προκειμένου να εντοπίσει αντικείμενα σε μια εικόνα³⁸.

²⁵ OrientNet – Predictor-Net είναι προβλεπτικές αναλύσεις οι οποίες χρησιμοποιούν ιστορικά δεδομένα για να προβλέψουν μελλοντικά συμβάντα. Υλοποιούνται με στατιστικές τεχνικές συμπεριλαμβανομένης της εξόρυξης δεδομένων, της Μηχανικής Μάθησης και της προβλεπτικής μοντελοποίησης³⁹.

²⁶ RPN – Risk Priority Number είναι ο αριθμός προτεραιότητας αποτυχίας του συστήματος με βάση την επικινδυνότητα⁴⁰.

²⁷ INS – Inertial Navigation System είναι μια αυτόνομη τεχνική πλοήγησης στην οποία οι μετρήσεις που παρέχονται από επιταχυντόμετρα και γυροσκόπια χρησιμοποιούνται για την παρακολούθηση της θέσης και του προσανατολισμού ενός αντικείμενου σε σχέση με ένα γνωστό σημείο εκκίνησης, προσανατολισμού και ταχύτητας⁴¹.

²⁸ GNSS – Global Navigation Satellite Systems είναι συστήματα τα οποία χρησιμοποιούνται για την τοποθεσία και την πλοήγηση Αυτόνομων Οχημάτων⁴².

²⁹ IoT – Internet of Things λέγεται ένα δίκτυο φυσικών συσκευών, οχημάτων ή άλλων φυσικών αντικειμένων στα οποία υφίστανται ενσωματωμένοι αισθητήρες, λογισμικό και συνδεσιμότητα δικτύου επιτρέποντας τα να συλλέξουν και να διαμοιραστούν δεδομένα.

Η εμφάνιση των ΑΟ προσφέρει γόνιμο έδαφος για συζητήσεις σχετικά με θέματα ηθικότητας τα τελευταία χρόνια. Τέτοιες συζητήσεις καταγράφονται σε μεγάλο βαθμό στην επιστημονική βιβλιογραφία και περιστρέφονται κυρίως γύρω από ακραίες καταστάσεις κυκλοφορίας που απεικονίζονται ως ηθικά διλήμματα, δηλαδή καταστάσεις στις οποίες το ΑΟ καλείται να λάβει μια δύσκολη ηθικά επιλογή. Παραδόξως ελάχιστα ζητήματα είναι γνωστά τα οποία βρίσκονται στο επίκεντρο της βιομηχανίας ΑΟ. Έχουν διατυπωθεί γενικοί ισχυρισμοί οι οποίοι στερούνται κατάλληλη τεκμηρίωση. Καθώς οι ιδιωτικές εταιρείες έχουν μεγάλη επιρροή στην ανάπτυξη και την αποδοχή των τεχνολογιών ΑΟ, μια ουσιώδης συζήτηση σχετικά με θέματα δεοντολογίας των ΑΟ, πρέπει να λαμβάνει υπόψιν τα ηθικά θέματα που θέτει ως προτεραιότητα η βιομηχανία. Προκειμένου να εκτιμηθεί η ευαισθητοποίηση και η δέσμευση της βιομηχανίας σχετικά με την δεοντολογία των ΑΟ, έχουν επιθεωρηθεί οι αφηγήσεις από τις επίσημες επιχειρηματικές και τεχνικές εκθέσεις των εταιρειών με άδεια δοκιμών.

Τα αυτόνομα οχήματα είναι εξοπλισμένα με Προηγμένα Συστήματα Υποβοήθησης Οδηγού – ΠΣΥΟ. Τα συστήματα ΤΝ που σχετίζονται με τον Έλεγχο και την Δημιουργία Πορείας αποτελούν συστήματα τα οποία χαρακτηρίζονται από υψηλή αυτονομία – επίπεδο 4 έως πλήρη αυτονομία – επίπεδο 5. Πρακτικά, ο βαθμός αυτονομίας είναι τέτοιος που σχεδόν σε όλες τις περιπτώσεις το όχημα μπορεί να διαχειριστεί όλες τις οδηγικές απαιτήσεις χωρίς καμία ανθρώπινη παρέμβαση. Τα ΑΟ με τέτοια επίπεδα αυτονομίας δυνητικά μπορούν να θέσουν σε κίνδυνο την ζωή και την υγεία των πολιτών. Συμπεριλαμβάνονται δηλαδή στις μεταφορές οι οποίες αποτελούν υποδομές κριτικής σημασίας.

Συμπερασματικά, σύμφωνα με το ΑΙ Νομικό Πλαίσιο κατηγοριοποιούνται σε συστήματα ΤΝ υψηλού ρίσκου και είναι αντικείμενα σε αυστηρές επιταγές προτού γίνουν δημοσίως διαθέσιμα στις αγορές. Σε τέτοιες περιπτώσεις το ΑΙ Νομικό Πλαίσιο επιτάσσει επαρκή συστήματα αξιολόγησης και μετριάσμου κινδύνου, υψηλής ποιότητας συνόλων δεδομένων που τροφοδοτούν το σύστημα με στόχο την ελαχιστοποίηση των κινδύνων και των αποτελεσμάτων που προκαλούν διακρίσεις. Επίσης, κρίνεται απαραίτητη η καταγραφή των δραστηριοτήτων προκειμένου να διασφαλίζεται η ιχνηλασιμότητα των αποτελεσμάτων, η λεπτομερής καταγραφή σε έγγραφα που παρέχει όλες τις απαραίτητες πληροφορίες για το σύστημα και το σκοπό του έτσι ώστε να μπορεί να αξιολογηθεί από

τις αρχές ο βαθμός συμμόρφωσής του. Χρειάζονται επίσης, τα κατάλληλα μέτρα ανθρώπινης εποπτείας για την ελαχιστοποίηση του κινδύνου και υψηλά επίπεδα αξιοπιστίας, ασφάλειας και ακρίβειας.

3.3.3 Βιομηχανική και Βιβλιογραφική Ανασκόπηση Αυτοκινητοβιομηχανίας

Τα ευρήματα της βιβλιογραφικής και βιομηχανικής ανασκόπησης υποδηλώνουν τα εξής:

I) Δεδομένης της πληθώρας ηθικών ζητημάτων που εξετάζονται στις εκθέσεις, οι εταιρείες αυτόνομης οδήγησης έχουν επίγνωση και ασχολούνται με την ηθική της ΑΟ.

II) Η επιστημονική βιβλιογραφία και οι εκθέσεις του κλάδου δίνουν προτεραιότητα στην ασφάλεια.

III) Οι επιστημονικές και βιομηχανικές κοινότητες συμφωνούν ότι τα ΑΟ δεν εξαλείφουν τον κίνδυνο των ατυχημάτων.

IV) Η επιστημονική βιβλιογραφία σχετικά με την δεοντολογία των ΑΟ κυριαρχείται από συζητήσεις σχετικά με το «πρόβλημα του τρόλεϊ».

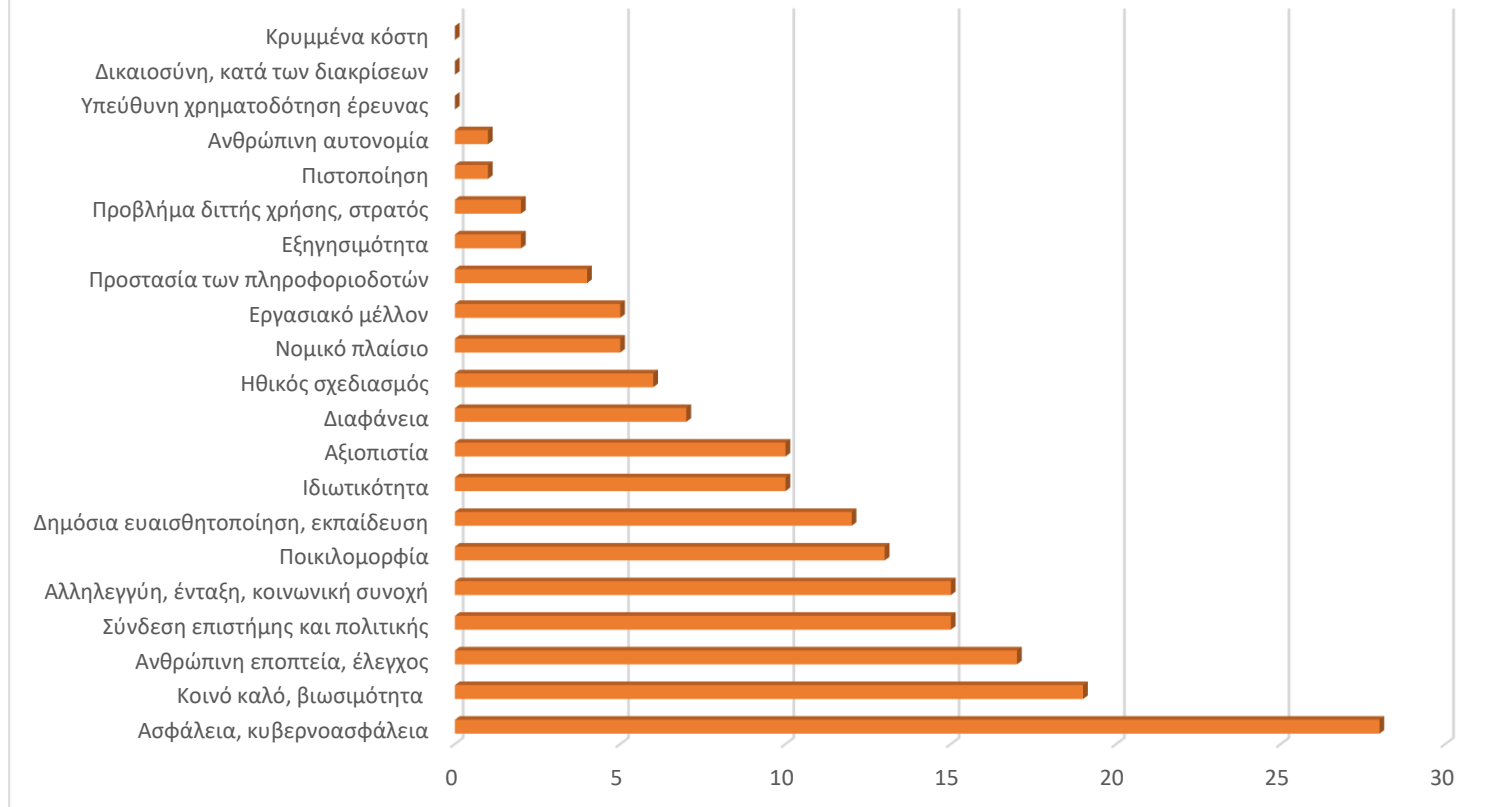
V) Ηθικά διλήμματα που ομοιάζουν με τις περιπτώσεις του «προβλήματος του τρόλεϊ» δεν εξετάζονται στις εκθέσεις της βιομηχανίας αλλά υπάρχουν αναφορές που ξετυλίγουν τις υποκείμενες ανησυχίες σχετικά με αυτές τις ακραίες κυκλοφοριακές καταστάσεις.

VI) Οι εταιρείες ΑΟ έχουν διαφορετικές προσεγγίσεις όσον αφορά την εξουσία των απομακρυσμένων χειριστών.

VII) Οι εταιρείες φαίνεται να επενδύουν σε σχεδιασμό με χαμηλότερο κίνδυνο ευθύνης, στρατηγικές που βασίζονται σε κανόνες και κανονισμούς, επιτάχυνση των ερευνών και αλγορίθμους αποφυγής σύγκρουσης.

Στην παρούσα μελέτη, έχουν χρησιμοποιηθεί 86 έγγραφα από 29 εταιρείες στο μέτωπο της τεχνολογίας ΑΟ. Συνολικά, οι αναφορές που ανασκοπήθηκαν δείχνουν σημαντικό θετικό τόνο για την τεχνολογία των ΑΟ, με αποτέλεσμα να είναι αναγκαίο να έρθουν στο φως της δημοσιότητας για επενδυτές, καταναλωτές και υπηρεσίες διακυβέρνησης. Ακολουθεί γράφημα 21 ηθικών θεμάτων υπό εταιρική διαχείριση στις αναφορές τους.

Εταιρείες που διαχειρίζονται θέματα ηθικότητας



Εικόνα 4: Ηθικά ζητήματα στην βιομηχανία ΑΟ^[46]

Είναι εμφανές πως η Ασφάλεια και η κυβερνοασφάλεια είναι το θέμα το οποίο οι περισσότερες εταιρείες αντιμετωπίζουν, ακολουθούμενο από το κοινό καλό, την βιωσιμότητα, την ανθρώπινη εποπτεία, τον έλεγχο και την σύνδεση επιστήμης και πολιτικής.

ALTAI Ερωτηματολόγιο - Ασφάλεια και Εμπιστοσύνη

Η ασφάλεια και η προστασία σχετίζονται με την ακεραιότητα των συστημάτων. Η ασφάλεια αφορά την επαρκή λειτουργία ενός συστήματος και η προστασία αφορά την ικανότητα ενός συστήματος να αντιστέκεται σε σκόπιμα κακόβουλες ενέργειες. Υπάρχουν

αδιευκρίνιστες εκτιμήσεις σχετικά με τα αποδεκτά επίπεδα ασφάλειας και κυβερνοασφάλειας των ΑΟ τόσο σε καθημερινές όσο και σε ακραίες καταστάσεις προκειμένου να διασφαλιστεί η ευημερία των χρηστών και των άλλων παραγόντων κυκλοφορίας.

Λαμβάνοντας υπόψιν ότι η εμπορική επιτυχία της τεχνολογίας ΑΟ εξαρτάται σε μεγάλο βαθμό από την εμπιστοσύνη των καταναλωτών, δεν αποτελεί έκπληξη το γεγονός ότι οι αφηγήσεις της βιομηχανίας επικεντρώνονται κυρίως σε θέματα ασφάλειας. Πριν γίνουν ευρέως αποδεκτά τα ΑΟ χωρίς οδηγό, οι άνθρωποι πρέπει να είναι πρόθυμοι να εμπιστευτούν τις ζωές τους και τις ζωές εκείνων που τους ενδιαφέρουν, επομένως τα ΑΟ πρέπει να συμπεριφέρονται, να αντιδρούν και να επικοινωνούν με τρόπους που να διευκολύνουν τους ανθρώπους να τα εμπιστευτούν – όχι μόνο τους επιβάτες στο εσωτερικό τους αλλά και τους πεζούς και τους οδηγούς που συναντούν στον δρόμο.

Ως αποτέλεσμα της ανασκόπησης των αναφορών των ΑΟ, συμπεραίνεται πως τα ηθικά διλήμματα που μοιάζουν με το «πρόβλημα του τρόλεϊ» δεν αντιμετωπίζονται σε αυτές τις εκθέσεις με τους όρους που περιγράφονται σε επιστημονικό υλικό και στις δημοσιεύσεις στα μέσα ενημέρωσης, αλλά υπάρχουν υπαινιγμοί που ξετυλίγουν υποκείμενες ανησυχίες σχετικά με ακραίες καταστάσεις κυκλοφορίας. Όσον αφορά τις λύσεις που προτείνονται από την βιομηχανία, οι εταιρείες προς το παρόν βασίζονται σε ραντάρ και σε περιορισμό της ταχύτητας για την αντιμετώπιση των τυφλών σημείων που συχνά αποτελούν την ρίζα ατυχημάτων.

ALTAI Ερωτηματολόγιο - Ανθρώπινη Εποπτεία

Η ανθρώπινη εποπτεία και ο έλεγχος των ΑΟ προϋποθέτουν την επιτήρηση της ανάπτυξης της τεχνολογίας. Αναμένεται ότι η εξ αποστάσεως εποπτεία των επιδόσεων της αυτόνομης οδήγησης διασφαλίζει την εμπιστοσύνη και την ασφάλεια σε αυτή την τεχνολογία, καθώς οι άνθρωποι χειριστές είναι σε θέση να αναλάβουν τον έλεγχο του οχήματος. Ωστόσο έχει αναφερθεί ότι αρκετά τεχνικά προβλήματα λαμβάνουν χώρα κατά την διάρκεια της μεταφοράς του ελέγχου επί του ΑΟ.

Οι ειδικοί αποστολών εκπαιδεύονται στον επιχειρησιακό σχεδιασμό που διέπει τα ΑΟ και είναι προετοιμασμένοι να αναλάβουν τον χειροκίνητο έλεγχο του οχήματος όταν τους παρουσιάζεται ένα σενάριο που δεν περιλαμβάνεται στον τομέα επιχειρησιακού

σχεδιασμού. Τα κέντρα υποστήριξης απομακρυσμένων επιχειρήσεων διαθέτουν χειριστές που μπορούν να καθοδηγούν τα οχήματα εξ αποστάσεως ανά πάσα στιγμή, ημέρα και νύχτα όταν ένα όχημα αντιμετωπίζει μια αβέβαιη οδηγική κατάσταση, όπως διακοπή φωτεινής σηματοδότησης ή εμπόδιο στο δρόμο.

Οι δηλώσεις που εντοπίστηκαν στα έγγραφα σχετικά με την ανθρώπινη εποπτεία αφορούν περισσότερο την πρώτη προϋπόθεση της θεωρίας του Ανθρώπινου Ουσιαστικού Ελέγχου – ΑΟΕ ή Human Meaningful Control – HMC³⁰ και με την συνθήκη παρακολούθησης σε σχέση με την κατανόηση του συστήματος. Οι χειριστές δοκιμών αντιμετωπίζουν τις δικές τους μοναδικές προκλήσεις. Η αποσφαλμάτωση ενός πολύπλοκου αυτόνομου συστήματος είναι ακατανόητη για το ανεκπαιδευτο μάτι. Οι μηχανικοί και οι προγραμματιστές έχουν βαθιά κατανόηση του κώδικα που λειτουργεί στα πρωτότυπα, επιτρέποντας τους να προβλέψουν πότε το σύστημα αποτυγχάνει. Αυτό επιτρέπει στα συνεργεία δοκιμών να ανακτήσουν τον έλεγχο του οχήματος προληπτικά, με ελεγχόμενο τρόπο.

ALTAI Ερωτηματολόγιο - Αξιοπιστία

Τα ζητήματα λογοδοσίας που σχετίζονται με την τεχνολογία ΑΟ έχουν λάβει σημαντική προσοχή στην επιστημονική βιβλιογραφία. Γίνεται αναφορά στην λογοδοσία με ευρείς όρους, καλύπτοντας έτσι συναφείς έννοιες, όπως η ευθύνη και η υπευθυνότητα. Διευκρινίζεται ότι η λογοδοσία συνεπάγεται ευθύνη, αλλά σε αντίθεση με την τελευταία απαιτεί εξηγήσεις σχετικά με τις πράξεις και ό,τι δεν μπορεί να διαμοιραστεί. Η ευθύνη για μια πράξη παραδοσιακά απαιτεί τουλάχιστον μια συνθήκη ελέγχου, δηλαδή ένας πράκτορας είναι υπεύθυνος εάν είναι ο φορέας της πράξης και μια συνθήκη, δηλαδή επίγνωση ή γνώση του πράκτορα σχετικά με την πράξη – και η ευθύνη είναι νομική ή οικονομική. Αυτά τα θέματα αποτελούν πρόκληση στον τομέα των ΑΟ κυρίως λόγω του κατακερματισμού της τεχνολογικής δράσης, η οποία μπορεί να οδηγήσει σε κενά ευθύνης.

³⁰ Human Meaningful Control theory – HMC: Αποτελεί μια πολιτικά βεβαρυμένη ιδέα η οποία προήλθε από συζήτηση για τα αυτόνομα όπλα. Γενικά, αναφέρεται στην διατήρηση της ανθρώπινης κρίσης και την είσοδο της σε οποιοδήποτε σύστημα αυτόνομης φύσης^[47].

Οι εταιρείες έχουν δηλώσει την ανάγκη να τεθούν εκ των προτέρων σαφείς κανόνες και νομικά πλαίσια προκειμένου να αποσαφηνιστεί πού βρίσκεται η ευθύνη σε περίπτωση ατυχήματος μετά την υλοποίηση της πλήρους αυτοματοποιημένης οδήγησης. Παράλληλα παρουσιάζονται τεχνικές λύσεις που αποσκοπούν στην επιτάχυνση της έρευνας, όπως η χρήση ενός «μαύρου-κουτιού» που μοιάζει με ένα καταγραφικό πτήσης, σχεδιασμένο να αποθηκεύει δεδομένα ατυχήματος ή αλγόριθμοι ευθύνης που βασίζονται σε μαθηματικά μοντέλα.

Καθώς η τεχνολογία αυτόνομης οδήγησης αναπτύσσεται και η χρήση της διαδίδεται, η ευθύνη των κατασκευαστών αυτοκινήτων μπορεί να τίθεται υπό αμφισβήτηση σε σχέση με την μείωση των οδηγών που ασχολούνται με την οδήγηση.

Επίσης, καθώς οι ανακλήσεις οχημάτων που έχουν εφαρμοστεί προς όφελος των πελατών αυξηθούν σε πλήθος, πρόσθετα έξοδα επιβαρύνουν τις εταιρείες και υπόκεινται δυνητικές ζημιές στην εικόνα της μάρκας τους^[46].

Οι εφαρμογές ΤΝ στην αυτοκινητοβιομηχανία που κυμαίνονται από το επίπεδο 0 έως το επίπεδο 3 αποτελούν συστήματα περιορισμένου ρίσκου σύμφωνα με την ιεραρχία της πυραμίδας του ΑΙ Νομικού Πλαισίου καθώς διατρέχεται ο κίνδυνος της ελλιπούς διαφάνειας. Το ΑΙ Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας ώστε οι άνθρωποι να είναι ενημερωμένοι όταν κρίνεται απαραίτητο ενισχύοντας έτσι την εμπιστοσύνη. Οι πάροχοι των συστημάτων ΤΝ πρέπει να διασφαλίζουν ότι οι αποφάσεις που παράγονται ταυτοποιούνται.

Όταν τα επίπεδα αυτονομίας στα ΑΟ κυμαίνονται στα επίπεδα 4 και 5 τότε πρόκειται για συστήματα ΤΝ υψηλού ρίσκου καθώς αποτελούν μέρη κριτικής υποδομής – μεταφοράς τα οποία δυνητικά μπορούν να θέσουν την ζωή και την υγεία των πολιτών σε κίνδυνο. Σύμφωνα με το ΑΙ Νομικό Πλαίσιο τέτοια συστήματα είναι αντικείμενα αυστηρών επιταγών πριν γίνουν διαθέσιμα στην αγορά. Πιο συγκεκριμένα απαιτούνται επαρκείς έλεγχοι ρίσκου και συστήματα αντιμετώπισης, υψηλό επίπεδο συνόλων δεδομένων τα οποία τροφοδοτούν τα συστήματα ΤΝ ώστε να ελαχιστοποιηθούν τα ρίσκα και τα κοινωνικώς διαχωριστικά αποτελέσματα. Επίσης, χρειάζεται καταγραφή των δραστηριοτήτων για να διασφαλιστεί η ανιχνευσιμότητα των αποτελεσμάτων, η λεπτομερής τεκμηρίωση που παρέχει όλες τις απαραίτητες πληροφορίες για το σύστημα

και το σκοπό του έτσι ώστε οι αρχές να αξιολογούν την συμμόρφωσή του. Καθαρές και επαρκείς πληροφορίες στον εφαρμοστή, κατάλληλη ανθρώπινη εποπτεία για να ελαχιστοποιείται ο κίνδυνος και υψηλά επίπεδα ευρωστίας, ασφάλειας και ακρίβειας.

Κεφάλαιο 4: Σύνθεση Μεθοδολογίας για την Ανάπτυξη Ηθικών Λύσεων Τεχνητής Νοημοσύνης

4.1 Μεθοδολογικό Πλαίσιο Αξιολόγησης και Ανάπτυξης Ηθικών Λύσεων Τεχνητής Νοημοσύνης

Η μεθοδολογία αξιολόγησης των αναφερθέντων συστημάτων ΤΝ απαρτίζεται από τρία επίπεδα θεωρητικού και τεχνικού φιλτραρίσματος και κατηγοριοποίησης.

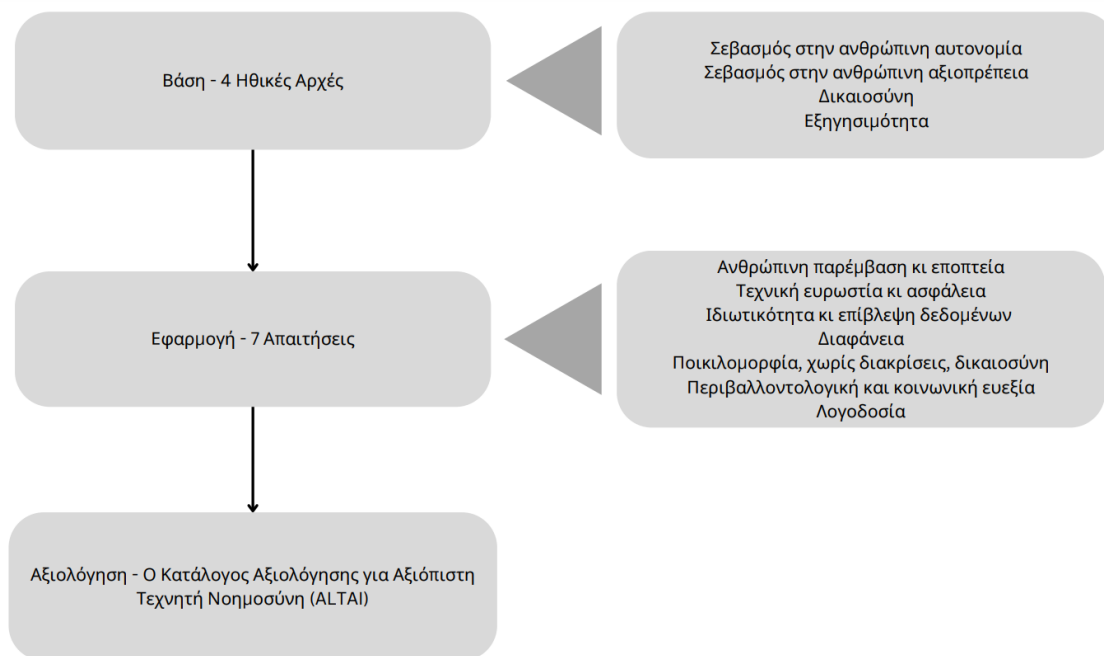
- ΑΙ Νομικό Πλαίσιο και εκτίμηση ρίσκου του εκάστοτε συστήματος ΤΝ.
- Συμπλήρωση του ψηφιακού του εργαλείου The Assessment List for Trustworthy Artificial Intelligence^[48], με βάση τα εκάστοτε τεχνικά χαρακτηριστικά του συστήματος ΤΝ, έλεγχος των προαπαιτήσεων και κατηγοριοποίησή του ως προς την ηθικότητα με βάση τα αποτελέσματα τα οποία παράγονται.
- Προτάσεις για βελτίωση της ηθικότητας του συστήματος, αξιολόγηση των προτάσεων και κριτική.

Είναι σημαντικό να αναφερθεί πως κάθε εταιρεία η οποία αναπτύσσει λογισμικό ΤΝ χρειάζεται να διαβάζει τα νομικά πλαίσια και την πιθανή βαθμίδα ρίσκου εκ προοιμίου. Είναι απαραίτητο να λαμβάνονται υπόψιν τα προαπαιτούμενα και ποια είναι η δομή του από την αρχή μέχρι το τέλος. Επίσης, αποτελεί κρίσιμη αρχή σχεδίασης το να χρησιμοποιούνται ψηφιακά εργαλεία όπως το The Assessment List for Trustworthy Artificial Intelligence προκειμένου να ελέγχεται η τήρηση της ηθικότητας του συστήματος ΤΝ. Το βιομηχανικό λογισμικό κατηγοριοποιείται συνήθως ως λογισμικό μηδαμινού έως περιορισμένου ρίσκου καθώς το εταιρικό πλαίσιο είναι προστατευμένο.

Ο Κατάλογος Αξιολόγησης για Αξιόπιστη Τεχνητή Νοημοσύνη – ΚΑΑΤΝ ή The Assessment List for Trustworthy Artificial Intelligence - ALTAI

Το ALTAI Ερωτηματολόγιο θέτει την δεοντολογία στο επίκεντρο της ανάπτυξης συστημάτων ΤΝ. Λειτουργεί σαν ένα εργαλείο αυτο-αξιολόγησης για την εκτίμηση της ηθικότητας ενός συστήματος ΤΝ υπό το πρίσμα των βασικών απαιτήσεων που ορίζονται στις κατευθυντήριες γραμμές. Ο κατάλογος περιέχει ένα σύνολο ερωτήσεων σχετικών με

τις απαιτήσεις που παρέχουν καθοδήγηση για την πρακτική εφαρμογή τους. Επιπλέον, ο κατάλογος αυτός ευαισθητοποιεί το κοινό για τις πιθανές επιπτώσεις και τους κινδύνους προτεινόμενων συστημάτων ΤΝ και το είδος των μέτρων που μπορούν να ληφθούν για τον μετριασμό αυτών των κινδύνων. Ο κατάλογος αξιολόγησης αξιόπιστης ΤΝ επιτρέπει την επικύρωση του τελικού συστήματος ΤΝ, ολοκληρώνοντας έτσι τον κύκλο ζωής ανάπτυξης αξιόπιστης ΤΝ όπως καθορίστηκε από την Ομάδα Εμπειρογνομώνων Υψηλού Επιπέδου – ΟΕΥΕ.



Εικόνα 5: Το πλαίσιο αξιόπιστης Τεχνητής Νοημοσύνης όπως καθορίζεται από την ΟΕΥΕ

Σύμφωνα με το ALTAI Κέντρο Πληροφόρησης, η αξιόπιστη ΤΝ ορίζεται από τρεις συμπληρωματικές έννοιες, την Νόμιμη ΤΝ, την Ηθική ΤΝ και την Εύρωστη ΤΝ. Οι κατευθυντήριες γραμμές έχουν ανθρωποκεντρική προσέγγιση ως προς την ΤΝ και προσδιορίζουν 4 ηθικές αρχές και 7 απαιτήσεις που πρέπει να ακολουθούν οι εταιρείες για να επιτύχουν αξιόπιστη ΤΝ. Το ερωτηματολόγιο συμπληρώνεται με ένα σύνολο ερωτήσεων για κάθε μία από τις 7 απαιτήσεις, οι οποίες έχουν ως στόχο την λειτουργικοποίηση των απαιτήσεων. Οι ερωτήσεις αυτές είναι χρωματικά κωδικοποιημένες όπως εξηγείται παρακάτω:

1. Οι ερωτήσεις με λευκό φόντο αποσκοπούν στην περιγραφή των χαρακτηριστικών του συστήματος TN.
2. Οι απαντήσεις στις μπλε ερωτήσεις συμβάλλουν στις συστάσεις.
3. Το κείμενο με κόκκινο χρώμα επιτρέπει την αυτοαξιολόγηση της συμμόρφωσης ενός οργανισμού με τις αντίστοιχες απαιτήσεις.

4.2 Εφαρμογή Μεθοδολογίας

Κατά την συμπλήρωση των ερωτηματολογίων για την εκάστοτε περίπτωση χρήσης έχουν συμπληρωθεί με την εξής λογική:

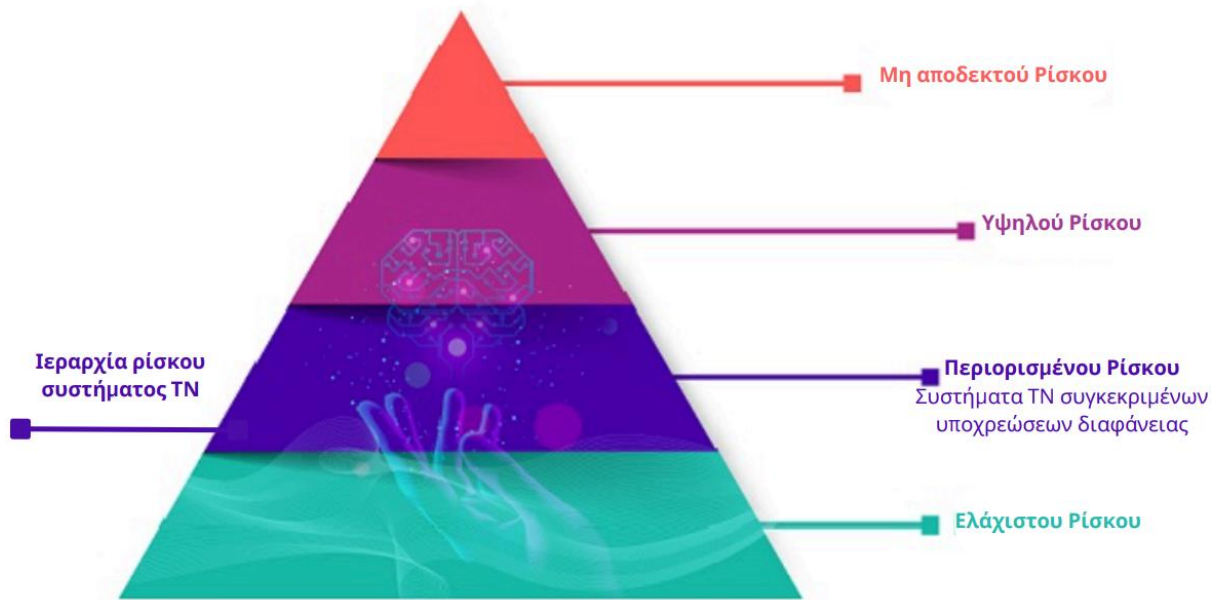
- Οι ερωτήσεις στις οποίες τα τεχνικά χαρακτηριστικά των συστημάτων TN είναι σαφή στην βιβλιογραφία συμπληρώθηκαν αυτούσια.
- Οι ερωτήσεις στις οποίες τα τεχνικά χαρακτηριστικά του εκάστοτε συστήματος TN είναι επαρκή αλλά δεν υπάρχει εφαρμογή, απαντήθηκαν με την επιλογή «μη-εφαρμόσιμη».
- Οι ερωτήσεις στις οποίες υπήρχαν ελλείψεις τεχνικών χαρακτηριστικών των συστημάτων στις ερευνητικές δημοσιεύσεις ανασκόπησης απαντήθηκαν είτε με βάση την ανάλυση χειρίστης περίπτωσης, είτε με την επιλογή «δεν είναι γνωστό», ανάλογα με την διαθέσιμη απάντηση στο ερωτηματολόγιο.

4.2.1 Περίπτωση Χρήσης στην Ναυτιλία

4.2.1.1 Εφαρμογή Μεθοδολογίας στους Ψηφιακούς Πλοηγούς - ΨΠ

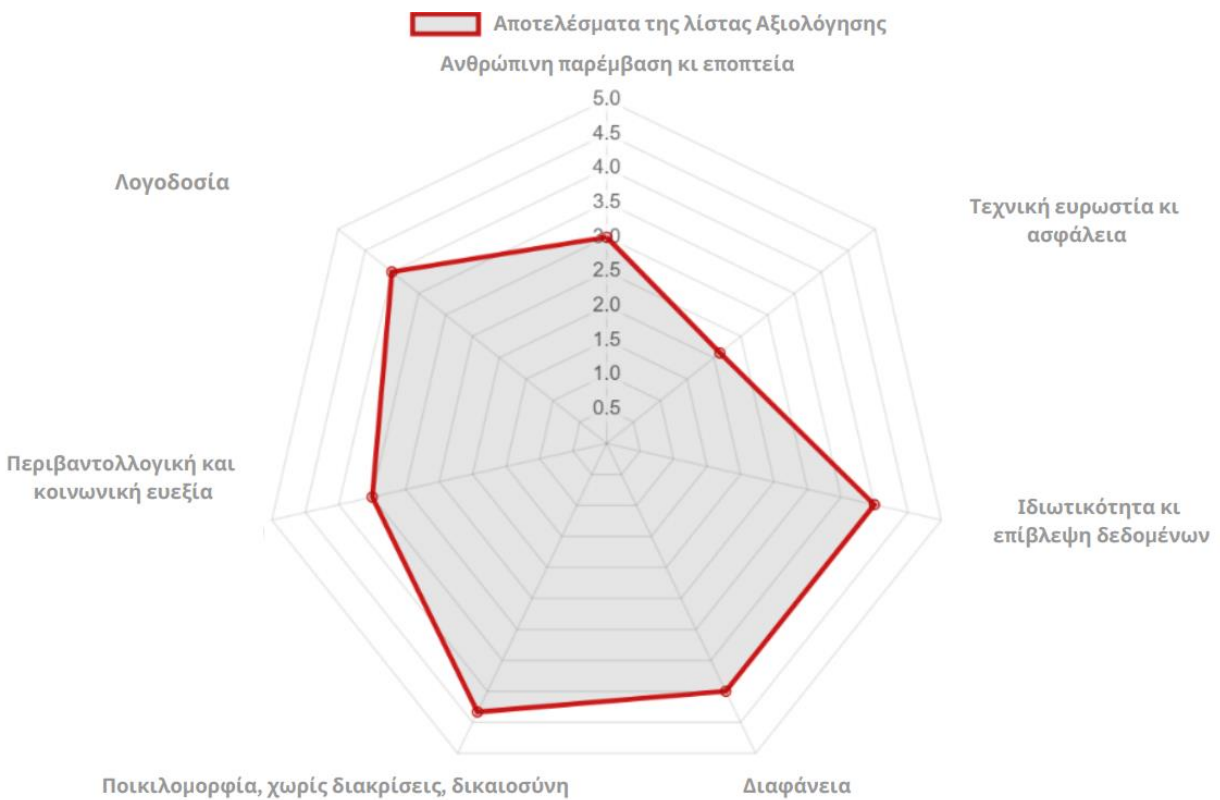
Ι) ΑΙ Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

Οι ψηφιακοί πλοηγοί^[16] κατατάσσονται σε συστήματα TN με περιορισμένο ρίσκο καθώς με συγκεκριμένες επιταγές διαφάνειας μπορούν να μην θέτουν ξεκάθαρο κίνδυνο στις ζωές των ανθρώπων, στην ασφάλεια και στα δικαιώματά τους εφόσον φυσικά τηρούνται οι απαιτήσεις αξιοπιστίας.



Εικόνα 6: Ιεραρχία ΑΙ Νομικού Πλαισίου: Ψηφιακοί Πλοηγοί – ΨΠ

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 7: ALTAI Αξιολόγηση: Ψηφιακών Πλοηγών - ΨΠ

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 9: ALTAI Προτάσεις για Ψηφιακούς Πλοηγούς - ΨΠ.

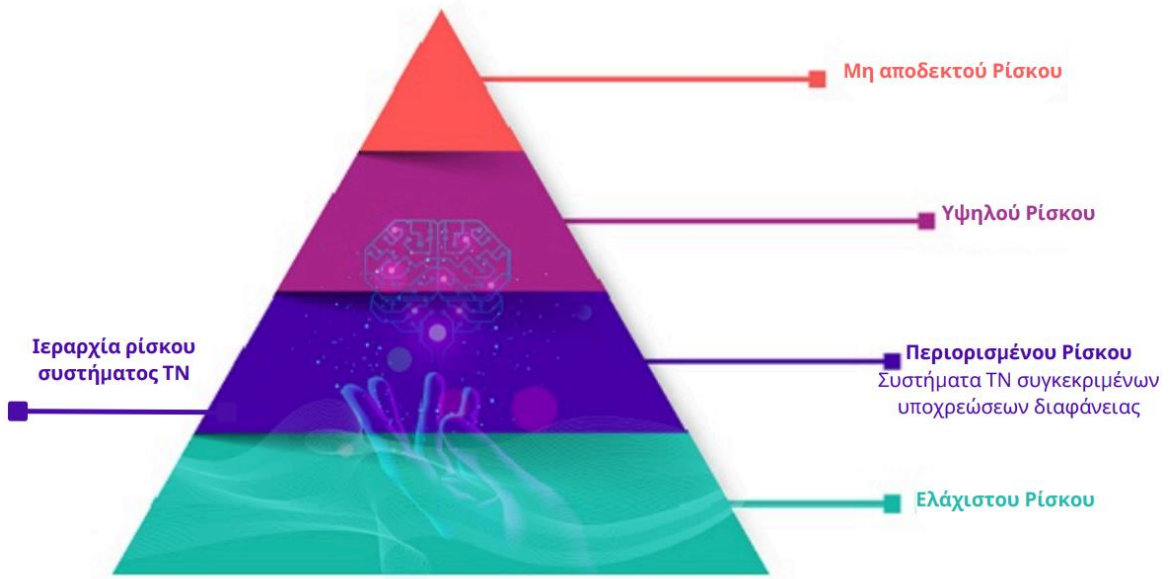
Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	Λήψη μέτρων για την αντιμετώπιση των πιθανών αρνητικών συνεπειών για τους τελικούς χρήστες ή τα υποκείμενα σε περίπτωση που αναπτύξουν προσκόλληση. Παροχή μέτρων ώστε ο χρήστης να έχει τον έλεγχο των αλληλεπιδράσεων - <i>Οι ψηφιακοί πλοηγοί απευθύνονται σε ναυτικούς οι οποίοι διατρέχουν μικρότερο κίνδυνο προσκόλλησης λόγω εκπαίδευσης και εμπειρίας.</i>
Τεχνική ευρωστία και ασφάλεια	<p>Διασφάλιση ότι τα δεδομένα, συμπεριλαμβανομένων των δεδομένων εκπαίδευσης, που χρησιμοποιούνται για την ανάπτυξη του συστήματος TN είναι επικαιροποιημένα, υψηλής ποιότητας, πλήρη και αντιπροσωπευτικά του περιβάλλοντος στο οποίο θα αναπτυχθεί το σύστημα.</p> <p>Έλεγχος κατά πόσον η λειτουργία του συστήματος TN μπορεί να ακυρώσει τα δεδομένα ή τις υποθέσεις με βάση τις οποίες εκπαιδεύτηκε και πώς αυτό μπορεί να οδηγήσει σε αντιφατικά αποτελέσματα.</p>
Ιδιωτικότητα και επίβλεψη δεδομένων	Προστασία και κρυπτογράφηση των προσωπικών δεδομένων ναυτικών – τελικών χρηστών στα εν χρήσει σύνολα δεδομένων.
Διαφάνεια	<p>Παροχή κειμένων με τον τρόπο λειτουργίας και την διαδικασία λήψης απόφασης με στόχο την εκπαίδευση του ναυτικού – τελικού χρήστη.</p> <p>Επικοινωνία στους χρήστες ότι αλληλεπιδρούν με μια μηχανή - <i>Οι τελικοί χρήστες ήδη γνωρίζουν πως αλληλεπιδρούν με σύστημα TN.</i></p>
Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη	<p>Εφαρμογή διαδικασιών για τον έλεγχο και την παρακολούθηση πιθανών μεροληψιών καθ' όλη τη διάρκεια του κύκλου ζωής του συστήματος TN (π.χ. μεροληψίες λόγω πιθανών περιορισμών που απορρέουν από τη σύνθεση των εν χρήσει συνόλων δεδομένων, έλλειψη ποικιλομορφίας, μη αντιπροσωπευτικότητα).</p> <p>Πρωτοβουλίες εκπαίδευσης και ευαισθητοποίησης (εργαστήρια και εκπαίδευση, διαδικτυακά μαθήματα, μελέτες περιπτώσεων, διαλέξεις, συνεχής μάθηση) για να εξοπλιστούν οι σχεδιαστές και οι προγραμματιστές TN με μεγαλύτερη επίγνωση των πιθανών προκαταλήψεων που μπορεί να εισάγουν κατά το σχεδιασμό και την ανάπτυξη του συστήματος TN.</p>

Περιβαλλοντολογική και κοινωνική ευεξία	Μέτρα για τη μείωση των περιβαλλοντολογικών επιπτώσεων του κύκλου ζωής του συστήματος ΤΝ (μέθοδοι ανάπτυξης, εκπαίδευσης, μοντελοποίησης με μικρότερο περιβαλλοντολογικό αντίκτυπο) και συμμετοχή σε διαγωνισμούς για την ανάπτυξη λύσεων ΤΝ που αντιμετωπίζουν αυτό το πρόβλημα.
Λογοδοσία	<p>Ο έλεγχος και η καθοδήγηση από τρίτο μέρος μπορεί να βοηθήσει τόσο στην ποιοτική όσο και στην ποσοτική ανάλυση κινδύνου. Επιπλέον, μπορεί να συμβάλλει στη δημιουργία εμπιστοσύνης στην τεχνολογία και στο ίδιο το προϊόν.</p> <p>Ανάπτυξη με προληπτική προσέγγιση των κινδύνων και με τρόπο ώστε να συμπεριφέρονται αξιόπιστα όπως προβλέπεται, ελαχιστοποιώντας παράλληλα τις ακούσιες και απροσδόκητες βλάβες και αποτρέποντας τις μη αποδεκτές βλάβες. Κατά συνέπεια, οι προγραμματιστές και οι φορείς ανάπτυξης θα πρέπει να λαμβάνουν κατάλληλη εκπαίδευση σχετικά με το νομικό πλαίσιο που ισχύει για τα συστήματα που αναπτύσσονται.</p>

4.2.1.2 Εφαρμογή Μεθοδολογίας στην Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ

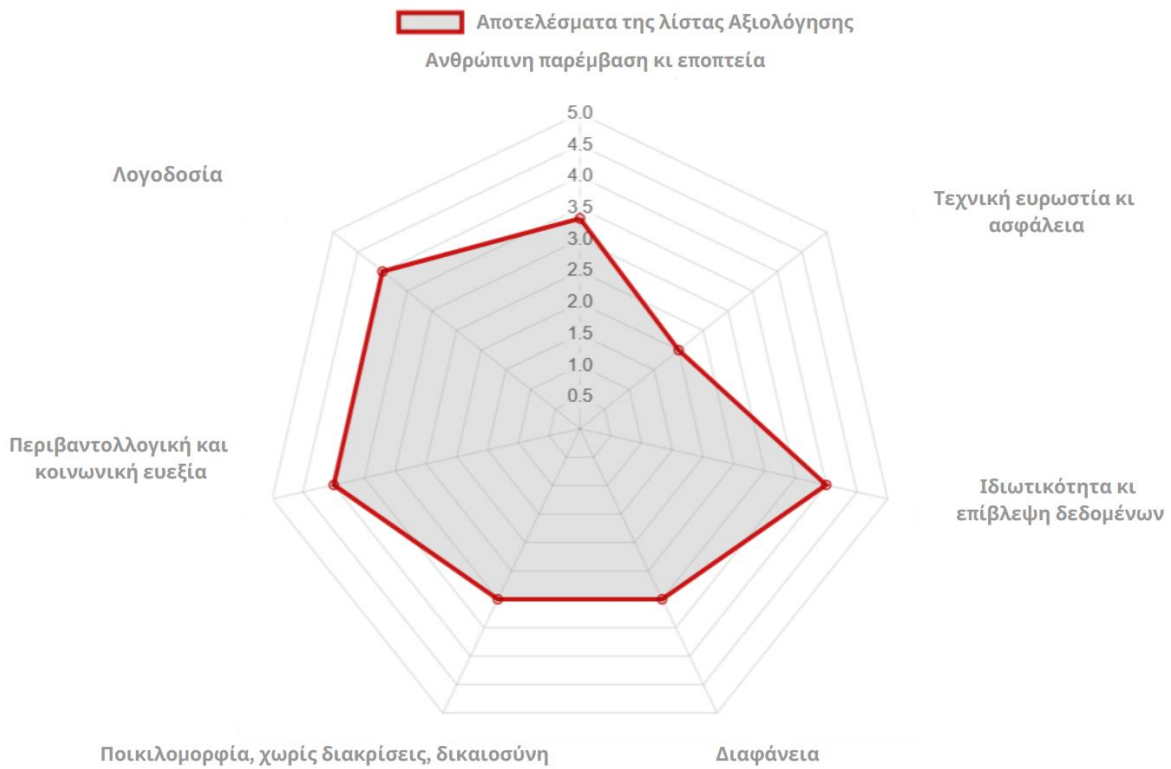
1) ΑΙ Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

Η Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ αποτελεί ένα έργο που είναι η βάση για συστήματα ΤΝ στην ναυτιλία, το οποίο κατατάσσεται στην κλάση περιορισμένου ρίσκου στην πυραμίδα του ΑΙ Νομικού Πλαισίου. Δεδομένου των θεμάτων διαφάνειας του έργου, το ΑΙ Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας οι οποίες διαβεβαιώνουν ότι οι άνθρωποι πληροφορούνται όταν είναι απαραίτητο, ενισχύοντας την εμπιστοσύνη του έργου και των δυνητικών αποκημάτων του^[19].



Εικόνα 8: Ιεραρχία AI Νομικού Πλαισίου: Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης – ΘΟΕΕ

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 9: ALTAI Αξιολόγηση Θαλάσσιας Ολοκληρωμένης Επίγνωσης της Επιτήρησης - ΘΟΕΕ

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 10: ALTAI Προτάσεις για Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ.

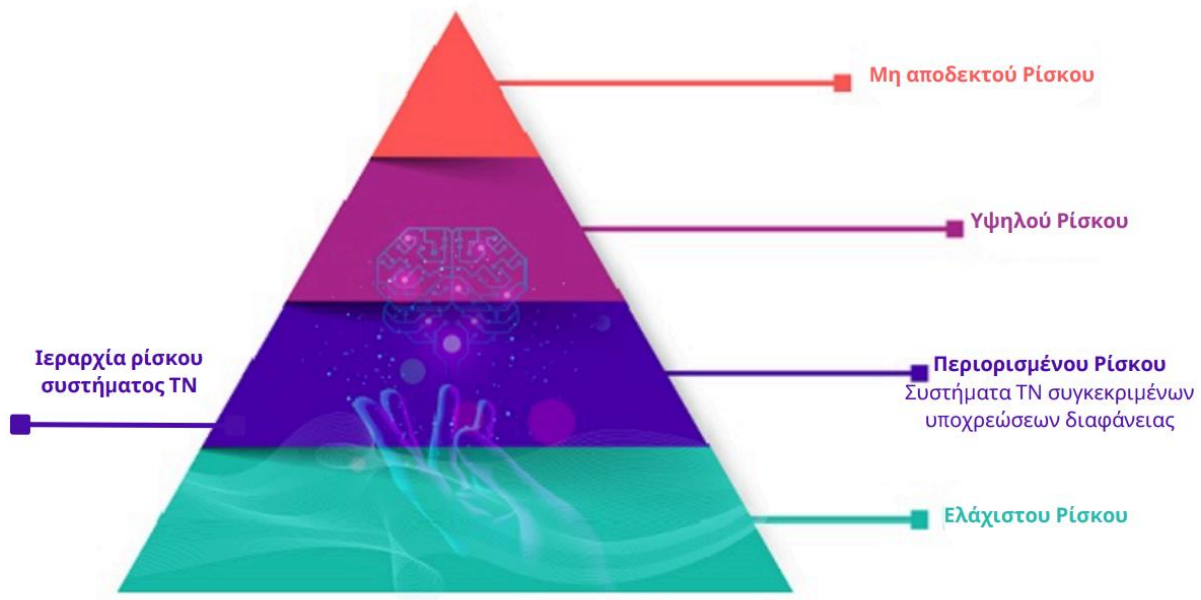
Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	Λήψη μέτρων για την αντιμετώπιση των πιθανών αρνητικών συνεπειών για τους τελικούς χρήστες ή τα υποκείμενα σε περίπτωση που αναπτύξουν προσκόλληση στα συστήματα TN που βασίζονται στην ΘΟΕΕ.
Τεχνική ευρωστία και ασφάλεια	<p>Μέτρα για τη διασφάλιση της ακεραιότητας, της ευρωστίας και της συνολικής ασφάλειας του συστήματος TN έναντι πιθανών επιθέσεων κατά τη διάρκεια του κύκλου ζωής του.</p> <p>Προσδιορισμός των πιθανών απειλών για το σύστημα TN (σφάλματα σχεδιασμού, τεχνικά σφάλματα, περιβαλλοντολογικές απειλές) και των πιθανών συνεπειών που προκύπτουν.</p> <p>Εκτίμηση του κινδύνου πιθανής κακόβουλης χρήσης, κακής χρήσης ή ακατάλληλης χρήσης του συστήματος TN.</p>
Ιδιωτικότητα και επίβλεψη δεδομένων	Διασφάλιση ότι τα δεδομένα, συμπεριλαμβανομένων των δεδομένων εκπαίδευσης, που χρησιμοποιούνται για την ανάπτυξη του συστήματος TN είναι επικαιροποιημένα, υψηλής ποιότητας, πλήρη και αντιπροσωπευτικά του περιβάλλοντος στο οποίο θα αναπτυχθεί το σύστημα.
Διαφάνεια	Δημιουργία και σύνταξη κειμένων με στόχο την ενημέρωση του τελικού χρήστη για τον τρόπο λειτουργίας και την διαδικασία λήψης απόφασης του συστήματος TN.
Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη	<p>Πρωτοβουλίες εκπαίδευσης και ευαισθητοποίησης για να εξοπλιστούν οι σχεδιαστές και οι προγραμματιστές TN με μεγαλύτερη επίγνωση της πιθανής προκατάληψης που μπορεί να εισάγουν κατά το σχεδιασμό και την ανάπτυξη του συστήματος TN.</p> <p>Διαβούλευση με τις κοινότητες που επηρεάζονται σχετικά με τον ορθό ορισμό της δικαιοσύνης.</p>
Περιβαλλοντολογική και κοινωνική ευεξία	Μέτρα για τη μείωση των περιβαλλοντολογικών επιπτώσεων του κύκλου ζωής του συστήματος TN και συμμετοχή σε διαγωνισμούς για την ανάπτυξη λύσεων TN που αντιμετωπίζουν αυτό το πρόβλημα.

	Μέτρα διασφάλισης ότι οι εργασιακές επιπτώσεις του συστήματος ΤΝ είναι καλά κατανοητές βάσει ανάλυσης των εργασιακών διαδικασιών και ολόκληρου του κοινωνικο-τεχνικού συστήματος.
Λογοδοσία	<p>Μια χρήσιμη μη τεχνική μέθοδος για να εξασφαλιστεί η εφαρμογή αξιόπιστης ΤΝ είναι η συμμετοχή διαφόρων ενδιαφερόμενων μερών, π.χ. συγκεντρωμένων σε μια "επιτροπή ηθικού ελέγχου" για την παρακολούθηση και την υποστήριξη της διαδικασίας ανάπτυξης.</p> <p>Η συμμετοχή τρίτων μερών για την αναφορά των τρωτών σημείων και των κινδύνων συμβάλλει στον εντοπισμό και τον μετριασμό των πιθανών παγίδων.</p> <p>Μια διαδικασία διαχείρισης κινδύνων πρέπει πάντα να περιλαμβάνει νέα ευρήματα, καθώς οι αρχικές υποθέσεις σχετικά με την πιθανότητα εμφάνισης ενός συγκεκριμένου κινδύνου μπορεί να είναι εσφαλμένες και συνεπώς, η ποσοτική ανάλυση κινδύνου δεν είναι σωστή και πρέπει να αναθεωρηθεί με τα νέα ευρήματα.</p>

4.2.1.3 Εφαρμογή Μεθοδολογίας στα Θαλάσσια Αυτόνομα Πλοία Επιφάνειας – ΘΑΠΕ

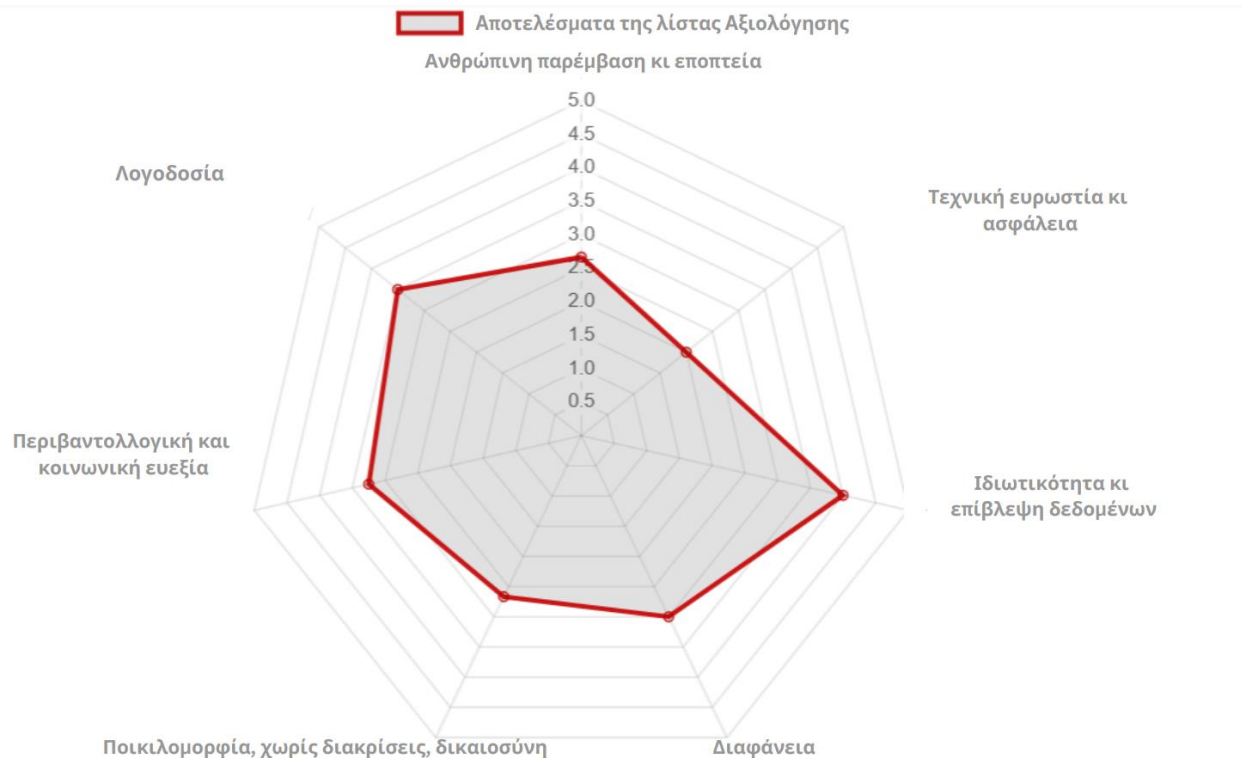
1) ΑΙ Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

Τα ΘΑΠΕ κατηγοριοποιούνται στα συστήματα περιορισμένου ρίσκου του ΑΙ Νομικού Πλαισίου καθώς με συγκεκριμένες επιταγές διαφάνειας ενισχύεται η εμπιστοσύνη και η λήψη απόφασης δεν γίνεται παρακινδυνευμένα^[20].



Εικόνα 10: Ιεραρχία ΑΙ Νομικού Πλαισίου: Θαλάσσια Αυτόνομα Πλοία Επιφάνειας – ΘΑΠΕ

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 11: ALTAI Αξιολόγηση: Θαλάσσια Αυτόνομα Πλοία Επιφάνειας - ΘΑΠΕ

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 11: ALTAI Προτάσεις για Θαλάσσια Αυτόνομα Πλοία Επιφάνειας - ΘΑΠΕ.

Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	<p>Αποφυγή προσκόλλησης και επιρροή της ανθρώπινης αυτονομίας.</p> <p>Πρόβλεψη στο σύστημα TN ώστε ο άνθρωπος - χειριστής να μπορεί να ανακτήσει τον έλεγχο σε περίπτωση λανθασμένης λειτουργίας του συστήματος.</p>
Τεχνική ευρωστία και ασφάλεια	<p>Προσδιορισμός των πιθανών απειλών για το σύστημα TN (σχεδιαστικά σφάλματα, τεχνικά σφάλματα, περιβαλλοντολογικές απειλές) και των πιθανών συνεπειών που προκύπτουν.</p> <p>Αξιολόγηση κινδύνου πιθανής κακόβουλης χρήσης, κατάχρησης ή ακατάλληλης χρήσης του συστήματος TN.</p>
Ιδιωτικότητα και επίβλεψη δεδομένων	<p>Κατά περίπτωση, εφαρμογή του δικαιώματος ανάκλησης της συγκατάθεσης, του δικαιώματος αντίρρησης και του δικαιώματος λήθης στο σύστημα TN.</p> <p>Όποτε είναι δυνατόν και σχετικό, ευθυγράμμιση του συστήματος TN με τα σχετικά πρότυπα (π.χ. Ινστιτούτο Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών - ΙΗΗΜ) ή τα ευρέως υιοθετημένα πρωτόκολλα για την καθημερινή διαχείριση και διακυβέρνηση δεδομένων.</p>
Διαφάνεια	<p>Ενημέρωση των χρηστών πως αλληλεπιδρούν με ένα σύστημα TN.</p>
Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη	<p>Έρευνα και χρήση δημοσίων διαθέσιμων τεχνικών εργαλείων, τα οποία είναι τελευταίας τεχνολογίας, για καλύτερη κατανόηση των δεδομένων, των μοντέλων και των επιδόσεων.</p> <p>Διαδικασίες για τον έλεγχο και την παρακολούθηση πιθανών προκαταλήψεων κατά τη διάρκεια ολόκληρου του κύκλου ζωής του συστήματος TN (π.χ. προκαταλήψεις λόγω πιθανών περιορισμών που απορρέουν από τη σύνθεση των εν χρήσει συνόλων δεδομένων, έλλειψη ποικιλομορφίας, μη αντιπροσωπευτικότητα).</p>
Περιβαλλοντολογική και κοινωνική ευεξία	<p>Πιθανές θετικές και αρνητικές επιπτώσεις του συστήματος TN στο περιβάλλον και μηχανισμοί για την αξιολόγηση των επιπτώσεων αυτών.</p>

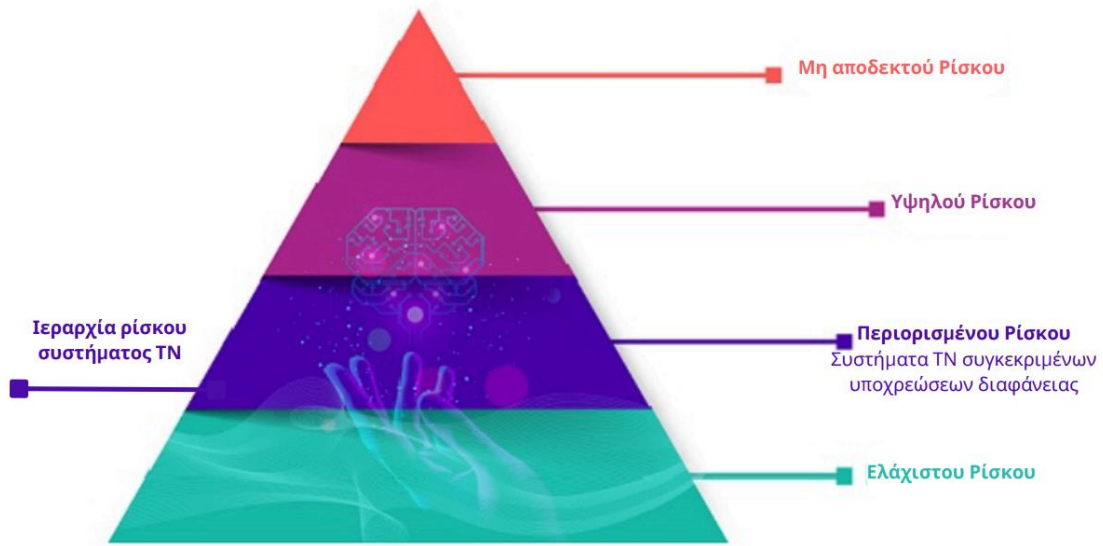
	Μέτρα για τη μείωση των περιβαλλοντολογικών επιπτώσεων του κύκλου ζωής του συστήματος TN και συμμετοχή σε διαγωνισμούς για την ανάπτυξη λύσεων TN που αντιμετωπίζουν αυτό το πρόβλημα.
Λογοδοσία	<p>Σχεδιασμός ενός συστήματος με τρόπο που να μπορεί να ελεγχθεί αργότερα, οδηγεί σε μια πιο σπονδυλωτή και στιβαρή αρχιτεκτονική συστήματος. Συνεπώς, συνιστάται ιδιαίτερα η διασφάλιση της αρθρωτότητας, της ιχνηλασιμότητας της ροής ελέγχου και δεδομένων και των κατάλληλων μηχανισμών καταγραφής.</p> <p>Η διευκόλυνση του ελέγχου από τρίτους μπορεί να συμβάλλει στη δημιουργία εμπιστοσύνης στην τεχνολογία των ΘΑΠΕ. Επιπλέον, αποτελεί ισχυρή ένδειξη εφαρμογής της δέουσας προσοχής κατά την ανάπτυξη και της τήρησης βέλτιστων πρακτικών και βιομηχανικών προτύπων.</p>

4.2.2 Περίπτωση Χρήσης στην Γενετική Τεχνητή Νοημοσύνη

4.2.2.1 Εφαρμογή Μεθοδολογίας στην Διαχείριση Προϊόντων Λογισμικού - ΔΠΛ

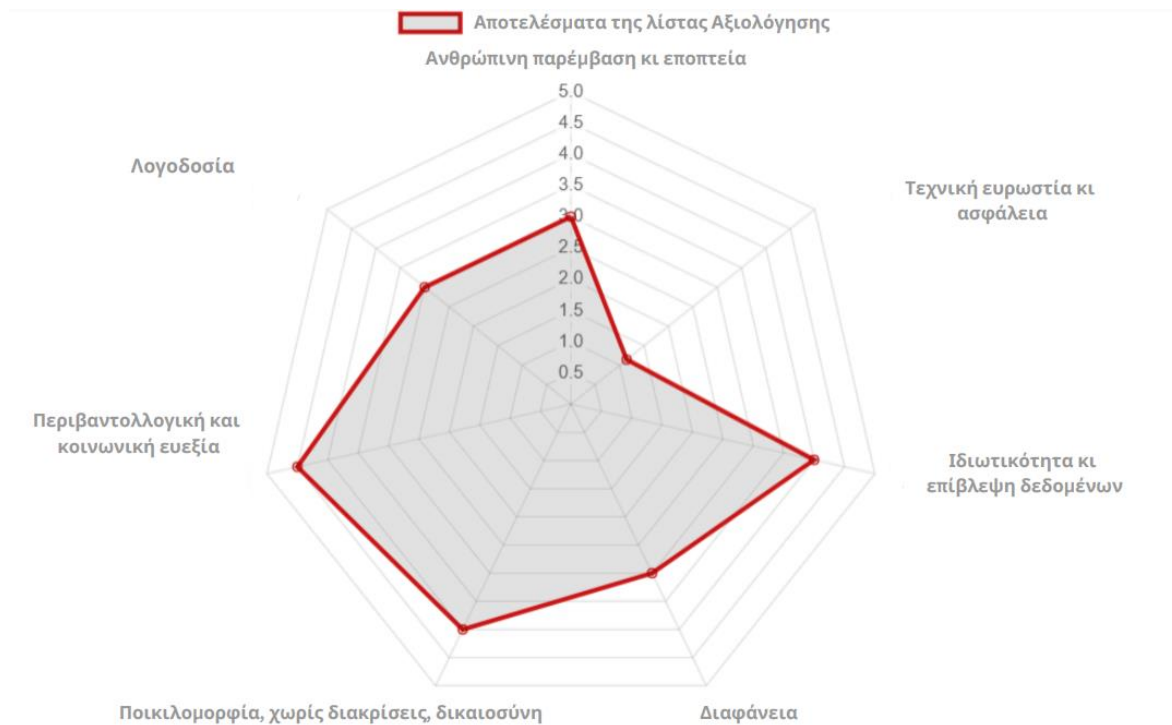
1) AI Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

Η χρήση λύσεων ΓΤΝ στην Διαχείριση Προϊόντων Λογισμικού – ΔΠΛ δεν έρχεται χωρίς πιθανά όρια ή εμπόδια, τα οποία πρέπει να αναγνωριστούν και να αντιμετωπιστούν. Είναι σημαντικό να δοθεί ιδιαίτερη προσοχή στην τρέχουσα πρακτική όσο και σε μελλοντική μελέτη σε θέματα μεροληψίας, διαφάνειας, προστασίας της ιδιωτική ζωής, λογοδοσίας, δεοντολογίας και προκλήσεων ενσωμάτωσης. Περιεχόμενο που δημιουργήθηκε από ΓΤΝ εξακολουθεί να πρέπει να ελέγχεται από τους διαχειριστές προϊόντων για να επαληθεύεται η αυθεντικότητα και η ορθότητα του. Συμπερασματικά, κατηγοριοποιείται στην ιεραρχία του AI Νομικού Πλαισίου ως λύση περιορισμένου ρίσκου η οποία συσχετίζεται κατά κύριο λόγο με την έλλειψη διαφάνειας στην χρήση της ΓΤΝ. Σε αυτήν την περίπτωση, το AI Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας ώστε να διασφαλιστεί πως οι άνθρωποι ενημερώνονται όταν χρειάζεται, ενισχύοντας έτσι την εμπιστοσύνη. Οι πάροχοι λύσεων ΓΤΝ στην διαχείριση προϊόντων λογισμικού πρέπει να διασφαλίζουν πως το παραγόμενο περιεχόμενο από TN είναι ταυτοποιήσιμο.



Εικόνα 12: Ιεραρχία AI Νομικού Πλαισίου: Διαχείριση Προϊόντων Λογισμικού – ΔΠΛ

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 13: ALTAI Αξιολόγηση: Διαχείριση Προϊόντων Λογισμικού – ΔΠΛ

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 12: ALTAI Προτάσεις για Διαχείριση Προϊόντων Λογισμικού – ΔΠΛ.

Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	<p>Ενσωμάτωση διαδικασίας κατά την οποία οι τελικοί χρήστες ή/και τα υποκείμενα γνωρίζουν επαρκώς ότι ένα σύστημα ΓΤΝ επηρέασε την απόφαση, το περιεχόμενο, τη συμβουλή ή το αποτέλεσμα.</p> <p>Διασφάλιση ότι οι τελικοί χρήστες ή τα υποκείμενα ενημερώνονται επαρκώς ότι αλληλεπιδρούν με ένα σύστημα ΓΤΝ.</p>
Τεχνική ευρωστία και ασφάλεια	Έλεγχος της ανθεκτικότητας του συστήματος σε κυβερνοεπιθέσεις.
Ιδιωτικότητα και επίβλεψη δεδομένων	<p>Εξέταση ενδεχομένου δημιουργίας μηχανισμών που επιτρέπουν την επισήμανση ζητημάτων που σχετίζονται με την προστασία της ιδιωτικής ζωής ή την προστασία των δεδομένων όσον αφορά το σύστημα ΓΤΝ.</p> <p>Κατά περίπτωση, εφαρμογή του δικαιώματος ανάκλησης της συγκατάθεσης, του δικαιώματος της αντίρρησης και του δικαιώματος της διαγραφής στο σύστημα ΓΤΝ.</p> <p>Εξέταση των επιπτώσεων στην προστασία της ιδιωτικής ζωής και των δεδομένων που συλλέγονται, παράγονται ή υποβάλλονται σε επεξεργασία κατά τη διάρκεια του κύκλου ζωής του συστήματος ΓΤΝ.</p>
Διαφάνεια	<p>Εξέταση του ενδεχομένου εξήγησης στους τελικούς χρήστες της απόφασης που υιοθετεί ή προτείνει το σύστημα ΓΤΝ.</p> <p>Εξέταση του ενδεχομένου διενέργειας συνεχών ερευνών στους χρήστες για τον βαθμό κατανόησης των αποφάσεων του συστήματος ΓΤΝ.</p> <p>Εξέταση του ενδεχομένου ενημέρωσης χρηστών σχετικά με το σκοπό, τα κριτήρια και τους περιορισμούς των αποφάσεων που παράγει το σύστημα ΓΤΝ.</p>
Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη	Συμβουλή των ειδικά επηρεαζόμενων κοινοτήτων σχετικά με τον σωστό ορισμό της δικαιοσύνης, όπως οι εκπρόσωποι των ηλικιωμένων ατόμων ή ατόμων με αναπηρίες.

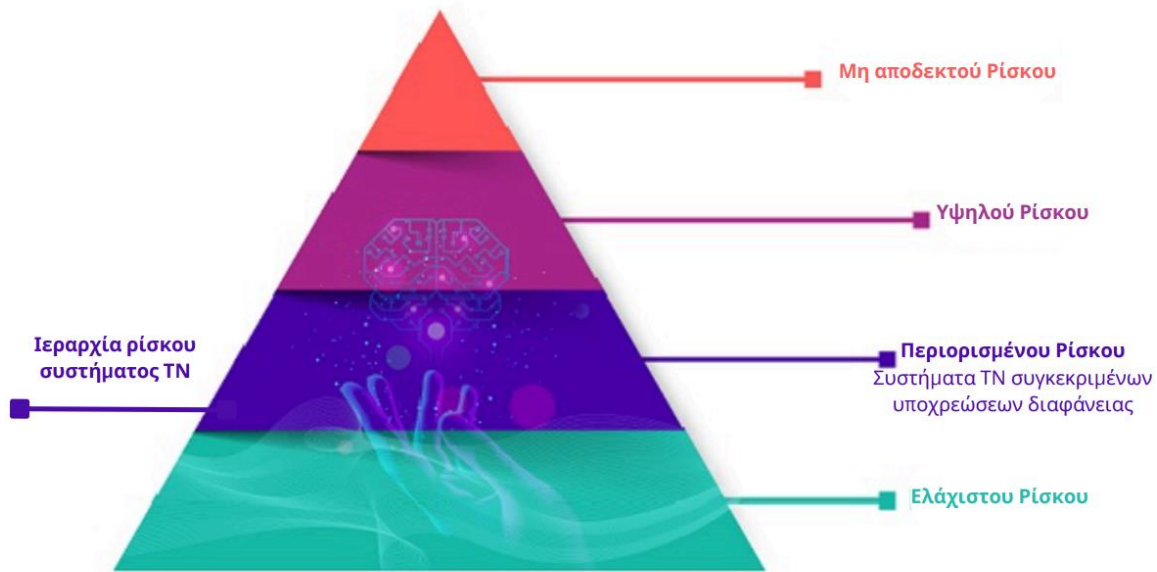
	<p>Εξασφάλιση ποσοτικής ανάλυσης και μετρήσεων, και δοκιμή του εφαρμοσμένου ορισμού της δικαιοσύνης.</p> <p>Καθιέρωση μηχανισμών διασφάλισης δικαιοσύνης στο σύστημα GTN.</p> <p>Συμμετοχή ή συμβουλή τελικών χρηστών για θέματα ανάγκης βοηθητικής τεχνολογίας κατά τη διάρκεια του σχεδιασμού και της φάσης ανάπτυξης του συστήματος GTN.</p>
Περιβαλλοντολογική και κοινωνική ευεξία	<p>Μέτρα μείωσης των περιβαλλοντολογικών επιπτώσεων του κύκλου ζωής του συστήματος GTN και συμμετοχή σε διαγωνισμούς για την ανάπτυξη λύσεων GTN που αντιμετωπίζουν αυτό το πρόβλημα.</p> <p>Ευκαιρίες κατάρτισης και υλικό για μέτρα επανεκπαίδευσης και εξέλιξης των εργαζομένων.</p>
Λογοδοσία	<p>Ο σχεδιασμός ενός συστήματος με τρόπο που να μπορεί να ελεγχθεί αργότερα, οδηγεί σε μια πιο σπονδυλωτή και στιβαρή αρχιτεκτονική συστήματος. Συνεπώς, συνίσταται ιδιαίτερα η διασφάλιση της αρθρωτότητας, της ιχνηλασιμότητας, της ροής δεδομένων και των κατάλληλων μηχανισμών καταγραφής.</p> <p>Ο έλεγχος από τρίτους μπορεί να συμβάλλει στη δημιουργία εμπιστοσύνης στην τεχνολογία και στο ίδιο το προϊόν. Επιπλέον, αποτελεί ισχυρή ένδειξη εφαρμογής της δέουσας προσοχής κατά την ανάπτυξη και την τήρηση βέλτιστων πρακτικών και βιομηχανικών προτύπων.</p> <p>Η πρόβλεψη ελέγχου ή καθοδήγηση από τρίτους μπορεί να βοηθήσει τόσο στην ποιοτική όσο και στην ποσοτική ανάλυση κινδύνου. Επιπλέον, μπορεί να συμβάλλει στην δημιουργία εμπιστοσύνης στο σύστημα GTN και το ίδιο το προϊόν.</p>

4.2.2.2 Εφαρμογή Μεθοδολογίας στην Μηχανική Λογισμικού – ΜΛ και Κυβερνο-Φυσικά Συστήματα - ΚΦΣ

1) ΑΙ Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

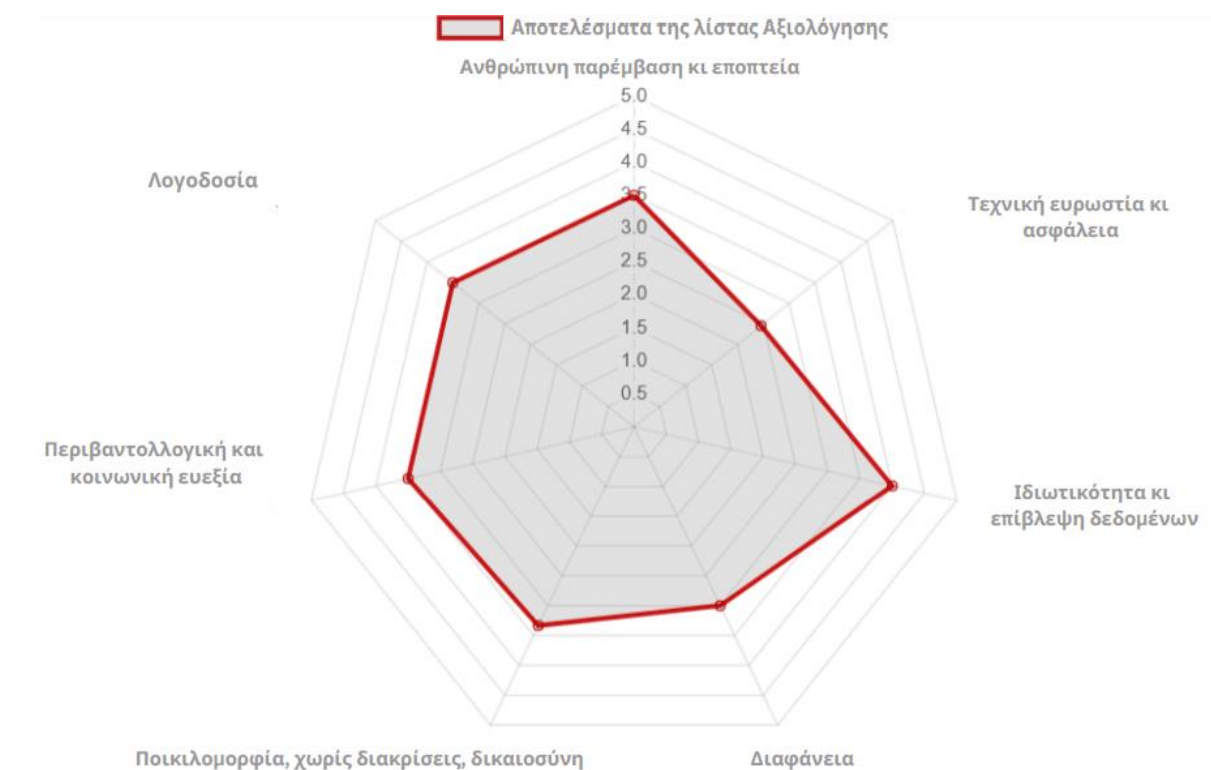
Η Ανάπτυξη Λογισμικού και τα Κυβερνο-Φυσικά Συστήματα αποτελούν συστήματα ΤΝ περιορισμένου ρίσκου τα οποία συνδέονται με την έλλειψη διαφάνειας κατά κύριο λόγο. Το ΑΙ Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας. Για παράδειγμα, όταν χρησιμοποιούνται αυτόματοι συνομιλητές ΤΝ, οι άνθρωποι πρέπει να ενημερώνονται πως αλληλεπιδρούν με ένα μηχάνημα ώστε να λάβουν πληροφορημένες αποφάσεις. Οι

πάροχοι πρέπει επίσης να διασφαλίσουν ότι το περιεχόμενο που έχει δημιουργηθεί μέσω ΤΝ είναι ταυτοποιήσιμο. Κείμενα τα οποία έχουν δημιουργηθεί μέσω ΤΝ και είναι δημόσιου ενδιαφέροντος πρέπει να συνοδεύονται με ενημέρωση του δημόσιου καλού και της φύσης τους. Το ίδιο πρέπει να εφαρμόζεται για οπτικοακουστικό περιεχόμενο που αποτελεί ψευδές υλικό.



Εικόνα 14: Ιεραρχία AI Νομικού Πλαισίου: Μηχανική Λογισμικού – ΜΛ και Κυβερνο-Φυσικά Συστήματα – ΚΦΣ

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 15: ALTAI Αξιολόγηση: Μηχανική Λογισμικού και Κυβερνο-Φυσικά Συστήματα

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 13: Προτάσεις ALTAI για Μηχανική Λογισμικού – ΜΛ και Κυβερνο-Φυσικά Συστήματα – ΚΦΣ.

Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	Εφαρμογή διαδικασιών έτσι ώστε τελικοί χρήστες – προγραμματιστές να μην βασίζονται υπερβολικά στο σύστημα TN. Διαδικασίες αποφυγής ακούσιας επιρροής της ανθρώπινης αυτονομίας από το σύστημα TN.
Τεχνική ευρωστία και ασφάλεια	Αξιολόγηση πιθανών μορφών επιθέσεων στις οποίες το σύστημα TN μπορεί να είναι ευάλωτο. Μέτρα για την διασφάλιση της ακεραιότητας, της ευρωστίας και της συνολικής ασφάλειας του συστήματος TN έναντι πιθανών επιθέσεων κατά την διάρκεια του κύκλου ζωής του.

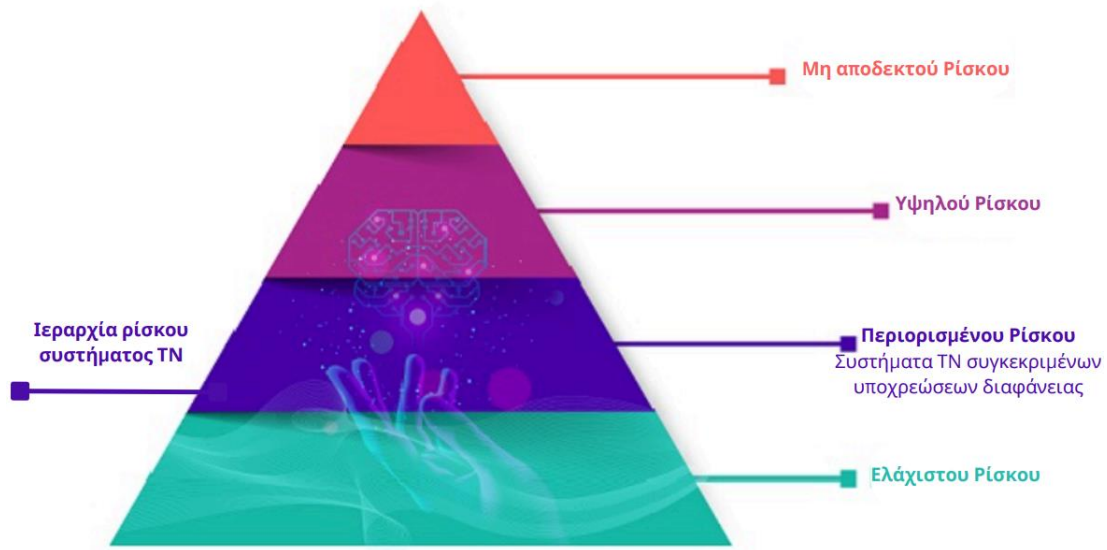
	<p>Προσδιορισμός πιθανών απειλών κατά του συστήματος TN. Για παράδειγμα, σχεδιαστικά σφάλματα, τεχνικά σφάλματα, περιβαλλοντολογικές απειλές και πιθανές συνέπειες οι οποίες προκύπτουν.</p>
<p>Ιδιωτικότητα και επίβλεψη δεδομένων</p>	<p>Κατά περίπτωση, εφαρμογή του δικαιώματος της ανάκλησης της συγκατάθεσης, το δικαίωμα αντίρρησης και το δικαίωμα διαγραφής στο σύστημα TN.</p>
<p>Διαφάνεια</p>	<p>Εξέταση του ενδεχομένου εξήγησης στους τελικούς χρήστες της απόφασης που υιοθετεί ή προτείνει το σύστημα TN.</p> <p>Εξέταση του ενδεχομένου διενέργειας συνεχών ερευνών στους χρήστες για τον βαθμό κατανόησης των αποφάσεων του συστήματος TN.</p>
<p>Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη</p>	<p>Έρευνα και χρήση δημόσια διαθέσιμων τεχνικών εργαλείων, τα οποία είναι τελευταίας τεχνολογίας για την βελτίωση της κατανόησης των δεδομένων, του μοντέλου και των επιδόσεων.</p> <p>Αξιολόγηση και θέσιμο διαδικασιών για την παρακολούθηση δυναμικών προκαταλήψεων κατά την διάρκεια όλης της ζωής ενός συστήματος TN. Για παράδειγμα, προκαταλήψεις λόγω πιθανών περιορισμών που απορρέουν από την σύνθεση των εν χρήσει συνόλων δεδομένων, όπως έλλειψη ποικιλομορφίας και μη αντιπροσωπευτικότητα.</p> <p>Εφαρμογή πρωτοβουλιών εκπαίδευσης και ευαισθητοποίησης για να βοηθηθούν οι σχεδιαστές και οι προγραμματιστές TN έτσι ώστε να έχουν μεγαλύτερη επίγνωση των πιθανών προκαταλήψεων που μπορούν να εισάγουν κατά τον σχεδιασμό και την ανάπτυξη ενός συστήματος TN.</p>
<p>Περιβαλλοντολογική και κοινωνική ευεξία</p>	<p>Εξέταση πιθανών θετικών και αρνητικών επιπτώσεων του συστήματος TN στο περιβάλλον και αξιολόγηση των επιπτώσεων αυτών.</p> <p>Μέτρα για την μείωση των περιβαλλοντολογικών επιπτώσεων του κύκλου ζωής του συστήματος TN και συμμετοχή σε διαγωνισμούς για την ανάπτυξη λύσεων TN που αντιμετωπίζουν αυτό το πρόβλημα.</p> <p>Ευκαιρίες κατάρτισης και υλικό για σεμινάρια επανεκπαίδευσης και εξέλιξης.</p>
<p>Λογοδοσία</p>	<p>Ο έλεγχος ή η καθοδήγηση από τρίτο μέρος μπορεί να βοηθήσει τόσο στην ποιοτική όσο και στην ποσοτική ανάλυση κινδύνου. Επιπλέον, μπορεί να συμβάλλει στη δημιουργία εμπιστοσύνης στην τεχνολογία και στο ίδιο το προϊόν.</p>

	<p>Τα συστήματα TN θα πρέπει να αναπτύσσονται με προληπτική προσέγγιση των κινδύνων και με τέτοιο τρόπο ώστε να συμπεριφέρονται αξιόπιστα όπως προβλέπεται, ελαχιστοποιώντας παράλληλα τις ακούσιες και απροσδόκητες βλάβες. Κατά συνέπεια, οι προγραμματιστές και οι φορείς ανάπτυξης θα πρέπει να λαμβάνουν κατάλληλη εκπαίδευση σχετικά με το νομικό πλαίσιο που ισχύει για τα συστήματα που αναπτύσσονται.</p> <p>Μια χρήσιμη μη-τεχνική μέθοδος για να εξασφαλιστεί η εφαρμογή αξιόπιστης TN είναι η συμμετοχή διαφόρων ενδιαφερόμενων μερών. Για παράδειγμα η συγκέντρωση των ενδιαφερόμενων μερών σε μια "επιτροπή ηθικού ελέγχου" για την παρακολούθηση και την υποστήριξη της διαδικασίας ανάπτυξης.</p> <p>Η συμμετοχή τρίτων μερών για την αναφορά των τρωτών σημείων και των κινδύνων συμβάλλει στον εντοπισμό και τον μετριασμό των πιθανών σφαλμάτων.</p>
--	---

4.2.2.3 Εφαρμογή Μεθοδολογίας στο Διαδίκτυο των Πραγμάτων - ΔΤΠ

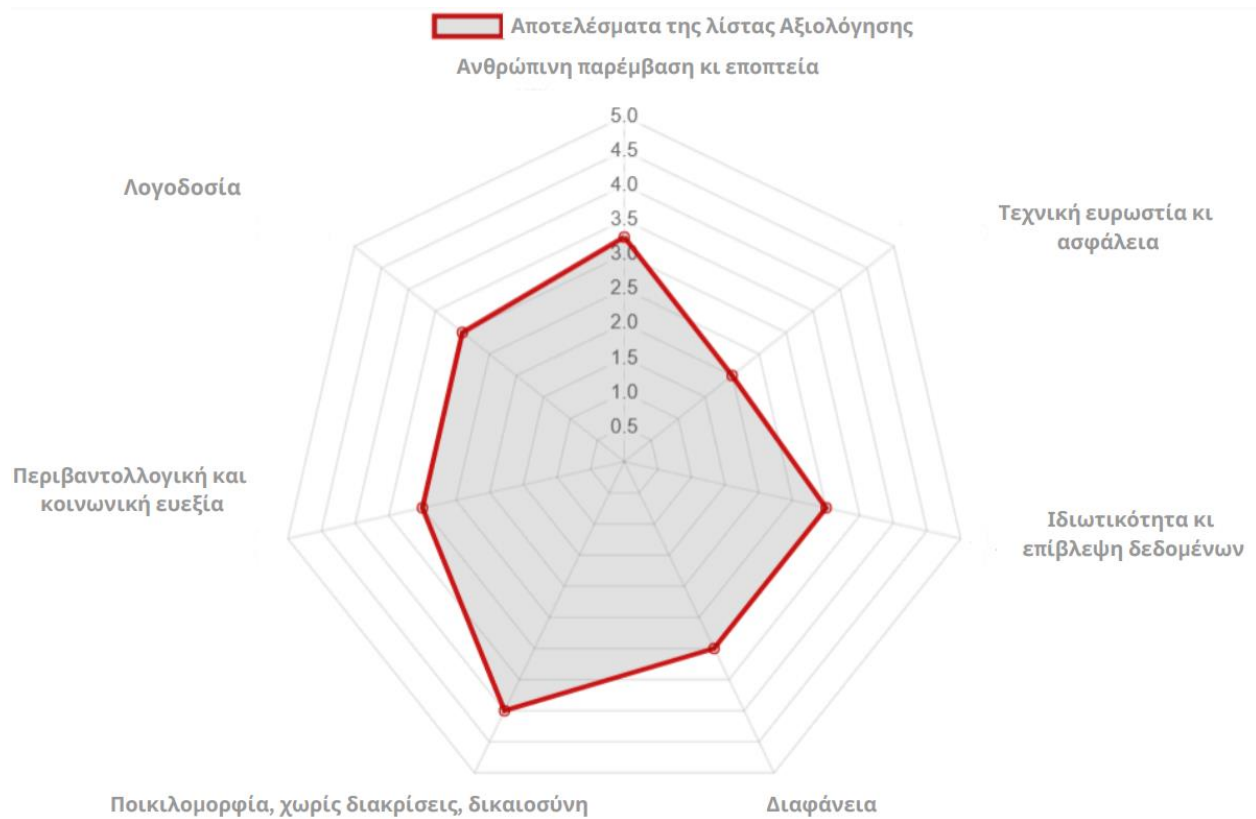
1) AI Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

Η σύγκλιση της ΓΤΝ με τα συστήματα ΔΤΠ αποτελεί σημαντικό βήμα για την αξιοποίηση του τεράστιου δυναμικού των δεδομένων που παράγονται από έξυπνες συσκευές. Η σύγκλιση αυτή αποτελεί ένα σύστημα TN περιορισμένου ρίσκου καθώς σχετίζεται κατά βάση με ρίσκα που συνδέονται με την έλλειψη διαφάνειας στον τρόπο χρήσης του. Σε τέτοιες περιπτώσεις το AI Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας προκειμένου να διασφαλίσει πως οι άνθρωποι πληροφορούνται όταν χρειάζεται, ενισχύοντας την εμπιστοσύνη στο σύστημα TN και διασφαλίζοντας την εύρυθμη λειτουργία του.



Εικόνα 16: Ιεραρχία ΑΙ Νομικού Πλαισίου: Διαδίκτυο των Πραγμάτων - ΔΤΠ

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 17: ALTAI Αξιολόγηση: Διαδίκτυο των Πραγμάτων – ΔΤΠ

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 14: ALTAI Προτάσεις για Διαδίκτυο των Πραγμάτων – ΔΤΠ.

Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	<p>Εφαρμογή διαδικασιών έτσι ώστε να αποφευχθεί ότι το σύστημα επηρεάζει ακούσια την ανθρώπινη αυτονομία.</p> <p>Μέτρα για τον μετριασμό του κινδύνου χειραγώγησης, συμπεριλαμβανομένης της παροχής σαφών πληροφοριών σχετικά με την ιδιοκτησία και τους στόχους του συστήματος, της αποφυγής αδικαιολόγητης επιτήρησης και της διατήρησης της αυτονομίας και της ψυχικής υγείας των χρηστών.</p>
Τεχνική ευρωστία και ασφάλεια	<p>Αξιολόγηση της εξάρτησης των κρίσιμων αποφάσεων του συστήματος από τη σταθερή και αξιόπιστη συμπεριφορά του.</p> <p>Σχεδιαστική αρχή ανοχής σφαλμάτων μέσω διπλών συστημάτων ή ενός άλλου παράλληλου συστήματος, βασισμένο στην TN ή "συμβατικού" τύπου.</p> <p>Έλεγχος προκειμένου να διερευνηθεί για το αν η λειτουργία του συστήματος TN μπορεί να ακυρώσει τα δεδομένα ή τις υποθέσεις με βάση τις οποίες εκπαιδεύτηκε και πώς αυτό μπορεί να οδηγήσει σε αντιφατικά αποτελέσματα. Για παράδειγμα μεροληπτικοί εκτιμητές κ.λπ.</p>
Ιδιωτικότητα και επίβλεψη δεδομένων	<p>Δημιουργία μηχανισμών που επιτρέπουν την επισήμανση ζητημάτων που σχετίζονται με την προστασία της ιδιωτικής ζωής ή την προστασία των δεδομένων όσον αφορά το σύστημα TN.</p>
Διαφάνεια	<p>Εξέταση του ενδεχομένου εξήγησης στους τελικούς χρήστες της απόφασης που υιοθετεί ή προτείνει το σύστημα TN.</p> <p>Εξέταση του ενδεχομένου διενέργειας συνεχών ερευνών στους χρήστες για το κατά πόσο κατανοούν την απόφαση ή τις αποφάσεις του συστήματος TN.</p> <p>Σε περίπτωση διαδραστικού συστήματος TN, εξέταση του ενδεχομένου γνωστοποίησης στους χρήστες ότι αλληλεπιδρούν με μια μηχανή.</p>
Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη	<p>Δημιουργία μηχανισμών διασφάλισης της δικαιοσύνης στο σύστημα TN.</p> <p>Διασφάλιση ότι το σύστημα TN ανταποκρίνεται στην ποικιλία προτιμήσεων και ικανοτήτων της κοινωνίας.</p>

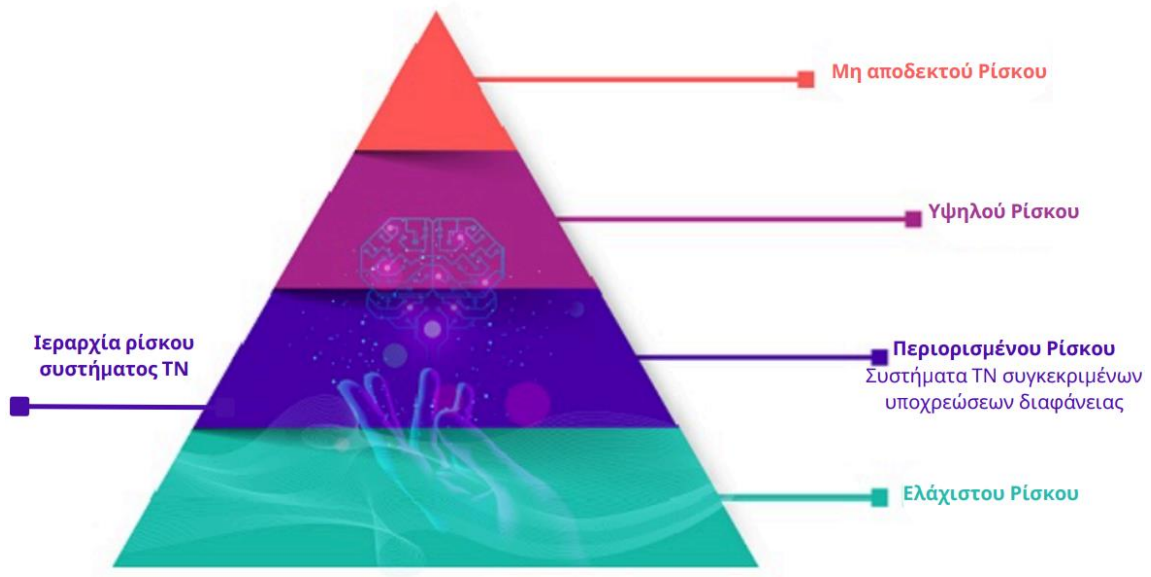
	<p>Αξιολόγηση του κατά πόσον η διεπαφή χρήστη του συστήματος TN είναι εύχρηστη από άτομα με ειδικές ανάγκες ή αναπηρίες ή άτομα που κινδυνεύουν από αποκλεισμό.</p> <p>Συνυπολογισμός του αντικτύπου του συστήματος TN στους δυνητικούς τελικούς χρήστες ή/και υποκείμενα.</p> <p>Αξιολόγηση του κατά πόσον η ομάδα που συμμετέχει στην κατασκευή του συστήματος TN ασχολήθηκε με τους πιθανούς τελικούς χρήστες ή/και υποκείμενα.</p> <p>Εξέταση μηχανισμού για να συμπεριλαμβάνεται η συμμετοχή του ευρύτερου δυνατού φάσματος ενδιαφερομένων μερών στο σχεδιασμό και την ανάπτυξη του συστήματος TN.</p>
<p>Περιβαλλοντολογική και κοινωνική ευεξία</p>	<p>Παροχή ευκαιριών κατάρτισης και υλικών για μέτρα επανακατάρτισης και εξέλιξης των δεξιοτήτων των εργαζομένων.</p>
<p>Λογοδοσία</p>	<p>Μια χρήσιμη μη-τεχνική μέθοδος για να εξασφαλιστεί η εφαρμογή αξιόπιστης TN είναι η συμμετοχή διαφόρων ενδιαφερόμενων μερών. Για παράδειγμα, συγκέντρωση των ενδιαφερόμενων μερών σε μια "επιτροπή ηθικού ελέγχου" για την παρακολούθηση και την υποστήριξη της διαδικασίας ανάπτυξης.</p> <p>Εφόσον τα συστήματα TN χρησιμοποιούνται όλο και περισσότερο για την υποστήριξη λήψης αποφάσεων ή για τη λήψη αποφάσεων από μόνα τους, πρέπει να διασφαλιστεί ότι είναι δίκαια όσον αφορά τον αντίκτυπό τους στη ζωή των ανθρώπων, ότι είναι σύμφωνα με αξίες που δεν πρέπει να διακυβεύονται και ότι μπορούν να ενεργούν αναλόγως. Πρέπει επίσης οι κατάλληλες διαδικασίες λογοδοσίας να το διασφαλίζουν αυτό. Κατά συνέπεια, όλες οι συγκρούσεις αξιών ή οι συμβιβασμοί θα πρέπει να τεκμηριώνονται και να εξηγούνται πλήρως.</p> <p>Η εμπλοκή τρίτων μερών για την αναφορά τρωτών σημείων και κινδύνων συμβάλλει στον εντοπισμό και τον μετριασμό πιθανών παγίδων.</p> <p>Η διαδικασία διαχείρισης κινδύνων θα πρέπει πάντα να περιλαμβάνει νέα ευρήματα, καθώς οι αρχικές υποθέσεις σχετικά με την πιθανότητα εμφάνισης ενός συγκεκριμένου κινδύνου μπορεί να είναι εσφαλμένες και συνεπώς, η ποσοτική ανάλυση κινδύνου δεν είναι ακόμα σωστή και θα πρέπει να αναθεωρηθεί συμπεριλαμβάνοντας τα νέα δεδομένα.</p>

4.2.3 Περίπτωση Χρήσης στην Αυτοκινητοβιομηχανία

4.2.3.1 Εφαρμογή Μεθοδολογίας στον Ταυτόχρονο Εντοπισμό και Χαρτογράφηση – TEX (0 – 3)

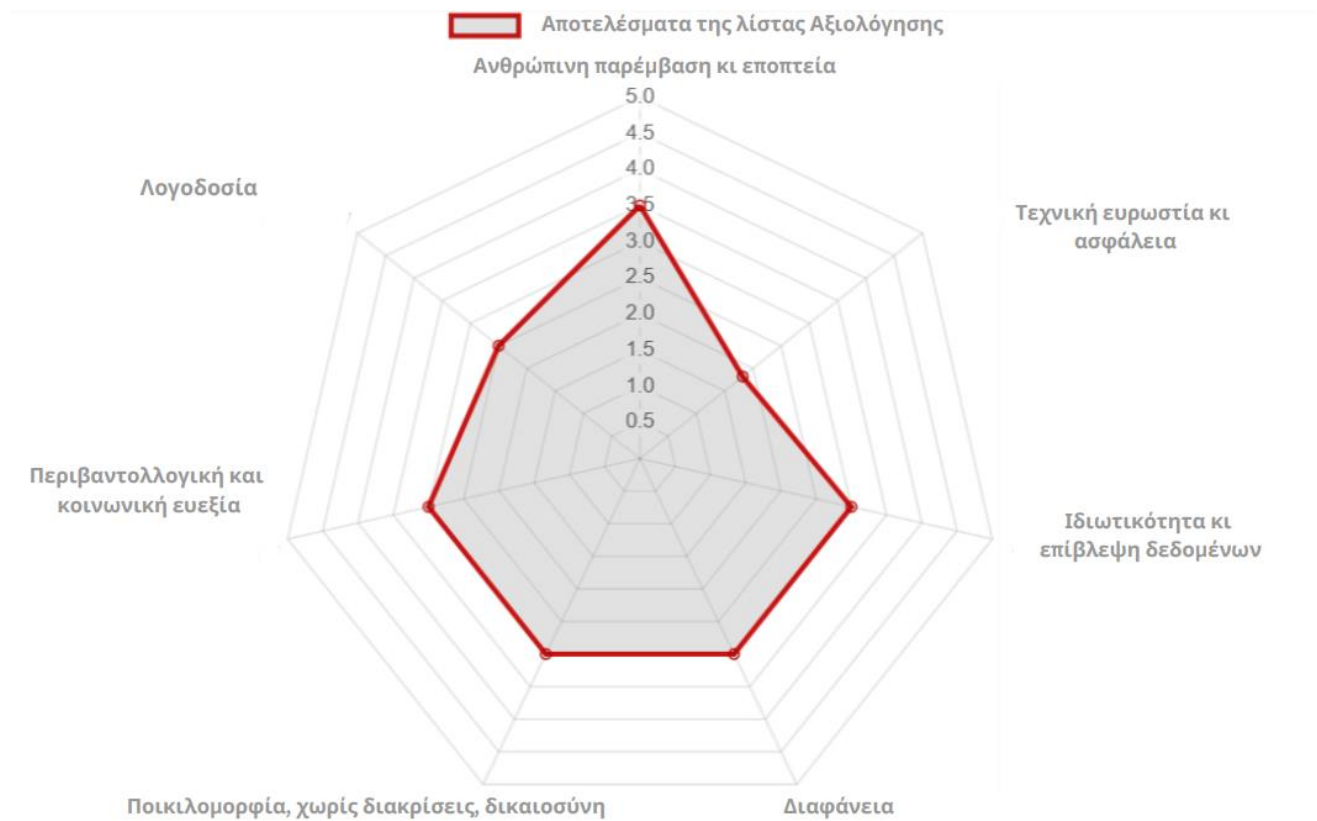
1) AI Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

Τα αυτόνομα οχήματα είναι εξοπλισμένα με Προηγμένα Συστήματα Υποβοήθησης Οδηγού – ΠΣΥΟ. Τα συστήματα TN που σχετίζονται με τον Εντοπισμό και την Χαρτογράφηση αποτελούν συστήματα τα οποία χαρακτηρίζονται από μηδενική αυτονομία – επίπεδο 0 έως υπό συνθήκη αυτονομία – επίπεδο 3. Πρακτικά, ο βαθμός αυτονομίας είναι τέτοιος που χρειάζεται στις περισσότερες περιπτώσεις να αναλαμβάνουν τον έλεγχο οι οδηγοί του ΑΟ. Συμπερασματικά, τα συστήματα Εντοπισμού και Χαρτογράφησης, ταξινομούνται ως συστήματα TN περιορισμένου ρίσκου σύμφωνα με την διαβάθμιση του AI Νομικού Πλαισίου. Το περιορισμένο ρίσκο σχετίζεται με την έλλειψη διαφάνειας. Σε τέτοιες περιπτώσεις το AI Νομικό Πλαίσιο εισάγει συγκεκριμένες επιταγές διαφάνειας προκειμένου να διασφαλίσει πως οι άνθρωποι ενημερώνονται όταν χρειάζεται. Οι πάροχοι πρέπει επίσης να διασφαλίζουν πως τα αποτελέσματα τα οποία είναι TN-παραγόμενα είναι ταυτοποιήσιμα.



Εικόνα 18: Ιεραρχία AI Νομικού Πλαισίου: Ταυτόχρονος Εντοπισμός και Χαρτογράφηση (0 – 3)

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 19: ALTAI Αξιολόγηση: Ταυτόχρονος Εντοπισμός και Χαρτογράφηση (0 – 3)

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 15: ALTAI Προτάσεις για Ταυτόχρονο Εντοπισμό και Χαρτογράφηση (0 – 3).

Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	<p>Λήψη μέτρων για την αντιμετώπιση των πιθανών αρνητικών συνεπειών για τους τελικούς χρήστες ή τα υποκείμενα σε περίπτωση που αναπτύξουν προσκόλληση. Παροχή μέσων ώστε ο χρήστης να έχει τον έλεγχο των αλληλεπιδράσεων όπου κρίνεται απαραίτητο.</p> <p>Λήψη μέτρων για την ελαχιστοποίηση του κινδύνου εθισμού με τη συμμετοχή εμπειρογνομώνων από άλλους κλάδους, όπως οι ψυχολόγοι και οι κοινωνικοί λειτουργοί.</p> <p>Μέτρα για την μείωση του κινδύνου χειραγώγησης, συμπεριλαμβανομένης της παροχής σαφών πληροφοριών σχετικά με την ιδιοκτησία και τους στόχους του συστήματος, της αποφυγής αδικαιολόγητης επιτήρησης και της διατήρησης της αυτονομίας και της ψυχικής υγείας των χρηστών.</p>
Τεχνική ευρωστία και ασφάλεια	<p>Ανοχή σε σφάλματα μέσω διπλού συστήματος ή παράλληλου συστήματος, βασισμένο σε TN ή "συμβατικού" τύπου.</p>
Ιδιωτικότητα και επίβλεψη δεδομένων	<p>Εξέταση του ενδεχομένου δημιουργίας μηχανισμών που επιτρέπουν την επισήμανση ζητημάτων που σχετίζονται με την προστασία της ιδιωτικής ζωής ή την προστασία των δεδομένων όσον αφορά το σύστημα TN.</p> <p>Κατά περίπτωση, εφαρμογή του δικαιώματος της ανάκλησης της συγκατάθεσης, του δικαιώματος της αντίρρησης και του δικαιώματος στη λήθη στο σύστημα TN.</p> <p>Ευθυγράμμιση του συστήματος όποτε είναι εφικτό και τελέσφορο σύμφωνα με τα πρότυπα του Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών - ΙΗΗΜ ή τα ευρέως υιοθετημένα πρωτόκολλα για την διαχείριση και διακυβέρνηση δεδομένων.</p>
Διαφάνεια	<p>Σε περίπτωση διαδραστικού συστήματος TN, εξέταση του ενδεχομένου επικοινωνίας στους χρήστες ότι αλληλεπιδρούν με μια μηχανή - <i>Οι οδηγοί ήδη γνωρίζουν πως αλληλεπιδρούν με μια μηχανή</i></p>
Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη	<p>Έρευνα και χρήση δημόσια διαθέσιμων τεχνικών εργαλείων, τα οποία είναι τελευταίας τεχνολογίας, για την βελτίωση της κατανόησης των δεδομένων, των μοντέλων και των επιδόσεων.</p>

	<p>Αξιολόγηση και εφαρμογή διαδικασιών για τον έλεγχο και την παρακολούθηση πιθανών μεροληψιών καθ' όλη την διάρκεια του κύκλου ζωής του συστήματος TN. Για παράδειγμα μεροληψίες λόγω πιθανών περιορισμών που απορρέουν από τη σύνθεση των εν χρήσει συνόλων δεδομένων, έλλειψη ποικιλομορφίας και μη αντιπροσωπευτικότητα.</p> <p>Εφαρμογή πρωτοβουλιών εκπαίδευσης και ευαισθητοποίησης για να για να βοηθηθούν οι σχεδιαστές και οι προγραμματιστές TN ώστε να έχουν μεγαλύτερη επίγνωση των πιθανών προκαταλήψεων που μπορεί να εισάγουν κατά τον σχεδιασμό και την ανάπτυξη του συστήματος TN.</p> <p>Αξιολόγηση του κατά πόσο υπάρχουν ομάδες που δυνητικά επηρεάζονται δυσανάλογα από τα αποτελέσματα του συστήματος.</p> <p>Αξιολόγηση του κινδύνου πιθανής αδικίας του συστήματος έναντι των κοινοτήτων των τελικών χρηστών ή των υποκειμένων.</p>
<p>Περιβαλλοντολογική και κοινωνική ευεξία</p>	<p>Εξέταση πιθανών θετικών και αρνητικών επιπτώσεων του συστήματος TN στο περιβάλλον και θέσπιση μηχανισμών για την αξιολόγηση των επιπτώσεων αυτών.</p> <p>Καθορισμός μέτρων για την μείωση των περιβαλλοντολογικών επιπτώσεων του κύκλου ζωής του συστήματος TN και συμμετοχή σε διαγωνισμούς για την ανάπτυξη λύσεων TN που αντιμετωπίζουν αυτό το πρόβλημα.</p> <p>Ενημέρωση των εργαζομένων που επηρεάζονται και τους εκπροσώπους τους, ταυτόχρονη εμπλοκή και άλλων ενδιαφερόμενων φορέων. Εφαρμογή στην επικοινωνία, στην εκπαίδευση και στην κατάρτιση σε επιχειρησιακό και διοικητικό επίπεδο.</p> <p>Λήψη μέτρων για την αντιμετώπιση της μείωσης της οδηγικής ικανότητας μέσω της συνεχούς κατάρτισης, ιδίως σε ευαίσθητες περιοχές από άποψη ασφάλειας και προστασίας.</p>
<p>Λογοδοσία</p>	<p>Ο σχεδιασμός συστήματος με τρόπο που μπορεί να ελεγχθεί αργότερα, οδηγεί σε μια πιο σπονδυλωτή και στιβαρή αρχιτεκτονική συστήματος. Συνεπώς, συνιστάται ιδιαίτερα η διασφάλιση της αρθρωτότητας, της ιχνηλασιμότητας της ροής ελέγχου και δεδομένων και των κατάλληλων μηχανισμών καταγραφής.</p> <p>Η διευκόλυνση του ελέγχου από τρίτους μπορεί να συμβάλλει στην δημιουργία εμπιστοσύνης στην Αυτόνομη Οδήγηση και στα</p>

ίδια τα ΑΟ. Επιπλέον, αποτελεί ισχυρή ένδειξη εφαρμογής της δέουσας προσοχής κατά την ανάπτυξη και της τήρησης βέλτιστων πρακτικών και βιομηχανικών προτύπων.

Ο προβλεπόμενος έλεγχος ή η καθοδήγηση από τρίτο μέρος μπορεί να βοηθήσει τόσο στην ποιοτική όσο και στην ποσοτική ανάλυση κινδύνου. Επιπλέον, μπορεί να συμβάλλει στην δημιουργία εμπιστοσύνης στην Αυτόνομη Οδήγηση και στο ίδιο το προϊόν.

Τα συστήματα ΤΝ πρέπει να αναπτύσσονται με προληπτική προσέγγιση των κινδύνων και με τρόπο ώστε να συμπεριφέρονται αξιόπιστα όπως προβλέπεται, ελαχιστοποιώντας παράλληλα τις ακούσιες και απροσδόκητες βλάβες και αποτρέποντας τις ανθρώπινες απώλειες. Κατά συνέπεια, οι προγραμματιστές και οι φορείς ανάπτυξης θα πρέπει να λαμβάνουν κατάλληλη εκπαίδευση σχετικά με το νομικό πλαίσιο που ισχύει για τα συστήματα ΤΝ που αναπτύσσονται.

Μια χρήσιμη μη-τεχνική μέθοδος για να εξασφαλιστεί η εφαρμογή αξιόπιστης ΤΝ είναι η συμμετοχή διαφόρων ενδιαφερόμενων μερών. Για παράδειγμα, συγκέντρωση ενδιαφερόμενων μερών σε μια "επιτροπή ηθικού ελέγχου" για την παρακολούθηση και την υποστήριξη της διαδικασίας ανάπτυξης.

Η συμμετοχή τρίτων μερών για την αναφορά των τρωτών σημείων και των κινδύνων συμβάλλει στον εντοπισμό και τον μετριασμό των πιθανών παγίδων.

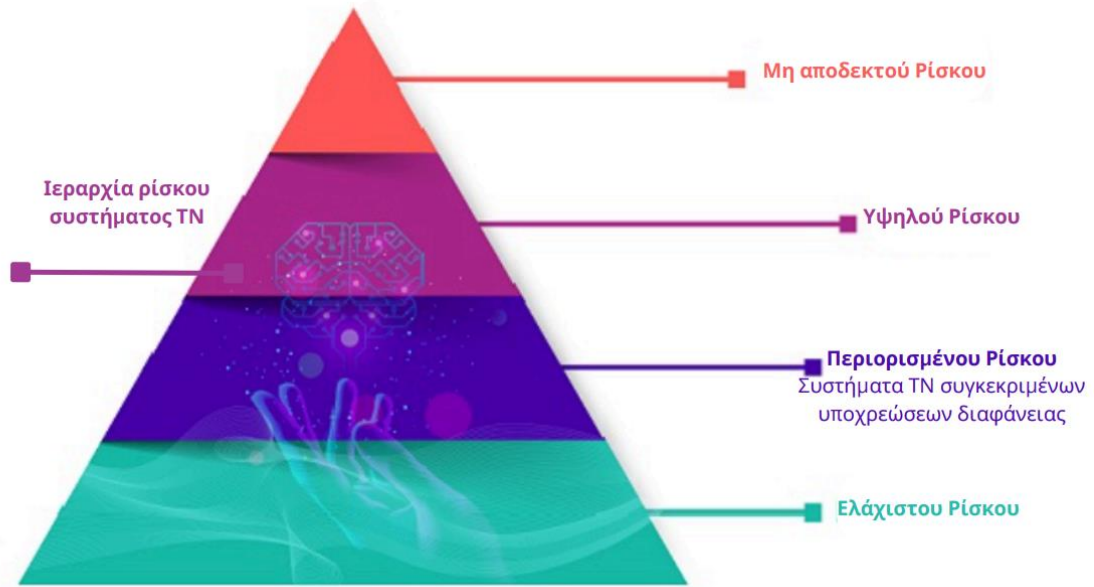
Μια διαδικασία διαχείρισης κινδύνων θα πρέπει πάντα να περιλαμβάνει νέα ευρήματα, καθώς οι αρχικές παραδοχές σχετικά με την πιθανότητα εμφάνισης ενός συγκεκριμένου κινδύνου μπορεί να είναι εσφαλμένες και ως εκ τούτου η ποσοτική ανάλυση κινδύνου δεν είναι σωστή και θα πρέπει να αναθεωρηθεί με νέα ευρήματα.

4.2.3.2 Εφαρμογή Μεθοδολογίας στον Έλεγχο και Δημιουργία Πορείας – ΕΔΠ (4 – 5)

Ι) ΑΙ Νομικό Πλαίσιο – Αρχική Κατηγοριοποίηση

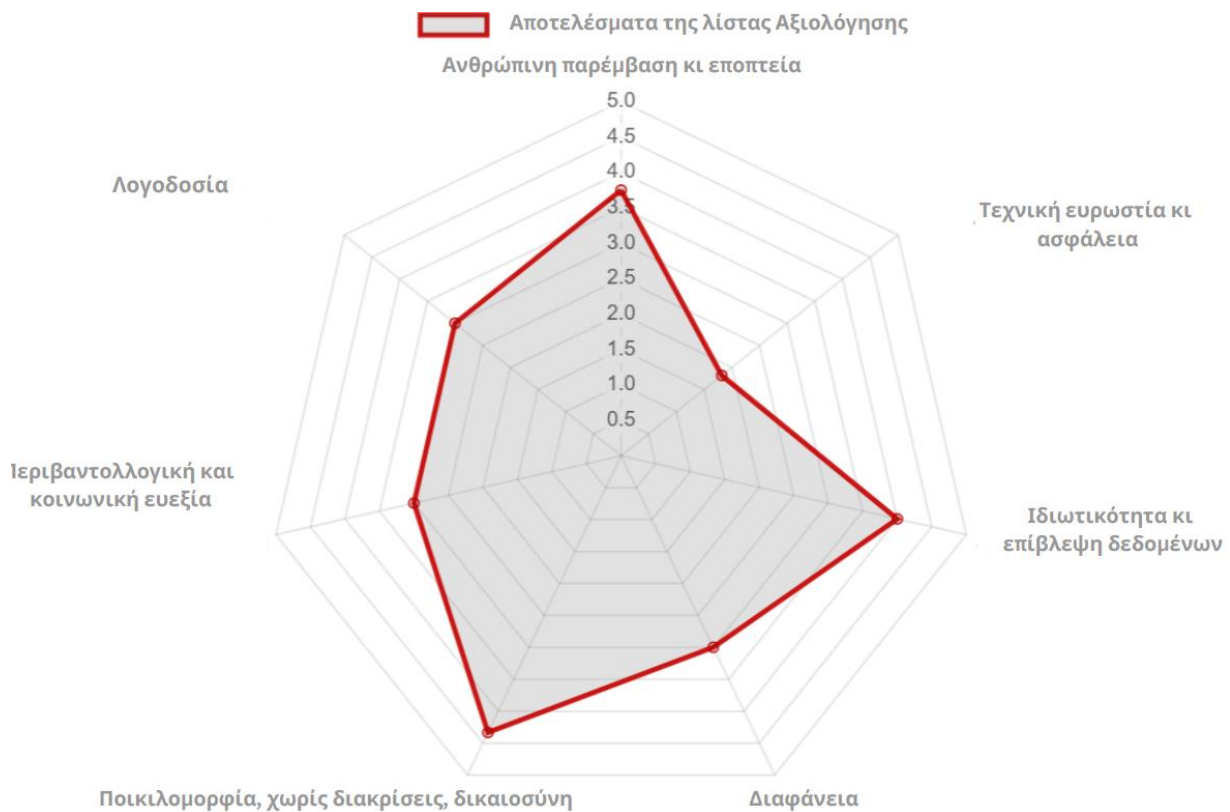
Τα αυτόνομα οχήματα είναι εξοπλισμένα με Προηγμένα Συστήματα Υποβοήθησης Οδηγού – ΠΣΥΟ. Τα συστήματα ΤΝ που σχετίζονται με τον Έλεγχο και την Δημιουργία Πορείας αποτελούν συστήματα τα οποία χαρακτηρίζονται από υψηλή αυτονομία – επίπεδο 4 έως πλήρη αυτονομία – επίπεδο 5. Πρακτικά, ο βαθμός αυτονομίας είναι τέτοιος που σχεδόν σε όλες τις περιπτώσεις το όχημα μπορεί να διαχειριστεί όλες τις οδηγικές απαιτήσεις χωρίς καμία ανθρώπινη παρέμβαση. Τα ΑΟ με τέτοια επίπεδα αυτονομίας δυνητικά μπορούν να θέσουν σε κίνδυνο την ζωή και την υγεία των πολιτών. Συμπεριλαμβάνονται δηλαδή, στις μεταφορές οι οποίες αποτελούν υποδομές κριτικής σημασίας.

Συμπερασματικά, σύμφωνα με το ΑΙ Νομικό Πλαίσιο κατηγοριοποιούνται σε συστήματα ΤΝ υψηλού ρίσκου και είναι αντικείμενα αυστηρών επιταγών προτού γίνουν δημοσίως διαθέσιμα στις αγορές. Σε τέτοιες περιπτώσεις το ΑΙ Νομικό Πλαίσιο επιτάσσει επαρκή συστήματα αξιολόγησης και μετριασμού κινδύνου, υψηλής ποιότητας σύνολα δεδομένων που τροφοδοτούν το σύστημα με στόχο την ελαχιστοποίηση των κινδύνων και των αποτελεσμάτων που προκαλούν διακρίσεις. Επίσης, κρίνεται απαραίτητη η καταγραφή των δραστηριοτήτων προκειμένου να διασφαλίζεται η ιχνηλασιμότητα των αποτελεσμάτων, η λεπτομερής καταγραφή σε έγγραφα που παρέχει όλες τις απαραίτητες πληροφορίες για το σύστημα και το σκοπό του έτσι ώστε να μπορεί να αξιολογηθεί από τις αρχές ο βαθμός συμμόρφωσής του. Χρειάζονται επίσης, τα κατάλληλα μέτρα ανθρώπινης εποπτείας για την ελαχιστοποίηση του κινδύνου και υψηλά επίπεδα αξιοπιστίας, ασφάλειας και ακρίβειας.



Εικόνα 20: Ιεραρχία ΑΙ Νομικού Πλαισίου: Έλεγχος και Δημιουργία Πορείας - ΕΔΠ (4 – 5)

II) ALTAI Ερωτηματολόγιο - Εργαλείο Αξιολόγησης



Εικόνα 21: ALTAI Αξιολόγηση: Έλεγχος και Δημιουργία Πορείας – ΕΔΠ (4 – 5)

III) Προτάσεις ALTAI και Σχόλια

Πίνακας 16: ALTAI Προτάσεις για Έλεγχο και Δημιουργία Πορείας - ΕΔΠ (4 – 5).

Απαιτήσεις	Προτάσεις
Ανθρώπινη εκπροσώπηση και εποπτεία	<p>Εφαρμογή διαδικασιών για την αποφυγή υπερβολικής εξάρτησης των οδηγών από το σύστημα TN.</p> <p>Εφαρμογή οποιασδήποτε διαδικασίας για την αποφυγή ότι το σύστημα TN επηρεάζει ακούσια την ανθρώπινη αυτονομία.</p> <p>Μέτρα για την αντιμετώπιση των πιθανών αρνητικών συνεπειών για τους οδηγούς σε περίπτωση που αναπτύξουν προσκόλληση. Παροχή μέσων ώστε ο χρήστης να έχει τον έλεγχο των αλληλεπιδράσεων όταν κρίνεται απαραίτητο.</p> <p>Μέτρα για την ελαχιστοποίηση του κινδύνου εθισμού με τη συμμετοχή εμπειρογνομόνων από άλλους κλάδους, όπως ψυχολόγοι και κοινωνικοί λειτουργοί.</p> <p>Μέτρα για τον μετριασμό του κινδύνου χειραγώγησης, συμπεριλαμβανομένης της παροχής σαφών πληροφοριών σχετικά με την ιδιοκτησία και τους στόχους του συστήματος TN, την αποφυγή αδικαιολόγητης επιτήρησης και την διατήρηση της αυτονομίας και της ψυχικής υγείας των χρηστών.</p>
Τεχνική ευρωστία και ασφάλεια	<p>Προσδιορισμός των πιθανών απειλών για το σύστημα TN, σχεδιαστικών σφαλμάτων, τεχνικών σφαλμάτων, περιβαλλοντολογικών απειλών και των πιθανών συνεπειών που προκύπτουν.</p> <p>Αξιολόγηση του κινδύνου πιθανής κακόβουλης χρήσης, κατάχρησης ή ακατάλληλης χρήσης του συστήματος TN.</p> <p>Αξιολόγηση των κρίσιμων αποφάσεων του συστήματος TN από τη σταθερή και αξιόπιστη συμπεριφορά του.</p> <p>Σχεδιαστική ανοχή σε σφάλματα μέσω διπλού συστήματος TN ή άλλου παράλληλου συμβατικού τύπου συστήματος.</p> <p>Έλεγχος για την εφαρμογή συγκεκριμένων πλαισίων ή συνθηκών για την διασφάλιση της αναπαραξιμότητας.</p>
Ιδιωτικότητα και επίβλεψη δεδομένων	<p>Εξέταση του ενδεχομένου δημιουργίας μηχανισμών που επιτρέπουν την επισήμανση ζητημάτων που σχετίζονται με την προστασία της ιδιωτικής ζωής ή την προστασία των δεδομένων όσον αφορά το σύστημα TN.</p>

	<p>Εφαρμογή του δικαιώματος ανάκλησης της συγκατάθεσης, του δικαιώματος της αντίρρησης και του δικαιώματος στην λήθη.</p> <p>Όποτε είναι δυνατόν και σκόπιμο, ευθυγράμμιση του συστήματος TN με τα σχετικά πρότυπα του Ινστιτούτου Ηλεκτρολόγων και Ηλεκτρονικών Μηχανικών - ΙΗΗΜ ή τα ευρέως υιοθετημένα πρωτόκολλα για την διαχείριση και διακυβέρνηση δεδομένων.</p>
Διαφάνεια	Εξέταση του ενδεχομένου διενέργειας συνεχών ερευνών στους χρήστες για το κατά πόσο κατανοούν τις αποφάσεις του συστήματος TN.
Ποικιλομορφία, χωρίς διακρίσεις, δικαιοσύνη	<p>Εξέταση της ποικιλομορφίας και της αντιπροσωπευτικότητας των οδηγών στα δεδομένα.</p> <p>Έρευνα και χρήση δημοσίως διαθέσιμων τεχνικών εργαλείων, τα οποία είναι τελευταίας τεχνολογίας για την βελτίωση της κατανόησης των δεδομένων, των μοντέλων και των επιδόσεων.</p> <p>Αξιολόγηση και θέσπιση διαδικασιών για τον έλεγχο και την παρακολούθηση πιθανών προκαταλήψεων κατά τη διάρκεια ολόκληρου του κύκλου ζωής του συστήματος TN. Για παράδειγμα, προκαταλήψεις λόγω πιθανών περιορισμών που απορρέουν από τη σύνθεση των εν χρήσει συνόλων δεδομένων, έλλειψη ποικιλομορφίας, μη αντιπροσωπευτικότητα.</p> <p>Θέσπιση πρωτοβουλιών εκπαίδευσης και ευαισθητοποίησης για την ενίσχυση γνώσεων στους σχεδιαστές και στους προγραμματιστές TN έτσι ώστε να έχουν μεγαλύτερη επίγνωση της πιθανής προκατάληψης που μπορούν να εισάγουν κατά τον σχεδιασμό και την ανάπτυξη του συστήματος TN.</p> <p>Ανάλογα με την περίπτωση χρήσης, εξασφάλιση ενός μηχανισμού που επιτρέπει την επισήμανση ζητημάτων που σχετίζονται με προκατάληψη, τις διακρίσεις ή τις χαμηλότερες του αποδεκτού επιδόσεις του συστήματος TN.</p> <p>Καθορισμός ξεκάθαρων βημάτων και τρόπων επικοινωνίας σχετικά με το πώς και σε ποιον μπορούν να αναφερθούν τέτοια ζητήματα.</p>
Περιβαλλοντολογική και κοινωνική ευεξία	<p>Ενημέρωση και διαβούλευση με τους θιγόμενους εργαζόμενους – οδηγούς ταξί και συμμετοχή άλλων ενδιαφερόμενων φορέων. Επικοινωνία, εκπαίδευση και κατάρτιση σε επιχειρησιακό και διοικητικό επίπεδο.</p> <p>Λήψη μέτρων για την διασφάλιση ότι οι επιπτώσεις του συστήματος TN στην εργασία είναι καλά κατανοητές βάσει</p>

	<p>ανάλυσης των διαδικασιών εργασίας και ολόκληρου του κοινωνικο-τεχνικού συστήματος.</p> <p>Μέτρα για την αντιμετώπιση της μείωσης της οδηγικής ικανότητας μέσω της συνεχούς κατάρτισης, ιδίως σε ευαίσθητες περιοχές από άποψη ασφάλειας και προστασίας.</p> <p>Παροχή ευκαιριών κατάρτισης και υλικών με στόχο την εξέλιξη των οδηγικών ικανοτήτων παρέμβασης.</p>
<p>Λογοδοσία</p>	<p>Διευκόλυνση του ελέγχου από τρίτους με στόχο να συμβάλλει στη δημιουργία εμπιστοσύνης στην Αυτόνομη Οδήγηση και στα ίδια τα οχήματα. Επιπλέον, αποτελεί ισχυρή ένδειξη της εφαρμογής της δέουσας προσοχής κατά την ανάπτυξη και τήρησης των βέλτιστων πρακτικών και βιομηχανικών προτύπων.</p> <p>Ο σχεδιασμός ενός συστήματος TN με τρόπο που να μπορεί να ελεγχθεί αργότερα, οδηγεί σε μια πιο αρθρωτή και στιβαρή αρχιτεκτονική συστήματος. Συνεπώς, συνιστάται ιδιαίτερα η διασφάλιση της αρθρωτότητας, της ιχνηλασιμότητας της ροής ελέγχου και δεδομένων και των κατάλληλων μηχανισμών καταγραφής.</p> <p>Ο προβλεπόμενος έλεγχος ή η καθοδήγηση από τρίτο μέρος μπορεί να βοηθήσει τόσο στην ποιοτική όσο και στην ποσοτική ανάλυση κινδύνου. Επιπλέον, μπορεί να συμβάλλει στη δημιουργία εμπιστοσύνης στην τεχνολογία και στο ίδιο το προϊόν.</p> <p>Τα συστήματα TN πρέπει να αναπτύσσονται με προληπτική προσέγγιση των κινδύνων και με τρόπο ώστε να συμπεριφέρονται αξιόπιστα όπως προβλέπεται, ελαχιστοποιώντας παράλληλα τις ακούσιες και απροσδόκητες βλάβες. Κατά συνέπεια, οι προγραμματιστές και οι φορείς ανάπτυξης θα πρέπει να λαμβάνουν κατάλληλη εκπαίδευση σχετικά με το νομικό πλαίσιο που ισχύει για τα συστήματα που αναπτύσσονται.</p> <p>Μια χρήσιμη μη-τεχνική μέθοδος για την εξασφάλιση αξιόπιστης TN είναι η συμμετοχή διαφόρων ενδιαφερόμενων μερών. Για παράδειγμα, η συγκέντρωση των ενδιαφερόμενων μερών σε μια "επιτροπή ηθικού ελέγχου" για την παρακολούθηση και την υποστήριξη της διαδικασίας ανάπτυξης.</p>

	<p>Η συμμετοχή τρίτων μερών για την αναφορά των τρωτών σημείων και των κινδύνων συμβάλλει στον εντοπισμό και τον μετριασμό των πιθανών παγίδων.</p> <p>Μια διαδικασία διαχείρισης κινδύνων θα πρέπει πάντα να περιλαμβάνει νέα ευρήματα, καθώς οι αρχικές παραδοχές σχετικά με την πιθανότητα εμφάνισης ενός συγκεκριμένου κινδύνου μπορεί να είναι εσφαλμένες και ως εκ τούτου η ποσοτική ανάλυση κινδύνου δεν είναι σωστή και θα πρέπει να αναθεωρηθεί με νέα δεδομένα.</p>
--	---

Κεφάλαιο 5: Συμπεράσματα και Μελλοντικές Προοπτικές

5.1 Συμπεράσματα

5.1.1 Ανάπτυξη ALTAI Ερωτηματολογίου

Η ηθικότητα των συστημάτων TN με το υπάρχον ALTAI Ερωτηματολόγιο ελέγχεται σε ικανοποιητικό βαθμό αλλά δεν παύει να αποτελεί μια λύση η οποία είναι αρκετά γενικευμένη.

Ένα φιλο-εξελικτικό και προτεινόμενο σενάριο ανάπτυξης του Ερωτηματολογίου ALTAI είναι να πραγματοποιηθεί περαιτέρω ανάπτυξη ανά κλάδο της οικονομίας με βάση την επίδραση που έχει σε αυτόν η Τεχνητή Νοημοσύνη.

Η προσθήκη γνώσης από εξειδικευμένους επιστήμονες μέσω της προώθησης της συνεργασίας με τους προγραμματιστές βοηθάει στον ακριβέστερο έλεγχο της ηθικότητας των συστημάτων TN.

Προτείνεται, η εξειδίκευση και ανάπτυξη του Ερωτηματολογίου ALTAI με στόχο τον εμπλουτισμό και την ακριβέστερη μορφή ερωτήσεων και προτάσεων στους παρακάτω κλάδους - πυλώνες της οικονομίας.

- Εμπόριο
- Διασκέδαση
- Σύστημα Υγείας
- Εκπαίδευση
- Παραγωγή και Κατασκευές
- Ενέργεια
- Προώθηση Προϊόντων
- Οικονομικά και Τραπεζικά
- Εφοδιαστική Αλυσίδα - Ναυτιλιακά – Αυτοκινητοβιομηχανία
- Ασφαλιστικές Υπηρεσίες

- Τεχνολογία – Γενετική Τεχνητή Νοημοσύνη

5.1.2 Τομέας Ερευνητικής Ανάπτυξης: Ναυτιλιακά

Ψηφιακοί Πλοηγοί – ΨΠ

Η εφαρμογή Ψηφιακών Πλοηγών – ΨΠ σε αυτόνομα πλοία επιφέρει διάφορες δεοντολογικές προκλήσεις, όπως η αξιοπιστία της ΤΝ. Ένας αξιόπιστος ΨΠ πρέπει να μπορεί να εξηγηθεί και η διαδικασία λήψης αποφάσεων πρέπει να είναι διαφανής. Επιπλέον, ο ΨΠ πρέπει να ενεργεί δίκαια και υπεύθυνα κατά την πλοήγηση του πλοίου.

Για να διασφαλιστεί ότι ο ΨΠ είναι δίκαιος, προτείνεται παραπάνω ένα μοντέλο επτά επιπέδων για τη μελέτη της δικαιοσύνης σε κάθε βήμα της ανάπτυξής του, από το επίπεδο της έννοιας έως το επίπεδο της υλοποίησης, ώστε να εντοπιστούν και να εξαλειφθούν όλες οι πιθανές προκαταλήψεις.

Αναπτύσσεται ένα πλαίσιο για τη μελέτη της λογοδοσίας και της ευθύνης των πλοηγών. Για να είναι δίκαιοι και υπεύθυνοι, οι αλγόριθμοι μηχανικής μάθησης για τους πλοηγούς που βασίζονται στην ΤΝ, πρέπει να είναι διαφανείς και εξηγήσιμοι.

Συμπερασματικά, το ALTAI Ερωτηματολόγιο μπορεί να εμπλουτιστεί με τον εξής δεοντολογικό έλεγχο:

- Έλεγχος της περιγραφικής διαδικασίας λήψης απόφασης ο οποίος συγκρίνεται και διασταυρώνεται με πραγματικά μη-πολωμένα σενάρια λήψης απόφασης προκειμένου να ελέγχεται πληρέστερα η ηθικότητα του συστήματος ΤΝ.

Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης - ΘΟΕΕ

Τα ηθικά ζητήματα που αφορούν την Νοημοσύνη Ανοιχτού Κώδικα - ΝΑΚ, την Νοημοσύνη Κοινωνικών Μέσων - ΝΚΜ και τα Μεγάλα Δεδομένα είναι ποικίλα και εξελίσσονται συνεχώς. Ο αντίκτυπός τους στην ΘΟΕΕ αφορά τόσο την τεχνολογία, όσο και τις διαδικασίες των χρηστών και το επιχειρηματικό - διοικητικό μοντέλο είναι εμφανή.

Παρ' όλο που υπάρχουν διεθνείς κανονιστικές κατευθυντήριες γραμμές, οι ειδικές άδειες, απαγορεύσεις και εξαιρέσεις προέρχονται κυρίως από την Εθνική νομοθεσία. Ο Γενικός Κανονισμός για την Προστασία Δεδομένων - ΓΚΠΔ^[8], εναρμονίζει εν μέρει τη ρύθμιση για την προστασία των δεδομένων στις χώρες της ΕΕ - Αρχή Προστασίας Δεδομένων – ΑΠΔ

ή Data Protection Authority – DPA - αλλά εξακολουθεί να αφήνει τη δυνατότητα διαφοροποίησης σε Εθνικό επίπεδο. Μια μεγάλη πρόκληση είναι η αντιμετώπιση της Επίδρασης Μωσαϊκού – Mosaic Effect. Η Επίδραση Μωσαϊκού, αποτελεί την επίδραση διαφορετικών κομματιών δεδομένων ή πληροφορίας – παρά το γεγονός ότι από μόνα τους δεν έχουν ιδιαίτερη χρησιμότητα – που γίνονται σημαντικά όταν συνδυάζονται με άλλους τύπους πληροφοριών. Η προστασία των δεδομένων θέτει ισχυρές απαιτήσεις στην τεχνολογία ΘΟΕΕ που χρησιμοποιεί διάφορες πηγές δεδομένων και εκτελεί συγχωνεύσεις δεδομένων σε διάφορα επίπεδα.

Μια άλλη μεγάλη πρόκληση αφορά την εξάρτηση από την αυτοματοποιημένη ανάλυση και το πώς μπορούν να αναπτυχθούν αλγόριθμοι συγχώνευσης δεδομένων που είναι αξιόπιστοι και διαφανείς για τον τελικό χρήστη. Σε συνδυασμό με την ΝΑΚ, τα Μεγάλα Δεδομένα αφορούν τη δυνατότητα χαρτογράφησης συμπεριφορών και τάσεων. Ωστόσο, η επιστήμη των δεδομένων είναι απαραίτητη στην ΝΑΚ λόγω της έλλειψης - χαμηλής ποιότητας των Μεγάλων Δεδομένων, για να βρεθούν οι σωστές απαντήσεις και να καταγραφούν τα σωστά δεδομένα.

Στο πλαίσιο των Μεγάλων Δεδομένων, είναι επίσης αξιοσημείωτο ότι οι τρέχουσες ακαδημαϊκές και δημόσιες συζητήσεις απασχολούνται με την ιδέα της μετατόπισης της έμφασης από τη συλλογή δεδομένων στην ανάλυση δεδομένων και τη χρήση δεδομένων. Υπάρχουν μελετητές που υπογραμμίζουν την ανάγκη για "αλγοριθμική λογοδοσία". Ως εκ τούτου, αναμένεται ότι οι νομικές απαιτήσεις σχετικά με την επεξεργασία δεδομένων προσωπικού χαρακτήρα και τα Μεγάλα Δεδομένα να εξελιχθούν προς αυτή την κατεύθυνση. Ο διαχωρισμός της ηθικής της συλλογής δεδομένων από τη δεοντολογία της επεξεργασίας και της χρήσης των δεδομένων είναι απαραίτητος.

Συμπερασματικά, το ALTAI Ερωτηματολόγιο μπορεί να εμπλουτιστεί με τους παρακάτω ηθικούς ελέγχους:

- Ο βαθμός που η Επίδραση Μωσαϊκού – Mosaic Effect αντιμετωπίζεται στην Θαλάσσια Ολοκληρωμένη Επίγνωση της Επιτήρησης – ΘΟΕΕ.
- Η ηθικότητα των αλγορίθμων συγχώνευσης δεδομένων που είναι αξιόπιστοι και διαφανείς για τον τελικό χρήστη όταν συγκρίνονται με μη-πολωμένους και ίδιας λειτουργικότητας αλγορίθμους.

Θαλάσσια Αυτόνομα Πλοία Επιφάνειας – ΘΑΠΕ

Έχουν εντοπιστεί πολλές προκλήσεις για τους ανθρώπους στην εξ' αποστάσεως λειτουργία των ΘΑΠΕ λόγω της αλλαγής του ρόλου τους σε εποπτικό. Η παρακολούθηση αποτελεί πρόκληση, καθώς η αποτελεσματικότητα της αυτοματοποίησης εξαρτάται από τον βαθμό εξάρτησης του ανθρώπου από αυτήν και το φόρτο εργασίας που υφίσταται ο ανθρώπινος χειριστής. Υπάρχουν επίσης προκλήσεις όσον αφορά την υποστήριξη της επικοινωνίας και της συνεργασίας στο πλαίσιο των Ομάδων Ανθρώπων – Μηχανών - ΟΑΜ.

Η συστηματική ανασκόπηση της βιβλιογραφίας για τις ΟΑΜ συνοψίζεται σε επτά θέματα: Συστήματα Υποστήριξης Απόφασης - ΣΥΑ, Εμπιστοσύνη, Διαφάνεια, Ομάδες Ανθρώπων - Μηχανών, Κατανομή Καθηκόντων - Ρόλων, Υπευθυνότητα και Επίγνωση της Κατάστασης. Είναι σημαντικό να εξεταστεί πώς κάθε ένα από αυτά τα θέματα επηρεάζει τη λήψη αποφάσεων στο πλαίσιο των ΟΑΜ. Πρέπει να εξεταστεί πώς μπορούν να σχεδιαστούν τα ΣΥΑ των ΘΑΠΕ ώστε να βοηθήσουν τους χειριστές να διατηρήσουν την Επίγνωση της Κατάστασης και να υποστηρίξουν την επικοινωνία και την κατανόηση εντός των ΟΑΜ, έτσι ώστε οι χειριστές να είναι σε θέση να λαμβάνουν ορθές και λογικές αποφάσεις.

Συμπερασματικά το ALTAI Ερωτηματολόγιο μπορεί να εμπλουτιστεί ως εξής:

- Έλεγχος του κατά πόσο οι επτά κύριες περιοχές για τα ΘΑΠΕ, δηλαδή τα Συστήματα Υποστήριξης Απόφασης - ΣΥΑ, η Εμπιστοσύνη, η Διαφάνεια, οι Ομάδες Ανθρώπων – Μηχανών, η Κατανομή Καθηκόντων - Ρόλων, η Υπευθυνότητα και η Επίγνωση της Κατάστασης, υλοποιούνται ηθικά επί του πρακτέου μέσω περιγραφικού ελέγχου και σύγκρισης με πραγματικά μη-πολωμένα σενάρια.

5.1.3 Τομέας Ερευνητικής Ανάπτυξης: Γενετική Τεχνητή Νοημοσύνη

Διαχείριση Προϊόντων Λογισμικού – ΔΠΛ

Για την πλήρη αξιοποίηση των πλεονεκτημάτων της TN, είναι απαραίτητο να διερευνηθεί ο τρόπος με τον οποίο η GTN μπορεί να ενσωματωθεί στις προϋπάρχουσες ευέλικτες τεχνικές. Οι ιστορίες χρηστών, τα σχέδια έργων και οι τεχνικές κατανομής πόρων που δημιουργούνται από την TN ενδέχεται να χρειαστεί να τροποποιηθούν έτσι ώστε να είναι συνεπείς με τις έννοιες και τις πρακτικές της ευέλικτης ανάπτυξης.

Οι διαχειριστές προϊόντων λογισμικού μπορούν να επωφεληθούν από εξατομικευμένη υποστηρικτική TN. Καθώς βελτιώνονται τα μοντέλα GTN, μπορούν να προσαρμοστούν ώστε να ανταποκρίνονται στις ειδικές απαιτήσεις συγκεκριμένων διαχειριστών προϊόντων.

Αξιολόγηση των μακροπρόθεσμων επιπτώσεων της GTN στις διαδικασίες διαχείρισης προϊόντων λογισμικού, εστιάζοντας στην παραγωγικότητα, το κόστος και την ποιότητα. Καθώς οι τεχνολογίες GTN εξελίσσονται, είναι πιθανό να αλλάξουν δραστικά την τρέχουσα κατάσταση της Διαχείρισης Προϊόντων Λογισμικού. Οι διαχειριστές προϊόντων λογισμικού μπορούν να αξιοποιήσουν την υπόσχεση της GTN για την προώθηση της δημιουργικότητας, της αποτελεσματικότητας και της επιτυχίας στα έργα τους, παρακολουθώντας τις νεότερες εξελίξεις, υιοθετώντας νέες εφαρμογές και αντιμετωπίζοντας τυχόν περιορισμούς και ηθικά ζητήματα.

Συμπερασματικά, το ALTAI Ερωτηματολόγιο μπορεί να εμπλουτιστεί ως εξής:

- Έλεγχος του κατά πόσο οι ιστορίες χρηστών, τα σχέδια έργων και οι τεχνικές κατανομής πόρων που δημιουργούνται από την TN τροποποιούνται ηθικά έτσι ώστε να είναι συνεπείς με τις έννοιες και τις πρακτικές της ευέλικτης ανάπτυξης.
- Έλεγχος του κατά πόσο μακροπρόθεσμα επηρεάζονται σε ηθικά αποδεκτό βαθμό οι διαδικασίες διαχείρισης προϊόντων λογισμικού, εστιάζοντας στην παραγωγικότητα, το κόστος και την ποιότητα.

Μηχανική Λογισμικού – ΜΛ και Κυβερνο-Φυσικά Συστήματα – ΚΦΣ

Καθώς η Προβλεπτική ΤΝ και ΓΤΝ διεισδύουν σε κάθε τομέα της κοινωνίας μας, από τις τέχνες και τη δημοσιογραφία μέχρι την επιστήμη και την εκπαίδευση, οι ηθικές εκτιμήσεις έχουν καταστεί υψίστης σημασίας. Τα ζητήματα της πνευματικής ιδιοκτησίας, της προκατάληψης, συμπεριλαμβανομένων των λανθασμένων συμβουλών, των παραισθήσεων και της αυθεντικότητας του παραγόμενου περιεχομένου, της πιθανότητας παραπληροφόρησης και της υπονόμησης της ανθρώπινης εργασίας βρίσκονται στο προσκήνιο αυτής της συζήτησης. Έτσι, προκύπτουν σοβαρά ζητήματα στην ΜΛ και στα ΚΦΣ.

Σε συνδυασμό με μοντέλα πρόβλεψης και άλλα εργαλεία και προσεγγίσεις ΤΝ, υπάρχουν τεράστιες δυνατότητες για κοινωνικά θετικές εφαρμογές. Τα σημαντικά κοινωνικά και οικονομικά οφέλη είναι καλά κατανοητά και είναι πιθανό, όπως και με κάθε άλλη πτυχή της ΓΤΝ και Προβλεπτικής ΤΝ, να αυξηθούν δραματικά, οδηγώντας σε μεγαλύτερη αποδοτικότητα, καλύτερες υπηρεσίες και υψηλότερη ποιότητα και εξατομικευμένα αποτελέσματα. Η ηθική θέση, ακόμη και αν ήταν εφικτή, δεν μπορεί επομένως να επιδιώκει την εγκατάλειψη της χρήσης της ΓΤΝ και άλλων προσεγγίσεων που βασίζονται στην ΤΝ. Από την παροχή βοήθειας στους ερευνητές στη δημιουργία καινοτόμων λύσεων έως την προώθηση της απaráμιλλης δημιουργικότητας, αυτή η τεχνολογία αποτελεί καταλύτη για πρωτοφανή επιτεύγματα.

Τα αποτελέσματα χρήσης πρέπει να υπόκεινται σε προσεκτική επίβλεψη, δοκιμή, επικύρωση και επαλήθευση, εκφραζόμενα και μέσω επαγγελματικών προτύπων και κυβερνητικών συστάσεων και πολιτικών κανονισμών. Το επίπεδο παρέμβασης ποικίλλει ανάλογα με τον τομέα και το πρόβλημα, με τον χρήστη και την προβλεπόμενη χρήση των αποτελεσμάτων. Μια σημαντική δικλείδα ασφαλείας είναι η χρήση κατάλληλων και αντιπροσωπευτικών συνόλων δεδομένων εκπαίδευσης.

Για παράδειγμα, κάποια περιστασιακή ή δημιουργική χρήση από ενήλικες για ατομική ή μικρή ομαδική κατανάλωση μπορεί να χρειάζεται ελάχιστη εποπτεία. Από την άλλη πλευρά, η χρήση σε διαδικτυακές εφαρμογές, στο σχεδιασμό ενός κρίσιμου για την ασφάλεια κυβερνοφυσικού αεροδιαστημικού συστήματος μπορεί να απαιτεί έντονη και επαναλαμβανόμενη επανεξέταση, επικύρωση και αξιολόγηση.

Συμπερασματικά το ALTAI Ερωτηματολόγιο μπορεί να εμπλουτιστεί ως εξής:

- Ο τύπος της εφαρμογής, δηλαδή κατά πόσο το σύστημα ΜΛ και ΚΦΣ χρησιμοποιείται ιδιωτικά ή μη και ο ηθικός κοινωνικός βαθμός επιρροής του.
- Το κατά πόσο τηρείται η ασφαλιστική δικλίδα ανά εξειδικευμένο τύπο συστήματος ΤΝ κατάλληλων και αντιπροσωπευτικών συνόλων δεδομένων εκπαίδευσης.

Διαδίκτυο των Πραγμάτων – ΔΤΠ

Η σύγκλιση της ΓΤΝ με τα συστήματα ΔΤΠ αντιπροσωπεύει ένα σημαντικό βήμα για την αξιοποίηση των τεράστιων δυνατοτήτων των δεδομένων που παράγονται από έξυπνες συσκευές. Αξιοποιώντας τις ικανότητες της ΓΤΝ, υπάρχει η δυνατότητα να αποκαλυφθούν πολύτιμες γνώσεις, να προωθηθεί η καινοτομία, να ενισχυθεί η ασφάλεια και να δημιουργηθούν προσαρμοστικά συστήματα ΤΝ, ωθώντας το τοπίο του ΔΤΠ με αποτέλεσμα την αυξημένη απόδοση και παραγωγικότητα. Αυτός ο μετασχηματισμός οδηγεί σε ένα τοπίο απεριόριστων δυνατοτήτων για τον τομέα του ΔΤΠ, ανοίγοντας το δρόμο για έναν πιο έξυπνο και διασυνδεδεμένο κόσμο.

Είναι απαραίτητο να συμβαδίζουν οι εξελίξεις στην ΓΤΝ και στο ΔΤΠ. Είναι κρίσιμη η εξοικείωση με τις προσπάθειες που καθοδηγούνται από οργανισμούς και ιδρύματα που ερευνούν την ΓΤΝ και το ΔΤΠ, όπως η Διεθνής Ένωση Τηλεπικοινωνιών – ΔΕΤ ή International Telecommunication Union – ITU, που διερευνά τους κινδύνους της ΓΤΝ σε έξυπνες πόλεις καθώς και η Εκπαιδευτική Επιστημονική και Πολιτιστική Οργάνωση των Ηνωμένων Εθνών – ΕΕΠΟΗΕ ή United Nations Educational Scientific and Cultural Organization - UNESCO που δημοσίευσε πρόσφατα ένα έγγραφο σχετικά με τις κατευθυντήριες γραμμές για την ΓΤΝ στην Εκπαίδευση και την Έρευνα. Η καλλιέργεια της ευαισθητοποίησης και η κατανόηση του μελλοντικού αντίκτυπου της ΓΤΝ και του ΔΤΠ στις επιχειρήσεις, τα άτομα και την κοινωνία εξαρτάται από τον τρόπο αντιμετώπισης των κινδύνων που παρουσιάζει.

Χρειάζονται κανόνες για τη συλλογή, την αποθήκευση και την επεξεργασία δεδομένων, συμβάλλοντας στον μετριασμό πιθανών παραβιάσεων του απορρήτου και τρωτών σημείων ασφαλείας συστημάτων ΓΤΝ και ΔΤΠ. Επιπλέον, πρέπει να αναπτυχθούν οδηγίες για υπεύθυνη ανάπτυξη μοντέλων που δίνουν προτεραιότητα στην ευημερία των χρηστών και αποτρέπουν τις αρνητικές συνέπειες. Επίσης, η εξισορρόπηση της

αυτοματοποίησης και της ανθρώπινης συμμετοχής είναι σημαντική για την αξιοποίηση του πλήρους δυναμικού της ΓΤΝ, ενώ ταυτόχρονα μετριάζονται οι πιθανές αρνητικές συνέπειες.

Η ευαισθητοποίηση και η εκπαίδευση των χρηστών σχετικά με το περιεχόμενο που δημιουργείται από την ΓΤΝ και το ΔΤΠ είναι ζωτικής σημασίας για την προώθηση υπεύθυνων αλληλεπιδράσεων ΓΤΝ σε περιβάλλοντα ΔΤΠ. Οι χρήστες πρέπει να ενημερώνονται για να μπορούν να λαμβάνουν αποφάσεις σχετικά με τις αλληλεπιδράσεις τους με συστήματα που βασίζονται στη διαφάνεια, την πλήρη κατανόηση των δυνατοτήτων, των πλεονεκτημάτων και των πιθανών κινδύνων του περιεχομένου που δημιουργείται από ΓΤΝ στα περιβάλλοντα ΔΤΠ.

Συμπερασματικά, το ALTAI Ερωτηματολόγιο μπορεί να εμπλουτιστεί ως εξής:

- Ο βαθμός που οι κανόνες για τη συλλογή, την αποθήκευση και την επεξεργασία δεδομένων, συμβάλλοντας στον μετριασμό πιθανών παραβιάσεων του απορρήτου και τρωτών σημείων ασφαλείας, πετυχαίνουν τον ηθικό τους στόχο.
- Έλεγχος του βαθμού ηθικής εξισορρόπησης της αυτοματοποίησης και της ανθρώπινης συμμετοχής καθώς αποτελεί σημείο κλειδί για την αξιοποίηση του πλήρους δυναμικού της ΓΤΝ.

5.1.4 Τομέας Ερευνητικής Ανάπτυξης: Αυτοκινητοβιομηχανία

Ταυτόχρονος Εντοπισμός και Χαρτογράφηση – TEX (0 – 3)

Παρουσιάζεται παραπάνω μια ολοκληρωμένη ανάλυση του ρόλου της ΤΝ και των αλγορίθμων εκμάθησης στα ΑΟ, διεισδύοντας σε διάφορες πτυχές, όπως οι μετατοπίσεις από συστήματα βασισμένα σε κανόνες σε συστήματα ΤΝ με Βαθιά Νευρωνικά Δίκτυα – ΒΝΔ λόγω βελτιωμένων δυνατοτήτων μοντέλων και υπολογιστικής ισχύος.

Οι ανάγκες συστημάτων Ταυτόχρονου Εντοπισμού και Χαρτογράφησης – TEX (0 – 3) για φορητά συγκριτικά με εκείνες των αυτοκινήτων είναι διαφορετικές καθώς τα φορητά δίνουν προτεραιότητα στις μεταφορές από κόμβο σε κόμβο και τα αποτελεσματικά ταξίδια μεγάλων αποστάσεων, ενώ τα επιβατικά αυτοκίνητα στοχεύουν στην αυτονομία από την πηγή προς τον προορισμό. Υπάρχει αυξανόμενη πολυπλοκότητα από την ανίχνευση

βασικού αντικειμένου στην τρισδιάστατη χαρτογράφηση και την πρόβλεψη προσαρμοστικής συμπεριφοράς.

Υφίστανται προκλήσεις και επιπτώσεις, όπως περιορισμένη ισχύς αποθήκευσης και επεξεργασίας, ανησυχίες για ενημέρωση λογισμικού και αυξημένες τρωτότητες ασφαλείας σε υψηλότερα επίπεδα αυτονομίας. Οι προηγμένες τεχνικές όπως η Βαθιά Μάθηση και Ενισχυτική Μάθηση είναι απαραίτητες για υψηλότερα επίπεδα με περίπλοκη διαδικασία λήψης αποφάσεων και προσαρμοστική συμπεριφορά.

Έλεγχος και Δημιουργία Πορείας – ΕΔΠ (4 - 5)

Οι εφαρμογές Ελέγχου και Δημιουργίας Πορείας – ΕΔΠ (4 – 5) για φορτηγά και αυτοκίνητα έχουν διαφορετικές απαιτήσεις. Τα φορτηγά επικεντρώνονται στη βελτιστοποίηση της διαδρομής και την απόδοση καυσίμου, ενώ τα επιβατικά αυτοκίνητα δίνουν προτεραιότητα στην άνεση των επιβατών και στη δυναμική προσαρμογή στα αστικά περιβάλλοντα. Το μέγεθος του πακέτου λογισμικού αυξάνεται με το επίπεδο αυτονομίας. Αυτό δημιουργεί προκλήσεις για την αποθήκευση, την επεξεργασία και τις ενημερώσεις λογισμικού, υπογραμμίζοντας την ανάγκη για αποτελεσματικές αρχιτεκτονικές και ισχυρά μέτρα ασφαλείας. Απαιτείται περαιτέρω έρευνα για τα αυτόνομα φορτηγά παρά τις υποσχόμενες επιχειρηματικές δυνατότητές τους. Αυτός ο τομέας υστερεί σε σχέση με την έρευνα επιβατικών αυτοκινήτων, αλλά η ανάπτυξη του μπορεί να βελτιστοποιήσει σημαντικά την εφοδιαστική αλυσίδα και να αντιμετωπίσει τις ελλείψεις οδηγών.

Ορισμένες πτυχές δεν περιλαμβάνονται στο πεδίο αυτής της εργασίας, συμπεριλαμβανομένων των αναδυόμενων τεχνολογιών και αλγορίθμων τεχνητής νοημοσύνης όπως η Κβαντική ΤΝ, η Μεταφορική Μάθηση και η Μέτα-μάθηση. Επί του παρόντος, αυτές οι εφαρμογές στη ρομποτική και τα φυσικά συστήματα, όπως τα ΑΟ, είναι περιορισμένες.

Διασαφηνίζεται η εικόνα του εξελισσόμενου τοπίου της ΤΝ στα ΑΟ, τονίζοντας τον κρίσιμο ρόλο της σε αποτελεσματικές και ασφαλείς λύσεις μεταφοράς. Προσδιορίζονται βασικές προκλήσεις και προτείνονται τομείς για μελλοντική έρευνα, συμβάλλοντας σε έναν οδικό χάρτη για ερευνητές, επαγγελματίες και λάτρεις που ενδιαφέρονται για τη δυναμική σχέση μεταξύ της ΤΝ, των αλγορίθμων μάθησης και της πρώτης γραμμής των σύγχρονων μεταφορών.

Συμπερασματικά, το ALTAI Ερωτηματολόγιο μπορεί να εμπλουτιστεί διαφοροποιώντας και διαχωρίζοντας τις ερωτήσεις στον τομέα της αυτοκινητοβιομηχανίας με βάση τον τύπο του ΑΟ. Δηλαδή, αν πρόκειται για συστήματα TN σε επιβατικά ΑΟ ή σε φορτηγά ΑΟ.

- Στο ερωτηματολόγιο για επιβατικά ΑΟ, όσα είναι εξοπλισμένα με συστήματα Ταυτόχρονου Εντοπισμού και Χαρτογράφησης – TEX (0 – 3) που στοχεύουν στην αυτονομία από την πηγή προς τον προορισμό, υφίσταται αυξανόμενη πολυπλοκότητα από την ανίχνευση βασικού αντικειμένου στην τρισδιάστατη χαρτογράφηση και την πρόβλεψη προσαρμοστικής συμπεριφοράς. Έτσι, συνίσταται να ελέγχεται ο βαθμός που πετυχαίνονται δεοντολογικά οι παραπάνω τεχνικοί στόχοι.
- Στο ερωτηματολόγιο για επιβατικά ΑΟ, με συστήματα Ελέγχου και Δημιουργία Πορείας – ΕΔΠ (4 – 5), οι ερωτήσεις πρέπει να επικεντρώνονται στην άνεση των επιβατών και στο κατά πόσο πετυχαίνεται η ηθική δυναμική προσαρμογή στα αστικά περιβάλλοντα.
- Στο ερωτηματολόγιο για φορτηγά ΑΟ, τα συστήματα Ταυτόχρονου Εντοπισμού και Χαρτογράφησης – TEX (0 - 3) για φορτηγά ΑΟ εστιάζουν στις μεταφορές από κόμβο σε κόμβο και τα αποτελεσματικά ταξίδια μεγάλων αποστάσεων. Οπότε, πρέπει να ελέγχεται η ηθική τεχνική υλοποίησή τους.
- Στο ερωτηματολόγιο για φορτηγά ΑΟ που είναι εξοπλισμένα με συστήματα Ελέγχου και Δημιουργία Πορείας – ΕΔΠ (4 – 5), οι ερωτήσεις πρέπει να επικεντρώνονται στην ηθική βελτιστοποίηση της διαδρομής και την δεοντολογική απόδοση καυσίμων.

5.2 Προτάσεις για περαιτέρω έρευνα

5.2.1 Θεσμοθέτηση Ηθικής Τεχνητής Νοημοσύνης

Η οικοδόμηση ενός ισχυρού πλαισίου για την ΤΝ περιλαμβάνει διάφορα βασικά στοιχεία, καθένα από τα οποία συμβάλλει σε μια ολιστική προσέγγιση:

Στιβαρά Μοντέλα Διακυβέρνησης

Η ισχυρή διακυβέρνηση είναι απαραίτητη, όπως αναδεικνύεται από εταιρικά μοντέλα διακυβέρνησης, τα οποία παρέχουν ένα πρότυπο για την ηθική συμπεριφορά. Μια τέτοια προσέγγιση παρέχει καθοδήγηση για προγραμματιστές, μηχανικούς, διευθυντές, εμπόρους και συνεργάτες. Τα μοντέλα λειτουργούν ως μικρόκοσμος για τον ευρύτερο κλάδο και γενικότερα θέτοντας ένα πρότυπο που οι υπόλοιποι μπορούν να υιοθετήσουν.

Ηθικά Εργαλεία Τεχνολογίας

Εργαλεία που επιβάλλουν και υποστηρίζουν την ηθική ΤΝ είναι ύψιστης σημασίας. Οι κάρτες μοντέλων για παράδειγμα, περικλείουν το ηθικό προφίλ των μοντέλων ΤΝ, χρησιμεύοντας ως διαφανή μέσα καταγραφής του σκοπού, των δυνατοτήτων και των περιορισμών τους, συμβάλλοντας έτσι στην διατήρηση μιας ηθικής στάσης.

Ρυθμιστική Ηγεσία

Η επιρροή της νομοθεσίας είναι μεγάλο κατόρθωμα, ωστόσο οι ηγέτες του κλάδου έχουν αναλάβει αυτόν τον ρόλο, συμβάλλοντας στην διαμόρφωση μιας παγκόσμιας συζήτησης που επιδιώκει να καθιερώσει συνεπή εφαρμόσιμα πρότυπα για ηθικές πρακτικές ΤΝ.

5.2.2 Όραμα

Τεχνητή Νοημοσύνη με Επίκεντρο τον Ανθρώπο

Το μέλλον της Τεχνητής Νοημοσύνης είναι στην ουσία ανθρωποκεντρικό. Αυτή η αντίληψη προβλέπει την Τεχνητή Νοημοσύνη όχι ως αντικαταστάτη της ανθρώπινης εφευρετικότητας αλλά ως μέσο αύξησής της, βελτιώνοντας τις ανθρώπινες εμπειρίες και αλληλεπιδράσεις με βαθείς και ποικίλους τρόπους. Εστιάζει σε τεχνολογίες που βελτιώνουν ουσιαστικά την ποιότητα ζωής. Η ανθρωποκεντρική προσέγγιση διασφαλίζει

ότι το κύμα της Τεχνητής Νοημοσύνης ανυψώνει όλες τις βάρκες, ενισχύοντας την ανθρώπινη ικανότητα χωρίς να εκτοπίζει την ανθρώπινη αξία.

Καθώς το τοπίο της Τεχνητής Νοημοσύνης επεκτείνεται σε νέα εδάφη αυτονομίας και επιρροής, μας ενθαρρύνει να περιηγηθούμε σε αυτό το έδαφος με ηθική πυξίδα, διασφαλίζοντας ότι κάθε βήμα που κάνουμε είναι σε αρμονία με τις αρχές της ισότητας και του σεβασμού της ανθρώπινης αξιοπρέπειας. Με την καλλιέργεια μιας κουλτούρας ισχυρής διακυβέρνησης και ρυθμιστικής πρόβλεψης. Η συλλογική επιδίωξη είναι σαφής για κάθε οργανισμό: ευθυγράμμιση της μετασχηματιστικής δύναμης της Τεχνητής Νοημοσύνης με τον κοινωνικό ιστό, διασφάλιση της ιδιωτικής ζωής, καλλιέργεια της ποικιλομορφίας και εκδημοκρατισμός των πλεονεκτημάτων της Τεχνητής Νοημοσύνης για όλους.

Βιβλιογραφία

- [1] European Commission, Joint Research Centre, Samoili, S., López Cobo, M., Delipetrev, B. et al., AI watch, defining artificial intelligence 2.0 – Towards an operational definition and taxonomy for the AI landscape, Publications Office of the European Union, 2021, <https://data.europa.eu/doi/10.2760/019901>
- [2] European Commission, Directorate-General for Communications Networks, Content and Technology, Sectoral Considerations on the Policy and Investment Recommendations for Trustworthy Artificial Intelligence, Publications Office, 2020, <https://data.europa.eu/doi/10.2759/733662>
- [3] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21). Association for Computing Machinery, New York, NY, USA, 624–635. <https://doi.org/10.1145/3442188.3445923>
- [4] ETHICS GUIDELINES FOR TRUSTWORTHY AI, High-Level Expert Group on Artificial Intelligence, 2019, https://www.europarl.europa.eu/cmsdata/196377/AI%20HLEG_Ethics%20Guidelines%20for%20Trustworthy%20AI.pdf
- [5] Pin-Yu Chen, Cho-Jui Hsieh, Chapter 12 - Adversarial training, Editor(s): Pin-Yu Chen, Cho-Jui Hsieh, Adversarial Robustness for Machine Learning, Academic Press, 2023, Pages 119-125, ISBN 9780128240205, <https://doi.org/10.1016/B978-0-12-824020-5.00023-5>.

[6] Transfer learning: a friendly introduction, Journal of Big Data, 2022,
<https://doi.org/10.1186/s40537-022-00652-w>

[7] The Charter of Fundamental Rights of the European Union (CFREU) is the EU's bill of rights, European Union Agency for Fundamental Rights, 2022
https://fra.europa.eu/sites/default/files/fra_uploads/fra-2022-charter-case-studies-trainers-manual_en.pdf

[8] Regulation of the European Parliament and of the Council, Protection of natural persons with regard to the processing of personal data and on the free movement of such data, General Data Protection Regulation, 2016,
<http://data.europa.eu/eli/reg/2016/679/oj>

[9] Council Directive 85/374/EEC of 25 July 1985 on the approximation of the laws, regulations and administrative provisions of the Member States concerning liability for defective products, <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=celex%3A31985L0374>

[10] Regulation (EU) 2018/1807 of the European Parliament and of the Council on a framework for the free flow of non-personal data in the European Union, 2018, <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32018R1807>

[11] Council Directive implementing the principle of equal treatment between persons irrespective of racial or ethnic origin, European Commission, Non-discrimination, 2000, <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:32000L0043:en:HTML>

[12] Directive of the European Parliament and of the Council on consumer rights, amending Council Directive 93/13/EEC and Directive 1999/44/EC of the European

Parliament and of the Council and repealing Council Directive 85/577/EEC and Directive 97/7/EC of the European Parliament and of the Council Text with EEA relevance, 2011, <http://data.europa.eu/eli/dir/2011/83/oj>

[13] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of Regions on the practical implementation of the provisions of the Health and Safety at Work Directives 89/391 (Framework), 89/654 (Workplaces), 89/655 (Work Equipment), 89/656 (Personal Protective Equipment), 90/269 (Manual Handling of Loads) and 90/270 (Display Screen Equipment), <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52004DC0062>

[14] Center for Security and Emerging Technology, CSET, 2023,

What Does AI Red-Teaming Actually Mean? <https://cset.georgetown.edu/article/what-does-ai-red-teaming-actually-mean/>

[15] The AI Act is the first-ever legal framework on AI, which addresses the risks of AI and positions Europe to play a leading role globally, <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

[16] Namazi, Hosna & Perera, Lokukaluge. (2023). Trustworthiness Evaluation Framework for Digital Ship Navigators in Bridge Simulator Environments. 10.1115/OMAE2023-104863. https://www.researchgate.net/profile/Lokukaluge-Perera/publication/369453764_Trustworthiness_Evaluation_Framework_for_Digital_Ship_Navigators_in_Bridge_Simulator_Environments/links/648e32bc95bbbe0c6ecfadde/Trustworthiness-Evaluation-Framework-for-Digital-Ship-Navigators-in-Bridge-Simulator-Environments.pdf

[17] Mosaic effect, <https://centre.humdata.org/glossary-2/mosaic-effect/>

[18] IPR Legislation, https://commission.europa.eu/business-economy-euro/doing-business-eu/intellectual-property-rights_en

[19] Rajamäki, J. ; Sarlio-Siintola, S. ; Simola, J. (2018) Ethics of Open Source Intelligence Applied by Maritime Law Enforcement Authorities. In Audun Josang (Ed.) Proceedings of the 17th European Conference on Cyber Warfare and Security ECCWS 2018, 28-29 June 2018, Oslo, Norway. Academic Conferences and Publishing International Limited, 424-431,
https://www.academia.edu/72590203/Ethics_of_Open_Source_Intelligence_Applied_by_Maritime_Law_Enforcement_Authorities

[20] Kirsty M. Lynch, Victoria A. Banks, Aaron P. J. Roberts, Stewart Radcliffe & Katherine L. Plant (2024) What factors may influence decision-making in the operation of Maritime autonomous surface ships? A systematic review, *Theoretical Issues in Ergonomics Science*, 25:1, 98-142, DOI: 10.1080/1463922X.2022.2152900, <https://doi.org/10.1080/1463922X.2022.2152900>

[21] Gupta, S. ., Sharma, P. ., Chaudhary, S. ., Kumar, V. ., Singh, S. P. ., Lourens, M. ., & Beri, N. . (2023). Study on the Beneficial Impacts and Ethical Dimensions of Generative AI in Software Product Management. *International Journal of Intelligent Systems and Applications in Engineering*, 12(8s), 251–264. Retrieved from <https://ijisae.org/index.php/IJISAE/article/view/4116>

[22] Sprint and Scrum Planning, <https://www.theserverside.com/blog/Coffee-Talk-Java-News-Stories-and-Opinions/Sprint-vs-Scrum-Whats-the-difference>

[23] KIROVA, Vassilka D., et al. "The Ethics of Artificial Intelligence in the Era of Generative AI." *Journal of Systemics, Cybernetics and Informatics* 21.4 (2023): 42-50., <https://www.iiisci.org/Journal/PDV/sci/pdfs/SA445KR23.pdf>

[24] UNESCO, <https://www.un.org/youthenvoy/2013/08/unesco-united-nations-educational-scientific-and-cultural-organization/>

[25] IEEE, <https://www.ieee.org/>

[26] IBM, <https://www.ibm.com/us-en>

[27] Belmont Report, <https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html>

[28] Chimbga, B. (2023). Exploring the Ethical and Societal Concerns of Generative AI in Internet of Things (IoT) Environments. In: Pillay, A., Jembere, E., J. Gerber, A. (eds) *Artificial Intelligence Research. SACAIR 2023. Communications in Computer and Information Science*, vol 1976. Springer, Cham. <https://rdcu.be/dElTM>

[29] Kumar, A. (2024). Exploring Ethical Considerations in AI-driven Autonomous Vehicles: Balancing Safety and Privacy. *Journal of Artificial Intelligence General Science (JAIGS)* ISSN:3006-4023, 2(1), 125-138. <https://doi.org/10.60087/jaigs.v2i1.p138>

[30] SAE International, <https://www.sae.org/>

[31] Shetiya, Sneha Sudhir, and Divya Garikapati. "Autonomous Vehicles: Evolution of Artificial Intelligence and Learning Algorithms."
<https://doi.org/10.48550/arXiv.2402.17690> (2024).

[32] Trolley problem, https://en.wikipedia.org/wiki/Trolley_problem

[33] Software-in-the-Loop, <https://www.opal-rt.com/software-in-the-loop/>

[34] Hardware-in-the-Loop, <https://www.ni.com/en/solutions/transportation/hardware-in-the-loop/what-is-hardware-in-the-loop-.html>

[35] Human-in-the-Loop, <https://www.telusinternational.com/glossary/human-in-the-loop>

[36] YOLOv3, <https://viso.ai/deep-learning/yolov3-overview/>

[37] K-means clustering, <https://deepai.org/machine-learning-glossary-and-terms/k-means>

[38] RCNN, <https://towardsdatascience.com/object-detection-explained-r-cnn-a6c813937a76>

[39] OrientNet/Predictor-Net, <https://www.automotivemastermind.com/blog/predictive-analytics/future-of-predictive-analytics-for-dealerships/>

[40] RPN, <http://www.applied-statistics.org/fmea-automotive.html>

[41] INS, <https://www.sciencedirect.com/topics/engineering/inertial-navigation-system>

[42] GNSS, <https://novatel.com/tech-talk/an-introduction-to-gnss/what-are-global-navigation-satellite-systems-gnss>

[43] 3D-SLAM, <https://www.flyability.com/simultaneous-localization-and-mapping>

[44] TARA, <https://www.cyres-consulting.com/automotive-cybersecurity-online-courses/threat-analysis-and-risk-assessment-in-automotive-cyber-security/>

[45] Grey modelling, <https://www.sciencedirect.com/topics/computer-science/gray-model>

[46] Andreia Martinho, Nils Herber, Maarten Kroesen & Caspar Chorus (2021) Ethical issues in focus by the autonomous vehicles industry, *Transport Reviews*, 41:5, 556-577, <https://doi.org/10.1080/01441647.2020.1862355>

[47] Human Meaningful Control theory, <https://mwi.westpoint.edu/what-is-meaningful-human-control-anyway-cracking-the-code-on-autonomous-weapons-and-human-judgment/>

[48] The Assessment List for Trustworthy Artificial Intelligence, <https://altai.insight-centre.org/>