



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Τίτλος Πτυχιακής Εργασίας	Η Εξέλιξη της Τεχνητής Νοημοσύνης από τη Μηχανική Μάθηση στην Παραγωγική Τεχνητή Νοημοσύνη The Evolution of Artificial Intelligence from Machine Learning to Generative AI
Ονοματεπώνυμο Φοιτητή	Φώτιος Τσάβαλος
Πατρώνυμο	Παναγιώτης
Αριθμός Μητρώου	Π/06140
Επιβλέπων	Γεώργιος Τσιχριντζής, Καθηγητής

Ημερομηνία Παράδοσης / Σεπτέμβριος 2026

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν αποκλειστικά τον συγγραφέα και δεν αντιπροσωπεύουν τις επίσημες θέσεις του Πανεπιστημίου Πειραιώς.

Ως συγγραφέας της παρούσας εργασίας δηλώνω πως η παρούσα εργασία δεν αποτελεί προϊόν λογοκλοπής και δεν περιέχει υλικό από μη αναφερόμενες πηγές.

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον επιβλέποντα καθηγητή μου για την καθοδήγηση, τις πολύτιμες παρατηρήσεις και τη συνεχή υποστήριξή του καθ' όλη τη διάρκεια εκπόνησης της παρούσας πτυχιακής εργασίας. Επιπλέον, ευχαριστώ την οικογένεια και τους φίλους μου για την ενθάρρυνση και την υπομονή τους.

Περίληψη

Η Τεχνητή Νοημοσύνη (TN) αποτελεί έναν από τους πλέον δυναμικούς και ταχέως εξελισσόμενους επιστημονικούς τομείς της σύγχρονης εποχής. Η μετάβαση από τη Μηχανική Μάθηση (Machine Learning) και τη Βαθιά Μάθηση (Deep Learning) προς τη Παραγωγική Τεχνητή Νοημοσύνη (generative AI) έχει επιφέρει σημαντικές αλλαγές στον τρόπο με τον οποίο τα υπολογιστικά συστήματα αλληλεπιδρούν με τον άνθρωπο, λαμβάνουν αποφάσεις και δημιουργούν περιεχόμενο.

Η παρούσα εργασία εξετάζει την εξέλιξη της TN δίνοντας έμφαση σε κρίσιμες έννοιες όπως η εμπιστοσύνη (trust), η επεξηγησιμότητα (explainability), η ανθρωποκεντρική σχεδίαση και η αλληλεπίδραση ανθρώπου-μηχανής. Παρουσιάζονται εφαρμογές της TN σε τομείς όπως η εκπαίδευση, οι έξυπνες πόλεις, η ενέργεια και η υγεία, ενώ αναλύονται οι προκλήσεις που σχετίζονται με epistemic risks και hallucinations στα μεγάλα γλωσσικά μοντέλα.

Τέλος, γίνεται συγκριτική ανάλυση των πλεονεκτημάτων και περιορισμών της TN, καθώς και των ηθικών και κοινωνικών επιπτώσεών της, με στόχο την ανάδειξη μελλοντικών κατευθύνσεων έρευνας.

Λέξεις Κλειδιά: Τεχνητή Νοημοσύνη, Παραγωγική Τεχνητή Νοημοσύνη (generative AI), Μηχανική Μάθηση (Machine Learning), Εξηγήσιμη Τεχνητή Νοημοσύνη (Explainable AI)

Abstract

Artificial Intelligence (AI) is one of the most rapidly evolving scientific fields. The transition from Machine Learning and Deep Learning to Generative AI has significantly transformed human-machine interaction, decision-making processes, and content generation.

This thesis explores the evolution of AI, focusing on trust, explainability, human-centered design, and human-AI interaction. Applications in education, smart cities, energy, and healthcare are analyzed, along with challenges such as epistemic risks and hallucinations in large language models.

Finally, the study discusses advantages, limitations, and ethical implications, highlighting future research directions.

Keywords: Artificial Intelligence, Generative AI, Trust, Explainability

Περιεχόμενα

Copyright ©	2
Ευχαριστίες	3
Περίληψη	4
Abstract	4
1. Εισαγωγή	8
1.1 Τεχνητή Νοημοσύνη	8
1.2 Σημασία	8
1.3 Από τη Μηχανική Μάθηση στη Παραγωγική Τεχνητή Νοημοσύνη (generative AI)	8
2. Εμπιστοσύνη, Αυτονομία και Ανθρωποκεντρική Τεχνητή Νοημοσύνη	9
2.1 Η έννοια της εμπιστοσύνης στην Τεχνητή Νοημοσύνη	9
2.2 Ανθρωποκεντρική ΤΝ και Human-AI Interaction	9
2.3 Το μοντέλο εμπιστοσύνης VIRTSL	10
2.4 Αυτονομία και Ηθική στην ΤΝ	11
2.5 Επιστημικοί Κίνδυνοι (epistemic risks) και Ψευδαισθήσεις (hallucinations)	11
2.6 Επεξηγησιμότητα και Επεξηγήσιμη Τεχνητή Νοημοσύνη (Explainable AI)	13
2.7 Συμπεράσματα Κεφαλαίου	14
3. Παραγωγική Τεχνητή Νοημοσύνη (generative AI) και Ευφυή Συστήματα στην Εκπαίδευση και τον Πολιτισμό	15
3.1 Η εμφάνιση της Παραγωγική Τεχνητή Νοημοσύνη (generative AI) στην εκπαίδευση	15
3.2 Παραγωγική Τεχνητή Νοημοσύνη (generative AI) και ανάπτυξη δεξιοτήτων του 21ου αιώνα	15
3.3 Ψευδαισθήσεις (hallucinations) και Επιστημικοί Κίνδυνοι (epistemic risks) στην εκπαίδευση	16
3.4 Πολιτιστικές εφαρμογές της ΤΝ	17
3.5 Παιχνίδια, ευφυή περιβάλλοντα και εκπαιδευτική ΤΝ	17
3.6 E-learning και εξατομικευμένη αξιολόγηση	18

3.7 Συμπεράσματα Κεφαλαίου	19
4. Τεχνητή Νοημοσύνη σε Έξυπνες Πόλεις, Ενέργεια και Αστικά Οικοσυστήματα Απόφασης	20
4.1 Η μετάβαση προς τις έξυπνες πόλεις.....	20
4.2 Αλληλεπίδραση Ανθρώπου-Έξυπνης Πόλης και ανθρωποκεντρικός σχεδιασμός	20
4.3 Προσωπικά δεδομένα, GDPR και αλγοριθμική διαφάνεια	21
4.4 TN και ενεργειακά συστήματα	22
4.5 Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) και αξιοπιστία ενεργειακών προβλέψεων	23
4.6 Συνθετικά Δεδομένα (Synthetic Data) και Παραγωγικά Αντιπαραθετικά Δίκτυα (Generative Adversarial Networks – GANs) σε ενεργειακά οικοσυστήματα.....	23
4.7 Συμπεράσματα Κεφαλαίου	24
5. Τεχνητή Νοημοσύνη στην Υγεία, τη Βιοϊατρική και τα Εξατομικευμένα Συστήματα	25
5.1 Η εξέλιξη της TN στον χώρο της υγείας	25
5.2 Μεγάλα Γλωσσικά Μοντέλα (Large Language Models – LLMs) και ιατρική διάγνωση	26
5.3 Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) και αξιοπιστία ιατρικών συστημάτων	26
5.4 Εξατομικευμένη διατροφή και Προγνωστική Αναλυτική (Predictive Analytics)	27
5.5 Ανάλυση Ακραίων Τιμών (Extreme Value Analysis – EVA) και ανάλυση βιοϊατρικών δεδομένων	28
5.6 Microservices και αξιόπιστη Ιατρική TN	28
5.7 Συμπεράσματα Κεφαλαίου	29
6. Τεχνικές και Αλγοριθμικές Βάσεις της Τεχνητής Νοημοσύνης	30
6.1 Η εξέλιξη των αλγοριθμικών προσεγγίσεων στην TN.....	30
6.2 Μηχανική Μάθηση (Machine Learning) και Βαθιά Μάθηση (Deep Learning – DL).....	31
6.3 Ενισχυτική Μάθηση (Reinforcement Learning – RL) και αυτόνομα συστήματα	32
6.4 Πολυτροπική TN και ενοποίηση δεδομένων	33

6.5 Συναισθηματική Υπολογιστική και αναγνώριση συναισθήματος.....	34
6.6 Ευφυή πολυμεσικά συστήματα και Περιβάλλοντα Συστάσεων (Recommendation Environments).....	35
6.7 Συμπεράσματα Κεφαλαίου	36
7. Συζήτηση.....	36
7.1 Η μετάβαση από τη Μηχανική Μάθηση στη Generative AI.....	36
7.2 Συγκριτική ανάλυση εφαρμογών της TN.....	37
7.3 Εμπιστοσύνη και ανθρώπινος παράγοντας	38
7.4 Εξηγησιμότητα και το πρόβλημα των black-box models.....	39
7.5 Ηθικές και κοινωνικές προεκτάσεις της TN	39
7.6 Μελλοντικές κατευθύνσεις της TN	40
8. Συμπεράσματα.....	41
Βιβλιογραφία.....	44

1. Εισαγωγή

1.1 Τεχνητή Νοημοσύνη

Η Τεχνητή Νοημοσύνη (TN) αποτελεί έναν διεπιστημονικό τομέα που συνδυάζει επιστήμες όπως η πληροφορική, τα μαθηματικά, η γνωσιακή επιστήμη και η μηχανική, με στόχο την ανάπτυξη συστημάτων ικανών να εκτελούν εργασίες που απαιτούν ανθρώπινη νοημοσύνη. Οι εργασίες αυτές περιλαμβάνουν τη μάθηση, τη λήψη αποφάσεων, την κατανόηση φυσικής γλώσσας και την αναγνώριση προτύπων.

Σύμφωνα με σύγχρονες προσεγγίσεις, η TN δεν περιορίζεται πλέον στην εκτέλεση προκαθορισμένων εντολών, αλλά εξελίσσεται σε ένα σύστημα λήψης αποφάσεων με αυτονομία, δημιουργώντας ένα νέο παράδειγμα αλληλεπίδρασης ανθρώπου-μηχανής

1.2 Σημασία

Τα τελευταία χρόνια, η TN έχει αποκτήσει καθοριστικό ρόλο σε πλήθος εφαρμογών, από την καθημερινή χρήση (π.χ. ψηφιακοί βοηθοί) έως κρίσιμα συστήματα όπως η υγεία, η ενέργεια και οι έξυπνες πόλεις. Η αυξανόμενη διαθεσιμότητα δεδομένων και η ανάπτυξη προηγμένων αλγορίθμων έχουν επιταχύνει την ενσωμάτωση της TN σε όλα τα επίπεδα της κοινωνίας.

Ιδιαίτερα, η ανάπτυξη τεχνολογιών όπως τα μεγάλα γλωσσικά μοντέλα (LLMs) έχει μετασχηματίσει τη χρήση της TN, επιτρέποντας την παραγωγή κειμένου, εικόνας και γνώσης σε πραγματικό χρόνο. Αυτή η εξέλιξη συνδέεται με την έννοια της Generative AI, η οποία αποτελεί το επόμενο στάδιο μετά τη μηχανική μάθηση.

1.3 Από τη Μηχανική Μάθηση στη Παραγωγική Τεχνητή Νοημοσύνη (generative AI)

Η Μηχανική Μάθηση (Machine Learning) αποτελεί τον πυρήνα της σύγχρονης TN, επιτρέποντας στα συστήματα να μαθαίνουν από δεδομένα χωρίς ρητό προγραμματισμό. Η εξέλιξη προς τη Βαθιά Μάθηση (Deep Learning) ενίσχυσε σημαντικά την ικανότητα των μοντέλων να επεξεργάζονται πολύπλοκες πληροφορίες.

Η μετάβαση στη Παραγωγική Τεχνητή Νοημοσύνη (generative AI) σηματοδοτεί μια θεμελιώδη αλλαγή: τα συστήματα δεν περιορίζονται πλέον στην ανάλυση δεδομένων, αλλά μπορούν να δημιουργούν νέο περιεχόμενο. Ωστόσο, αυτή η ικανότητα συνοδεύεται από προκλήσεις όπως οι Ψευδαισθήσεις (hallucinations) και οι Επιστημικοί Κίνδυνοι (epistemic risks), τα οποία επηρεάζουν την αξιοπιστία των αποτελεσμάτων.

2. Εμπιστοσύνη, Αυτονομία και Ανθρωποκεντρική Τεχνητή Νοημοσύνη

2.1 Η έννοια της εμπιστοσύνης στην Τεχνητή Νοημοσύνη

Η εμπιστοσύνη (trust) αποτελεί έναν από τους σημαντικότερους παράγοντες επιτυχούς αλληλεπίδρασης μεταξύ ανθρώπου και συστημάτων Τεχνητής Νοημοσύνης (TN). Η ραγδαία ανάπτυξη της Παραγωγικής Τεχνητής Νοημοσύνης (generative AI) και των μεγάλων γλωσσικών μοντέλων (Large Language Models – LLMs) έχει αυξήσει σημαντικά την ανάγκη δημιουργίας αξιόπιστων, διαφανών και επεξηγήσιμων συστημάτων.

Σύμφωνα με τη Virvou (2022a), η σύγχρονη εποχή χαρακτηρίζεται από τη μετάβαση των υπολογιστικών συστημάτων από παθητικά εργαλεία εκτέλεσης εντολών σε ενεργούς γνωστικούς παράγοντες με δυνατότητες λήψης αποφάσεων. Η εξέλιξη αυτή δημιουργεί νέες μορφές αλληλεπίδρασης ανθρώπου-μηχανής, στις οποίες η εμπιστοσύνη αποκτά καθοριστικό ρόλο.

Οι Virvou και Tsihrintzis (2023a) επισημαίνουν ότι η αυξανόμενη χρήση συστημάτων όπως το ChatGPT αναδεικνύει τη σημασία των ανθρώπινων AI-experts, οι οποίοι λειτουργούν ως μηχανισμοί ελέγχου, αξιολόγησης και επαλήθευσης των αποτελεσμάτων της TN. Η ανάγκη αυτή προκύπτει κυρίως λόγω της πιθανότητας εμφάνισης λανθασμένων ή παραπλανητικών απαντήσεων από τα μοντέλα γενετικής TN.

Η εμπιστοσύνη επηρεάζεται από πολλούς παράγοντες, όπως:

- η ακρίβεια των απαντήσεων,
- η διαφάνεια του συστήματος,
- η δυνατότητα επεξήγησης των αποφάσεων,
- η αξιοπιστία της συμπεριφοράς,
- η συνέπεια των αποτελεσμάτων,
- η εμπειρία του χρήστη.

Ιδιαίτερα σε κρίσιμα περιβάλλοντα, όπως η υγεία, η εκπαίδευση και η ενέργεια, η εμπιστοσύνη αποτελεί βασική προϋπόθεση αποδοχής και αποτελεσματικής χρήσης των συστημάτων TN.

2.2 Ανθρωποκεντρική TN και Human-AI Interaction

Η ανθρωποκεντρική TN (Human-Centered AI) αποτελεί σύγχρονη σχεδιαστική προσέγγιση, σύμφωνα με την οποία τα συστήματα TN πρέπει να σχεδιάζονται με επίκεντρο τον άνθρωπο, τις ανάγκες του και τις γνωστικές του δυνατότητες.

Η Virvou (2022a) υποστηρίζει ότι, παρά την αυξανόμενη αυτονομία των συστημάτων TN, ο ανθρώπινος παράγοντας παραμένει το σημαντικότερο στοιχείο της αλληλεπίδρασης

ανθρώπου-μηχανής. Οι χρήστες εξακολουθούν να χρειάζονται κατανόηση, διαφάνεια και συναισθηματική ασφάλεια κατά τη χρήση ευφυών συστημάτων.

Παράλληλα, οι Tsihrintzis et al. (2024a) αναφέρουν ότι τα σύγχρονα συστήματα υποστήριξης αποφάσεων βασίζονται όλο και περισσότερο σε τεχνικές TN, γεγονός που καθιστά αναγκαία την ανάπτυξη ανθρωποκεντρικών μηχανισμών αξιολόγησης και ελέγχου.

Η ανθρωποκεντρική προσέγγιση επιδιώκει:

- τη διατήρηση ανθρώπινης εποπτείας,
- τη δυνατότητα παρέμβασης του χρήστη,
- την επεξηγησιμότητα των αποτελεσμάτων,
- την αποφυγή αλγοριθμικών προκαταλήψεων,
- την ηθική χρήση της TN.

Στο πλαίσιο αυτό, η αλληλεπίδραση ανθρώπου-TN δεν αντιμετωπίζεται ως πλήρης αντικατάσταση της ανθρώπινης κρίσης, αλλά ως συνεργατική διαδικασία μεταξύ ανθρώπου και υπολογιστικού συστήματος.

2.3 Το μοντέλο εμπιστοσύνης VIRTSI

Μία από τις σημαντικότερες σύγχρονες προσεγγίσεις αξιολόγησης εμπιστοσύνης αποτελεί το μοντέλο VIRTSI, το οποίο παρουσιάστηκε από τους Virvou και Tsihrintzis (2024a).

Το VIRTSI προτείνει ένα δυναμικό μοντέλο εμπιστοσύνης που βασίζεται στη συνεχή αλληλεπίδραση μεταξύ χρήστη και συστήματος TN. Η εμπιστοσύνη δεν θεωρείται σταθερή μεταβλητή, αλλά μεταβάλλεται ανάλογα με:

- την ποιότητα των απαντήσεων,
- την εμπειρία του χρήστη,
- τη σοβαρότητα των συνεπειών,
- το επίπεδο εξειδίκευσης του χρήστη,
- την αξιοπιστία προηγούμενων αλληλεπιδράσεων.

Οι Virvou και Tsihrintzis (2024b) εξετάζουν ειδικότερα την επίδραση των συνεπειών (consequences) στη δυναμική εμπιστοσύνης. Σύμφωνα με τα αποτελέσματα της έρευνας, οι χρήστες τείνουν να χάνουν γρηγορότερα την εμπιστοσύνη τους όταν ένα σφάλμα σχετίζεται με κρίσιμο τομέα, όπως η υγεία ή η οικονομία.

Το μοντέλο VIRTSI εφαρμόστηκε και στον ενεργειακό τομέα από τους Tsihrintzis et al. (2025b), όπου αξιολογήθηκε η αλληλεπίδραση ειδικών και μη ειδικών χρηστών με συστήματα TN μέσω του ChatGPT. Τα αποτελέσματα έδειξαν ότι οι ειδικοί χρήστες

παρουσιάζουν μεγαλύτερη κριτική ικανότητα και χαμηλότερη αρχική εμπιστοσύνη σε σύγκριση με τους μη ειδικούς.

Η συμβολή του VIRTSI είναι ιδιαίτερα σημαντική, καθώς προσφέρει θεωρητικό και πρακτικό πλαίσιο αξιολόγησης της συνεργασίας ανθρώπου-TN σε πραγματικά περιβάλλοντα εφαρμογών.

2.4 Αυτονομία και Ηθική στην TN

Η αυξανόμενη αυτονομία των συστημάτων TN αποτελεί μία από τις σημαντικότερες προκλήσεις της σύγχρονης πληροφορικής. Τα συστήματα αυτά είναι πλέον ικανά να λαμβάνουν αποφάσεις, να δημιουργούν περιεχόμενο και να εκτελούν σύνθετες εργασίες χωρίς άμεση ανθρώπινη παρέμβαση.

Ωστόσο, η πλήρης αυτοματοποίηση δημιουργεί σημαντικά ζητήματα ηθικής, λογοδοσίας και ελέγχου. Η Virvou (2025a) αναφέρει ότι η μηχανική λογισμικού οφείλει να ισορροπήσει ανάμεσα στην αυτονομία της TN και στις απαιτήσεις ηθικής και ανθρωποκεντρικής σχεδίασης.

Η ανάγκη αυτή είναι ιδιαίτερα έντονη σε εφαρμογές όπου:

- οι αποφάσεις επηρεάζουν ανθρώπινες ζωές,
- χρησιμοποιούνται προσωπικά δεδομένα,
- υπάρχει υψηλό επίπεδο ρίσκου,
- απαιτείται ανθρώπινη ευθύνη.

Για τον λόγο αυτό, προτείνονται υβριδικά μοντέλα συνεργασίας ανθρώπου-TN, στα οποία η TN λειτουργεί υποστηρικτικά και όχι αντικαταστατικά προς τον άνθρωπο.

Οι Virvou και Tsihrintzis (2023b) προτείνουν επίσης state-chart μοντέλο μηχανικής απαιτήσεων για αξιόπιστα συστήματα τεχνητής νοημοσύνης, με στόχο τη διατήρηση ελέγχου και προβλεψιμότητας στη συμπεριφορά των συστημάτων.

2.5 Επιστημικοί Κίνδυνοι (epistemic risks) και Ψευδαισθήσεις (hallucinations)

Ένα από τα σημαντικότερα προβλήματα της σύγχρονης Παραγωγικής Τεχνητής Νοημοσύνης (generative AI) είναι το φαινόμενο των Ψευδαισθήσεων (hallucinations), δηλαδή η παραγωγή πληροφοριών που παρουσιάζονται ως έγκυρες και αξιόπιστες, ενώ στην πραγματικότητα είναι ανακριβείς, παραπλανητικές ή ακόμη και πλήρως λανθασμένες. Το ιδιαίτερα ανησυχητικό στοιχείο είναι ότι τα μεγάλα γλωσσικά μοντέλα διατυπώνουν

αυτές τις απαντήσεις με υψηλό βαθμό γλωσσικής συνοχής και αυτοπεποίθησης, γεγονός που δυσκολεύει σημαντικά τον χρήστη να εντοπίσει το σφάλμα.

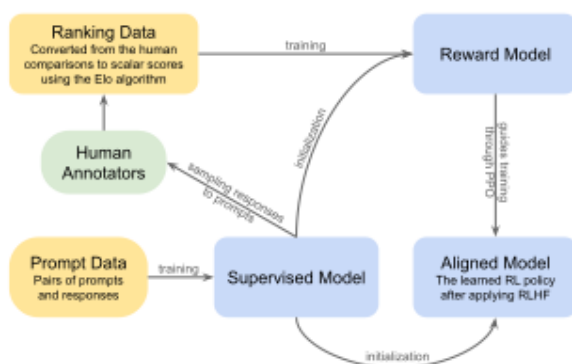
Οι Virvou et al. (2025c) προσεγγίζουν το φαινόμενο αυτό μέσα από την έννοια των Επιστημικών Κινδύνων (epistemic risks), δηλαδή των κινδύνων που σχετίζονται με την αξιοπιστία της γνώσης που παράγεται από τα συστήματα ΤΝ. Οι κίνδυνοι αυτοί αποκτούν ιδιαίτερη σημασία σε εκπαιδευτικά περιβάλλοντα, όπου οι μαθητές και οι εκπαιδευόμενοι συχνά αντιμετωπίζουν τις απαντήσεις των LLMs ως έγκυρες πηγές πληροφόρησης χωρίς να προχωρούν σε διαδικασία επαλήθευσης.

Σύμφωνα με τη μελέτη, οι Ψευδαισθήσεις (hallucinations) σχετίζονται άμεσα με τον τρόπο λειτουργίας των μεγάλων γλωσσικών μοντέλων. Τα LLMs βασίζονται κυρίως σε στατιστική πρόβλεψη επόμενων λέξεων και όχι σε πραγματική κατανόηση της γνώσης ή της αλήθειας. Ως αποτέλεσμα, είναι δυνατόν να παράγουν κείμενα που εμφανίζονται απολύτως λογικά και καλογραμμένα, αλλά περιέχουν ανακρίβειες ή ανύπαρκτες πληροφορίες. Το πρόβλημα επιτείνεται όταν τα δεδομένα εκπαίδευσης είναι ελλιπή, προκατειλημμένα ή περιέχουν αντικρουόμενες πληροφορίες. Παράλληλα, τα συστήματα αυτά διαθέτουν περιορισμένη δυνατότητα λογικής επαλήθευσης και αδυνατούν να αξιολογήσουν ουσιαστικά την εγκυρότητα των απαντήσεών τους.

Οι Επιστημικοί Κίνδυνοι (epistemic risks) μπορούν να επηρεάσουν άμεσα κρίσιμους τομείς εφαρμογής της ΤΝ. Στην εκπαίδευση, η διάδοση λανθασμένης γνώσης μπορεί να οδηγήσει σε παρανοήσεις και εσφαλμένη μαθησιακή διαδικασία. Στην ιατρική, μια ανακριβής πληροφορία ενδέχεται να επηρεάσει τη διαγνωστική ή θεραπευτική διαδικασία, ενώ στον ενεργειακό ή νομικό τομέα οι λανθασμένες προβλέψεις ή συμβουλές μπορούν να οδηγήσουν σε σοβαρές επιχειρησιακές ή διοικητικές συνέπειες.

Η προβληματική αυτή συνδέεται άμεσα με την εμπιστοσύνη των χρηστών απέναντι στα συστήματα ΤΝ. Όσο περισσότερο οι άνθρωποι χρησιμοποιούν LLMs ως εργαλεία καθημερινής λήψης αποφάσεων, τόσο σημαντικότερη γίνεται η ανάγκη διασφάλισης της αξιοπιστίας των παραγόμενων αποτελεσμάτων. Για τον λόγο αυτό, οι Virvou et al. (2025a) προτείνουν την ενσωμάτωση trust performance indicators σε συστήματα Παραγωγική Τεχνητή Νοημοσύνη (generative AI) που χρησιμοποιούνται στον ενεργειακό τομέα, με στόχο τη συνεχή αξιολόγηση της αξιοπιστίας και της ποιότητας των απαντήσεων.

Η αντιμετώπιση των Επιστημικών Κινδύνων (epistemic risks) απαιτεί συνδυασμό τεχνικών και ανθρωποκεντρικών προσεγγίσεων. Η Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) μπορεί να συμβάλει στη μεγαλύτερη διαφάνεια των μοντέλων, ενώ οι μηχανισμοί επαλήθευσης γεγονότων (fact-checking) και αξιολόγησης αξιοπιστίας μπορούν να περιορίσουν τη διάδοση λανθασμένων πληροφοριών. Παράλληλα, η ανθρώπινη εποπτεία παραμένει απαραίτητη, ιδιαίτερα σε τομείς υψηλού κινδύνου, όπου η πλήρης εμπιστοσύνη σε αυτόνομα συστήματα ΤΝ εξακολουθεί να θεωρείται προβληματική.



Εικόνα 1: Διαδικασία Ενισχυτικής Μάθησης από Ανθρώπινη Ανατροφοδότηση (Reinforcement Learning from Human Feedback – RLHF) σε μεγάλα γλωσσικά μοντέλα.

2.6 Επεξηγησιμότητα και Επεξηγήσιμη Τεχνητή Νοημοσύνη (Explainable AI)

Η Επεξηγησιμότητα (explainability) αποτελεί βασική απαίτηση για τη δημιουργία αξιόπιστων συστημάτων ΤΝ. Οι χρήστες επιθυμούν να κατανοούν όχι μόνο το αποτέλεσμα ενός αλγορίθμου, αλλά και τη διαδικασία μέσω της οποίας προέκυψε η απόφαση.

Παρότι η Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) θεωρείται απαραίτητη σε ιατρικά συστήματα, αρκετές προσεγγίσεις εξακολουθούν να αντιμετωπίζουν δυσκολίες ως προς την ισορροπία μεταξύ ακρίβειας και ερμηνευσιμότητας

Οι Panagoulas et al. (2023b) εφαρμόζουν τη θεωρία DOI για τον καθορισμό του απαιτούμενου επιπέδου Επεξηγησιμότητας (explainability) σε ιατρικές εφαρμογές ΤΝ. Τα αποτελέσματα έδειξαν ότι όσο υψηλότερος είναι ο κίνδυνος μιας απόφασης, τόσο μεγαλύτερη ανάγκη υπάρχει για επεξηγησιμότητα και ανθρώπινο έλεγχο.

Παράλληλα, οι Panagoulas, Tsihrintzis και Virvou (2025c) προτείνουν αρχιτεκτονική microservices για αξιόπιστη τεχνητή νοημοσύνη σε ιατρικές εφαρμογές, ενισχύοντας τη διαφάνεια και τη δυνατότητα παρακολούθησης των αποφάσεων.

Η Επεξηγησιμότητα (explainability) δεν σχετίζεται μόνο με την τεχνική κατανόηση των αλγορίθμων, αλλά και με την ψυχολογική αποδοχή της ΤΝ από τον άνθρωπο.

2.7 Συμπεράσματα Κεφαλαίου

Η εμπιστοσύνη, η επεξηγησιμότητα και η ανθρωποκεντρική σχεδίαση αποτελούν θεμελιώδεις προϋποθέσεις για τη βιώσιμη και ασφαλή ανάπτυξη της Τεχνητής Νοημοσύνης.

Η ανάπτυξη μοντέλων όπως το VIRTISI συμβάλλει σημαντικά στην κατανόηση της δυναμικής εμπιστοσύνης μεταξύ ανθρώπου και ΤΝ, ενώ η μελέτη των Επιστημικών Κινδύνων (epistemic risks) και των Ψευδαισθήσεων (hallucinations) αναδεικνύει τις προκλήσεις της σύγχρονης Παραγωγικής Τεχνητής Νοημοσύνης (generative AI).

Παράλληλα, η ισορροπία μεταξύ αυτονομίας και ανθρώπινου ελέγχου αναδεικνύεται ως κρίσιμο ζήτημα για τη μελλοντική εξέλιξη των ευφυών συστημάτων, ιδιαίτερα σε τομείς υψηλής κοινωνικής και ηθικής σημασίας.

3. Παραγωγική Τεχνητή Νοημοσύνη (generative AI) και Ευφυή Συστήματα στην Εκπαίδευση και τον Πολιτισμό

3.1 Η εμφάνιση της Παραγωγική Τεχνητή Νοημοσύνη (generative AI) στην εκπαίδευση

Η ανάπτυξη της Παραγωγικής Τεχνητής Νοημοσύνης (generative AI) και των μεγάλων γλωσσικών μοντέλων (Large Language Models – LLMs) έχει μετασχηματίσει σημαντικά τον χώρο της εκπαίδευσης. Συστήματα όπως το ChatGPT παρέχουν τη δυνατότητα παραγωγής φυσικής γλώσσας, επεξήγησης εννοιών, δημιουργίας εκπαιδευτικού υλικού και υποστήριξης μαθησιακών διαδικασιών σε πραγματικό χρόνο.

Σύμφωνα με τους Virvou και Tsihrintzis (2023c), η αξιολόγηση του ChatGPT στην εκπαίδευση δείχνει ότι τα LLMs μπορούν να λειτουργήσουν ως ευφυείς εκπαιδευτικοί βοηθοί, παρέχοντας εξατομικευμένη υποστήριξη στους μαθητές. Η δυνατότητα άμεσης αλληλεπίδρασης και παραγωγής περιεχομένου ενισχύει τη μαθησιακή εμπειρία και επιτρέπει την προσαρμογή του εκπαιδευτικού υλικού στις ανάγκες του χρήστη.

Ωστόσο, οι ερευνητές επισημαίνουν ότι η αποτελεσματικότητα της Παραγωγικής Τεχνητής Νοημοσύνης (generative AI) εξαρτάται σε μεγάλο βαθμό από:

- την ποιότητα των παραγόμενων απαντήσεων,
- την αξιοπιστία των πληροφοριών,
- την ικανότητα επεξήγησης,
- την ύπαρξη ανθρώπινης εποπτείας.

Η εισαγωγή της ΤΝ στην εκπαίδευση δεν περιορίζεται στην αυτοματοποίηση διαδικασιών, αλλά επηρεάζει συνολικά τον τρόπο με τον οποίο οι μαθητές αποκτούν γνώση, αλληλεπιδρούν με το εκπαιδευτικό περιεχόμενο και αναπτύσσουν δεξιότητες.

3.2 Παραγωγική Τεχνητή Νοημοσύνη (generative AI) και ανάπτυξη δεξιοτήτων του 21ου αιώνα

Η σύγχρονη εκπαίδευση δίνει ιδιαίτερη έμφαση στις δεξιότητες του 21ου αιώνα, όπως:

- η κριτική σκέψη,
- η δημιουργικότητα,
- η επίλυση προβλημάτων,
- η συνεργασία,
- η ψηφιακή παιδεία.

Οι Virvou et al. (2025b) υποστηρίζουν ότι η Παραγωγική Τεχνητή Νοημοσύνη (generative AI) μπορεί να συμβάλει σημαντικά στην ανάπτυξη αυτών των δεξιοτήτων μέσω

διαδραστικών και εξατομικευμένων μαθησιακών περιβαλλόντων. Παράλληλα, τονίζεται ότι οι μαθητές πρέπει να αποκτήσουν και νέες δεξιότητες σχετικές με την αξιολόγηση και υπεύθυνη χρήση της ΤΝ.

Η έννοια των bi-directional skills που προτείνουν οι Virvou et al. (2025b) αναφέρεται στη διπλή κατεύθυνση μάθησης:

1. οι χρήστες αξιοποιούν την ΤΝ ως εργαλείο μάθησης,
2. οι χρήστες μαθαίνουν να αξιολογούν, να ελέγχουν και να επιβλέπουν την ΤΝ.

Η ανάγκη αυτή προκύπτει λόγω της αυξανόμενης πολυπλοκότητας των LLMs και της πιθανότητας εμφάνισης Επιστημικοί Κίνδυνοι (epistemic risks) και Ψευδαισθήσεων (hallucinations).

3.3 Ψευδαισθήσεις (hallucinations) και Επιστημικοί Κίνδυνοι (epistemic risks) στην εκπαίδευση

Παρά τα σημαντικά πλεονεκτήματα της Παραγωγικής Τεχνητής Νοημοσύνης (generative AI), η χρήση της στην εκπαίδευση συνοδεύεται από σοβαρές προκλήσεις. Μία από τις σημαντικότερες είναι το φαινόμενο των Ψευδαισθήσεων (hallucinations).

Οι Virvou et al. (2025c) αναφέρουν ότι οι Ψευδαισθήσεις (hallucinations) αποτελούν Επιστημικούς Κινδύνους (epistemic risks), καθώς τα LLMs μπορούν να παράγουν λανθασμένες ή ανύπαρκτες πληροφορίες με πειστικό τρόπο. Το πρόβλημα αυτό είναι ιδιαίτερα επικίνδυνο σε εκπαιδευτικά περιβάλλοντα, όπου οι μαθητές ενδέχεται να αποδεχθούν τις απαντήσεις ως αληθείς χωρίς επαλήθευση.

Οι βασικοί κίνδυνοι περιλαμβάνουν:

- διάδοση παραπληροφόρησης,
- ανάπτυξη λανθασμένων γνωστικών μοντέλων,
- περιορισμό κριτικής σκέψης,
- υπερβολική εξάρτηση από την ΤΝ.

Σύμφωνα με τους Virvou et al. (2025c), η αντιμετώπιση των epistemic risks απαιτεί:

- ανάπτυξη Εξηγήσιμης Εκπαιδευτικής Τεχνητής Νοημοσύνης (explainable educational AI),
- ενσωμάτωση μηχανισμών επαλήθευσης γνώσης,
- ανθρώπινη εποπτεία,
- εκπαίδευση των μαθητών στη σωστή χρήση της ΤΝ.

Η κριτική αξιολόγηση των αποτελεσμάτων της Παραγωγικής Τεχνητής Νοημοσύνης (generative AI) αποτελεί πλέον βασική δεξιότητα της σύγχρονης εκπαίδευσης.

3.4 Πολιτιστικές εφαρμογές της TN

Η TN χρησιμοποιείται πλέον εκτενώς και σε πολιτιστικά περιβάλλοντα, όπου αξιοποιείται για:

- ανάλυση πολιτισμικού περιεχομένου,
- δημιουργία εκπαιδευτικού υλικού,
- επεξεργασία φυσικής γλώσσας,
- διατήρηση πολιτιστικής κληρονομιάς.

Οι Virvou et al. (2023) εξετάζουν τη χρήση του ChatGPT σε e-learning περιβάλλοντα πολιτιστικής κληρονομιάς, με έμφαση στην ανάλυση στίχων τραγουδιών και ποιημάτων. Η μελέτη δείχνει ότι τα LLMs μπορούν να ενισχύσουν τη συμμετοχή και αλληλεπίδραση των μαθητών μέσω διαδραστικών μορφών μάθησης.

Παράλληλα, οι Panagoulas et al. (2024a) παρουσιάζουν το σύστημα LYRICEL, το οποίο συνδυάζει:

- Large Language Models,
- Machine Learning,
- Knowledge Graphs,
- κανόνες πολιτισμικής ανάλυσης.

Το LYRICEL χρησιμοποιείται για διαπολιτισμική ανάλυση ελληνικών τραγουδιών και ανάπτυξη πολιτιστικού e-learning περιεχομένου.

Σύμφωνα με τους Panagoulas et al. (2025c), η ενσωμάτωση Γράφων Γνώσης (Knowledge Graphs – KGs) με Μεγάλα Γλωσσικά Μοντέλα (Large Language Models – LLMs) βελτιώνει:

- τη σημασιολογική κατανόηση,
- την πολιτισμική ακρίβεια,
- τη συνοχή των παραγόμενων αποτελεσμάτων.

Η αξιοποίηση της TN στον πολιτισμό δημιουργεί νέες δυνατότητες πρόσβασης στη γνώση και ενισχύει τη διατήρηση και προβολή πολιτιστικής κληρονομιάς.

3.5 Παιχνίδια, ευφυή περιβάλλοντα και εκπαιδευτική TN

Τα εκπαιδευτικά παιχνίδια αποτελούν έναν ακόμη σημαντικό τομέα εφαρμογής της TN. Σύμφωνα με τους Papadimitriou και Virvou (2025a), τα παιχνίδια μπορούν να λειτουργήσουν ως ολιστικά μαθησιακά περιβάλλοντα που ενισχύουν:

- τη δημιουργικότητα,
- τη συνεργασία,
- τη λήψη αποφάσεων,
- τη βιωματική μάθηση.

Η ενσωμάτωση TN στα εκπαιδευτικά παιχνίδια επιτρέπει:

- δυναμική προσαρμογή δυσκολίας,
- εξατομικευμένη εμπειρία,
- ανάλυση συμπεριφοράς χρήστη,
- συναισθηματική αλληλεπίδραση.

Η χρήση Συναισθηματική Υπολογιστική (Affective Computing) σε τέτοια περιβάλλοντα επιτρέπει στα συστήματα να αναγνωρίζουν και να ανταποκρίνονται στη συναισθηματική κατάσταση του χρήστη.

Οι Virvou et al. (2025d) συγκρίνουν συναισθηματικές συμβουλές που παρέχονται από affective intelligent tutoring systems και από το ChatGPT. Τα αποτελέσματα δείχνουν ότι η ανθρώπινη συναισθηματική κατανόηση εξακολουθεί να αποτελεί πρόκληση για τα σύγχρονα LLMs.

3.6 E-learning και εξατομικευμένη αξιολόγηση

Η TN έχει συμβάλει σημαντικά στην ανάπτυξη συστημάτων e-learning με δυνατότητες εξατομικευσης και προσαρμοστικής αξιολόγησης.

Οι Chrysafiadi et al. (2023a) εξετάζουν την αποδοχή και αποτελεσματικότητα των e-learning συστημάτων από φοιτητές πληροφορικής στην Ελλάδα κατά την περίοδο COVID-19. Τα αποτελέσματα δείχνουν ότι η προσωποποίηση και η προσαρμογή του περιεχομένου ενισχύουν σημαντικά την εκπαιδευτική εμπειρία.

Παράλληλα, οι Chrysafiadi et al. (2022a) προτείνουν fuzzy-based μηχανισμό αυτόματης εξατομικευμένης αξιολόγησης σε περιβάλλον εκμάθησης προγραμματισμού. Η χρήση fuzzy logic επιτρέπει:

- καλύτερη προσαρμογή στις δυνατότητες του μαθητή,
- δυναμική αξιολόγηση,
- ευέλικτη ανατροφοδότηση.

Η ενσωμάτωση τεχνικών TN στα e-learning περιβάλλοντα δημιουργεί περισσότερο προσωποποιημένες και αποτελεσματικές μαθησιακές εμπειρίες.

3.7 Συμπεράσματα Κεφαλαίου

Η Παραγωγική Τεχνητή Νοημοσύνη (generative AI) έχει επιφέρει σημαντικό μετασχηματισμό στον χώρο της εκπαίδευσης και του πολιτισμού. Τα μεγάλα γλωσσικά μοντέλα επιτρέπουν νέες μορφές εξατομικευμένης μάθησης, πολιτιστικής ανάλυσης και διαδραστικής εκπαιδευτικής εμπειρίας.

Παράλληλα, η ανάπτυξη αυτών των τεχνολογιών συνοδεύεται από σημαντικές προκλήσεις, όπως οι Ψευδαισθήσεις (hallucinations), η αξιοπιστία της γνώσης και η ανάγκη ανάπτυξης κριτικής σκέψης στους χρήστες.

Η συνδυαστική αξιοποίηση Εξηγήσιμης Τεχνητής Νοημοσύνης (explainable AI), Συναισθηματικής Υπολογιστικής (Affective Computing) και ανθρωποκεντρικής σχεδίασης αναμένεται να διαδραματίσει καθοριστικό ρόλο στη μελλοντική εξέλιξη των ευφών εκπαιδευτικών συστημάτων.

4. Τεχνητή Νοημοσύνη σε Έξυπνες Πόλεις, Ενέργεια και Αστικά Οικοσυστήματα Απόφασης

4.1 Η μετάβαση προς τις έξυπνες πόλεις

Η έννοια της «έξυπνης πόλης» (Smart City) αποτελεί μία από τις σημαντικότερες τεχνολογικές και κοινωνικές εξελίξεις των τελευταίων ετών. Οι σύγχρονες πόλεις καλούνται να αντιμετωπίσουν σύνθετα προβλήματα, όπως η αύξηση του πληθυσμού, η ενεργειακή κατανάλωση, η κυκλοφοριακή συμφόρηση, η περιβαλλοντική επιβάρυνση και η ανάγκη αποτελεσματικότερης διαχείρισης πόρων. Η Τεχνητή Νοημοσύνη αναδεικνύεται ως βασικός μηχανισμός αντιμετώπισης αυτών των προκλήσεων, καθώς επιτρέπει την ανάλυση τεράστιου όγκου δεδομένων και τη λήψη αποφάσεων σε πραγματικό χρόνο.

Σύμφωνα με τους Tsihrintzis et al. (2026a), οι έξυπνες πόλεις δεν πρέπει να αντιμετωπίζονται αποκλειστικά ως τεχνολογικά συστήματα αυτοματοποίησης, αλλά ως πολύπλοκα κοινωνικο-τεχνικά οικοσυστήματα στα οποία συνυπάρχουν τεχνολογικές υποδομές, ανθρώπινη δραστηριότητα και μηχανισμοί λήψης αποφάσεων. Η λειτουργία μιας smart city εξαρτάται από τη συνεχή αλληλεπίδραση μεταξύ ανθρώπων και ευφύων συστημάτων, γεγονός που καθιστά αναγκαία την ανάπτυξη ανθρωποκεντρικών προσεγγίσεων σχεδίασης.

Η ΤΝ χρησιμοποιείται σε πληθώρα εφαρμογών έξυπνων πόλεων. Στα συστήματα μεταφορών, αλγόριθμοι μηχανικής μάθησης αναλύουν δεδομένα κυκλοφορίας και προσαρμόζουν δυναμικά τη λειτουργία των φωτεινών σηματοδοτών, μειώνοντας τη συμφόρηση και τους χρόνους μετακίνησης. Στον ενεργειακό τομέα, predictive analytics χρησιμοποιούνται για πρόβλεψη κατανάλωσης και βελτιστοποίηση ενεργειακών δικτύων. Παράλληλα, αισθητήρες IoT συλλέγουν συνεχώς δεδομένα για την ποιότητα αέρα, την κατανάλωση νερού και τη λειτουργία δημοσίων υποδομών.

Ωστόσο, η τεχνολογική πρόοδος δημιουργεί και νέες απαιτήσεις. Όσο περισσότερο οι πόλεις βασίζονται σε αλγοριθμικά συστήματα, τόσο μεγαλύτερη σημασία αποκτούν ζητήματα όπως η εμπιστοσύνη των πολιτών, η διαφάνεια των αποφάσεων και η προστασία προσωπικών δεδομένων.

4.2 Αλληλεπίδραση Ανθρώπου-Έξυπνης Πόλης και ανθρωποκεντρικός σχεδιασμός

Η αλληλεπίδραση ανθρώπου-έξυπνης πόλης (Human-Smart City Interaction) αποτελεί κεντρικό στοιχείο της σύγχρονης αστικής πληροφορικής. Σύμφωνα με τους Tsihrintzis et

al. (2026a), η αποτελεσματικότητα μιας έξυπνης πόλης δεν καθορίζεται αποκλειστικά από την τεχνολογική υπεροχή των υποδομών της, αλλά κυρίως από το κατά πόσο οι πολίτες αισθάνονται ότι οι υπηρεσίες λειτουργούν προς όφελός τους και υπό τον έλεγχό τους.

Στα πρώτα στάδια ανάπτυξης των smart cities, οι πολίτες αντιμετωπίζονταν κυρίως ως παθητικοί αποδέκτες υπηρεσιών. Η σύγχρονη προσέγγιση, όμως, θεωρεί ότι οι χρήστες πρέπει να συμμετέχουν ενεργά στον σχεδιασμό και στη λειτουργία των ψηφιακών υποδομών. Η ΤΝ καλείται επομένως να λειτουργήσει όχι ως μηχανισμός αντικατάστασης του ανθρώπου, αλλά ως εργαλείο ενίσχυσης των ανθρώπινων δυνατοτήτων και υποστήριξης αποφάσεων.

Η ανάγκη ανθρωποκεντρικού σχεδιασμού είναι ιδιαίτερα σημαντική επειδή οι αποφάσεις που λαμβάνονται από τα συστήματα ΤΝ επηρεάζουν άμεσα την καθημερινότητα των πολιτών. Για παράδειγμα, ένα σύστημα έξυπνης κυκλοφορίας μπορεί να καθορίζει τις διαδρομές μετακίνησης χιλιάδων ανθρώπων, ενώ ένα σύστημα ενεργειακής διαχείρισης μπορεί να επηρεάζει την κατανάλωση και το οικονομικό κόστος για ολόκληρες περιοχές.

Για τον λόγο αυτό, οι πολίτες απαιτούν όλο και περισσότερο διαφάνεια και δυνατότητα κατανόησης των αλγοριθμικών αποφάσεων. Η έλλειψη επεξηγησιμότητας μπορεί να οδηγήσει σε μειωμένη εμπιστοσύνη, ακόμη και όταν τα συστήματα είναι τεχνικά αποτελεσματικά. Επιπλέον, η αίσθηση απώλειας ελέγχου ενδέχεται να δημιουργήσει κοινωνική αντίδραση απέναντι στις τεχνολογίες ΤΝ.

Οι έξυπνες πόλεις του μέλλοντος επομένως δεν θα πρέπει να σχεδιάζονται μόνο με γνώμονα τη βελτιστοποίηση των υποδομών, αλλά και με στόχο τη δημιουργία σχέσης εμπιστοσύνης μεταξύ πολιτών και ευφυών συστημάτων.

4.3 Προσωπικά δεδομένα, GDPR και αλγοριθμική διαφάνεια

Η λειτουργία των smart cities βασίζεται στη συνεχή συλλογή και επεξεργασία μεγάλου όγκου δεδομένων. Οι αισθητήρες, τα κινητά τηλέφωνα, τα έξυπνα οχήματα και τα συστήματα παρακολούθησης παράγουν τεράστιες ποσότητες πληροφοριών σχετικά με τη συμπεριφορά των πολιτών και τη λειτουργία της πόλης.

Η πραγματικότητα αυτή δημιουργεί σημαντικά ζητήματα ιδιωτικότητας και ασφάλειας δεδομένων. Οι Theocharis και Tsihrintzis (2026a) επισημαίνουν ότι η ενσωμάτωση ΤΝ σε έξυπνες πόλεις πρέπει να συνοδεύεται από αυστηρούς μηχανισμούς συμμόρφωσης με τον GDPR, ώστε να προστατεύονται τα θεμελιώδη δικαιώματα των πολιτών.

Ένα από τα βασικά προβλήματα είναι ότι οι περισσότεροι πολίτες δεν γνωρίζουν ακριβώς ποια δεδομένα συλλέγονται, πώς χρησιμοποιούνται και ποια αλγοριθμικά μοντέλα λαμβάνουν αποφάσεις βασισμένες σε αυτά. Η αδιαφάνεια αυτή δημιουργεί κινδύνους όχι μόνο για την ιδιωτικότητα αλλά και για τη δημοκρατική λειτουργία των πόλεων.

Επιπλέον, τα συστήματα TN ενδέχεται να αναπαράγουν αλγοριθμικές προκαταλήψεις. Εάν τα δεδομένα εκπαίδευσης περιέχουν κοινωνικές ανισότητες ή στρεβλώσεις, τα αποτελέσματα των αλγορίθμων μπορεί να οδηγήσουν σε άδικες ή μεροληπτικές αποφάσεις. Το πρόβλημα αυτό είναι ιδιαίτερα κρίσιμο σε εφαρμογές δημόσιας ασφάλειας, κοινωνικών υπηρεσιών και διαχείρισης πόρων.

Για τον λόγο αυτό, η Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) αποκτά ιδιαίτερη σημασία στα αστικά οικοσυστήματα. Οι πολίτες πρέπει να έχουν τη δυνατότητα να γνωρίζουν γιατί ένα σύστημα πήρε μια συγκεκριμένη απόφαση και ποιοι παράγοντες επηρέασαν το αποτέλεσμα. Η διαφάνεια αποτελεί πλέον όχι μόνο τεχνική απαίτηση, αλλά και κοινωνική και ηθική αναγκαιότητα.

4.4 TN και ενεργειακά συστήματα

Ο ενεργειακός τομέας αποτελεί ένα από τα σημαντικότερα πεδία εφαρμογής της TN στις έξυπνες πόλεις. Η ανάγκη για βιώσιμη ανάπτυξη και ενεργειακή μετάβαση έχει οδηγήσει στη δημιουργία σύνθετων ενεργειακών οικοσυστημάτων που βασίζονται σε Προγνωστική Αναλυτική (Predictive Analytics) και Συστήματα Υποστήριξης Αποφάσεων βασισμένα στην Τεχνητή Νοημοσύνη (AI-driven Decision Support Systems).

Οι Tsihrintzis et al. (2024a) υποστηρίζουν ότι τα σύγχρονα ενεργειακά δίκτυα παράγουν τεράστιες ποσότητες δεδομένων, οι οποίες μπορούν να αξιοποιηθούν μέσω μηχανικής μάθησης για πρόβλεψη κατανάλωσης, βελτιστοποίηση παραγωγής και ανίχνευση ανωμαλιών.

Η TN χρησιμοποιείται ήδη για:

- πρόβλεψη ενεργειακής ζήτησης,
- διαχείριση φωτοβολταϊκών πάρκων,
- έξυπνη κατανομή φορτίου,
- βελτιστοποίηση αποθήκευσης ενέργειας,
- ανίχνευση ενεργειακών βλαβών.

Η χρήση προγνωστικών μοντέλων επιτρέπει την προσαρμογή των ενεργειακών συστημάτων σε πραγματικό χρόνο, μειώνοντας απώλειες και βελτιώνοντας τη συνολική αποδοτικότητα.

Παράλληλα, οι Virvou et al. (2025a) εξετάζουν την ενσωμάτωση Παραγωγικής Τεχνητής Νοημοσύνης (generative AI) και των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models – LLMs) στον ενεργειακό ψηφιακό μετασχηματισμό. Η χρήση των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models – LLMs) στον ενεργειακό τομέα επιτρέπει την αλληλεπίδραση ανθρώπου-συστήματος μέσω φυσικής γλώσσας, διευκολύνοντας την πρόσβαση σε πολύπλοκες ενεργειακές πληροφορίες ακόμη και από μη ειδικούς χρήστες.

Ωστόσο, η χρήση Generative AI σε κρίσιμες ενεργειακές υποδομές συνοδεύεται από σημαντικούς κινδύνους. Οι Ψευδαισθήσεις (hallucinations) και οι Επιστημικοί Κίνδυνοι (epistemic risks) μπορούν να οδηγήσουν σε λανθασμένες αποφάσεις με σοβαρές επιπτώσεις για την ενεργειακή σταθερότητα. Για τον λόγο αυτό, απαιτούνται μηχανισμοί αξιολόγησης αξιοπιστίας και συνεχούς ανθρώπινης εποπτείας.

4.5 Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) και αξιοπιστία ενεργειακών προβλέψεων

Η επεξηγησιμότητα αποτελεί κρίσιμο στοιχείο των ενεργειακών εφαρμογών ΤΝ. Οι ειδικοί δεν αρκούνται πλέον μόνο στην ακρίβεια των προβλέψεων, αλλά απαιτούν κατανόηση του τρόπου με τον οποίο παράγεται το αποτέλεσμα.

Οι Panagoulas et al. (2025a) προτείνουν explainable optimization framework για προβλέψεις φωτοβολταϊκής παραγωγής μέσω Extreme Value Analysis και δυναμικής ανάλυσης συσχετίσεων. Η έρευνα αναδεικνύει ότι η explainability επιτρέπει στους ειδικούς να εντοπίζουν αδυναμίες των μοντέλων και να αξιολογούν το επίπεδο αξιοπιστίας των προβλέψεων.

Η Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) είναι ιδιαίτερα σημαντική επειδή τα ενεργειακά δεδομένα χαρακτηρίζονται από υψηλή πολυπλοκότητα και αβεβαιότητα. Μεταβολές καιρικών συνθηκών, ακραία φαινόμενα και διακυμάνσεις κατανάλωσης μπορούν να επηρεάσουν σημαντικά τα αποτελέσματα των προγνωστικών μοντέλων.

Η δυνατότητα επεξήγησης των προβλέψεων ενισχύει την εμπιστοσύνη των ενεργειακών stakeholders και βελτιώνει τη συνεργασία ανθρώπων και αλγοριθμικών συστημάτων.

4.6 Συνθετικά Δεδομένα (Synthetic Data) και Παραγωγικά Αντιπαραθετικά Δίκτυα (Generative Adversarial Networks – GANs) σε ενεργειακά οικοσυστήματα

Ένα σημαντικό πρόβλημα των ενεργειακών συστημάτων ΤΝ είναι η περιορισμένη διαθεσιμότητα ποιοτικών δεδομένων. Σε πολλές περιπτώσεις, τα πραγματικά datasets είναι ελλιπή, ανισόρροπα ή περιέχουν περιορισμένο αριθμό ακραίων συμβάντων.

Οι Panagoulas et al. (2026) προτείνουν framework παραγωγής synthetic data μέσω των Παραγωγικών Αντιπαραθετικών Δικτύων (Generative Adversarial Networks – GANs) για δεδομένα φόρτισης ηλεκτρικών οχημάτων. Το μοντέλο χρησιμοποιεί Ανάλυση Ακραίων Τιμών (Extreme Value Analysis – EVA) ώστε να προσομοιώνει ρεαλιστικά ακραίες καταστάσεις ενεργειακής κατανάλωσης.

Η χρήση Συνθετικών Δεδομένων (synthetic data) επιτρέπει:

- καλύτερη εκπαίδευση αλγορίθμων,
- αντιμετώπιση data imbalance,
- προσομοίωση κρίσιμων σεναρίων,
- προστασία προσωπικών δεδομένων.

Τα GAN-based μοντέλα αποτελούν πλέον βασικό εργαλείο ανάπτυξης αξιόπιστων ενεργειακών predictive systems, ιδιαίτερα σε περιβάλλοντα όπου η πρόσβαση σε πραγματικά δεδομένα είναι περιορισμένη.

4.7 Συμπεράσματα Κεφαλαίου

Η Τεχνητή Νοημοσύνη μετασχηματίζει ριζικά τον τρόπο λειτουργίας των έξυπνων πόλεων και των ενεργειακών οικοσυστημάτων. Η δυνατότητα ανάλυσης μεγάλου όγκου δεδομένων και λήψης αποφάσεων σε πραγματικό χρόνο δημιουργεί σημαντικές δυνατότητες βελτιστοποίησης υποδομών και υπηρεσιών.

Παράλληλα, η αυξανόμενη εξάρτηση από αλγοριθμικά συστήματα δημιουργεί νέες κοινωνικές, ηθικές και τεχνικές προκλήσεις. Η προστασία προσωπικών δεδομένων, η explainability, η εμπιστοσύνη των πολιτών και η αντιμετώπιση epistemic risks αποτελούν πλέον κρίσιμες απαιτήσεις για τη βιώσιμη ανάπτυξη των smart cities.

Η μελλοντική εξέλιξη των έξυπνων πόλεων αναμένεται να βασιστεί σε ανθρωποκεντρικά, explainable και trustworthy AI systems που θα λειτουργούν συνεργατικά με τον άνθρωπο και όχι ανταγωνιστικά προς αυτόν.

5. Τεχνητή Νοημοσύνη στην Υγεία, τη Βιοϊατρική και τα Εξατομικευμένα Συστήματα

5.1 Η εξέλιξη της TN στον χώρο της υγείας

Η Τεχνητή Νοημοσύνη έχει επηρεάσει σημαντικά τον χώρο της υγείας κατά την τελευταία δεκαετία, μεταβάλλοντας τον τρόπο με τον οποίο αναλύονται τα ιατρικά δεδομένα, λαμβάνονται οι κλινικές αποφάσεις και παρέχονται οι υπηρεσίες περίθαλψης. Η συνεχής αύξηση του όγκου βιοϊατρικών δεδομένων, σε συνδυασμό με την πρόοδο της υπολογιστικής ισχύος και των τεχνικών μηχανικής μάθησης, δημιούργησε το κατάλληλο υπόβαθρο για την ανάπτυξη προηγμένων ευφύων συστημάτων ικανών να υποστηρίξουν τους επαγγελματίες υγείας σε σύνθετες διαδικασίες διάγνωσης και θεραπείας.

Σε αντίθεση με παλαιότερα πληροφοριακά συστήματα, τα οποία λειτουργούσαν κυρίως ως εργαλεία αποθήκευσης και ανάκτησης δεδομένων, τα σύγχρονα συστήματα TN είναι σε θέση να αναλύουν πρότυπα, να εντοπίζουν ανωμαλίες και να παράγουν προβλέψεις βασισμένες σε τεράστιες ποσότητες πληροφορίας. Η μετάβαση αυτή σηματοδοτεί μια ουσιαστική αλλαγή στη φιλοσοφία της ιατρικής πληροφορικής, καθώς τα συστήματα TN δεν περιορίζονται πλέον σε βοηθητικό ρόλο, αλλά συμμετέχουν ενεργά στη διαδικασία λήψης αποφάσεων.

Οι Panagoulis et al. (2023a) επισημαίνουν ότι τα μεγάλα γλωσσικά μοντέλα και ειδικότερα το ChatGPT παρουσιάζουν σημαντικές δυνατότητες στην υποστήριξη ιατρικών διαδικασιών. Τα μοντέλα αυτά μπορούν να επεξεργάζονται ιατρικά κείμενα, να συνοψίζουν κλινικές πληροφορίες και να παρέχουν απαντήσεις σε ερωτήματα φυσικής γλώσσας. Η δυνατότητα αυτή διευκολύνει την πρόσβαση των επαγγελματιών υγείας σε ιατρική γνώση και επιτρέπει την ταχύτερη αξιολόγηση μεγάλου όγκου πληροφοριών.

Παρόλα αυτά, η χρήση της TN στην υγεία συνοδεύεται από αυξημένες απαιτήσεις αξιοπιστίας. Σε αντίθεση με άλλους τομείς εφαρμογής, ακόμη και ένα μικρό ποσοστό σφάλματος μπορεί να έχει σοβαρές συνέπειες για την ανθρώπινη ζωή. Ένα λανθασμένο αποτέλεσμα σε ένα recommendation system ψυχαγωγίας μπορεί να θεωρηθεί αμελητέο· μια λανθασμένη ιατρική πρόβλεψη όμως μπορεί να οδηγήσει σε καθυστερημένη διάγνωση ή ακατάλληλη θεραπεία. Για τον λόγο αυτό, η ιατρική TN αποτελεί έναν από τους πλέον απαιτητικούς τομείς ανάπτυξης αξιόπιστων συστημάτων Εξηγήσιμης Τεχνητής Νοημοσύνης (explainable AI).

5.2 Μεγάλα Γλωσσικά Μοντέλα (Large Language Models – LLMs) και ιατρική διάγνωση

Η ανάπτυξη των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models – LLMs) δημιούργησε νέες προοπτικές για την αυτοματοποίηση και υποστήριξη της ιατρικής διάγνωσης. Τα LLMs έχουν τη δυνατότητα να αναλύουν φυσική γλώσσα, να συσχετίζουν ιατρικές έννοιες και να επεξεργάζονται πληροφορίες από διαφορετικές πηγές δεδομένων. Η ικανότητα αυτή τα καθιστά ιδιαίτερα χρήσιμα σε περιβάλλοντα όπου οι επαγγελματίες υγείας καλούνται να διαχειριστούν μεγάλο όγκο πληροφορίας σε περιορισμένο χρονικό διάστημα.

Οι Panagoulas et al. (2025b) προτείνουν framework αξιολόγησης και fine-tuning LLMs για πολυτροπική ιατρική διάγνωση. Η έρευνα βασίζεται στην ιδέα ότι η διάγνωση δεν εξαρτάται από μία μόνο μορφή δεδομένων, αλλά από τον συνδυασμό διαφορετικών πληροφοριών, όπως ιατρικές εικόνες, εργαστηριακές εξετάσεις, ιστορικό ασθενούς και κλινικές παρατηρήσεις. Η Πολυτροπική Τεχνητή Νοημοσύνη (multimodal AI) επιδιώκει να ενοποιήσει αυτές τις πηγές δεδομένων, δημιουργώντας πιο ολοκληρωμένα και ακριβή συστήματα υποστήριξης αποφάσεων.

Η χρήση LLMs στην ιατρική παρουσιάζει σημαντικά πλεονεκτήματα, κυρίως ως προς την ταχύτητα επεξεργασίας γνώσης και τη δυνατότητα αλληλεπίδρασης μέσω φυσικής γλώσσας. Ένας ιατρός μπορεί να διατυπώσει σύνθετα ερωτήματα και να λάβει άμεσα οργανωμένες πληροφορίες, χωρίς να απαιτείται αναζήτηση σε πολλαπλές βάσεις δεδομένων ή επιστημονικές δημοσιεύσεις.

Ωστόσο, τα ίδια χαρακτηριστικά που καθιστούν τα LLMs ιδιαίτερα ισχυρά δημιουργούν και σημαντικούς κινδύνους. Οι Panagoulas et al. (2023a) επισημαίνουν ότι τα μοντέλα αυτά εξακολουθούν να παρουσιάζουν hallucinations, δηλαδή παραγωγή ανακριβών ή ανύπαρκτων πληροφοριών με ιδιαίτερα πειστικό τρόπο. Στον χώρο της ιατρικής, το πρόβλημα αυτό αποκτά κρίσιμη σημασία, καθώς μια λανθασμένη πληροφορία μπορεί να επηρεάσει άμεσα τη θεραπευτική διαδικασία.

Για τον λόγο αυτό, η χρήση των LLMs στην υγεία προτείνεται κυρίως ως υποστηρικτικό εργαλείο και όχι ως αυτόνομος μηχανισμός λήψης αποφάσεων. Η ανθρώπινη εποπτεία παραμένει απαραίτητη, ιδιαίτερα σε περιπτώσεις υψηλού κλινικού κινδύνου.

5.3 Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) και αξιοπιστία ιατρικών συστημάτων

Η Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) αποτελεί μία από τις σημαντικότερες απαιτήσεις των σύγχρονων ιατρικών εφαρμογών. Σε αντίθεση με άλλους τομείς όπου ένα αδιαφανές μοντέλο μπορεί να θεωρηθεί αποδεκτό εφόσον εμφανίζει υψηλή απόδοση, στην ιατρική η δυνατότητα κατανόησης της απόφασης είναι εξίσου σημαντική με την ακρίβεια του αποτελέσματος.

Οι Panagoulas et al. (2023b) εξετάζουν το επίπεδο εξηγησιμότητας που απαιτείται σε ιατρικά συστήματα TN, εφαρμόζοντας τη θεωρία της Διάχυσης της των Καινοτομιών (Diffusion of Innovations Theory). Η μελέτη αναδεικνύει ότι όσο μεγαλύτερος είναι ο κίνδυνος μιας απόφασης, τόσο αυξάνεται η ανάγκη επεξήγησης του τρόπου με τον οποίο το σύστημα κατέληξε στο αποτέλεσμα.

Η εξηγησιμότητα είναι απαραίτητη για πολλούς λόγους. Πρώτον, επιτρέπει στους επαγγελματίες υγείας να αξιολογούν την αξιοπιστία των προβλέψεων του μοντέλου. Δεύτερον, διευκολύνει τη λογοδοσία και την αιτιολόγηση των αποφάσεων προς τον ασθενή. Τρίτον, συμβάλλει στον εντοπισμό αλγοριθμικών προκαταλήψεων και πιθανών σφαλμάτων εκπαίδευσης.

Ένα black-box σύστημα μπορεί να παράγει ιδιαίτερα ακριβείς προβλέψεις, όμως εάν ο ιατρός δεν είναι σε θέση να κατανοήσει τη λογική της απόφασης, η αποδοχή του συστήματος γίνεται ιδιαίτερα δύσκολη. Η ιατρική πράξη βασίζεται σε μεγάλο βαθμό στην εμπιστοσύνη και στην επαγγελματική ευθύνη· συνεπώς, η αδυναμία ερμηνείας μιας αλγοριθμικής πρόβλεψης δημιουργεί σοβαρά ηθικά και πρακτικά προβλήματα.

Η Εξηγήσιμη Τεχνητή Νοημοσύνη (explainable AI) επομένως δεν λειτουργεί μόνο ως τεχνικός μηχανισμός διαφάνειας, αλλά και ως εργαλείο ενίσχυσης της εμπιστοσύνης μεταξύ ανθρώπου και συστήματος TN.

5.4 Εξατομικευμένη διατροφή και Προγνωστική Αναλυτική (Predictive Analytics)

Η εξατομικευμένη διατροφή αποτελεί έναν από τους πιο δυναμικούς τομείς εφαρμογής της βιοϊατρικής TN. Η παραδοσιακή προσέγγιση της διατροφής βασιζόταν κυρίως σε γενικές οδηγίες που απευθύνονταν στον συνολικό πληθυσμό. Ωστόσο, οι σύγχρονες έρευνες δείχνουν ότι οι διατροφικές ανάγκες διαφοροποιούνται σημαντικά μεταξύ των ατόμων λόγω γενετικών, μεταβολικών και περιβαλλοντικών παραγόντων.

Οι Panagoulas et al. (2021) αναλύουν τη χρήση Βαθιάς Μάθησης (Deep Learning – DL) και βιοδεικτών για τη δημιουργία συστημάτων personalized nutrition. Μέσω Μηχανικής Μάθησης (Machine Learning – ML), τα συστήματα μπορούν να εντοπίζουν πρότυπα που σχετίζονται με τον μεταβολισμό, τη διατροφική συμπεριφορά και την πιθανότητα εμφάνισης ασθενειών.

Η ανάλυση βιοδεικτών επιτρέπει τη δημιουργία προδιαγνωστικών μοντέλων που μπορούν να προσαρμόζουν τις διατροφικές συστάσεις στις πραγματικές ανάγκες κάθε ατόμου. Η δυνατότητα αυτή θεωρείται ιδιαίτερα σημαντική για τη διαχείριση χρόνιων νοσημάτων, όπως ο διαβήτης και τα καρδιαγγειακά προβλήματα.

Οι Panagoulas, Tsihrintzis και Virvou (2025a) επισημαίνουν ότι η TN μπορεί να συμβάλει όχι μόνο στη θεραπεία αλλά και στην πρόληψη ασθενειών, μέσω έγκαιρης ανίχνευσης διατροφικών κινδύνων και προσωποποιημένων προτάσεων παρέμβασης.

Η εξατομικευμένη διατροφή αποτελεί χαρακτηριστικό παράδειγμα μετάβασης από τη γενικευμένη ιατρική στη Ιατρική Ακριβείας (Precision Medicine), όπου οι αποφάσεις βασίζονται στα ιδιαίτερα χαρακτηριστικά κάθε ασθενούς.

5.5 Ανάλυση Ακραίων Τιμών (Extreme Value Analysis – EVA) και ανάλυση βιοϊατρικών δεδομένων

Η ανάλυση ακραίων τιμών (Extreme Value Analysis – EVA) αποτελεί σημαντική τεχνική στη μελέτη βιοϊατρικών και διατροφικών δεδομένων. Σε πολλές περιπτώσεις, τα πιο κρίσιμα ιατρικά περιστατικά σχετίζονται όχι με τις μέσες τιμές, αλλά με ακραίες αποκλίσεις και σπάνια γεγονότα.

Οι Panagoulas, Tsihrintzis και Virvou (2025b) εφαρμόζουν EVA σε διατροφικά δεδομένα, με στόχο την ανάλυση ακραίων μεταβολικών και διατροφικών συμπεριφορών. Η τεχνική αυτή επιτρέπει την αναγνώριση περιπτώσεων που δύσκολα εντοπίζονται μέσω συμβατικών στατιστικών προσεγγίσεων.

Η σημασία της EVA είναι ιδιαίτερα μεγάλη σε ιατρικά περιβάλλοντα, καθώς οι ακραίες περιπτώσεις συχνά σχετίζονται με υψηλό κίνδυνο εμφάνισης σοβαρών παθολογικών καταστάσεων. Η δυνατότητα πρόβλεψης τέτοιων συμβάντων μέσω machine learning μπορεί να συμβάλει στην έγκαιρη παρέμβαση και πρόληψη.

Η ενσωμάτωση Προηγμένης Αναλυτικής (Advanced Analytics) και Προγνωστικής Τεχνητής Νοημοσύνης (predictive AI) δημιουργεί νέες δυνατότητες κατανόησης σύνθετων βιοϊατρικών φαινομένων που μέχρι πρόσφατα θεωρούνταν δύσκολο να αναλυθούν υπολογιστικά.

5.6 Microservices και αξιόπιστη Ιατρική ΤΝ

Η αυξανόμενη πολυπλοκότητα των συστημάτων Ιατρικής ΤΝ δημιουργεί ανάγκη ανάπτυξης πιο ευέλικτων και αξιόπιστων αρχιτεκτονικών λογισμικού. Οι Panagoulas, Tsihrintzis και Virvou (2025c) προτείνουν επαναληπτική Αρχιτεκτονική Μικροϋπηρεσιών (microservices) για Εξηγήσιμα και Αξιόπιστα Ιατρικά Συστήματα Τεχνητής Νοημοσύνης .

Η προσέγγιση Αρχιτεκτονικής Μικροϋπηρεσιών (microservices) βασίζεται στον διαχωρισμό του συστήματος σε μικρότερα ανεξάρτητα υποσυστήματα, καθένα από τα οποία εκτελεί συγκεκριμένη λειτουργία. Η αρχιτεκτονική αυτή επιτρέπει καλύτερη συντήρηση, ευκολότερη αναβάθμιση και μεγαλύτερη αξιοπιστία.

Στον χώρο της υγείας, η modular προσέγγιση θεωρείται ιδιαίτερα σημαντική επειδή επιτρέπει την απομόνωση σφαλμάτων και τη σαφέστερη κατανόηση της λειτουργίας κάθε επιμέρους υποσυστήματος. Παράλληλα, ενισχύει τη δυνατότητα explainability, καθώς κάθε component μπορεί να αναλυθεί και να αξιολογηθεί ξεχωριστά.

Η ανάπτυξη Αξιόπιστων Ιατρικών Συστημάτων απαιτεί επομένως όχι μόνο ισχυρούς αλγορίθμους, αλλά και κατάλληλες αρχιτεκτονικές λογισμικού που να υποστηρίζουν διαφάνεια, επεκτασιμότητα και ασφαλή λειτουργία.

5.7 Συμπεράσματα Κεφαλαίου

Η Τεχνητή Νοημοσύνη μετασχηματίζει ριζικά τον χώρο της υγείας και της βιοϊατρικής, εισάγοντας νέες δυνατότητες διάγνωσης, πρόβλεψης και εξατομίκευσης υπηρεσιών υγείας. Η ανάπτυξη των LLMs και των προγνωστικών μοντέλων δημιουργεί ισχυρά εργαλεία υποστήριξης των επαγγελματιών υγείας και συμβάλλει στη μετάβαση προς πιο προσωποποιημένες μορφές ιατρικής φροντίδας.

Παράλληλα, η αξιοπιστία, η εξηγησιμότητα και η ανθρώπινη εποπτεία παραμένουν κρίσιμες απαιτήσεις. Οι Ψευδαισθήσεις (hallucinations), οι αλγοριθμικές προκαταλήψεις και η αδιαφάνεια των μοντέλων αναδεικνύουν ότι η πλήρης αυτονομία της ΤΝ στην ιατρική εξακολουθεί να αποτελεί σημαντική πρόκληση.

Η μελλοντική εξέλιξη της ιατρικής ΤΝ αναμένεται να βασιστεί στη συνεργασία ανθρώπου και ευφυών συστημάτων, με στόχο τη δημιουργία αξιόπιστων και ανθρωποκεντρικών Οικοσυστημάτων Βιοϊατρικής Τεχνητής Νοημοσύνης (Biomedical AI Ecosystems) .

6. Τεχνικές και Αλγοριθμικές Βάσεις της Τεχνητής Νοημοσύνης

6.1 Η εξέλιξη των αλγοριθμικών προσεγγίσεων στην ΤΝ

Η σύγχρονη Τεχνητή Νοημοσύνη βασίζεται σε ένα ευρύ σύνολο αλγοριθμικών και υπολογιστικών τεχνικών, οι οποίες εξελίχθηκαν σημαντικά κατά τις τελευταίες δεκαετίες. Από τα πρώτα rule-based expert systems έως τα σημερινά μεγάλα γλωσσικά μοντέλα και τα πολυτροπικά συστήματα μάθησης, η ΤΝ έχει μετατραπεί σε έναν ιδιαίτερα σύνθετο και διεπιστημονικό τομέα.

Η αρχική ανάπτυξη της ΤΝ στηρίχθηκε κυρίως σε συμβολικές προσεγγίσεις, όπου η γνώση αναπαριστάτο μέσω κανόνων και λογικών δομών. Τα συστήματα αυτά μπορούσαν να επιλύουν συγκεκριμένα προβλήματα, αλλά παρουσίαζαν περιορισμένη δυνατότητα προσαρμογής σε νέα δεδομένα ή άγνωστες καταστάσεις. Η εμφάνιση της Μηχανικής μάθησης (Machine Learning) άλλαξε ριζικά αυτή την προσέγγιση, επιτρέποντας στα υπολογιστικά συστήματα να «μαθαίνουν» από δεδομένα αντί να βασίζονται αποκλειστικά σε προκαθορισμένους κανόνες.

Σύμφωνα με τους Tsihrintzis et al. (2024a), η εξέλιξη αυτή οδήγησε στη δημιουργία AI-empowered decision support systems, τα οποία είναι σε θέση να αναλύουν πολύπλοκα δεδομένα και να υποστηρίζουν διαδικασίες λήψης αποφάσεων σε δυναμικά περιβάλλοντα. Η μετάβαση αυτή θεωρείται ιδιαίτερα σημαντική, καθώς επιτρέπει στα σύγχρονα

συστήματα TN να λειτουργούν σε πραγματικά περιβάλλοντα με υψηλό επίπεδο αβεβαιότητας.

Παράλληλα, η πρόοδος της υπολογιστικής ισχύος και η αύξηση της διαθεσιμότητας δεδομένων συνέβαλαν στην ανάπτυξη πιο σύνθετων μοντέλων, όπως τα deep neural networks, τα reinforcement learning systems και τα multimodal AI architectures.

6.2 Μηχανική Μάθηση (Machine Learning) και Βαθιά Μάθηση (Deep Learning – DL)

Η μηχανική μάθηση αποτελεί τον πυρήνα της σύγχρονης TN. Η βασική της αρχή είναι ότι τα συστήματα μπορούν να εξάγουν πρότυπα και γνώση μέσω της ανάλυσης δεδομένων, χωρίς να απαιτείται ρητός προγραμματισμός για κάθε πιθανό σενάριο.

Οι αλγόριθμοι Μηχανικής Μάθησης (Machine Learning) διακρίνονται συνήθως σε προσεγγίσεις Επιβλεπόμενης Μάθησης (supervised learning), Μη Επιβλεπόμενης Μάθησης (unsupervised learning) και Ενισχυτικής Μάθησης (reinforcement learning). Στη Επιβλεπόμενη Μάθηση το μοντέλο εκπαιδεύεται χρησιμοποιώντας δεδομένα με γνωστές απαντήσεις, ενώ στην Μη Επιβλεπόμενη Μάθηση επιχειρεί να εντοπίσει κρυφές δομές ή συσχετίσεις χωρίς προκαθορισμένες ετικέτες.

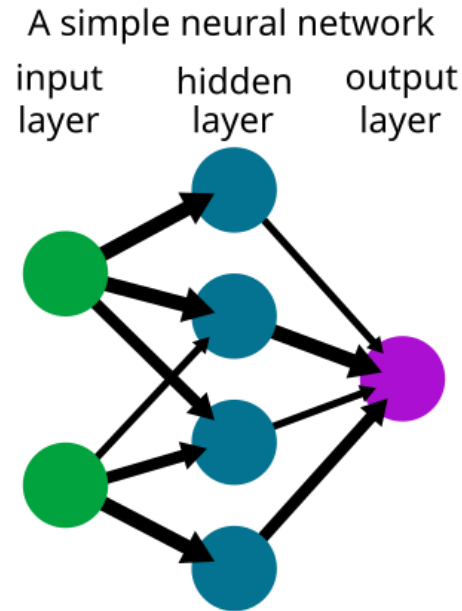
Η ανάπτυξη της Βαθιάς Μάθησης (Deep Learning – DL) αποτέλεσε μία από τις σημαντικότερες τεχνολογικές εξελίξεις της τελευταίας δεκαετίας. Τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks – DNNs) βασίζονται σε πολυεπίπεδες αρχιτεκτονικές τεχνητών νευρωνικών δικτύων, οι οποίες επιτρέπουν την επεξεργασία ιδιαίτερα πολύπλοκων δεδομένων, όπως εικόνες, ομιλία και φυσική γλώσσα.

Οι Panagoulis et al. (2021) εφαρμόζουν Βαθιά Μάθηση (Deep Learning – DL) σε βιοϊατρικές εφαρμογές και συγκεκριμένα σε συστήματα Εξατομικευμένης Διατροφής (personalized nutrition) βασισμένα σε διατροφικούς βιοδείκτες. Η έρευνα δείχνει ότι τα deep neural networks μπορούν να εντοπίζουν πολύπλοκες σχέσεις μεταξύ βιοδεικτών και διατροφικών προτύπων, επιτρέποντας τη δημιουργία εξατομικευμένων Συστημάτων Συστάσεων (recommendation systems).

Η ισχύς των μοντέλων Βαθιάς Μάθησης οφείλεται κυρίως στην ικανότητά τους να εξάγουν αυτόματα χαρακτηριστικά από τα δεδομένα. Σε αντίθεση με παλαιότερες τεχνικές, όπου η εξαγωγή χαρακτηριστικών γινόταν χειροκίνητα, τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks – DNNs) μπορούν να ανακαλύπτουν πολυεπίπεδες αναπαραστάσεις πληροφορίας με ελάχιστη ανθρώπινη παρέμβαση.

Παρόλα αυτά, η αυξανόμενη πολυπλοκότητα των μοντέλων Βαθιάς Μάθησης δημιουργεί σημαντικά προβλήματα explainability. Τα περισσότερα Βαθιά Νευρωνικά Δίκτυα (Deep

Neural Networks – DNNs) λειτουργούν ως black-box systems, γεγονός που καθιστά δύσκολη την κατανόηση της διαδικασίας λήψης αποφάσεων.



Εικόνα 2: Παράδειγμα τεχνητού νευρωνικού δικτύου

6.3 Ενισχυτική Μάθηση (Reinforcement Learning – RL) και αυτόνομα συστήματα

Η Ενισχυτική Μάθηση (Reinforcement Learning – RL) αποτελεί μία από τις σημαντικότερες προσεγγίσεις της σύγχρονης ΤΝ, ιδιαίτερα σε περιβάλλοντα όπου απαιτείται λήψη αποφάσεων μέσω αλληλεπίδρασης με το περιβάλλον.

Σε αντίθεση με τη supervised learning, όπου το μοντέλο εκπαιδεύεται μέσω παραδειγμάτων με γνωστές απαντήσεις, στο RL ο αλγόριθμος μαθαίνει μέσω δοκιμής και σφάλματος. Ο agent εκτελεί ενέργειες, λαμβάνει feedback από το περιβάλλον και σταδιακά βελτιστοποιεί τη συμπεριφορά του με στόχο τη μεγιστοποίηση της συνολικής ανταμοιβής.

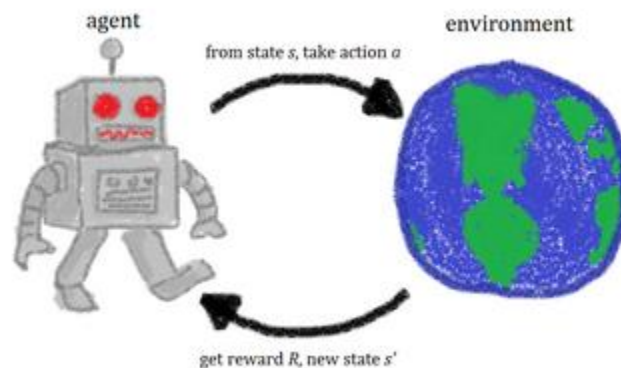
Οι Almalis et al. (2025a) παρουσιάζουν ένα αποκεντρωμένο πλαίσιο ενισχυτικής μάθησης (Decentralized Reinforcement Learning Framework) για επιχειρήσεις μη επανδρωμένων εναέριων οχημάτων (UAV) σε περιβάλλοντα χωρίς διαθεσιμότητα GNSS (GNSS-denied environments). Η έρευνα εστιάζει στην πλοήγηση και παρακολούθηση μη επανδρωμένων αεροσκαφών σε συνθήκες όπου δεν υπάρχει πρόσβαση σε συμβατικά συστήματα γεωεντοπισμού.

Η σημασία του RL σε τέτοιες εφαρμογές είναι ιδιαίτερα μεγάλη, καθώς τα αυτόνομα συστήματα καλούνται να λαμβάνουν αποφάσεις σε δυναμικά και αβέβαια περιβάλλοντα. Η Ενισχυτική Μάθηση (Reinforcement Learning – RL) επιτρέπει στους agents να προσαρμόζονται συνεχώς σε νέες συνθήκες και να αναπτύσσουν στρατηγικές συμπεριφοράς που δεν έχουν προγραμματιστεί εκ των προτέρων.

Η χρήση RL επεκτείνεται πλέον σε πλήθος εφαρμογών, όπως:

- αυτόνομα οχήματα,
- ρομποτικά συστήματα,
- έξυπνα ενεργειακά δίκτυα,
- recommendation systems,
- στρατηγικά παιχνίδια.

Ωστόσο, η αυτονομία των συστημάτων RL δημιουργεί και νέες προκλήσεις, ιδιαίτερα σε ζητήματα ασφάλειας, προβλεψιμότητας και ανθρώπινου ελέγχου. Ένα reinforcement learning μοντέλο μπορεί να αναπτύξει απρόβλεπτες στρατηγικές εάν ο objective function δεν έχει σχεδιαστεί σωστά, γεγονός που καθιστά απαραίτητη την ανάπτυξη trustworthy RL architectures.



Εικόνα 3: Αλληλεπίδραση reinforcement learning agent με το περιβάλλον.

6.4 Πολυτροπική ΤΝ και ενοποίηση δεδομένων

Η Πολυτροπική Τεχνητή Νοημοσύνη (Multimodal AI) αποτελεί μία από τις πλέον εξελιγμένες μορφές σύγχρονης ΤΝ. Τα πολυτροπικά συστήματα έχουν τη δυνατότητα να επεξεργάζονται και να συνδυάζουν διαφορετικές μορφές δεδομένων, όπως:

- κείμενο,
- εικόνα,

- ήχο,
- βίντεο,
- αισθητηριακά δεδομένα.

Η σημασία της Πολυτροπικής Τεχνητής Νοημοσύνης (Multimodal AI) είναι ιδιαίτερα μεγάλη επειδή ο πραγματικός κόσμος δεν λειτουργεί μέσω μίας μόνο μορφής πληροφορίας. Ο άνθρωπος αντιλαμβάνεται το περιβάλλον μέσω συνδυασμού πολλαπλών αισθήσεων και γνωστικών μηχανισμών. Η πολυτροπική ΤΝ επιχειρεί να προσεγγίσει αυτή τη διαδικασία μέσω ενοποίησης διαφορετικών καναλιών πληροφορίας.

Οι Panagoulas et al. (2025b) εφαρμόζουν την Πολυτροπική Τεχνητή Νοημοσύνη (Multimodal AI) σε ιατρική διάγνωση, συνδυάζοντας κλινικά δεδομένα, ιατρικές εικόνες και φυσική γλώσσα. Η προσέγγιση αυτή επιτρέπει τη δημιουργία πιο ολοκληρωμένων diagnostic systems και βελτιώνει την ακρίβεια των προβλέψεων.

Παράλληλα, η Πολυτροπική Τεχνητή Νοημοσύνη (Multimodal AI) χρησιμοποιείται εκτενώς και στα σύγχρονα LLMs, τα οποία πλέον δεν περιορίζονται στην επεξεργασία κειμένου αλλά μπορούν να αναλύουν εικόνες, ήχο και βίντεο. Η εξέλιξη αυτή θεωρείται καθοριστική για τη μετάβαση προς πιο «ανθρώπινα» ευφυή συστήματα.

Ωστόσο, η πολυτροπική μάθηση δημιουργεί σημαντικές υπολογιστικές και μεθοδολογικές προκλήσεις. Η ενοποίηση ετερογενών δεδομένων απαιτεί σύνθετες αρχιτεκτονικές και ιδιαίτερα μεγάλους υπολογιστικούς πόρους. Επιπλέον, αυξάνει σημαντικά το πρόβλημα επεξηγησιμότητας, καθώς καθίσταται δυσκολότερη η κατανόηση του τρόπου με τον οποίο κάθε μορφή δεδομένων επηρεάζει την τελική απόφαση.

6.5 Συναισθηματική Υπολογιστική και αναγνώριση συναισθήματος

Η Συναισθηματική Υπολογιστική (Affective Computing) αποτελεί έναν από τους σημαντικότερους τομείς ανθρωποκεντρικής ΤΝ. Βασικός της στόχος είναι η ανάπτυξη συστημάτων ικανών να αναγνωρίζουν, να ερμηνεύουν και να ανταποκρίνονται σε ανθρώπινα συναισθήματα.

Η σημασία της Συναισθηματικής Υπολογιστικής (Affective Computing) σχετίζεται με το γεγονός ότι η ανθρώπινη επικοινωνία δεν βασίζεται αποκλειστικά στη λογική πληροφορία, αλλά και στη συναισθηματική έκφραση. Η δυνατότητα κατανόησης συναισθημάτων θεωρείται επομένως απαραίτητη για τη δημιουργία πιο φυσικών και αποτελεσματικών Συστημάτων Αλληλεπίδρασης Ανθρώπου-Τεχνητής Νοημοσύνης (Human-AI interaction systems).

Οι Lampropoulos και Tsihrintzis (2012a) εξετάζουν την αναγνώριση συναισθήματος από ομιλία μέσω MPEG-7 descriptors. Η ανάλυση φωνητικών χαρακτηριστικών επιτρέπει στα συστήματα να εντοπίζουν συναισθηματικές καταστάσεις, όπως άγχος, χαρά ή θυμό.

Παράλληλα, οι Stathopoulou και Tsihrintzis (2011a) μελετούν την αναγνώριση συναισθήματος μέσω κινήσεων σώματος και χειρονομιών, ενώ οι Stathopoulou και Tsihrintzis (2004a) παρουσιάζουν neural-network-based σύστημα αναγνώρισης εκφράσεων προσώπου.

Η ενοποίηση πολλαπλών μορφών συναισθηματικής πληροφορίας παρουσιάζεται από τους Alepis et al. (2010a), οι οποίοι προτείνουν bi-modal interaction system που συνδυάζει φωνητική και οπτική πληροφορία.

Η Συναισθηματική Υπολογιστική (Affective Computing) χρησιμοποιείται πλέον σε πλήθος εφαρμογών, όπως:

- ευφυή εκπαιδευτικά συστήματα,
- recommendation systems,
- ιατρικές εφαρμογές,
- κοινωνικά ρομπότ,
- conversational agents.

Παρά τις σημαντικές προόδους, η ακριβής κατανόηση ανθρώπινου συναισθήματος εξακολουθεί να αποτελεί ιδιαίτερα δύσκολο πρόβλημα, καθώς τα συναισθήματα επηρεάζονται από πολιτισμικούς, κοινωνικούς και ψυχολογικούς παράγοντες που δύσκολα μοντελοποιούνται υπολογιστικά.

6.6 Ευφυή πολυμεσικά συστήματα και Περιβάλλοντα Συστάσεων (Recommendation Environments)

Η ανάπτυξη ευφυών πολυμεσικών συστημάτων αποτέλεσε σημαντικό στάδιο στην εξέλιξη της Τεχνητής Νοημοσύνης. Οι Tsihrintzis, Virvou και Jain (2013a) αναλύουν υπηρεσίες συστάσεων πολυμεσικού περιεχομένου (multimedia recommendation services) σε ευφυή περιβάλλοντα (intelligent environments), αναδεικνύοντας τον ρόλο των συστημάτων συστάσεων (recommendation systems) στη διαχείριση της πληροφορίας και του περιεχομένου.

Τα συστήματα συστάσεων (recommendation systems) βασίζονται σε Μηχανικής Μάθησης (Machine Learning) τεχνικές και χρησιμοποιούνται ευρέως σε:

- πλατφόρμες ψυχαγωγίας,
- ηλεκτρονικό εμπόριο,

- εκπαιδευτικά περιβάλλοντα,
- μέσα κοινωνικής δικτύωσης (social media) ,
- έξυπνα περιβάλλοντα.

Η βασική λειτουργία τους είναι η ανάλυση συμπεριφοράς χρήστη και η πρόβλεψη προτιμήσεων, με στόχο την παροχή εξατομικευμένων προτάσεων.

Η εξέλιξη των Συστημάτων Συστάσεων (Recommendation Systems) συνδέεται άμεσα με τη μετάβαση προς περισσότερο ανθρωποκεντρικά και προσαρμοστικά συστήματα Τεχνητής Νοημοσύνης (Adaptive AI Systems), τα οποία προσαρμόζονται δυναμικά στις ανάγκες και τις προτιμήσεις των χρηστών. .

6.7 Συμπεράσματα Κεφαλαίου

Οι τεχνικές και αλγοριθμικές βάσεις της σύγχρονης Τεχνητής Νοημοσύνης αποτελούν το θεμέλιο πάνω στο οποίο αναπτύσσονται οι σύγχρονες εφαρμογές Παραγωγικής Τεχνητής Νοημοσύνης (generative AI), έξυπνων πόλεων και βιοϊατρικών συστημάτων.

Η εξέλιξη από τα παραδοσιακά μοντέλα Μηχανικής Μάθησης (Machine Learning Models) προς τα Βαθιά Νευρωνικά Δίκτυα (Deep Neural Networks), τα συστήματα Ενισχυτικής Μάθησης (Reinforcement Learning Systems) και την Πολυτροπική Τεχνητή Νοημοσύνη (Multimodal AI) επέτρεψε τη δημιουργία ολοένα και πιο σύνθετων και αυτόνομων συστημάτων. Παράλληλα, η Συναισθηματική Υπολογιστική (Affective Computing) και τα ανθρωποκεντρικά Περιβάλλοντα Συστάσεων (Recommendation Environments) ανέδειξαν τη σημασία της συναισθηματικής και κοινωνικής διάστασης της αλληλεπίδρασης ανθρώπου-TN.

Ωστόσο, όσο αυξάνεται η πολυπλοκότητα των αλγοριθμικών μοντέλων, τόσο εντονότερες γίνονται οι προκλήσεις explainability, αξιοπιστίας και ηθικής χρήσης. Η μελλοντική εξέλιξη της TN αναμένεται να βασιστεί όχι μόνο στη βελτίωση της απόδοσης των αλγορίθμων, αλλά και στη δημιουργία trustworthy και ανθρωποκεντρικών ευφυών συστημάτων.

7. Συζήτηση

7.1 Η μετάβαση από τη Μηχανική Μάθηση στη Generative AI

Η εξέλιξη της Τεχνητής Νοημοσύνης κατά την τελευταία δεκαετία υπήρξε ιδιαίτερα ραγδαία και πολυδιάστατη. Η μετάβαση από τα παραδοσιακά συστήματα Μηχανικής

Μάθησης (Machine Learning) προς την Παραγωγική Τεχνητή Νοημοσύνη (Generative AI) δεν αποτελεί απλώς τεχνολογική βελτίωση, αλλά ουσιαστική αλλαγή στον τρόπο με τον οποίο τα υπολογιστικά συστήματα αλληλεπιδρούν με τον άνθρωπο και παράγουν γνώση.

Τα πρώτα συστήματα Μηχανικής Μάθησης (Machine Learning) είχαν κυρίως αναλυτικό χαρακτήρα. Ο βασικός τους στόχος ήταν η ταξινόμηση, η πρόβλεψη και η αναγνώριση προτύπων μέσα από δεδομένα. Η Παραγωγική Τεχνητή Νοημοσύνη (Generative AI), αντίθετα, εισάγει μια νέα διάσταση: τη δυνατότητα δημιουργίας περιεχομένου. Τα σύγχρονα Μεγάλα Γλωσσικά Μοντέλα (Large Language Models – LLMs) δεν περιορίζονται πλέον στην αναγνώριση και επεξεργασία πληροφορίας, αλλά είναι σε θέση να παράγουν κείμενο και κώδικα, καθώς και σύνθετες απαντήσεις σε φυσική γλώσσα.

Η αλλαγή αυτή επηρεάζει ουσιαστικά τη σχέση ανθρώπου-μηχανής. Στα παραδοσιακά πληροφοριακά συστήματα, ο άνθρωπος λειτουργούσε ως αποκλειστικός φορέας λήψης αποφάσεων, ενώ το σύστημα είχε βοηθητικό ρόλο. Στο πλαίσιο της σύγχρονης Παραγωγικής Τεχνητής Νοημοσύνης, τα συστήματα αποκτούν περισσότερο «συνεργατικό» ρόλο, συμμετέχοντας ενεργά στη διαδικασία παραγωγής περιεχομένου, υποστήριξης της γνώσης και επίλυσης προβλημάτων.

Η μετάβαση αυτή, ωστόσο, συνοδεύεται από αυξημένη πολυπλοκότητα και νέες μορφές αβεβαιότητας. Όσο περισσότερο τα συστήματα λειτουργούν αυτόνομα και δημιουργικά, τόσο δυσκολότερη γίνεται η πλήρης κατανόηση και ο έλεγχος της συμπεριφοράς τους. Το ζήτημα αυτό αναδείχθηκε επανειλημμένα στα προηγούμενα κεφάλαια, ιδιαίτερα στις εφαρμογές της εκπαίδευσης, της υγείας και της ενέργειας.

7.2 Συγκριτική ανάλυση εφαρμογών της TN

Στον χώρο της εκπαίδευσης, η Παραγωγική Τεχνητή Νοημοσύνη (Generative AI) αξιοποιείται κυρίως ως εργαλείο υποστήριξης της μάθησης, εξατομίκευσης του περιεχομένου και ανάπτυξης διαδραστικών μαθησιακών εμπειριών. Η δυνατότητα των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models – LLMs) να αλληλεπιδρούν μέσω φυσικής γλώσσας δημιουργεί νέες μορφές ηλεκτρονικής μάθησης (e-learning) και ευφών συστημάτων διδασκαλίας (Intelligent Tutoring Systems – ITS). Ωστόσο, οι επιστημικοί κίνδυνοι (Epistemic Risks) και οι ψευδαισθήσεις (Hallucinations) αναδεικνύονται ως ιδιαίτερα σημαντικά ζητήματα, καθώς οι μαθητές συχνά δυσκολεύονται να αξιολογήσουν την αξιοπιστία των παραγόμενων πληροφοριών.

Στον τομέα της υγείας, η TN λειτουργεί κυρίως ως μηχανισμός υποστήριξης κλινικών αποφάσεων και ανάλυσης βιοϊατρικών δεδομένων. Οι εφαρμογές Μηχανικής Μάθησης (Machine Learning) και Πολυτροπικής Τεχνητής Νοημοσύνης (Multimodal AI) προσφέρουν

σημαντικές δυνατότητες βελτίωσης της διάγνωσης και της εξατομίκευσης της θεραπείας. Παράλληλα, όμως, οι απαιτήσεις εξηγησιμότητας (Explainability) και αξιοπιστίας είναι πολύ υψηλότερες συγκριτικά με άλλους τομείς, καθώς τα σφάλματα ενδέχεται να επηρεάσουν άμεσα την ανθρώπινη υγεία και ασφάλεια.

Στις έξυπνες πόλεις και στα ενεργειακά οικοσυστήματα, η ΤΝ χρησιμοποιείται κυρίως για Προγνωστική Αναλυτική (Predictive Analytics), αυτοματοποίηση και βελτιστοποίηση πολύπλοκων υποδομών. Τα συστήματα αυτά επεξεργάζονται μεγάλες ποσότητες δεδομένων σε πραγματικό χρόνο και μπορούν να επηρεάσουν μεγάλο αριθμό πολιτών. Για τον λόγο αυτό, η εμπιστοσύνη, η διαφάνεια και η προστασία των προσωπικών δεδομένων αποκτούν ιδιαίτερη σημασία.

Παρότι οι εφαρμογές διαφέρουν σημαντικά μεταξύ τους, παρατηρούνται ορισμένα κοινά χαρακτηριστικά. Σε όλους τους τομείς αναδεικνύεται η ανάγκη ανάπτυξης αξιόπιστων συστημάτων Τεχνητής Νοημοσύνης (Trustworthy AI Systems), η σημασία της εξηγησιμότητας (Explainability) και η απαίτηση διατήρησης της ανθρώπινης εποπτείας. Η βασική διαφοροποίηση έγκειται κυρίως στον βαθμό κινδύνου και στις πιθανές συνέπειες που συνεπάγεται ένα σφάλμα του συστήματος.

7.3 Εμπιστοσύνη και ανθρώπινος παράγοντας

Η έννοια της εμπιστοσύνης αναδείχθηκε ως κεντρικός άξονας της παρούσας εργασίας. Η αυξανόμενη αυτονομία των συστημάτων ΤΝ καθιστά απαραίτητη τη δημιουργία μηχανισμών που να ενισχύουν τη συνεργασία ανθρώπου-μηχανής και να επιτρέπουν στον χρήστη να αισθάνεται ότι διατηρεί τον έλεγχο της διαδικασίας.

Το μοντέλο VIRTSI που παρουσιάστηκε από τους Virvou et al. (2024a) αποτελεί χαρακτηριστικό παράδειγμα προσπάθειας κατανόησης της δυναμικής εμπιστοσύνης. Η έρευνα έδειξε ότι η εμπιστοσύνη δεν είναι σταθερή μεταβλητή, αλλά μεταβάλλεται ανάλογα με την εμπειρία χρήσης, τις συνέπειες των σφαλμάτων και το επίπεδο εξειδίκευσης του χρήστη.

Η σημασία του ανθρώπινου παράγοντα γίνεται ακόμη πιο εμφανής στις εφαρμογές υψηλού ρίσκου. Στην υγεία, για παράδειγμα, ακόμη και ιδιαίτερα ακριβή συστήματα ΤΝ δεν μπορούν να λειτουργήσουν αποτελεσματικά χωρίς ανθρώπινη αξιολόγηση και κλινική κρίση. Αντίστοιχα, στις έξυπνες πόλεις, οι πολίτες απαιτούν διαφάνεια και δυνατότητα κατανόησης των αλγοριθμικών αποφάσεων που επηρεάζουν την καθημερινότητά τους.

Η εμπιστοσύνη επομένως δεν εξαρτάται αποκλειστικά από την τεχνική απόδοση ενός συστήματος, αλλά και από ψυχολογικούς, κοινωνικούς και ηθικούς παράγοντες. Ένα

σύστημα μπορεί να εμφανίζει υψηλή ακρίβεια, αλλά να αποτυγχάνει να γίνει αποδεκτό εάν οι χρήστες δεν κατανοούν τον τρόπο λειτουργίας του ή αισθάνονται ότι δεν έχουν δυνατότητα παρέμβασης.

7.4 Εξηγησιμότητα και το πρόβλημα των **black-box models**

Η ανάπτυξη των Βαθιών Νευρωνικών Δικτύων (Deep Neural Networks) και των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models – LLMs) συνοδεύτηκε από σημαντική αύξηση της υπολογιστικής πολυπλοκότητας των μοντέλων. Παρότι τα σύγχρονα μοντέλα επιτυγχάνουν ιδιαίτερα υψηλές επιδόσεις, η εσωτερική διαδικασία λήψης αποφάσεων συχνά παραμένει αδιαφανής ακόμη και για τους ίδιους τους δημιουργούς τους.

Το πρόβλημα αυτό οδήγησε στην ανάπτυξη της Εξηγήσιμης Τεχνητής Νοημοσύνης (Explainable Artificial Intelligence – XAI), η οποία επιδιώκει να καταστήσει τις αλγοριθμικές αποφάσεις περισσότερο κατανοητές και διαφανείς. Η εξηγησιμότητα (Explainability) αποτελεί ιδιαίτερα κρίσιμη απαίτηση σε τομείς όπως η ιατρική, η εκπαίδευση και η ενέργεια, όπου η λήψη αποφάσεων μπορεί να συνδέεται άμεσα με την ανθρώπινη ασφάλεια και ευημερία.

Η ανάγκη για εξηγησιμότητα δεν αφορά μόνο τεχνικούς λόγους. Η δυνατότητα κατανόησης των αποφάσεων ενισχύει την εμπιστοσύνη των χρηστών και επιτρέπει τον εντοπισμό αλγοριθμικών προκαταλήψεων ή λανθασμένων προβλέψεων. Παράλληλα, συμβάλλει στη διατήρηση της ανθρώπινης λογοδοσίας, καθώς οι επαγγελματίες που χρησιμοποιούν συστήματα ΤΝ μπορούν να κατανοούν και να αιτιολογούν καλύτερα τις αποφάσεις που λαμβάνονται με την υποστήριξή τους.

Ωστόσο, η εξηγησιμότητα παραμένει ιδιαίτερα δύσκολη σε πολύπλοκα συστήματα Βαθιάς Μάθησης (Deep Learning Systems). Όσο αυξάνεται η πολυπλοκότητα ενός μοντέλου, τόσο δυσκολότερη γίνεται η πλήρης ερμηνεία της λειτουργίας του. Η πρόκληση αυτή αποτελεί ένα από τα σημαντικότερα ανοιχτά ζητήματα της σύγχρονης έρευνας στην Τεχνητή Νοημοσύνη.

7.5 Ηθικές και κοινωνικές προεκτάσεις της ΤΝ

Η ευρεία διάδοση της ΤΝ δημιουργεί σημαντικές κοινωνικές και ηθικές προεκτάσεις. Η αυτοματοποίηση διαδικασιών, η αυξανόμενη χρήση προσωπικών δεδομένων και η δυνατότητα λήψης αποφάσεων από αλγοριθμικά συστήματα επηρεάζουν άμεσα τον τρόπο λειτουργίας της κοινωνίας.

Ένα από τα σημαντικότερα ζητήματα αφορά την ιδιωτικότητα και τη διαχείριση προσωπικών δεδομένων. Στις έξυπνες πόλεις, για παράδειγμα, η συνεχής συλλογή δεδομένων δημιουργεί ανησυχίες σχετικά με την παρακολούθηση πολιτών και την

πιθανότητα κατάχρησης πληροφοριών. Η συμμόρφωση με κανονισμούς όπως ο GDPR θεωρείται πλέον απαραίτητη προϋπόθεση για την αποδοχή των smart city systems.

Παράλληλα, η ανάπτυξη της Παραγωγικής Τεχνητής Νοημοσύνης (Generative AI) δημιουργεί νέα ζητήματα παραπληροφόρησης και επιστημικών κινδύνων (Epistemic Risks). Οι ψευδαισθήσεις (Hallucinations) των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models – LLMs) μπορούν να οδηγήσουν στην παραγωγή πειστικών αλλά ανακριβών ή ανυπόστατων πληροφοριών, γεγονός που αποκτά ιδιαίτερη σημασία σε περιβάλλοντα εκπαίδευσης και υγείας.

Επιπλέον, η αυξανόμενη αυτονομία των συστημάτων δημιουργεί προβληματισμούς σχετικά με τη λογοδοσία και την ευθύνη. Όταν μια απόφαση λαμβάνεται ή υποστηρίζεται από ένα αλγοριθμικό σύστημα, καθίσταται συχνά δύσκολο να προσδιοριστεί με σαφήνεια ποιος φέρει την ευθύνη σε περίπτωση σφάλματος.

Η αντιμετώπιση αυτών των ζητημάτων απαιτεί τον συνδυασμό τεχνικών, νομικών και κοινωνικών προσεγγίσεων. Η μελλοντική ανάπτυξη της TN δεν μπορεί να βασιστεί αποκλειστικά στη βελτίωση της απόδοσης των αλγορίθμων, αλλά πρέπει να συνοδεύεται από ένα ισχυρό ηθικό και ανθρωποκεντρικό πλαίσιο.

7.6 Μελλοντικές κατευθύνσεις της TN

Η μελλοντική εξέλιξη της TN αναμένεται να κινηθεί προς την ανάπτυξη περισσότερο ανθρωποκεντρικών, εξηγήσιμων (Explainable) και αξιόπιστων (Trustworthy) συστημάτων. Η αυξανόμενη χρήση της Πολυτροπικής Τεχνητής Νοημοσύνης (Multimodal AI), της Συναισθηματικής Υπολογιστικής (Affective Computing) και των Προσαρμοστικών Αρχιτεκτονικών (Adaptive Architectures) δείχνει ότι η TN εξελίσσεται προς συστήματα ικανά να λαμβάνουν υπόψη περισσότερες πτυχές της ανθρώπινης επικοινωνίας, συμπεριφοράς και αλληλεπίδρασης.

Παράλληλα, η ενοποίηση διαφορετικών μορφών δεδομένων και η ανάπτυξη πιο προηγμένων μοντέλων συλλογιστικής (Reasoning) αναμένεται να οδηγήσουν σε ακόμη πιο ισχυρά συστήματα υποστήριξης αποφάσεων. Η χρήση Συνθετικών Δεδομένων (Synthetic Data), Ομοσπονδιακής Μάθησης (Federated Learning) και Εξηγήσιμων Αρχιτεκτονικών (Explainable Architectures) θεωρείται επίσης ιδιαίτερα σημαντική για την αντιμετώπιση ζητημάτων ιδιωτικότητας, διαφάνειας και αξιοπιστίας.

Ωστόσο, η συνεχής αύξηση της πολυπλοκότητας των μοντέλων καθιστά αναγκαία την ενίσχυση των μηχανισμών ανθρώπινης εποπτείας. Η πλήρης αυτονομία της TN εξακολουθεί να αποτελεί αμφιλεγόμενο ζήτημα, ιδιαίτερα σε τομείς υψηλού κινδύνου.

Η μελλοντική πρόκληση, επομένως, δεν θα είναι μόνο η δημιουργία «ισχυρότερων» αλγορίθμων, αλλά κυρίως η ανάπτυξη συστημάτων που θα μπορούν να συνεργάζονται αποτελεσματικά, με διαφάνεια και με σεβασμό στις ηθικές αρχές, με τον άνθρωπο.

8. Συμπεράσματα

Η παρούσα εργασία εξερεύνησε την εξέλιξη της Τεχνητής Νοημοσύνης από τις παραδοσιακές τεχνικές Μηχανικής Μάθησης (Machine Learning) έως τη σύγχρονη εποχή της Παραγωγικής Τεχνητής Νοημοσύνης (Generative AI) και των Μεγάλων Γλωσσικών Μοντέλων (Large Language Models – LLMs). Μέσα από τη μελέτη της σύγχρονης βιβλιογραφίας αναδείχθηκε ότι η ΤΝ δεν αποτελεί πλέον ένα περιορισμένο τεχνολογικό εργαλείο εξειδικευμένων εφαρμογών, αλλά μια οριζόντια τεχνολογία που επηρεάζει ολόένα και περισσότερους τομείς της ανθρώπινης δραστηριότητας.

Η μετάβαση από τα κλασικά συστήματα Μηχανικής Μάθησης (Machine Learning Systems) προς τα μοντέλα Βαθιάς Μάθησης (Deep Learning) και την Παραγωγική Τεχνητή Νοημοσύνη άλλαξε ουσιαστικά τον ρόλο των υπολογιστικών συστημάτων. Τα σύγχρονα μοντέλα δεν περιορίζονται στην αναγνώριση προτύπων ή στην εκτέλεση προκαθορισμένων λειτουργιών, αλλά συμμετέχουν ενεργά στην παραγωγή περιεχομένου, στην επεξεργασία και αξιοποίηση της πληροφορίας και στην υποστήριξη της λήψης αποφάσεων. Η δυνατότητα των Μεγάλων Γλωσσικών Μοντέλων να αλληλεπιδρούν μέσω φυσικής γλώσσας και να επεξεργάζονται σύνθετες πληροφορίες δημιούργησε νέες μορφές συνεργασίας ανθρώπου-μηχανής, μετασχηματίζοντας τομείς όπως η εκπαίδευση, η υγεία, η ενέργεια και οι έξυπνες πόλεις.

Στον τομέα της εκπαίδευσης, η Παραγωγική Τεχνητή Νοημοσύνη εισήγαγε νέες δυνατότητες εξατομικευμένης μάθησης και ανάπτυξης Ευφυών Συστημάτων Διδασκαλίας (Intelligent Tutoring Systems – ITS), τα οποία επιτρέπουν την προσαρμογή του εκπαιδευτικού περιεχομένου στις ανάγκες του μαθητή. Παράλληλα, η αξιοποίηση της ΤΝ σε πολιτιστικά περιβάλλοντα και συστήματα ηλεκτρονικής μάθησης (e-learning) ανέδειξε τη δυνατότητα των Μεγάλων Γλωσσικών Μοντέλων να λειτουργούν ως μηχανισμοί υποστήριξης διαδραστικών πολιτισμικών και γνωστικών εμπειριών. Ωστόσο, οι ψευδαισθήσεις (Hallucinations) και οι επιστημικοί κίνδυνοι (Epistemic Risks) αναδείχθηκαν ως ιδιαίτερα σημαντικές προκλήσεις, κυρίως λόγω της πιθανότητας παραγωγής και διάδοσης ανακριβών ή ανυπόστατων πληροφοριών σε μαθησιακά περιβάλλοντα.

Στον χώρο της υγείας και της βιοϊατρικής, η ΤΝ προσφέρει σημαντικές δυνατότητες ανάλυσης βιοϊατρικών δεδομένων, υποστήριξης της διάγνωσης και ανάπτυξης εξατομικευμένων θεραπευτικών προσεγγίσεων. Η χρήση Πολυτροπικής Τεχνητής Νοημοσύνης (Multimodal AI) και Προγνωστικής Αναλυτικής (Predictive Analytics) επιτρέπει τη δημιουργία πιο ολοκληρωμένων συστημάτων υποστήριξης αποφάσεων, ενώ η αξιοποίηση Διατροφικών Βιοδεικτών (Nutritional Biomarkers) και τεχνικών Μηχανικής Μάθησης ενισχύει τη μετάβαση προς την Ιατρική Ακριβείας (Precision Medicine). Παράλληλα, όμως, η ιατρική ΤΝ αναδεικνύει με ιδιαίτερη ένταση την ανάγκη εξηγησιμότητας (Explainability), αξιοπιστίας και ανθρώπινης εποπτείας, καθώς τα σφάλματα σε ιατρικά περιβάλλοντα μπορούν να έχουν άμεσες συνέπειες για την ανθρώπινη υγεία και ασφάλεια.

Οι έξυπνες πόλεις και τα ενεργειακά οικοσυστήματα αποτέλεσαν επίσης σημαντικά πεδία εφαρμογής της ΤΝ. Η χρήση προγνωστικών μοντέλων (Predictive Models), Συνθετικών Δεδομένων (Synthetic Data) και εξηγήσιμων συστημάτων ΤΝ στον ενεργειακό τομέα δημιουργεί νέες δυνατότητες βελτιστοποίησης των αστικών υποδομών και της ενεργειακής διαχείρισης. Ταυτόχρονα, όμως, η συνεχής συλλογή δεδομένων και η αυξανόμενη αυτονομία των αλγοριθμικών συστημάτων δημιουργούν σημαντικές προκλήσεις που σχετίζονται με την ιδιωτικότητα, τη διαφάνεια και την εμπιστοσύνη των πολιτών.

Ένα από τα σημαντικότερα συμπεράσματα της εργασίας είναι ότι η τεχνολογική εξέλιξη της ΤΝ συνοδεύεται από κοινωνικές, ηθικές και επιστημικές προκλήσεις. Η εμπιστοσύνη αναδείχθηκε ως βασικός παράγοντας επιτυχούς αλληλεπίδρασης ανθρώπου-ΤΝ, ενώ μοντέλα όπως το VIRTSL ανέδειξαν ότι η εμπιστοσύνη αποτελεί δυναμική και πολυπαραγοντική διαδικασία, η οποία επηρεάζεται από την εμπειρία χρήσης, τις συνέπειες των σφαλμάτων και το επίπεδο κατανόησης του συστήματος από τον άνθρωπο.

Παράλληλα, η Εξηγήσιμη Τεχνητή Νοημοσύνη (Explainable Artificial Intelligence – XAI) αναδείχθηκε ως μία από τις πλέον κρίσιμες απαιτήσεις της σύγχρονης ΤΝ. Όσο τα μοντέλα γίνονται πιο σύνθετα και αυτόνομα, τόσο αυξάνεται η ανάγκη για διαφάνεια και δυνατότητα κατανόησης των αλγοριθμικών αποφάσεων. Η ανάπτυξη συστημάτων «μαύρου κουτιού» (Black-Box Systems) χωρίς επαρκή δυνατότητα ερμηνείας ενδέχεται να περιορίσει την εμπιστοσύνη των χρηστών και να δημιουργήσει σημαντικά ζητήματα λογοδοσίας και ηθικής ευθύνης.

Η εργασία ανέδειξε επίσης ότι η πλήρης αντικατάσταση του ανθρώπου από την ΤΝ δεν αποτελεί ρεαλιστικό ούτε επιθυμητό στόχο στο παρόν στάδιο εξέλιξης της τεχνολογίας. Αντίθετα, μια περισσότερο βιώσιμη προσέγγιση βασίζεται στην ανάπτυξη συνεργατικών μοντέλων Αλληλεπίδρασης Ανθρώπου-Τεχνητής Νοημοσύνης (Human-AI Interaction), στα οποία η ΤΝ λειτουργεί ως εργαλείο ενίσχυσης των ανθρώπινων δυνατοτήτων και όχι ως πλήρως αυτόνομος μηχανισμός λήψης αποφάσεων.

Όσον αφορά τις μελλοντικές κατευθύνσεις, η έρευνα αναμένεται να εστιάσει σε τομείς όπως η Πολυτροπική Τεχνητή Νοημοσύνη, η Συναισθηματική Υπολογιστική (Affective

Computing), τα Αξιόπιστα Συστήματα Τεχνητής Νοημοσύνης (Trustworthy AI Systems) και οι Εξηγήσιμες Αρχιτεκτονικές (Explainable Architectures). Η ανάπτυξη πιο προηγμένων μηχανισμών συλλογιστικής (Reasoning) και η ενσωμάτωση μηχανισμών αναγνώρισης και επεξεργασίας συναισθηματικών και κοινωνικών σημάτων στα ευφυή συστήματα ενδέχεται να οδηγήσουν σε περισσότερο φυσικές και αποτελεσματικές μορφές αλληλεπίδρασης ανθρώπου-μηχανής.

Τέλος, καθίσταται σαφές ότι η μελλοντική εξέλιξη της Τεχνητής Νοημοσύνης δεν θα εξαρτηθεί αποκλειστικά από την αύξηση της υπολογιστικής ισχύος ή τη βελτίωση των αλγορίθμων. Η ουσιαστική πρόκληση θα είναι η δημιουργία συστημάτων που να είναι αξιόπιστα, διαφανή, ανθρωποκεντρικά και κοινωνικά αποδεκτά. Η επιτυχία της ΤΝ στο μέλλον θα καθοριστεί όχι μόνο από τις δυνατότητες των αλγορίθμων, αλλά κυρίως από τον βαθμό στον οποίο οι άνθρωποι θα μπορούν να συνεργάζονται με τα συστήματα αυτά με ασφάλεια, κατανόηση και εμπιστοσύνη.

Βιβλιογραφία

Panagoulas, D. P., Palamidas, A., Virvou, M., & Tsihrintzis, G. A. (2025). *A framework for evaluation and requirement extraction for fine-tuning of large language models in multimodal medical diagnosis*. *Knowledge-Based Systems*, 326, 113975. **(Panagoulas et al., 2025b)**

Virvou, M., Tsihrintzis, G. A., Marinakis, V., Sarvas, E., Panagoulas, D. P., & Tsihrintzi, E. A. (2025). *Trustworthy integration of generative AI and LLMs in the energy digital spine: Epistemic risk models and trust performance indicators*. In *2025 IEEE 37th International Conference on Tools with Artificial Intelligence (ICTAI)* (pp. 130–138). IEEE. **(Virvou et al., 2025a)**

Virvou, M., Tsihrintzis, G. A., & Tsihrintzi, E. A. (2025). *Hallucinations in generative AI: Epistemic risks for learners in educational applications*. In *2025 16th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–8). IEEE. **(Virvou et al., 2025c)**

Virvou, M., Tsihrintzis, G. A., Chrysafiadi, K., Sakkopoulos, E., & Tsihrintzi, E. A. (2025). *Bi-directional 21st century skills for generative AI in educational software: A requirements engineering perspective*. In *2025 16th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–10). IEEE. **(Virvou et al., 2025b)**

Virvou, M., Tsihrintzis, G. A., & Alepis, E. (2025). *Emotional advice compared between an affective intelligent authoring tool and generative AI: A cognitive walkthrough in medical mobile tutor and ChatGPT*. In *2025 16th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–8). IEEE. **(Virvou et al., 2025d)**

Virvou, M., & Tsihrintzis, G. A. (2024). *VIRTISI: A novel trust dynamics model enhancing artificial intelligence collaboration with human users – Insights from a ChatGPT evaluation study*. *Information Sciences*, 675, 120759. **(Virvou & Tsihrintzis, 2024a)**

Virvou, M., & Tsihrintzis, G. A. (2024). *Impact of consequences on human trust dynamics in artificial intelligence responses*. In *2024 15th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–6). IEEE. **(Virvou & Tsihrintzis, 2024b)**

Virvou, M., & Tsihrintzis, G. A. (2023). *Pre-made empowering artificial intelligence and ChatGPT: The growing importance of human AI-experts*. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–8). IEEE. **(Virvou & Tsihrintzis, 2023a)**

Virvou, M., & Tsihrintzis, G. A. (2023). *A novel trust state-chart model for requirements engineering of trustful AI-empowered software*. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–6). IEEE. **(Virvou & Tsihrintzis, 2023b)**

Virvou, M., & Tsihrintzis, G. A. (2023). *Is ChatGPT beneficial to education? A holistic evaluation framework based on intelligent tutoring systems*. In *2023 14th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1–8). IEEE. **(Virvou & Tsihrintzis, 2023c)**

Panagoulas, D. P., Sarvas, E., Marinakis, V., Virvou, M., & Tsihrintzis, G. A. (2025). *Explainable optimization of extreme value analysis for photovoltaic prediction: Introducing dynamic correlation shifts and weighted benchmarking*. *Electronics*, 14(22), 4484. **(Panagoulas et al., 2025a)**

Panagoulas, D. P., Sotiropoulos, D., Chrysafiadi, K., Tsihrintzi, E. A., Sakkopoulos, E., Tsihrintzis, G. A., & Virvou, M. (2025). *LYRICEL: Knowledge graphs combined with large language models and machine learning for cross-cultural analysis of lyrics – The case of Greek songs*. *IEEE Access*. **(Panagoulas et al., 2025c)**

Panagoulas, D. P., Tsihrintzis, G. A., & Virvou, M. (2025). *Personalized nutrition applications using nutritional biomarkers and machine learning*. In *Artificial Intelligence-Empowered Bio-medical Applications* (Vol. 49). Springer Nature Switzerland. **(Panagoulas, Tsihrintzis & Virvou, 2025a)**

Panagoulas, D. P., Tsihrintzis, G. A., & Virvou, M. (2025). *Extreme value analysis applied in dietary data*. In *Artificial Intelligence-Empowered Bio-medical Applications* (Vol. 49). Springer Nature Switzerland. **(Panagoulas, Tsihrintzis & Virvou, 2025b)**

Panagoulas, D. P., Tsihrintzis, G. A., & Virvou, M. (2025). *Iterative microservices approach for explainable and reliable AI in medical applications*. In *Artificial Intelligence-Empowered Bio-medical Applications* (Vol. 49). Springer Nature Switzerland. **(Panagoulas, Tsihrintzis & Virvou, 2025c)**

Tsihrintzis, G. A., Sarvas, E., Marinakis, V., Panagoulas, D. P., Tsihrintzi, E. A., & Virvou, M. (2025). *Which AI and how trusted by human energy stakeholders? Comparing generative AI ChatGPT with domain specific AI-ENERGIA-SYS*. In *2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 7066–7073). IEEE. **(Tsihrintzis et al., 2025a)**

Tsihrintzis, G. A., Sarvas, E., Marinakis, V., Panagoulas, D. P., Tsihrintzi, E. A., & Virvou, M. (2025). *Energy urban domain: Personalized evaluation of expert and non-expert stakeholder interaction with artificial intelligence through ChatGPT using the VIRTSL model*. *IEEE Access*. **(Tsihrintzis et al., 2025b)**

Tsihrintzis, G. A., Virvou, M., Doukas, H., & Jain, L. C. (2024). *Introduction to advances in artificial intelligence-empowered decision support systems*. In *Advances in Artificial Intelligence-Empowered Decision Support Systems* (pp. 1–11). Springer Nature Switzerland. **(Tsihrintzis et al., 2024a)**

Tsihrintzis, G. A., Marinakis, V., Virvou, M., & Jain, L. C. (2026). *Introduction to human-smart city interactions and user-citizen experiences*. In *Human-Smart City Interactions and User-Citizen Experiences* (pp. 1–13). Springer Nature Switzerland. **(Tsihrintzis et al., 2026a)**