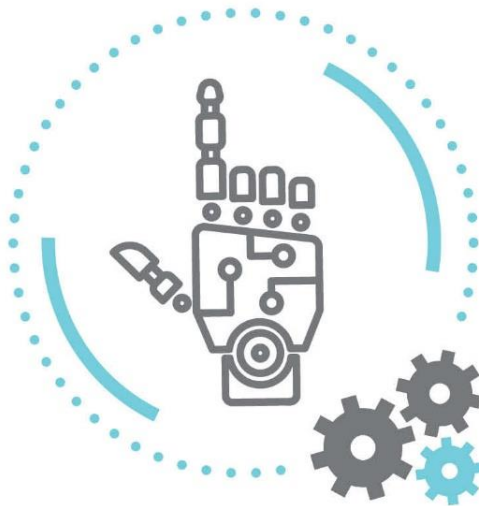


Εντοπισμός και ανάλυση
των ελληνικών επιστημονικών
δημοσιεύσεων στον τομέα της
Τεχνητής Νοημοσύνης με τεχνικές
Μηχανικής Μάθησης



Εντοπισμός και ανάλυση
των ελληνικών επιστημονικών
δημοσιεύσεων στον τομέα της
Τεχνητής Νοημοσύνης με τεχνικές
Μηχανικής Μάθησης



ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ &
ΗΛΕΚΤΡΟΝΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ

Η παρούσα έκδοση υλοποιήθηκε από το Εθνικό Κέντρο Τεκμηρίωσης & Ηλεκτρονικού Περιεχομένου (ΕΚΤ), Εθνική αρχή του Ελληνικού Στατιστικού Συστήματος.

Την επιστημονική επίβλεψη της έκδοσης είχε η Δρ. Εύη Σαχίνη, Διευθύντρια του ΕΚΤ.

Στον εντοπισμό και την ανάλυση των δεδομένων, την παρουσίαση των αποτελεσμάτων και τη συγγραφή της έκδοσης συμμετείχαν οι Κωνσταντίνος Σιουμάλας-Χριστοδούλου, Στέφανος Χριστόπουλος, Δρ. Χαράλαμπος Χρυσομαλλίδης, Δρ. Νίκος Καραμπέκιος. Τον έλεγχο της μεθοδολογίας και της εγκυρότητας των στοιχείων πραγματοποίησε η Δρ. Νένα Μάλλιου.

Η Δρ. Μαρία Μαλαγκονιάρη πραγματοποίησε την τελική γλωσσική και νοηματική επιμέλεια του κειμένου.

ΠΡΟΤΕΙΝΟΜΕΝΟΣ ΤΡΟΠΟΣ ΑΝΑΦΟΡΑΣ

Σαχίνη, Ε., Μάλλιου Ε., Χρυσομαλλίδης, Χ., Καραμπέκιος, Ν., Σιουμάλας-Χριστοδούλου, Κ., Χριστόπουλος, Σ. (2022). *Εντοπισμός και ανάλυση των ελληνικών επιστημονικών δημοσιεύσεων στον τομέα της Τεχνητής Νοημοσύνης με τεχνικές Μηχανικής Μάθησης*. Αθήνα: Εθνικό Κέντρο Τεκμηρίωσης και Ηλεκτρονικού Περιεχομένου.

ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ

Copyright © 2022 Εθνικό Κέντρο Τεκμηρίωσης και Ηλεκτρονικού Περιεχομένου

δ: Ζεφύρου 56, 17564 Παλαιό Φάληρο • τ: 210 2204900 • e: ekt@ekt.gr • www.ekt.gr



Το έργο αυτό διατίθεται με άδεια Creative Commons

Αναφορά Δημιουργού - Μη Εμπορική Χρήση - Όχι Παράγωγα Έργα 4.0 Διεθνές

Προκειμένου να δείτε αντίγραφο της άδειας επισκεφθείτε:

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.el>

Η έκδοση διατίθεται ηλεκτρονικά στη διεύθυνση <http://metrics.ekt.gr/>

ISBN: 978-618-5557-35-5 (pdf)

ISBN: 978-618-5557-34-8 (print)

Ο σχεδιασμός της έκδοσης πραγματοποιήθηκε από την κα Δήμητρα Πελεκάνου και την κα Μαργαρίτα Σκανδάλη.

Πρόλογος



Η παραγωγή και η ανάδειξη στοιχείων και δεδομένων που συμβάλλουν στην ουσιαστική αποτύπωση των ερευνητικών και επιστημονικών επιδόσεων της χώρας μας, αποτελούν βασικές δραστηριότητες του Εθνικού Κέντρου Τεκμηρίωσης και Ηλεκτρονικού Περιεχομένου (ΕΚΤ), ως Εθνική Αρχή του Ελληνικού Στατιστικού Συστήματος για τις ευρωπαϊκές στατιστικές Έρευνας, Ανάπτυξης και Καινοτομίας.

Η παρούσα έκδοση επιχειρεί την πρώτη αναλυτική αποτύπωση των επιστημονικών δημοσιεύσεων των ελληνικών φορέων στο δυναμικά αναπτυσσόμενο επιστημονικό και τεχνολογικό πεδίο της Τεχνητής Νοημοσύνης. Στον 21^ο αιώνα, η Τεχνητή Νοημοσύνη ανοίγει νέους ορίζοντες, απασχολεί όλο και περισσότερο την έρευνα σε πολλές επιστημονικές περιοχές, ενισχύει τις καινοτόμες εφαρμογές σε πολλούς κλάδους, ενώ παράλληλα είναι το επίκεντρο εθνικών και δημόσιων πολιτικών.

Το ΕΚΤ, ως ένας φορέας με ερευνητικό και αναπτυξιακό χαρακτήρα, που από τον θεσμικό του ρόλο συγκεντρώνει και διαθέτει πλήθος δεδομένων, χρησιμοποιεί τεχνικές Τεχνητής Νοημοσύνης. Ένα πεδίο αφορά τις ανθρωπιστικές επιστήμες και συγκεκριμένα τη δημιουργία αξιόπιστου Μοντέλου Μηχανικής/Βαθιάς Μάθησης για την εκτίμηση των θεματικών των επιστημονικών πεδίων των τεκμηρίων που συγκεντρώνονται στις πύλες Searchculture.gr και Openarchives.gr, βάσει των μετα-δεδομένων τους (Classifier), και την αξιοποίηση του μοντέλου για αυτόματο σημασιολογικό εμπλουτισμό.

Ένα επίσης ενδιαφέρον πεδίο διερεύνησης, αποτέλεσμα του οποίου είναι και η παρούσα έκδοση, αφορά την εφαρμογή τεχνικών μηχανικής μάθησης για την ανάδειξη περισσότερης πληροφορίας από τα στατιστικά δεδομένα του ΕΚΤ. Ειδικότερα στην έκδοση αυτή, οι τεχνικές μηχανικής μάθησης αξιοποιήθηκαν για τον εντοπισμό των επιστημονικών δημοσιεύσεων Ελλήνων ερευνητών στον τομέα της Τεχνητής Νοημοσύνης από τις διεθνείς βιβλιομετρικές βάσεις Web of Science και Scopus.

Ακόμα, στην έκδοση παρουσιάζονται οι κυριότεροι βιβλιομετρικοί δείκτες των επιστημονικών αυτών δημοσιεύσεων και η χρονική τους εξέλιξη, καταγράφονται οι ελληνικοί φορείς που δραστηριοποιούνται στον τομέα της Τεχνητής Νοημοσύνης, και αναδεικνύονται τα δίκτυα που έχουν σχηματίσει οι Έλληνες ερευνητές στο διεθνές στερέωμα για ερευνητικούς σκοπούς σε αυτό το πεδίο. Για τον υπολογισμό των βιβλιομετρικών δεικτών αξιοποιήθηκε η πολύχρονη συσσωρευμένη γνώση σε στατιστικό και αναλυτικό επίπεδο που υποστηρίζεται από το εξειδικευμένο ανθρώπινο δυναμικό και τις τυποποιημένες στατιστικές διαδικασίες του ΕΚΤ.

Η συγκεκριμένη έκδοση αποτελεί την απαρχή μιας νέας σειράς εκδόσεων που παρουσιάζουν βιβλιομετρικές αναλύσεις με θεματικά εξειδικευμένη προσέγγιση καλύπτοντας τομείς που βρίσκονται στην αιχμή τεχνολογικών εξελίξεων αλλά και του ευρύτερου ενδιαφέροντος.

A handwritten signature in blue ink, consisting of a stylized, cursive script. The signature appears to be 'Σαχίνη' (Sachini) with a horizontal line underneath.

Δρ Εύη Σαχίνη
Διευθύντρια ΕΚΤ

Περιεχόμενα

1. Επισκόπηση εννοιών, εφαρμογών και πολιτικών σχετικών με την Τεχνητή Νοημοσύνη	7
1.1 Επιστημονικά πεδία	13
1.2 Παραγωγική δραστηριότητα	15
1.3 Δημόσιες παρεμβάσεις και μέτρα πολιτικής	18
2. Μεθοδολογία	21
Βήμα 1: Προσδιορισμός των επιστημονικών δημοσιεύσεων	21
Βήμα 2: Εντοπισμός των επιστημονικών άρθρων και των πρακτικών συνεδρίων	23
Βήμα 4: Κατασκευή του μοντέλου εντοπισμού των δημοσιεύσεων	24
Επιλογή, κατασκευή και προσαρμογή του μοντέλου	25
Εκπαιδύοντας το μοντέλο (training set)	25
Βήμα 5: Προσδιορισμός του τελικού πληθυσμού των δημοσιεύσεων	27
Βήμα 6: Συλλογή των βιβλιομετρικών χαρακτηριστικών των επιστημονικών δημοσιεύσεων και των ιδρυμάτων	27
3. Αποτελέσματα για την ελληνική επιστημονική παραγωγή στον τομέα της Τεχνητής Νοημοσύνης	29
3.1 Αριθμός δημοσιεύσεων, συνολικά και ανά ίδρυμα	29
3.2 Κυρίαρχοι όροι στις δημοσιεύσεις	34
3.3 Ανάλυση διεθνών δικτύων για τη συγγραφή επιστημονικών δημοσιεύσεων	35
Βιβλιογραφία	39

1. Επισκόπηση εννοιών, εφαρμογών και πολιτικών σχετικών με την Τεχνητή Νοημοσύνη

Ο όρος “Βιομηχανία 4.0 (Industry 4.0)”, που έχει επικρατήσει διεθνώς, υποδηλώνει την τρέχουσα τάση της αυτοματοποίησης και της ανταλλαγής δεδομένων στις τεχνολογίες παραγωγής. Η “Βιομηχανία 4.0” αναφέρεται συνήθως και ως Τέταρτη Βιομηχανική Επανάσταση, όπου προεξάρχοντα ρόλο έχουν οι νέες δυνατότητες που προσφέρει η ικανότητα επεξεργασίας των υπολογιστών, η διαρκώς ταχύτερη ανάλυση και η μεγαλύτερη αποθήκευση και επεξεργασία όγκου δεδομένων σε συνδυασμό με την εκθετική πορεία της διείσδυσης του Internet και τις υψηλές ταχύτητες που προσφέρουν οι ενσύρματες και ασύρματες υποδομές.

Στοιχείο που διαφοροποιεί σε σημαντικό βαθμό το σύγχρονο πλαίσιο σε σχέση με παλιότερες καθιερωμένες πρακτικές σχετίζεται με το πλήθος των επιστημονικών και τεχνολογικών εφαρμογών και χρήσεων, οι οποίες αξιοποιούν πρακτικά στη σύγχρονη καθημερινότητα και παραγωγική πραγματικότητα τις ανωτέρω νέες δυνατότητες και συνεχώς τις εμπλουτίζουν. Χαρακτηριστικά παραδείγματα αποτελούν οι εφαρμογές που επιτρέπουν την ανάδυσση του υπολογιστικού νέφους (cloud computing), δηλαδή, τη διάθεση υπολογιστικών πόρων μέσω του διαδικτύου από κεντρικά συστήματα, τα οποία βρίσκονται απομακρυσμένα από τον τελικό χρήστη, την επικοινωνία συσκευών, οικιακών συσκευών, αυτοκινήτων καθώς και κάθε αντικειμένου που ενσωματώνει ηλεκτρονικά μέσα ώστε να επιτρέπεται η σύνδεση και η ανταλλαγή δεδομένων (internet of things), καθώς και της Τεχνητής Νοημοσύνης (εφεξής «ΤΝ») (artificial intelligence) η οποία λογίζεται ως «επιστήμη και μεθοδολογία της δημιουργίας νοημών μηχανών» (McCarthy, 1998).

Οι χρήστες των νέων αυτών εφαρμογών δεν περιορίζονται στον χώρο των επιχειρήσεων και των παραγωγών νέων τεχνολογιών, καθώς, σήμερα, και ο δημόσιος τομέας συμπεριλαμβάνεται δυναμικά στους χρήστες τους. Ο βασικός λόγος για αυτό σχετίζεται με τη δυνατότητα αξιοποίησής τους και τη δυνατότητα που παρέχουν στην επίλυση ζητημάτων δημόσιας πολιτικής, όπως η αξιοποίηση των δημόσιων δεδομένων, η βέλτιστη κατανομή δημόσιων πόρων, κτλ. (World Bank 2020; Misuraca and Van Noordt 2020). Συνολικά, οι εμπλεκόμενοι δρώντες (ακαδημαϊκή κοινότητα, επιχειρήσεις και δημόσιος τομέας) δεν είναι, απλώς, κοινωνικοί σχηματισμοί οι οποίοι αξιοποιούν ή συμμετέχουν στην παραγωγή της υποδομής που απαιτείται για αυτό το νέο «τεχνολογικό παράδειγμα», αλλά είναι ταυτόχρονα χρήστες και παραγωγοί.

Αυτή η πολλαπλότητα των ρόλων αποτελεί ένα –ακόμα– διακριτικό χαρακτηριστικό της σημερινής πρακτικής που αναδεικνύεται, σε σχέση με παλαιότερα, αφού η παράλληλη ανάπτυξη και η χρήση των τεχνολογιών αυτών δημιουργεί έναν ενάρετο κύκλο ταχείας τεχνολογικής υιοθέτησης (technology adoption) και αξιοποίησης. Σε αυτό το πλαίσιο, όροι, όπως η Βιομηχανία 4.0, ο ψηφιακός μετασχηματισμός, η αυτοματοποίηση της παραγωγής, η (επαν)απόκτηση ψηφιακών δεξιοτήτων,

αποτελούν μέσα-τρόπους και επιχειρησιακές έννοιες, που εκφράζονται από φορείς άσκησης πολιτικής, οι οποίοι στοχεύουν στην αξιοποίηση των τεχνολογικών εξελίξεων καθώς εκτιμάται πως αποτελούν τρόπους επίλυσης σύνθετων κοινωνικών, επιχειρησιακών, επιχειρηματικών και οικονομικών φαινομένων. Αυτό ισχύει σε όλες τις αναπτυγμένες οικονομίες του κόσμου. Πλέον, συντάσσονται και υλοποιούνται εθνικές στρατηγικές αξιοποίησης αυτής της δέσμης των τεχνολογιών¹. Αντίστοιχα, στρατηγικές και τρόποι υλοποίησης του τεχνολογικού αυτού δυναμικού (potential) συντάσσονται στο κλαδικό, τομεακό και άλλως επιμεριζόμενο παραγωγικό και κοινωνικό συνεχές. Στο ίδιο μήκος κύματος, ευρωπαϊκοί² και διεθνείς οργανισμοί καταγράφουν σχετιζόμενες τάσεις, ερευνούν τις τεχνολογικές και οικονομικές συνέπειες και προσπαθούν να σκιαγραφήσουν το μέλλον συναρτήσει αυτής της εξέλιξης (WIPO, 2019; Zhang et al., 2021). Το ίδιο ισχύει και για τις επιχειρήσεις που επιδιώκουν να προσδιορίσουν τον εταιρικό τους προσανατολισμό δεδομένων αυτών των τεχνολογιών.

Σε αυτό το πλαίσιο, η ΤΝ είναι ένα σύγχρονο πεδίο δράσης και δημιουργίας το οποίο διευρύνεται διαρκώς. Αν και είναι κοινά αποδεκτό, ότι η ΤΝ, ως όρος, περιγράφει και συμπεριλαμβάνει μηχανές και συστήματα που εκτελούν ανθρώπινες γνωστικές λειτουργίες (π.χ. μάθηση, κατανόηση, συλλογισμός ή αλληλεπίδραση), αξιοποιώντας διάφορες μορφές γνώσης και κατανόησης νοήματος, όπως η αναγνώριση ομιλίας και η γλωσσική επεξεργασία, αλλά και δραστηριότητες ανθρώπινης αλληλεπίδρασης, διά της ανίχνευσης σημάτων, προσομοιώσεων κ.λπ., λόγω της φύσης του συγκεκριμένου πεδίου, δεν υπάρχει ενιαίος και κοινά αποδεκτός ορισμός. Παρ' όλα αυτά, σύμφωνα με τον Έλληνα καθηγητή Α. Τσιρίγκο, Καθηγητή στο Ινστιτούτο Υπολογιστικής Ιατρικής του NYU Grossman School of Medicine και Διευθυντή των Εργαστηρίων Εφαρμοσμένης Βιοπληροφορικής του ίδιου Ιδρύματος, η ΤΝ σχετίζεται άμεσα με σημαντικά διεθνή φαινόμενα και τάσεις που σημειώνονται τα τελευταία χρόνια, όπως η επιταχυνόμενη αστικοποίηση, η παγκοσμιοποίηση και η αφθονία των προϊόντων, καθώς και των διαθέσιμων πληροφοριών που μεταβάλλουν τη σύγχρονη κοινωνία, καθιστώντας πιο σύνθετες και απαιτητικές τις συνθήκες γύρω μας. Εν μέρει, ως απάντηση σε αυτό το περιβάλλον και τις προκλήσεις που αυτό συνεπάγεται, η ΤΝ μπορεί να ιδωθεί ως πλέγμα εφαρμογών οι οποίες αναπτύσσονται για αυτοματοποιημένες διαδικασίες και υπηρεσίες, στη βάση ισχυρών αλγόριθμων που σταδιακά ενεργούν προσομοιάζοντας σε θεμελιώδεις ανθρώπινες δυνατότητες, όπως η όραση, η ομιλία και η πλοήγηση (WIPO 2019).

Συνολικά, η ΤΝ αναφέρεται σε υπολογιστικά συστήματα που παρουσιάζουν έξυπνη συμπεριφορά μέσα από την ανάλυση του περιβάλλοντός τους, αλλά και την υλοποίηση ενεργειών με κάποιο βαθμό αυτονομίας, στην προσπάθεια επίτευξης συγκεκριμένων στόχων. Τα παραπάνω αποδίδουν μια ευρεία θεώρηση της ΤΝ, η οποία πέρα από τη μίμηση κοινών ανθρώπινων δυνατοτήτων, προχωρά και σε πιο εξειδικευμένες δραστηριότητες που πραγματοποιούνται από ειδικούς σε συγκεκριμένους τομείς (OECD 2021).

Σε κάθε περίπτωση, ένα σύστημα ΤΝ αποτελείται, τουλάχιστον, από αισθητήρες, λογική λειτουργίας και ενεργοποιητές, όπου καθένα από αυτά τροφοδοτεί και επανατροφοδοτεί την αλυσίδα που δημιουργείται, συνιστώντας ένα ενιαίο σύστημα (OECD 2019). Για αυτό τον λόγο, πρακτικά η ΤΝ μπορεί να θεωρηθεί ένας ετερογενής -ως προς την τεχνολογική του βάση- τομέας, καθώς περιλαμβάνει τόσο πτυχές που

¹ Ενδεικτικά, για ΗΠΑ, βλ. <https://www.ai.gov/>, για ΗΒ, βλ. <https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version>, για Γερμανία, βλ. <https://www.ki-strategie-deutschland.de/home.html>, Γαλλία, βλ. <https://www.aiforhumanity.fr/en/>. Η ΕΕ και ο ΟΟΣΑ έχουν δομές παρακολούθησης των διαφόρων εθνικών πρωτοβουλιών στις οποίες ευετηριάζονται συστηματικά τα επιμέρους εθνικά έγγραφα, βλ. https://knowledge4policy.ec.europa.eu/ai-watch_en και <https://oecd.ai/en/>, αντίστοιχα.

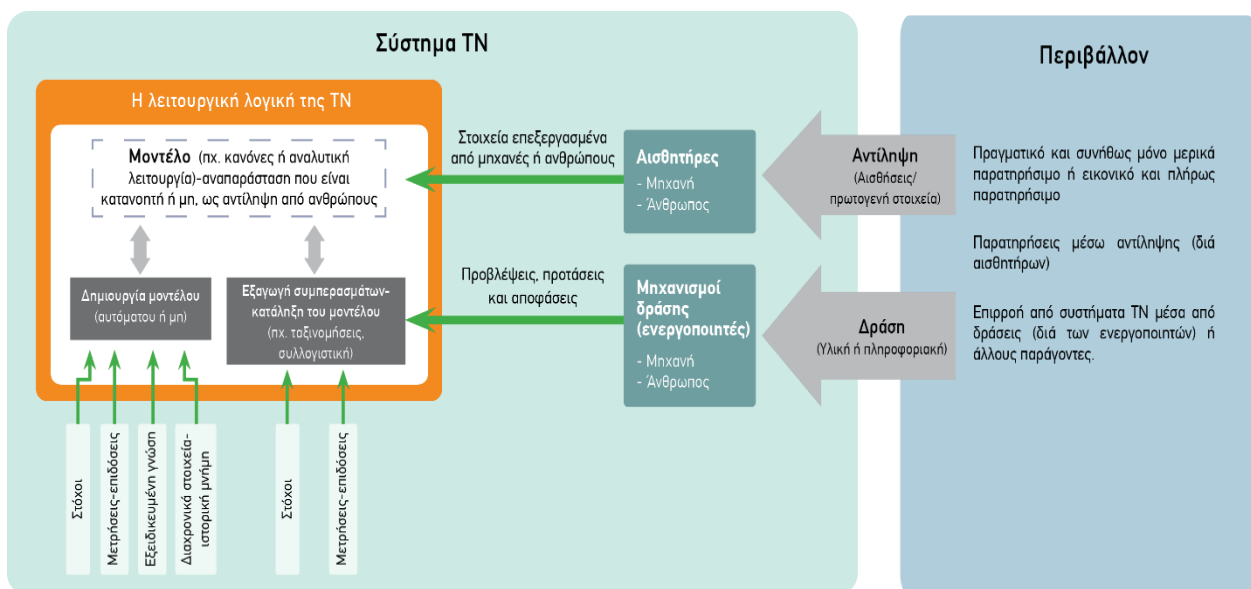
² Βλ. <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>.

αφορούν κυρίως λογισμικό (π.χ. αισθητήρες, τσιπ), όσο και εφαρμογές ενσωματωμένες σε μηχανές που μπορούν να ιδωθούν ως συστήματα (π.χ. ρομπότ, αυτόνομη οδήγηση).

Σε αυτή τη βάση, κύριο χαρακτηριστικό και σκοπός της ΤΝ είναι η ανάπτυξη μηχανών και συστημάτων που μπορούν να διεκπεραιώνουν δραστηριότητες, για τις οποίες είναι αναγκαία η χρήση της ανθρώπινης νοημοσύνης και ευφυΐας. Με άλλα λόγια, η ΤΝ ασχολείται με τη σχεδίαση και την υλοποίηση υπολογιστικών –κυρίως– συστημάτων, τα οποία μιμούνται στοιχεία της ανθρώπινης συμπεριφοράς, βασιζόμενα στη μάθηση, την προσαρμοστικότητα, την εξαγωγή συμπερασμάτων και την επίλυση προβλημάτων, χωρίς να έχει προηγηθεί ακριβής και σχεδιασμένος προγραμματισμός, όπως συνθίζεται στον κλάδο της Πληροφορικής³. Αποτέλεσμα είναι η δημιουργία και λειτουργία ενός συστήματος που μπορεί να κάνει προβλέψεις, συστάσεις για ένα δεδομένο σύνολο ανθρώπινων δραστηριοτήτων και στόχων ή να προτείνει αποφάσεις οι οποίες επηρεάζουν πραγματικά ή εικονικά περιβάλλοντα. Μια γενικότερη δομή των συστημάτων αυτών αποτυπώνεται στο Σχήμα 1, όπου αξιοποιούνται μηχανικές ή/και ανθρώπινες εισροές για:

- την αντίληψη (πραγματικών και/ή εικονικών) περιβαλλόντων,
- την αφαίρεση ή τη σύνοψη δραστηριοτήτων με αυτοματοποιημένο τρόπο και
- την αξιοποίηση/τη χρήση συμπερασμάτων βάσει του μοντέλου που έχει δομηθεί για τη διατύπωση επιλογών και ενεργειών.

ΣΧΗΜΑ 1: Αναλυτική εννοιολογική θεώρηση ενός αντιπροσωπευτικού συστήματος ΤΝ



Πηγή: OECD (2019)

³ <https://www.europarl.europa.eu/news/el/headlines/society/20200827STO85804/ti-einai-i-techniti-noimosuni-kai-pos-chrisimopoietai>

Η πιο χαρακτηριστική διάκριση ως προς την ΤΝ είναι μεταξύ της λεγόμενης «στενής» (ή ασθενούς) ΤΝ, η οποία αφορά μεμονωμένες εργασίες και της «γενικής» (ή ισχυρής) ΤΝ, η οποία σχετίζεται με τη συνολικότερη ικανότητα μιας μηχανής να κατανοεί ή να μαθαίνει οποιαδήποτε πνευματική εργασία που μπορεί να εκτελέσει ένας άνθρωπος. Μεταξύ άλλων, οι σημαντικότερες εξελίξεις στην ΤΝ, σήμερα, εντοπίζονται στα ακόλουθα υποπεδία:

Μηχανική μάθηση (Machine learning)

Η μηχανική μάθηση αναφέρεται στη δημιουργία συστημάτων, διά της μελέτης και κατασκευής αλγόριθμων και μοτίβων που μπορούν να μαθαίνουν από τα δεδομένα και να κάνουν προβλέψεις σχετικά με αυτά. Με άλλα λόγια, οι αλγόριθμοι «εκπαιδεύονται» ώστε να εντοπίζουν μοτίβα και χαρακτηριστικά σε μεγάλο όγκο δεδομένων, προκειμένου να είναι σε θέση να επιτυγχάνουν μείωση των αναγκαίων παραμετροποιήσεων, μέσω ταξινομήσεων και ομαδοποιήσεων. Η μηχανική μάθηση είναι μία μέθοδος η οποία επιτρέπει και διευκολύνει την παραγωγή αξιόπιστων αποφάσεων και αποτελεσμάτων προκειμένου να αναδεικνύονται αλληλοσυσχετίσεις μέσω αυτής από σχέσεις και τάσεις στα δεδομένα, με την ελάχιστη ανθρώπινη παρέμβαση⁴.

Ενισχυτική μάθηση (Reinforcement learning)

Μάθηση για τη λειτουργία συστημάτων που «εκπαιδεύονται» αλληλεπιδρώντας άμεσα με το περιβάλλον δραστηριοποίησης η οποία εφαρμόζεται στον έλεγχο κίνησης ρομπότ, στη βελτιστοποίηση εργασιών στα εργοστάσια, κ.ά.

Βαθιά μάθηση (Deep learning)

Τεχνική – προσέγγιση μηχανικής μάθησης μέσω της οποίας οι υπολογιστές «εκπαιδεύονται» ώστε να δρουν «φυσικά», δηλαδή να αναλύουν δεδομένα με μία συγκεκριμένη διαδικασία, όπως κάνουν οι άνθρωποι. Με άλλα λόγια, οι υπολογιστές «μαθαίνουν» να δρουν έξυπνα, αυτοβούλως, καθώς το σύστημα είναι σε θέση να «μάθει» και να «λάβει» έξυπνες αποφάσεις χωρίς ανθρώπινη παρέμβαση.

Επεξεργασία φυσικής γλώσσας (Natural language processing – NLP)

Διεπιστημονικός κλάδος της επιστήμης της πληροφορικής, της ΤΝ και της υπολογιστικής γλωσσολογίας που ασχολείται με τις αλληλεπιδράσεις μεταξύ των υπολογιστών και των ανθρώπινων (φυσικών) γλωσσών⁵.

⁴ <https://www.europarl.europa.eu/news/el/headlines/society/20200827ST085804/ti-einai-i-techniti-noimosuni-kai-pos-chrisimopoiieitai> , <https://www.ibm.com/cloud/learn/what-is-artificial-intelligence>

⁵ Προκλήσεις στον συγκεκριμένο χώρο περιλαμβάνουν την κατανόηση φυσικής γλώσσας, δηλαδή την προσπάθεια να καταστούν ικανοί οι υπολογιστές να εξάγουν νοήματα από ανθρώπινα ή γλωσσικά δεδομένα, αλλά και την παραγωγή φυσικής γλώσσας.

Σε αυτό το πλαίσιο, πρόοδος έχει συντελεστεί:

α) στο πεδίο των γενικών εξελίξεων, με την έννοια ότι τα συστήματα ΤΝ μπορούν πλέον να συνθέτουν κείμενο, ήχο και εικόνες σε αρκετά υψηλά πρότυπα.

β) στο εγχείρημα της εκβιομηχάνισης της όρασης του υπολογιστή, όπου έχει σημειωθεί σημαντική πρόοδος την τελευταία δεκαετία, κυρίως λόγω της χρήσης τεχνικών βαθιάς μάθησης. Νέα δεδομένα δείχνουν ότι τα συστήματα όρασης υπολογιστών αποτελούν τομέα που μπορεί να εκβιομηχανιστεί περαιτέρω, ενώ παράλληλα οι τεχνολογίες ανεπτυγμένων συστημάτων ανίχνευσης αντικειμένων εξελίσσονται διαρκώς.

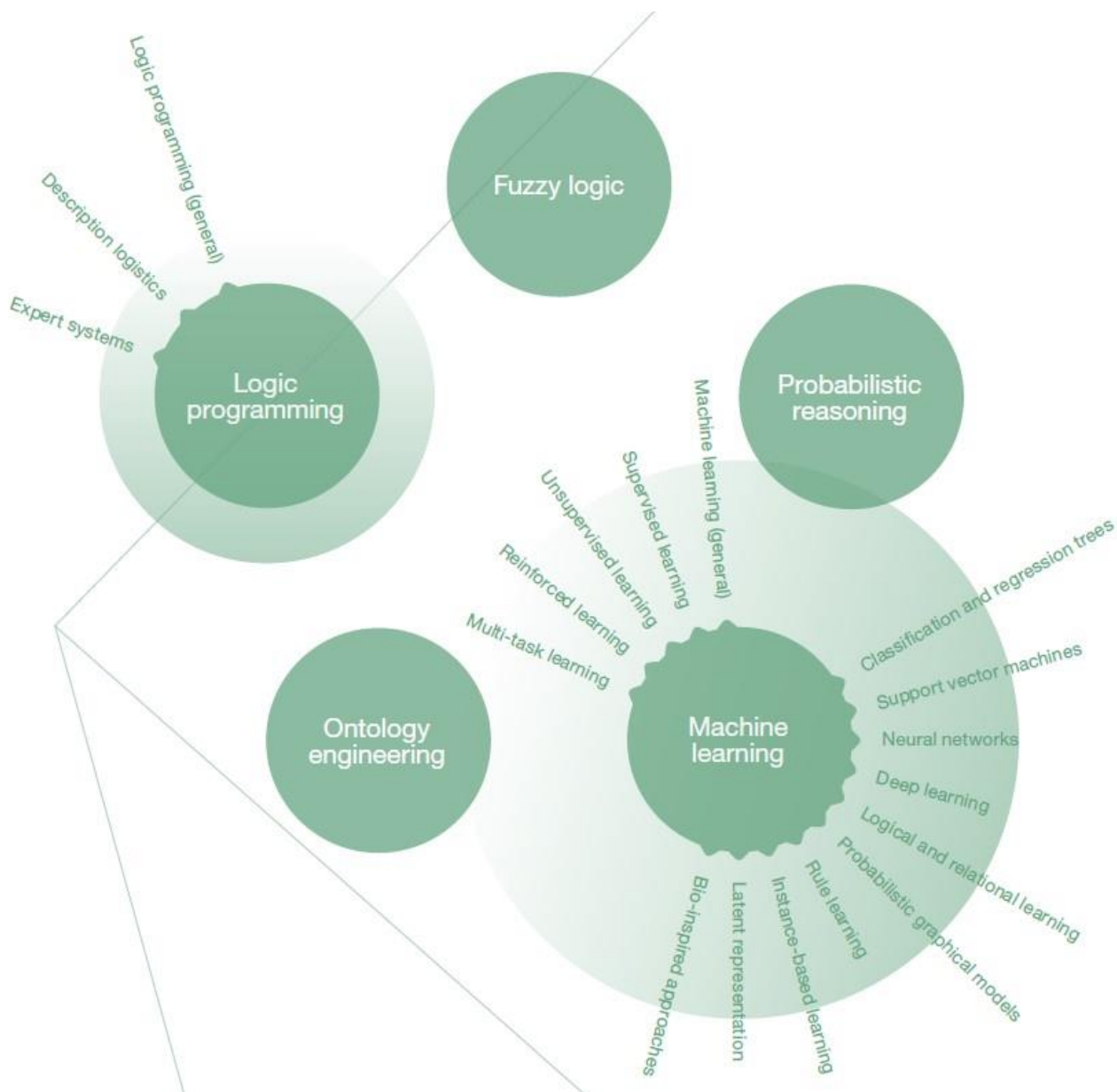
γ) στη φυσική επεξεργασία γλωσσών, όπου η ταχεία πρόοδος στα συστήματα ΤΝ με σημαντικά βελτιωμένες γλωσσικές δυνατότητες έχει πλέον οικονομικό αντίκτυπο, καθώς μηχανές αναζήτησης έχουν ενσωματώσει και αναπτύξει περαιτέρω τις σχετικές δυνατότητες.

δ) στις νέες αναλύσεις στο πεδίο της συλλογιστικής-σκεπτικού, καθώς με απαρχή την ύπαρξη πλήθους δεδομένων επιταχύνονται οι εξελίξεις στην ανάπτυξη λογικών συνειρμών και συλλογιστικής από συστήματα ΤΝ.

ε) στην ίδια τη μηχανική μάθηση, όπου ενδεικτικά αναφέρεται ότι αυτή συμβάλλει στην εξέλιξη της υγειονομικής περίθαλψης και της βιολογίας, καθώς η ανάπτυξη συστημάτων ΤΝ που αναπαριστούν χημικά μόρια συμβάλλουν στον πιο αποτελεσματικό σχεδιασμό χημικής σύνθεσης φαρμάκων (Zhang et al. 2021).

Στο Σχήμα 2 αποτυπώνονται οι πλέον διαδεδομένες και σύγχρονες τεχνικές που χαρακτηρίζονται ως απόλυτα συναφείς με την ΤΝ.

ΣΧΗΜΑ 2: Σύγχρονες τεχνικές, σχετικές με την ΤΝ



Πηγή: WIPO (2019)

1.1 Επιστημονικά πεδία

Τα συστήματα ΤΝ εξαρτώνται σε μεγάλο βαθμό από την ύπαρξη και αξιοποίηση δεδομένων, ως εισροές και αναμενόμενες εκροές (δομημένα δεδομένα και δεδομένα εκμάθησης ή εκπαίδευσης), με αποτέλεσμα η ακρίβειά τους να είναι ευθέως ανάλογη του αριθμού των υπαρχόντων δεδομένων. Σήμερα, κυρίαρχες έννοιες για την ΤΝ είναι η Μηχανική Μάθηση και η Βαθιά Μάθηση. Η συγκεκριμένη περιοχή αποτελεί υποπεδίο της Επιστήμης Ηλεκτρονικών Υπολογιστών (κατά Frascati, μέρος του ευρύτερο πεδίου Φυσικές Επιστήμες). Ιδίως η πρώτη χρησιμοποιεί σε μεγάλο βαθμό τη Στατιστική, προκειμένου να βελτιωθεί η ικανότητα των μηχανών να κάνουν προβλέψεις με βάση ιστορικά δεδομένα που οδηγούν με τη σειρά τους σε σχετικά μοντέλα. Αυτά τα μοντέλα προσομοιάζουν στο ανθρώπινο νευρικό σύστημα και είναι γνωστά ως «νευρωνικά δίκτυα» στη σχετική ορολογία, τα οποία συμβάλλουν στην επέκταση και την περαιτέρω ανάπτυξη της ΤΝ από την αξιοποίηση δεδομένων μεγάλης έκτασης και υπολογιστικής ισχύος (OECD 2019).

Εντός αυτών των επιστημονικών πεδίων, κρίσιμης σημασίας για τις δραστηριότητες και τις εξελίξεις που θεωρούνται σχετικές με την ΤΝ είναι υπο-πεδία που σχετίζονται με τη διαχείριση μεγάλων δεδομένων, τη λειτουργία και αξιοποίηση των αλγορίθμων, τη μηχανική μάθηση, την επεξεργασία της γλώσσας, το λειτουργικό σύστημα των ΗΥ και την υπολογιστική όραση (αναγνώριση αντικειμένων βάσει όρασης, όπως συμβαίνει και από τους ανθρώπους). Αντίστοιχα, ερευνητικά πεδία, τα οποία συγκαταλέγονται μεταξύ των πιο διαδεδομένων, όντας συμβατά με την ΤΝ, είναι τα συστήματα ελέγχου και παρακολούθησης, τα οποία βοηθούν στην επίλυση προβλημάτων, βάσει συνεχούς και επαναλαμβανόμενης παρατήρησης, η ρομποτική, τα – προερχόμενα από την επιστήμη της διοίκησης - συστήματα υποστήριξης αποφάσεων, καθώς και η αναγνώριση μοτίβων, τα οποία συμβάλλουν στην αύξηση της αντιληπτικής ικανότητας των υπολογιστικών συστημάτων (Zhang & Lu 2021).

Αποτιμώντας μετρτικά την επιστημονική παραγωγή στον τομέα της ΤΝ, οι Baruffaldi et al. (2020) αναφέρουν ότι από το σύνολο των σχετικών δημοσιεύσεων, το ένα τρίτο περίπου αφορά τον τομέα της Επιστήμης Η/Υ, το ένα τέταρτο τις Επιστήμες Μηχανικού και περίπου το ένα δέκατο το πεδίο των Μαθηματικών, ενώ δημοσιεύσεις σχετικές με την ΤΝ συναντώνται με σχετικά υψηλή συγκέντρωση και συχνότητα και σε άλλους τομείς, όπως η Μηχανική Υλικών, η Ιατρική και η Χημεία. Επίσης, σχετικά αναμενόμενο είναι το ότι η έρευνα στον συγκεκριμένο τομέα εξελίσσεται, με συνέπεια οι «λέξεις-κλειδιά» που εμφανίζονται, σήμερα, συχνότερα σε δημοσιεύσεις σχετικές με την ΤΝ να είναι νέες, συγκριτικά με εκείνες από τα τέλη της δεκαετίας του '10 και με αντίστοιχες στα τέλη της δεκαετίας του '90. Χαρακτηριστικές τέτοιες περιπτώσεις είναι οι ακόλουθες: «support vector machine», «wireless sensor network», «particle swarm optimization», κ.ά. Αντίθετα, παλαιότερα, επαναλαμβανόμενες και πιο γενικές «λέξεις-κλειδιά» ήταν οι εξής: «artificial neural network» και «neural network model» οι οποίες συνεχίζουν να εμφανίζονται, και σήμερα, στις σχετικές δημοσιεύσεις.

Στο Σχήμα 3 αποτυπώνονται κυρίαρχες λειτουργικές αναφορές για τις σύγχρονες εξελίξεις στην ΤΝ. Στη βάση αυτών βρίσκονται τα κύρια επιστημονικά πεδία της δραστηριότητας και της ανάπτυξης της ΤΝ.

ΣΧΗΜΑ 3: Σύγχρονες λειτουργικές αναφορές γύρω από την ΤΝ



Πηγή: WIPO (2019)

Εδώ, αξίζει να σημειωθεί ότι πεδίο δημόσιας παρέμβασης με οριζόντιο χαρακτήρα αποτελεί η προώθηση της καινοτομίας σε συστήματα ΤΝ, η οποία οδηγεί σε καινοτόμες υπηρεσίες και προϊόντα που ενσωματώνουν τεχνολογία ΤΝ, συμβάλλοντας στην αύξηση της ψηφιακής αυτονομίας. Συνολικά, τα παραπάνω συνδέονται με την ευρύτερη θεματική της διακυβέρνησης δεδομένων, τη διαχείριση των ανοικτών δεδομένων και τη διασύνδεση δεδομένων που είναι απαραίτητα για την ανάπτυξη υψηλής ποιότητας μοντέλων, βασισμένων σε δεδομένα, ΤΝ. Παρόλα αυτά, είναι -σε μεγάλο βαθμό- αποδεκτό ότι η άνθηση της καινοτομίας στο χώρο αυτό συνδυάζεται με την ύπαρξη ενός νομικού πλαισίου που είναι αφενός φιλικό προς την καινοτομία, και αφετέρου εξασφαλίζει τα ζητήματα ασφάλειας και ηθικής τα οποία είναι άκρως απαραίτητα στη συζήτηση περί ΤΝ. Προς τούτο, η δημιουργία «ρυθμιστικών δοκιμαστηρίων» (regulatory sandboxes) θεωρείται σημαντική, όπως αναφέρει η ΕΕ, τα οποία πρόκειται να θέσουν το πλαίσιο για τη διακυβέρνηση, την εποπτεία και την ευθύνη, προκειμένου να εξασφαλίζεται η δοκιμή καινοτόμων τεχνολογιών ΤΝ, για περιορισμένο χρονικό διάστημα και με τρόπο που δεν θα επιβαρύνει έμμεσα –διά της κανονιστικής οδού– τις μικρομεσαίες και τις νεοφυείς επιχειρήσεις (Ευρωπαϊκή Επιτροπή 2021).

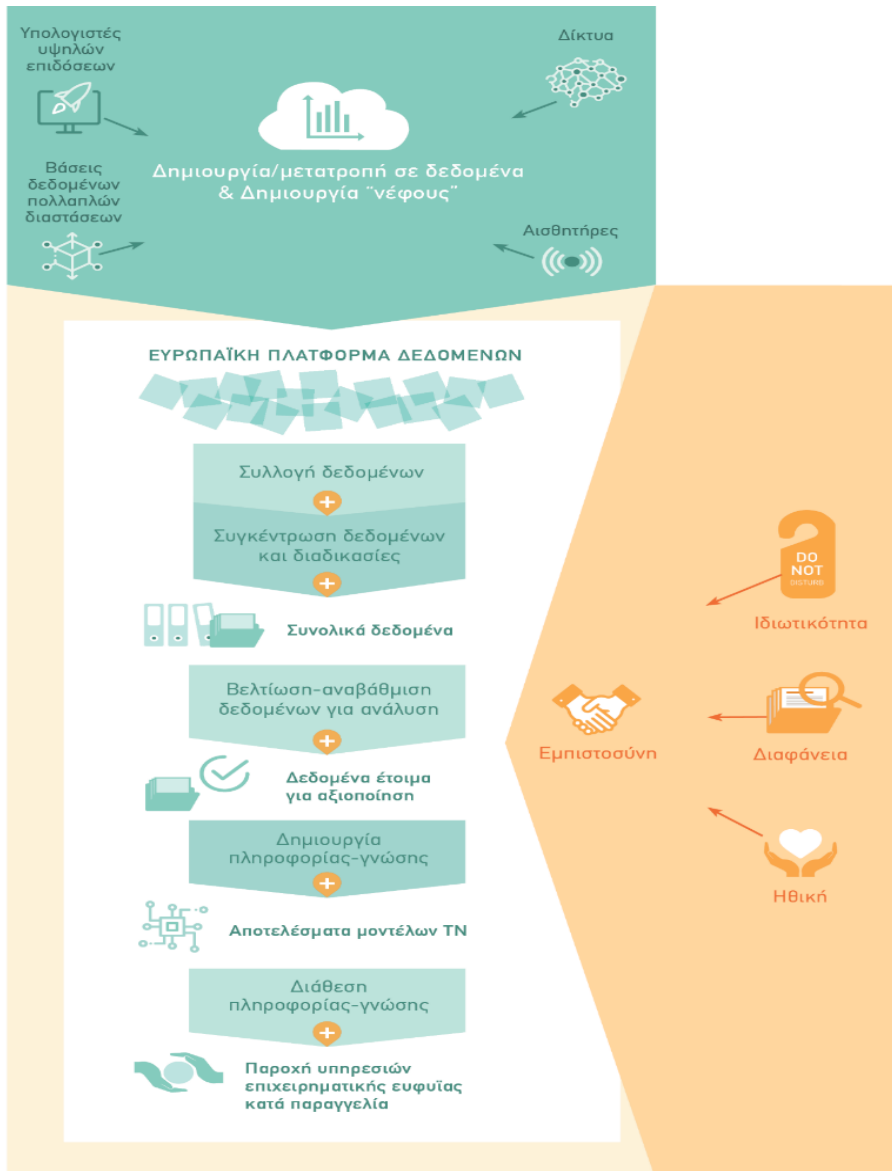
Μια πρόσθετη κι ενδιαφέρουσα διάσταση, η οποία άπτεται της ΤΝ, σχετίζεται με την επίδραση που η ίδια έχει στην εκπαιδευτική δραστηριότητα των πανεπιστημίων, διεθνώς, κυρίως στο επίπεδο της εξειδίκευσης των σπουδών, από τις προπτυχιακές στις μεταπτυχιακές (και τις διδακτορικές) (Zhang et al. 2021).

1.2 Παραγωγική δραστηριότητα

Η ανίχνευση μοτίβων μεταξύ πλήθους, φαινομενικά ανεξάρτητων / ασυμβίβαστων δεδομένων, είναι δυνατό να συμβάλλουν –μεταξύ άλλων- στην ενίσχυση της βιομηχανικής παραγωγικότητας μέσω της ΤΝ. Σε κλαδικό επίπεδο, στοιχεία σχετικά με την ΤΝ που προκύπτουν από τις αιτήσεις για διπλώματα ευρεσιτεχνίας, φανερώνουν τη σύνδεσή της με συγκεκριμένους κλάδους, οι οποίοι αποδεικνύεται ότι εφαρμόζουν πρακτικές και συστήματα σχετικά με την ίδια, διευρύνοντας με αυτό τον τρόπο την εμπορική εκμετάλλευσή της. Συγκεκριμένα, οι κλάδοι αυτοί είναι οι εξής: Τηλεπικοινωνίες, Μεταφορές (π.χ. αυτόνομη οδήγηση), Επιστήμες Ζωής και Ιατρική (π.χ. ιατρική βοήθεια, διάγνωση, ανάπτυξη φαρμάκων), Ατομικές συσκευές, Υπολογιστές και Αλληλεπίδραση ανθρώπου-υπολογιστή. Άλλοι τομείς που, επίσης, εμφανίζουν –ελαφρώς μικρότερη- σχέση με την ΤΝ είναι της τραπεζικής και χρηματοοικονομικής (π.χ. έλεγχος και πρόβλεψη ρίσκου, συμβουλευτικές υπηρεσίες), της ψυχαγωγίας (π.χ. βιομηχανία του θεάματος, αποτύπωση της κοινής γνώμης, προώθηση εξατομικευμένων υπηρεσιών δημόσιων σχέσεων), της ασφάλειας, της βιομηχανίας και της μεταποίησης, της γεωργίας, των δικτύων (συμπεριλαμβανομένων των κοινωνικών δικτύων, των «έξυπνων πόλεων» και του «διαδικτύου των πραγμάτων – Internet of Things»), του εμπορίου (π.χ. αύξηση ταχύτητας εκτέλεσης παραγγελιών λιανικού εμπορίου) (WIPO 2019, Zhang & Lu 2021).

Η ΕΕ εντοπίζει μια σειρά από άλλους παραγωγικούς και τεχνολογικούς τομείς που επηρεάζονται άμεσα από την ΤΝ, όπως η κυβερνοασφάλεια, όπου αναμένεται να υπάρξουν περαιτέρω αναβαθμίσεις, εξαιτίας της προόδου στην ΤΝ. Εξελίξεις που θα επηρεάσουν την παραγωγική δραστηριότητα και την κλαδική εξειδίκευση σε θέματα συναφή με την ΤΝ αναμένονται, όπως η παροχή και η άντληση δεδομένων, οι υπολογιστικές δυνατότητες, η ενίσχυση του ενεργειακού αποτυπώματος της ΤΝ και της υπολογιστικής διαδικασίας (π.χ. εξαιτίας της επεξεργασίας τεράστιου όγκου δεδομένων και της διασφάλισης της αειφορίας μέσα από τεχνολογικά και υπολογιστικά προηγμένες λύσεις), η διάθεση και η διαχείριση των δημόσιων δεδομένων (λαμβάνοντας υπόψη τους στόχους, τις ανάγκες και τις προτεραιότητες παρέμβασης του Δημοσίου), η προώθηση κοινών δεξαμενών και πλατφορμών δεδομένων, καθώς και η προώθηση των ανοικτών και έξυπνα διαμοιραζόμενων δεδομένων (χαρακτηριστικό, ως προς αυτό, είναι το Σχήμα 4) (Craglia 2018).

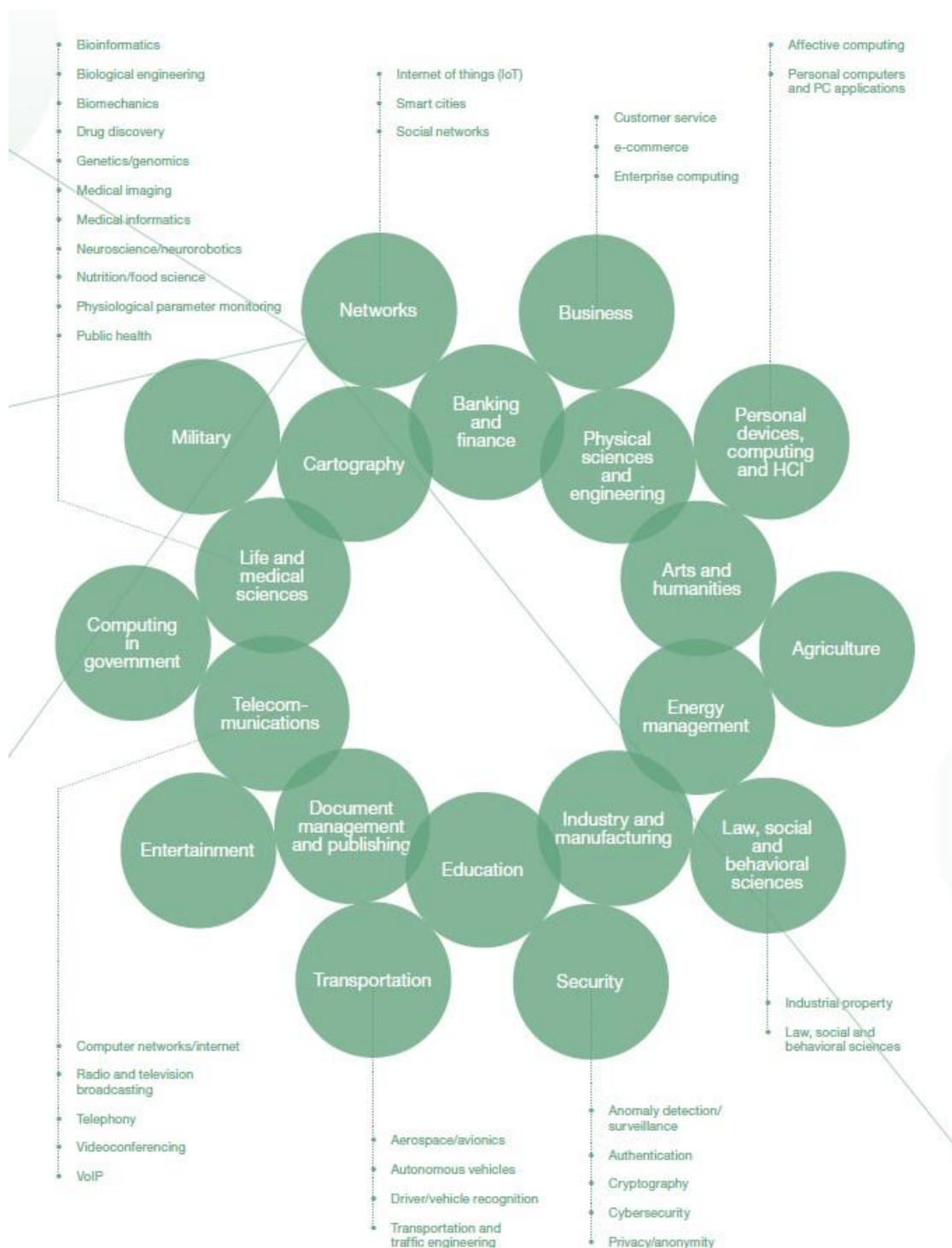
ΣΧΗΜΑ 4: Αναδραστική σχέση μεταξύ δεδομένων και συστημάτων ΤΝ



Πηγή: Craglia (2018)

Στο Σχήμα 5 αποτυπώνονται τα ευρύτερα πεδία εφαρμογής της ΤΝ στη σύγχρονη παραγωγική δραστηριότητα.

ΣΧΗΜΑ 5: Ευρύτερα πεδία εφαρμογής της ΤΝ



Πηγή: WIPO (2019)

1.3 Δημόσιες παρεμβάσεις και μέτρα πολιτικής

Η χάραξη δημόσιας πολιτικής και το μείγμα πολιτικών σχετικών με την ΤΝ είναι δυνατό να καθορίσουν και το μέγεθος των οφελών από αυτή. Ενδεικτικά αναφέρονται η ανάγκη προετοιμασίας στρατηγικής, νόμων και κανονισμών που σχετίζονται με την ΤΝ για την κάλυψη συναφών νομικών και ηθικών ζητημάτων, καθώς και η επεξεργασία ζητημάτων που άπτονται των σχετικών πνευματικών δικαιωμάτων, της πρόσβασης και ιδιοκτησίας ψηφιακών δεδομένων. Η λήψη μέτρων για την άσκηση στοχευμένων πολιτικών, αφενός για την ύπαρξη κατάλληλα εκπαιδευμένου και εξειδικευμένου ανθρώπινου δυναμικού, και αφετέρου για την εξασφάλιση των αναγκαίων πόρων, είναι ιδιαίτερα σημαντική βάσει ενός σχετικού πλαισίου χρηματοδότησης αφιερωμένου στην ΤΝ και μιας αντίστοιχης επενδυτικής στρατηγικής (WIPO 2019).

Σε αυτό το πλαίσιο, το 2017 ήταν η χρονιά που δημοσιεύθηκε η πρώτη εθνική στρατηγική για την ΤΝ στον Καναδά. Μέσα σε λιγότερα από 5 χρόνια, άλλες 30 χώρες έχουν ακολουθήσει μια αντίστοιχη πρακτική, ενώ και διεθνείς οργανισμοί, ασχολούνται συστηματικά σε αυτό το αντικείμενο. Ενδεικτικά αναφέρεται ότι ο ΟΟΣΑ έχει δημιουργήσει Παρατηρητήριο Δημόσιας Πολιτικής για την ΤΝ (AI Policy Observatory), έχει συγκροτήσει Δίκτυο Ειδικών για την ΤΝ (Network of Experts on AI), και έχει οργανώσει ένα νέο Working Party στο πλαίσιο της Επιτροπής για την Πολιτική για την Ψηφιακή Οικονομία (Committee on Digital Economy Policy – CDEP). Επίσης, με υπερεθνικό χαρακτήρα, έχει συγκροτηθεί η Παγκόσμια Συμμαχία για την ΤΝ (Global Partnership on AI - GPAI) η οποία αποσκοπεί στην ενίσχυση της διεθνούς συνεργασίας και της ευρύτερης δραστηριότητας σχετικά με την ΤΝ (Zhang et al. 2021).

Συνολικά, η ρύθμιση είναι σημαντική, κατά την άσκηση δημόσιας πολιτικής, γύρω από τα θέματα της ΤΝ, με βασικό ζητούμενο τα συστήματά της να είναι αξιόπιστα και ανθρωποκεντρικά, καλύπτοντας την ηθική διάσταση, τη δικαιοσύνη, τον σεβασμό των δημοκρατικών αξιών και της ιδιωτικότητας, καθώς και την αποφυγή του κινδύνου μεταφοράς υπαρχουσών προκαταλήψεων από τον πραγματικό στον ψηφιακό κόσμο (π.χ. ως προς το φύλο και τη φυλή). Ως πολιτικές, οι οποίες προωθούν αξιόπιστα συστήματα ΤΝ θεωρούνται, κυρίως, εκείνες που ενθαρρύνουν τις επενδύσεις στην «υπεύθυνη» έρευνα και ανάπτυξη (E&A) σχετικά με την ΤΝ. Τα αναγκαία μέτρα πολιτικής (πρέπει να) στοχεύουν στην αποφυγή του αποκλεισμού για τους μικρότερους –σε μέγεθος και ειδικό βάρος– οικονομικούς δρώντες, στην προώθηση του ανταγωνισμού, στη διασφάλιση της πνευματικής ιδιοκτησίας και στη διευκόλυνση της κινητικότητας στην αγορά εργασίας, δεδομένου ότι οι συναφείς θέσεις απασχόλησης (τείνουν να) εξελίσσονται, ενώ άλλες χαρακτηρίζονται ως ξεπερασμένες ως απόρροια των εξελίξεων στην ΤΝ και των εφαρμογών της.

Επίσης, το δημόσιο είναι σε θέση να θεσπίσει και να προάγει συνεργασίες με τον ιδιωτικό τομέα, γύρω από τη λειτουργία των συστημάτων ΤΝ, κατά περίπτωση. Τέτοιου είδους δραστηριότητες μπορούν να περιλαμβάνουν ελέγχους συμβατότητας, αξιολογήσεις πιστότητας και σχήματα πιστοποιήσεων, λαμβάνοντας υπόψη τις ανάγκες ή τους περιορισμούς των ΜμΕ, οι οποίες, αναμφίβολα, αποτελούν τη ραχοκοκαλιά της ευρωπαϊκής βιομηχανίας και μεταποίησης. Επιπροσθέτως, άξιο αναφοράς είναι το ότι σημαντικές παράμετροι που καθορίζουν την άσκηση δημόσιας πολιτικής στον συγκεκριμένο τομέα είναι οι υφιστάμενες συνθήκες. Καθοριστικοί παράγοντες ως προς αυτό αποτελούν η χωρική διάσταση και το διαθέσιμο ανθρώπινο κεφάλαιο. Ιδιαίτερο ρόλο έχει, επίσης, η «προσφορά γνώσης» από πλευράς της ερευνητικής κοινότητας και η ζήτηση για συναφείς υπηρεσίες, τόσο από το δημόσιο (π.χ. σε όρους εθνικών προτεραιοτήτων) όσο και από τον ιδιωτικό τομέα, καθώς και η υπάρχουσα βιομηχανική και

παραγωγική διάρθρωση της χώρας. Τέλος, η ύπαρξη κατάλληλων (ηλεκτρονικών) υποδομών και η ενίσχυση της δομής, της στρατηγικής και του ανταγωνισμού μεταξύ των επιχειρήσεων, με μέσα πολιτικής και δημόσιας παρέμβασης (π.χ. κίνητρα για την ενίσχυση της ιδιωτικής επένδυσης, αύξηση της ελκυστικότητας για ΑΞΕ στον τομέα της ΤΝ κ.ά.) συνιστούν προσδιοριστικούς παράγοντες αυτής της συζήτησης (OECD 2019).

Επιπροσθέτως, όπως αναφέρουν οι Baruffaldi et al. (2020), ιδίως σε τομείς, όπως η ΤΝ, οι τεχνολογικές και οι επιστημονικές εξελίξεις είναι ταχείες και διαρκώς μεταβαλλόμενες. Αυτό τεκμηριώνεται από την επιταχυνόμενη χρήση συγκεκριμένων εννοιών και «λέξεων-κλειδιών» στις σχετικές με την ΤΝ επιστημονικές και τεχνολογικές εκροές και υποδηλώνει αφενός τις σύγχρονες τάσεις στον τομέα και αφετέρου τη χρησιμότητα στην άσκηση δημόσιας πολιτικής, ιδίως ως προς την υποστήριξη συγκεκριμένων τεχνολογικών και επιστημονικών δραστηριοτήτων και έρευνας σε επιμέρους υπο-πεδία, όπως τα εξής: artificial neural network, convolutional neural network, unmanned aerial vehicles, neural network model. Οι προαναφερθείσες περιοχές, με τα σχετικά πεδία λειτουργίας συστημάτων ΤΝ, υποδεικνύουν – συν τοις άλλοις – τους κλάδους της οικονομίας και της παραγωγής που εμπλέκονται καθώς και τις δραστηριότητες που εξυπηρετούνται από την ΤΝ, άρα και τα συναφή (στοχευμένα) μέτρα πολιτικής.

Η ΕΕ, από την πλευρά της, επισημαίνει την ανάγκη χάραξης και υλοποίησης μιας στρατηγικής για τη διαχείριση των δεδομένων, που αποδεικνύονται κρίσιμης σημασίας για την ΤΝ, ώστε και η ίδια να συμβάλλει στην άσκηση δημόσιας πολιτικής, με τέτοιο τρόπο που να εξειδικεύει τις παρεμβάσεις με βάση τα χαρακτηριστικά κάθε περιοχής ή ομάδας πληθυσμού (Craglia 2018). Σε αυτό το πλαίσιο, η ΕΕ έχει σχεδιάσει παρεμβάσεις στις ακόλουθες περιοχές:

- Ενίσχυση των Κρατών Μελών για τον σχεδιασμό εθνικών στρατηγικών
- Μεγιστοποίηση των επενδύσεων μέσω συνεργασιών
- Μεταφορά ερευνητικών αποτελεσμάτων στην αγορά
- Δεξιότητες και διά βίου μάθηση
- Ύπαρξη/οργάνωση δεδομένων και πρόσβαση σε αυτά
- Αρχές και κανόνες για το ρυθμιστικό πλαίσιο
- Αξιοποίηση της ΤΝ από τον Δημόσιο τομέα
- Διεθνής συνεργασία

Βάσει αυτών, η ΕΕ έχει επισημάνει τους βασικούς στόχους γύρω από τους εξής άξονες: α) καλλιέργεια της πρωτοπορίας σε τομείς σχετικούς με την ΤΝ, β) λειτουργία διευκολυντικών παραγόντων, γ) βελτίωση της ελκυστικότητας της Ευρώπης για σχετικές επενδύσεις, δ) εξασφάλιση της παραδοχής ότι η ΤΝ λειτουργεί προς όφελος των ανθρώπων. Κατά την ΕΕ, διακεκριμένοι τομείς για την ανάπτυξη της ΤΝ είναι οι εξής:

- Κλίμα και περιβάλλον
- Υγεία
- Ρομποτική
- Το Δημόσιο ως χρήστης της ΤΝ
- Τήρηση του νόμου, μετανάστευση και άσυλο
- Ασφαλέστερη κινητικότητα, φιλική για το περιβάλλον
- Αειφόρος γεωργία (EC 2021).

Στην Ελλάδα, καταγράφονται μία σειρά από εθνικές πολιτικές και πρωτοβουλίες οι οποίες στοχεύουν στην ταχεία και βέλτιστη ενσωμάτωση της ΤΝ. Πιο συγκεκριμένα, αυτές αφορούν στα εξής:

- Βίβλος Ψηφιακού Μετασχηματισμού, η οποία εστιάζει (α) στις προϋποθέσεις για την ανάπτυξη της ΤΝ και την οριοθέτηση των εθνικών προτεραιοτήτων, και (β) στην αξιοποίηση των οφελών της ΤΝ ως προς την οικονομική ανάπτυξη και την αντιμετώπιση των κοινωνικών προκλήσεων.
- Εθνική Στρατηγική για την ΤΝ, η οποία θέτει ένα πλαίσιο για μια ολιστική πολιτική σχετική με τη μελλοντική ανάπτυξη και εφαρμογή της ΤΝ στην Ελλάδα.

2. Μεθοδολογία

Στην παρούσα ενότητα παρουσιάζεται η μεθοδολογία που ακολουθήθηκε για τους σκοπούς της έκδοσης, δηλαδή ο τρόπος με τον οποίο έγινε η συλλογή των στοιχείων, η επεξεργασία και η ανάλυσή τους. Πιο συγκεκριμένα, έγιναν τα εξής:

- Συγκέντρωση πρωτογενούς υλικού (άρθρα σε επιστημονικά περιοδικά και παρουσιάσεις συνεδρίων, συνοδεία των περιλήψεών τους) σχετικού με την ΤΝ προς ανάλυση - βλ. Βήματα #1 έως #3.
- Κατασκευή, προσαρμογή και εφαρμογή του μοντέλου μηχανικής Μάθησης BERT (Sachini et al., 2022) με σκοπό την ταξινόμηση των επιστημονικών δημοσιεύσεων στην θεματική περιοχή της ΤΝ - βλ. Βήματα #4 και #5.
- Άντληση των βιβλιομετρικών χαρακτηριστικών των άρθρων - βλ. Βήμα #6.
- Ανάλυση και παρουσίαση των δεικτών που αφορούν τις συγκεκριμένες δημοσιεύσεις - βλ. ενότητα 4.

Βήμα 1: Προσδιορισμός των επιστημονικών δημοσιεύσεων

Ο προσδιορισμός και η αποτύπωση των επιστημονικών περιοδικών και των πρακτικών των συνεδρίων, τα οποία σχετίζονται με την ΤΝ, αποτελεί το πρωταρχικό βήμα για την συλλογή βιβλιομετρικών (μετα)δεδομένων σχετικών με τις ελληνικές επιστημονικές εκροές. Η αναζήτηση, η αναγνώριση και η καταγραφή των συγγενών με τη θεματική της ΤΝ επιστημονικών περιοδικών και πρακτικών συνεδρίων τόσο στον εθνικό όσο και στο διεθνή χώρο, ως πρώτο βήμα, κρίθηκαν απαραίτητες ενέργειες, διότι βάσει αυτών ήταν δυνατό να προσδιοριστούν οι Έλληνες συγγραφείς που εξειδικεύονται σε αυτό το αντικείμενο.

Για την εύρεση όλων των επιστημονικών περιοδικών και των πρακτικών των συνεδρίων, που σχετίζονται με την ΤΝ, διερευνήθηκαν οι κοινώς αποδεκτές βιβλιομετρικές βάσεις δεδομένων καθώς και οι εξειδικευμένες αναφορές. Πιο συγκεκριμένα, εξετάστηκαν οι εξής βάσεις δεδομένων: Web of Science και Scopus. Στη συνέχεια, αναζητήθηκαν περιοδικά και πρακτικά συνεδρίων ταξινομημένα ως "AI" στην πύλη SCImago Journal & Country Rank. Επιπλέον, εξετάστηκαν οι σχετικές εκδόσεις, όπως το Artificial Intelligence Index Report (AI Index Report, 2021) και η έκθεση μέτρησης AI του ΟΟΣΑ (Baruffaldi et al., 2020). Έτσι, διαπιστώθηκε ότι το Artificial Intelligence Index Report έκανε χρήση της βάσης δεδομένων Scopus, των βάσεων δεδομένων Microsoft Academic Graph (MAG), καθώς και δεδομένων από το αποθετήριο arXiv και το NESTA (AI Index Report, 2021). Ο ΟΟΣΑ χρησιμοποίησε δεδομένα αποκλειστικά από τη βάση δεδομένων Scopus.

Για τους σκοπούς της παρούσας έκδοσης, λήφθηκαν υπόψη επιστημονικά περιοδικά τα οποία έχουν ευρετηριαστεί ως σχετιζόμενα με την ΤΝ στις διάφορες βιβλιομετρικές βάσεις. Ως εκ τούτου, αξιοποιώντας τη βάση δεδομένων Scopus, επιδιώχθηκε ο εντοπισμός όλων των περιοδικών και των πρακτικών των συνεδρίων συναφών με την ΤΝ. Η Scopus υιοθετώντας μία συγκεκριμένη εννοιολογική ομαδοποίηση έχει προχωρήσει στη συγκρότηση συγκεκριμένου επιστημονικού υποπεδίου «Τεχνητή Νοημοσύνη» (πεδίο με κωδικό 1702). Έτσι, αντλήθηκε άμεσα μια λίστα 243 περιοδικών και πρακτικών από συνέδρια.

Αντίστοιχα με τη βάση δεδομένων Scopus, η βάση δεδομένων Web of Science εφαρμόζει τη δική της εννοιολογική ομαδοποίηση επιστημονικών περιοδικών για τη συγκρότηση συγκεκριμένου επιστημονικού υποπεδίου «Τεχνητή Νοημοσύνη». Προκειμένου να εμπλουτιστεί το σετ δεδομένων λαμβάνοντας υπόψιν τα διαφορετικά εννοιολογικά ταξινομικά σχήματα πραγματοποιήθηκε αντίστοιχη αναζήτηση στη βάση δεδομένων Web of Science. Εδώ, βρέθηκαν διαθέσιμα 188 σχετικά περιοδικά και πρακτικά από συνέδρια.

Εξερευνώντας τα δεδομένα και προκειμένου να βρεθούν τα κοινά έγγραφα μεταξύ των ομαδοποιήσεων Scopus και Web of Science, προέκυψε ότι, από τα 188 που ευρετηριάστηκαν ως «AI»⁶ στο Web of Science, τα 120 ήταν ήδη καταχωρημένα στο Scopus ως «AI». Τα υπόλοιπα 68, αν και ευρετηριασμένα στο Scopus, δεν ήταν ταξινομημένα ως «AI». Εντούτοις, κι αυτά συμπεριλήφθηκαν στο τελικό σύνολο δεδομένων στοχεύοντας σε μία – όσο το δυνατόν – πληρέστερη αναφορά. Κατά συνέπεια, το νέο σύνολο των επιστημονικών περιοδικών και πρακτικών από συνέδρια είναι $N = 243 + 68 = 311$.

Επιπλέον, εξετάστηκε το SCImago ως συμπληρωματική πηγή δεδομένων. Εδώ, η αναζήτηση περιορίστηκε σε περιοδικά και πρακτικά από συνέδρια τα οποία είχαν δείκτες κατάταξης Q1 και Q2⁷, δεδομένου ότι αυτοί θεωρούνται ως οι πιο αντιπροσωπευτικοί και επιστημονικά με την υψηλότερη απήκηση (Cardoso et al., 2020). Ωστόσο, συνυπολογίζοντας ότι η πηγή δεδομένων της SCImago είναι το Scopus⁸, 243 περιοδικά και πρακτικά από συνέδρια «AI», τα οποία προσδιορίστηκαν εξ αρχής στο Scopus, θα ήταν και ο πληθυσμός της. Όμως, εκτελώντας βασικές εντολές αναζήτησης προέκυψαν, συνολικά, 157 εκ των οποίων τα 102 αποτελούσαν μέρος των 243 του Scopus. Μετά την αφαίρεση των επικαλύψεων, εντοπίστηκαν επιπλέον 55 περιοδικά και πρακτικά από συνέδρια «AI» - βελτιώνοντας, έτσι, την ακολουθούμενη μεθοδολογία του ΟΟΣΑ (Baruffaldi et al., 2020). Συμπερασματικά, τα περιοδικά και τα πρακτικά από συνέδρια που κατηγοριοποιούνται ως σχετικά με την ΤΝ δεν ταυτίζονται υποχρεωτικά μεταξύ SCImago και Scopus. Ωστόσο, για τους σκοπούς της παρούσας έκδοσης, τα περιοδικά και τα πρακτικά από συνέδρια αθροίστηκαν στα 366 ($N=311+55$). Αυτά απαρτίζουν τον πληθυσμό της έρευνας. Επίσης, από αυτά αναζητήθηκαν δημοσιεύσεις επιστημόνων ελληνικών φορέων και ιδρυμάτων (πανεπιστήμια, ερευνητικά κέντρα, κ.ά.) σχετικές με την ΤΝ (βλ. Πίνακα 1).

⁶ Όπως αναφέρεται στην σχετική έρευνα του ΟΟΣΑ (Baruffaldi κ.ά., 2020).

⁷ Πληροφορίες σχετικά με τον τρόπο ταξινόμησης των επιστημονικών δημοσιεύσεων σε quantiles (Q1, Q2, Q3, Q4) είναι διαθέσιμες στον παρακάτω σύνδεσμο: <https://www.scimagojr.com/help.php>

⁸ βλ. <https://www.scimagojr.com/aboutus.php>

ΠΙΝΑΚΑΣ 1: Επιστημονικές εκροές (περιοδικά και πρακτικά συνεδρίων) που εντοπίστηκαν στις βάσεις δεδομένων

Βάσεις Δεδομένων	Αριθμός επιστημονικών περιοδικών
Scopus (a)	<u>243</u>
Web of Science	188
Ανιχνεύτηκαν στη Scopus και στο Web of Science	93
Ανιχνεύτηκαν μόνο στο Web of Science (b)	<u>68</u>
Scimago TN	157
Scimago TN ταξινομημένα στο Scopus ως μη TN (c)	<u>55</u>
Συνολικός αριθμός επιστημονικών δημοσιεύσεων (a+b+c)	366

Βήμα 2: Εντοπισμός των επιστημονικών άρθρων και των πρακτικών συνεδρίων

Λόγω του ότι η έρευνα αφορά σε επιστημονικές εκροές ερευνητών των οποίων η γεωγραφική διεύθυνση επικοινωνίας (affiliation) ήταν η “Ελλάδα”, χρησιμοποιήθηκε η βάση Scopus, στην οποία πραγματοποιήθηκε η εξής αναζήτηση (query):

AFFILIATION COUNTRY = ‘Greece’ and

JOURNALS = “All 366 AI related journals and conferences”

Με αυτό τον τρόπο, συλλέχθηκαν, συνολικά, 7.596 επιστημονικά άρθρα και πρακτικά από συνέδρια τα οποία εντοπίζονται χρονικά μεταξύ του 1980 και του 2021 (τέλος Οκτωβρίου). Επιπλέον, συγκεντρώθηκαν στοιχεία σχετικά με άλλα χαρακτηριστικά των δημοσιεύσεων, όπως τα ονοματεπώνυμα των συγγραφέων, τα έτη δημοσίευσης και οι αναγνωριστικοί κωδικοί (electronic identifier - eid).

Βήμα 3: Συλλογή περιλήψεων

Το βήμα αυτό αφορά στην συλλογή των περιλήψεων των 7.596 άρθρων και πρακτικών συνεδρίων. Σημειώνεται δε, ότι η περίληψη (abstract) επιλέχθηκε ως το πλέον κατάλληλο μέρος της δομής μιας δημοσίευσης για να διερευνηθεί εάν αυτή σχετίζεται με την ΤΝ. Ωστόσο, επισημαίνεται ότι δεν επιλέχθηκαν οι λέξεις-κλειδιά (keywords), διότι λόγω του πολύ μικρού τους αριθμού (έως 5), οι συγγραφείς προτιμούν να βάζουν λέξεις συναφείς με την προσέλκυση του ενδιαφέροντος των αναγνωστών που είναι και εύκολα ανακτήσιμες από τις μηχανές αναζήτησης (Vincent-Lamarre και Larivière, 2021). Συνεπώς, η περίληψη, η οποία περιγράφει συνοπτικά το υπόβαθρο, τις επιδιώξεις και τα αποτελέσματα μίας δημοσίευσης, αποτελεί το πιο αξιόπιστο στοιχείο, προκειμένου να προσδιοριστεί η σχέση κάθε δημοσίευσης με την ΤΝ.

Προς τούτο, έγινε χρήση του δυναμικού περιβάλλοντος αναζήτησης (interactive documentation) της Scopus. Με άλλα λόγια, αξιοποιώντας το Scopus API⁹, ενεργοποιήθηκαν οι σχετικές αναζητήσεις μέσα από το περιβάλλον της Pythοn. Συνολικά, περατώθηκε η συλλογή 7.361 περιλήψεων. Ωστόσο, 235 (= 7.596 - 7.361) δημοσιεύσεις δεν διέθεταν περιλήψεις διότι ήταν άρθρα γνώμης (editorials). Αυτά δεν συμπεριλήφθηκαν στο τελικό σετ δεδομένων (dataset).

Βήμα 4: Κατασκευή του μοντέλου εντοπισμού των δημοσιεύσεων

Έχοντας προσεγγίσει τον αρχικό πληθυσμό-στόχο, δηλαδή τις δημοσιεύσεις επιστημόνων ελληνικών ιδρυμάτων και φορέων που σχετίζονται με την ΤΝ, πρέπει να διερευνηθεί ποιες εξ αυτών έχουν, εν τέλει, άμεση συνάφεια. Επιπλέον, στο πλαίσιο της κατασκευής ενός μοντέλου εντοπισμού των σχετικών με την ΤΝ δημοσιεύσεων, σημειώνεται ότι για τις δημοσιεύσεις σε επιστημονικά περιοδικά πρέπει να γίνει η εξής επισήμανση: επειδή στις περισσότερες βιβλιομετρικές βάσεις, ο τρόπος ταξινόμησης γίνεται σε επίπεδο *περιοδικού* (journal based) και όχι σε επίπεδο *εγγράφου* (document based), δηλαδή άρθρου, παρατηρείται ότι ορισμένα άρθρα τα οποία δεν είναι άμεσα σχετιζόμενα με την ΤΝ, έχουν δημοσιευτεί σε περιοδικά που έχουν κατηγοριοποιηθεί ως σχετικά με την ΤΝ και ευρετηριάζονται ως τέτοια. Ωστόσο, είναι δυνατό να συμβαίνει και το αντίστροφο, δηλαδή, άρθρα τα οποία σχετίζονται με την ΤΝ να μην έχουν ευρετηριαστεί με αυτό το τρόπο στις βιβλιομετρικές βάσεις διότι δεν έχουν ενταχθεί στα αντίστοιχα περιοδικά. Συνεπώς, για τον προσδιορισμό του τελικού πληθυσμού-στόχου, πρέπει να λαμβάνονται υπόψη τέτοιες περιπτώσεις.

Για τον εντοπισμό των σχετικών με την ΤΝ δημοσιεύσεων έγινε χρήση και κατάλληλη προσαρμογή παραμέτρων του μοντέλου μηχανικής μάθησης BERT το οποίο αναπτύχθηκε από την Google (Sachini et al., 2022). Βάσει αυτού, επιδιώκεται η ανίχνευση των περιλήψεων που σχετίζονται αποκλειστικά με την ΤΝ έχοντας ως σημείο αναφοράς τον πληθυσμό που έχει συλλεχθεί (7.361 περιλήψεις).

⁹ <https://dev.elsevier.com/documentation/AbstractRetrievalAPI.wadl>

Επιλογή, κατασκευή και προσαρμογή του μοντέλου

Στον τομέα της μηχανικής μάθησης, οι μέθοδοι που βασίζονται σε νευρωνικά δίκτυα έχουν έρθει στο προσκήνιο τα τελευταία χρόνια, καθώς η απόδοσή τους υπερβαίνει σημαντικά όλες τις προηγούμενες τεχνικές. Συγκεκριμένα, τα μοντέλα Transformer έχουν εξαιρετική απόδοση σε πολλαπλά πεδία εφαρμογής, συμπεριλαμβανομένης της ταξινόμησης κειμένου (Vaswani et al., 2017). Αυτό επιτυγχάνεται με την εφαρμογή μιας στοιβαγμένης αρχιτεκτονικής κωδικοποιητή-αποκωδικοποιητή με μηχανισμούς αυτοπροσοχής. Από τότε, έχουν αναπτυχθεί πολλά μοντέλα που βασίζονται στα μοντέλα Transformer. Τα πιο αξιόπιστα και δημοφιλή είναι τα μοντέλα GPT του OpenAI (Radford & Narasimhan, 2018) και το BERT της Google (Devlin et al., 2019). Το τελευταίο χρησιμοποιείται ευρέως για ταξινόμηση κειμένου (Liu et al., 2019; Lan et al., 2020). Συνεπώς, στο πλαίσιο της παρούσας έκδοσης, γίνεται αξιοποίησή του προκειμένου να περατωθεί η διαδικασία της ταξινόμησης των ελληνικών επιστημονικών δημοσιεύσεων στον τομέα της TN.

Το BERT της Google είναι ένα μοντέλο αναπαράστασης γλώσσας το οποίο βασίζεται στην αρχιτεκτονική Transformer. Αυτό που το διακρίνει είναι η χρήση αμφίδρομων επιπέδων αυτοπροσοχής (bidirectional self-attention layers) που μιμούνται καλύτερα τον τρόπο με τον οποίο οι άνθρωποι επεξεργάζονται κείμενο. Η αρχιτεκτονική του BERT καθιστά επιτακτική τη χρήση του ως βασικού επιπέδου αναπαράστασης γλώσσας που μπορεί να χρησιμοποιηθεί χωρίς να απαιτείται εκπαίδευση του γλωσσικού μοντέλου από την αρχή. Αυτό, μαζί με το γεγονός ότι υπάρχουν πολλές παραλλαγές προεκπαιδευμένων δικτύων BERT, το καθιστά ιδανικό δίκτυο για εφαρμογές Επεξεργασίας Φυσικής Γλώσσας (Devlin et al., 2019). Ως εκ τούτου, για την εν λόγω μελέτη, χρησιμοποιείται το ανωτέρω μοντέλο.

Εκπαιδεύοντας το μοντέλο (training set)

Με στόχο την συλλογή επιστημονικών δημοσιεύσεων οι οποίες αφορούν στην TN, η παρούσα έρευνα, λαμβάνοντας υπόψη τις δύο διαφορετικές σκοπιές που προαναφέρθηκαν ως προς τον τρόπο ταξινόμησης επιστημονικών δημοσιεύσεων σε θεματικές περιοχές, αναλύει τις επιστημονικές δημοσιεύσεις οι οποίες αφορούν στην TN χρησιμοποιώντας και τους δύο τρόπους ταξινόμησης.

Η πρώτη αφορά δεδομένα που έχουν ταξινομηθεί ως σχετιζόμενα με την TN σε επίπεδο περιοδικού (journal based), ενώ η δεύτερη σε επίπεδο εγγράφου (document based). Εξετάστηκαν δύο βιβλιομετρικές βάσεις δεδομένων. Πιο συγκεκριμένα, πραγματοποιήθηκε αναζήτηση στη βάση δεδομένων Scopus για δημοσιεύσεις σε περιοδικά που περιλάμβαναν στον τίτλο τους «Artificial Intelligence», «Machine Learning», «Neural Networks» και «Neural Computing». Συνολικά, συλλέχθηκαν 49.584 περιλήψεις δημοσιεύσεων από αυτά τα περιοδικά για την κατηγορία TN. Εκτός από τη βάση δεδομένων Scopus, χρησιμοποιήθηκε και η βάση δεδομένων arXiv –σε αυτή η διαδικασία ταξινόμησης πραγματοποιείται σε επίπεδο εγγράφου- για την εξαγωγή περαιτέρω δεδομένων για την κατηγορία TN. Από το arXiv συμπεριλήφθηκαν όλες οι περιλήψεις (122.911) από τις εξής κατηγορίες: cs.AI (Τεχνητή Νοημοσύνη) και cs.LG (Μηχανική Μάθηση). Συνολικά, συλλέχθηκαν 172.495 περιλήψεις, από τις οποίες, οι 129.371 χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου (training) ενώ οι υπόλοιπες 43.124 χρησιμοποιήθηκαν για την επαλήθευση του μοντέλου (validation) (βλ. Γραμμή 1 Πίνακα 2).

Για τη συλλογή των απαραίτητων περιλήψεων της κατηγορίας που δεν ταξινομείται ως σχετιζόμενη με την TN (στο εξής «όχι TN»), χρησιμοποιήθηκε η βάση δεδομένων Scopus. Η ανάλυση πραγματοποιείται για καθένα από τα 26 θεματικά πεδία¹⁰ - εξαιρουμένου του Πολυθεματικού πεδίου (Multidisciplinary Sciences). Τυχόν περιοδικά ευρετηριασμένα ως σχετικά με την TN παραλείφθηκαν. Συνολικά, συγκεντρώθηκαν 249.552 περιλήψεις. Από αυτές (βλ. Πίνακα 2), 187.164 περιλήψεις χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου (training), ενώ οι υπόλοιπες 62.388 περιλήψεις χρησιμοποιήθηκαν για την επαλήθευση του αυτού μοντέλου (validation).

ΠΙΝΑΚΑΣ 2: Κατανομή των σχετιζόμενων (και μη) με την TN περιλήψεων στα έγγραφα που εντοπίζονται στο Scopus και το Arxiv.

Dataset	Training set	Validation set
TN	129.371	43.124
Όχι TN	187.164	62.388

Για την ταξινόμηση των ελληνικών επιστημονικών δημοσιεύσεων στον τομέα της TN χρησιμοποιήθηκε μια μικρότερη έκδοση του BERT (Sachini et al., 2022). Το μοντέλο αποτελείται από: 6 Transformer Blocks, Hidden Layer μεγέθους 768 και χρησιμοποιεί 12 κεφαλές προσοχής ($L = 6$, $H = 768$, $A = 12$). Επιπλέον, χρησιμοποιήθηκε ένα Fully Connected Layer, ακολουθούμενο από δυαδική απώλεια διασταυρούμενης εντροπίας (binary cross entropy loss). Ο βελτιστοποιητής AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$) χρησιμοποιήθηκε με αρχικό ρυθμό μάθησης $3e-5$ που μειώνεται σταδιακά σε $1e-6$ και περίοδο προθέρμανσης 10% των συνολικών βημάτων εκπαίδευσης.

Το μοντέλο εκπαιδεύτηκε μόνο για 2 epochs, διότι το σύνολο δεδομένων εκπαίδευσης (training set) είναι αρκετά μεγάλο (105.512 περιλήψεις). Στον Πίνακα Σύγκυσης (βλ. Πίνακα 3), που παρουσιάζεται ο αριθμός των περιλήψεων που επιλέχθηκε για το σετ επαλήθευσης επί του συνόλου, παρατηρείται επιτυχία με ακρίβεια 96.5% [= (TP+TN)/(TP+FP+TN+FN)].

¹⁰ Βλ. https://service.elsevier.com/app/answers/detail/a_id/14882/supporthub/scopus/~/-/what-are-the-most-frequent-subject-area-categories-and-classifications-used-in/

ΠΙΝΑΚΑΣ 3: Πίνακας Σύγχυσης (Confusion Matrix)

Πίνακας Σύγχυσης (Confusion Matrix)	Συχνότητα (# περιλήψεων)	Ποσοστό (%)
True Positive (TP)	41.809	96.95%
False Negative (FN)	1.315	3.05%
True Negative (TN)	60.019	96.20%
False Positive (FP)	2.369	3.80%

Βήμα 5: Προσδιορισμός του τελικού πληθυσμού των δημοσιεύσεων

Ο Πίνακας 4 παραθέτει τα αποτελέσματα που προκύπτουν από την εφαρμογή του μοντέλου πρόβλεψης στις 7.361 περιλήψεις των επιστημονικών δημοσιεύσεων. Από αυτές, 5.306 (72.1%) περιλήψεις ταξινομήθηκαν ως σχετιζόμενες με την TN απαρτίζοντας τον τελικό πληθυσμό-στόχο.

ΠΙΝΑΚΑΣ 4: "TN" επιστημονικές δημοσιεύσεις.

Μεθοδολογικό σχήμα	Συχνότητα	Μερίδιο επί του συνόλου
Σύνολο επιστημονικών δημοσιεύσεων	7.361	100%
Τελικός αριθμός επιστημονικών δημοσιεύσεων σχετικών με TN	5.306	72,1%

Βήμα 6: Συλλογή των βιβλιομετρικών χαρακτηριστικών των επιστημονικών δημοσιεύσεων και των ιδρυμάτων

Με κύριους στόχους την αποτύπωση τόσο της επιστημονικής επίδοσης και απήχησης, όσο και του δικτύου συνεργασίας, σε επίπεδο ατόμων (συγγραφέων) και είδους ιδρυμάτων-φορέων (επιχειρήσεις, ιδρύματα ανώτατης εκπαίδευσης, φορείς δημόσιου τομέα, κλπ.), συλλέχθηκαν, από τις βιβλιομετρικές βάσεις, τα ακόλουθα χαρακτηριστικά των υπό εξέταση δημοσιεύσεων: ο αριθμός ανά φορέα/ίδρυμα, το έτος δημοσίευσης, ο αριθμός των ετεροαναφορών και η γεωγραφική διεύθυνση (χώρα) συνεργασίας για τη συγγραφή (όπου υπάρχει).

3.

Αποτελέσματα για την ελληνική επιστημονική παραγωγή στον τομέα της Τεχνητής Νοημοσύνης

Σε αυτή την ενότητα, παρουσιάζονται τα αποτελέσματα της έρευνας. Η παρουσίαση έχει ως σημεία αναφοράς τον αριθμό των δημοσιεύσεων και των ετεροαναφορών που αυτές έλαβαν -συνολικά, ανά φορέα και ανά κατηγορία φορέα (Πανεπιστήμια, Επιχειρήσεις, Φορείς κρατικού τομέα, Τομέας ιδιωτικών μη κερδοσκοπικών ιδρυμάτων)-, τις κύριες έννοιες (σχετικές με την TN) που ανιχνεύονται στις δημοσιεύσεις καθώς και το δίκτυο των χωρών συνεργασίας.

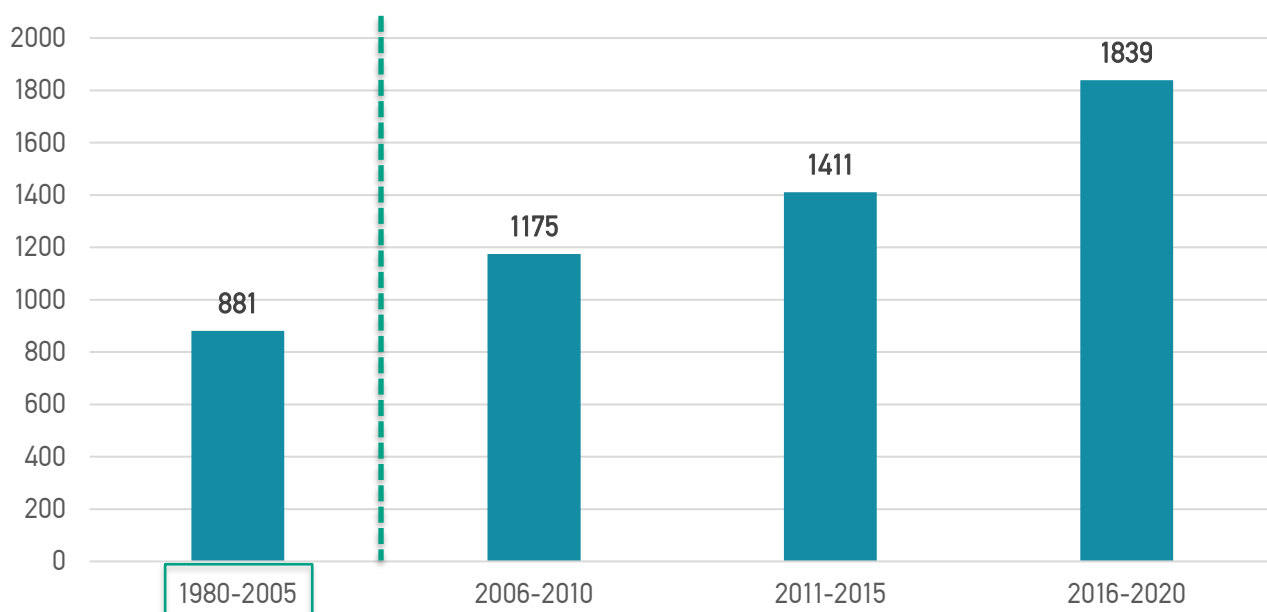
3.1 Αριθμός δημοσιεύσεων, συνολικά και ανά ίδρυμα

Σύμφωνα με τα αποτελέσματα, από το 1980 έως και το 2020, ο αριθμός των επιστημονικών δημοσιεύσεων ελληνικών φορέων στον τομέα της TN ανέρχεται σε 5.306 (Διάγραμμα 1). Είναι χαρακτηριστικό ότι ενώ στη διάρκεια των πρώτων 26 ετών (1980-2005) παρατηρούμε μια πενιχρή απόδοση με συνολικά 881 δημοσιεύσεις (ετήσιος μέσος όρος 33), στη διάρκεια των επόμενων 15 ετών (2006-2020) θα έχουμε μια αλματώδη αύξηση με συνολικά 4.425 δημοσιεύσεις (ετήσιος μέσος όρος 295).

Αναλυτικότερα, ήδη από τη δεκαετία του 1980-1989 εντοπίζονται σε ελληνικούς φορείς δημοσιεύσεις σχετικές με την TN (24 δημοσιεύσεις). Η ακόλουθη δεκαετία (1990 – 1999) συγκεντρώνει σχεδόν δωδεκαπλάσιο αριθμό (308 δημοσιεύσεις), ενώ την πενταετία 2000-2004 σημειώνονται 380 δημοσιεύσεις. Ωστόσο, ο κύριος αριθμός δημοσιεύσεων σχετικών με την TN πραγματοποιείται μετά το 2005, οπότε και παρατηρείται μία συστηματική ετήσια παραγωγή. Συγκεκριμένα, την πενταετία 2006-2010 καταγράφονται 1.175 δημοσιεύσεις, την πενταετία 2011-2015 καταγράφονται 1.411 και την περίοδο 2016-2020 καταγράφονται 1.839.

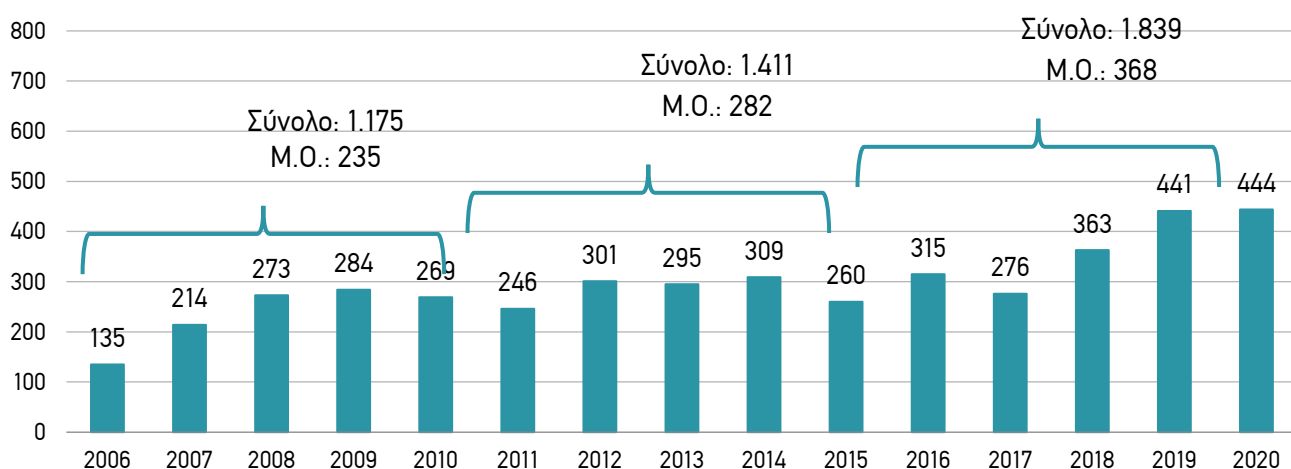
Η παρούσα έκδοση επικεντρώνεται στις σχετικές με την TN δημοσιεύσεις της περιόδου 2006-2020, ακολουθώντας την πρακτική της πιο πρόσφατης έκδοσης του ΕΚΤ στο πλαίσιο της εθνικής στατικής για τα βιβλιομετρικά στοιχεία. Την περίοδο αυτή, 2006-2020 που καλύπτει η παρούσα ανάλυση εντοπίζονται συνολικά, όπως ήδη αναφέραμε, 4.425 επιστημονικές δημοσιεύσεις.

ΔΙΑΓΡΑΜΜΑ 1: Αριθμός επιστημονικών δημοσιεύσεων από ελληνικούς φορείς σχετικών με ΤΝ, 1980-2020.



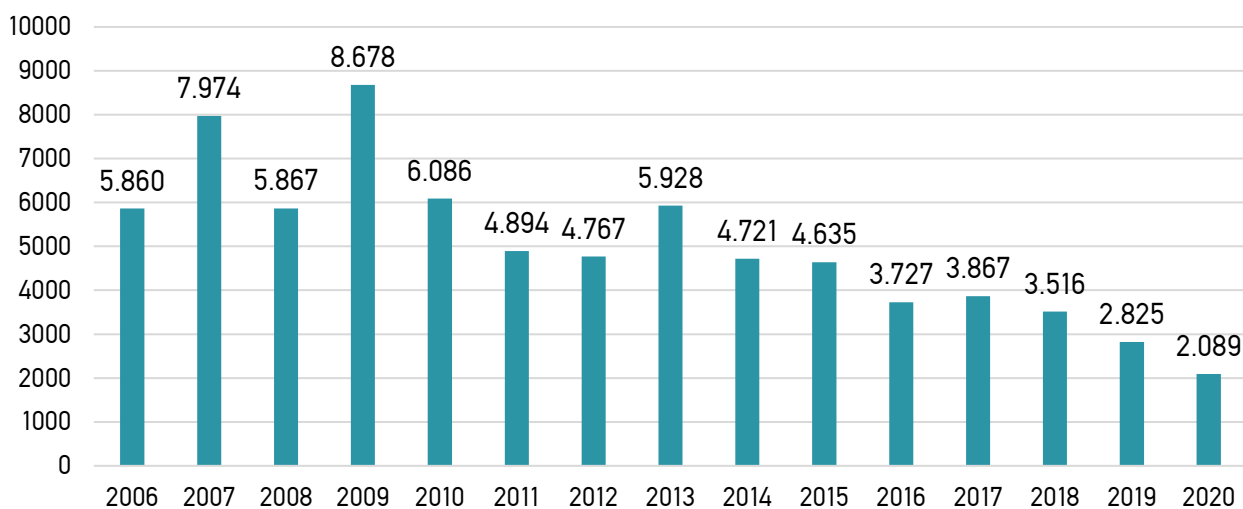
Το ακόλουθο διάγραμμα (Διάγραμμα 2) παρουσιάζει την ετήσια κατανομή των 4.425 επιστημονικών δημοσιεύσεων από το 2006 ως το 2020. Έχει ενδιαφέρον να επισημανθεί ότι οι ελληνικές επιστημονικές δημοσιεύσεις που σχετίζονται με την Τεχνητή Νοημοσύνη παρουσιάζουν μια σταθερά αυξητική τάση την τελευταία δεκαετία (περίοδος αναφοράς της έρευνας του ΕΚΤ). Ειδικότερα, από 1.175 δημοσιεύσεις την πρώτη πενταετία (2006-2010, ετήσιος μέσος όρος 235) αυξάνονται αρχικά σε 1.411 τη δεύτερη (2011-2015, ετήσιος μέσος όρος 282) και σε 1.839 την τρίτη (2016-2020, ετήσιος μέσος όρος 368).

ΔΙΑΓΡΑΜΜΑ 2: Ετήσια κατανομή δημοσιεύσεων ελληνικών φορέων σχετικών με ΤΝ, 2006-2020.



Το Διάγραμμα 3 παρουσιάζει τον αριθμό των ετεροαναφορών που έλαβαν οι δημοσιεύσεις κάθε έτους την χρονική περίοδο 2006 – 2020. Επισημαίνεται ωστόσο ότι ο αριθμός των αναφορών σε μια επιστημονική εργασία εξαρτάται από το χρονικό διάστημα που έχει παρέλθει μετά τη δημοσίευσή της. Ως εκ τούτου, οι παλαιότερες δημοσιεύσεις έχουν και περισσότερες αναφορές χωρίς αυτό να συνδέεται πάντα με την απήχσή τους στην επιστημονική κοινότητα¹¹.

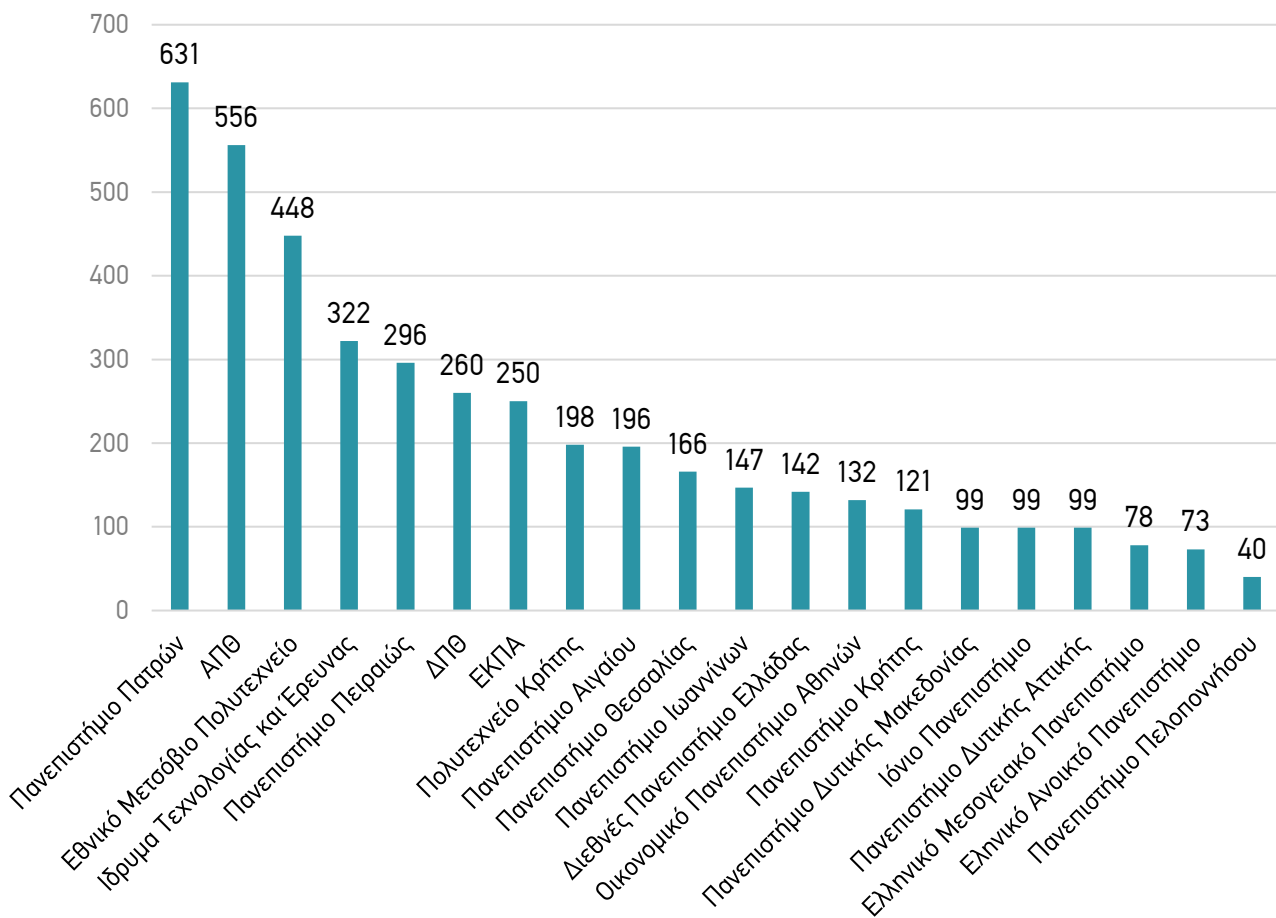
ΔΙΑΓΡΑΜΜΑ 3: Αριθμός αναφορών στις σχετικές με ΤΝ επιστημονικές δημοσιεύσεις κάθε έτους, 2006-2020.



Το επόμενο διάγραμμα (Διάγραμμα 4) παρουσιάζει την κατανομή του αριθμού των σχετικών με την ΤΝ επιστημονικών δημοσιεύσεων ανά ελληνικό ερευνητικό φορέα κατά την περίοδο 2006-2020. Τα πανεπιστήμια καταλαμβάνουν τις 9 από τις πρώτες δέκα θέσεις. Πιο συγκεκριμένα, το Πανεπιστήμιο της Πάτρας κατατάσσεται στην πρώτη θέση με 631 δημοσιεύσεις, ενώ ακολουθεί το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης με 556 δημοσιεύσεις. Έπονται το Εθνικό Μετσόβιο Πολυτεχνείο με 448 δημοσιεύσεις, το Ίδρυμα Τεχνολογίας και Έρευνας (ΙΤΕ) με 322 δημοσιεύσεις (ο μοναδικός δημόσιος ερευνητικός φορέας στη λίστα), το Πανεπιστήμιο Πειραιώς με 296 δημοσιεύσεις, το Δημοκρίτειο Πανεπιστήμιο Θράκης με 260 δημοσιεύσεις, το Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών (ΕΚΠΑ) με 250 δημοσιεύσεις, το Πολυτεχνείο Κρήτης με 198 δημοσιεύσεις, το Πανεπιστήμιο Αιγαίου με 196 δημοσιεύσεις και το Πανεπιστήμιο Θεσσαλίας με 166 δημοσιεύσεις.

¹¹ Για την εξομάλυνση των διαφορών που συνδέονται με τη φυσιολογική αύξηση του αριθμού των αναφορών σε παλαιότερες δημοσιεύσεις, η ανάλυση των αναφορών στις στατιστικές που παράγει το ΕΚΤ πραγματοποιείται με τη βιβλιομετρική τεχνική μέτρησης των αναφορών με χρήση μεταβλητού χρονικού παραθύρου ανά άρθρο και συγκεκριμένα σε επικαλυπτόμενα χρονικά διαστήματα πέντε ετών (overlapping 5 years periods). Τα διαθέσιμα στοιχεία στην παρούσα μελέτη δεν ήταν επαρκή για τη σχετική ανάλυση.

ΔΙΑΓΡΑΜΜΑ 4. Αριθμός επιστημονικών δημοσιεύσεων σχετικών με ΤΝ, ανά ελληνικό ερευνητικό φορέα, 2006-2020.



Σημείωση: Το διάγραμμα απεικονίζει τα πρώτα 20 ιδρύματα βάσει του αριθμού των επιστημονικών δημοσιεύσεων

Η κατάταξη των ιδρυμάτων μεταξύ διαφορετικών κατηγοριών¹² παρουσιάζεται στο επόμενο διάγραμμα (Διάγραμμα 5). Με δεδομένο, ότι για τις ανάγκες αυτής της ανάλυσης, ο υπολογισμός της κατανομής των δημοσιεύσεων ανά κατηγορία φορέα στη βάση της κλασματικής μέτρησης (fractional counting) εμφανίζει σημαντικά πλεονεκτήματα (CWTS, 2015), τα δεδομένα που παρατίθενται στο Διάγραμμα 5 έχουν προκύψει με την αξιοποίηση αυτής της μεθόδου.

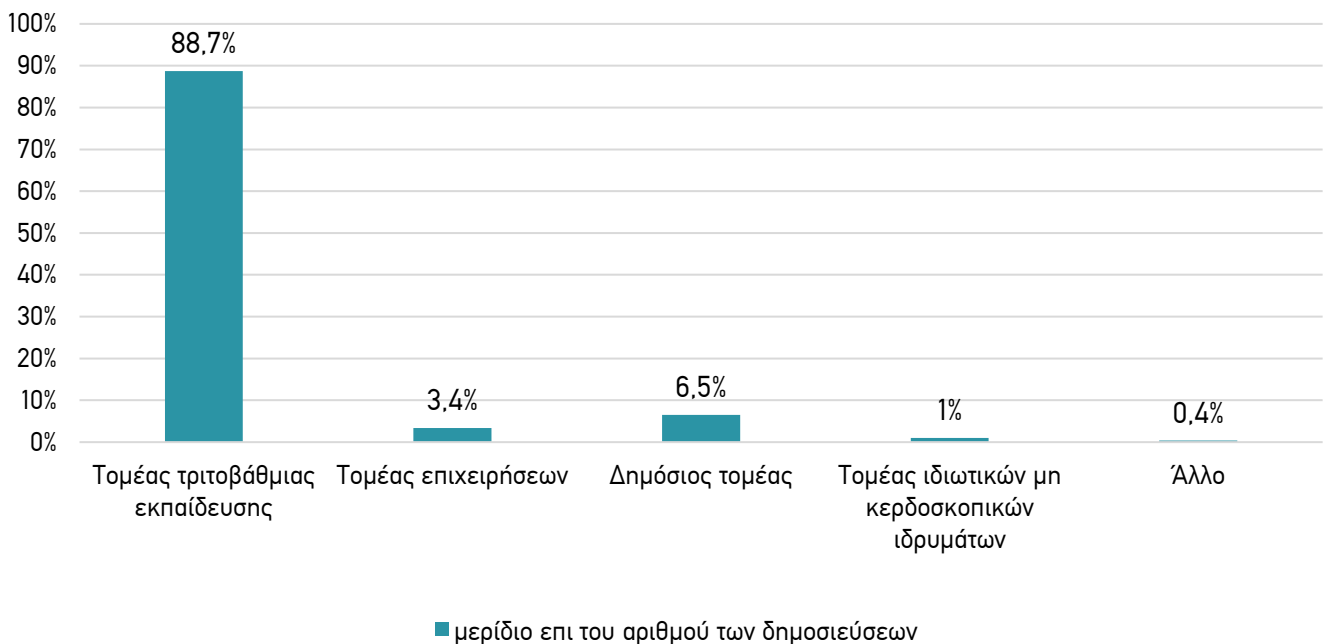
¹² Η κατηγοριοποίηση των ελληνικών φορέων γίνεται σύμφωνα με τα κριτήρια ταξινόμησης του εγχειριδίου Frascati και ακολουθεί την ταξινόμηση των φορέων που εφαρμόζει το ΕΚΤ στις επίσημες ευρωπαϊκές και εθνικές στατιστικές που παράγει. Η κατηγοριοποίηση Frascati για την εκτέλεση δραστηριοτήτων Ε&Α διακρίνει τους ακόλουθους τέσσερις τομείς: Τομέας τριτοβάθμιας εκπαίδευσης (HES - Higher Education Sector), Κρατικός Δημόσιος τομέας (GOV - Government Sector), Τομέας επιχειρήσεων (BES - Business Sector) και Ιδιωτικός μη κερδοσκοπικός τομέας (PNP - Private Non Profit Sector). Η κατηγορία «Άλλο» αφορά ένα ποσοστό φορέων (5.3%) οι οποίοι ήταν καταγεγραμμένοι στις βιβλιομετρικές βάσεις με τέτοιο τρόπο που καθιστούσε αδύνατη την ταξινόμησή τους σε κάποιον από τους προαναφερθέντες τομείς.

Συγκεκριμένα, το Διάγραμμα 5 απεικονίζει την ποσοστιαία κατανομή δημοσιεύσεων ανά κατηγορία ελληνικών ιδρυμάτων-φορέων. Στα ιδρύματα της τριτοβάθμιας εκπαίδευσης (44 ιδρύματα, συμπεριλαμβανομένων των ΑΕΙ και άλλων φορέων, όπως τα πανεπιστημιακά νοσοκομεία) αντιστοιχεί το 88,7% των σχετικών με την ΤΝ δημοσιεύσεων, με το Πανεπιστήμιο Πάτρας, το Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης και το Εθνικό Μετσόβιο Πολυτεχνείο να συγκεντρώνουν τις περισσότερες δημοσιεύσεις.

Σε φορείς του δημόσιου τομέα (σύνολο 32 φορείς) αποδίδεται το 6,5% των δημοσιεύσεων με πρώτο στη σχετική κατάταξη το Ίδρυμα Τεχνολογίας και Έρευνας (ΙΤΕ), ακολουθούμενο από το Ινστιτούτο Τεχνολογίας Υπολογιστών και Εκδόσεων “Διόφαντος” και το Εθνικό Κέντρο Έρευνας Φυσικών Επιστημών “ΔΗΜΟΚΡΙΤΟΣ”.

Στις επιχειρήσεις, οι οποίες συνολικά ανέρχονται στις 90, αναλογεί το 3,4% των δημοσιεύσεων, με επιχειρήσεις, όπως η Athens Technology Center, IRIDA Labs και Albio Data να συγκεντρώνουν τις περισσότερες δημοσιεύσεις.

ΔΙΑΓΡΑΜΜΑ 5. Ποσοστιαία κατανομή δημοσιεύσεων ελληνικών ερευνητικών φορέων ανά κατηγορία, 2006-2020.

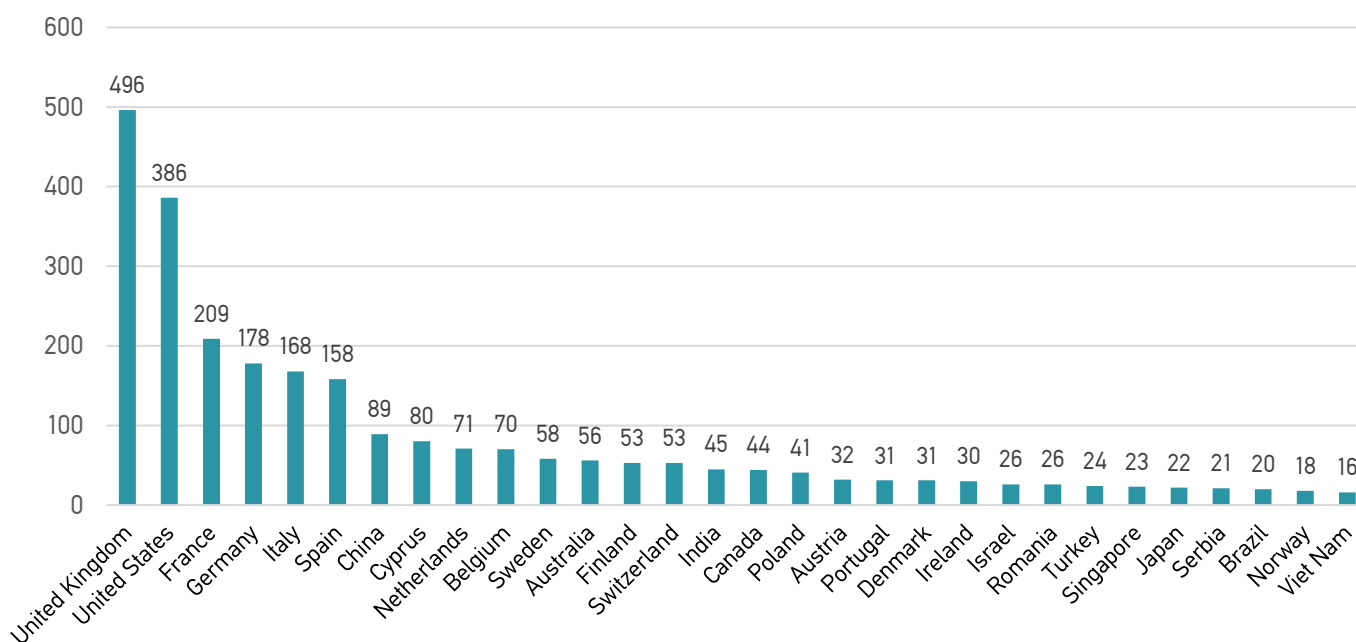


3.3 Ανάλυση διεθνών δικτύων για τη συγγραφή των επιστημονικών δημοσιεύσεων ΤΝ

Το ακόλουθο διάγραμμα (Διάγραμμα 6) παρουσιάζει τις 30 πρώτες, ως προς τη συχνότητα εμφάνισής τους, χώρες με τις οποίες συνεργάζονται οι ελληνικοί φορείς στις σχετικές με την ΤΝ επιστημονικές δημοσιεύσεις της περιόδου 2006-2020.

Συνολικά, οι χώρες που αντιστοιχούν στη γεωγραφική διεύθυνση των συν-συγγραφέων εκτός Ελλάδας ανέρχονται σε 71. Αξιοποιώντας τις γεωγραφικές διευθύνσεις (affiliations) των συγγραφέων, καταγράφεται ότι η Μεγάλη Βρετανία, οι Η.Π.Α., η Γαλλία, Γερμανία, η Ιταλία, Ισπανία, η Κίνα, η Κύπρος, η Ολλανδία και το Βέλγιο αποτελούν τις χώρες στις οποίες βρίσκονται τα ιδρύματα με τα οποία είναι συνδεδεμένοι οι περισσότεροι επιστήμονες ελληνικών φορέων και ιδρυμάτων στον τομέα της ΤΝ.

ΔΙΑΓΡΑΜΜΑ 6. Χώρες με τις οποίες συνεργάζονται οι ελληνικοί φορείς για τη συγγραφή επιστημονικών δημοσιεύσεων ΤΝ (αριθμός συνεργασιών ανά χώρα, 2006-2020).



Με σκοπό να παρουσιαστεί το διεθνές δίκτυο συνεργασιών συγγραφής ελληνικών επιστημονικών δημοσιεύσεων σχετικών με την ΤΝ χρησιμοποιήθηκαν τεχνικές από την Θεωρία Γράφων. Με σημείο αναφοράς την Ελλάδα (ego network), γίνεται απεικόνιση των μοτίβων συνεργασίας τόσο σε σχέση με την ίδια (ego) όσο και μεταξύ των λοιπών χωρών (alters). Συγκεκριμένα, η κατασκευή του δικτύου (βλ. Σχήμα 4¹³) συμμορφώθηκε με τον εξής κανόνα: Δύο κόμβοι (χώρες) συνδέονται μεταξύ τους με μια ακμή (edge) αν μοιράζονται (έχουν συγγράψει) τουλάχιστον μία επιστημονική δημοσίευση. Το μέγεθος των κόμβων, το πάχος των ακμών και της γραμματοσειράς των χωρών είναι ανάλογο του αριθμού των επιστημονικών δημοσιεύσεων κάθε χώρας. Με άλλα λόγια, όσο πιο μεγάλος είναι ο αριθμός επιστημονικών δημοσιεύσεων της χώρας, τόσο πιο μεγάλος εμφανίζεται ο κόμβος σε μέγεθος και γραμματοσειρά καθώς και πιο παχιά είναι η ακμή μεταξύ δύο κόμβων.

Στο ακόλουθο σχήμα (Σχήμα 4), η Ελλάδα αποτελεί τον “κεντρικό” κόμβο του δικτύου (ego) διότι και οι 4.425 επιστημονικές δημοσιεύσεις έχουν την Ελλάδα ως γεωγραφική διεύθυνση επικοινωνίας. Συνεπώς, ο κόμβος (πράσινος) ο οποίος αφορά την Ελλάδα έχει το μεγαλύτερο μέγεθος σε σχέση με τους υπόλοιπους καθώς το σύνολο των επιστημονικών δημοσιεύσεων που αφορούν την ΤΝ έχει την Ελλάδα ως τη χώρα με την μεγαλύτερη συχνότητα εμφάνισης σε σχέση με οποιαδήποτε άλλη χώρα.

Το δίκτυο του Σχήματος 4 αποτυπώνει τις πρώτες 29 χώρες (πέρα από την Ελλάδα) με τις οποίες οι Έλληνες συγγραφείς έχουν την μεγαλύτερη δικτύωση. Σε σχέση με την τοποθέτηση των κόμβων στο δίκτυο, οι χώρες που είναι τοποθετημένες σε κοντινή απόσταση μεταξύ τους έχουν ισχυρούς “επιστημονικούς δεσμούς” (π.χ. Ελλάδα - “Greece” και Ηνωμένο Βασίλειο - “United Kingdom”). Το αντίστροφο ισχύει για τους κόμβους (χώρες) οι οποίοι βρίσκονται σε μακρινή απόσταση μεταξύ τους (π.χ. Βραζιλία - “Brazil” και ΗΠΑ - “United States”). Με άλλα λόγια, η απόσταση μεταξύ δύο κόμβων είναι αντιστρόφως ανάλογη της συνεργασίας των δύο χωρών αναφορικά με την παραγωγή επιστημονικών δημοσιεύσεων.

Συνολικά, το δίκτυο αποτελείται από πέντε (5) συστάδες (clusters), κάθε μια από τις οποίες αποτελείται από τουλάχιστον τρεις (3) χώρες (βλ. Πίνακα 3). Αυτή η επιλογή αποτελεί μια βασική παράμετρο για την συνολική κατασκευή και οριστικοποίηση του συνολικού αριθμού των συστάδων¹⁴. Κάθε μια ομάδα χωρών (συστάδα) αποτυπώνει συγκεκριμένα μοτίβα συνεργασίας τα οποία προκύπτουν από τα διαφορετικά προφίλ συνεργιών μεταξύ των συγγραφέων-ερευνητών.

Λαμβάνοντας υπόψη το Διάγραμμα 6 καθώς και το Σχήμα 4, επιβεβαιώνεται ότι η Μεγάλη Βρετανία, οι Η.Π.Α., η Γαλλία, η Γερμανία, η Ιταλία και η Ισπανία αποτελούν τις χώρες με τη μεγαλύτερη συχνότητα αναφορικά με την συγγραφή επιστημονικών δημοσιεύσεων.

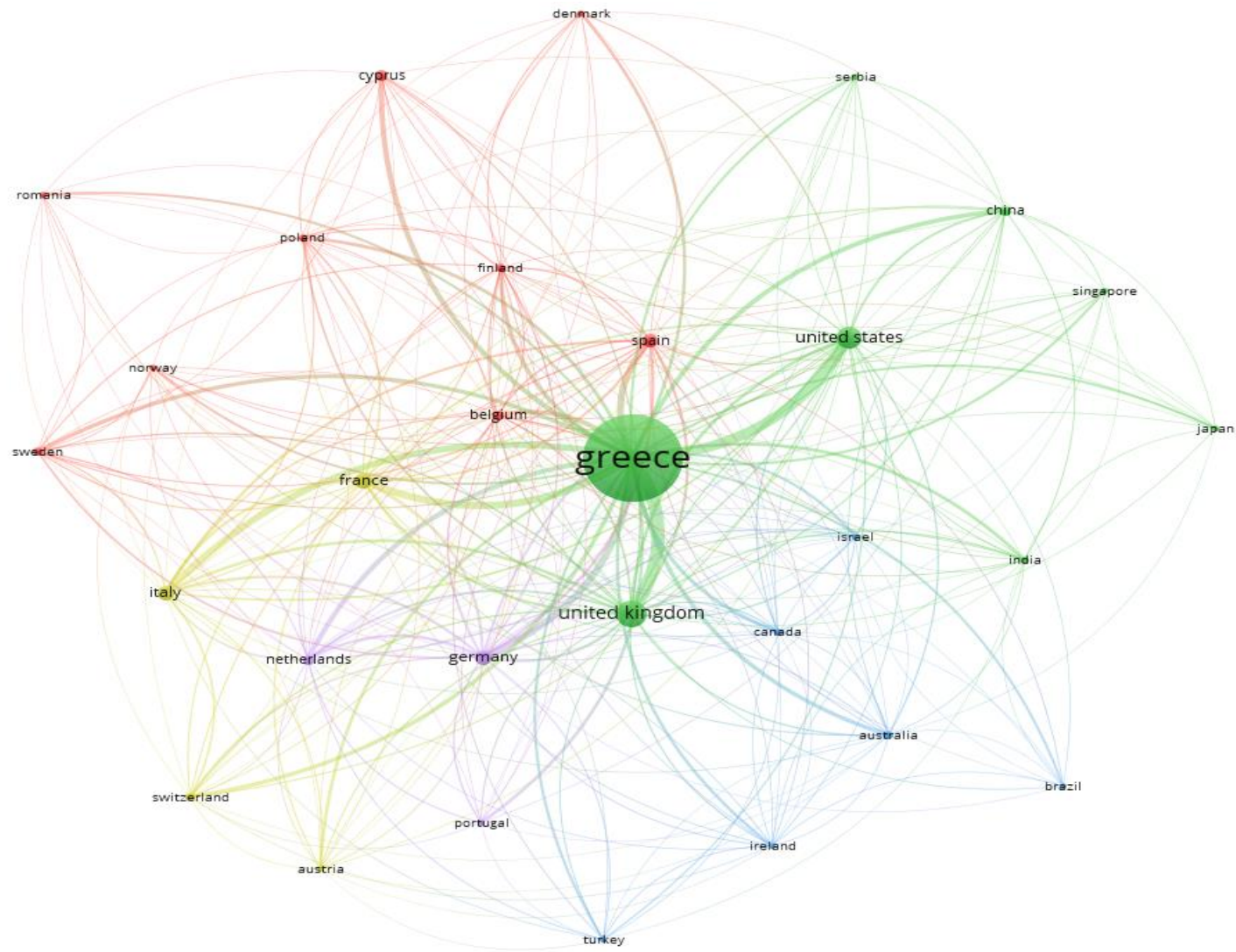
¹³ Παρεμφερή γραφήματα-σχήματα, σε πιο απλή μορφή, είναι η οπτική αναπαράσταση των σχέσεων που αναπτύσσουν ορισμένες ποσότητες, σχεδιασμένες σε σχέση με ένα σύνολο αξόνων. Πιο συγκεκριμένα, πρόκειται για μια απεικόνιση αποτελούμενη από ένα σύνολο σημείων (κόμβων) που συνδέονται με γραμμές (ακμές).

¹⁴ Για περισσότερες πληροφορίες σχετικά με τον τρόπο κατασκευής του δικτύου βλ. Sachini et al., (2021), υποενοότητα “Methodology and Network”.

ΠΙΝΑΚΑΣ 3: Κατανομή των συστάδων του διεθνούς δικτύου συνεργασιών/χωρών των ελληνικών φορέων στις σχετικές με ΤΝ επιστημονικές δημοσιεύσεις.

Συστάδες	Χώρες	Αριθμός
Κόκκινη	Βέλγιο, Ισπανία, Κύπρος, Δανία, Φινλανδία, Νορβηγία, Πολωνία, Ρουμανία, Σουηδία	9
Πράσινη	Ελλάδα, ΗΠΑ, Ηνωμένο Βασίλειο, Κίνα, Ιαπωνία, Ινδία, Σιγκαπούρη, Σερβία	8
Μπλε	Καναδάς, Ισραήλ, Αυστραλία, Βραζιλία, Ιρλανδία, Τουρκία	6
Κίτρινη	Γαλλία, Ιταλία, Αυστρία, Ελβετία	4
Μωβ	Γερμανία, Ολλανδία, Πορτογαλία	3

ΣΧΗΜΑ 4. Διεθνές δίκτυο συνεργασιών ελληνικών φορέων σε επιστημονικές δημοσιεύσεις σχετικές με ΤΝ



Βιβλιογραφία

Baruffaldi, S., van Beuzekom, B., Dernis, H., Harhoff, D., Rao, N., Rosenfeld, D. & Squicciarini, M. (2020). Identifying and measuring developments in artificial intelligence: Making the impossible possible. OECD Science, Technology and Industry Working Papers 2020/05.

Cardoso, L., Silva, R., Almeida, G. G. F. D., & Lima Santos, L. (2020). A bibliometric model to analyze country research performance: SciVal topic prominence approach in tourism, leisure and hospitality. *Sustainability*, 12(23), 9897.

Craglia (Edit.) (2018). *Artificial Intelligence: A European Perspective*. Luxembourg: EU Publications Office.

CWTS. (2015). Waltman, L., & Van Eck, N.J. Field-normalized citation impact indicators and the choice of an appropriate counting method. *Journal of Informetrics*, 9(4), 872-894. Available at: <https://www.cwts.nl/research/publications>

Devlin, J., Chang, M., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *ArXiv*, abs/1810.04805.

EC (2021). Annexes to the Communication from the Commission to the European Parliament, the European Council, the Council, the European Economic and Social Committee and the Committee of the Regions Fostering a European approach to Artificial Intelligence. COM(2021) 205 final.

Ευρωπαϊκή Επιτροπή (2021). Πρόταση. Κανονισμός του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου για τη θέσπιση εναρμονισμένων κανόνων σχετικά με την Τεχνητή Νοημοσύνη (Πράξη για την Τεχνητή Νοημοσύνη) και για την τροποποίηση ορισμένων νομοθετικών πράξεων της Ένωσης. COM(2021) 206 final.

Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., & Soricut, R. (2020). ALBERT: A Lite BERT for Self-supervised Learning of Language Representations. *ArXiv*, abs/1909.11942.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., & Stoyanov, V. (2019). RoBERTa: A Robustly Optimized BERT Pretraining Approach. *ArXiv*, abs/1907.11692.

McCarthy, J. (2007). What is artificial intelligence?, διαθέσιμο στο http://35.238.111.86:8080/jspui/bitstream/123456789/274/1/McCarthy_John_What%20is%20artificial%20intelligence.pdf

Misuraca, G. & Van Noordt, C., (2020). AI Watch - Artificial Intelligence in public services, Publications Office of the European Union, Luxembourg, 2020. <https://publications.jrc.ec.europa.eu/repository/handle/JRC120399>

OECD (2019). Artificial Intelligence in Society. Paris: OECD Publishing.

OECD (2021). AI measurement in ICT usage surveys. A review. OECD Digital Economy Papers No. 308.

Radford, A., & Narasimhan, K. (2018). Improving Language Understanding by Generative Pre-Training. Διαθέσιμο στο: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf

Sachini, E., Sioumalas-Christodoulou, K., Chrysomallidis, C., Siganos, G., Bouras, N., & Karampekios, N. (2021). COVID-19 enabled co-authoring networks: a country-case analysis. *Scientometrics*, 126(6), 5225-5244.

Sachini, E., Sioumalas-Christodoulou, K., Christopoulos, S., & Karampekios, N. (2022). AI for AI: Using AI methods for classifying AI science documents. *Quantitative Science Studies*, 1-19.

Vaswani, A., Shazeer, N.M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. ArXiv, abs/1706.03762.

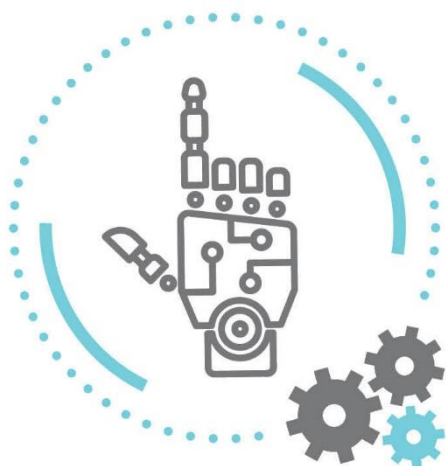
Vincent-Lamarre, P., & Larivière, V. (2021). Textual analysis of artificial intelligence manuscripts reveals features associated with peer review outcome. *Quantitative Science Studies*, 1-27.

WIPO (2019). Artificial Intelligence. WIPO Technology Trends 2019. Geneva: World Intellectual Property Organization.

World Bank. (2020). Artificial Intelligence in the Public Sector: Maximizing Opportunities, Managing Risks. Equitable Growth, Finance and Institutions Insight. World Bank, Washington, DC. <https://openknowledge.worldbank.org/handle/10986/35317>

Zhang, C. and Lu, Y. (2021). Study on artificial intelligence: The state of the art and future prospects, *Journal of Industrial Information Integration*, 23, doi.org/10.1016/j.jii.2021.100224.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J.C., Sellitto, M., Shoham, Y., Clark, J. and Perrault, R. (2021). The AI Index 2021 Annual Report. Stanford: Human-Centered AI Institute, Stanford University



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
Υπουργείο Ψηφιακής Διακυβέρνησης



**ΕΘΝΙΚΟ ΚΕΝΤΡΟ ΤΕΚΜΗΡΙΩΣΗΣ &
ΗΛΕΚΤΡΟΝΙΚΟΥ ΠΕΡΙΕΧΟΜΕΝΟΥ**

ISBN: 978-618-5557-35-5 (pdf)

ISBN: 978-618-5557-34-8 (print)