

Παραγωγική Τεχνητή Νοημοσύνη: Μολύνοντας την Επιστημονική Σκέψη

Το παρόν κείμενο ανακοίνωσης παρουσιάστηκε στο:

31ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών – 22-24/10/2025

Ιωάννινα

Βιβλιογραφική αναφορά ανακοίνωσης στα ελληνικά & αγγλικά:

- Χαλεπλιόγλου, Α. & Ζαμπάκογλου, Χ. (2025). Παραγωγική Τεχνητή Νοημοσύνη: Μολύνοντας την Επιστημονική Σκέψη. *31ο Πανελλήνιο Συνέδριο Ακαδημαϊκών Βιβλιοθηκών*. Ακαδημαϊκή ευημερία, ελευθερία και ακεραιότητα: Από το ΑΒ στο ΑΙ, Ιωάννινα. <https://olympias.lib.uoi.gr/jspui/handle/123456789/39311>
- Chaleplioglou, A., & Zampakolas, C. (2025). The generative ai: Contaminating scientific thought. *31st Panhellenic Academic Libraries Conference*. Academic Well-being, Freedom, and Integrity: From AB to AI, Ioannina. <https://olympias.lib.uoi.gr/jspui/handle/123456789/39311>

Περιεχόμενα ηλεκτρονικού αρχείου:

- Πλήρες κείμενο εργασίας / full-text in Greek
- Στοιχεία ανακοίνωσης στην ελληνική γλώσσα
- Conference paper's details in English

Για φιλικότερη πλοήγηση συνιστάται εμφάνιση του δυναμικού παραθύρου περιήγησης (επικεφαλίδων/σελιδοδεικτών).

Το περιεχόμενο διατίθεται **ελεύθερα** με άδεια χρήσης: Creative Commons Attribution CC **BY-NC-ND** 4.0 International License

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



Πλήρες κείμενο εργασίας / full-text in Greek

Εισαγωγή

Η δημόσια διανομή του μοντέλου Παραγωγικής Τεχνητής Νοημοσύνης (ΠΤΝ), ChatGPT (Chat Generative Pretrained Transformer), έκδοση GPT-3.5, στις 30 Νοεμβρίου 2022, προκάλεσε ένα μαζικό κύμα κοινωνικών, οικονομικών, πολιτικών και φιλοσοφικών συζητήσεων σχετικά με τη χρήση του και τις επιπτώσεις σε όλους τους τομείς της ανθρώπινης δραστηριότητας. Ο σάλος αυτός προκλήθηκε εξαιτίας των γλωσσικών και γνωσιακών δυνατοτήτων του συστήματος, καθιστώντας το ικανό να μιμείται την ανθρώπινη εκφορά ή γραφή λόγου στις απαντήσεις του στα ερωτήματα των χρηστών, μάλιστα με διαφορετικό ύφος, απλό, σύνθετο, επιστημονικό, τοπικό και σε διαφορετικές γλώσσες. Επιχειρηματικοί κολοσσοί στον χώρο της πληροφορικής ανέπτυξαν τα δικά τους μοντέλα ΠΤΝ, ώστε σήμερα είναι διαθέσιμα στους χρήστες του Παγκόσμιου Ιστού, χωρίς χρέωση ή με συνδρομή, πολλά ανάλογα μοντέλα.

Οι δυνατότητες εκφοράς γραπτού επιστημονικού λόγου από συστήματα ΠΤΝ, στην αγγλική γλώσσα, είναι τέτοιες, ώστε πολύ σύντομα εμφανίστηκαν επιστημονικά άρθρα που ονομάτιζαν τα μοντέλα αυτά σαν συν-συγγραφείς, προκαλώντας την αντίδραση μεγάλου μέρους της επιστημονικής κοινότητας (Stokel-Walker, 2023; Thorp, 2023). Οι αντιδράσεις αυτές αφορούσαν κυρίως το σκέλος της συγγραφικής ευθύνης, της αδυναμίας των συστημάτων να αναλάβουν την ευθύνη για τα γραφόμενά ή τα σφάλματά τους. Οι προγραμματιστές των ΠΤΝ συστημάτων εξέδωσαν προειδοποιήσεις αναφορικά με την πιθανότητα λανθασμένων ή παράλογων απαντήσεων, ευαισθησίας στη διατύπωση της ερώτησης, υπερβολικής φλυαρίας, εσφαλμένων υποθέσεων και επιβλαβούς ή προκατειλημμένης συμπεριφοράς. Ωστόσο, παρά τις προειδοποιήσεις ή τις αντιδράσεις έγινε φανερό πως δεν υπήρχε δυνατότητα διάκρισης επιστημονικού κειμένου που έχει παραχθεί από την μηχανή από αυτό που έχει συγγράψει άνθρωπος (Curtis, 2023). Θα μπορούσε να χρησιμοποιηθεί, χωρίς να αποκαλυφθεί η χρήση του, σε όλες τις φάσεις σύνθεσης ενός επιστημονικού

έργου, από τη σχεδίαση, την πραγματοποίηση, την ανάλυση των αποτελεσμάτων, τη συγγραφή της επιστημονικής εργασίας ή κατά την κρίση μιας επιστημονικής εργασίας, ακόμη και για την αποδοχή της ή όχι για δημοσίευση (Lin, 2025).

Στον αντίποδα, τα συστήματα ΠΤΝ προσφέρουν μια πρωτοφανή ευκαιρία να γεφυρωθεί το γλωσσικό χάσμα που έχει δημιουργηθεί από τα μέσα του 20ού αιώνα στη διεθνή ακαδημαϊκή δημοσίευση, με την αγγλική να είναι η κυρίαρχη γλώσσα. Η πρόκληση που αντιμετωπίζουν οι επιστημονικοί συγγραφείς που χρησιμοποιούν τα αγγλικά ως δεύτερη ή τρίτη γλώσσα, μέσα σε μια αγγλόφωνη επιστημονική εκδοτική κοινότητα που απαιτεί κείμενα συγγραφέα επιπέδου μητρικής γλώσσας, επηρεάζει σημαντικά την επαγγελματική τους εξέλιξη, μειώνει τις ατομικές δημοσιεύσεις του, και αποκλείει πιθανώς σημαντικές συνεισφορές, οι οποίες μένουν αδημοσίευτες. Ειδικές εφαρμογές ΠΤΝ μπορούν να διορθώνουν τη γραμματική ή το συντακτικό στην αγγλική γραφή ή μπορούν να μεταφράζουν με ακρίβεια κείμενο από άλλες γλώσσες στα αγγλικά. Έτσι, αίρονται τα εμπόδια για τους συγγραφείς που δεν μιλούν αγγλικά σαν μητρική γλώσσα για να δημοσιεύουν σε διεθνώς αναγνωρισμένα επιστημονικά περιοδικά (Katsnelson, 2022). Τέτοια συστήματα διαθέτουν επιπλέον ένα σημαντικό εκπαιδευτικό πλεονέκτημα βοηθώντας τους συγγραφείς επιστημονικών άρθρων που έχουν τα αγγλικά δεύτερη γλώσσα να βελτιώσουν τις συγγραφικές τους δεξιότητες (Ingley and Pack, 2023). Λειτουργώντας πάλι ως «μεταφραστές» με την ευρεία έννοια, συστήματα ΠΤΝ μπορούν να βοηθήσουν επιστήμονες, χωρίς γνωστικό υπόβαθρο πληροφορικής, να κάνουν προγραμματισμό για την κατασκευή εμπειρικού λογισμικού για την εκπόνηση συγκεκριμένου επιστημονικού έργου. Στην περίπτωση αυτή οι επιστήμονες χρήστες περιγράφουν σε συστήματα ΠΤΝ την αποστολή του εμπειρικού λογισμικού και τις επιθυμητές λειτουργίες του και τα μοντέλα παράγουν κώδικα προγράμματος επιταχύνοντας το επιστημονικό έργο (Aygün et al., 2025).

Ένα επιπλέον πλεονέκτημα των ΠΤΝ μοντέλων είναι οι αναλυτικές τους δυνατότητες, ιδιαίτερα κατά την διαχείριση μεγάλων και σύνθετων συνόλων

δεδομένων. Τα συστήματα PTN μπορούν να δημιουργήσουν αυτοματοποιημένες περιλήψεις δεδομένων και, το σημαντικότερο, να ανιχνεύσουν τάσεις ή μοτίβα και να εξάγουν συμπεράσματα από τα δεδομένα (Curtis, 2023; Amano et al., 2023). Έτσι, είναι δυνατό να ενισχυθεί και να επιταχυνθεί η την επιστημονική παραγωγικότητα, συνοψίζοντας ογκώδεις επιστημονικές βιβλιογραφίες, μεταφράζοντας άρθρα στις μητρικές γλώσσες των ερευνητών, αναλύοντας διεπιστημονικά σύνολα δεδομένων σε επιστήμονες διαφορετικών πεδίων, δημιουργώντας ερευνητικά ερωτήματα προς εξέταση, ανακαλύπτοντας κενά (gaps) στη βιβλιογραφία και προτείνοντας άξονες επιστημονικής διερεύνησης.

Για τους εκδότες επιστημονικών περιοδικών, τα συστήματα PTN προσφέρουν μοναδικούς τρόπους ξεπεράσματος των γλωσσικών ή διεπιστημονικών φραγμών, λειτουργώντας ως ευέλικτο εργαλείο επεξήγησης της πληροφορίας για όλους τους χρήστες. Ακόμη, μπορούν να προσαρμόσουν βιβλιογραφικές παραπομπές στο στυλ του περιοδικού αναφοράς ή να αναθεωρήσει κείμενα (Kumar et al., 2024; Inala et al., 2024). Αδιαμφισβήτητα, μετά την επανάσταση της τυπογραφίας και τον ψηφιακό μετασχηματισμό, τα μοντέλα PTN μεταμορφώνουν τόσο τη δημιουργία όσο και τη διάδοση της επιστημονικής γνώσης (Bergstrom et al., 2024; Salazar & Kunc, 2025).

Η διάδοση της συμμετοχής συστημάτων PTN στη συγγραφή επιστημονικών άρθρων είναι εντυπωσιακή. Τεκμήριο αυτής της διάδοσης ακόμη και όταν δεν αναφέρεται από τους συγγραφείς των δημοσιεύσεων αποτελούν χαρακτηριστικές λέξεις τις οποίες κάποια συστήματα PTN χρησιμοποιούν σε μεγάλη συχνότητα. Μια τέτοια λέξη είναι η “delve,” που αποδίδεται στα ελληνικά ως σκάβω, χρησιμοποιείται όμως στα επιστημονικά άρθρα με την έννοια του εμβαθύνω. Έτσι, όπως φαίνεται από τα ευρετηριασμένα στο PubMed δημοσιευμένα άρθρα πριν και μετά το Νοέμβριο του 2022, όταν η πρόσβαση στο ChatGPT έγινε δημοφιλής, υπάρχει μια αύξηση τέσσερις φορές της συχνότητας χρήσης αυτής της λέξης. Άλλες πάλι λέξεις, όπως η λέξη “important,” που αποδίδεται στα ελληνικά για την σπουδαιότητα, εμφανίζουν μικρότερη συχνότητα χρήσης στις πλέον πρόσφατες επιστημονικές δημοσιεύσεις.

Βιβλιομετρικά το φαινόμενο αλλαγής συχνότητας χρήσης λέξεων ή επιστημονικών όρων δεν είναι ασυνήθιστο, καθώς σχετίζεται με γεγονότα ή νέες ανακαλύψεις. Για παράδειγμα, η επιδημία Ebola στην Αφρική το 2014 έφερε μια απότομη αύξηση αυτής της λέξης στις περιλήψεις επιστημονικών άρθρων, όπως και η λέξη “pandemic” μετά την πανδημία COVID-19, ή την λέξη “convolutional,” συνελκτικός, που εμφανίζεται μετά το 2016 με την ανάπτυξη των συνελκτικών νευρωνικών δικτύων, ειδικά σχεδιασμένων για την αναγνώριση και ταξινόμηση εικόνων (Convolutional Neural Networks, CNN). Λαμβάνοντας υπόψη συνολικά τις βιβλιομετρικές τάσεις χρήσης τέτοιων λέξεων στις περιλήψεις επιστημονικών άρθρων στην βιοϊατρική έρευνα, οι Kobak και συνεργάτες υπολόγισαν πως το 13.5% των περιλήψεων άρθρων που δημοσιεύτηκαν το 2024 είχαν υποστεί επεξεργασία από μοντέλα PTN (Kobak et al., 2025).

Ο μετασχηματισμός από τη χρήση συστημάτων PTN στην επιστήμη, όμως, ενέχει σημαντικούς κινδύνους που σχετίζονται με την ακεραιότητα των δεδομένων, τη λογική συνέπεια και τη δημιουργικότητα στην επιστήμη. Στην επιστημονική βιβλιογραφία εντοπίζονται ακραία προβληματικά παραδείγματα που έχουν οδηγήσει σε ανακλήσεις άρθρων, με δεδομένα, κείμενο, βιβλιογραφία ή εικόνες που έχουν παραχθεί από συστήματα PTN όμως είναι αποτέλεσμα παραισθήσεων (hallucinations) των συστημάτων. Όμως, υπάρχουν και περιπτώσεις παραποιήσεων είτε από σφάλμα PTN είτε με την πρόθεση των συγγραφέων που, περνώντας απαρατήρητες, διαμορφώνουν ψευδώς τεκμηριωμένη γνώση. Σε τέτοιες καταστάσεις, οι πληροφορίες αυτές που θεωρούνται αποδεδειγμένες, μπορεί να οδηγήσουν σε αστήρικτες επιστημονικές υποθέσεις και έρευνες χωρίς αντίκρισμα (Ferrara, 2024; Bjelobaba et al., 2025; Lidsky & Daves 2025).

Επιστημονικά πεδία που βασίζονται στη δευτερογενή επιστημονική επεξεργασία δεδομένων από προηγούμενες έρευνες, όπως η επιδημιολογία ή η βιοπληροφορική, αποδεικνύονται ιδιαίτερα ευπαθή εξαιτίας της χρήσης PTN ή της επαναχρησιμοποίησης PTN δημιουργημένων δεδομένων. Σε μια πρόσφατη έρευνα διαπιστώθηκε πως PTN μοντέλα χρησιμοποιήθηκαν σε μονοπαραγοντικές αναλυτικές προσεγγίσεις δευτερογενών αναλύσεων του

συνόλου δεδομένων NHANES (National Health and Nutrition Examination Survey) του Εθνικού Κέντρου Στατιστικής Υγείας (National Center for Health Statistics, NCHS) των ΗΠΑ, που δημοσιεύτηκαν από διαφορετικές ερευνητικές ομάδες σε επιστημονικά περιοδικά. Ένα πλήθος από 341 δημοσιευμένες βιοπληροφορικές εργασίες ισχυρίζονταν πως είχαν εντοπίσει κλινικά σημαντική συσχέτιση, όμως στην πραγματικότητα η στατιστική ισχύς αυτών των ερευνών ήταν χαμηλή. Αυτές οι μελέτες χρησιμοποιούν έναν τυποποιημένο (formulaic) τρόπο εκπόνησης έρευνας και τροφοδοτούν την επιστημονική κοινότητα με ανεπαρκώς τεκμηριωμένες και πλημμελώς ελεγμένες υποθέσεις (Suchak et al., 2025).

Με κίνητρο αυτές τις παρατηρήσεις στην εργασία αυτή εξετάζεται η επιστημονική βιβλιογραφία αναδρομικών βιοπληροφορικών ερευνών που εκπονήθηκαν, πριν και μετά το 2020, όταν προηγμένα μοντέλα ΠΤΝ έγιναν διαθέσιμα προς χρήση. Εξετάζονται ο βαθμός τυποποίησης των μεθοδολογιών και λεκτικά κοινά στοιχεία μεταξύ των δύο αυτών χρονικών περιόδων.

Υλικά και μέθοδοι

Η μελέτη αυτή έχει σαν αντικείμενο την ανάλυση της δημοσιευμένης σε περιοδικά επιστημονικής βιβλιογραφίας του επιστημονικού πεδίου της βιοπληροφορικής. Η έρευνα εστιάζεται σε πρωτότυπες αναδρομικές (retrospective) μελέτες, αποκλείοντας βιβλιογραφικές ανασκοπήσεις ή συντακτικά άρθρα, καθώς σε αυτό το ερευνητικό αντικείμενο δευτερογενούς ανάλυσης δεδομένων χρησιμοποιούνται συστήματα ΠΤΝ ήδη από το 2020 (Suchak et al., 2025). Οι αναζητήσεις βιβλιογραφίας πραγματοποιήθηκαν στις βάσεις δεδομένων PubMed (<https://pubmed.ncbi.nlm.nih.gov/>), iCite (<https://icite.od.nih.gov/>), Scopus (<https://www.scopus.com>) και Web of Science (<https://www.webofscience.com/wos/woscc/smart-search>).

Το ερώτημα αναζήτησης (query) που χρησιμοποιήθηκε στα PubMed και iCite, των οποίων οι αλγόριθμοι αναζήτησης μοιράζονται την ίδια σύνταξη ερωτημάτων

καθώς αποτελούν βάσεις δεδομένων του Εθνικού Κέντρου Βιοτεχνολογικής Πληροφορίας των ΗΠΑ (National Center for Biotechnology, NCBI), ήταν:

```
((("bioinformatical"[All Fields] OR "bioinformatically"[All Fields]  
OR "computational biology"[MeSH Terms] OR ("computational"[All  
Fields] AND "biology"[All Fields]) OR "computational biology"[All  
Fields] OR "bioinformatic"[All Fields] OR "bioinformatics"[All  
Fields]) AND ("retrospective studies"[MeSH Terms] OR  
("retrospective"[All Fields] AND "studies"[All Fields]) OR  
"retrospective studies"[All Fields] OR "retrospective"[All Fields] OR  
"retrospectively"[All Fields] OR "retrospectives"[All Fields])) NOT  
"review"[Publication Type]) NOT "editorial"[Publication Type]
```

Το ερώτημα αναζήτησης (query) που χρησιμοποιήθηκε στο Scopus ήταν:

```
TITLE-ABS-KEY ( bioinformatics ) AND TITLE-ABS-KEY ( retrospective )  
AND ( LIMIT-TO ( DOCTYPE , "ar" ) )
```

Το ερώτημα αναζήτησης (query) που χρησιμοποιήθηκε στο Web of Science ήταν:

```
bioinformatics (All Fields) AND retrospective (All Fields) and Review  
Article (Exclude - Document Types) and Article (Document Types) and  
Article (Document Types)
```

Πραγματοποιήθηκε εξαγωγή των αποτελεσμάτων σε πίνακες είτε ως αρχεία τιμών διαχωρισμένων με κόμμα (.csv) είτε ως αρχεία κειμένου (.txt).

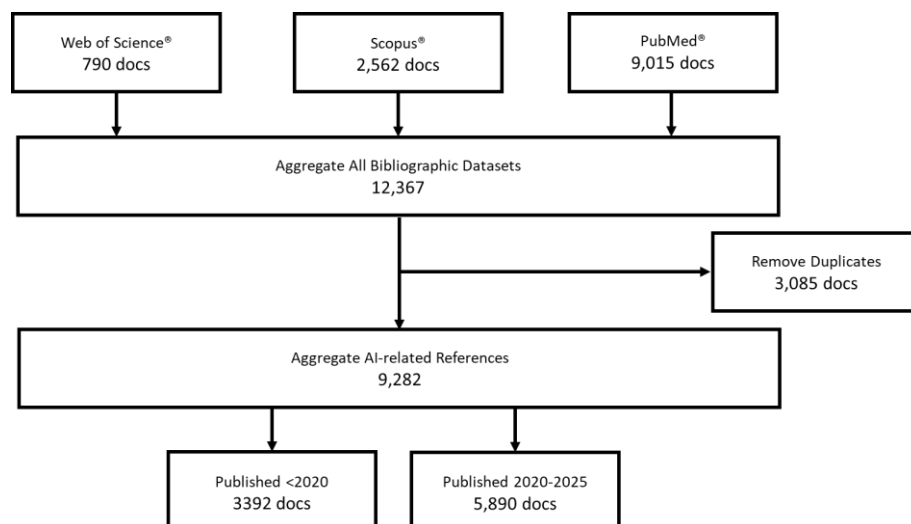
Περιγραφική στατιστική ανάλυση των δεδομένων που ελήφθησαν πραγματοποιήθηκε σε φύλλα υπολογισμού (.xlsx). Η απομάκρυνση διπλότυπων πραγματοποιήθηκε μέσω DOI των άρθρων, ενώ ακολούθησε περαιτέρω έλεγχος σχετικότητας με το θέμα. Οι συλλεγείσες βιβλιογραφικές αναφορές συμπεριλήφθηκαν στην ανάλυση χωρίς εξαιρέσεις ή αποκλεισμούς. Τελευταία ενημέρωση των βιβλιογραφικών δεδομένων πραγματοποιήθηκε 8^η Ιουλίου 2025. Η ανάλυση στο VOSviewer πραγματοποιήθηκε σύμφωνα με τις οδηγίες των δημιουργών και τον τύπο της εισόδου. Οι τίτλοι και οι περιλήψεις τους για τον εντοπισμό της συχνότητας εμφάνισης επιστημονικών όρων έγινε με τον αλγόριθμο Λανθάνουσας Κατανομής Dirichlet (Latent Dirichlet Allocation, LDA) στο πρόγραμμα VOSviewer. Ως κριτήριο ομαδοποίησης των δημοσιεύσεων

χρησιμοποιήθηκε ο χρόνος δημοσίευσής τους (publication year). Τα δεδομένα παρουσιάζονται ως μέσος όρος $\pm$ τυπική απόκλιση. Τιμή $p < 0,05$ θεωρήθηκε στατιστικώς σημαντική.

Αποτελέσματα

Από τις αναζητήσεις που πραγματοποιήθηκαν βρέθηκαν αναδρομικές μελέτες βιοπληροφορικής από τις βιβλιογραφικές βάσεις δεδομένων Web of Science, Scopus, και PubMed. Στο διάγραμμα ροής της Εικόνας 1 παρουσιάζονται τα αποτελέσματα της αναζήτησης και η πορεία της έρευνας.

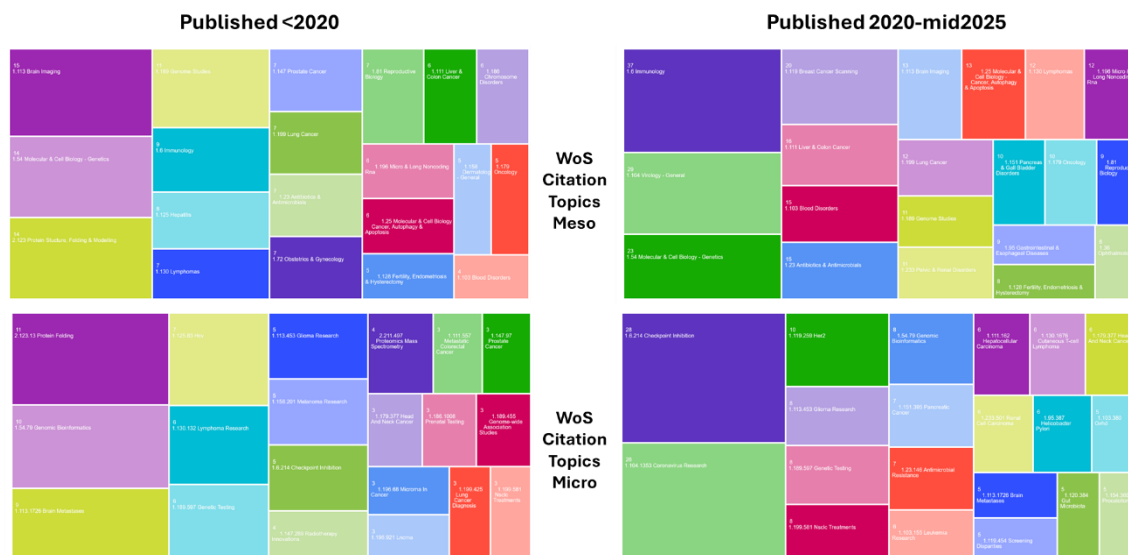
Εικόνα 1. Διάγραμμα ροής των βιβλιογραφικών αναζητήσεων που πραγματοποιήθηκαν



Από τα αποτελέσματα που ανακτήθηκαν από το Web of Science παρατηρούνται σημαντικές διαφοροποιήσεις που αφορούν την κατηγοριοποίηση των άρθρων (Web of Science Categories, Citation Topics Meso, and Citation Topics Micro), όπως φαίνεται στην Εικόνα 2. Ενώ η κατηγορία της Ογκολογίας βρίσκεται στην πρώτη θέση των αποτελεσμάτων και στις δύο χρονικές περιόδους που εξετάστηκαν, η Γενετική και η Ανοσολογία εκτοπίζουν τις Βιοχημεία και Βιοτεχνολογία, από τη δεύτερη και τρίτη θέση, αντίστοιχα, μετά το 2020. Τα άρθρα που αφορούν τις κατηγορίες αυτές μειώνονται ποσοστιαία στο μισό από ότι πριν το 2020. Η τάση αυτή είναι ιδιαίτερα φανερή στα ειδικότερα θέματα. Σημαντική επίδραση στο φαινόμενο αυτό έχει η πανδημία COVID-19, καθώς

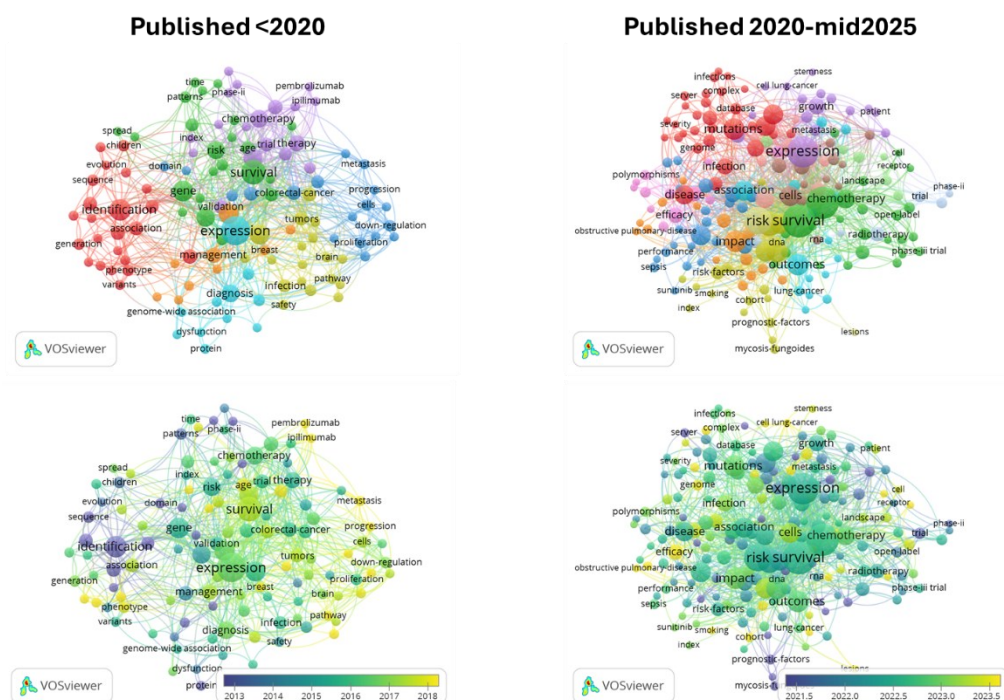
παρατηρείται μαζί με την αύξηση της Ανοσολογίας, η άνοδος της Ιολογίας και της έρευνας για του κορονοϊούς.

Εικόνα 2. Αλλαγές στη θεματολογία των άρθρων πριν και μετά το 2020 κατά το Web of Science



Ανάλυση των λέξεων κλειδιών (Keywords Plus®) των ανακτημένων τίτλων και περιλήψεων από το Web of Science στο VOSviewer δείχνει την αύξηση αναδρομικών βιοπληροφορικών εργασιών που αφορούν γενετικούς πολυμορφισμούς και τη συσχέτισή τους με ασθένειες, απόκριση στην θεραπεία, ανεπιθύμητες δράσεις φαρμάκων, και πρόγνωση πορείας νοσημάτων (Εικόνα 3). Ενώ η γονιδιακή έκφραση είναι η συχνότερη λέξη κλειδί πριν το 2020, χάνει την θέση της από την επιβίωση και τον καρκίνο μετά το 2020. Παρατηρείται ακόμη αύξηση των διαφορετικών ομάδων λέξεων κλειδιών από 7 πριν το 2020 σε 13 μετά, γεγονός που καταδεικνύει αυξημένη περιπλοκότητα στην αναζήτηση συσχετισμών μεταξύ βιολογικών και κλινικών οντοτήτων, προδιαθέσεων, νοσηρότητας και αποτελεσμάτων.

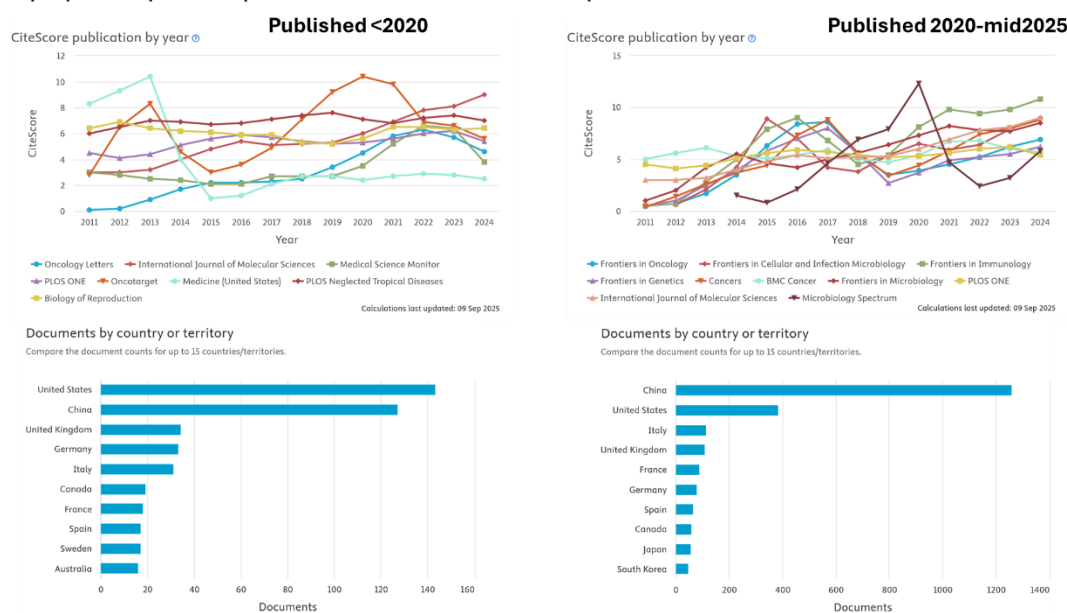
Εικόνα 3. Αλλαγές στον συν-εντοπισμό λέξεων κλειδιών (Keywords Plus®) πριν και μετά το 2020 κατά το Web of Science



Η αναζήτηση στο Scopus φανερώνει επίσης διαφοροποιήσεις στη θεματολογία των άρθρων, ωστόσο οι τάσεις διαφέρουν εξαιτίας των διαφορετικών κατηγοριών που χρησιμοποιούνται στο Scopus (Subject Area). Έτσι, ενώ η Ιατρική είναι η κυρίαρχη θεματική περιοχή και στις δύο χρονικές περιόδους, ακολουθείται από την κατηγορία Βιοχημεία, Γενετική, και Μοριακή Βιολογία, η οποία μάλιστα παρουσιάζει σημαντική ποσοστιαία αύξηση μετά το 2020. Λαμβάνοντας υπόψιν τις παρατηρήσεις από τις κατηγορίες του Web of Science, γίνεται φανερό πως η αύξηση αυτή οφείλεται στη Γενετική. Ακόμη αυξάνονται σημαντικά οι θεματικές κατηγορίες Ανοσολογία και Μικροβιολογία, Φαρμακολογία, Τοξικολογία και Φαρμακευτική, Νευροπεπιστήμη και Πληροφορική, καθώς και η Διεπιστημονική κατηγορία. Ενδιαφέρον έχει η αλλαγή που παρατηρείται στις εκδοτικές τάσεις όσον αφορά τα επιστημονικά περιοδικά που δημοσιεύουν τις περισσότερες σχετικές εργασίες. Ενώ πριν το 2020 τα περιοδικά με τα περισσότερα άρθρα στο θέμα αναζήτησης, αναδρομικές μελέτες βιοπληροφορικής, δημοσιεύονται στο Plos One, ακολουθούμενο από τα Oncology Letters, International Journal of Molecular Sciences (IJMS), Medical Science Monitor, και Oncotarget, μετά το 2020 τα σχετικά άρθρα δημοσιεύονται περισσότερο συχνά σε περιοδικά των εκδόσεων Frontiers και συγκεκριμένα στα Frontiers in Oncology, Frontiers in Cellular and Infection Microbiology, Frontiers

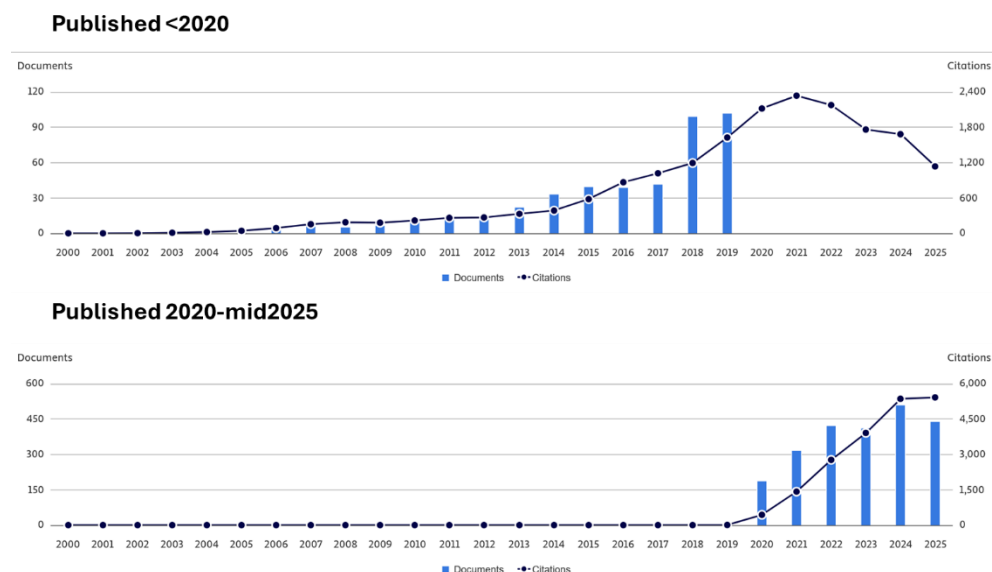
in Immunology και Frontiers in Genetics, τα οποία ακολουθούνται από τα Cancers, BMC Cancer και Frontiers in Microbiology (Εικόνα 4, επάνω). Παράλληλα, σημαντική είναι η διαφοροποίηση στις χώρες προέλευσης των επιστημονικών άρθρων. Πριν το 2020 φαίνεται πως οι εργασίες αναδρομικών μελετών βιοπληροφορικής προέρχονται σχεδόν ισοτίμα από ΗΠΑ και Κίνα, καλούθουμενες από Μεγάλη Βρετανία, Γερμανία και Ιταλία. Μετά το 2020, όμως, η αναλογία δημοσιεύσεων μεταξύ Κίνας και ΗΠΑ διαμορφώνεται σε 3 προς 1, με τρίτη χώρα προέλευσης την Ιταλία και έπειτα την Μεγάλη Βρετανία, Γαλλία και Γερμανίας (Εικόνα 4, κάτω). Πριν το 2020 τα άρθρα που προέρχονταν από την Κίνα αποτελούσαν το 30% του συνόλου, ενώ μετά το 2020, το 55% του συνόλου των δημοσιεύσεων.

Εικόνα 4. Αλλαγές στα επιστημονικά περιοδικά και στις χώρες προέλευσης των άρθρων πριν και μετά το 2020 κατά το Scopus



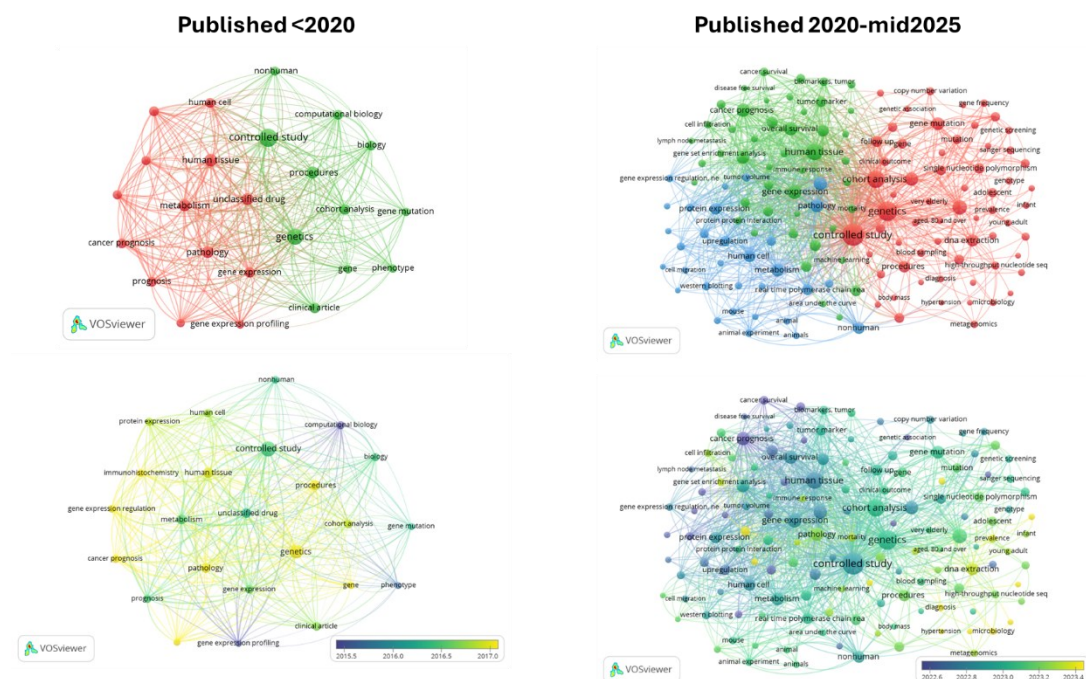
Τα δεδομένα που αντλούνται από το Scopus δείχνουν πως οι αλλαγές στο χώρο των επιστημονικών εκδόσεων και χωρών προέλευσης δεν επηρεάζουν τις αυξητικές τάσεις για βιβλιογραφικές αναφορές (Εικόνα 5). Η καμπύλη αύξησης των βιβλιογραφικών αναφορών με το χρόνο έχει σχεδόν την ίδια κλήση πριν και μετά το 2020. Συνεπώς, τα άρθρα μεταξύ των δύο χρονικών περιόδων διατηρούν την ίδια απήχηση στην παγκόσμια επιστημονική κοινότητα.

Εικόνα 5. Δεν παρατηρούνται αλλαγές τάσεων στις βιβλιογραφικές αναφορές πριν και μετά το 2020 κατά το Scopus



Ενώ η κατηγοριοποίηση στο Scopus δεν φανερώνει ανακατατάξεις στη σειρά των θεματικών κατηγοριών, ο συν-εντοπισμός ευρετηριασμένων λέξεων κλειδιών γίνεται εξαιρετικά σύνθετος μετά το 2020 (Εικόνα 6). Η επιλογή συμπερίληψης στην ανάλυση μόνο των ευρετηριασμένων και όχι και των λέξεων κλειδιών που αποδίδονται από τους συγγραφείς έγινε για την τήρηση ενός ομοιογενούς σχήματος ανάλυσης. Πριν το 2020 ομαδοποιούνται με το VOSviewer 8 ομάδες λέξεων κλειδιών, με 1086 να σχηματίζουν δίκτυα τριών συν-εντοπισμών. Μετά το 2020, όμως, σχηματίζεται ένα περίπλοκο δίκτυο αλληλεπιδράσεων μεταξύ των βιολογικών και κλινικών οντοτήτων από 1588 όρους σε συνδυασμούς δέκα συν-εντοπισμών. Αυτή η περιπλοκότητα συνάδει με όσα έχουν αναφερθεί περί των συσχετίσεων που αναζητούνται στην βιοπληροφορική με τη χρήση PTN μοντέλων από τους Suchak et al. (Suchak et al., 2025). Η ασθενής στατιστική ισχύς αυτών των μελετών, με δεδομένο πως διατηρούν σημαντική βιβλιογραφική απήχηση δείχνει τον κίνδυνο εξάπλωσης στην βιβλιογραφία αδύναμων υποθέσεων, που μάλλον προκαλούν σύγχυση και κύκλους αυτο-αναφοράς.

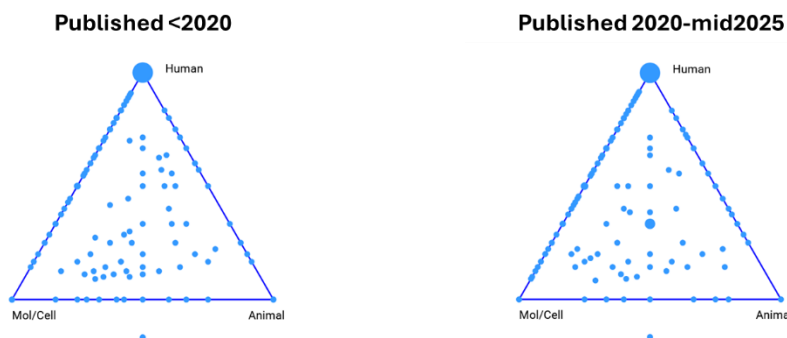
Εικόνα 6. Ανάλυση συν-εντοπισμένων ευρετηριασμένων λέξεων κλειδιών δείχνει σημαντικές διαφοροποιήσεις πριν και μετά το 2020 κατά το Scopus



Μια ακόμη ενδιαφέρουσα παρατήρηση αφορά την μεταφραστική έρευνα, έρευνα που από τον πάγκο των εργαστηρίων και τα ζωικά πρότυπα φτάνει σε κλινικό επίπεδο στον άνθρωπο ασθενή. Το iCite του NCBI είναι ειδικά σχεδιασμένο για αυτή την ανάλυση καθώς πέρα από την καταγραφή των βιβλιογραφικών αναφορών ανιχνεύει λέξεις κλειδιά που δείχνουν το είδος μεταφραστικής έρευνας που πραγματοποιείται καθώς και πόσο συχνά αναφέρεται από άλλες κλινικές μελέτες. Με άξονα τις αλλαγές που προέκυψαν μέσω της εφαρμογής ΠΤΝ μοντέλων στην αναδρομική βιοπληροφορική έρευνα παρατηρείται αλλαγή στην επικέντρωση των άρθρων μεταξύ μοριακών και κυτταρικών βιολογικών μελετών, πειραματοζώων και ανθρώπου (Εικόνα 7). Σε αντίθεση με τη διάχυτη διαμόρφωση στο εσωτερικό του διαγράμματος σχήματος τριγώνου πριν το 2020, στο οποίο διακρίνεται πως οι περισσότερες μελέτες κινούνται στον άξονα Μοριακής Βιολογίας και Κλινικής έρευνας, μετά το 2020 διαμορφώνεται μια εστία δημοσιευμάτων που συμπίπτουν με το βαρύκεντρο του τριγώνου, σε ίση απόσταση μεταξύ Μοριακής, Πειραματοζώων και Κλινικής έρευνας. Οι μελέτες που τοποθετούνται σε αυτό το ιδιαίτερο σημείο εμφανίζουν ισόρροπα χαρακτηριστικά χρησιμοποιώντας δεδομένα και από τις τρεις κατευθύνσεις, όπως οι επιλογές των συστημάτων ΠΤΝ μεταξύ των κατανομών πιθανοτήτων που ακολουθούν. Ο αριθμός των εργασιών αυτών είναι περίπου το

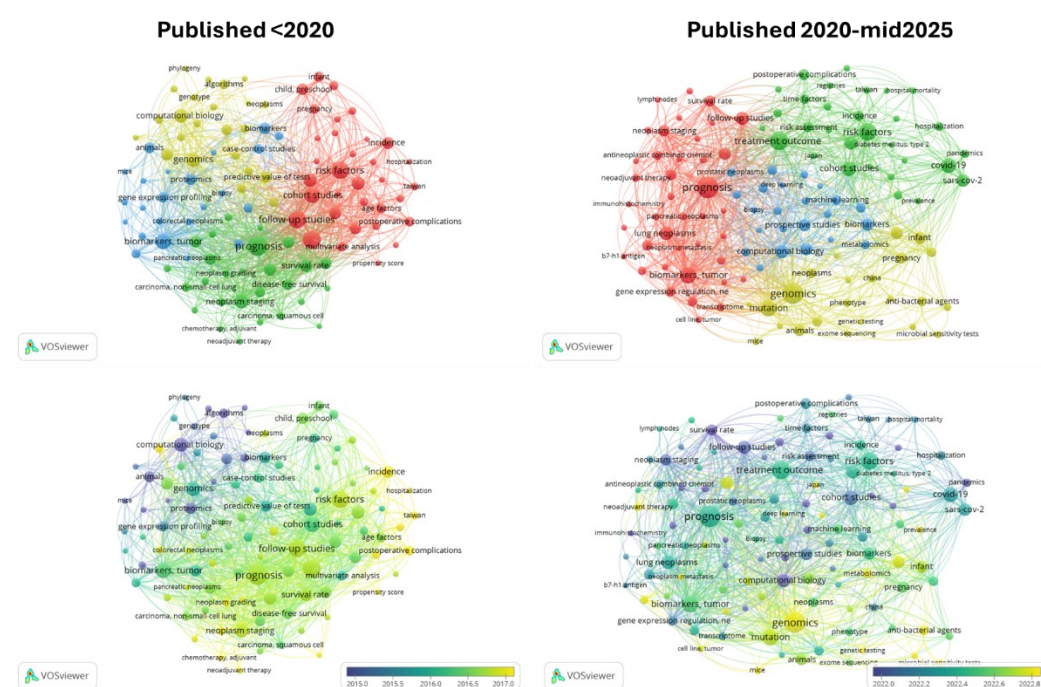
17.85% του συνόλου των αναδρομικών βιοπληροφορικών μελετών που εντοπίζονται από το PubMed μετά το 2020.

Εικόνα 7. Αλλαγές στις τάσεις εστίασης της μεταφραστικής έρευνας πριν και μετά το 2020 σύμφωνα με το iCite



Ανάλυση με το VOSviewer στο σύνολο των άρθρων δείχνει 16 ομάδες συν-εντοπισμένων λέξεων κλειδιών για την χρονική περίοδο πριν το 2020, με 2120 σε σύνολο 5817 (36.5%) να σχηματίζουν δίκτυα τριών συν-εντοπισμών. Μετά το 2020, 2193 λέξεις κλειδιά σε σύνολο 5834 (37.6%) σχηματίζουν επίσης δίκτυα τριών συν-εντοπισμών, οργανωμένες σε 21 ομάδες (Εικόνα 8). Η δυναμική αύξηση της περιπλοκότητας των αλληλεπιδράσεων βιολογικών και κλινικών οντοτήτων στη δευτερογενή αναδρομική βιοπληροφορική έρευνα φαίνεται πως οφείλεται στη χρήση PTN μοντέλων που τείνουν στην ισόρροπη χρήση πηγών για την υποστήριξη συσχετίσεων με στατιστικό τρόπο.

Εικόνα 8. Αλλαγές στις συν-εντοπισμένες λέξεις κλειδιά στο σύνολο των επιστημονικών άρθρων που εντοπίστηκαν πριν και μετά το 2020



Στους Πίνακες 1 και 2, παρουσιάζονται συνοπτικά οι σημαντικότερες διαφορές στη συχνότητα εμφάνισης όρων στην αναδρομική βιοπληροφορική βιβλιογραφία πριν και μετά το 2020. Συνολικά 690 όροι που εμφανίζονταν στη σχετική βιβλιογραφία πριν το 2020, με συχνότητα κατά μέσο όρο 4.58 και διάμεσο τιμή 4, δεν εμφανίζονται μετά. Παράλληλα, 761 νέοι όροι εμφανίζονται στη βιβλιογραφία μετά το 2020, χωρίς να είχαν αναφερθεί νωρίτερα. Η μέση συχνότητα των νέων όρων είναι 5.4 με διάμεσο τιμή 4. Στους νέους όρους συμπεριλαμβάνονται ασθένειες που προκάλεσαν πανδημίες ή επιδημίες όπως COVID-19 και Zika virus. Επιπλέον, παρατηρούνται σημαντικές αλλαγές στη συχνότητα εμφάνισης στη βιβλιογραφία κοινά εμφανιζόμενων όρων μεταξύ των δύο διαφορετικών χρονικών περιόδων. Βρέθηκαν 266 όροι με σχεδόν διπλάσια αύξηση συχνότητας αναφοράς και 328 με πολλαπλάσια ελάττωση αναφορών στη βιβλιογραφία. Κάποιοι από τις αυξήσεις συχνότητας όρων αποδίδονται στα μοντέλα ΠΤΝ και στην τάση επαναχρησιμοποίησης συγκεκριμένων κλινικών ερευνών μεγάλων δεδομένων. Η απάλειψη μετά το 2020 σημαντικών βιολογικών εργαστηριακών μεθόδων από την κλασσική αλληλούχιση βιομορίων, την εξέταση της διαμόρφωσής τους, την ηλεκτροφόρηση δύο διαστάσεων ή ανοσοαποτύπωσης, αλλά και των αναφορών σε τιμές αναφοράς ή γενετικούς δείκτες οφείλει να προβληματίσει για τα ποιοτικά στοιχεία των νεώτερων ερευνών που ακολουθούν

αποκλειστικά στατιστική επαναχρησιμοποίηση προηγούμενων εργαστηριακών ή κλινικών ερευνών.

Πίνακας 1. Λέξεις κλειδιά που δεν εμφανίζονται μετά το 2020 και νέες λέξεις που δεν υπήρχαν πριν

<2020	Εμφανίσεις	Συχνότητα	2020-mid2025	Εμφανίσεις	Συχνότητα
molecular sequence data	31	1.36	covid-19	318	12.98
protein conformation	31	1.36	sars-cov-2	236	9.63
electrophoresis, gel, two-dimensional	22	0.96	immune checkpoint inhibitors	81	3.31
reference values	21	0.92	multiomics	44	1.80
genetic markers	20	0.88	east asian people	29	1.18
blotting, western	19	0.83	ischemic stroke	28	1.14
statistics, nonparametric	19	0.83	pneumonia, viral	28	1.14
chi-square distribution	16	0.70	covid-19 drug treatment	24	0.98
polymorphism, genetic	16	0.70	radiomics	20	0.82
confidence intervals	15	0.66	ranibizumab	20	0.82
down-regulation	15	0.66	covid-19 vaccines	19	0.78
pattern recognition, automated	15	0.66	rna-seq	18	0.73
radiotherapy, conformal	15	0.66	macular edema	17	0.69
sequence alignment	15	0.66	betacoronavirus	14	0.57
crystallography, x-ray	14	0.61	telemedicine	14	0.57

Η ομαλή συνέχεια των βιβλιογραφικών αναφορών που δέχονται οι μελέτες μετά το 2020 δείχνει μια συνέχεια ως προς την εμπιστοσύνη της επιστημονικής κοινότητας σε αναδρομικές βιοπληροφορικές μελέτες. Ωστόσο, η αύξηση της περιπλοκότητας των συσχετίσεων μεταξύ οντοτήτων δεν συνάδει με πραγματικές ανακαλύψεις με δεδομένο τον αναδρομικό χαρακτήρα των μελετών. Με την εξαίρεση της πανδημίας του COVID-19, της επιδημίας Zika και τις εμφάνισης κάποιων νέων φαρμακευτικών ουσιών ή τον επαναπροσδιορισμό της χρήσης τους, η σύνθετη εικόνα των δικτύων λέξεων κλειδιών μετά το 2020 οφείλεται σε μεγάλο βαθμό σε συσχετίσεις από μοντέλα ΠΤΝ, οι οποίες όταν δημοσιεύονται ανατροφοδοτούνται από άλλες ανάλογες μελέτες με βιβλιογραφικές αναφορές.

Πίνακας 2. Λέξεις κλειδιά με τις μεγαλύτερες αυξήσεις και μειώσεις πριν και μετά το 2020

Αυξημένες	Αυξήσεις Εμφανίσεων	Ποσοστό Αύξησης	Ελαττωμένες	Μειώσεις Εμφανίσεων	Ποσοστό Μείωσης
pandemics	103	96.26%	analysis of variance	-28	-933.33%
lipidomics	29	90.63%	oligonucleotide array sequence analysis	-51	-850.00%
deep learning	51	89.47%	regression analysis	-40	-800.00%
intravitreal injections	25	89.29%	amino acid sequence	-23	-766.67%
diabetic retinopathy	23	88.46%	drug design	-21	-700.00%
coronavirus infections	19	86.36%	models, molecular	-32	-640.00%
tumor microenvironment	86	83.50%	binding sites	-19	-633.33%
carbapenems	14	82.35%	databases, protein	-18	-600.00%
klebsiella pneumoniae	27	81.82%	reverse transcriptase polymerase chain reaction	-23	-575.00%

Αυξημένες	Αυξήσεις Εμφανίσεων	Ποσοστό Αύξησης	Ελαττωμένες	Μειώσεις Εμφανίσεων	Ποσοστό Μείωσης
exome sequencing	45	78.95%	protein binding	-23	-460.00%
immunotherapy	52	78.79%	ligands	-24	-400.00%
klebsiella infections	22	78.57%	cause of death	-27	-385.71%
neoplasms, unknown primary	11	78.57%	multivariate analysis	-112	-373.33%
overweight	11	78.57%	dose-response relationship, drug	-26	-371.43%
gastric bypass	14	77.78%	enzyme-linked immunosorbent assay	-22	-366.67%

Συζήτηση

Από δευτερεύουσες συνεισφορές όπως αξιολόγηση πληροφοριών, φιλτράρισμα και αποδοχή λογικών σφαλμάτων, μέχρι σκόπιμη κατασκευή δεδομένων και παρατυπίες από ανθρώπινους χρήστες, η εμπλοκή της ΠΤΝ στην πορεία της έρευνας, στο σχεδιασμό, στην ανάλυση και στη συγγραφική προετοιμασία άρθρων, οφείλει να γίνεται με προσοχή, επιμέλεια και διασταύρωση. Αυτή τη στιγμή είναι φανερό πως τα μοντέλα ΠΤΝ χρησιμοποιούνται σε σημαντικό βαθμό από επιστήμονες παγκόσμια.

Τα δεδομένα που παρουσιάσαμε δείχνουν μια πολύ ισχυρή διεϊσδυση που αγγίζει τη μια στις πέντε εργασίες που πραγματοποιούνται στο πλαίσιο των αναδρομικών βιοπληροφορικών ερευνών. Τα συστήματα ΠΤΝ φαίνεται πως παρουσιάζουν αυξημένη τάση να συνδυάζουν δεδομένα και να δημιουργούν υποθέσεις που μπορούν να υποστηριχθούν με ασθενείς στατιστικά δευτερογενείς αναλύσεις εργαστηριακών ή κλινικών δεδομένων. Με τον τρόπο αυτό αλλάζουν τον χάρτη της βιοϊατρικής έρευνας. Ο κίνδυνος οι αλλαγές αυτές να γίνουν παραμορφωτικές, σε βαθμό που να μην μπορεί να εντοπιστούν

πραγματικές βιολογικές συσχετίσεις μέσα από την επιστημονική βιβλιογραφία από μια περίοδο και έπειτα, είναι σημαντικός και ουσιαστικός.

Οι συντάκτες επιστημονικών περιοδικών, οι εκδότες και η ευρύτερη επιστημονική κοινότητα, συμμερίζονται τις ανησυχίες σχετικά με την αυξανόμενη χρήση μοντέλων ΠΤΝ στην επιστημονική έρευνα. Η κύρια κριτική είναι ότι τα μοντέλα ΠΤΝ, που συχνά αποκαλούνται "γλωσσικές μηχανές", μπορούν να μιμούνται το ανθρώπινο γράψιμο χωρίς πραγματική κατανόηση του νοήματος των φράσεων τους. Αυτό εγείρει σημαντικά ερωτήματα για τη φύση της νοημοσύνης αυτών των μοντέλων. Τα μοντέλα ΠΤΝ έχουν εκπαιδευτεί σε δισεκατομμύρια λέξεις αντλημένες από διάφορα πλαίσια, συμπεριλαμβανομένων βιβλίων, άρθρων και ιστοσελίδων. Μπορούν να παράγουν κατάλληλες απαντήσεις σε ανθρώπινα ερωτήματα, αλλά όχι με τον τρόπο της παραδοσιακής προγραμματιστικής λογικής (Hutson, 2021). Η δεύτερη γραμμή κριτικής αφορά την εκπαίδευση και τη διαμόρφωση προκαταλήψεων στα μοντέλα ΠΤΝ, ως αποτέλεσμα των πληροφοριών που λαμβάνουν. Με τον τρόπο αυτό μπορεί να ενισχυθούν στερεότυπα ή να επιχειρήσουν να γεμίσουν κενά πληροφορίας βασιζόμενα στο παρεχόμενο περιεχόμενο, παράγοντας έτσι προκατειλημμένες απαντήσεις (Ananya, 2024). Η τρίτη γραμμή κριτικής αφορά τους κύκλους ανατροφοδότησης και αυτό-επιβεβαίωσης, την παραμόρφωση των απαντήσεων συστημάτων ΠΤΝ όταν υλικό παραγόμενο από την ίδια ή άλλη ΠΤΝ χρησιμοποιείται στην εκπαίδευσή τους. Επειδή η εκπαίδευση των μοντέλων ΠΤΝ στηρίζεται σε περιεχόμενο του διαδικτύου και στην περίπτωση των επιστημονικών άρθρων, από αυτά που συγκεντρώνονται από ηλεκτρονικά περιοδικά ή βιβλιογραφικές βάσεις, όταν το περιβάλλον πλημμυρίσει με υλικό που έχει παραχθεί από τη ΠΤΝ, το αποτέλεσμα θα είναι ένας διαρκής αυτό-ενισχυόμενος κύκλος που μπορεί να οδηγήσει σε κατάρρευση των μοντέλων. Η κατάσταση αυτή αναφέρεται ως «δηλητηρίαση δεδομένων» (data poisoning) (Shumailov et al., 2022; Bohacek and Farid, 2025). Η προτίμηση σε όσο το δυνατό περισσότερο πρόσφατα άρθρα στη διαδικασία γνωστικών αναφορών, είναι μια προκατάληψη που μπορεί να εντείνει περισσότερο το πρόβλημα. Την ίδια στιγμή τα συστήματα ΠΤΝ δεν διαθέτουν την κρίση του ανθρώπου

επιστήμονα ως προς την επιλογή των πηγών που χρησιμοποιούν ανάλογα με την αξιοπιστία τους. Άρθρα με διαφορετικά ποιοτικά χαρακτηριστικά λαμβάνονται υπόψιν από τα συστήματα ΠΤΝ ισότιμα. Η τέταρτη γραμμή κριτικής αφορά αυτόν τον τρόπο σκέψης των μοντέλων ΠΤΝ και τα όρια της λογικής τους, ιδίως όταν καλούνται να αντιμετωπίσουν σύνθετα προβλήματα ή λογικούς γρίφους που δεν μπορούν να επιλύσουν (Biever, 2023; Hutson, 2024; Shojaee et al., 2025). Μια μη στοχαστική απάντηση σε ένα τέτοιο ερώτημα θα μπορούσε να προκαλέσει μια αλυσίδα λάθος υποθέσεων εφόσον ληφθεί υπόψιν ως τεκμηριωμένη άποψη. Οι δημιουργοί των ΠΤΝ υποστηρίζουν ότι τέτοιας προβλήματα προκύπτουν από εσφαλμένο έλεγχο μεταβλητών και απαιτητικά αιτήματα για εξαντλητικές λίστες επιλογών μέσω πολλαπλών πιθανοτήτων, που οδηγούν τα συστήματα σε αστοχίες και αποτυχίες (Orus and Lawsen, 2025). Το τελικό ζήτημα αφορά τις αντιλήψεις των χρηστών σχετικά με τη λειτουργική απόδοση και τις λογικές ικανότητες των ΠΤΝ συστημάτων που χρησιμοποιούν (Harper, 2025), καθώς και την επίγνωση των κινδύνων που μπορεί να προκύψουν από μια προβληματική ενημέρωση (Open AI, 2025) [24].

Η έρευνα αυτή έχει περιορισμούς, καθώς δεν εξετάστηκε η συγγραφική αναφορά περί της χρήσης ή μη συστημάτων ΠΤΝ κατά την πραγματοποίηση της έρευνας ή τη συγγραφή των άρθρων που εξετάστηκαν. Η προσέγγιση στο θέμα ήταν έμμεση, αξιοποιώντας τα βιβλιογραφικά και βιβλιομετρικά δεδομένα πριν και μετά τον χρόνο εμφάνισης τέτοιων συστημάτων. Ωστόσο, οι έμμεσοι παράμετροι που χρησιμοποιήθηκαν, ο όγκος της πληροφορίας, το είδος της πληροφορίας, η περιπλοκότητα της πληροφορίας, η συμπερίληψη άλλων γεγονότων μη σχετικών με τα μοντέλα ΠΤΝ που μπορεί να επηρεάσουν τη βιβλιογραφία και η απήχηση των δημοσιεύσεων μπορούν να διαμορφώσουν την εικόνα της έρευνας σήμερα και συγκριτικά με άλλες χρονικές περιόδους.

Συμπεράσματα

Στην εργασία αυτή εξετάστηκαν αναδρομικές έρευνες βιοπληροφορικής πριν και μετά την διάδοση συστημάτων ΠΤΝ. Εντοπίστηκαν πολλές διαφοροποιήσεις που

αφορούν στη θεματολογία, στην περιπλοκότητα, στην εκδοτική παραγωγή, στη μεταφραστική έρευνα και στις λέξεις κλειδιά των επιστημονικών άρθρων. Τα ευρήματα της έρευνας συνάδουν με την αυξημένη χρήση ΠΤΝ από βιοπληροφορικούς επιστήμονες, την αλλαγή του πλαισίου λειτουργίας της έρευνας στο πεδίο αυτό, από τον επιστημονικό πάγκο στην υπολογιστική βιολογία, στο σχηματισμό και έλεγχο υποθέσεων διαθέσιμων μεγάλων βιολογικών και κλινικών δεδομένων από μοντέλα ΠΤΝ. Παρά την υψηλή απόδοση και την ταχύτητα αυτών των συστημάτων είναι αναγκαία η επιβεβαίωση των πορισμάτων τους, για να αποφευχθεί η συγκέντρωση σφαλμάτων που μπορεί να υποβαθμίσουν την επιστημονική γνώση.

Βιβλιογραφία

- Amano, T., Ramírez-Castañeda, V., Berdejo-Espinola, V., Borokini, I., Chowdhury, S., Golivets, M., Gonzalez-Trujillo, J. D., Paudel, K. White, R. L., & Veríssimo, D. (2023). The manifold costs of being a non-native English speaker in science. *PLoS Biology*, 21(7), e3002184. <https://doi.org/10.1371/journal.pbio.3002184>.
- Ananya. (2024). AI image generators often give racist and sexist results: can they be fixed? *Nature*, 627, 722-725. <https://doi.org/10.1038/d41586-024-00674-9>.
- Aygün, E., Belyaeva, A., Comanici, G., Coram, M., Cui, H., Garrison, J.,... & Brenner, M. P. (2025). An AI system to help scientists write expert-level empirical software. *arXiv preprint arXiv:2509.06503*. <https://doi.org/10.48550/arXiv.2509.06503>
- Bergstrom, T., Rieger, O. Y., & Schonfeld, R. C. (2024). The Second Digital Transformation of Scholarly Publishing. *Strategic Context and Shared Infrastructure*. Ithaka S+R. <https://doi.org/10.18665/sr.320210>
- Biever, C. (2023). The easy intelligence tests that AI chatbots fail. *Nature*, 619, 686-689. <https://doi.org/10.1038/d41586-023-02361-7>.
- Bjelobaba, S., Waddington, L., Perkins, M., Foltýnek, T., Bhattacharyya, S., & Weber-Wulff, D. (2025). Maintaining research integrity in the age of

GenAI: an analysis of ethical challenges and recommendations to researchers. *International Journal for Educational Integrity*, 21(1), 1-23.
<https://doi.org/10.1007/s40979-025-00191-w>

- Bohacek, M., & Farid, H. (2025). Nepotistically trained generative image models collapse. *arXiv*, arXiv:2311.12202v2.
- Curtis, N. (2023). To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. *The Pediatric infectious disease journal*, 42(4), 275. <https://doi.org/10.1097/INF.0000000000003852>.
- Curtis, N. (2023). To ChatGPT or not to ChatGPT? The impact of artificial intelligence on academic publishing. *The Pediatric infectious disease journal*, 42(4), 275. <https://doi.org/10.1097/INF.0000000000003852>.
- Ferrara, E. (2024). GenAI against humanity: Nefarious applications of generative artificial intelligence and large language models. *Journal of Computational Social Science*, 7(1), 549-569.
<https://doi.org/10.1007/s42001-024-00250-1>
- Harper, T. A. (2025). What Happens When People Don't Understand How AI Works. *The Atlantic*,
<https://www.theatlantic.com/culture/archive/2025/06/artificial-intelligence-illiteracy/683021/>
- Hutson, M. (2021). Robo-writers: the rise and risks of language-generating AI. *Nature*, 591, 22-25. <https://doi.org/10.1038/d41586-021-00530-0>.
- Hutson, M. (2024). How does ChatGPT 'think'? Psychology and neuroscience crack open AI large language models. *Nature*, 629, 986-988.
<https://doi.org/10.1038/d41586-024-01314-y>
- Inala, J. P., Wang, C., Drucker, S., Ramos, G., Dibia, V., Riche, N., Brown, D., Marshall, D., & Gao, J. (2024). Data Analysis in the Era of Generative AI. *arXiv preprint arXiv:2409.18475*.
<https://doi.org/10.48550/arXiv.2409.18475>.
- Ingley, S. J., & Pack, A. (2023). Leveraging AI tools to develop the writer rather than the writing. *Trends in ecology & evolution*, 38(9), 785-787.
<https://doi.org/10.1016/j.tree.2023.05.007>

- Katsnelson, A. (2022). Poor English skills? New AIs help researchers to write better. *Nature*, 609(7925), 208-209. <https://doi.org/10.1038/d41586-022-02767-9>.
- Kobak, D., González-Márquez, R., Horvát, E. Á., & Lause, J. (2025). Delving into LLM-assisted writing in biomedical publications through excess vocabulary. *Science Advances*, 11(27), eadt3813. <https://doi.org/10.1126/sciadv.adt3813>
- Kumar, A., Devi, M. L., & Saltz, J. S. (2024, December). GenAI Tools to Improve Data Science Project Outcomes. In *2024 IEEE International Conference on Big Data (BigData)* (pp. 3143-3152). IEEE. <https://doi.org/10.1109/BigData62323.2024.10825326>.
- Lidsky, L. B., & Daves, A. (2025). Inevitable Errors: Defamation by Hallucination in AI Reasoning Models (forthcoming in the *Journal of Free Speech Law* 2025). Available at SSRN 5362314. SSRN: <https://ssrn.com/abstract=5362314> or <http://dx.doi.org/10.2139/ssrn.5362314>
- Lin, Z. (2025). Hidden Prompts in Manuscripts Exploit AI-Assisted Peer Review. *arXiv preprint arXiv:2507.06185*. <https://arxiv.org/abs/2507.06185>
- OpenAI. (2025). Sycophancy in GTP-4o: what happened and what we're doing about it. <https://openai.com/index/sycophancy-in-gpt-4o/>
- Opus, C., & Lawsen, A. (2025). The illusion of the illusion of thinking: A comment on Shojaee et al. (2025). *arXiv*, arXiv:2506.09250v1.
- Salazar, A., & Kunc, M. (2025). The contribution of GenAI to business analytics. *Journal of Business Analytics*, 1-14. <https://doi.org/10.1080/2573234X.2024.2435835>.
- Shojaee, P., Mirzadeh, I., & Alizadeh, K. (2025). The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity. *arXiv*, arXiv:2506.06941v1, & Machine Learning Research at Apple, <https://machinelearning.apple.com/research/illusion-of-thinking>

- Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755-759. <https://doi.org/10.1038/s41586-024-07566-y>
- Stokel-Walker, C. 2023. "ChatGPT listed as author on research papers: many scientists disapprove." *Nature*, 613(7945), 620-621. <https://doi.org/10.1038/d41586-023-00107-z>.
- Suchak, T., Aliu, A. E., Harrison, C., Zwiggelaar, R., Geifman, N., & Spick, M. (2025). Explosion of formulaic research articles, including inappropriate study designs and false discoveries, based on the NHANES US national health database. *PLoS biology*, 23(5), e3003152. <https://doi.org/10.1371/journal.pbio.3003152>.
- Thorp, H. H. 2023. "ChatGPT is fun, but not an author." *Science*, 379(6630), 313-313. <https://doi.org/10.1126/science.adg7879E>.

Στοιχεία ανακοίνωσης στην ελληνική γλώσσα

Τίτλος εργασίας

Παραγωγική Τεχνητή Νοημοσύνη: Μολύνοντας την Επιστημονική Σκέψη

Στοιχεία συγγραφέα / συγγραφέων

Άρτεμις Χαλεπλιόγλου, Τμήμα Αρχειονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης, Πανεπιστήμιο Δυτικής Αττικής, artemischal@uniwa.gr

Χρήστος Ζαμπακόλας, Τμήμα Αρχειονομίας, Βιβλιοθηκονομίας και Συστημάτων Πληροφόρησης, Πανεπιστήμιο Δυτικής Αττικής, christoszampakolas@uniwa.gr

Λέξεις-κλειδιά

1. Βιβλιομετρία
2. Βιοπληροφορική
3. Μελέτη κοόρτης σε χρονοσειρά

4. Αναδρομικές μελέτες
5. Κατάρρευση Μοντέλων Παραγωγικής Τεχνητής Νοημοσύνης

Περίληψη

Η Παραγωγική Τεχνητή Νοημοσύνη (ΠΤΝ), βασισμένη σε μεγάλα γλωσσικά μοντέλα, νευρωνικά δίκτυα και μετασχηματιστές, παρουσιάζει μια εκπληκτική παραγωγικότητα κειμένων, εικόνων, ήχων και βίντεο, αναλύοντας μεγάλο όγκο δεδομένων. Προσομοιώνοντας και μιμούμενα πρότυπα κείμενα από μεγάλο αριθμό επιστημονικών δημοσιεύσεων που χρησιμοποιήθηκαν στην εκπαίδευσή τους, διάφορα μοντέλα ΠΤΝ έχουν δημιουργήσει επιστημονικό λόγο τόσο αληθοφανή, ώστε να μοιάζει πως προέρχεται από ειδικούς επιστήμονες. Όταν τα μοντέλα ΠΤΝ χρησιμοποιούνται για τη σύνταξη θεμελιωδών επιστημονικών υποθέσεων, είτε με τον εντοπισμό κενών, ασυνεπειών ή αντιθέσεων στη βιβλιογραφία, είτε με την ανάπτυξη μεθοδολογίας έρευνας, διαμορφώνεται ένα αυτοτροφοδοτούμενο περιβάλλον υποθέσεων, ενώ αυξάνεται η τυποποίηση εκπόνησης της έρευνας. Ενώ το φαινόμενο της τυποποίησης στην επιστημονική βιβλιογραφία δεν είναι ασυνήθιστο, η επανάπαυση στις προτεινόμενες ΠΤΝ λύσεις μειώνει την επιστημονική δημιουργικότητα. Παράλληλα, η απλοποίηση του σχηματισμού θεμελιωδών ερευνητικών ερωτημάτων ενέχει τον κίνδυνο διαμόρφωσης ακραίων υποθέσεων, αναζητώντας απομακρυσμένες συσχετίσεις εκεί που δεν υπάρχουν. Στην εργασία αυτή εξετάζονται τα παραπάνω φαινόμενα σε δημοσιευμένα επιστημονικά άρθρα αναδρομικών βιοπληροφορικών ερευνών, πριν και μετά το 2020, όταν αναπτύχθηκαν προηγμένα μοντέλα ΠΤΝ με μετασχηματιστές. Εξετάζονται πρωτότυπα επιστημονικά άρθρα ευρετηριασμένα στις βάσεις δεδομένων PubMed, iCite, Scopus και Web of Science. Η απομάκρυνση διπλότυπων πραγματοποιήθηκε μέσω DOI των άρθρων, ενώ ακολούθησε περαιτέρω έλεγχος σχετικότητας με το θέμα. Βρέθηκαν 12.367 άρθρα σε όλες τις βάσεις που εξετάστηκαν, 9.282 χωρίς τα διπλότυπα, από τα οποία 5890 δημοσιεύτηκαν από το 2020 και μετά. Αναλύθηκαν οι τίτλοι και οι περιλήψεις τους για τον εντοπισμό της συχνότητας εμφάνισης επιστημονικών όρων με Λανθάνουσα Κατανομή Dirichlet στο πρόγραμμα VOSviewer. Μεταξύ

των επιστημονικών άρθρων που δημοσιεύτηκαν μετά το 2020 παρατηρήθηκε αυξημένη τυποποίηση μεθοδολογιών σε σχέση με πριν το 2020. Ακόμη διαπιστώθηκαν αρκετές περιπτώσεις ακραίων συσχετισμών μεταξύ βιολογικών παραμέτρων μετά το 2020 με προοδευτική αύξηση ανά έτος. Η αύξηση αυτή αποδίδεται στην αναδρομική επαναχρησιμοποίηση των θεμελιακών υποθέσεων που δημιουργήθηκαν από PTN μοντέλα από τα ίδια ή άλλα μοντέλα PTN. Η παραγωγή πορισμάτων, υποθέσεων και συμπερασμάτων ασύμβατα με την πραγματικότητα υποβαθμίζουν την επιστημονική γνώση

Conference paper's details in English

The paper was presented at the **31st Panhellenic Academic Libraries Conference**, which took place at Ioannina (Greece) during 22-24 of October 2025. Conference website: <https://en.31palc.lib.uoi.gr/>

Title

The Generative AI: Contaminating Scientific Thought

Details of author / authors

Artemis Chaleplioglou, Department of Archival, Libraries and Information Studies, University of West Attica, artemischal@uniwa.gr

Christos Zampakolas, Department of Archival, Libraries and Information Studies, University of West Attica, christoszampakolas@uniwa.gr

Keywords

1. Longitudinal cohort study
2. Bibliometrics
3. Bioinformatics
4. Retrospective studies
5. Generative Artificial Intelligence Models Collapse

Abstract

Based on Large Language Models (LLMs), neuronal networks, and transformers, Generative Artificial Intelligence (GenAI) generates an impressive amount of text, figures, sounds, and videos by analyzing massive volumes of training data. Through the simulation and mimicry of original documents derived from a wide array of scientific publications used in their pre-training, various GenAI models have already produced scientific texts of such quality that they can be considered as the work of scientists. However, when GenAI is applied to develop reasoning for fundamental research questions, identify gaps, inconsistencies, or contradictions in scientific literature, or design research methodologies, it establishes a self-sustaining environment for generating hypotheses while enhancing the formalization of research. This phenomenon is not unusual in the scientific literature, but relying solely on GenAI solutions reduces scientific creativity. Additionally, oversimplifying the formulation of scientific questions risks posing high-stakes inquiries that seek far-fetched correlations where none exist. These phenomena are explored in a portfolio of bioinformatics retrospective studies published prior to and following 2020, when the first GenAI models based on transformers were introduced. Original research articles indexed in PubMed, iCite, Scopus, and Web of Science were investigated. After removing duplicates by identifying matching DOIs, a relevance check was performed to determine inclusion based on the investigation topic. A total of 12,367 articles were found in all the databases examined, 9,282 after removing duplicates, of which 5,890 were published from 2020 onwards. The titles and abstracts of these papers were analyzed for the frequency of scientific terms related to Latent Dirichlet Allocation using VOSviewer software. There was an increase in the formulation of methods in papers published after 2020 compared to those published prior to this year. High-stakes scientific questions examining correlations between distant biological parameters became more prevalent in papers published after 2020, with a gradual increase over time. This growth is attributed to the retrospective reuse of GenAI-generated hypotheses by the same

or other GenAI models. The unrealistic deductive reasoning, hypothesis formulation, and conclusion extraction undermine scientific knowledge.