Διπλωματική Εργασία : Συστηματική βιβλιογραφική ανασκόπηση μεθόδων ανίχνευσης text-based ψευδών ειδήσεων

Επιβλέπων καθηγητής

Αλεξόπουλος Χαράλαμπος

Επιβλέπων υποψήφιος διδάκτορας

Παπαγεωργίου Γεώργιος

Προπτυχιακός Φοιτητής

Κουτσώνης Αναστάσιος

Στόχος

Η ανάλυση των αποδοτικών υφιστάμενων ερευνών στον τομέα της ανίχνευσης text-based ψευδών ειδήσεων.

Έναρξη διπλωματικής εργασίας : Νοέμβριος του 2024

Περιεχόμενα

Περιεχόμενα ---------- 1
Πίνακας εικόνων ---------- 2
1)Εισαγωγή Εργασίας ---------- 4
1.1)Περίληψη ---------- 4
1.2)Ερευνητική Μεθοδολογία SLR ---------- 4
1.3)Αναζήτηση Επιστημονικών Άρθρων ---------- 5
1.4)Κριτήρια Ένταξης/Αποκλεισμού ---------- 5
2)Εισαγωγή στα Fake News ---------- 8
2.1)Ορισμός των Fake News ---------- 8
2.2)Ιστορία των Fake News ---------- 9
2.3)Κατηγορίες των Fake News ---------- 10
2.4)Κοινωνικός Αντίκτυπος ---------- 11
2.4.1) Ψευδείς Ειδήσεις στα Μέσα Κοινωνικής Δικτύωσης ---------- 12
Παραδείγματα ---------- 14
2.4.2) Ψευδείς Ειδήσεις στην Πολιτική ---------- 16
Παραδείγματα ---------- 17
2.4.3) Ψευδείς Ειδήσεις στην Οικονομία ---------- 18
Παραδείγματα ---------- 18
2.4.4) Ψευδής Ειδήσεις στον Αθλητισμό ---------- 19
Παραδείγματα ---------- 20
2.4.5) Ψευδείς ειδήσεις στον Ιδιωτικό Τομέα- Βιομηχανίες ---------- 21
Παραδείγματα ---------- 21
2.4.6 Ψευδείς Ειδήσεις στον Υγειονομικό Τομέα ---------- 23
Παραδείγματα ---------- 23
3)Έρευνα και Μέθοδοι Ανίχνευσης Fake News ---------- 25

3.1)Έρευνα πάνω στα Fake News----------25
3.2)Μέθοδοι Ανίχνευσης ----------27
3.3) Ο Ρόλος της Επεξεργασίας Φυσικής Γλώσσας στην Ανίχνευση ----------29
3.4)Ανίχνευση μέσω Μηχανικής Μάθησης ----------32
3.5) Ανίχνευση με Βαθιά Μάθηση----------35
3.5.1) Ανίχνευση μέσω Νευρωνικών Δικτύων----------36
3.5.2) Ανίχνευση με χρήση Μετασχηματιστών ----------39
4) Πειράματα ----------42
4.1) Εισαγωγή στα Πειράματα----------42
4.2) Μέθοδοι Ανίχνευσης Ποσοτικά ----------42
4.3) Θεματολογίες των Συνόλων Δεδομένων----------43
4.4) Διαθεσιμότητα των Δεδομένων ----------45
4.5) Προέλευση Δεδομένων ----------46
4.6) Εξαγωγή Χαρακτηριστικών από τα Δεδομένα----------47
4.7) Χρήση των Διαθέσιμων Μέτα-Δεδομένων ----------48
4.8) Το Πρόβλημα του Overfitting----------49
4.9) Ανάγκες Υπολογιστικών Πόρων ----------50
4.10) Ικανότητα Κατανόησης των Συμφραζομένων ----------51
4.10) Τρόποι Αξιολόγησης των Πειραμάτων----------52
4.11) Σπάνια Χαρακτηριστικά των Πειραμάτων----------53
4.12) Στατιστική Απεικόνιση των Πειραμάτων ----------53
5)Συμπέρασμα ----------54
6)Προτεινόμενη Μελλοντική Επέκταση της Εργασίας ----------55
7)Βιβλιογραφία----------55
7.1Βιβλιογραφία πειραμάτων μέσω SLR----------58

Πίνακας εικόνων

Εικόνα 1: Διάγραμμα επιλογής άρθρων μετά την πρώτη κατηγορία κριτηρίων ένταξης/αποκλεισμού..... 6

Εικόνα 2 : Διάγραμμα επιλογής άρθρων μετά την εφαρμογή των 2ων κριτηρίων ένταξης/αποκλεισμού 7
Εικόνα 3 Σύνολο ειδήσεων και υποσύνολα φημών, νέων και ψευδών ειδήσεων [3] 8
Εικόνα 4 : Διάγραμμά ζωής μιας ψευδούς είδησης[9] 10
Εικόνα 5 : Οι συχνότερες 20 λέξεις σε αληθείς ειδήσεις [11] 12
Εικόνα 6 : Κορυφαίες 20 λέξεις σε ψευδείς ειδήσεις [11] 13
Εικόνα 7 : Ποσοστό αυτοπεποίθησης χρηστών να αναγνωρίσουν τις ψευδείς ειδήσεις [15] 16
Εικόνα 8 : Δημοσιευμένα papers που αφορούν Φήμες και Ψευδείς ειδήσεις[19] 26
Εικόνα 9 : Δημοσιευμένα papers που αφορούν Ανίχνευση Φημών και Ψευδών ειδήσεων [19] 26
Εικόνα 10 : Τεχνικές ανίχνευσης μέχρι το 2019 [19] 27
Εικόνα 11 : Κατηγορίες και υποκατηγορίες μεθόδων ανίχνευσης [26] 29
Εικόνα 12 : Κατηγορίες Machine Learning προσεγγίσεων [9] 33
Εικόνα 13 : Διαχωρισμός των Deep Learning τεχνικών για text-based tasks, σε Transformers και Neural Networks [9] 40
Εικόνα 14 : Ημερομηνίες κυκλοφορίας των Papers που περιλαμβάνουν πείραμα 42
Εικόνα 15 : Διαχωρισμός μεταξύ υβριδικών και μη προσεγγίσεων στο σύνολο των πειραμάτων 43
Εικόνα 17 : Ποσοτική απεικόνιση των διαθέσιμων datasets 45
Εικόνα 16 : Προέλευση των datasets από το Kaggle 46
Εικόνα 19 : Αυτόματη ή χειροκίνητη εξαγωγή χαρακτηριστικών 47
Εικόνα 20 : Χρήση επιπρόσθετων meta-data 48
Εικόνα 21: Αναγνώριση του Overfitting 49
Εικόνα 22 : Υπολογιστικές απαιτήσεις 50
Εικόνα 23 : Ικανότητα κατανόησης συμφραζομένων 51

1)Εισαγωγή Εργασίας

1.1)Περίληψη

Η παρούσα διπλωματική εργασία πραγματεύεται κατά κύριο λόγο τις αποδοτικές μεθόδους ανίχνευσης text-based ψευδών ειδήσεων. Αρχικά, ορίζεται η ύπαρξη των ψευδών ειδήσεων από την γέννηση μέχρι και την εδραίωση τους ως ένα μείζων κοινωνικό πρόβλημα της ψηφιακής εποχής, καθώς η πλειοψηφία της ανθρωπότητας διατηρεί προσωπικούς λογαριασμούς στα μέσα κοινωνικής δικτύωσης. Έπειτα αναφέρονται τα είδη ψευδών ειδήσεων αλλά και ο κοινωνικός τους αντίκτυπος. Στην συνέχεια, γίνεται και αναφορά στην ύπαρξη Fake News στην πολιτική, την οικονομία, τον αθλητισμό, τον ιδιωτικό και τον υγειονομικό τομέα. Στο επόμενο κεφάλαιο, αναλύεται η ακαδημαϊκή έρευνα. Αρχικά, αναφέρονται οι τεχνολογίες αιχμής που πλαισιώνουν τα μοντέλα ανίχνευσης αναλύοντας όλες τις πτυχές της χρησιμότητας τους. Έπειτα, υπάρχει το κεφάλαιο ανάλυσης των πειραμάτων όπου παρουσιάζονται χρήσιμες ποσοτικές πληροφορίες των πενήντα τριών ερευνών που αναλύθηκαν αλλά και συμπεράσματα που προέκυψαν. Στο τέλος, αναφέρονται οι πιθανές μελλοντικές επεκτάσεις της παρούσας έρευνας αλλά και ένα γενικό συμπέρασμα για τις ανάγκες του κλάδου.

1.2)Ερευνητική Μεθοδολογία SLR

Η διπλωματική εργασία έχει δομηθεί ακολουθώντας τη μέθοδο της Συστηματικής Βιβλιογραφικής Ανασκόπησης (Systematic Literature Review - SLR). Η μέθοδος αυτή, αποτελεί μια αξιόπιστη προσέγγιση για την ανασκόπηση υφιστάμενων ερευνητικών εργασιών. Η επιλογή της μεθοδολογίας αυτής, εξασφαλίζει την ακεραιότητα τη ς συγκέντρωσης, της ανάλυσης και της αξιολόγησης των υπαρχόντων επιστημονικών δεδομένων με έναν αντικειμενικό και επαναληπτικό τρόπο προσφέροντας διαφάνεια αλλά και μελλοντική επαναχρησιμοποίηση.

Η μεθοδολογία που ακολουθήθηκε περιλαμβάνει τα εξής στάδια :

1. Καθορισμός λέξεων κλειδιών που εμφανίζονται με μεγάλη συχνότητα στα επιστημονικά άρθρα όπου έχουν ως κύριο πεδίο την ανίχνευση ψευδών ειδήσεων.

2. Καθορισμός του ερευνητικού Query με σκοπό την λήψη των καταλληλότερων αποτελεσμάτων. Το παρόν query δημιουργήθηκε με σκοπό να μας οδηγήσει σε αποτελέσματα που εστιάζουν στην τεχνική ανίχνευσης και στην αποτελεσματικότητα.

3.Καθορισμός επιστημονικών βάσεων δεδομένων όπως το IEEE , Sciece Direct και Scopus. Η επιλογή των papers έγινε με βάση το query, ενώ παράλληλα υπήρχε και φιλτράρισμα με βάση τα κριτήρια ένταξης/αποκλεισμού.

Λέξεις κλειδιά

Fake News Detection, Machine Learning, Deep Learning , Natural Language Processing, Neural Networks, Transformers, Efficiency, Accuracy.

1.3)Αναζήτηση Επιστημονικών Άρθρων

Βάσεις Δεδομένων : IEEE , Scopus , Science direct

Query : ("Fake news detection " AND ("machine learning" OR "deep learning" OR "natural language processing") AND ("misinformation" OR "disinformation" OR " malinformation ") AND ("evaluation metrics" OR "accuracy"))

Η επιλογή του συγκεκριμένου **Query** προέκυψε από την ανάγκη της εργασίας να επικεντρωθεί στην ανίχνευση ψευδών ειδήσεων χρησιμοποιώντας αρχικά το Fake news detection. Στην συνέχεια, πλαισιώθηκε με τους τεχνολογικούς κλάδους ML,DL και NLP όπου ο συνδυασμός τους περιλαμβάνει και τα τεχνολογικά τους παρακλάδια όπως τα Neural Networks και οι Transformers όπου αποτελούν τεχνολογίες αιχμής στο τομέα ανίχνευσης. Επιπρόσθετα, για να βρεθούν οι καταλληλότεροι τρόποι ανίχνευσης με βάση το είδος των ψευδών ειδήσεων αλλά και την ακρίβεια τους, χρησιμοποιήθηκε το κομμάτι που αναφέρει τα είδη ψευδών ειδήσεων και την αποτελεσματικότητα.

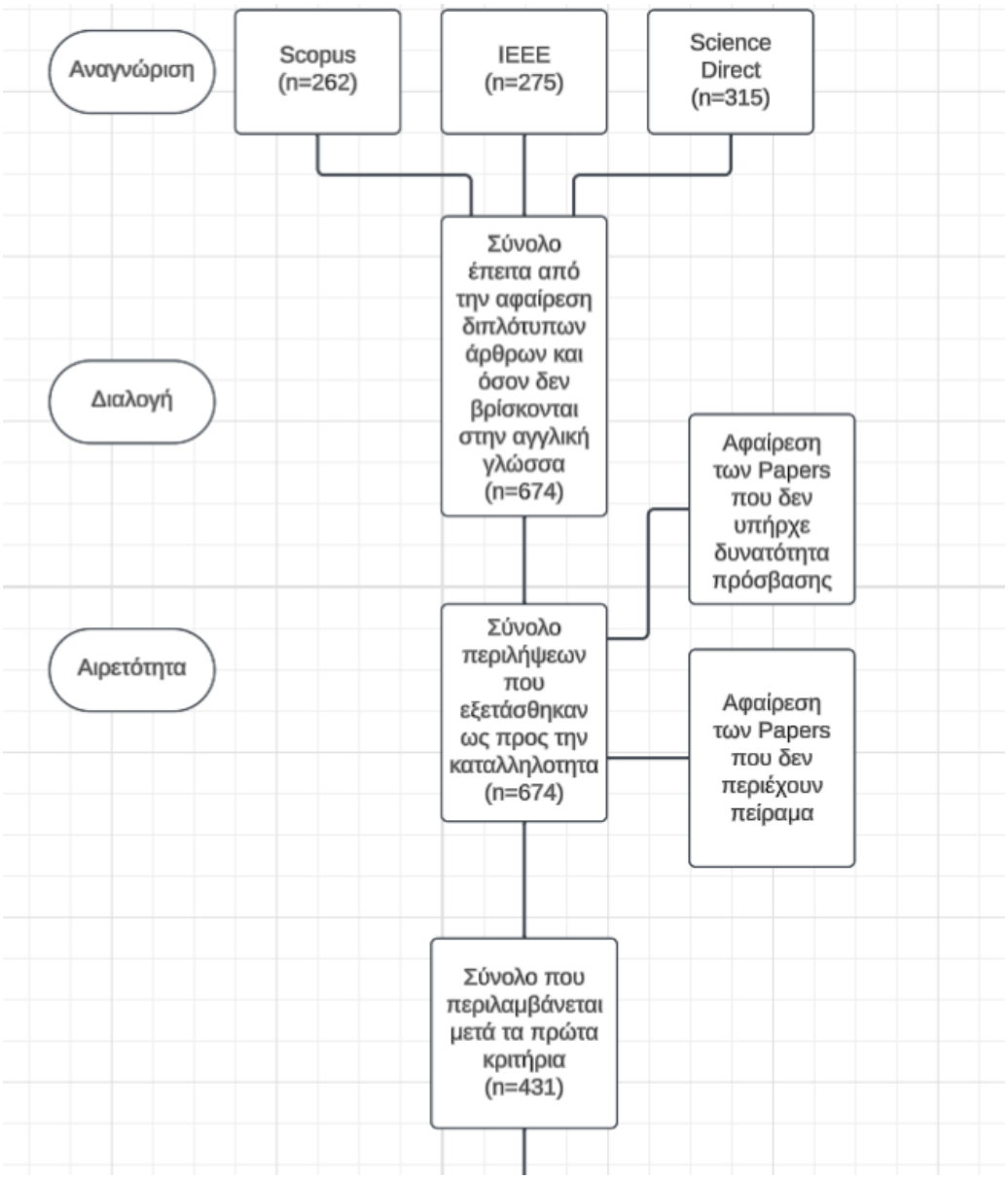
Τέλος, το συγκεκριμένο query αναζητά υποχρεωτικά την ακολουθία ‘fake news detection’, τουλάχιστον μια από τις τεχνολογίες NLP,DL,ML , τουλάχιστον ένα από τα είδη misinformation, disinformation, malinformation και τουλάχιστον ένα από τα evaluation metrics και accuracy. Οι λέξεις αυτές αναζητούνται είτε στον τίτλο, είτε στις λέξεις κλειδιά, είτε στην περίληψη της υφιστάμενης έρευνας.

1.4)Κριτήρια Ένταξης/Αποκλεισμού

Περιγραφή κριτηρίων ένταξης/αποκλεισμού: Τα επιστημονικά άρθρα που επιλέγονται για ανάλυση πρέπει να πληρούν τα ακόλουθα κριτήρια:

- **Ψευδείς ειδήσεις**: Η δημοσίευση πρέπει να εστιάζει στις ψευδείς ειδήσεις.
- **Ανίχνευση**: Η δημοσίευση πρέπει να εστιάζει στην μέθοδο ανίχνευσης.
- **Γλώσσα**: Οι δημοσιεύσεις πρέπει να είναι στην αγγλική γλώσσα.
- **Χρονικό πλαίσιο**: Οι δημοσιεύσεις πρέπει να έχουν δημοσιευθεί μεταξύ 2018 και 2024.

- **Δυνατότητα πρόσβασης :** Οι δημοσιεύσεις θα πρέπει να διαθέτουν δυνατότητα πρόσβασης.
- **Υλοποίηση Πειράματος** : Οι δημοσιεύσεις θα πρέπει να παρουσιάζουν υλοποιημένο πείραμα ανίχνευσης Fake News.

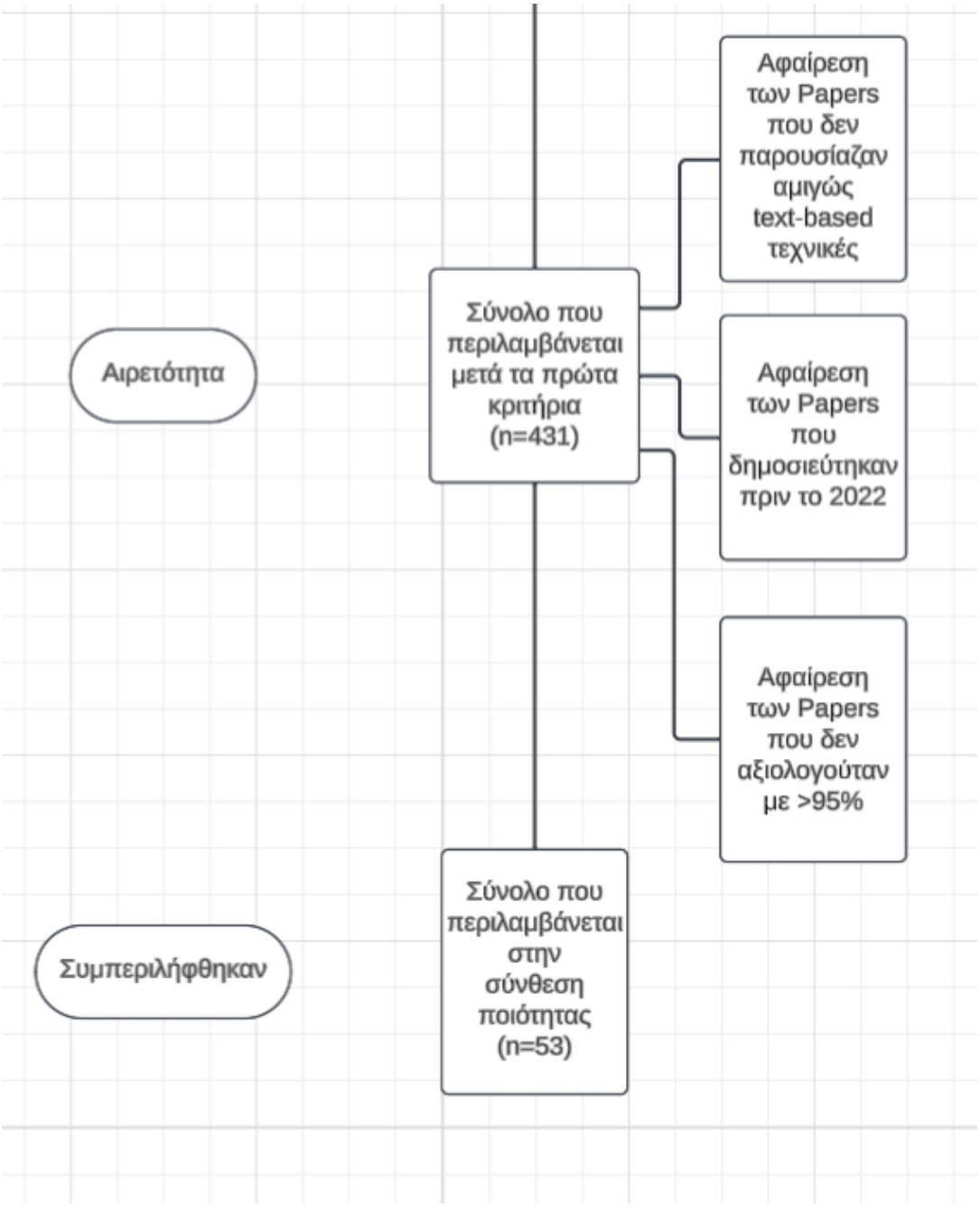


Εικόνα 1: Διάγραμμα επιλογής άρθρων μετά την πρώτη κατηγορία κριτηρίων ένταξης/αποκλεισμού

Επιπρόσθετα κριτήρια ένταξης/αποκλεισμού

Με σκοπό την ποιοτικότερη εξαγωγή επιστημονικών ερευνών αλλά και σύγχρονων/αποδοτικών πειραμάτων προστέθηκαν τα εξής κριτήρια.

- **Text-based** : Τα papers επικεντρώνονται σε πειράματα text-based τεχνικών ανίχνευσης.
- **Απόδοση** : Τα πειράματα πρέπει να έχουν πετύχει τιμές αξιολόγησης μεγαλύτερες του 95% κατά το μέσο όρο των μετρικών.
- **Επίκαιρα** : Τα papers θα πρέπει να έχουν δημοσιευθεί μεταξύ 2022-2024.

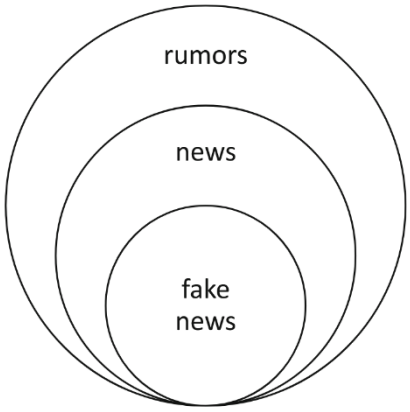


Εικόνα 2 : Διάγραμμα επιλογής άρθρων μετά την εφαρμογή των 2ων κριτηρίων ένταξης/αποκλεισμού

2)Εισαγωγή στα Fake News

2.1)Ορισμός των Fake News

Οι ψευδείς ειδήσεις αποτελούν ειδησεογραφικό περιεχόμενο το οποίο είναι **ψευδώς κατασκευασμένο**, **εσκεμμένα παραποιημένο** ή **απλά αναληθές**[1]. Οι δημιουργοί του, το κατασκευάζουν και το διαμοιράζουν με διάφορους σκοπούς όπως να βλάψουν, να αποπροσανατολίσουν ή να χειραγωγήσουν απόψεις. Παρατηρούνται σε όλους τους τομείς όπως η πολιτική, η οικονομία, η υγεία και ο αθλητισμός. Επιπρόσθετα, οι ψευδείς ειδήσεις φιλοξενούνται από διαδικτυακές ειδησεογραφικές σελίδες αλλά κατά κύριο λόγο εμφανίζονται στα μέσα κοινωνικής δικτύωσης[2]. Έχουν υψηλή επιζήμια ικανότητα και διαμοιράζονται ταχύτατα μέσω της κοινοποίησης και αναπαραγωγής από ανυποψίαστους χρήστες.



Εικόνα 3 Σύνολο ειδήσεων και υποσύνολα φημών, νέων και ψευδών ειδήσεων [3]

2.2)Ιστορία των Fake News

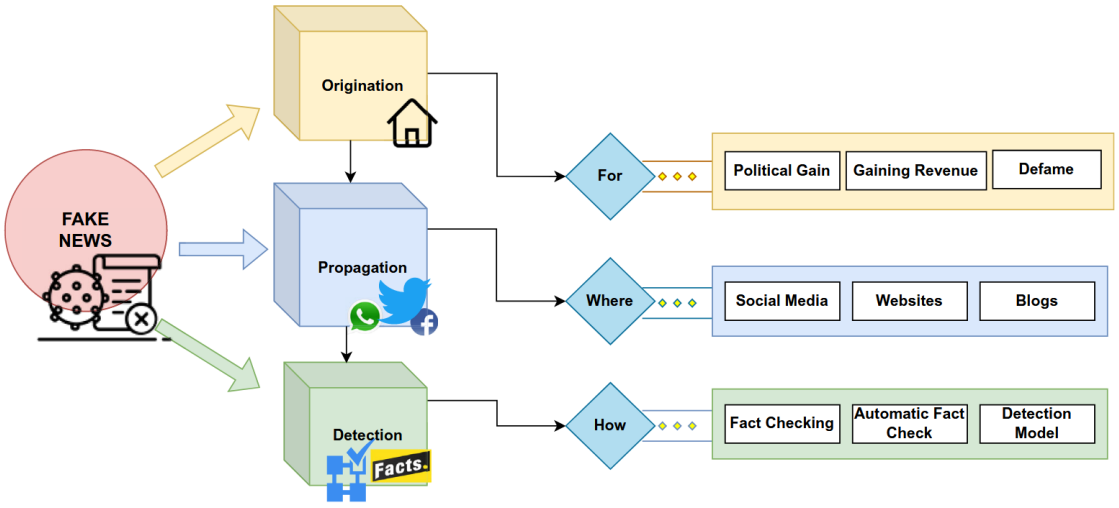
Η ιστορία των **Fake News** αρχίζει περίπου κατά τα μέσα του 15ου αιώνα[4], παρόλο που έγιναν δημοφιλή και ορίστηκαν στις αρχές του 21ου αιώνα. Το πρώτο παράδειγμα εμφάνισης και διάδοσης ψευδών ειδήσεων χρονολογείται κατά τον 15ο αιώνα όπου στην εποχή εκείνη, είχε ήδη εμφανιστεί η εκτυπωτική τεχνολογία του **Γουτεμβέργιου**[4]. Η τεχνολογική πρόοδος αυτή, συνέβαλε στο να εμφανιστεί για πρώτα φορά χρονικά ο εύκολος διαμοιρασμός του έντυπου τύπου αλλά και η εκτύπωση πανό. Δυστυχώς όμως, κατά την τότε εποχή της **Αναγέννησης** αλλά και της **Μεταρρύθμισης**, η τεχνολογία του Γουτεμβέργιου χρησιμοποιήθηκε καταχρηστικά καθώς για πρώτη φορά κυκλοφόρησαν ψευδείς ειδήσεις που περιείχαν φήμες και πλαστά νέα[4]. Αφορούσαν τις τότε πολιτικές και θρησκευτικές φιγούρες με σκοπό να τις δυσφημίσουν ή να τις βλάψουν. Η εκκλησία είχε αντιδράσει έντονα αναγνωρίζοντας την επικινδυνότητα τους. Η αναταραχή αυτή, ώθησε τον βασιλιά της Γαλλίας **Φραγκίσκο Ά** να επιβάλλει νόμο εφαρμογής της θανατικής ποινής σε όσους χρησιμοποιούσαν χωρίς άδεια την εκτυπωτική τεχνολογία του Γουτεμβέργιου. Ο βασικός σκοπός του βασιλιά ήταν η αποτροπή της διάδοσης αλλά ο νόμος απέτυχε διότι η προσβασιμότητα στην πληροφορία είχε ήδη αναπτυχθεί ριζικά[4]. Ένα ευρέως γνωστό παραδείγματα έντυπων ψευδών ειδήσεων αποτελούν τα **Canards** τον 17ο αιώνα όπου περιείχαν ιστορίες για τέρατα από την Χιλή και γελοιογραφίες της Μαρίας Αντουανέτας με σκοπό την υποβάθμιση της γαλλικής επανάστασης.[4] Άρθρο πληροφοριών

Με το πέρασμα των αιώνων, η φαινομενική άνοδος της δημοσιογραφίας παρείχε στους συντάκτες σημαντικές ευκαιρίες για κερδοσκοπία μέσω συναισθηματικά φορτισμένων ψευδών ειδήσεων όπου κέντριζαν το ενδιαφέρον του κόσμου[5]. Κατά τον 19ο αιώνα, πρωτοεμφανίστηκε ο **"κίτρινος τύπος"** στις Ηνωμένες Πολιτείες της Αμερικής, όπου δημοφιλείς εφημερίδες όπως των δημοσιογράφων **Χερστ** και **Πούλιτζερ** εκμεταλλευόντουσαν δραματοποιημένες ιστορίες με αποτέλεσμα να εκτοξεύσουν τις πωλήσεις τους[5]. Τέτοιου είδους αφηγήματα, δεν ήταν απαραίτητα εντελώς ψευδή, ωστόσο παρουσίαζαν σκοπίμως παραμορφωμένα τα γεγονότα και τις καταστάσεις ώστε να επιτευχθεί η πώληση[5].

Κατά την σύγχρονη εποχή, η εμφάνιση του διαδικτυακού ειδησεογραφικού τύπου και των μέσων κοινωνικής δικτύωσης άλλαξε δραστικά τη δυναμική των fake news[6]. Αρχικά, από την πρώτη δεκαετία του 2000, τα μέσα κοινωνικής δικτύωσης όπως το Facebook και το Twitter διευκόλυναν την άμεση εξάπλωσή τους[6]. Σημαντικότερο παράδειγμα, οι **προεδρικές εκλογές** των **ΗΠΑ** το 2016, όπου θεωρούνται σημείο καμπής και εναρκτήρια αφορμή για περαιτέρω ακαδημαϊκή έρευνα ως προς την

ανίχνευση[7] [2]. Είναι η πρώτη φορά στην ιστορία που αναδείχθηκε ο ρόλος των ψευδών ειδήσεων στη χειραγώγηση της κοινής γνώμης και στην πολιτική πόλωση[2]. Το αποτέλεσμα ήταν να κλονιστεί η δημοκρατία καθώς είχαν ήδη επηρεαστεί οι πολιτικές απόψεις των ψηφοφόρων. Τα **Fake News** προέρχονταν από οργανωμένες καμπάνιες προπαγάνδας στα social media αλλά και από bots τα οποία είχαν spamming ικανότητες με σκοπό την μέγιστη διάδοση του ψεύδους[8]. Τα γεγονότα αυτά, διευκόλυναν τους ερευνητές να κατανοήσουν την επιζήμια τους ικανότητα ενώ παράλληλα τους ανέδειξαν την επιτακτική ανάγκη για εκτεταμένη ακαδημαϊκή έρευνα στο κομμάτι της ανίχνευσης[7].

Στις μέρες μας, τα **Fake News** αποτελούν παγκόσμιο πρόβλημα και ερευνητική πρόκληση, με την επιστημονική κοινότητα να αναζητά τρόπους ανίχνευσης και πρόληψης. Η παραπληροφόρηση υπονομεύει την εμπιστοσύνη στην ενημέρωση και τις δημοκρατικές διαδικασίες. Κάθε είδους ψευδούς είδησης περιέχει κάποιον σκοπό όπως ο πολιτικός και ο υγειονομικός κλονισμός. Διακινείται διαδικτυακά μέσω των μέσων κοινωνικής δικτύωσης αλλά και ειδησεογραφικών διαδικτυακών σελιδών[6], [7].



Εικόνα 4 : Διάγραμμά ζωής μιας ψευδούς είδησης[9]

2.3)Κατηγορίες των Fake News

Ψευδή Πληροφορία- Misinformation: Είναι αναληθή πληροφορία που διαδίδεται χωρίς να έχει απόλυτο σκοπό να βλάψει, όπως ειδήσεις που αναφέρουν ανακριβή στοιχεία λόγω ημιμάθειας ή δημοσιογραφικής ανακρίβειας.[1]

Παραπληροφόρηση - Disinformation: Είναι δημιουργημένη ψευδή πληροφορία που διαδίδεται εσκεμμένα με σκοπό να παραπλανήσει, να χειραγωγήσει και να προκαλέσει ζημιά σε άτομα, οργανισμούς ή και ολόκληρη την κοινωνία.[1]

Επιβλαβή Είδηση - Malinformation : Είναι αληθή πληροφορία η οποία διαδίδεται και παρουσιάζεται επιλεκτικά με σκοπό να βλάψει, όπως η ανάρτηση προσωπικών στοιχείων ενός μη καταδικασμένου εγκληματία. [1]

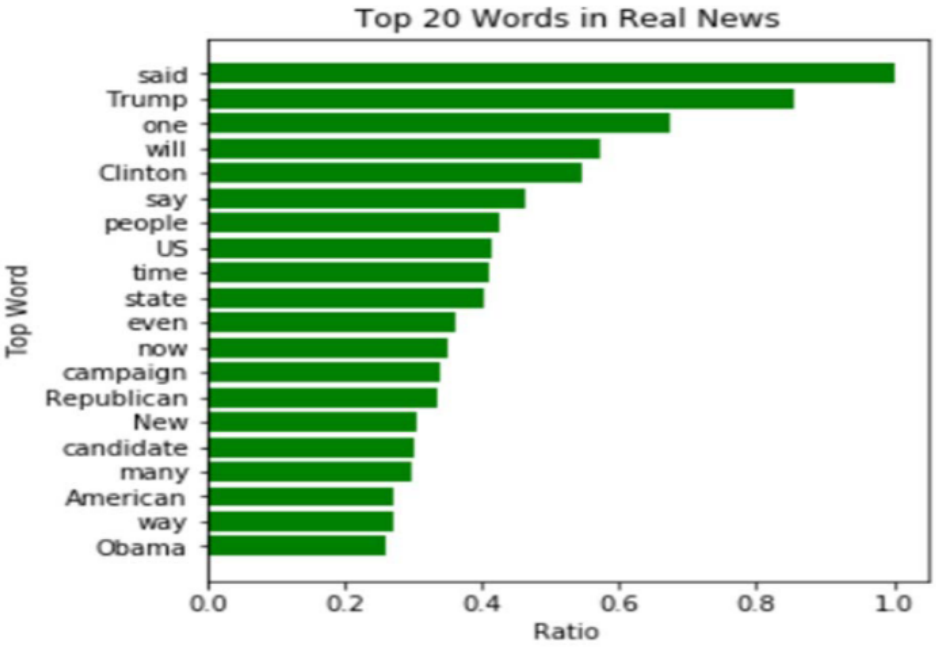
2.4)Κοινωνικός Αντίκτυπος

Τα **Fake News** συνοδευόμενα από τις εξαιρετικά επιζήμιες ικανότητες τους μπορούν να προκαλέσουν σημαντικό κοινωνικό αντίκτυπο[7], [10]. Κατέχουν την δύναμη να διαστρεβλώνουν την πραγματικότητα και να υπονομεύουν την εμπιστοσύνη της κοινωνίας στις θεσμικές και δημοσιογραφικές πηγές. Επίσης, είναι ικανά να ενισχύσουν τη διάσπαση της κοινωνικής συνοχής και να προκαλέσουν διχασμό μέσω της παραπληροφόρησης και της δημιουργίας φανατισμένων απόψεων[2]. Ειδικότερα, σε κρίσιμες περιόδους, όπως οι εκλογές ή οι υγειονομικές κρίσεις, τα Fake News ενισχύουν τις καταστάσεις πανικού, επηρεάζουν τη λήψη αποφάσεων και προκαλούν καταστάσεις σύγχυσης και αποδιοργάνωσης. Η ανάγκη για χρήση μεθόδων μετρίασης του φαινομένου κρίνεται επιτακτική καθώς πέρα από την ανάπτυξη τεχνικών ανίχνευσης που παρέχει η επιστημονική κοινότητα, είναι αναγκαία η εκπαίδευση της κριτικής σκέψης των αναγνωστών αλλά και η καλλιέργεια της ικανότητας της εγρήγορσης καθώς λειτουργούν ως κόμβοι αναπαραγωγής με το να κοινοποιούν μια είδηση όπου δεν έχουν αντιληφθεί ότι είναι ψευδή[10].
Επιπρόσθετα, τα Fake news βλάπτουν σημαντικά όλες τις κοινωνικές ομάδες και όλα τα πεδία όπως η πολιτική, η οικονομία, η υγεία, ο αθλητισμός και οι ιδιωτικές εταιρίες-βιομηχανίες.[1], [10]

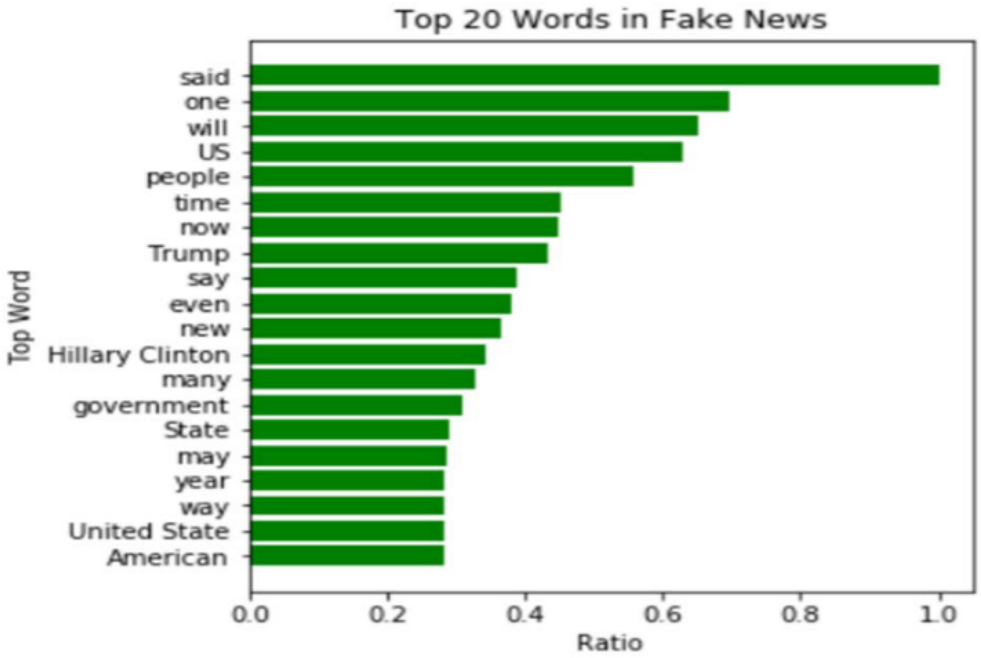
Συμπερασματικά, η ύπαρξη δημοκρατικών θεσμών που προάγουν την συμμετοχή και την ισότητα (**inclusive institutions**) αποτελούν θεμελιώδη προϋπόθεση για την σκλήρυνση της ανθεκτικότητα μιας κοινωνίας απέναντι στην εξάπλωση των Fake News. Όπως επισημαίνουν οι **Daron Acemoglu** και **James A. Robinson** στο εμπορικό τους βιβλίο '**Why Nations Fail**', οι θεσμοί αυτοί ενισχύουν τη συμμετοχικότητα των πολιτών στις διαδικασίες λήψης αποφάσεων, ενώ παράλληλα εξασφαλίζουν την προσβασιμότητα σε αξιόπιστη και τεκμηριωμένη πληροφόρηση. Προάγουν τη διαφάνεια με σκοπό τον περιορισμό της επιρροής της παραπληροφόρησης. Why NationsFail (βιβλίο)

2.4.1) Ψευδείς Ειδήσεις στα Μέσα Κοινωνικής Δικτύωσης

Τα μέσα κοινωνικής δικτύωσης όπως το Facebook και το Twitter χαρακτηρίζονται ως οι πιο άμεσοι και ισχυρότεροι κόμβοι διάδοσης ειδήσεων[2]. Οι δημιουργοί των ψευδών ειδήσεων χρησιμοποιούν καταχρηστικά τα social media, εκμεταλλευόμενοι του πόσο δημοφιλή είναι σε όλες τις κοινωνικές ομάδες.[10] Η αντιμετώπιση του προβλήματος αυτού, έχει προσελκύσει έντονο ερευνητικό ενδιαφέρον. Οι ακαδημαϊκοί ερευνητές έχουν ήδη αναπτύξει social media-based συστήματα ανίχνευσης εκμεταλλευόμενοι τα διαθέσιμα σύνολα δεδομένων που περιλαμβάνουν ψευδή tweets ή αναρτήσεις που έχουν εμφανιστεί κατά καιρούς στις πλατφόρμες αυτές[10].



Εικόνα 5 : Οι συχνότερες 20 λέξεις σε αληθείς ειδήσεις [11]



Εικόνα 6 : Κορυφαίες 20 λέξεις σε ψευδείς ειδήσεις [11]

Οι μέθοδοι ανίχνευσης ψευδών ειδήσεων που εμφανίζονται στα μέσα κοινωνικής δικτύωσης, βασίζονται σε προσεγγίσεις που αξιοποιούν τα 'propagation-based features'. Τα χαρακτηριστικά αυτά αφορούν τα μοτίβα διάδοσης, των αριθμό των χρηστών που συμμετέχουν στην διάδοση (retweets,shares) αλλά και την ταχύτητα διάδοσης. Τα 'behavioral user signals' όπου αναφέρονται στις αντιδράσεις των χρηστών όπως likes,comments αλλά και στα 'crowd signals' όπου αφορούν την ανάλυση συλλογικών δεδομένων συμπεριφοράς του πλήθους.[10]

Η Γεωμετρική βαθιά μάθηση απεδείχθη ως μια πολλά υποσχόμενη τεχνική για την ανάλυση των δικτύων διάδοσης ψευδών ειδήσεων καθώς επιτρέπει την μοντελοποίηση των σχέσεων μεταξύ χρηστών και περιεχομένου με σκοπό την δημιουργία μοτίβων αναγνώρισης αλλά και συνηθειών[12]. Τα νευρωνικά δίκτυα διαθέτουν την ικανότητα να ενσωματώνουν δομικές και σημασιολογικές πληροφορίες του κειμένου με σκοπό να βελτιώσουν την ακρίβεια στην ανίχνευση ψευδών ειδήσεων[11], [12].

Με την παρακολούθηση της συνεχούς εξέλιξης των τεχνικών ανίχνευσης, διακρίνουμε την ανάγκη για άμεση ανίχνευση ή έστω τον ταχύτατο δυνατό περιορισμό τους πριν επεκταθούν. Εξίσου σημαντική, είναι η ανάπτυξη επεξηγήσιμων μοντέλων (explainable models) όπου μέσω της τεχνητής νοημοσύνης, παρουσιάζονται στον χρήστη οι λόγοι της απόφασης να χαρακτηριστεί μια είδηση ως ψευδής εξαλείφοντας το φαινόμενο των 'μαύρων κουτιών'[7], [13]. Παρόλα αυτά, ακόμα μια πρόκληση που παραμένει και στα πιο σύγχρονα συστήματα είναι αυτό της δια λειτουργικότητας. Είναι ανάγκη λοιπόν, να αναπτυχθούν συστήματα ανίχνευσης τα οποία να διαθέτουν ικανότητες λειτουργίας

μεταξύ διαφορετικών πλατφορμών κοινωνικής δικτύωσης που έχουν ενιαία κριτήρια[14].

Οι χρήστες των social media παίζουν εξίσου σημαντικό ρόλο στον τομέα της ανίχνευσης. Πέρα από την εξέλιξη του πεδίου και την συνεχή ενσωμάτωση τεχνολογικών παραμέτρων θα πρέπει να ενισχυθεί η ενσωμάτωση κοινωνικών παραμέτρων' social factors'. Οι χρήστες θα πρέπει να εκπαιδευτούν στην αναγνώριση των ύποπτων ειδήσεων με σκοπό την αποφυγή της διάδοσης τους.[14]

Παραδείγματα

1) Meta fact-checkers

Αρχικά, είναι αναγκαίο να αναφέρουμε μια πρόσφατη δυσάρεστη είδηση που αφορά την εταιρία Meta όπου της ανήκει το **Instagram,** το **Facebook** και το **Whatsapp**. Τους πρώτους μήνες του 2025 λοιπόν, αποφασίστηκε από την εταιρία να αφαιρεθούν οι fact-checkers όπου βρισκόταν σε 60 διαφορετικές χώρες και κατείχαν καίριο ρόλο στην αντιμετώπιση της προπαγάνδας κυρίως στο Facebook. Οι ρόλος τους ήταν να μειώνουν την εμβέλεια της δημοσίευσης όταν αυτή θεωρούταν ψευδής αλλά και να προειδοποιούν τον χρήστη πριν την κοινοποίηση της, με σκοπό να μην αναπαραχθεί περαιτέρω ο διαμοιρασμός της. Παράλληλα, οι χρήστες του Instagram και του Facebook είχαν ανέκαθεν την δυνατότητα να κάνουν αναφορά δημοσίευσης για διάφορους λόγους όπως ο χαρακτηρισμός τους ως ψευδείς ειδήσεις.

Ο διευθύνων σύμβουλος της **Meta**, **Μαρκ Ζούκερμπεργκ**, ανέφερε ότι αυτή η αλλαγή στοχεύει στη μείωση της λογοκρισίας και στην προώθηση της ελευθερίας του λόγου στα μέσα κοινωνικής δικτύωσης που ανήκουν στην εταιρεία. Επιπρόσθετα, η Meta σκοπεύει να προωθήσει περισσότερο πολιτικό περιεχόμενο και να μεταφέρει τις ομάδες εποπτείας περιεχομένου(moderators) από την Καλιφόρνια στο Τέξας όπου θεωρείται πιο ευδόκιμο πολιτικό περιβάλλον για τις επιχειρήσεις.

Η απόφαση αυτή έχει προκαλέσει αρκετές αντιδράσεις. Η πλειοψηφία των χρηστών εξέφρασε την ανησυχία του και αναφέρει ότι η κατάργηση των ελεγκτών γεγονότων(fact-checkers) θα οδηγήσει σε αύξηση της παραπληροφόρησης και του επιβλαβούς περιεχομένου στις πλατφόρμες καθώς υπεύθυνοι για τον περιορισμό τους θα είναι μόνο οι χρήστες και καμία τεχνολογική μέθοδος.

Ο επίτροπος του συμβουλίου της Ευρώπης, υπεύθυνος για τα δικαιώματα του ανθρώπου, Μάικλ Ο' Φλάχερτι, προειδοποίησε ότι αυτή η κίνηση θα έχει αρνητικές επιπτώσεις στα ανθρώπινα δικαιώματα. Επιπλέον, οργανισμοί όπως το Full Fact, μια ανεξάρτητη φιλανθρωπική οργάνωση ελέγχου γεγονότων, εξέφρασε την απογοήτευσή της για την απόφαση της Meta, υποστηρίζοντας ότι αποτελεί ένα πισωγύρισμα στην καταπολέμηση των ψευδών ειδήσεων καθώς πέρα από το σύστημα 'σημειώσεων κοινότητας' (community notes), όπου οι ίδιοι οι χρήστες θα μπορούν να επισημαίνουν

και να διορθώνουν αναρτήσεις ως προς την ακρίβειά τους , η ίδια η πλατφόρμα αρνείται να τους προστατεύει. **Meta fact-checking announcement**

2) Youtube Policie

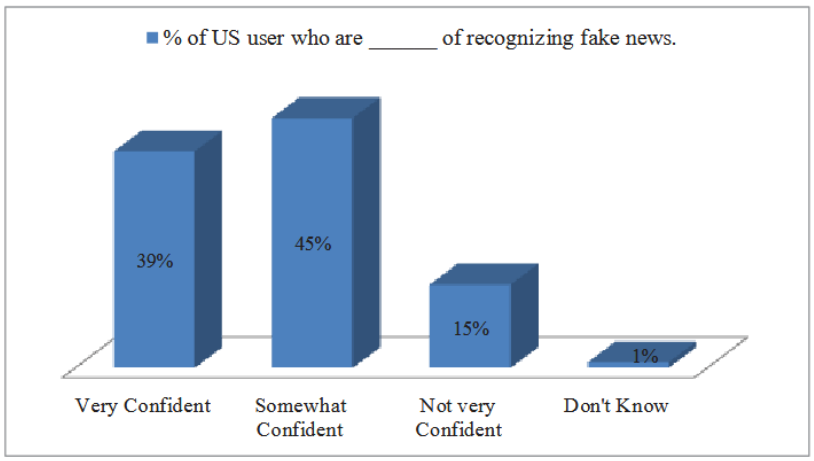
Το **YouTube**, αντιμετωπίζει τα fake news μέσω του αλγορίθμου προώθησης περιεχομένου. Ο αλγόριθμος αυτός, φροντίζει να μην υπάρχουν στα προτεινόμενα βίντεο επιλογές που έχουν αναφερθεί ως ψευδείς ή παραπλανητικές. Επιπλέον, συνεργάζεται με οργανισμούς ελέγχου γεγονότων (fact-checking) όπου φροντίζουν να προστίθενται προειδοποιήσεις στα βίντεο με σκοπό την εύκολη αναγνώριση από τον χρήστη. Υπάρχουν και περιπτώσεις που αφορούν διάδοση θεωριών συνωμοσίας ή επικίνδυνων ιατρικών συμβουλών, με την πλατφόρμα αφαιρεί το βίντεο και να τιμωρεί τους δημιουργούς με αποκλεισμό. **YouTube Policies**

3) WhatsApp Policie

Η **WhatsApp**, είναι υπηρεσία ανταλλαγής γραπτών και φωνητικών μηνυμάτων και όχι κοινωνικό δίκτυο με συνεχή ροή δημοσιεύσεων. Παρόλα αυτά, αναγνωρίζει τον κίνδυνο της διάδοσης των fake news. Συγκεκριμένα, έχει μειώσει τον αριθμό προώθησης μηνυμάτων, ώστε να αποτρέψει τη μαζική διανομή παραπλανητικών πληροφοριών, ενώ τα "forwarded messages" που προωθούνται από άγνωστους χρήστες επισημαίνονται ως ύποπτα. Τέλος, η πλατφόρμα έχει ξεκινήσει ενημερωτικές καμπάνιες για τους χρήστες, οι οποίες σχετίζονται με τη σημασία της επαλήθευσης των ειδήσεων πριν από την κοινοποίησή τους. **WhatsApp Policie**

4) TikTok Policie

Το **TikTok** αντιμετωπίζει την παραπληροφόρηση μέσω αφαίρεσης τους ψευδούς περιεχομένου από τις ροές των χρηστών. Συνεργάζεται με περισσότερους από 15 οργανισμούς ελέγχου γεγονότων όπως π.χ. το Reuters και το Politifact. Επισυνάπτει το AI-generated content και υποστηρίζει την εκπαίδευση των χρηστών κατά της παραπληροφόρησης, μέσω συνεργασιών με οργανισμούς όπως η National Association for Media Literacy Education (NAMLE). Τέλος, παρέχει την δυνατότητα στους χρήστες να αναφέρουν ως ψευδές το περιεχόμενο που παρακολούθησαν. **TikTok Harmful Misinformation Policie**



Εικόνα 7 : Ποσοστό αυτοπεποίθησης χρηστών να αναγνωρίσουν τις ψευδείς ειδήσεις [15]

2.4.2) Ψευδείς Ειδήσεις στην Πολιτική

Τα Fake News έχουν αποδειχθεί ιδιαίτερα επιζήμια στην πολιτική. Ο βασικότερος λόγος είναι ότι μπορούν να χειραγωγήσουν τον τρόπο που οι πολίτες ψηφίζουν, με αποτέλεσμα να υπονομευθεί ο θεσμός της δημοκρατίας. Οι ψευδείς ειδήσεις πολιτικού είδους, έχουν χρησιμοποιηθεί με σκοπό να κλονίσουν την κοινή γνώμη, να επηρεάσουν εκλογές και να πλήξουν την αξιοπιστία σε πολιτικά πρόσωπα.[2]

Το φαινόμενο αυτό δεν είναι καινούργιο. Η παραπληροφόρηση υπήρχε πάντα, αλλά με την ταχύτητα του διαδικτύου και των social media η εξάπλωσή της είναι πια ανεξέλεγκτη. Σε προεκλογικές περιόδους εμφανίζονται ειδήσεις που υποστηρίζουν ότι ένας υποψήφιος εμπλέκεται σε σκάνδαλα ή ότι μια κυβέρνηση σχεδιάζει μέτρα που δεν ισχύουν. Μέχρι να διαψευστούν, η ζημιά έχει ήδη γίνει. Ο κόσμος συχνά θυμάται την πρώτη αίσθηση που του προκάλεσε μια είδηση παρά την ίδια αλήθεια.[1], [4]

Γι' αυτό, είναι κρίσιμη η ανίχνευση και η αντιμετώπιση των ψευδών ειδήσεων στον πολιτικό χώρο. Άλλωστε, όπως αναφέρθηκε και προηγουμένως η εκλογές στην Αμερική το 2016 ήταν από τους σημαντικότερους λόγους όπου η επιστημονική κοινότητα άρχισε να επενδύει και να αναπτύσσει ραγδαία τις τεχνολογίες ανίχνευσης ψευδών ειδήσεων.[8]

Παραδείγματα

1) Brexit – Ηνωμένο Βασίλειο (2016)

Το 2016, κυκλοφόρησαν πολιτικές καμπάνιες εκστρατείας υπέρ του **Brexit**. Το αποτέλεσμα ήταν να διαδοθούν ψευδείς ειδήσεις όπως ότι η Ευρωπαϊκή Ένωση χρεώνει 350 εκατομμύρια λίρες την εβδομάδα το Ηνωμένο Βασίλειο υπερκοστολογόντας της παρουσία του. Η πληροφορία δεν επιβεβαιώνεται από καμία έγκυρη οικονομική πηγή. Επίσης, διαδόθηκαν και ψευδείς φήμες που ανέφεραν ότι σε περίπτωση ένταξης στην Ευρωπαϊκή ένωση θα πραγματοποιούταν μαζικές μεταναστεύσεις από την Τουρκία αλλά και από άλλες χώρες της Ανατολής. Το αποτέλεσμα ήταν να ενισχυθεί ο λαϊκισμός και το εθνικό φρόνιμα όπου επηρέασε την κοινή γνώμη υπέρ του Brexit, κάτι που απέφερε μακροχρόνιες πολιτικές και οικονομικές συνέπειες. **Πηγή 1ου παραδείγματος**

2) Μακεδονικό Ζήτημα – Ελλάδα (2018–2019)

Το 2018, κατά την περίοδο των πολιτικών συζητήσεων για την συμφωνία των Πρεσπών (μετονομασία της Π.Γ.Δ.Μ σε Βόρεια Μακεδονία) κυκλοφόρησαν ψευδείς πληροφορίες περί παραχώρησης εδαφών και μυστικής προδοσίας της κυβέρνησης. Οι ειδήσεις αυτές, προκάλεσαν τον κοινωνικό πανικό και τα νέα που κυκλοφορούσαν είχαν χρωματιστεί με εθνικισμό. Οι ειρηνικές διαδηλώσεις έχασαν τον εθνισμό τους που παρερμηνεύτηκε σε εθνικισμό και κάθε δημόσια πολιτική τοποθέτηση εκλαμβανόταν καχύποπτα από του πολίτες. **Πηγή 2ου παραδείγματος**

3) Εκλογές Βραζιλίας (2018)

Πριν την διαδικασία των εκλογών της Βραζιλίας το 2018, ο Ζαΐρ Μπολσονάρο και οι υποστηρικτές του, χρησιμοποίησαν το WhatsApp και το Facebook για να διαδώσουν βαρύτατους ψευδείς ισχυρισμούς περί διαφθοράς των πολιτικών αντιπάλων. Οι ψευδείς ισχυρισμοί αυτοί κλιμακώθηκαν όταν διαδόθηκε πως το Εργατικό κόμμα θα υποστήριζε την νομιμοποίηση αλλαγής φύλλου σε μικρή ηλικία. Η ανήθικη αυτή προπαγάνδα, ενίσχυσε τα αποτελέσματα υπέρ του Μπολσονάρο καθώς οι ισχυρισμοί προκάλεσαν χειραγώγηση και μεταστροφή της πολιτικής γνώμης των ψηφοφόρων άλλων κομμάτων. **Πηγή 3ου παραδείγματος**

2.4.3) Ψευδείς Ειδήσεις στην Οικονομία

Οι ψευδείς ειδήσεις βλάπτουν εξίσου σημαντικά τον τομέα της οικονομίας. Προκαλούν αναταράξεις στις αγορές, επηρεάζουν την εμπιστοσύνη καταναλωτών και επενδυτών και οδηγούν σε επιχειρηματικές αποφάσεις υπό το πρίσμα του πανικού[16]. Η ταχύτατη διάδοση τους μέσω των social media, έχει καταστήσει την οικονομία ευάλωτη στην παραπληροφόρηση. Οι ψευδείς πληροφορίες που σχετίζονται με οικονομικούς δείκτες, εταιρικές επιδόσεις ή κρίσεις των αγορών, οδηγούν τις αγορές σε πανικό προκαλώντας απότομες διακυμάνσεις στις τιμές των μετοχών και άλλων χρηματοοικονομικών προϊόντων όπως τα κρυπτονομίσματα. [16]

Παραδείγματα

1) Tweet για "έκρηξη στον Λευκό Οίκο" (2013)

Το 2013, έγινε κατάχρηση του χακαρισμένου λογαριασμού της Associated Press στο Twitter. Οι επιτήδειοι, προχώρησαν στην δημοσίευση ενός tweet που αποκάλυπτε την υποτιθέμενη έκρηξη στον Λευκό οίκο, αλλά και τον τραυματισμό του προέδρου Μπαράκ Ομπάμα. Η είδηση προκάλεσε πτώση 1% του δείκτη S&P 500 (500 κορυφαίες μετοχές της Αμερικής) δηλαδή πτώση 136 δισεκατομμυρίων της υποκειμενικής αξίας. Όταν αποκαταστάθηκε η αλήθεια , ο δείκτης επανήλθε στην πραγματική του τιμή. **Πηγή 1ου παραδείγματος**

2) Fake news για συνεργασία Litecoin με Walmart (2021)

Το 2021, κυκλοφόρησαν ψευδείς ειδήσεις που υποστήριζαν ότι η κολοσσιαία εταιρία Walmart θα δεχόταν το κρυπτονόμισμα **LiteCoin** ως μια διαθέσιμη στο κοινό μέθοδο πληρωμής. Το LTC ανέβηκε κατά 20+% μέσα σε λίγες ώρες, ενώ με την αποκατάσταση της πραγματικότητας το νόμισμα κατέρρευσε άμεσα σε τιμές μικρότερες και από την αρχική του αξία. **Πηγή 2ου παραδείγματος**

3) Ψευδείς φήμες για θάνατο του Vitalik Buterin (2017)

Το 2017 κυκλοφόρησαν ψευδείς ειδήσεις στο Reddit και το 4chan όπου ανέφεραν τον υποτιθέμενο θάνατο του Vitalik Buterin σε τροχαίο. Ο Buterin είναι γνωστός ως ο εφευρέτης του κρυπτονομίσματος Etherium. Οι δημοσιεύσεις προκάλεσαν άμεση πτώση 20% του Etherium (4 δισεκατομμύρια πτώση της συνολικής αγοράς). Η μερική αποκατάσταση απήλθε από τον ίδιο τον Buterin, ο οποίος διέψευδε τις ειδήσεις με δημοσίευση φωτογραφίας του ίδιου στην τότε σημερινή ημερομηνία και ώρα. **Πηγή 3ου παραδείγματος**

4) COVID-19 και πανικός στις αγορές (2020)

Το 2020, κατά το πρώτο τρίμηνο της πανδημίας κυκλοφόρησαν ειδήσεις που ανέφεραν ότι ο ιός είναι τεχνητός και τα εμβόλια που θα δημιουργηθούν θα έχουν ως σκοπό την εξόντωση της ανθρωπότητας. Σίγουρα, ειδήσεις τέτοιου είδους δεν είναι εύκολα πιστευτές αλλά σε περιόδους υγειονομικών κρίσεων και με τον κόσμο να αναζητά πληροφορίες μέσα στην αβεβαιότητα του μέλλοντος, δημιουργείται χώρος για διάφορες θεωρίες συνομωσίας. Το αποτέλεσμα των συνομωσιών ήταν η πρόκληση ακραίας μεταβλητότητας σε μετοχές αλλά και ανοδική εκτόξευση των τιμών των κρυπτονομισματών. **Πηγή 4ου παραδείγματος**

5) Elon Musk και Dogecoin / Bitcoin (2021–2022)

Κατά την διάρκεια των ετών 2021 και 2022 υπήρχε μια τάση να δημοσιεύονται παραπλανητικά tweets που αφορούσαν το memecoin Doge. Γινόντουσαν αναφορές σε άλλα πετυχημένα memes παρουσιάζοντας τα ως νέα τάση με αποτέλεσμα το DogeCoin να ανεβοκατεβαίνει κατά 20% για αρκετές συναπτές ημέρες. Επίσης, υπήρχαν και παραπλανητικά tweets που ανέφεραν ότι η εταιρία του Elon Musk Tesla, δεν θα δεχόταν το BitCoin ως τρόπο πληρωμής, κάτι που προφανώς επηρέασε την μεταβλητότητα του αλλά και την αξία του. **Πηγή 5ου παραδείγματος**

2.4.4) Ψευδής Ειδήσεις στον Αθλητισμό

Ένας ακόμα κλάδος που εμφανίζονται τα Fake News είναι ο αθλητικός ειδησεογραφικός τομέας. Μια ψευδή είδηση που αφορά τον αθλητισμό, είναι ικανή να χειραγωγήσει τις απόψεις φιλάθλων και οπαδών. Το αποτέλεσμα είναι να επηρεαστούν αρνητικά ομάδες, παίκτες, διοικήσεις αλλά και οι οικογένειες των προαναφερθέντων. Ειδήσεις τέτοιου τύπου συνήθως αφορούν ψευδείς μεταγραφές, σκάνδαλα, τραυματισμούς, ακόμα και στημένα παιχνίδια. Είναι αρκετά συνηθισμένο φαινόμενο τα αθλητικά ειδησεογραφικά sites να διαδίδουν ανακριβείς πληροφορίες χωρίς να είναι διασταυρωμένες ή τεκμηριωμένες και να χρησιμοποιούν clickbait τίτλους[17]. Όλα τα παραπάνω, είναι ικανά να οδηγήσουν σε μειωμένη απόδοση των αθλητών και συνολικά των ομάδων τους, να επηρεάσουν στοιχηματικές αγορές αλλά και να αμαυρώσουν την φήμη των επαγγελματιών. Εξίσου σημαντική, είναι η αναφορά των μηχανισμών προπαγάνδας που συντηρούν ισχυρές ομάδες με σκοπό την χειραγώγηση απόψεων και την παραποίηση των καταστάσεων. Οι δράσεις αυτές, πιθανόν να οδηγήσουν σε πρόκλησης βίας αλλά και ‘δολοφονία’ χαρακτήρων αθλητών και παραγόντων όπου οι δηλώσεις τους και οι πράξεις τους μετασχηματίζονται ανάλογα με τα συμφέροντα που εξυπηρετεί κάθε αρθρογράφος. Μέσα στο πέρασμα των χρόνων και με δεδομένο τον φανατισμό που επικρατεί στις τάξεις του αθλητισμού, επιτήδειοι βρίσκουν πάτημα να διαδώσουν

ψευδείς αφηγήματα με σκοπό να βλάψουν προσωπικότητες παιχτών, προέδρων , προπονητών ακόμα και συνδέσμων φιλάθλων.

Παραδείγματα

1) Barcelona Football Club - I3 Ventures(2020)

Το 2020, αποκαλύφθηκε ότι η διοίκηση της FC Barcelona είχε προσλάβει την εταιρεία I3 Ventures για να διαχειρίζεται τους λογαριασμούς της ομάδας στα μέσα κοινωνικής δικτύωσης. Ο σκοπός τους, ήταν να προωθήσουν θετικά την εικόνα του προέδρου Josep Maria Bartomeu και ταυτόχρονα να δυσφημήσουν πρώην και νυν ποδοσφαιριστές του συλλόγου, όπως οι Lionel Messi, Gerard Piqué και Xavi. Η υπόθεση αυτή, γνωστή ως "Barçagate", οδήγησε σε σύλληψη και παραίτηση του Bartomeu με τις αντιδράσεις να είναι έντονες στην ποδοσφαιρική κοινότητα. **Πηγή 1ου παραδείγματος**

2) Lucas Podolski και Breitbart (2017)

Το 2017, δημοσιεύτηκε στην ειδησεογραφική σελίδα Breitbart ένα ψευδές δημοσίευμα που αφορούσε την παράνομη μετακίνηση μεταναστών μέσω χρήσης jet ski. Η δημοσίευση συνοδευόταν με μια φωτογραφία του διεθνή ποδοσφαιριστή της εθνικής Γερμανίας Lucas Podolski να οδηγεί ένα jet ski. Η φωτογραφία ήταν κλεμμένη από τους προσωπικούς λογαριασμούς του Podolski στα social media και δεν είχε καμία σχέση με την είδηση. Ο ποδοσφαιριστής κινήθηκε νομικά για να αποκαταστήσει τις ψευδείς φήμες γύρω από το όνομα του. **Πηγή 2ου παραδείγματος**

3) Deflategate NFL (2015)

Το 2015, κατά την διάρκεια του τελικού της διοργάνωσης NFL, κυκλοφόρησαν δημοσιεύματα όπου κατηγορούσαν τους New England Patriots για παράβαση των κανονισμών. Κατηγορήθηκαν ότι χρησιμοποιούσαν μπάλες με λιγότερο αέρα από όσο επιτάσσουν οι κανονισμοί και ο σκοπός τους ήταν να αποκτήσουν τακτικό πλεονέκτημα έναντι των αντιπάλων τους. Η είδηση προκάλεσε φανατισμένες απόψεις στο φίλαθλο κοινό, ενώ αργότερα αποδείχθηκε ότι η NFL είχε παραποιήσει τα δεδομένα για να συντάξει τις ψευδείς κατηγορίες. **Πηγή 3ου παραδείγματος**

4) Γεωργιανοί ποδοσφαιριστές Euro (2024)

Κατά την διάρκεια του Euro 2024 που φιλοξενήθηκε στην χώρα της Γερμανίας, αρκετοί ποδοσφαιριστές την εθνικής ομάδας της Γεωργίας στοχοποιήθηκαν από ψευδή νέα. Αρχικά, ο ποδοσφαιριστής Georges Mikautadze κατηγορήθηκε ότι πανηγύρισε με σατανιστικό χαιρετισμό έπειτα από γκολ που πέτυχε, ενώ στην πραγματικότητα επρόκειτο για κοινή χειρονομία πανηγυρισμού. Η αποκατάσταση της αλήθειας προβλήθηκε άμεσα από τηλεοπτικά κανάλια καθώς ο ποδοσφαιριστής αποτελεί

πρότυπο για πολλούς μικρούς ηλικιακά ποδοσφαιρόφιλους από την Γεωργία αλλά και ολόκληρη την ποδοσφαιρική κοινότητα. Επίσης, υπήρξαν ψευδή tweets που υποστήριζαν ότι ο προπονητής της ομάδας Temur Ketsbaia είχε αποκλείσει τον ποδοσφαιριστή Budu Zivzivadze λόγω των πολιτικών του πεποιθήσεων. Η είδηση καταρρίφθηκε καθώς ο παίχτης αγωνίστηκε κανονικά στο τουρνουά με την εθνική του ομάδα. **Πηγή 4ου παραδείγματος**

2.4.5) Ψευδείς ειδήσεις στον Ιδιωτικό Τομέα- Βιομηχανίες

Ο ιδιωτικός τομέας και η βιομηχανίες απειλούνται σημαντικά από τα Fake News. Έχουν καταγραφεί γεγονότα όπου στοχευμένες ψευδείς εκστρατείες πλήττουν την φήμη, την οικονομική σταθερότητα και την εμπιστοσύνη των εταιριών. Η δημοσιοποίηση ψευδών πληροφοριών για υποτιθέμενες τεχνικές βλάβες των προϊόντων, προκαλούν άμεση πτώση των μετοχών, ακυρώσεις παραγγελιών, νομικές συνέπειες και πρόκληση αβεβαιότητας στους εργαζομένους της εκάστοτε εταιρίας. Επιπρόσθετα, τονίζεται η ανάγκη ύπαρξης πλάνων crisis managment αλλά και προληπτικά συστήματα ανίχνευσης για να μειώσουν τον κίνδυνο επιζήμιων καταστάσεων[16]. Τέλος, σε πολλές περιπτώσεις η αποκατάσταση των συνεπειών των ψευδών ειδήσεων αντιμετωπίζεται με εξαιρετικά κοστοβόρες διαφημιστικές καμπάνιες αποκατάστασης της αλήθειας.[16]

Παραδείγματα

1)Η περίπτωση της Coca-Cola και της Pepsi 2003

Το 2003, δημοσιεύτηκαν ψευδείς ειδήσεις στην Ινδία όπου ισχυριζόταν ότι τα προϊόντα της Coca-Cola και της Pepsi περιείχαν υψηλά επίπεδα φυτοφαρμάκων καθιστώντας τα επικίνδυνα για κατανάλωση. Το αφήγημα εξαπλώθηκε ταχύτατα προκαλώντας πανικό στους καταναλωτές και τους πωλητές των προϊόντων. Οι πωλήσεις των δύο εταιρειών μειώθηκαν έως και 40% στην ινδική αγορά ενώ ακόμα και το κοινοβούλιο της Ινδίας εξέτασε την απαγόρευση των προϊόντων. Οι εταιρείες αναγκάστηκαν να δαπανήσουν εκατομμύρια δολάρια σε διαφημιστικές καμπάνιες για να αποκαταστήσουν τη φήμη των προϊόντων τους. Οι δοκιμές απέδειξαν ότι τα επίπεδα των χημικών ουσιών ήταν εντός των επιτρεπτών ορίων αλλά η φήμη των εταιρειών υπέστη σοβαρό πλήγμα αποδεικνύοντας τη δύναμη των ψευδών ειδήσεων στην εμπορική επιτυχία μιας επιχείρησης. Πηγή 1ου παραδείγματος

2)Η περίπτωση της Tesla και του Έλον Μασκ 2020

Το 2020, κυκλοφόρησαν ψευδείς δημοσιεύματα στα μέσα κοινωνικής δικτύωσης όπου ισχυριζόταν ότι ο Έλον Μασκ, CEO της Tesla είχε παραιτηθεί από τα καθήκοντά του και ότι η εταιρεία αντιμετώπιζε σοβαρά οικονομικά προβλήματα. Η μετοχή της Tesla παρουσίασε πτώση έως και 10% με χιλιάδες επενδυτές να πανικοβάλλονται πουλώντας τις μετοχές τους. Η Tesla οδηγήθηκε σε επίσημες ανακοινώσεις για να διαψεύσει την είδηση. Το περιστατικό απέδειξε πως οι ψευδείς ειδήσεις μπορούν να έχουν άμεσο οικονομικό αντίκτυπο σε μια εταιρεία, ειδικά όταν αφορούν τα κορυφαία στελέχη της στην ιεραρχία. Η Tesla χρησιμοποίησε μέσα ανίχνευσης της παραπληροφόρησης για να αποτρέψει περαιτέρω ζημιά. Πηγή 2ου παραδείγματος

3) McDonald's Fake video με ανθρώπινο κρέας (2017)

Το 2017, ένα ψεύτικο βίντεο κυκλοφόρησε στα κοινωνικά δίκτυα, ισχυριζόμενο ότι η McDonald's χρησιμοποιούσε ανθρώπινο κρέας στα προϊόντα της. Παρόλο που η είδηση ήταν εντελώς ψευδής, το βίντεο έγινε viral, ζημιώνοντας την φήμη της εταιρείας. Προκλήθηκε μείωση πωλήσεων με την McDonald's να αναγκάζεται να απαντήσει επίσημα και να διαψεύσει τη φήμη. Πηγή 3ου παραδείγματος

4) Starbucks Fake καμπάνια για μετανάστες(2018)

Το 2018, μια ψεύτικη διαφημιστική καμπάνια κυκλοφόρησε στα social media. Ισχυριζόταν ότι η Starbucks προσέφερε δωρεάν καφέ σε παράνομους μετανάστες στις Η.Π.Α. κάτι που προκάλεσε οργισμένες αντιδράσεις με καταναλωτές να προσκαλούν τον κόσμο σε μποϊκοτάζ. Το hashtag γνωστό ως #BoycottStarbucks αποτέλεσε μόδα στο Twitter και τις προτεινόμενες αναζητήσεις με την εταιρία να αναγκάζεται να δαπανήσει εκατομμύρια δολάρια σε damage control και αποκατάσταση της αλήθειας. Πηγή 4ου παραδείγματος

5) Arla Foods και το πρόσθετο Bovaer (2024)

Το 2024, η γαλακτοκομική εταιρεία Arla Foods ανακοίνωσε τη δοκιμή του Bovaer, ενός συμπληρώματος ζωοτροφών που έχει την ικανότητα να μειώνει τις εκπομπές μεθανίου από τις αγελάδες αλλά και τα ζώα γενικότερα. Παρά την έγκριση από τις ευρωπαϊκές αρχές που φροντίζουν για την ασφάλεια των τροφίμων, κυκλοφόρησαν ψευδείς φήμες

στα μέσα κοινωνικής δικτύωσης κατηγορώντας την εταιρεία ότι προσπαθεί να δηλητηριάσει τους καταναλωτές. Αυτό, οδήγησε αρκετούς πελάτες να σταματήσουν να καταναλώνουν τα προϊόντα της Arla. Η εταιρεία τοποθετήθηκε με δηλώσεις και συνεντεύξεις στα μέσα μαζικής ενημέρωσης για να αντικρούσει τους ψευδείς ισχυρισμούς και να τονίσει την ασφάλεια του Bovaer με σκοπό την αποκατάσταση της φήμης της εταιρίας και των προϊόντων της. Πηγή 5ου παραδείγματος

2.4.6 Ψευδείς Ειδήσεις στον Υγειονομικό Τομέα

Η παραπληροφόρηση γύρω από θέματα υγείας αποτελεί μια επικίνδυνη πληγή για τη δημόσια υγεία. Σύμφωνα με τη μελέτη της Papanikou και των συνεργατών της (2025)[18], ψευδείς δημοσιεύματα που αφορούν εμβόλια ή θεραπείες του COVID-19 διαδίδονται ταχύτερα από την ιατρικά τεκμηριωμένη πληροφόρηση. Η παραδοσιακή χειροκίνητη επαλήθευση δεν δύναται να ανταποκριθεί στον όγκο και την ταχύτητα της διασποράς με αποτέλεσμα να διαμοιράζονται πληροφορίες επικίνδυνες για την σωματική και ψυχική υγεία των πολιτών. Η κατάσταση αυτή, τονίζει την ανάγκη η επιστημονική κοινότητα να στραφεί στην ανάπτυξη αυτοματοποιημένων εργαλείων, ικανά για γρήγορη ανίχνευση ή περιορισμό της είδησης.[18]

Παραδείγματα

1)Θεραπείες κατά του καρκίνου 2024

Το 2024, χρήστες του TikTok δημοσίευσαν βίντεο που σχετιζόντουσαν με υποτιθέμενες θεραπείες σοβαρών ασθενειών όπως ο καρκίνος. Η Belle Gibson ισχυριζόταν ψευδώς ότι θεραπεύτηκε από τον καρκίνο μέσω της διατροφής που ακολουθούσε εξαπατώντας και θέτοντας σε κίνδυνο χιλιάδες ασθενείς. Αποκόμισε τεράστια κέρδη από την σειρά των βίντεο με το γεγονός αυτό να τονίζει την επίδραση των ψευδών ειδήσεων σε θέματα υγείας. Πηγή 1ου παραδείγματος

2) Θεραπείες κατά του COVID-19 Εποχή του Covid

Κατά την διάρκεια της πανδημίας, κυκλοφόρησαν στα μέσα κοινωνικής δικτύωσης υποτιθέμενες θεραπείες κατά του Covid. Μια εξ αυτών, πρότεινε την κατανάλωση μεθανόλης. Το αποτέλεσμα ήταν οι ανυποψίαστοι πολίτες να οδηγηθούν σε δηλητηριάσεις αλλά ακόμα και σε θανάτους. Μέσω του γεγονότος αυτού, αναδείχθηκε για άλλη μια φορά η επιτακτική ανάγκη ανάπτυξης ή παραμετροποίησης των αλγορίθμων προώθησης υλικού με σκοπό την ασφάλεια των χρηστών. Πηγή 2ου παραδείγματος

3) Paloma Shemirani 2023

Το 2023, η άτυχη απόφοιτος του Cambridge Paloma Shemirani διαγνώστηκε με λέμφωμα. Η επιστήμη θεωρεί ότι τα ποσοστά ίασης αγγίζουν το 80% αν ο ασθενής υποβληθεί σε χημειοθεραπείες. Η μητέρα της Paloma, πρώην νοσηλεύτρια, έχοντας ήδη αποβληθεί και διαγραφεί από το νοσηλευτικό μητρώο με κατηγορίες για επικίνδυνες νοσηλευτικές πρακτικές, διαμοιρασμό θεωριών συνομωσίας κατά των εμβολίων αλλά και γενικότερα την προώθηση παραϊατρικών προσεγγίσεων ίασης, δεν της επέτρεψε να ακολουθήσει την θεραπεία όπου η επιστήμη της ιατρικής πρότεινε. Το τραγικό αυτό γεγονός, οδήγησε στον θάνατο της Paloma από καρδιακή προσβολή αναδεικνύοντας την επικινδυνότητα των Fake News ακόμα και εντός της οικογενείας. Πηγή 3ου Παραδείγματος

4) PCOS 2024

Το σύνδρομο πολυκυστικών ωοθηκών (PCOS) εμφανίζεται κατά μέσω όρο σε μια στις δέκα γυναίκες. Προκαλεί συμπτώματα όπως η ακμή, αστάθεια του κύκλου της περιόδου και αυξημένη τριχοφυΐα. Στα μέσα κοινωνικής δικτύωσης προωθείται πληθώρα ψευδών πληροφοριών από ανειδίκευτους και επιτήδειους χρήστες που συμβουλεύουν τις γυναίκες να ακολουθήσουν υποτιθέμενες προτεινόμενες δίαιτες και κατανάλωση αμφιβόλων συμπληρωμάτων διατροφής για την καταπολέμηση του. Οι συμβουλές αυτές, θεωρούνται ιατρικά ανεπαρκείς και αντιφατικές με την μέχρι τώρα ιατρική έρευνα, μετατρέποντας τις ανυποψίαστες διαγνωσμένες γυναίκες σε πειραματόζωα επικίνδυνων παραϊατρικών πρακτικών μεθόδων. Πηγή 4ου παραδείγματος

5) Επιδημία ιλαράς στη Σαμόα 2019

Το 2019, η αρρώστια της ιλαράς είχε πλήξει καταστροφικά την Σαμόα με 83 καταγεγραμμένους θανάτους. Η επιδημία αφορούσε κυρίως παιδιά, με την κρίση να επιδεινώνεται έπειτα από την επίσκεψη του Robert Kennedy Jr. Ο Kennedy γνωστός για τις αντιεμβολιαστικές του απόψεις, κλόνισε την εμπιστοσύνη των πολιτών της Σαμόα χρησιμοποιώντας το παράδειγμα του θανάτου 2 βρεφών η οποία όμως είχε προκληθεί από λανθασμένη ποσοτική χρήση του εμβολίου την οποία απέκρυψε. Οι καταστροφικές συνέπειες που προκάλεσε, αφορούσαν την καταστολή του εμβολιαστικού προγράμματος για 10 μήνες με την επιδημία να εξαπλώνεται μαζικά. Πηγή 5ου παραδείγματος

3)Έρευνα και Μέθοδοι Ανίχνευσης Fake News

3.1)Έρευνα πάνω στα Fake News

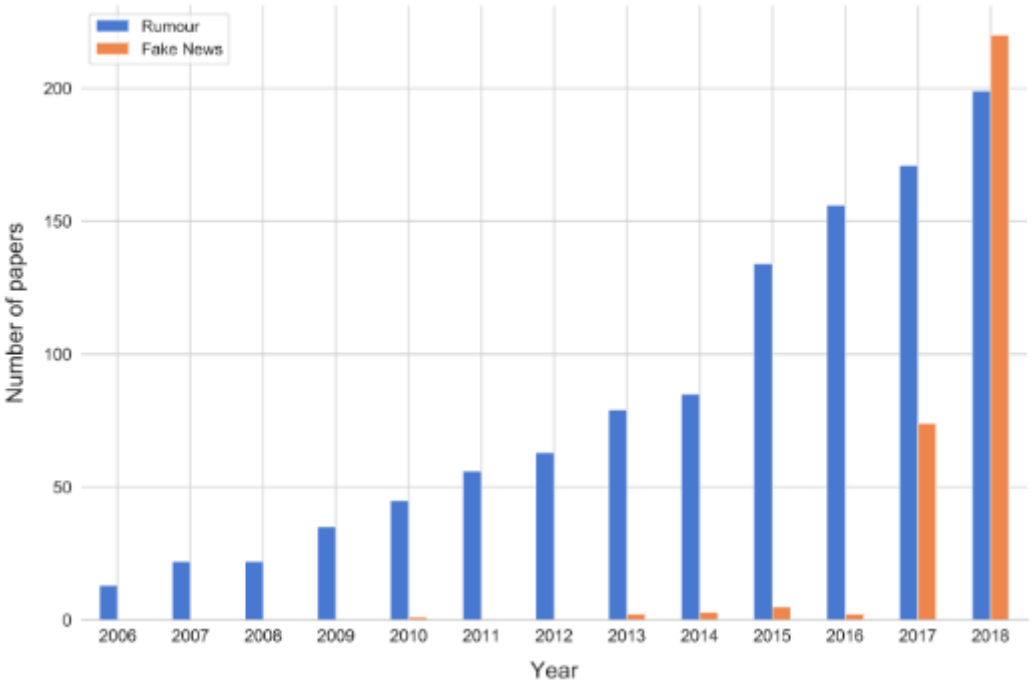
Η ακαδημαϊκή κοινότητα, έπειτα από την αναγνώριση του προβλήματος των **Fake News** που εδραιώθηκε το **2016** κατά την διάρκεια των εκλογών των Η.Π.Α.[2] άρχισε να διαθέτει πόρους και έρευνες στην δημιουργία συστημάτων ανίχνευσης των ψευδών ειδήσεων. Οι προσεγγίσεις τους, για να δομηθούν κατάλληλα, έπρεπε να εξετάσουν ενδελεχώς τα εξής 3 πεδία.

Διασπορά : Πραγματοποιήθηκαν έρευνες και μελέτες όπου εξέταζαν τους τρόπους που οι ψευδείς ειδήσεις διαμοιραζόταν μέσω των social media και τον ειδησεογραφικών σελιδών '**Network-based Features**'[19]. Μελετήθηκε το κομμάτι των '**Temporal Characteristics of Propagation'**[19] όπου έδωσε την δυνατότητα εξαγωγής μοτίβων που αφορούσαν τους 'κόμβους' δηλαδή του χρήστες διάδοσης αλλά και τον χρόνο διάδοσης.

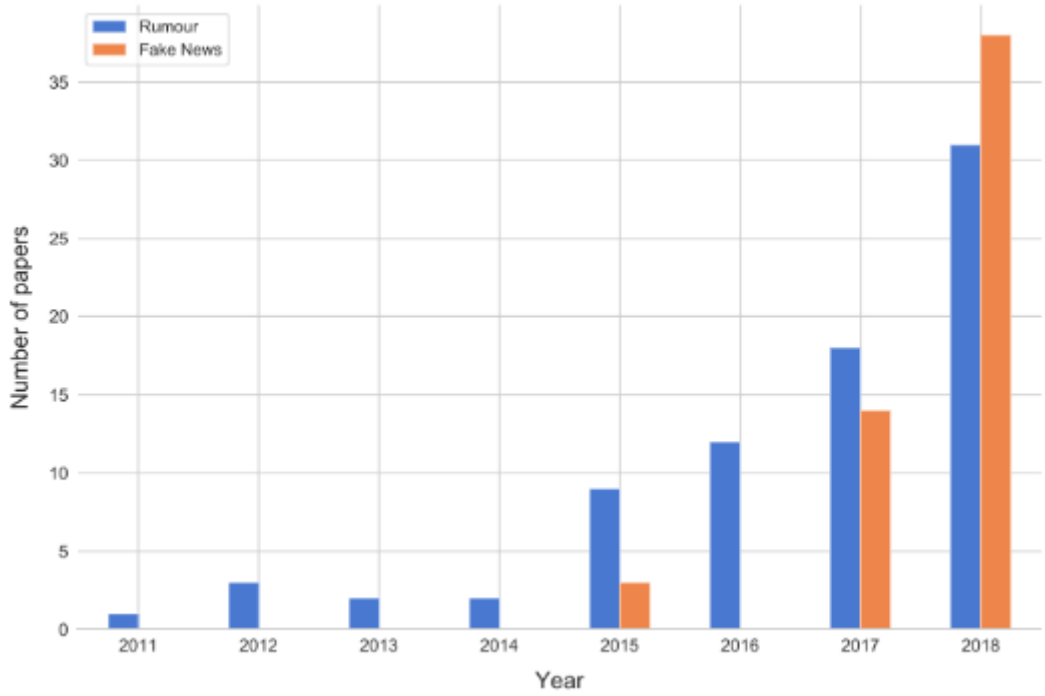
Αντίκτυπος : Μελέτησαν και αξιολόγησαν τις επιζήμιες δυνατότητες των Fake News σε ευάλωτες κοινωνικές ομάδες, σε ισχυρές κοινωνικές ομάδες αλλά και σε κάθε κλάδο όπως η πολίτική και η υγεία[10], [18]. Η αναγνώριση των πιθανών συνεπειών, ώθησε τους ερευνητές στο να καταλάβουν ποιοι κλάδοι έχουν την μεγαλύτερη ανάγκη για συστήματα ανίχνευσης όπου θα ειδικεύονται σε συγκεκριμένες θεματολογίες. [14]

Τεχνολογίες : Τα Fake News συνηθίζεται να σχεδιάζονται με τέτοιους τρόπους ώστε να μοιάζουν αξιόπιστα και να ξεγελούν τους αναγνώστες. Οι ερευνητές λοιπόν, κλήθηκαν να επιλέξουν αλλά και να προσαρμόσουν τις κατάλληλες υπάρχουσες τεχνολογίες ώστε να είναι εφικτό να δημιουργήσουν συστήματα ανίχνευσης υψηλής ακρίβειας. Με το πέρασμα των χρόνων και την εμφάνιση προηγμένων τεχνολογιών όπου θα μπορούσαν να χρησιμοποιηθούν και στον τομέα της ανίχνευσης, τα συστήματα ισχυροποιήθηκαν ακόμα περισσότερο έχοντας ως ακρογωνιαίους λίθους προσεγγίσεις που βασίζονται στο **Machine Learning, Natural Language Processing** και **Deep Learning** όπου περιέχει τα **Neural Networks** και τους **Transformers**.[7]

Παρακάτω, διακρίνουμε με την βοήθεια των γραφημάτων **των Alessandro Bondielli** και **Francesco Marcelloni** την τάση εμφάνισης ερευνών που αφορούν τις φήμες και τα Fake News(εικόνα 8)[19] όπου οι έρευνες που επικεντρώνονται στις φήμες έχουν καταγράψει ανοδική τάση από το 2006 ενώ τα Fake News αποκτούν σημαντική υπόσταση από το 2017. Το δεύτερο γράφημα περιγράφει την τάση δημοσιεύσεων ερευνών που αφορούν την ανίχνευση φημών και ψευδών ειδήσεων(εικόνα 9)[19]. Τα fake news papers εκτινάσσονται μετά τις εκλογές την Αμερικής το 2016 παράλληλα με τον τομέα ανίχνευσης των fake news.



Εικόνα 8 : Δημοσιευμένα papers που αφορούν Φήμες και Ψευδείς ειδήσεις[19]

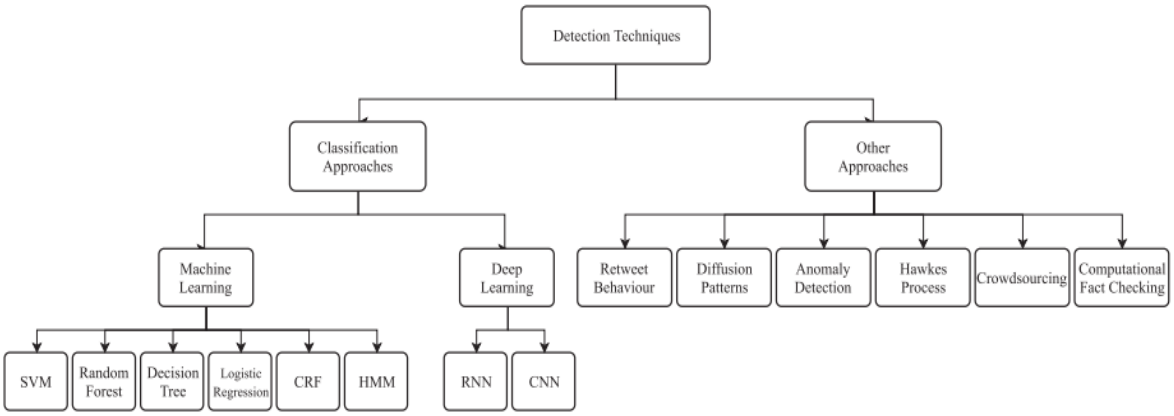


Εικόνα 9 : Δημοσιευμένα papers που αφορούν Ανίχνευση Φημών και Ψευδών ειδήσεων [19]

3.2)Μέθοδοι Ανίχνευσης

Τα έτη από το 2016 μέχρι το 2019 αποτελούν την εκκίνηση του πεδίου ανίχνευσης ψευδών ειδήσεων[19]. Οι τότε προσεγγίσεις, αποτέλεσαν παρακαταθήκη για τους επόμενους ερευνητές όπου στην συνέχεια εξελίχθηκαν αλλά και συνεχίζουν να εξελίσσονται χάρη στην τεχνητή νοημοσύνη.

Παρακάτω το γράφημα αναφέρει τις τεχνικές ανίχνευσης το 2019.



Εικόνα 10 : Τεχνικές ανίχνευσης μέχρι το 2019 [19]

Οι μέθοδοι ανίχνευσης ψευδών ειδήσεων εξελίσσονται ραγδαία περιλαμβάνοντας τεχνολογίες αιχμής και πολυδιάστατες προσεγγίσεις:

Επεξεργασία Φυσικής Γλώσσας (NLP): Το **NLP** αποτελεί την χρησιμότερη τεχνολογία στον τομέα ανίχνευσης των ψευδών ειδήσεων καθώς από την φύση του το πρόβλημα της text-based ανίχνευσης είναι NLP task[20]. Σε κάθε πείραμα, οποιαδήποτε και αν είναι η κατηγορία της επιλεγμένης τεχνολογίας όπως ML,DL,Transformer, το NLP θα εμφανιστεί στην προ επεξεργασία / καθαρισμό των δεδομένων μάθησης με τις δεκάδες παραδοσιακές NLP τεχνικές που υπάρχουν. Επίσης, σε πιο προηγμένα συστήματα ανίχνευσης βλέπουμε πιο εξελιγμένες μορφές NLP ικανοτήτων όπως η semantic analysis και η text analysis. Τέλος, το NLP υπάρχει και σε μορφές προ εκπαίδευσης όπως τις περιπτώσεις του μετασχηματιστή BERT.

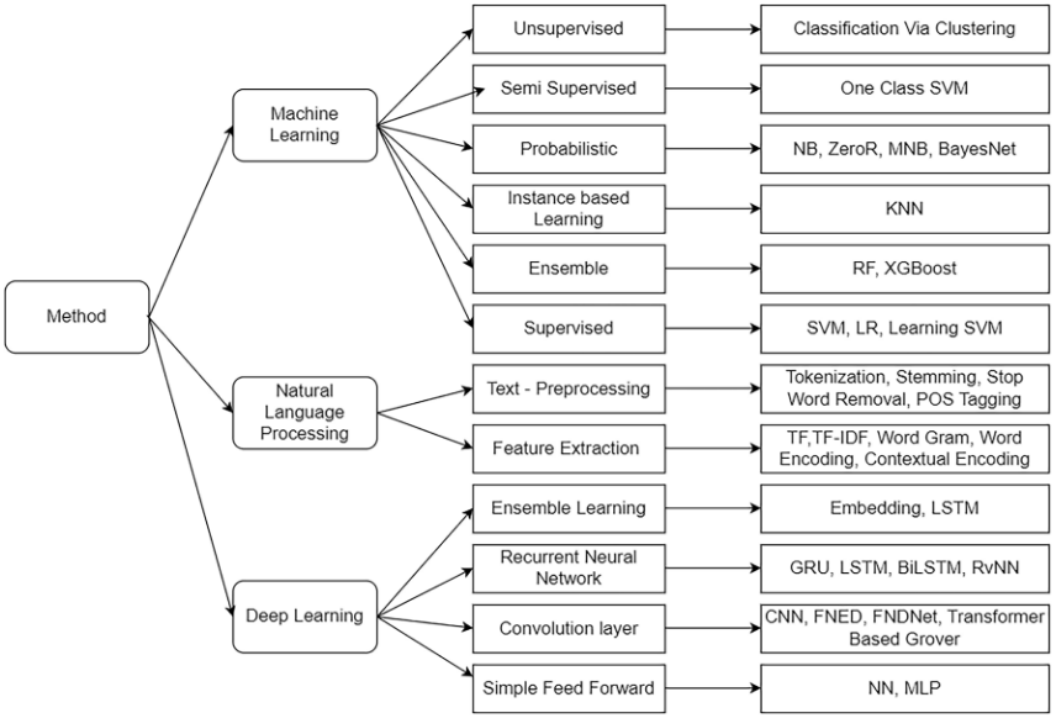
Μηχανική Μάθηση (Machine Learning): Ο τομέας της **μηχανικής μάθησης** αξιοποιήθηκε και συνεχίζει να αξιοποιείται από τους ερευνητές. Οι παραδοσιακοί αλγόριθμοι ταξινόμησης (LR,SVM,DT,RF) , έχουν χρησιμοποιηθεί σε αρκετά συστήματα που πέτυχαν υψηλές αποδόσεις[21] αλλά τεχνολογικά τείνουν να ξεπεραστούν. Τα συστήματα αυτά, έχουν υψηλές ανάγκες καθαρισμού των δεδομένων. Επιπλέον, έχουν ανάγκη από μεθόδους μετατροπής των δεδομένων σε αριθμητική μορφή με τεχνικές όπως TF-IDF, BoW[22]. Τέλος, συνηθίζεται να έχουν την ανάγκη από χειροκίνητα

εξαγόμενα χαρακτηριστικά από τα δεδομένα[23] όπου είναι μια χρονοβόρα διαδικασία για τους ερευνητές[23]. Από όλες τις τεχνολογίες που συναντήσαμε στα 53 πειράματα, τα ML συστήματα ήταν αυτά που δεν είχαν ικανότητες για να κατανοήσουν τα συμφραζόμενα και το βαθύτερο νόημα του κειμένου[23].

Βαθιά Μάθηση (Deep Learning): Η **Βαθιά Μάθηση** προσφέρει περισσότερες ικανότητες στα συστήματα ανίχνευσης ψευδών ειδήσεων καθώς μέσω τον Νευρωνικών Δικτύων τα συστήματα μπορούν να αξιοποιήσουν μεγαλύτερες ποσότητες δεδομένων, να εξάγουν και να μαθαίνουν αυτόματα τα καταλληλότερα χαρακτηριστικά και μοτίβα από τα δεδομένα[24] αλλά και να πετυχαίνουν υψηλές αξιολογήσεις ανίχνευσης. Επίσης, είναι ικανά να αποκτούν NLP ικανότητες μέσω της χρήσης word embeddings (Glove, Word2Vec) , καθώς αντιλαμβάνονται τα συμραζόμενα, το βαθύτερο νόημα , την γραμματική και το συντακτικό του κειμένου[11], [24].

Μετασχηματιστές (Transformers) : Μια από τις πιο σύγχρονες προσεγγίσεις είναι οι **Transformers** καθώς θεωρούνται και αυτά Νευρωνικά Δίκτυα αλλά πιο εξελιγμένα. Η εξέλιξη τους έγκειται στην ικανότητα να διαχειρίζονται παράλληλα σχέσεις του κειμένου μέσω των μηχανισμών προσοχής που διαθέτουν[24]. Έχουν υψηλές ικανότητες να κατανοούν συμφραζόμενα και σημασιολογικές σχέσεις μέσω των contextual embeddings[25]. Επίσης, είναι pre-trained μοντέλα με NLP tasks και κατά την χρησιμοποίησης τους χρειάζονται fine-tuning πάνω στο εκάστοτε datashet που χρησιμοποιείται στο πείραμα[25], η πιο συχνή μορφή που συναντήσαμε ήταν το BERT.

Συνδυαστική Ανίχνευση: Είναι εύκολα κατανοητό πως οι τεχνολογίες αυτές μπορούν να συνδυαστούν. Αρκετοί ακαδημαϊκοί ερευνητές έχουν υλοποιήσει προσεγγίσεις που συνδυάζουν τις τεχνολογίες με σκοπό να βρουν την καταλληλότερη για το εκάστοτε σύνολο δεδομένων και θεματολογία που πρόκειται να ερευνήσουν.



Εικόνα 11 : Κατηγορίες και υποκατηγορίες μεθόδων ανίχνευσης [26]

3.3) Ο Ρόλος της Επεξεργασίας Φυσικής Γλώσσας στην Ανίχνευση

Η **Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP)** παρέχει μια πληθώρα τεχνικών για την ποιοτική ανάλυση και επεξεργασία των δεδομένων εκπαίδευσης.

Βέβαια, είναι απαραίτητο να επαναλάβουμε πως το NLP έχει την ανάγκη να συνδυαστεί και με άλλες τεχνολογίες όπως το machine learning και το deep learning για να δομήσει συστήματα ανίχνευσης ψευδών ειδήσεων[20]. Οι προσεγγίσεις που βασίζονται στο NLP χαρακτηρίζονται από αυτοματοποιημένες διαδικασίες που συντελούν στην ταχύτητα επεξεργασίας των δεδομένων στα συστήματα ανίχνευσης ψευδών ειδήσεων[20] καθώς ο όγκος περιεχομένου στο διαδίκτυο και η αδυναμία των χρηστών των κοινωνικών δικτύων να αναγνωρίσουν τις ψευδής ειδήσεις έχουν δημιουργήσει την επιτακτική αυτή ανάγκη.

Το NLP ουσιαστικά είναι υπεύθυνο για την επεξεργασία των δεδομένων κειμένου που πρόκειται να χρησιμοποιηθούν σε πειράματα που αφορούν την ανίχνευση ψευδών ειδήσεων. Οι πληθώρα των τεχνικών αυτών χωρίζεται σε τέσσερις κατηγορίες όπου αφορούν την **επεξεργασία κειμένου** (Text Preprocessing), την **ανάλυση κειμένου** (Text

Analysis), την **ανάλυση σημασίας** (Semantic Analysis) **και την αναπαράσταση κειμένου** (Vectorization)[27].

1)Αρχικά, η **προεπεξεργασία** των δεδομένων αποτελεί το πρώτο θεμελιώδες στάδιο επεξεργασίας μέσω Φυσικής Επεξεργασίας της Γλώσσας (NLP). Μετατρέπει το ακατέργαστο κείμενο σε δομημένη μορφή χωρίς ανωμαλίες. Το στάδιο αυτό περιλαμβάνει μια σειρά από τεχνικές που στοχεύουν στη μείωση της περιττής πληροφορίας και του θορύβου των δεδομένων με σκοπό τη βελτιστοποίηση της ποιότητας του κειμένου[20]. Παρακάτω, επεξηγούνται οι πιο συνηθισμένες μορφές προ επεξεργασίας που υπήρχαν στα 53 πειράματα.

1.1)Tokenization, είναι η διαδικασία τεμαχισμού των λέξεων και φράσεων σε tokens, διευκολύνοντας την ατομική ανάλυση κάθε στοιχείου. [27]

1.2)Stopword Removal, είναι η διαδικασία αφαίρεσης λέξεων όπως "και", "είναι", οι οποίες εμφανίζονται με μεγάλη συχνότητα στο κείμενο αλλά δεν προσφέρουν ουσιαστική πληροφορία. [27]

1.3)Lowercasing, είναι η διαδικασία αφαίρεσης μετατροπής όλων των χαρακτήρων σε πεζούς. Με την τεχνική αυτή, δίνεται μια βοήθεια στο σύστημα να μην καταλαβαίνει την μια λέξη ως δύο διαφορετικές λόγω της διαφορετικής μορφής (π.χ. "Εργασία" και "εργασία").[23]

1.4)Stemming, αναγνωρίζει και επιστρέφει την ρίζα της λέξης μέσω αφαίρεσης της κατάληξης π.χ. "τρέχοντας", "τρέχει", "έτρεχε" → "τρέχ". [23]

1.5)Lemmatization πραγματοποιεί την μετατροπή της λέξης στην αρχική της μορφή (λήμμα) π.χ. "έτρεχε" → "τρέχω".[28]

Η αποτελεσματική εφαρμογή όλων αυτών των σταδίων προ επεξεργασίας είναι απαραίτητη ώστε το κείμενο να αποκτήσει καθαρή μορφή, χωρίς θόρυβο ώστε να είναι ιδανική για περαιτέρω ανάλυση.

2) Η **ανάλυση κειμένου**, αφορά την ικανότητα της σε βάθος ανάλυσης όπου αναγνωρίζεται η γραμματική και το συντακτικό. Ο σκοπός του σταδίου αυτού είναι να αναγνωριστούν οι εσωτερικές σχέσεις του κειμένου και οι διασυνδέσεις των λέξεων όπου κάθε λέξη λαμβάνει κάποιο ρόλο στην εκάστοτε πρόταση[28]. Αυτή η NLP ικανότητα αποκτάται από τα συστήματα είτε με χειροκίνητα εξαγόμενα χαρακτηριστικά από τους ερευνητές είτε μέσω των word και contextual embeddings.

2.1) Syntactic Parsing, αναλαμβάνει την αναγνώριση της γραμματικής και της συντακτικής δομής των προτάσεων. Επίσης, δημιουργεί συντακτικό δέντρο της πρότασης το οποίο περιέχει την γραμματική τους λειτουργία. [28]

2.2) Part-of-Speech Tagging (POS Tagging) αναγνωρίζει και προσδίδει σε κάθε λέξη το γραμματικό της ρόλο μέσα σε μια πρόταση όπως υποκείμενο, ρήμα, αντικείμενο, επίθετο και επίρρημα. [28]

2.3) Dependency Parsing είναι η τεχνική κατά την οποία το σύστημα μαθαίνει την αλληλεξάρτηση των λέξεων μέσα σε μια πρόταση αλλά και την γραμματική τους υπόσταση π.χ. το ρήμα 'παίζει΄ εξαρτάται από το αντικείμενο 'κιθάρα' και ο υποκείμενο 'ο Γιάννης'. [28]

2.4)Named Entity Recognition (NER). Πραγματοποιεί τον εντοπισμό και την κατηγοριοποίηση των οντοτήτων. Ως οντότητες, χαρακτηρίζονται τα ονόματα προσώπων, τοποθεσιών και ημερομηνιών. Η αναγνώριση τους θεωρείται σημαντικό κομμάτι στην ανίχνευση ψευδών ειδήσεων καθώς οι οντότητες μπορούν να διασταυρωθούν μέσω αξιόπιστων πηγών. [28]

Συμπερασματικά, η ανάλυση κειμένου επιτρέπει στα NLP συστήματα να κατανοούν τη γλωσσική δομή(γραμματική & συντακτικό) και την αλληλεξάρτηση των λέξεων σε βαθύτερο επίπεδο.

3) Η **σημασιολογική ανάλυση** είναι και αυτή NLP ικανότητα που αποκτάται είτε με χειροκίνητα εξαγόμενα χαρακτηριστικά από τους ερευνητές είτε μέσω των word και contextual embeddings. Αυτού του είδους η ανάλυση, είναι υπεύθυνη να μάθει στο σύστημα το νόημα του κειμένου, τα συμφραζόμενα αλλά και να αναλύσει τα συναισθήματα των κειμένων όπως μια καταφατική & αρνητική πρόταση.[29]

3.1) Sentiment Analysis προσφέρει την δυνατότητα κατανόησης του συναισθήματος όπως καταφατικός, αρνητικός αλλά και ουδέτερος τόνος σε μια πρόταση.[29]

3.2) Text Summarization ο σκοπός του είναι το σύστημα να μπορεί να συμπυκνώσει το νόημα του κειμένου σε μια μικρότερη περιληπτική μορφή η οποία περιλαμβάνει τις σημαντικότερες πληροφορίες.[29]

4) Τέλος, υπάρχουν τεχνικές NLP που επιτρέπουν στο κείμενο να χρησιμοποιηθεί σε ανίχνευση ψευδών ειδήσεων και αφορά την μετατροπή ή αλλιώς **αναπαράσταση του κειμένου σε αριθμητική μορφή** όπως διανύσματα, με σκοπό την διευκόλυνση των συστημάτων ως προς την ταξινόμηση.[30]

4.1) Bag of Words (BoW), ουσιαστικά μετράει τις εμφανίσεις κάθε λέξης στο κείμενο και χρησιμοποιεί τις πληροφορίες αυτές για να δημιουργήσει πίνακα συχνοτήτων εμφάνισης.[30]

4.2) TF-IDF (Term Frequency - Inverse Document Frequency), κατανοεί την συχνότητα εμφάνισης των λέξεων μέσα στο κείμενο και στην συνέχεια βαθμολογεί κάθε λέξη βάση της σπανιότητας της ώστε να δημιουργηθούν νοηματικά συμπεράσματα χάρη στην σημαντικότητα των σπάνιων λέξεων.[30]

4.3) Word embeddings, έχουν 2 ρόλους, καθώς πέρα από την αναπαράσταση των λέξεων σε διανυσματικό πίνακα, έχουν την ικανότητα να τοποθετούν σε κοντινές θέσεις του πίνακα τις σημασιολογικά παρόμοιες λέξεις προσφέροντας πληροφορίες ως προς το νόημα και τα συμφραζόμενα του κειμένου. Τα πιο δημοφιλή embeddings είναι τα GloVe, Word2Vec και FastText.[30]

4.4)N-grams είναι ικανά να κατανοήσουν την αλληλουχία των λέξεων ανάλογα με το είδος τους, δηλαδή Bigram -> Golden State, Trigram->Golden State Warriors. [30]

Συμπερασματικά, η αναπαράσταση των δεδομένων κειμένου σε διανύσματα θεωρείται απαραίτητη διαδικασία για τα ML και DL μοντέλα καθώς υποβοηθούνται τα συστήματα ανίχνευσης και ουσιαστικά γίνεται εφικτή η διαδικασία ταξινόμησης.

3.4)Ανίχνευση μέσω Μηχανικής Μάθησης

Η χρήση των αλγορίθμων ταξινόμησης **Machine Learning** είναι αρκετά διαδεδομένη σε διαδικασίες δημιουργίας συστημάτων ανίχνευσης ψευδών ειδήσεων. Τα συστήματα πλαισιώνονται από τεχνικές NLP ώστε να εκπαιδευτούν και να δοκιμαστούν. Τον τελικό ρόλο της λήψης απόφασης αν μια είδηση θα ταξινομηθεί ως αληθινή ή ψεύτικη, τον αναλαμβάνει ο εκάστοτε αλγόριθμος.[31]

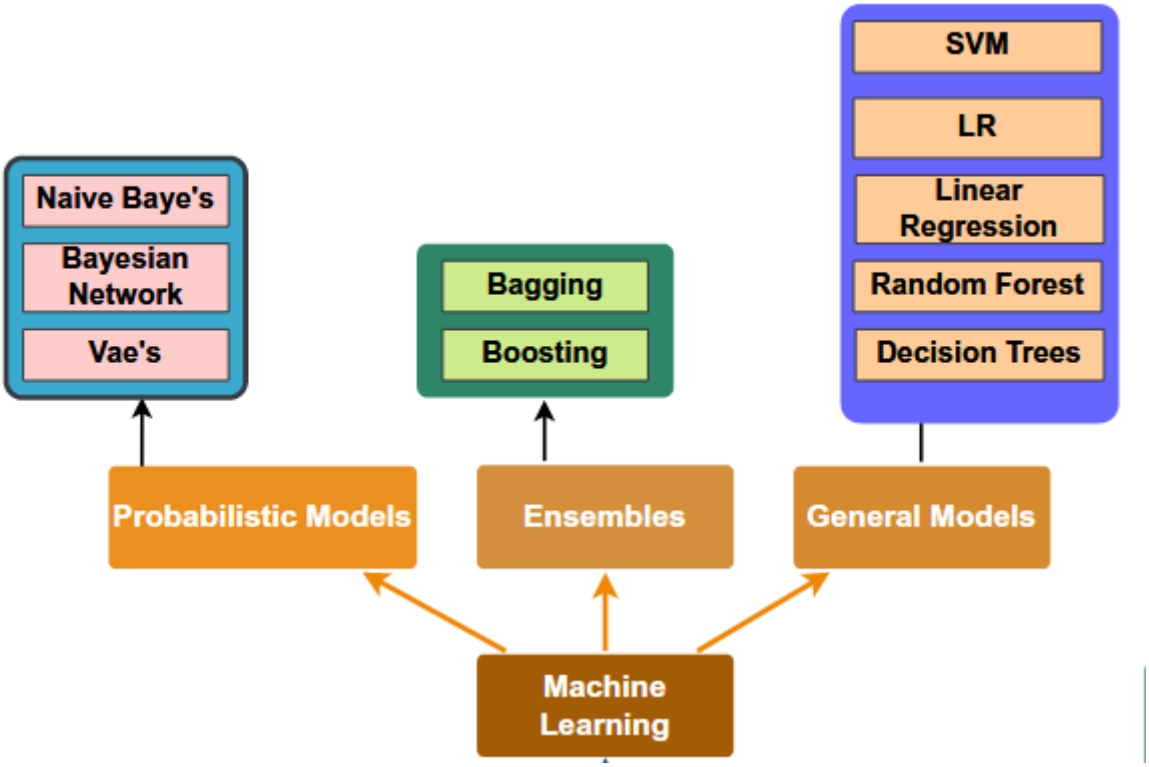
Εξίσου σημαντικό, είναι να αναφέρουμε πως οι δυνατότητες των Machine Learning based συστημάτων εξαρτώνται από την ποιότητα των εξαγόμενων χαρακτηριστικών από το dataset[31]. Η πιο συνηθισμένη διαδικασία είναι η επιλογή της θεματολογίας μέσω του κατάλληλου dataset που πρόκειται να δοκιμαστεί το σύστημα[32]. Έπειτα απαιτείται ποιοτικός καθαρισμός μέσω NLP τεχνικών , εκπαίδευση και δοκιμή του συστήματος και στο τέλος προκύπτουν οι πίνακες σύγχυσης (confusion matrix) όπου εξασφαλίζουν στατιστικά μετρικών αξιολόγησης όπως F1 score, Accuracy , Recall και Precision.[31], [33]

Εποπτευόμενη Μηχανική Μάθηση (Supervised Learning)

Η μηχανική μάθηση χρησιμοποιείται κατά κύριο λόγο μέσω αλγορίθμων όπου ονομάζονται εποπτευόμενοι καθώς εκπαιδεύονται σε δεδομένα όπου έχουν ήδη κατηγοριοποιηθεί μέσω των labels τους ως αληθινά ή ψεύτικα νέα. Με αυτόν τον τρόπο ο εκάστοτε αλγόριθμος ταξινόμησης μαθαίνει τις συσχετίσεις των εξαγόμενων χαρακτηριστικών και ταξινομεί.[33]

Οι πιο γνωστοί αλγόριθμοι εποπτευόμενης μηχανικής μάθησης που συναντούμε σε πειράματα ανίχνευσης είναι οι **Support Vector Machine, Naive-Bayes, Random Forest, Decision Tree,K-Nearest Neighbors , PAC(Passive Aggressive Classifier)και Logistic Regression** όπου αναλύονται παρακάτω.

Επίσης γίνεται και αναφορά στους boosting αλγορίθμους όπως **XGBoost, AdaBoost** όπου και αυτοί ανήκουν στους supervised αλγορίθμους.



Εικόνα 12 : Κατηγορίες Machine Learning προσεγγίσεων [9]

1) Ο **Naive Bayes** είναι από τους ταχύτερους ως προς την εκπαίδευση και πιο απλοϊκούς ως προς την πολυπλοκότητα αλγορίθμους ταξινόμησης. Βασίζεται στο θεώρημα του Bayes, δηλαδή υποθέτει τις σχέσεις ανεξαρτησίας μεταξύ χαρακτηριστικών των δεδομένων. Παρά την απλότητά του, έχει δείξει ικανοποιητική ακρίβεια στην ανίχνευση ψευδών ειδήσεων.[32]

2) Ο **SVM** (Support Vector Machine) είναι ένας ισχυρός αλγόριθμος επιβλεπόμενης μάθησης όπου όταν του εισάγονται τα δεδομένα σε μορφή διανυσμάτων, βασίζεται στην εύρεση ενός υπέρ επίπεδου (hyperplane) όπου τον βοηθά να διαχωρίζει τις ψευδείς με τις αληθείς πληροφορίες και έπειτα αναλαμβάνει την ταξινόμηση ψευδών και αληθινών ειδήσεων.[34]

3) Τα δέντρα απόφασης(**Desicion Trees**) είναι δομές τύπου διαγράμματος ροής(flowchart). Διασπά το προ επεξεργασμένο κείμενο σε κόμβους και διακλαδώσεις. Κάθε κόμβος αποτελεί ένα 'test' ύπαρξης μιας λέξης και κάθε φύλλο αποτελεί μια ταξινόμηση μέχρι να φτάσει στο τελευταίο φύλλο που γίνεται η τελική λήψη απόφασης.[33]

4) Ο αλγόριθμος **Random Forest** βασίζεται σε ένα πλήθος δέντρων απόφασης (decision trees) τα οποία συνεργάζονται για την τελική πρόβλεψη. Το 'δάσος' βασίζεται σε τεχνικές όπως bootstrapping και feature randomness.[33] Κάθε δέντρο εκπαιδεύεται σε υποσύνολα των δεδομένων όπου συμπερασματικά καταλαβαίνουν πως με αυτόν τον τρόπο αποφεύγονται και οι πιθανότητες overfitting.

5) Ο K-NN (**K-Nearest Neighbors**)ταξινομεί τις ειδήσεις βάση της 'απόστασης' δηλαδή της ομοιότητας των πληροφοριών με άλλες επαληθευμένες πληροφορίες από το σύνολο δεδομένων. Όταν μία είδηση αποτελεί γείτονα πολλών αληθών ειδήσεων, το πιο πιθανό είναι να θεωρηθεί αληθή.[33]

6) Ο αλγόριθμος λογιστικής παλινδρόμησης (**Logistic Regression**), μέσω της χρησιμοποίησης των εξαγόμενων χαρακτηριστικών υπολογίζει την πιθανότητας αλήθειας ή ψεύδους. Κατανοεί την γραμμικότητα των χαρακτηριστικών χρησιμοποιώντας την σιγμοειδή (sigmoid) συνάρτηση λογιστικής ώστε να επιστρέψει την δυαδική του πρόβλεψη.[31]

7) Ο αλγόριθμος PAC (**Passive Aggressive Classifier**) έχει πάρει το όνομα του από τις ικανότητες του. Η passive στάση του σημαίνει σταθερότητα όταν προβλέπει σωστά ενώ η aggressive στάση του εννοεί την άμεση προσαρμογή ως προς τα βάρη της πληροφορίας όταν κάνει λάθος πρόβλεψη. Η διαφορά του με τους προηγούμενους είναι ότι διαθέτει online learning δηλαδή ενημερώνεται σε πραγματικό χρόνο για την είσοδο μιας νέας είδησης (streaming data) χωρίς να χρειάζεται επανεκπαίδευση.[35]

Αλγόριθμοι ενίσχυσης.

8) Ο **XGBoost** (Extreme Gradient Boosting) είναι αλγόριθμος ενίσχυσης. Στα πειράματα που τον συναντήσαμε περιείχε συνδυασμό decision trees όπου συνεργάζονται. Δηλαδή, γίνεται μια επαναληπτική διαδικασία όπου τα δέντρα εκπαιδεύονται ώστε να διορθώσουν τα λάθη. Ο σκοπός της επαναληπτικής διαδικασίας, είναι η συνεχή βελτίωση του συστήματος ανίχνευσης ως προς την ακρίβεια.[36]

9) Ο **AdaBoost** αλγόριθμος χρησιμοποιεί εξίσου decision trees με σκοπό την δόμηση ενός ισχυρού μοντέλου ανίχνευσης. Η χρησιμότητα του επίκειται στις δυνατότητες του να επικεντρώνεται στα λανθασμένα αποτελέσματα (ειδήσεις που μοιάζουν αληθείς αλλά δεν είναι) και αυξάνει την απόδοση του συστήματος μέσω ensembled συνδυασμών αλγορίθμων ταξινόμησης. Έχει ανάγκη από ποιοτικά χαρακτηριστικά δεδομένων μέσω NLP και παρουσιάζει ανθεκτικότητα στο overfitting.[37]

Μη εποπτευόμενη Μηχανική Μάθηση (unsupervised Machine Learning)

Ουσιαστικά οι αλγόριθμοι δεν εκπαιδεύονται σε labeled δεδομένα και προσπαθούν να αναγνωρίσουν από μόνα τους τα μοτίβα , τις συσχετίσεις αλλά και τα κρυφές σχέσεις

τους. Δεν εμφανίζονται συχνά σε πειράματα ανίχνευσης ψευδών ειδήσεων καθώς όπως είναι λογικό τα στατιστικά τους είναι χαμηλά.[33] Ένας γνωστός αλγόριθμος μη εποπτευόμενης μηχανικής μάθησης είναι ο clustering.[33]

Ενισχυτική Μηχανική Μάθηση (Reinforcement Machine Learning)

Αφορά τους αλγορίθμους μηχανικής μάθησης όπου περιέχουν έναν πράκτορα που ειδικεύεται στην λήψη αποφάσεων ανάλογα με την αλληλεπίδραση με το περιβάλλον. Το σύστημα είναι ικανό να βελτιώνεται συνεχώς μέσω της διαδικασία ανταμοιβής ή τιμωρίας έπειτα από κάθε του ενέργεια. Σκοπός του συστήματος είναι με το πέρασμα του χρόνου να πετύχει όσο τον δυνατόν πιο ψηλές τιμές ανταμοιβών. Δεν χρησιμοποιείται ακόμα ευρέως και δεν συναντήσαμε συστήματα ανίχνευσης ψευδών ειδήσεων που να βασίζονται πάνω του.

Συμπερασματικά, τα συστήματα ανίχνευσης με χρήση Machine Learning έχουν εμφανιστεί τα προηγούμενα χρόνια. Η πρόοδος της τεχνολογίας τείνει να τα αντικαταστήσει με νευρωνικά δίκτυα και μετασχηματιστές όπου μπορούν να χειριστούν μεγαλύτερες ποσότητες δεδομένων και να κατανοήσουν με περισσότερο βάθος το νόημα των κειμένων. Παρόλα αυτά, τα συστήματα με machine learning έχουν αποδείξει την χρησιμότητα τους καθώς μέχρι και σήμερα οι ερευνητές δημοσιεύουν ML προσεγγίσεις που επιτυγχάνουν σχεδόν άριστες μετρικές αξιολόγησης. Ένα ακόμα αρνητικό τους, είναι ότι κατέχουν την μεγαλύτερη εξάρτηση από ποιοτικές NLP τεχνικές σε σχέση με τα πιο σύγχρονα συστήματα.

3.5) Ανίχνευση με Βαθιά Μάθηση

Η Βαθιά Μάθηση (**Deep Learning)** αποτελεί κομμάτι και εξέλιξη του τομέα της Μηχανικής Μάθησης. Τα συστήματα της είναι ικανά να δομήσουν ισχυρές προσεγγίσεις ανίχνευσης ειδήσεων καθώς μπορούν να διαχειριστούν με καλύτερη ευχέρεια μεγαλύτερες ποσότητες δεδομένων. Τα συστήματα τους, είναι εξίσου ικανά να ανιχνεύουν αυτόματα τις ψευδείς ειδήσεις καθώς και να εξάγουν και μαθαίνουν μόνα τους τα κατάλληλα χαρακτηριστικά των δεδομένων. Προσθέτουν ικανότητες κατανόησης των συμφραζομένων και του βαθύτερου νοήματος του κειμένου [11]

Οι προσεγγίσεις που προσφέρει η επιστημονική κοινότητα μέσω του Deep Learning αφορά τα νευρωνικά δίκτυα (**CNN & RNN**) όπου περιέχουν και την εξέλιξη των RNN (**LSTM, Bi-LSTM, GRU, Bi-GRU**)[24]. Επίσης, κομμάτι εξέλιξης του κλάδου θεωρούνται και οι **Transformer** τύπου **BERT** και **GPT.** Η χρήση deep learning τεχνικών σε ένα πείραμα ανίχνευσης απαιτεί ελαφριές τεχνικές NLP πάνω στο dataset που θα εκπαιδευτούν όπως η προ επεξεργασία και η αριθμητική αναπαράσταση. Επίσης τα νευρωνικά δίκτυα μέσω των word embeddings όπως προ αναφέραμε αποκτούν

επιπρόσθετες NLP ικανότητες που του εξασφαλίζουν την σε βάθος νοηματική κατανόηση του κειμένου.[24]

3.5.1) Ανίχνευση μέσω Νευρωνικών Δικτύων

Τα **Neural Networks** , είναι υποτελές κομμάτι της τεχνολογίας της Βαθιάς Μάθησης (όπου είναι υποτελές κομμάτι της Μηχανικής Μάθησης) και μία από τις σημαντικότερες δομικές τεχνολογίες στη τεχνητή νοημοσύνη (AI). Διαθέτουν σημαντικές ικανότητες εξαγωγής χαρακτηριστικών , κάτι το οποίο ωθεί την επιστημονική κοινότητα να τα χρησιμοποιεί με μεγαλύτερη συχνότητα.[24] Τα νευρωνικά δίκτυα είναι εμπνευσμένα από την λειτουργία του ανθρώπινου εγκεφάλου, λόγω της πολύπλευρης συνεργασίας των νευρώνων. Η δομή τους περιέχει τον τεχνητό νευρώνα όπου δέχεται πολλαπλές εισόδους δεδομένων με σκοπό την εκτέλεση υπολογισμών και την παραγωγή εξόδου. Οι είσοδοι που εισέρχονται στον τεχνητό νευρώνα αποτελούν αποτελέσματα υπολογισμών άλλων υπό νευρώνων ή αλλιώς στρωμάτων (**layers**). Η αλληλένδετη αυτή σχέση, δίνει την δυνατότητα στα νευρωνικά δίκτυα να κατανοούν βαθύτερες σχέσεις των δεδομένων όπως το νόημα, το συναίσθημα και τα συμφραζόμενα[24]. Η εκπαίδευση τους γίνεται μέσω αλγορίθμων προώθησης (**forward-propagation**) και αντίστροφης διάδοσης (**backward**-**propagation**)[24]. Οι αλγόριθμοι αυτοί, προκαλούν συνεχή ανανέωση των βαρών του κάθε στρώματος μέχρι να βελτιστοποιηθεί με την χρήση της παραγώγου της συνάρτησης κόστους. [24] Οι βασικές κατηγορίες των Νευρωνικών Δικτύων που είδαμε στα πειράματα είναι τα **CNN**s και **RNN**s.

1) Convolutional Neural Networks (CNNs)

Τα Convolutional Neural Networks (**CNNs**) είναι πιο δημοφιλή στην ανίχνευση ειδήσεων μέσω εικόνας και βίντεο, αλλά χρησιμοποιούνται εξίσου και στην κειμενική ανίχνευση. Αποτελούνται από το επίπεδο εισόδου (**input layer**) , ένα σύνολο εσωτερικών στρωμάτων (**hidden layers**) και το στρώμα εξόδου (**output layer**). Τα πιο σημαντικά στρώματα είναι το συνελικτικό (**convolutional layer**) , τα στρώματα δειγματοληψίας (**pooling layers**),το στρώμα κανονικοποίησης (**regularization layer**) και στο τέλος το στρώμα ταξινόμησης[24]. Έχουν την ικανότητα να εφαρμόζουν εσωτερικά φίλτρα τα οποία επιτρέπουν την εξαγωγή των σημαντικότερων χαρακτηριστικών, τον εντοπισμό λέξεων κλειδιών και τα τοπικά μοτίβα πληροφορίας του περιεχομένου που σχετίζονται με παραπληροφόρηση, διατηρώντας παράλληλα την τοπική πληροφορία[24]. Τέλος, είναι αναγκαίο να αναφέρουμε πως σε όλες τις περιπτώσεις στα 53 πειράματα που αναλύθηκαν, τα CNN συνδυαζόντουσαν είτε με RNNs όπως το LSTM , είτε με Transformers όπως το BERT και ποτέ δεν δομούσαν αποκλειστικά την μέθοδο ανίχνευσης.

2) Recurrent Neural Networks (RNNs)

Τα Recurrent Neural Networks (**RNNs**) έχουν σειριακή αρχιτεκτονική παράταξης των στρωμάτων τους σχηματίζοντας ένα κατευθυνόμενο γράφημα (**directed graph**) όπου συγκρατεί πληροφορίες από τα προηγούμενα βήματα[24], η αρχιτεκτονική αυτή τονίζει την καταλληλόλητα τους για σειρές δεδομένων όπως είναι τα κειμενικά δεδομένα στον τομέα των ψευδών ειδήσεων. Η έξοδος του κάθε στρώματος λειτουργεί ως είσοδος του επόμενου μέσω της σειριακής τοποθέτησης των στρωμάτων[24]. Εκπαιδεύονται με την χρήση του αλγορίθμου αντίστροφης διάδοσης (**backpropagation**) και κατέχουν σημαντικές ικανότητες πρόβλεψης που βασίζονται στην χρονική αλληλουχία που προ αναφέρθηκε. Τα δεδομένα ακολουθιακής φύσης (**sequence-based data**) προσφέρουν εξίσου στην ενίσχυση της δυνατότητας πρόβλεψης[24]. Το μόνο αρνητικό που μπορούμε να αναφέρουμε για τα RNNs είναι ότι σε περιπτώσεις χρησιμοποίησης της sigmoid συνάρτησης ενεργοποίησης, προκαλείται μείωση της τιμής της παραγώγου σε κάθε στρώμα (αρχίζοντας από τα τελευταία) αυξάνοντας τον χρόνο εκπαίδευσης. Το πρόβλημα αυτό είναι γνωστό ως **vanishing-gradient problem**[24].Τα RNNs, αποτελούν τεχνολογία αιχμής στον τομέα ανίχνευσης ειδήσεων καθώς σε πάρα πολλά πειράματα από τα 53 που αναλύθηκαν, η βασικής μέθοδος δομούνταν αποκλειστικά από RNNs όπως LSTM, GRU αλλά και συνδυαστικές προσεγγίσεις όπως ο συνδυασμός των RNNs και CNNs μεταξύ τους ή με Transformers.

Περιπτώσεις RNNs

1)Long Short-Term Memory (LSTM)

Τα **LSTM** έχουν αποδείξει την χρησιμότητα τους στον τομέα ανίχνευσης text-based ψευδών ειδήσεων καθώς θεωρούνται κατάλληλα για NLP tasks. Είναι εξελιγμένα αναδρομικά νευρωνικά δίκτυα (RNNs) καθώς μπορούν να κατανοούν μακροχρόνιες εξαρτήσεις των δεδομένων, λόγω ότι η αντίστροφη διάδοση σφάλματος (back propagation) διαρκεί πολύ λιγότερο χρονικά. Επιπρόσθετα, είναι ικανά να διατηρούν βραχυπρόθεσμες μνήμες για όσο χρειάζεται το σύστημα. Η δομή τους αποτελείται από είσοδο (**input gate**) , έξοδο (**output gate**) , πύλη της λήθης (**forget gate**) και κυψέλη μνήμης (**cell**). Οι πύλες αυτές υπολογίζουν την κρυφή κατάσταση (**hidden state**) η οποία διατηρεί τις τιμές της στην μνήμη για μεγάλα χρονικά διαστήματα.[24]

2)Bidirectional LSTM (Bi-LSTM)

Τα **Bi-LSTM** είναι μια εξελιγμένη μορφή του LSTM καθώς αρχικά ισχύουν οι ίδιες πληροφορίες. Υπάρχει όμως μια βασική διαφορά μεταξύ τους, όπως καταλαβαίνουμε από τον τίτλο Bidirectional, τα Bi-LSTMs λοιπόν, μπορούν να επεξεργάζονται τις ακολουθίες των δεδομένων και προς τις δύο κατευθύνσεις. Δηλαδή, υπάρχει δυνατότητα πρόσβασης στα βήματα που πραγματοποιήθηκαν, από το μέλλον στο παρελθόν. Η ικανότητα αυτή προσφέρει στο Bi-LSTM μεγαλύτερες δυνατότητες κατανόησης του νοήματος, των συναισθημάτων, των οντοτήτων και των συμφραζομένων.[24]

3)Gated Recurrent Units (GRU)

Τα **GRU**s θεωρούνται εξίσου μορφές εξέλιξης των RNNs. Η δομή τους και η πολυπλοκότητα τους είναι πιο απλοϊκές από τα LSTM ενώ έχει παρατηρηθεί πως είναι και πιο ελαφριά ως προς την χρήση υπολογιστικών πόρων. Η δομή τους περιέχει μόνο δύο πύλες, την **reset gate** και την **update gate** όπου επιτρέπουν ταχύτερη εκπαίδευση και μικρότερη πιθανότητα overfitting. Η διαφορά τους με τα LSTMs, είναι ότι δεν χρειάζονται ξεχωριστή μονάδα μνήμης για να διαχειριστούν την πληροφορία. Τα GRUs αποκαλύπτουν όλο το εσωτερικό κρυφό περιεχόμενο χωρίς να πραγματοποιείται εσωτερικός έλεγχος. Τέλος, έχει παρατηρηθεί πως επιτυγχάνουν παρόμοιες αποδόσεις με τα LSTMs στις μετρικές αξιολόγησης.[24]

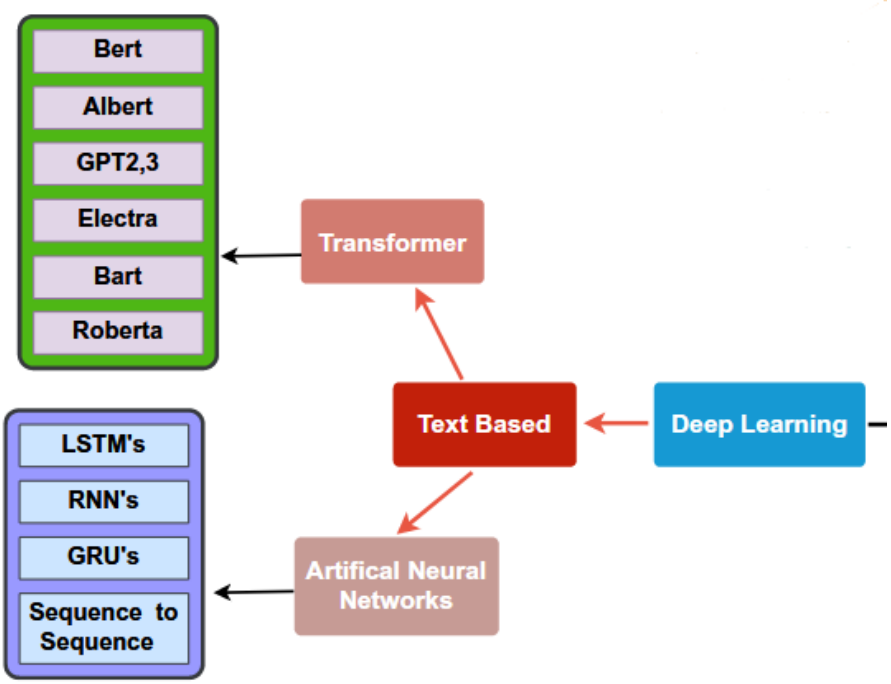
4)Bidirectional Gated Recurrent Units (Bi-GRU)

Ότι ίσχυε στις διαφορές των LSTMs με τα Bi-LSTMs ισχύει και εδώ. Δηλαδή, τα **Bi-GRU** έχουν δυνατότητες επεξεργασίας των δεδομένων και προς τις 2 κατευθύνσεις.[24]

Συμπερασματικά, από τα νευρωνικά δίκτυα πρέπει να κρατήσουμε τις αυξημένες δυνατότητες τους, δηλαδή, την διαχείριση περισσότερων δεδομένων, την αυτόματη εξαγωγή και μάθηση των χαρακτηριστικών, την καλύτερη ανάλυση του συναισθήματος και της υψηλότερης αξιολόγησης σε μεγάλα δεδομένα από τις προηγούμενες προσεγγίσεις ανίχνευσης.

3.5.2) Ανίχνευση με χρήση Μετασχηματιστών

Οι μετασχηματιστές (**Transformers**) θεωρούνται εξελιγμένα νευρωνικά δίκτυα, ακόμα πιο ικανά σε NLP tasks όπως η ανίχνευση ψευδών ειδήσεων[38]. Το κεντρικό τους γνώρισμα είναι ο μηχανισμός προσοχής (**Attention Mechanism**)[39], ο οποίος χωρίζεται στις υπό κατηγορίες **self-attention** και **multi-head attention**[38]**.** Οι μηχανισμοί αυτοί, τους επιτρέπουν να εστιάσουν προσεκτικά σε πολλά μέρη του κειμένου ταυτόχρονα. Έχουν την ικανότητα να εξάγουν αυτόματα τα κατάλληλα χαρακτηριστικά του κειμένου και να μαθαίνουν τις σχέσεις αλληλεξάρτησης των δεδομένα χωρίς να επηρεάζονται από την απόσταση τους μέσα στο κείμενο. Η αρχιτεκτονική τους αποτελείται από δύο κύρια μέρη, τον Encoder, δηλαδή, τον κωδικοποιητή που επεξεργάζεται τα δεδομένα εισόδου και τον Decoder, δηλαδή, τον αποκωδικοποιητή που παράγει έξοδο[39]. Βέβαια, οι transformers που συναντήσαμε στα 53 πειράματα ήταν BERT-based όπου χρησιμοποιούν μόνο τις encoder δυνατότητες τους[38]. Αποτελούνται από πάρα πολλά εσωτερικά στρώματα που περιέχουν **self-attention** στρώματα, **fully connected** νευρωνικά δίκτυα , **residual connections** και **normalization** στρώματα για περισσότερη σταθερότητα κατά την εκπαίδευση.[39]



Εικόνα 13 : Διαχωρισμός των Deep Learning τεχνικών για text-based tasks, σε Transformers και Neural Networks [9]

Οι πιο συχνές μορφές Tranformer-based προσεγγίσεων που συναντήσαμε στα 53 πειράματα, αφορούσαν το BERT αλλά και τις αρκετές παραλλαγές του. Μία προσέγγιση αφορούσε τον συνδυασμό GPT με BERT και υπήρχαν και εμφανίσεις ensembled BERT συστημάτων.

1.BERT (Bidirectional Encoder Representations from Transformers)

Το BERT είναι ένα εξελιγμένο προεκπαιδευμένο μοντέλο ενσωμάτωσης λέξεων (word-embedding) δημιουργημένο από την Google AI Language το 2018 και βασισμένο σε αρχιτεκτονική Transformer αλλά μόνο στην Encoder πλευρά του με χρήση multi-head self-attention μηχανισμών[38]. Η ικανότητα που το ξεχωρίζει από άλλους Transformers είναι ο διπλής κατεύθυνσης μηχανισμός προσοχής **Bidirectional Self-Attention**[38]**.** Τα embeddings εισόδου αφορούν Tokens όπως **CLS** (πραγματοποιεί ταξινόμηση) και **SEP** (διαχωρίζει της διαφορετικές προτάσεις)[38]. Κατά την προεκπαίδευση τους χρησιμοποιούνται 2 σημαντικά NLP tasks. Τα ονόματα τους είναι **Masked Language Modeling (MLM)** και **Next Sentence Prediction (NSP)**[38]**,** όπου μαθαίνουν στο σύστημα να προβλέπει τυχαία τις 'masked' λέξεις του κειμένου συνεισφέροντας στην σημασιολογική ανάλυση και του μαθαίνουν να αναγνωρίζει την αλληλεξάρτηση της λογικής των προτάσεων, μεταξύ προηγούμενης και επόμενης πρότασης, συνεισφέροντας στην βαθύτερη κατανόηση των σχέσεων. Εξίσου σημαντική είναι η αναφορά στα contextual embeddings[38], οι διαφορά του με τα word embeddings είναι ότι ακόμα και οι ίδιες λέξεις έχουν διαφορετικό embedding ανάλογα με το νόημα της

λέξης π.χ.(η λέξη καπνός θα είχε διαφορετικά embeddings για τον καπνό που χρησιμοποιείται στο κάπνισμα και τον καπνό που προκαλείται από φωτιά, τα word embeddings δεν το κάνουν αυτό). Το BERT για να χρησιμοποιηθεί σε ένα σύστημα ανίχνευσης ψευδών ειδήσεων χρειάζεται **fine-tuning** πάνω στο dataset που θα χρησιμοποιηθεί.[24], [38] Στα 53 πειράματα είδαμε αρκετές παραλλαγές του BERT.

1)Το **BERT** όπου περιέχει 24 layers, 16 attention heads και 340 εκατομμύρια εσωτερικές παραμέτρους.[38]

2)Το **BERT base** όπου περιέχει 12 layers, 12 attention heads και 110 εκατομμύρια εσωτερικές παραμέτρους.[38]

3)Το **RoBERTa** όπου αποτελεί optimized μορφή του BERT και αναπτύχθηκε από την Facebook AI το 2019. Η διαφορά τους έγκειται στην ανεξαρτητοποίηση του από το pre training task NSP, καθώς επικεντρώνεται στο MLM.Επίσης, η διαδικασία εκπαίδευσης του είναι διαφορετική από το BERT, καθώς εκπαιδεύεται με μεγαλύτερες ποσότητες δεδομένων και περισσότερα βήματα εκπαίδευσης.[40]

4) Το **DistilBERT** όπου είναι συμπυκνωμένη μορφή του BERT base κατά 60% καθώς και 40% πιο γρήγορο. Έχει 6 layers και παρουσιάστηκε από την Hugging Face στα μέσα του 2019. [41]

5)Το **CT-BERT** όπου είναι εξειδικευμένο σε covid-based tweets, καθώς έχει προ εκπαιδευτεί σε περισσότερα από 160 εκατομμύρια tweets. Επίσης, έχει την αρχιτεκτονική του BERT που αναφέρεται παραπάνω[42].

6) Το **mBERT** το οποίο έχει παρόμοια αρχιτεκτονική με το BERT base και είναι ένα multilingual μοντέλο που εξυπηρετεί 104 διαφορετικές γλώσσες.[43]

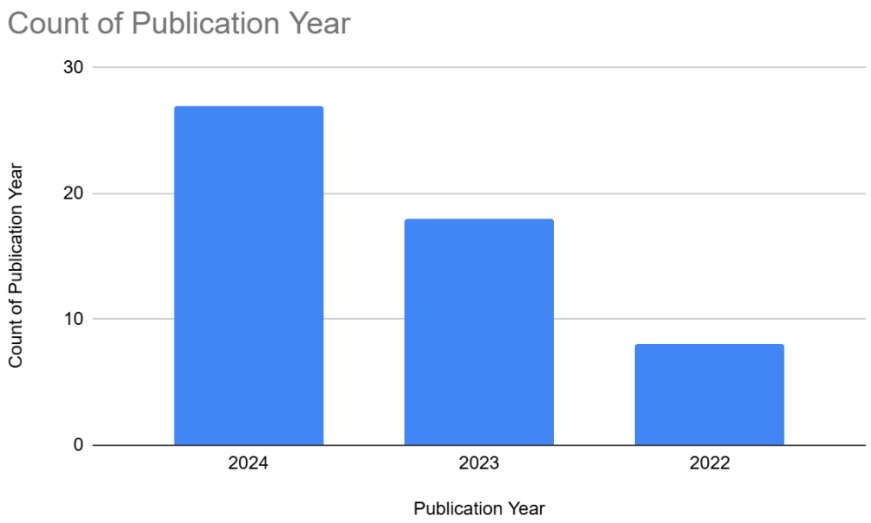
7) Το **AraBERT** όπου είναι προ εκπαιδευμένο στην αραβική γλώσσα, έχει την αρχιτεκτονική του BERT base και είναι εξίσου σημαντικό να αναφέρουμε πως πέρα από την επίσημη αραβική γλώσσα, μπορεί να επεξεργαστεί και τις τοπικές του αραβικές διαλέκτους. [44]

Συμπερασματικά, οι BERT προσεγγίσεις αποτελούν την σημαντική εξέλιξη του κλάδου και την ανάδειξη του πόσο πολύ έχει προχωρήσει η τεχνητή νοημοσύνη. Μπορεί να είναι τα πιο βαριά συστήματα από όσα είδαμε βάση πολυπλοκότητας και ανάγκης υπολογιστικών πόρων , αλλά έχουν πιο αναπτυγμένες ικανότητες στο να επεξεργαστούν μεγάλα ποσοτικά σύνολα δεδομένων και να κατανοήσουν σε μεγαλύτερο βαθμό την εσωτερική τους πληροφορία.

4) Πειράματα

4.1) Εισαγωγή στα Πειράματα

Στο κεφάλαιο αυτό παρέχονται πληροφορίες για τα 53 επιστημονικά πειραματικά άρθρα που αναλύθηκαν κατά την εκπόνηση της εργασίας. Αρχικά, όσον αφορά τις ημερομηνίες κυκλοφορίας παρατηρούμε την ανοδική τάση των δημοσιεύσεων αλλά και κατανοούμε πως η ραγδαία εξέλιξη της τεχνητής νοημοσύνης διευκολύνει τους ερευνητές να αναπτύσσουν όλο και πιο προηγμένα συστήματα ανίχνευσης. (1/1/2022-1/10/2024)



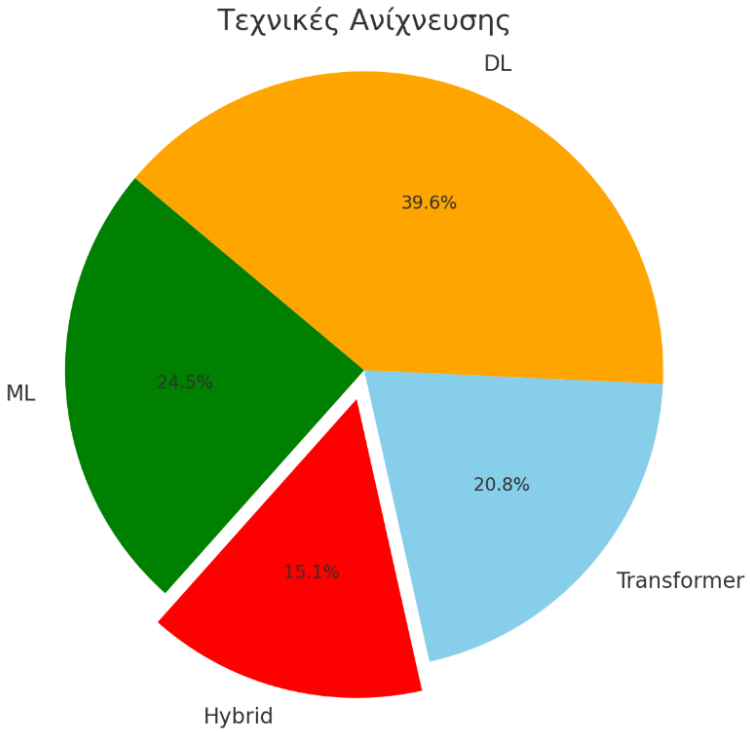
Εικόνα 14 : Ημερομηνίες κυκλοφορίας των Papers που περιλαμβάνουν πείραμα

4.2) Μέθοδοι Ανίχνευσης Ποσοτικά

Οι ερευνητές αρχικά απέδειξαν ότι τα Fake News μπορούν να ανιχνευτούν με πολλαπλούς τρόπους ανάλογα με το είδος και το ΄βάρος΄ σε πολυπλοκότητα της προσέγγισης. Υπάρχουν papers που χρησιμοποίησαν ως βασική τεχνική ανίχνευσης τις τεχνολογίες των Machine Learning αλγορίθμων ταξινόμησης (SVM, Decision Trees, LR, RF etc), των Deep Learning νευρωνικών δικτύων (CNN, RNN, LSTM,BiLSTM,GRU,BiGRU) και transformers(AraBERT,BERT,RoBERTa,mBERT).Εξίσου σημαντική είναι η αναφορά στον κλάδο του NLP όπου όπως είναι λογικό, δεν μπορεί να λείπει από προβλήματα text-based ανίχνευσης. Σχεδόν σε κάθε πείραμα είναι αναγκαία η χρήση των απλών μεθόδων NLP για προεπεξεργασία των δεδομένων ενώ το NLP εμφανίζεται και σε transformer-based προσεγγίσεις όπου NLP tasks, προεκπαιδεύουν εσωτερικά τους μετασχηματιστές όπως το BERT σε δυνατότητες επίλυσης NLP προβλημάτων όπως η ανίχνευση ψευδών ειδήσεων. Επιπρόσθετα, παρατηρούμε ότι στα πειράματα χρησιμοποιούνται και υβριδικές τεχνικές συνδυάζοντας για παράδειγμα μετασχηματιστές με νευρωνικά δίκτυα. Τέλος, είναι άξιο αναφοράς ότι οι ερευνητές

πέρα από την παρουσίαση της προσέγγισης που υλοποίησαν, συνηθίζουν να τις συγκρίνουν με υπάρχουσες έρευνες πάνω στα ίδια σύνολα δεδομένων.

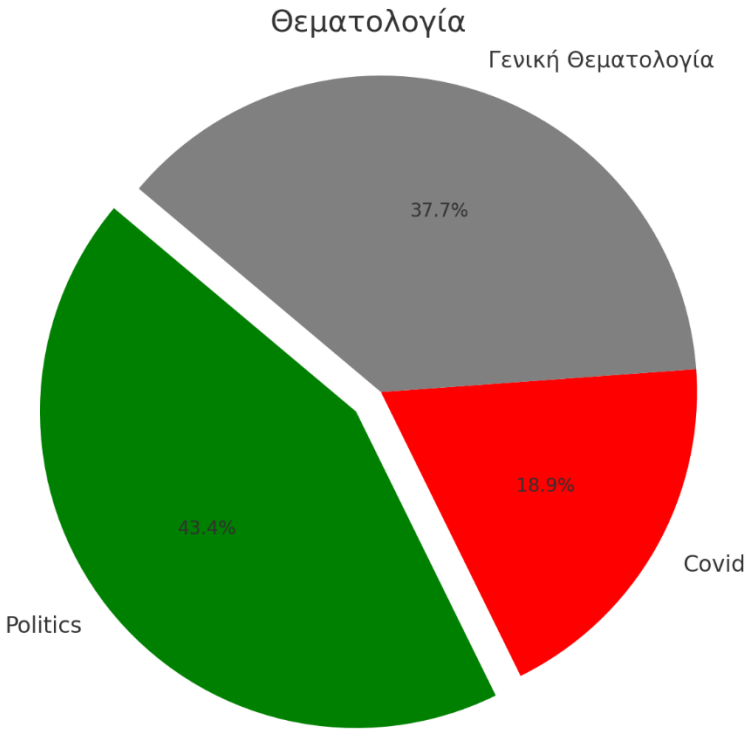
Το παρακάτω γράφημα , χωρίζεται σε Machine Learning, Deep Learning (Neural Networks) , Transformers αλλά και Hybrid τεχνικές που περιλαμβάνουν συνδυασμό DL τεχνικών όπως BERT-LSTM , BERT-BiGRU.



Εικόνα 15 : Διαχωρισμός μεταξύ υβριδικών και μη προσεγγίσεων στο σύνολο των πειραμάτων

4.3) Θεματολογίες των Συνόλων Δεδομένων

Το παρακάτω pie chart, ουσιαστικά παρουσιάζει την ανάγκη να περιοριστούν τα fake news στον υγειονομικό και τον πολιτικό τομέα. Για αυτό τον λόγο, έχουν δημιουργηθεί αρκετά political-based και covid-based datasets τα οποία τα έχει εκμεταλλευτεί ο ακαδημαϊκός χώρος στα πειράματα του. Όσον αφορά την γενική θεματολογία, τα datasets αυτά δεν έχουν κύρια θεματολογία αλλά ανάμεικτη όπως οικονομικά, κοινωνικά, πολιτικά κλπ.

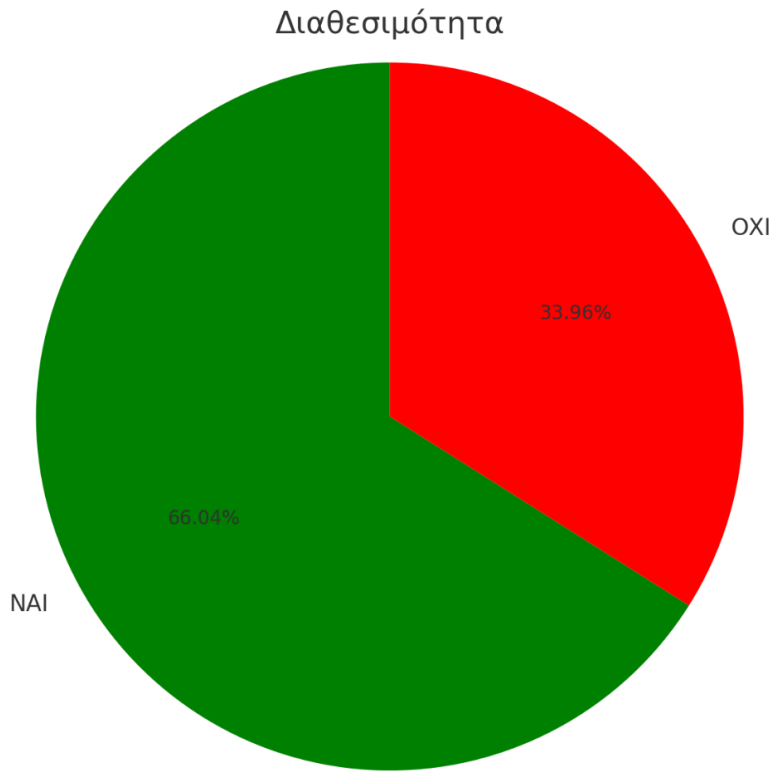


Εικόνα 16 : Θεματολογία των datasets

4.4) Διαθεσιμότητα των Δεδομένων

Στα 53 πειράματα που αναλύθηκαν, τα σύνολα των δεδομένων κατά κύριο λόγω ήταν διαθέσιμα προς το κοινό. Η πλειοψηφία των ερευνητών, ανέφερε τα ονόματα των datasets αλλά και την ιστοσελίδα προέλευσης ενώ σε πολλά custom made σύνολα, οι ερευνητές παρείχαν τον σύνδεσμο του repository στο Github ώστε να έχουν δυνατότητα πρόσβασης οι μελλοντικοί ερευνητές.

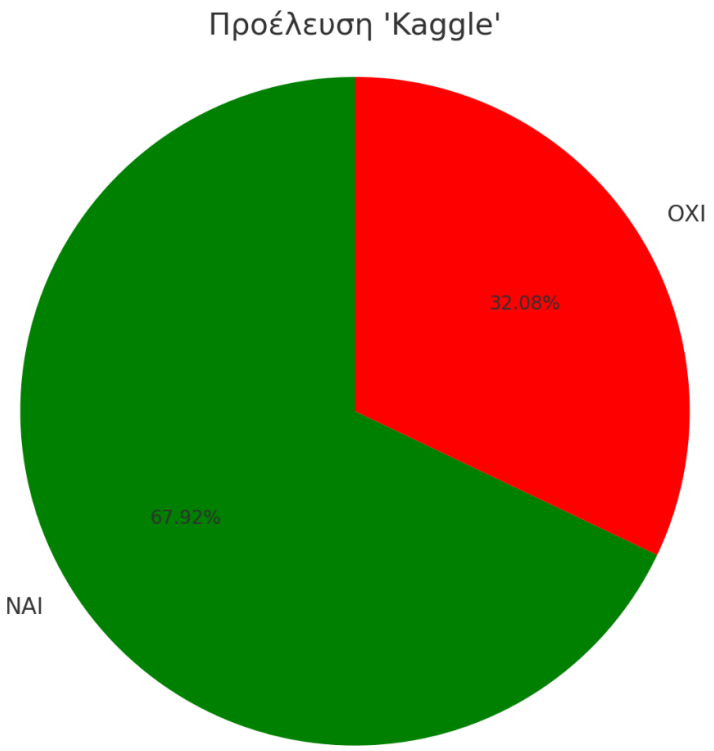
Αντιθέτως, υπήρχαν και σύνολα δεδομένων που δεν υπήρχε η δυνατότητα πρόσβασης για διάφορους λόγους. Ένας από τους λόγους αυτούς, είναι ότι σε μερικά πειράματα δεν γινόταν αναφορά στον τίτλο του dataset ή στην προέλευση ή στο μέγεθος του. Επίσης, υπήρχαν και σύνολα δεδομένων, κυρίως custom made, όπου οι ερευνητές ανέφεραν πως τα δεδομένα μπορούν να ζητηθούν μέσω αιτήματος στους ιδίους.



Εικόνα 16 : Ποσοτική απεικόνιση των διαθέσιμων datasets

4.5) Προέλευση Δεδομένων

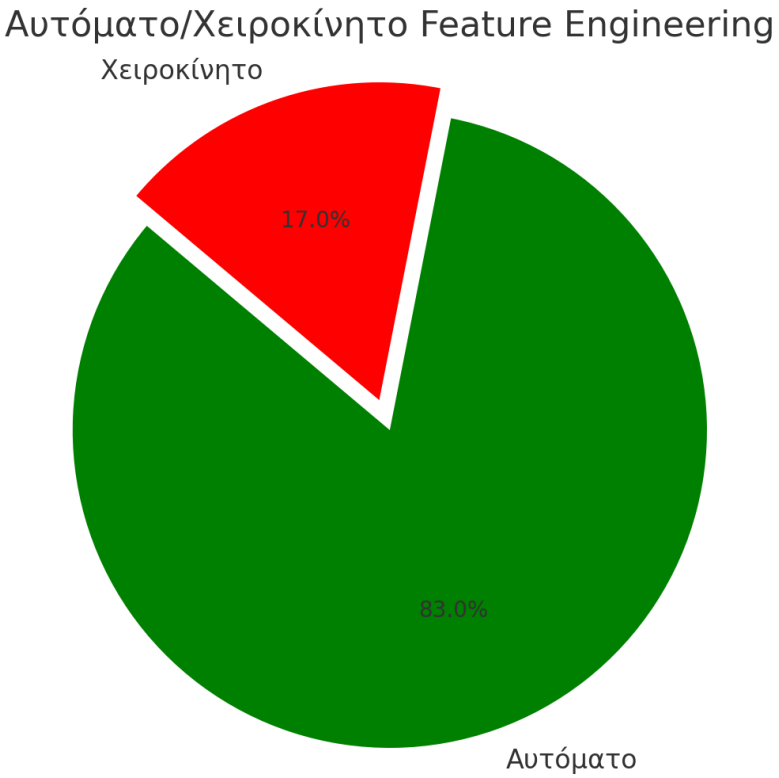
Στο παρακάτω διάγραμμα παρουσιάζεται ο σημαντικός ρόλος της Kaggle στον τομέα ανίχνευσης ψευδών ειδήσεων. Τα σύνολα δεδομένων διατίθενται στο κοινό για ερευνητικούς σκοπούς με τα πιο συνηθισμένα datasets να είναι το LIAR, το ISOT και το Fake News Net.



Εικόνα 17 : Προέλευση των datasets από το Kaggle

4.6) Εξαγωγή Χαρακτηριστικών από τα Δεδομένα

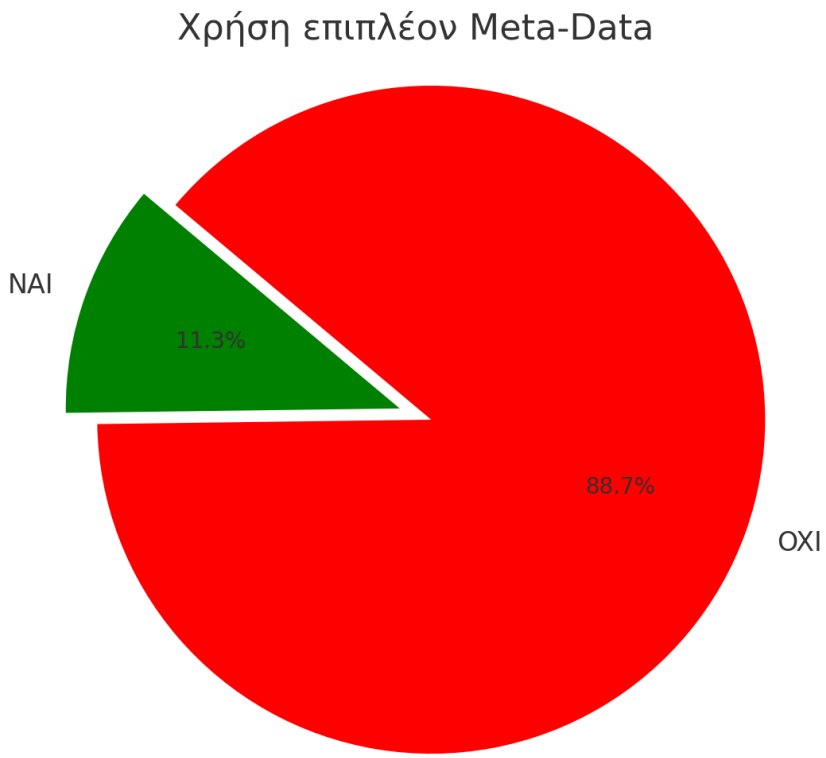
Το γράφημα περιγράφει πόσα από τα πειράματα είχαν αυτοματοποιημένη διαδικασία εξαγωγής χαρακτηριστικών σε αντίθεση με αυτά που οι ερευνητές τα καθόρισαν χειροκίνητα. Ουσιαστικά, μέσω του γραφήματος αυτού, κατανοούμε την εξέλιξη στον τομέα ανίχνευσης ψευδών ειδήσεων όπου τα νευρωνικά δίκτυα και οι μετασχηματιστές διαθέτουν δυνατότητες εξαγωγής των χαρακτηριστικών αυτοματοποιημένα.



Εικόνα 18 : Αυτόματη ή χειροκίνητη εξαγωγή χαρακτηριστικών

4.7) Χρήση των Διαθέσιμων Μέτα-Δεδομένων

Το παρακάτω pie-chart, περιγράφει την συντριπτική πλειοψηφία των ερευνητικών προσεγγίσεων όπου δεν γινόταν χρήση επιπρόσθετων meta-data όπως author, date, credibility ,title,url ,topic. Παράλληλα, οι ερευνητές κρατούσαν προς χρήση μόνο τον τίτλο, το σώμα του κείμενου και το label που χαρακτηρίζει την είδηση. Παρόλο που το ποσοστό χρησιμοποίησης που παρουσιάζεται είναι πολύ μικρό, οι προσεγγίσεις αυτές αφορούσαν εξιδεικευμένα συστήματα ανίχνευσης πάνω σε social media platforms όπως το Reddit και το Twitter. Τα δεδομένα αυτά, ήταν χρονικά μεταδεδομένα, αν ο χρήστης του Twitter είναι Verified, σχόλια και retweets. Μελλοντικά, όταν θα προκύψουν περισσότερες έρευνες για συγκεκριμένα μέσα κοινωνικής δικτύωσης, θεωρούμε απαραίτητη την χρήση μέτα-δεδομένων.

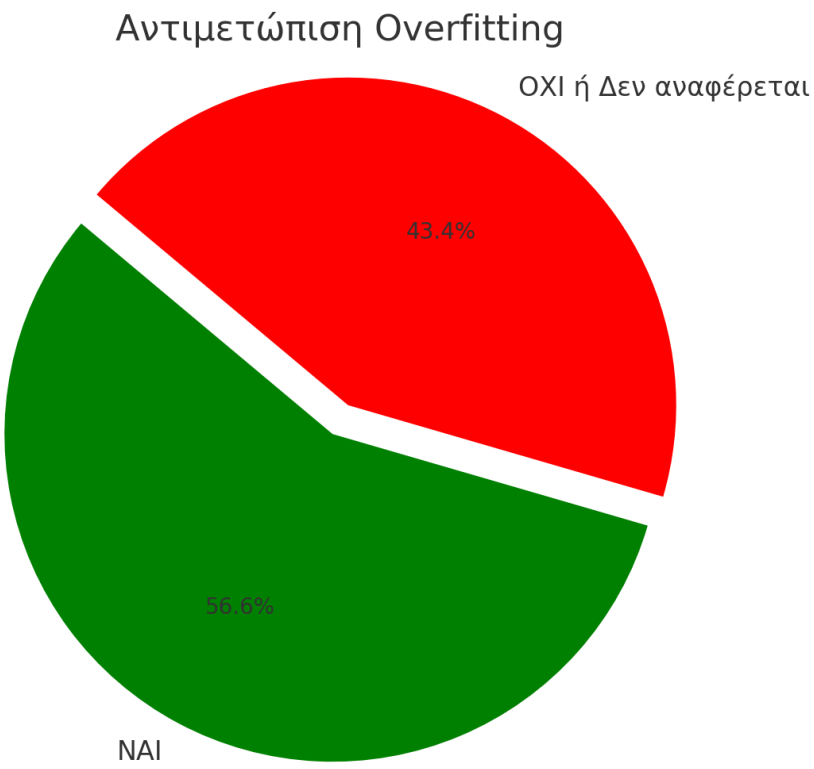


Εικόνα 19 : Χρήση επιπρόσθετων meta-data

4.8) Το Πρόβλημα του Overfitting

Το πρόβλημα του overfitting, αφορά την διαδικασία εκπαίδευσης των συστημάτων. Το πρόβλημα εμφανίζεται όταν ουσιαστικά τα συστήματα μαθαίνουν πολύ καλά τα δεδομένα (αγνοούν τον θόρυβο, μαθαίνουν τοπικά patters, αποκτούν μεγάλα βάρη) με αποτέλεσμα η αξιολόγηση τους να είναι σχεδόν τέλεια στο συγκεκριμένο σύνολο δεδομένων, αλλά σε κάποιο διαφορετικό σύνολο, η αξιολόγηση να αποκλίνει κατά πολύ. Συνήθως, ευπάθεια στο overfitting παρουσίαζαν τα συστήματα που χρησιμοποιούσαν μικρά σε μέγεθος σύνολα δεδομένων ή απλά η αρχιτεκτονική του συστήματος από την φύση της είναι ευπαθής στην υπέρ εκπαίδευση.

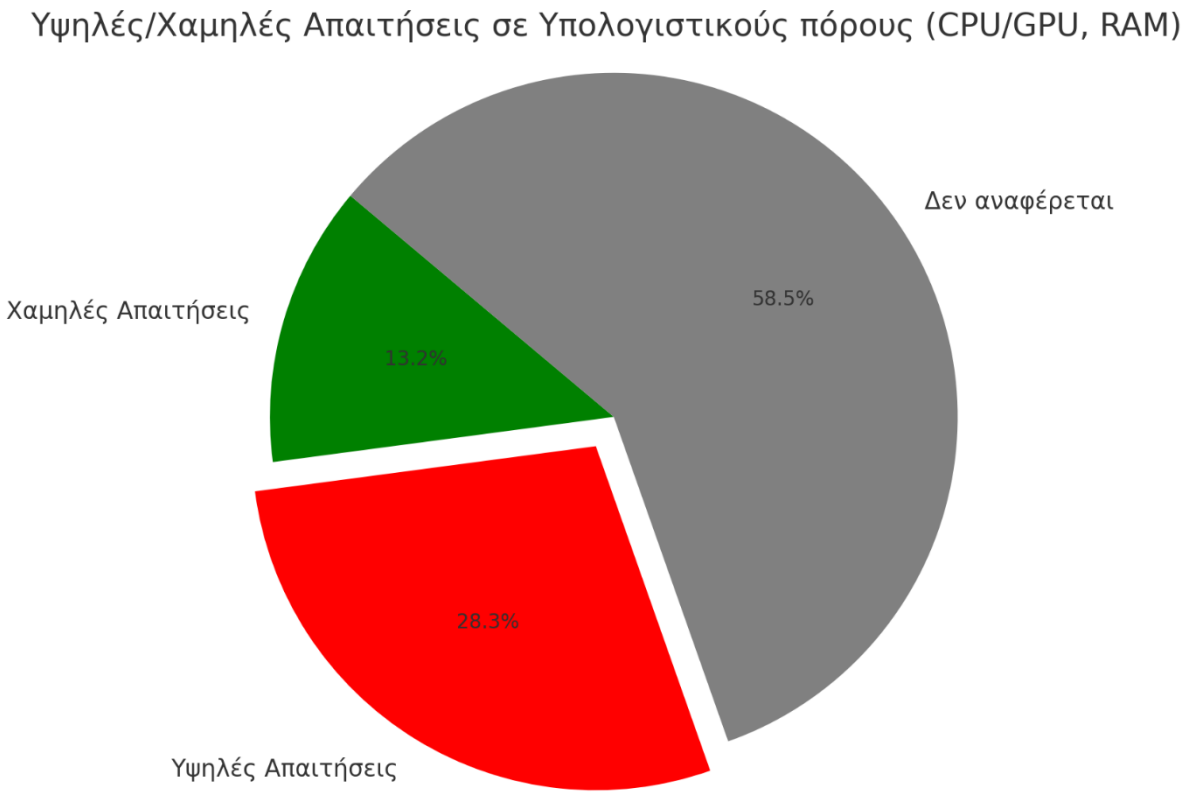
Στα 53 πειράματα που αναλύθηκαν, η πλειοψηφία των ερευνητών αναγνώριζε το πρόβλημα της υπέρ εκπαίδευσης. Μερικές από τις μεθόδους μετρίασης του φαινομένου ήταν το **dropout layer** (τυχαία απενεργοποίηση νευρώνων) ,το **regularization** (μια διαδικασία απόδοσης ποινών όταν το σύστημα αποκτούσε μεγάλα βάρη, με σκοπό να το αποτρέψει από την συνήθεια αυτή) , το **K-fold cross validation** (διαίρεση του συνόλου των δεδομένων σε ισόποσα τεμάχια όπου το σύστημα για Κ φορές θα εκπαιδεύεται στα Κ-1 τεμάχια και στο ένα fold που περισσεύει, θα δοκιμάζεται) και **early stopping** (σταματημός της διαδικασία εκπαίδευσης όταν η απόδοση στο validation set αρχίζει να μειώνεται).



Εικόνα 20: Αναγνώριση του Overfitting

4.9) Ανάγκες Υπολογιστικών Πόρων

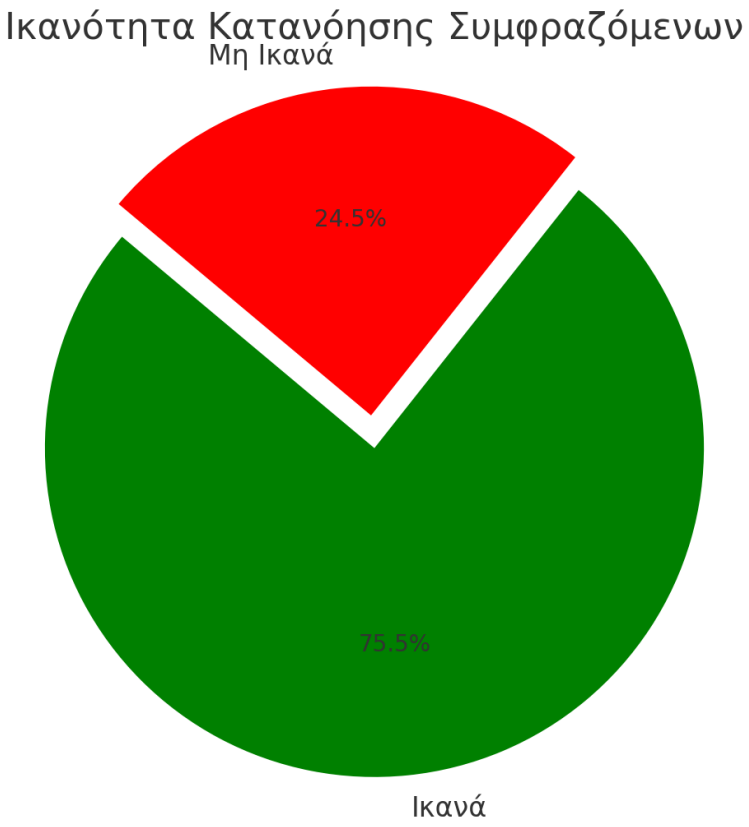
Στο επόμενο γράφημα περιγράφεται η ποσοτική εικόνα των πειραμάτων που χρειάζονται υψηλούς ή μη υπολογιστικούς πόρους π.χ. συστήματα Transformer-based που χρειάζονται fine tunning. Ενώ, συστήματα που βασίζονται σε αλγορίθμους ταξινόμησης machine learning είναι πιο ελαφριά. Υπάρχουν περιπτώσεις που έχουν χρησιμοποιεί πιο ελαφριές αρχιτεκτονικές, όπως το BERT-base ή DistilBERT, αντί του BERT ώστε να μην έχει υψηλές απαιτήσεις. Εξίσου σημαντικό, είναι να αναφέρουμε πως οι υψηλές υπολογιστικές απαιτήσεις δεν σχετίζονται μόνο με την αρχιτεκτονική των συστημάτων, αλλά και με το μέγεθος του συνόλου δεδομένων. Στο διάγραμμα έχουν συμπεριληφθεί μόνο οι περιπτώσεις που γίνεται ρητή αναφορά σε specifications των υπολογιστικών συστημάτων ή αναφέρονται απόψεις που αφορούν τον χρόνο εκπαίδευσης και πρόβλεψης.



Εικόνα 21 : Υπολογιστικές απαιτήσεις

4.10) Ικανότητα Κατανόησης των Συμφραζομένων

Το pie-chart περιγράφει το πόσα πειράματα έχουν την ικανότητα να αντιλαμβάνονται συμφραζόμενα αλλά και κατ' επέκταση το βαθύτερο νόημα και τις μακροπρόθεσμες σχέσεις των δεδομένων. Η κατανόηση των συμφραζομένων είναι μια σημαντική NLP ικανότητα, όπου μπορούν να την αποκτήσουν μέσω των embeddings τα Deep Learning συστήματα π.χ.(Neural Networks όπως LSTM,Bi-LSTM) και οι Transformers π.χ. (BERT,RoBERTa etc) σε αντίθεση με τα παραδοσιακά Machine Learning συστήματα ανίχνευσης π.χ.(SVM,DT,RF).



Εικόνα 22 : Ικανότητα κατανόησης συμφραζομένων

4.10) Τρόποι Αξιολόγησης των Πειραμάτων

Κατά την ανασκόπηση των πειραμάτων παρατηρούμε τις μετρικές αξιολόγησης Accuracy, F1, Recall και Precision. Οι ερευνητές στα συστήματα ανίχνευσης ψευδών ειδήσεων θεωρούσαν ως κλάση αναγνώρισης την ψευδή είδηση.

- **Accuracy** : Η ακρίβεια είναι μια μετρική αξιολόγησης όπου χρησιμοποιούταν στα περισσότερα πειράματα, δηλαδή ακόμα και στα πειράματα που οι ερευνητές παρουσίαζαν μόνο μια αξιολόγηση, αυτή θα ήταν το accuracy. Η καταλληλόλητα της έγκειται στην ισορροπία των δεδομένων καθώς για παράδειγμα σε ένα ανισόρροπο σύνολο 100 ειδήσεων με τις 90 αληθείς και τις 10 ψευδείς, αν το σύστημα προβλέψει 100 αληθείς θα έχει accuracy 90% αλλά παράλληλα το Recall και το Precision θα είναι μηδενικό.

$$\text{Accuracy} = \frac{\text{True Positives (TP)} + \text{True Negatives (TN)}}{\text{Σύνολο δείγματος}}$$

- **F1 Score** : Το F1 Score είναι μια μετρική αξιολόγησης όπου ουσιαστικά δεν επηρεάζεται ακόμα και από ανισόρροπα σύνολα δεδομένων, καθώς υπολογίζει τον αρμονικό μέσο όρο των μετρικών Recall και Precision.

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

- **Precision** : Το Precision είναι η μετρική αξιολόγησης όπου υπολογίζει το ποσοστό των πόσων ανιχνευμένων ψευδών ειδήσεων είναι όντως ψευδείς.

$$\text{Precision} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

- **Recall :** Το Recall είναι η μετρική αξιολόγησης όπου υπολογίζει το ποσοστό των ψευδών ειδήσεων που ανίχνευσε το σύστημα μέσα σε όλο το σύνολο των ψευδών ειδήσεων που βρίσκεται στο dataset.

$$\text{Recall} = \frac{\text{True Positives (TP)}}{\text{True Positives (TP)} + \text{False Negatives (FN)}}$$

True Positive ->	Σωστή πρόβλεψη σε ψευδή είδηση	**True Negative ->**	Σωστή πρόβλεψη σε αληθή είδηση
False Positive ->	Λάθος πρόβλεψη σε αληθή είδηση	**False Negative ->**	Λάθος πρόβλεψη σε ψευδή είδηση

4.11) Σπάνια Χαρακτηριστικά των Πειραμάτων

Κατά την ανάλυση των προσεγγίσεων, μόνο 4 συστήματα είχαν την ικανότητα της **Real Time** ή **Early Detection**. Οι επιζήμιες ικανότητες των ψευδών ειδήσεων γιγαντώνονται με το διαμοιρασμό τους και τον αργοπορημένο περιορισμό τους. Το στατιστικό αυτό, είναι ανησυχητικό και θα πρέπει να βελτιωθεί καθώς οι περισσότερες προσεγγίσεις αφορούν offline εκπαίδευση και δοκιμή των συστημάτων. Οι 4 προσεγγίσεις που είχαν ικανότητα πρώιμης ανίχνευσης ήταν social media based και αξιοποιούσαν είτε τα σχόλια μια δημοσίευσης, είτε χρησιμοποιούσαν temporal facts. Εξίσου σημαντική είναι και η αναφορά στον Passive Aggressive Classifier όπου έχει την ικανότητα να τροφοδοτείται με online δεδομένα. Τέλος, σε αρκετά πειράματα, η πρώιμη ανίχνευση αναφέρεται ως μελλοντική επέκταση.

Ένα ακόμα συμπέρασμα από τα στατιστικά των πειραμάτων, είναι ότι οι ερευνητές τείνουν να μην αναφέρουν τα specifications των υπολογιστικών τους συστημάτων ή τους χρόνους εκπαίδευσης με αποτέλεσμα να μην γνωρίζουν οι μελλοντικοί ερευνητές τις κατά προσέγγιση υπολογιστικές ανάγκες για τα συστήματα αυτά.

Η Γενίκευση είναι ακόμα μια ικανότητα που αναφέρεται συχνά από τους ερευνητές ως μελλοντική επέκταση των εργασιών τους. Βέβαια, υπήρχαν και πειράματα όπου τα συστήματα δοκιμαζόντουσαν και σε διαφορετικές θεματολογίες, με αποτέλεσμα να αποδεικνύουν τις ικανότητες γενίκευσης. Όσον αφορά την γενίκευση σε διαφορετικές γλώσσες, η δυνατότητα αυτή υπήρχε στο mBERT όπου είναι πολύ γλωσσικό μοντέλο.

4.12) Στατιστική Απεικόνιση των Πειραμάτων

Στα παρακάτω αρχεία αναφέρονται αναλυτικότερα τα 53 πειράματα που προέκυψαν έπειτα από το SLR. Αναφέρονται πληροφορίες όπως το Doi, η περίληψη των paper, η κατηγορία της μεθόδου ανίχνευσης, η μέθοδος ανίχνευσης, οι συγκρινόμενες μέθοδοι ανίχνευσης, το dataset που επιτεύχθηκε η καλύτερη αξιολόγηση, η διαθεσιμότητα του dataset, η προέλευση του dataset, η θεματολογία dataset, οι μετρικές αξιολογήσεις που κατέγραψαν οι ερευνητές και τα pros and cons που παρέχουν μεγάλο όγκο πληροφοριών. Οι πληροφορίες αυτές βρίσκονται στα 3 αρχεία που έχουν επισυναφθεί. Ένα Excel αρχείο και 2 pdf εξαγόμενα από το Google Spreadsheet.


Συστηματική Βιβλιογραφική Ανα


StatsCombined .pdf


Συστηματική%20Βιβλιογραφική%20Ανα

5)Συμπέρασμα

Ο τομέας της ανίχνευσης ψευδών ειδήσεων έχει προχωρήσει από τους παραδοσιακούς αλγορίθμους ταξινόμησης , στα νευρωνικά δίκτυα. Αρχικά υπήρχαν οι εξελιγμένες μορφές των RNNs και στην συνέχεια η μετεξέλιξη των νευρωνικων δικτύων όπου είναι οι Transformers. Ο τομέας της ανίχνευσης θεωρείται ένα πολύ δημοφιλές πεδίο με συνεχή ανοδική τάση σε δημοσιεύσεις ερευνών. Βάση της κοινής λογικής, οι ερευνητές θα στραφούν περισσότερο στην ανίχνευση πολυμορφικών ψευδών ειδήσεων καθώς είναι μια νέα ανάγκη που έχει εμφανιστεί στην εποχή που διανύουμε. Τα προηγμένα αυτά συστήματα όπως τα LLMs, πρόκειται να ισχυροποιηθούν περισσότερο. Όσον αφορά μερικά χρήσιμα συμπεράσματα στην μέχρι τώρα έρευνα στον τομέα ανίχνευσης text-based ψευδών ειδήσεων, θα πρέπει να αναπτυχθούν προσεγγίσεις ικανές για άμεση ή πρώιμη ανίχνευση και προσεγγίσεις όπου θα διαθέτουν δυνατότητες εξειδίκευσης στα μέσα κοινωνικής δικτύωσης. Εξίσου σημαντική, είναι η επιτακτική ανάγκη για μεγάλα σύνολα δεδομένων, τα πιο πρόσφατα συστήματα, βελτιώνουν τις ικανότητες τους όταν διαχειρίζονται υπέρογκα σύνολα δεδομένων παρόλο που προφανώς οι υπολογιστικές ανάγκες μεγαλώνουν. Επιπρόσθετα, τα Fake News παρόλο που αναγνωρίζονται ως πρόβλημα από την πλειοψηφία της κοινωνίας, συνεχίζουν να αναπτύσσονται[45]. Με το πέρασμα του χρόνου, οι επιτήδειοι βρίσκουν νέους τρόπου να κεντρίσουν το ενδιαφέρων των αναγνωστών και να τους ξεγελάσουν με παραπλανητικά κείμενα. Η ανάπτυξη της τεχνητής νοημοσύνης πέρα από το ότι δυναμώνει τον τομέα της ανίχνευσης, ισχυροποιεί και την παραγωγή των Fake News π.χ. η παραγωγή ψευδών αφηγήσεων με την χρήση του GPT. Μια τάση που παρατηρείται τους τελευταίους μήνες, είναι η παραγωγή ενημερωτικών ιστοσελίδων αλλά και λογαριασμών σε πλατφόρμες όπως το YouTube και το Twitter όπου δεν υπάρχει πρακτικά φυσική παρουσία content creator. Οι λογαριασμοί αυτοί, είναι αυτοματοποιημένα bots τα οποία δημιουργούν περιεχόμενο με βάση την επικαιρότητα και δημοσιεύουν υλικό ανά τακτά χρονικά διαστήματα.

Τέλος, είναι εύλογο να συζητηθεί ότι με τέτοιου είδους αξιολογήσεις όπως οι μέσοι όροι των 53^{ων} πειραμάτων που εμφανίζονται στο πίνακα παρακάτω, το πρόβλημα της ανίχνευσης των text-based ψευδών ειδήσεων έχει επιλυθεί.

AI Paradigms	Accuracy	F1	Recall	Precision
Hybrid	98.321	98.307	98.293	98.091
DL	98.709	98.535	98.493	98.347
Transformers	98.233	98.107	98.017	97.571
ML	98.350	98.201	98.048	98.755

6)Προτεινόμενη Μελλοντική Επέκταση της Εργασίας

Ο τομέας της τεχνητής νοημοσύνης εξελίσσεται ραγδαία. Αυτό προκαλεί την αυξητική τάση των δημοσιεύσεων που χρησιμοποιούν εργαλεία τεχνητής νοημοσύνης. Μια προτεινόμενη μελλοντική επέκταση της εργασίας θα μπορεί να είναι η επισκόπηση των επόμενων ερευνών από το έτος 2025 και έπειτα όπου όλο και περισσότερα LLMs θα αρχίσουν να χρησιμοποιούνται στον τομέα της ανίχνευση. Μια τέτοια έρευνα θα είναι ικανή να απαντήσει στο πόσο βελτιώθηκαν τα υπάρχοντα προβλήματα όπως η πρώιμη ανίχνευση και τι επιπρόσθετες ικανότητες έχουν ενταχθεί.

Ως μια δεύτερη επέκταση της εργασίας, προτείνεται η βιβλιογραφική ανασκόπηση των μεθόδων ανίχνευσης πολυμορφικών ειδήσεων όπως εικόνα , ήχος και βίντεο. Η αύξηση των AI generated videos αλλά και τον deep fakes είναι ένα γεγονός που προβληματίζει όλη την συνειδητοποιημένη κοινωνία πέρα από τους ακαδημαϊκούς ερευνητές.

7)Βιβλιογραφία

Παρακάτω υπάρχουν τα Papers που χρησιμοποιήθηκαν για την θεωρητική δόμηση της εργασίας.

[1] ‘Information disorder: Toward an interdisciplinary framework for research and policy making’, Council of Europe Publishing. Ημερομηνία πρόσβασης: 20 Ιούλιος 2025. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html

[2] H. Allcott και M. Gentzkow, ‘Social Media and Fake News in the 2016 Election’, *J. Econ. Perspect.*, τ. 31, τχ. 2, σελ. 211–236, Μαΐου 2017, doi: 10.1257/jep.31.2.211.

[3] A. Jarrahi και L. Safari, ‘Evaluating the effectiveness of publishers’ features in fake news detection on social media’, *Multimed. Tools Appl.*, τ. 82, τχ. 2, σελ. 2913–2939, Ιανουαρίου 2023, doi: 10.1007/s11042-022-12668-8.

[4] J. M. Burkhardt, ‘Chapter 1. History of Fake News’, *Libr. Technol. Rep.*, τ. 53, τχ. 8, Art. τχ. 8, Νοεμβρίου 2017.

[5] ‘(PDF) AN ANALYSIS OF SENSATIONALISM IN NEWS’, ResearchGate. Ημερομηνία πρόσβασης: 30 Ιούνιος 2025. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: https://www.researchgate.net/publication/374812735_AN_ANALYSIS_OF_SENSATIONALISM_IN_NEWS

[6] A. Doyle, ‘Tracing 'Fake News:' The Printing Press, Social Media, and Politics’.

[7] G. Fenza, V. Loia, C. Stanzione, και M. Di Gisi, ‘Robustness of models addressing Information Disorder: A comprehensive review and benchmarking study’, *Neurocomputing*, τ. 596, σελ. 127951, Σεπτεμβρίου 2024, doi: 10.1016/j.neucom.2024.127951.

[8] ‘(PDF) Algorithms, bots, and political communication in the US 2016 election: The challenge of automated political communication for election law and administration’, *ResearchGate*, Ιανουαρίου 2025, Ημερομηνία πρόσβασης: 30

Ιούνιος 2025. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: https://www.researchgate.net/publication/324464619_Algorithms_bots_and_political_communication_in_the_US_2016_election_The_challenge_of_automated_political_communication_for_election_law_and_administration
[9] S. Tufchi, A. Yadav, και T. Ahmed, 'A comprehensive survey of multimodal fake news detection techniques: advances, challenges, and opportunities', *Int. J. Multimed. Inf. Retr.*, τ. 12, τχ. 2, σελ. 28, Αυγούστου 2023, doi: 10.1007/s13735-023-00296-3.
[10] X. Zhou, R. Zafarani, K. Shu, και H. Liu, 'Fake News: Fundamental Theories, Detection Strategies and Challenges', στο *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*, στο WSDM '19. New York, NY, USA: Association for Computing Machinery, Ιανουαρίου 2019, σελ. 836–837. doi: 10.1145/3289600.3291382.
[11] S. H. Kong, L. M. Tan, K. H. Gan, και N. H. Samsudin, 'Fake News Detection using Deep Learning', στο *2020 IEEE 10th Symposium on Computer Applications & Industrial Electronics (ISCAIE)*, Απριλίου 2020, σελ. 102–107. doi: 10.1109/ISCAIE47305.2020.9108841.
[12] F. Monti, F. Frasca, D. Eynard, D. Mannion, και M. M. Bronstein, 'Fake News Detection on Social Media using Geometric Deep Learning', 10 Φεβρουάριος 2019, *arXiv*: arXiv:1902.06673. doi: 10.48550/arXiv.1902.06673.
[13] 'A survey of explainable AI techniques for detection of fake news and hate speech on social media platforms | Journal of Computational Social Science'. Ημερομηνία πρόσβασης: 30 Ιούνιος 2025. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: https://link.springer.com/article/10.1007/s42001-024-00248-9
[14] 'Fake news, disinformation and misinformation in social media: a review | Social Network Analysis and Mining'. Ημερομηνία πρόσβασης: 30 Ιούνιος 2025. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: https://link.springer.com/article/10.1007/s13278-023-01028-5
[15] K. Nagi, 'New Social Media and Impact of Fake News on Society', 6 Ιούνιος 2018, *Social Science Research Network, Rochester, NY*: 3258350. Ημερομηνία πρόσβασης: 20 Ιούλιος 2025. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: https://papers.ssrn.com/abstract=3258350
[16] M. C. Arcuri, G. Gandolfi, και I. Russo, 'Does fake news impact stock returns? Evidence from US and EU stock markets', *J. Econ. Bus.*, τ. 125–126, σελ. 106130, Μαΐου 2023, doi: 10.1016/j.jeconbus.2023.106130.
[17] J. L. Rojas Torrijos και M. S. Mello, 'Football Misinformation Matrix: A Comparative Study of 2020 Winter Transfer News in Four European Sports Media Outlets', *Journal. Media*, τ. 2, τχ. 4, Art. τχ. 4, Δεκεμβρίου 2021, doi: 10.3390/journalmedia2040037.
[18] '(PDF) Health Misinformation in Social Networks: A Survey of Information Technology Approaches', *ResearchGate*, Μαΐου 2025, doi: 10.3390/fi17030129.
[19] A. Bondielli και F. Marcelloni, 'A survey on fake news and rumour detection techniques', *Inf. Sci.*, τ. 497, σελ. 38–55, Σεπτεμβρίου 2019, doi: 10.1016/j.ins.2019.05.035.
[20] R. Oshikawa, J. Qian, και W. Y. Wang, 'A Survey on Natural Language Processing for Fake News Detection', 5 Μάρτιος 2020, *arXiv*: arXiv:1811.00770. doi: 10.48550/arXiv.1811.00770.

[21] Z. Khanam, B. N. Alwasel, H. Sirafi, και M. Rashid, ‘Fake News Detection Using Machine Learning Approaches’, *IOP Conf. Ser. Mater. Sci. Eng.*, τ. 1099, τχ. 1, σελ. 012040, Νοεμβρίου 2021, doi: 10.1088/1757-899X/1099/1/012040.
[22] J. Shaikh και R. Patil, ‘Fake News Detection using Machine Learning’, στο *2020 IEEE International Symposium on Sustainable Energy, Signal Processing and Cyber Security (iSSSC)*, Σεπτεμβρίου 2020, σελ. 1–5. doi: 10.1109/iSSSC50941.2020.9358890.
[23] S. I. Manzoor, J. Singla, και Nikita, ‘Fake News Detection Using Machine Learning approaches: A systematic Review’, στο *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)*, Απριλίου 2019, σελ. 230–234. doi: 10.1109/ICOEI.2019.8862770.
[24] M. F. Mridha, A. J. Keya, Md. A. Hamid, M. M. Monowar, και Md. S. Rahman, ‘A Comprehensive Review on Fake News Detection With Deep Learning’, *IEEE Access*, τ. 9, σελ. 156151–156170, 2021, doi: 10.1109/ACCESS.2021.3129329.
[25] J. Alghamdi, Y. Lin, και S. Luo, ‘Unveiling the hidden patterns: A novel semantic deep learning approach to fake news detection on social media’, *Eng. Appl. Artif. Intell.*, τ. 137, σελ. 109240, Νοεμβρίου 2024, doi: 10.1016/j.engappai.2024.109240.
[26] H. Thakar και B. Bhatt, ‘Fake news detection: recent trends and challenges’, *Soc. Netw. Anal. Min.*, τ. 14, τχ. 1, σελ. 176, Αυγούστου 2024, doi: 10.1007/s13278-024-01344-4.
[27] F. M. Costa και R. Guimaraes, ‘Fake News Detection using Natural Language Processing’.
[28] M. M. Jogi, ‘Syntactic and Semantic Analysis in Natural Language Processing: Unveiling the Underlying Mechanisms’.
[29] ‘(PDF) Systematic Review of Semantic Analysis Methods’, *ResearchGate*, doi: 10.52098/acj.2023346.
[30] ‘(PDF) Text Representations and Word Embeddings: Vectorizing Textual Data’, στο *ResearchGate*. doi: 10.1007/978-3-030-88389-8_16.
[31] A. A. A. Ahmed, A. Aljabouh, P. K. Donepudi, και M. S. Choi, ‘Detecting Fake News Using Machine Learning : A Systematic Literature Review’, 8 Φεβρουάριος 2021, *arXiv*: arXiv:2102.04458. doi: 10.48550/arXiv.2102.04458.
[32] A. Dubey, A. Lovewanshi, και N. Dhanda, ‘Fake News Detection Using Machine Learning’, στο *2024 4th International Conference on Technological Advancements in Computational Sciences (ICTACS)*, Νοεμβρίου 2024, σελ. 809–814. doi: 10.1109/ICTACS62700.2024.10841180.
[33] ‘(PDF) Fake News Detection Using Machine Learning Approaches’, *ResearchGate*, Ημερομηνία πρόσβασης: 1 Ιούλιος 2025. [Έκδοση σε ψηφιακή μορφή]. Διαθέσιμο στο: https://www.researchgate.net/publication/350082145_Fake_News_Detection_Using_Machine_Learning_Approaches
[34] A. Jain, A. Shakya, H. Khatter, και A. K. Gupta, ‘A smart System for Fake News Detection Using Machine Learning’, στο *2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT)*, Σεπτεμβρίου 2019, σελ. 1–4. doi: 10.1109/ICICT46931.2019.8977659.
[35] K. Crammer *κ.ά.*, ‘Online Passive-Aggressive Algorithms’.
[36] T. Chen και C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’, στο *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*

Discovery and Data Mining, Αυγούστου 2016, σελ. 785–794. doi: 10.1145/2939672.2939785.
[37] R. E. Schapire, 'Explaining AdaBoost', στο *Empirical Inference*, B. Schölkopf, Z. Luo, και V. Vovk, Επιμ., Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, σελ. 37–52. doi: 10.1007/978-3-642-41136-6_5.
[38] J. Devlin, M.-W. Chang, K. Lee, και K. Toutanova, 'BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding', 24 Μάιος 2019, *arXiv*: arXiv:1810.04805. doi: 10.48550/arXiv.1810.04805.
[39] A. Vaswani *κ.ά.*, 'Attention Is All You Need', 2 Αύγουστος 2023, *arXiv*: arXiv:1706.03762. doi: 10.48550/arXiv.1706.03762.
[40] Y. Liu *κ.ά.*, 'RoBERTa: A Robustly Optimized BERT Pretraining Approach', 26 Ιούλιος 2019, *arXiv*: arXiv:1907.11692. doi: 10.48550/arXiv.1907.11692.
[41] V. Sanh, L. Debut, J. Chaumond, και T. Wolf, 'DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter', 1 Μάρτιος 2020, *arXiv*: arXiv:1910.01108. doi: 10.48550/arXiv.1910.01108.
[42] M. Müller, M. Salathé, και P. E. Kummervold, 'COVID-Twitter-BERT: A Natural Language Processing Model to Analyse COVID-19 Content on Twitter', 15 Μάιος 2020, *arXiv*: arXiv:2005.07503. doi: 10.48550/arXiv.2005.07503.
[43] T. Pires, E. Schlinger, και D. Garrette, 'How multilingual is Multilingual BERT?', 4 Ιούνιος 2019, *arXiv*: arXiv:1906.01502. doi: 10.48550/arXiv.1906.01502.
[44] W. Antoun, F. Baly, και H. Hajj, 'AraBERT: Transformer-based Model for Arabic Language Understanding', 7 Μάρτιος 2021, *arXiv*: arXiv:2003.00104. doi: 10.48550/arXiv.2003.00104.
[45] N. L. Bragazzi και S. Garbarino, 'Understanding and Combating Misinformation: An Evolutionary Perspective', *JMIR Infodemiology*, τ. 4, τχ. 1, σελ. e65521, Δεκεμβρίου 2024, doi: 10.2196/65521.

7.1 Βιβλιογραφία πειραμάτων μέσω SLR

Παρακάτω υπάρχει η βιβλιογραφία των 53ων πειραμάτων που αναλύθηκαν, για περισσότερες πληροφορίες μπορείτε να χρησιμοποιήσετε τα αρχεία που βρίσκονται στην ενότητα 4.12.

1) Hossain, Md. Rajib, Mohammed Moshiul Hoque, Nazmul Siddique, και M. Ali Akber Dewan. 'AraCovTexFinder: Leveraging the transformer-based language model for Arabic COVID-19 text identification'. Engineering Applications of Artificial Intelligence 133 (1 Ιούλιος 2024): 107987. https://doi.org/10.1016/j.engappai.2024.107987.

2) 'Unveiling the hidden patterns: A novel semantic deep learning approach to fake news detection on social media'. Engineering Applications of Artificial Intelligence 137 (1 Νοέμβριος 2024): 109240.

https://doi.org/10.1016/j.engappai.2024.109240.

3) Yang, Chang, Xia Yu, JiaYi Wu, BoZhen Zhang, και HaiBo Yang. ‘Graph-aware multi-feature interacting network for explainable rumor detection on social network’. Expert Systems with Applications 249 (1 Σεπτέμβριος 2024): 123687. https://doi.org/10.1016/j.eswa.2024.123687.

4) Alnabrisi, Israa kh., και Motaz kh. Saad. ‘Detect Arabic fake news through deep learning models and Transformers’. Expert Systems with Applications 251 (1 Οκτώβριος 2024): 123997. https://doi.org/10.1016/j.eswa.2024.123997.

5) Dev, Deepali Goyal, Vishal Bhatnagar, Bhoopesh Singh Bhati, Manoj Gupta, και Aziz Nanthaamornphong. ‘LSTMCNN: A hybrid machine learning model to unmask fake news’. Heliyon 10, τχ. 3 (15 Φεβρουάριος 2024): e25244. https://doi.org/10.1016/j.heliyon.2024.e25244.

6) Dhiman, Pummy, Amandeep Kaur, Deepali Gupta, Sapna Juneja, Ali Nauman, και Ghulam Muhammad. ‘GBERT: A hybrid deep learning model based on GPT-BERT for fake news detection’. Heliyon 10, τχ. 16 (30 Αύγουστος 2024): e35865. https://doi.org/10.1016/j.heliyon.2024.e35865.

7) Bhardwaj, Akashdeep, Salil Bharany, και SeongKi Kim. ‘Fake social media news and distorted campaign detection framework using sentiment analysis & machine learning’. Heliyon 10, τχ. 16 (30 Αύγουστος 2024): e36049. https://doi.org/10.1016/j.heliyon.2024.e36049.

8) Sudhakar, M., και K. P. Kaliyamurthie. ‘Detection of fake news from social media using support vector machine learning algorithms’. Measurement: Sensors 32 (1 Απρίλιος 2024): 101028. https://doi.org/10.1016/j.measen.2024.101028.

9) Bellam, Thilak, και P Lakshmi Prasanna. ‘Novel DeepLearning based sentimental approach to identifying the fake news in social networking media based smart application’. Measurement: Sensors 33 (1 Ιούνιος 2024): 101148. https://doi.org/10.1016/j.measen.2024.101148.

10) Elfaik, Hanane, και El Habib Nfaoui. ‘Automatic Detection of Fake News Using Gated Recurrent Unit Deep Model’. Procedia Computer Science, 5th International Conference on Innovative Data Communication Technologies and Application (ICIDCA 2024), 233 (1 Ιανουάριος 2024): 474–80. https://doi.org/10.1016/j.procs.2024.03.237.

11) Hashmi, Ehtesham, Sule Yildirim Yayilgan, Muhammad Mudassar Yamin, Subhan Ali, και Mohamed Abomhara. ‘Advancing Fake News Detection: Hybrid Deep Learning With FastText and Explainable AI’. IEEE Access 12 (2024): 44462–80. https://doi.org/10.1109/ACCESS.2024.3381038.

12) Jian, Wang, Jian Ping Li, Muhammad Atif Akbar, Amin Ul Haq, Shakir Khan, Reemiah Muneer Alotaibi, και Saad Abdullah Alajlan. ‘SA-Bi-LSTM: Self Attention With Bi-Directional LSTM-Based Intelligent Model for Accurate Fake News Detection to Ensured Information Integrity on Social Media Platforms’. IEEE Access 12 (2024): 48436–52. https://doi.org/10.1109/ACCESS.2024.3382832.

13) Aljrees, Turki. ‘Improving Prediction of Arabic Fake News Using ELMO’s Features-Based Tri-Ensemble Model and LIME XAI’. IEEE Access 12 (2024): 63066–76. https://doi.org/10.1109/ACCESS.2024.3392297.

14) Al-Quayed, Fatima, Danish Javed, N. Z Jhanjhi, Mamoona Humayun, και Thanaa S. Alnusairi. ‘A Hybrid Transformer-Based Model for Optimizing Fake News Detection’. IEEE Access 12 (2024): 160822–34. https://doi.org/10.1109/ACCESS.2024.3476432.

15) Raza, Ali, Shafiq Ur Rehman Khan, Raja Sher Afgun Usmani, Ashok Kumar Das, και Shehzad Ashraf Chaudhry. ‘A Novel Temporal Footprints-Based Framework for Fake News Detection’. IEEE Access 12 (2024): 172419–28. https://doi.org/10.1109/ACCESS.2024.3490558.

16) Aggarwal, Nikhil, και Siddharth Kumar. ‘An Optimized TF-IDF Features based Deep Learning Enabled Automated Fake News Detection Scheme’. Στο 2024 Asia Pacific Conference on Innovation in Technology (APCIT), 1–9, 2024.https://doi.org/10.1109/APCIT62007.2024.10673703.

17) Bohra, Manvi, Indrajeet Kumar, και Kamred Udham Singh. ‘Fake News Detection Using Different Machine Learning Algorithms’. Στο 2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT),

45–50, 2024. https://doi.org/10.1109/DICCT61038.2024.10533120.

18) G, Elizabeth Rani, Kiruthikha D, και Sakthimohan M. ‘Boosting Fake News Detection Accuracy: A Deep Dive into LSTM Classifiers’. Στο 2024 10th International Conference on Advanced Computing and Communication Systems (ICACCS), 1:662–68,

2024. https://doi.org/10.1109/ICACCS60874.2024.10717086.

19) Khan, Ishmam, Azmain Yakin Srizon, Md. Farukuzzaman Faruk, Md. Al Mamun, S. M. Mahedy Hasan, Md Nazmul Islam, Md. Rakib Hossain, και Md Faruk Hossain. ‘Leveraging the Robust Capability of Modified Multi-Layer Gated Recurrent Units for Fake News Detection’. Στο 2024 6th International Conference on Electrical Engineering and Information & Communication Technology (ICEEICT), 986–91, 2024. https://doi.org/10.1109/ICEEICT62016.2024.10534407.

20) Mahmud, Tanjim, Tahmina Akter, Mohammad Tarek Aziz, Mohammad Kamal Uddin, Mohammad Shahadat Hossain, και Karl Andersson. ‘Integration of NLP and Deep Learning for Automated Fake News Detection’. Στο 2024 Second International Conference on Inventive Computing and Informatics (ICICI), 398–404, 2024. https://doi.org/10.1109/ICICI62254.2024.00072.

21) Asha, V., A. P Nirmala, A Arvind Raj, Akash Yadav, Abhijit Sen, και Abhinav Rajan. ‘An Effective Assessment of Machine Learning Approaches for Fake News Detection’. Στο 2024 International Conference on Inventive Computation Technologies (ICICT), 169–74, 2024. https://doi.org/10.1109/ICICT60155.2024.10544858.

22) Johnson, Deepika Roselind, R Srivats, Kalyanasundaram V, και Saimirra R. ‘A Multichannel CNN-LSTM Approach for Fake News Detection: A Comparative Study’. Στο 2024 International Conference on Integration of Emerging Technologies for the Digital World (ICIETDW), 1–6, 2024. https://doi.org/10.1109/ICIETDW61607.2024.10939885.

23) Kamble, Riya Ajay, Varda Pareek, και Tarun Jain. ‘Fake News Detection Using Machine Learning and Deep Learning’. Στο 2024 IEEE International Conference on Smart Power Control and Renewable Energy (ICSPCRE), 1–6, 2024. https://doi.org/10.1109/ICSPCRE62303.2024.10674937.

24) Mahmud, Tanjim, Imran Hasan, Mohammad Tarek Aziz, Taohidur Rahman, Mohammad Shahadat Hossain, και Karl Andersson. ‘Enhanced Fake News Detection through the Fusion of Deep Learning and Repeat Vector Representations’. Στο 2024 2nd International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), 654–60, 2024. https://doi.org/10.1109/IDCIoT59759.2024.10467839.

25) Balogun, Temitayo Elijah, Samson Adebowale Abosede, Shade Christiana Joda, Oluwaseyifunmi Balogun, Paul Kehinde Olotu, και Samuel Gbenga Faluyi. ‘Machine Learning Approaches for Detecting and Mitigating the Impact of Fake News in Online Information Ecosystems’. Στο 2024 International Conference on Science, Engineering and Business for Driving Sustainable Development Goals (SEB4SDG), 1–6, 2024. https://doi.org/10.1109/SEB4SDG60871.2024.10630248.

26) Sanida, Maria Vasiliki, Theodora Sanida, Argyrios Sideris, Michael Dossis, και Minas Dasygenis. ‘Fake News Detection Approach Using Hybrid Deep Learning Framework’. Στο 2024 9th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM), 81–84, 2024. https://doi.org/10.1109/SEEDA-CECNSM63478.2024.00023.

27) Aljawarneh, Shadi A., και Safa Ahmad Swedat. ‘Fake News Detection Using Enhanced BERT’. IEEE Transactions on Computational Social Systems 11, τχ. 4 (Δεκέμβριος 2024): 4843–50. https://doi.org/10.1109/TCSS.2022.3223786.

28) Roy, Pradeep Kumar, Asis Kumar Tripathy, Tien-Hsiung Weng, και Kuan-Ching Li. ‘Securing social platform from misinformation using deep learning’. Computer Standards & Interfaces 84 (1 Μάρτιος 2023): 103674. https://doi.org/10.1016/j.csi.2022.103674.

29) Puraivan, Eduardo, René Venegas, και Fabián Riquelme. ‘An empiric validation of linguistic features in machine learning models for fake news detection’. Data &

Knowledge Engineering 147 (1 Σεπτέμβριος 2023): 102207. https://doi.org/10.1016/j.datak.2023.102207.

30) Alghamdi, Jawaher, Yuqing Lin, και Suhuai Luo. 'Towards COVID-19 fake news detection using transformer-based models'. Knowledge-Based Systems 274 (15 Αύγουστος 2023): 110642. https://doi.org/10.1016/j.knosys.2023.110642.

31) Lakshman, V., V. Mounika, P. Tritisha, Sk. Firduas Fathima, και Nafisa Begum. 'Fake News Spotting using Interrelated Feature Selection Model using Logistic Regression'. Στο 2023 IEEE 12th International Conference on Communication Systems and Network Technologies (CSNT), 924–29, 2023. https://doi.org/10.1109/CSNT57126.2023.10134586.

32) Muduli, Debendra, Santosh Kumar Sharma, Dinesh Kumar, Akshat Singh, και Shubham Kumar Srivastav. 'Maithi-Net: A Customized Convolution Approach for Fake News Detection using Maithili Language'. Στο 2023 International Conference on Computer, Electronics & Electrical Engineering & their Applications (IC2E3), 1–6, 2023. https://doi.org/10.1109/IC2E357697.2023.10262664.

33) Rakib Mollah, Md. Abdur, Mir Md. Jahangir Kabir, Monika Kabir, και Md. Sazid Reza. 'Detection of Fake News with RoBERTa Based Embedding and Modified Deep Neural Network Architecture'. Στο 2023 26th International Conference on Computer and Information Technology (ICCIT), 1–6, 2023. https://doi.org/10.1109/ICCIT60459.2023.10441206.

34) Misra, Shamik, Prabhat Tandon, και Purna Chandra Panda. 'Optimization of Hugging Face Transformers for Fake News Detection'. Στο 2023 International Conference on Data Science, Agents & Artificial Intelligence (ICDSAAI), 1–5, 2023. https://doi.org/10.1109/ICDSAAI59313.2023.10452606.

35) Ramesh Reddy, B, T Tejaswi, P Naveen, J Giresh, και B Vamsi. 'Optimising the Detection of Fake News in Multilingual Exposition using Machine Learning Techniques'. Στο 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), 158–65, 2023. https://doi.org/10.1109/ICICCS56967.2023.10142901.

36) Chen, Mu-Yen, Guan-Ming Lin, Yi-Wei Lai, και Hsiu-Sen Chiang. ‘Cognitive COVID-19 Fake News Detection Model based on Machine Learning Approach’. Στο 2023 International Conference on Networking, Sensing and Control (ICNSC), 1:1–5, 2023. https://doi.org/10.1109/ICNSC58704.2023.10319025.

37) Kanchana, M, Vel Murugesh Kumar, T.P Anish., και Pv Gopirajan. ‘Deep Fake BERT: Efficient Online Fake News Detection System’. Στο 2023 International Conference on Networking and Communications (ICNWC), 1–6, 2023. https://doi.org/10.1109/ICNWC57852.2023.10127560.

38) Kai Xuan, Kelvin Liew, Mohaiminul Islam Bhuiyan, Nur Shazwani Kamarudin, Ahmad Fakhri Ab. Nasir, και Muhammad Zulfahmi Toh Abdullah. ‘COVID-19 Fake News Detection Model on Social Media Data Using Machine Learning Techniques’. Στο 2023 IEEE 8th International Conference On Software Engineering and Computer Systems (ICSECS), 28–34, 2023. https://doi.org/10.1109/ICSECS58457.2023.10256386.

39) Rohman, Shawly, Jannatul Ferdous, Sheikh Md. Rezone Ullah, και Md. Aminur Rahman. ‘IBFND: An Improved Dataset for Bangla Fake News Detection and Comparative Analysis of Performance of Baseline Models’. Στο 2023 International Conference on Next-Generation Computing, IoT and Machine Learning (NCIM), 1–6, 2023. https://doi.org/10.1109/NCIM59001.2023.10212799.

40) Sultana, Achhiya, Mahmudul Islam, Mahady Hasan, και Farruk Ahmed. ‘Fake News Detection Using Machine Learning Techniques’. Στο 2023 IEEE/ACIS 21st International Conference on Software Engineering Research, Management and Applications (SERA), 98–103, 2023. https://doi.org/10.1109/SERA57763.2023.10197712.

41) ‘Enhancing Fake News Detection Using Classification Algorithms and Deep Learning’. Στο ResearchGate. Ημερομηνία πρόσβασης 26 Ιούνιος 2025. https://doi.org/10.1109/UPCON59197.2023.10434370.

42) ‘Tackling Disinformation: Machine Learning Solutions for Fake News Detection | Proceedings of the 2023 13th International Conference on Communication and Network Security’. Ημερομηνία πρόσβασης 26 Ιούνιος 2025. https://dl.acm.org/doi/10.1145/3638782.3638790.

43) ‘(PDF) Automated COVID-19 Misinformation Checking System Using Encoder Representation with Deep Learning Models’. ResearchGate. Ημερομηνία πρόσβασης 26 Ιούνιος 2025. https://doi.org/10.11591/ijai.v12.i1.pp488-495.

44) ‘(PDF) A novel approach to fake news classification using LSTM-based deep learning models’. Ημερομηνία πρόσβασης 26 Ιούνιος 2025. https://www.researchgate.net/publication/377632641_A_novel_approach_to_fake_news_classification_using_LSTM-based_deep_learning_models.

45) ‘Fake News Detection Model on Social Media by Leveraging Sentiment Analysis of News Content and Emotion Analysis of Users’ Comments’. Ημερομηνία πρόσβασης 26 Ιούνιος 2025. https://www.mdpi.com/1424-8220/23/4/1748.

46) Hayawi, K., S. Shahriar, M. A. Serhani, I. Taleb, και S. S. Mathew. ‘ANTi-Vax: a novel Twitter dataset for COVID-19 vaccine misinformation detection’. Public Health 203 (1 Φεβρουάριος 2022): 23–30. https://doi.org/10.1016/j.puhe.2021.11.022.

47) ‘Improved Fake News Detection Method based on Deep Learning and Comparative Analysis with other Machine Learning approaches | IEEE Conference Publication | IEEE Xplore’. Ημερομηνία πρόσβασης 26 Ιούνιος 2025. https://ieeexplore.ieee.org/document/10007214.

48) Sultana, Raquiba, και Tetsuro Nishino. ‘Fake News Detection Using Transformer and Ensemble Learning Models’. Στο 2022 13th International Congress on Advanced Applied Informatics Winter (IIAI-AAI-Winter), 181–88, 2022. https://doi.org/10.1109/IIAI-AAI-Winter58034.2022.00044.

49) Reddy, S Sweta, Santanu Mandal, Varanasi L V S K B Kasyap, και Aswathy R K. ‘A Novel Approach to Detect Fake News Using eXtreme Gradient Boosting’. Στο 2022 10th International Symposium on Digital Forensics and Security (ISDFS), 1–4, 2022. https://doi.org/10.1109/ISDFS55398.2022.9800777.

50) *** TITLE NOT FOUND IN SOURCE FILE *** Fake News Detection in the Medical Field Using Machine Learning Techniques

51) Pavlov, Tashko, και Georgina Mirceva. ‘COVID-19 Fake News Detection by Using BERT and RoBERTa models’. Στο 2022 45th Jubilee International Convention on Information, Communication and Electronic Technology (MIPRO), 312–16, 2022. https://doi.org/10.23919/MIPRO55190.2022.9803414.

52) ‘Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique’. Ημερομηνία πρόσβασης 26 Ιούνιος 2025. https://www.mdpi.com/1424-8220/22/18/6970.

53) Mc-DNN: Fake News Detection Using MultiChannel Deep Neural Networks https://www.researchgate.net/publication/358794382_Mc-DNN_Fake_News_Detection_Using_Multi-Channel_Deep_Neural_Networks