ΠΑΝΕΠΙΣΤΗΜΙΟ
ΑΙΓΑΙΟΥ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΑΚΩΝ
ΚΑΙ ΕΠΙΚΟΙΝΩΝΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

Επισκόπηση Ανίχνευσης και Αντιμετώπισης Ψευδών Ειδήσεων

Ευάγγελος Σέρτης

321/2016177

Επιβλέπων: Αν. Καθηγητής Κυριάκος Κρητικός

Μέλη Επιτροπής Εξέτασης: Αν. Καθηγητής Παναγιώτης Συμεωνίδης, Αν. Καθηγητής Θεόδωρος Κωστούλας

Σάμος 2026

Πίνακας Περιεχομένων

Περίληψη 11
1. Εισαγωγή 13
1.1. Ψευδείς ειδήσεις και τεχνολογικές εξελίξεις 13
1.2. Συνεισφορά της εργασίας 14
1.3 Δομή αναφοράς 15
2. Θεωρητικό υπόβαθρο 17
2.1. Ορισμός ψευδών ειδήσεων 17
2.2. Τύποι ψευδών ειδήσεων 17
2.2.1. Παραπληροφόρηση 18
2.2.2. Παραποιημένο περιεχόμενο 22
2.3. Κύρια χαρακτηριστικά των ψευδών ειδήσεων 26
2.3.1. Κατασκευασμένες ειδήσεις και στρατηγικές παραπληροφόρησης 26
2.3.2. Συναισθηματισμοί και συναισθηματική έκκληση 26
2.3.3. Έλλειψη διαφάνειας και επαλήθευσης 27
2.3.4. Μεροληψία επιβεβαίωσης και θάλαμοι απήχησης 27
2.3.5. Ταχεία εξάπλωση και virality 27
2.3.6. Πρόθεση παραπλάνησης ή χειραγώγησης 27
2.3.7. Προκλήσεις για τη γνώση τρόπου χρήσης των μέσων ενημέρωσης 28
2.4. Γιατί οι ψευδείς ειδήσεις είναι πιο επίκαιρες από ποτέ; 28
2.4.1. Διάδοση των ψηφιακών μέσων 28
2.4.2. Ευκολία ανταλλαγής πληροφοριών 28
2.4.3. Αλγοριθμική ενίσχυση 29
2.4.4. Ψυχολογικοί παράγοντες 29
2.4.5. Κομματική πόλωση 29
2.4.6. Εκστρατείες επιρροής και τακτικές παραπληροφόρησης 30
2.4.7. Οικονομία της προσοχής 30
2.4.8. Παρακμή των παραδοσιακών “φρουρών” των μέσων ενημέρωσης 30
2.4.9. Παγκόσμιες κρίσεις 31
2.5. Τρόποι διάδοσης ψευδών ειδήσεων 31
2.5.1. Κοινοποίηση στα μέσα κοινωνικής δικτύωσης 31
2.5.2. Clickbait τίτλοι 32
2.5.3. Εκμετάλλευση των αλγορίθμων προώθησης περιεχομένου 32

2.5.4. Παραποιημένο οπτικοακουστικό υλικό 34
2.5.5. Αυτοματοποιημένα bots κοινωνικής δικτύωσης. 34
2.5.6. Φαινόμενο θαλάμου απήχησης 35
2.5.7. Στρατηγικές εκστρατείες παραπληροφόρησης 35
2.6. Μέσα διάδοσης ψευδών ειδήσεων 36
2.6.1. Πλατφόρμες κοινωνικής δικτύωσης 37
2.6.2. Botnets και αυτοματοποιημένοι λογαριασμοί 37
2.6.3. Ιστοσελίδες παραπληροφόρησης 37
2.6.4. Εφαρμογές μηνυμάτων 38
2.7. Υπαίτιοι ψευδών ειδήσεων 38
2.7.1. Κακόβουλοι δράστες και κρατικά υποστηριζόμενες εκστρατείες παραπληροφόρησης 38
2.7.2. Οντότητες με γνώμονα το κέρδος και clickbait farms 39
2.7.3. Κομματικές και ιδεολογικές ομάδες 39
2.7.4. Φάρσες, τρολ και παραπλανημένα άτομα 40
2.7.5. Αλγόριθμοι κοινωνικής δικτύωσης ως ενισχυτές ψευδών ειδήσεων 41
2.7.6. Ανεύθυνα μέσα ενημέρωσης 42
2.8. Κίνητρα και σκοποί στη διάδοση ψευδών ειδήσεων 43
2.8.1. Προώθηση πολιτικών γραμμών και ιδεολογιών 43
2.8.2. Οικονομικό κέρδος 44
2.8.3. Προσωπική προβολή και κοινωνική επιρροή 44
2.8.4. Πρόκληση χάους και σύγχυσης 45
2.8.5. Προκατάληψη της επιβεβαίωσης 46
2.9. Έλεγχος και τρόποι περιορισμού διάδοσης ψευδών ειδήσεων 47
2.9.1. Έλεγχος αληθείας αναφερόμενων γεγονότων 47
2.9.2. Κριτική αξιολόγηση στα μέσα ενημέρωσης 47
2.9.3. Αξιοποίηση τεχνολογίας και αυτοματοποιημένων εργαλείων 47
2.9.4. Αλγοριθμικές αλλαγές 48
2.9.5. Συνεργασία και συμπράξεις 48
2.9.6. Ρυθμιστικά μέτρα 48
2.9.7. Εκστρατείες ευαισθητοποίησης του κοινού 48
2.10. Συνέπειες, εμπόδια και αντίκτυπο των ψευδών ειδήσεων 48
2.10.1. Παραπληροφόρηση και εξαπάτηση 49
2.10.2. Διάβρωση της εμπιστοσύνης στα μέσα ενημέρωσης και τους θεσμούς 50
2.10.3. Κοινωνική πόλωση και διαίρεση 51

2.10.4. Κίνδυνοι για τη δημόσια υγεία 51
2.10.5. Πολιτική χειραγώγηση 52
3. Μεθοδολογία βιβλιογραφικής ανασκόπησης 53
3.1. Ερευνητικός στόχος και ερωτήματα 53
3.2. Διαδικασία αναζήτησης 54
3.3. Φιλτράρισμα άρθρων 55
3.4. Ποσοτική Ανάλυση. 58
3.5. Απαντήσεις στα ερευνητικά ερωτήματα 61
4. Ανάλυση 63
4.1.Κατηγοριοποίηση προσεγγίσεων ανίχνευσης ψευδών ειδήσεων 63
4.2. Ανίχνευση ψευδών ειδήσεων 64
4.2.1. Είδη ανίχνευσης ψευδών ειδήσεων 64
4.2.2. Είδη χαρακτηριστικών εστίασης 67
4.2.3. Τεχνική ανίχνευσης 73
4.3. Συγκριτική ανάλυση ερευνητικών προσεγγίσεων ανίχνευσης 75
4.4. Πρόληψη ψευδών ειδήσεων 80
4.4.1 Κατηγοριοποίηση προσεγγίσεων πρόληψης ψευδών ειδήσεων 80
4.4.2. Παρεμπόδιση επιρροής ατόμων σε κοινωνικά δίκτυα 81
4.4.3. Καμπάνιες διάδοσης της αλήθειας 83
4.4.4. Τεχνολογικές λύσεις 85
4.5. Συγκριτική ανάλυση ερευνητικών προσεγγίσεων πρόληψης 86
4.6. Σύνοψη συμπερασμάτων και συγκριτική αποτίμηση. 91
5. Ερευνητικά κενά 93
5.1. Ψευδές ειδήσεις παραγμένες από LLMs. 93
5.2. Επεξηγηματικότητα και διαφάνεια 94
5.3. Ανίχνευση σε πραγματικό χρόνο 96
5.4. Πολυγλωσσικότητα 99
5.5. Περιορισμοί των νευρωνικών δικτύων γράφων σε μεγάλα/δυναμικά κοινωνικά δίκτυα 100
6. Σύνολα Δεδομένων 102
6.1 Κατηγορίες συνόλων δεδομένων 102
6.1.1. LIAR dataset 102
6.1.2. FakeNewsNet dataset 102
6.1.3. BuzzFeed dataset 102
6.1.4. BS Detector 102
6.1.5. PHEME dataset 103

6.1.6. ISOT dataset 103
6.1.7. Fake News Corpus dataset 103
6.1.8. Hyperpartisan News dataset 103
6.2 Συγκριτική ανάλυση συνόλων δεδομένων 103
7. Συμπεράσματα και επίλογος 107
7.1 Συνεισφορά και συμπεράσματα από την εργασία 107
7.2 Σημαντικότερα ερευνητικά κενά και ερευνητικές κατευθύνσεις 108
7.3 Προσωπική εμπειρία που αποκομίσθηκε 108
Βιβλιογραφία 110

Κατάλογος Εικόνων:

Εικόνα 1: Ιεραρχικό δέντρο κατηγοριών ανίχνευσης……..……..63

Εικόνα 2: Συγκεντρωτική ανάλυση άρθρων για ανίχνευση ψευδών ειδήσεων……..……..76

Εικόνα 3: Ιεραρχικό δέντρο κατηγοριών πρόληψης……..80

Εικόνα 4: Συγκεντρωτική ανάλυση άρθρων για αντιμετώπιση ψευδών ειδήσεων…….....87

Κατάλογος Σχημάτων:

Σχήμα 1: Κατανομή πηγών ως προς το είδος τους (n=35)……59

Σχήμα 2: Εστιασμένη κατανομή πηγών (n=35)…….60

Σχήμα 3: Κατανομή πηγών ως προς το έτος (n=35)…….....61

Κατάλογος Πινάκων:

Πίνακας 1: Κατηγορίες ψευδών ειδήσεων……18

Πίνακας 2: Συγκριτική ανάλυση συνόλων δεδομένων……102

Πίνακας Ακρωνυμίων

Ακρωνύμιο	Full Term (English)	Επεξήγηση (Ελληνικά)
AI	Artificial Intelligence	Τεχνητή Νοημοσύνη
BERT	Bidirectional Encoder Representations from Transformers	Μοντέλο γλωσσικής κατανόησης βασισμένο σε Transformers
CNN	Convolutional Neural Network	Συνελικτικό Νευρωνικό Δίκτυο
CoT	Chain of Thought	Αλυσίδα σκέψης για ενίσχυση συλλογιστικής μοντέλων
GAN	Generative Adversarial Network	Παραγωγικό Ανταγωνιστικό Νευρωνικό Δίκτυο
GNN	Graph Neural Network	Νευρωνικό Δίκτυο Γραφημάτων
GPT	Generative Pre-trained Transformer	Παραγωγικό προεκπαιδευμένο μοντέλο Transformer
LCRB	Lexicon-based and Context Rule-Based	Υβριδική λεξικοκεντρική και συμφραστική μέθοδος
LIWC	Linguistic Inquiry and Word Count	Εργαλείο ανάλυσης γλωσσικού περιεχομένου
LLM	Large Language Model	Μεγάλο Γλωσσικό Μοντέλο
LSTM	Long Short-Term Memory	Τύπος Αναδρομικού Νευρωνικού Δικτύου με μακροχρόνια μνήμη
ML	Machine Learning	Μηχανική Μάθηση
NLI	Natural Language Inference	Συμπερασματολογία Φυσικής Γλώσσας
NLP	Natural Language Processing	Επεξεργασία Φυσικής Γλώσσας
NP-complete	Nondeterministic Polynomial-time Complete	Κλάση προβλημάτων ισοδύναμων δυσκολίας με τα NP
NP-hard	Nondeterministic Polynomial-time Hard	Προβλήματα τουλάχιστον τόσο δύσκολα όσο τα NP-complete
OSN	Online Social Network	Διαδικτυακό Κοινωνικό Δίκτυο
P	Polynomial Time	Κλάση προβλημάτων που επιλύονται σε πολυωνυμικό χρόνο
PCA	Principal Component Analysis	Ανάλυση Κύριων Συνιστωσών
POS	Part of Speech	Μέρος του Λόγου (γραμματική κατηγορία)
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses	Πρότυπο αναφοράς για συστηματικές ανασκοπήσεις
RAG	Retrieval-Augmented Generation	Παραγωγή κειμένου με ενσωμάτωση ανάκτησης πληροφορίας
RNN	Recurrent Neural Network	Αναδρομικό Νευρωνικό Δίκτυο
SEO	Search Engine Optimization	Βελτιστοποίηση για Μηχανές Αναζήτησης
SERPs	Search Engine Results Pages	Σελίδες Αποτελεσμάτων Μηχανών Αναζήτησης
SLR	Systematic Literature Review	Συστηματική Ανασκόπηση Βιβλιογραφίας
STEM	Science, Technology, Engineering, and Mathematics	Επιστήμη, Τεχνολογία, Μηχανική και Μαθηματικά
SVM	Support Vector Machine	Μηχανή Διανυσμάτων Υποστήριξης
TLDR	Too Long; Didn't Read	Πολύ μεγάλο· δεν το διάβασα (σύνοψη)
TR	Text Retrieval	Ανάκτηση Κειμένου
URL	Uniform Resource Locator	Διεύθυνση Πόρου στο Διαδίκτυο
XAI	Explainable Artificial Intelligence	Επεξηγήσιμη Τεχνητή Νοημοσύνη

Περίληψη

Η ανίχνευση και η πρόληψη ψευδών ειδήσεων αποτελούν σημαντικές προκλήσεις στο σημερινό ψηφιακό τοπίο, όπου η παραπληροφόρηση μπορεί να διαδοθεί γρήγορα και ευρέως μέσω των μέσων κοινωνικής δικτύωσης και των διαδικτυακών πλατφορμών. Συνεπώς, είναι επίμονη η ανάγκη μελέτης των διαφόρων προκλήσεων των ψευδών ειδήσεων σε συνδυασμό με τις επιτυχημένες τεχνικές καταπολέμησής τους, τόσο από τεχνολογική όσο και από κοινωνική άποψη. Η παρούσα διατριβή αποσκοπεί στην παροχή μιας ολοκληρωμένης ανάλυσης της τρέχουσας κατάστασης των ψευδών ειδήσεων, περιγράφοντας διεξοδικά το ιστορικό, τους τύπους και τα χαρακτηριστικά των ψευδών ειδήσεων, τις στρατηγικές διάδοσης και τις επιπτώσεις τους. Επιπλέον συνεισφέρει με μια βιβλιογραφική ανασκόπηση που επικεντρώνεται στην κατηγοριοποίηση, ανάλυση και σύγκριση των στρατηγικών ανίχνευσης και πρόληψης που σχετίζονται με τις ψευδείς ειδήσεις, προσδιορίζοντας σημαντικές εξελίξεις που σχετίζονται κυρίως με την ταχεία πρόοδο της τεχνητής νοημοσύνης και των αλγορίθμων μηχανικής μάθησης. Κατά τη διάρκεια της μελέτης μας, αναγνωρίσαμε ορισμένα ερευνητικά κενά, τα οποία αναλύσαμε και παρείχαμε για αυτά κατευθύνσεις έρευνας προς αντιμετώπισή τους. Συνολικά, φιλοδοξούμε ότι η παρούσα διατριβή να αποτελέσει σημείο αναφοράς για μελλοντικές προσπάθειες προς την κατεύθυνση της προόδου της κατάστασης στον τομέα της ανίχνευσης και της πρόληψης ψευδών ειδήσεων, ακολουθώντας την αδιαμφισβήτητη διευκόλυνση που προσφέρουν τα σύγχρονα νευρωνικά δίκτυα και τα μεγάλα γλωσσικά μοντέλα.

Λέξεις-κλειδιά: Ψευδείς Ειδήσεις, Παραπληροφόρηση, Συστηματική Βιβλιογραφική Ανασκόπηση, Μηχανική Μάθηση, Μεγάλα Γλωσσικά Μοντέλα (LLMs), Κοινωνικά Δίκτυα, Ανίχνευση, Πρόληψη, Σύνολα Δεδομένων.

Abstract

Fake news detection and prevention present significant challenges in today's digital landscape, where misinformation can spread rapidly and widely via social media and online platforms. As such, we acknowledge the persistent need of studying the various fake news challenges in conjunction with successful techniques to combat them, both from the technological as well as the social perspective. The current thesis aims to provide a comprehensive analysis of the current state of fake news, thoroughly describing the background, the types and characteristics of fake news, the dissemination strategies and their effects. Further, it contributes a literature review focusing on the categorization analysis and comparison of detection and prevention strategies related to fake news, determining major developments mostly associated with the rapid advancement of artificial intelligence and machine learning algorithms. Throughout our study, we have recognized some research gaps, which we analyse as well as correlate with research directions addressing them. Overall, we aspire that the current thesis will serve as a point of reference for future endeavours towards advancing the state of fake news detection and prevention, following the indisputable facilitation offered by contemporary neural networks and large language models.

Keywords: Fake News, Disinformation, Systematic Literature Review, Machine Learning, Large Language Models (LLMs), Social Media, Detection, Prevention, Datasets.

1. Εισαγωγή

1.1. Ψευδείς ειδήσεις και τεχνολογικές εξελίξεις

Η ταχεία εξέλιξη της τεχνολογίας τα τελευταία χρόνια έχει συμβάλει σημαντικά στη διάδοση ψευδών ειδήσεων (fake news), αναδιαμορφώνοντας το τοπίο της διάδοσης των πληροφοριών και θέτοντας ουσιαστικές προκλήσεις για την κοινωνία, τη δημοκρατία και την ατομική αυτονομία. Η παρούσα εργασία διερευνά πώς οι τεχνολογικές εξελίξεις έχουν διευκολύνει τη διάδοση της παραπληροφόρησης, εξετάζοντας βασικούς παράγοντες, όπως ο εκδημοκρατισμός της δημιουργίας περιεχομένου, η άνοδος των πλατφορμών των μέσων κοινωνικής δικτύωσης (social media), η χειραγώγηση των αλγορίθμων με σκοπό το κέρδος και την επιρροή, αλλά και η ανωνυμία που μπορεί να προσφέρει το διαδίκτυο με τη δημιουργία μη διαπιστευμένων προφίλ χρηστών. Ορισμένα παραδείγματα αλγορίθμων που η χειραγώγησή τους γίνεται με σκοπό το κέρδος και την επιρροή είναι οι αλγόριθμοι συστάσεων (Recommendation Algorithms) και οι αλγόριθμοι κατάταξης αναζητήσεων (Search Ranking Algorithms) [12, 111, 112, 113, 115] κ.α.

Ένας από τους κύριους τρόπους με τους οποίους η τεχνολογία έχει τροφοδοτήσει την εξάπλωση των ψευδών ειδήσεων είναι ο εκδημοκρατισμός της δημιουργίας περιεχομένου. Με την ευρεία διαθεσιμότητα της πρόσβασης στο διαδίκτυο και τα φιλικά προς τον χρήστη εργαλεία για την παραγωγή περιεχομένου, σχεδόν ο καθένας μπορεί να δημιουργήσει και να μοιραστεί πληροφορίες στο διαδίκτυο. Ενώ αυτός ο εκδημοκρατισμός έχει δώσει τη δυνατότητα στα άτομα να συμμετέχουν στο δημόσιο διάλογο, έχει επίσης μειώσει τα εμπόδια εισόδου για όσους επιδιώκουν να διαδώσουν ψευδείς ή παραπλανητικές πληροφορίες. Η ευκολία με την οποία μπορούν να δημιουργηθούν και να διαμοιραστούν κατασκευασμένες ιστορίες, εικόνες που έχουν υποστεί φωτομοντάζ και πειραγμένα βίντεο ενισχύει τη δυνατότητα της παραπληροφόρησης να φτάσει γρήγορα και ανεμπόδιστα σε μαζικά ακροατήρια.

Οι πλατφόρμες κοινωνικής δικτύωσης έχουν διαδραματίσει κεντρικό ρόλο στην ενίσχυση της διάδοσης των ψευδών ειδήσεων. Αυτές οι πλατφόρμες, οι οποίες έχουν σχεδιαστεί για να συνδέουν τους χρήστες και να διευκολύνουν την ανταλλαγή περιεχομένου, έχουν γίνει πρόσφορο έδαφος για την ταχεία διάδοση της παραπληροφόρησης. Οι αλγόριθμοι που δίνουν προτεραιότητα στην εμπλοκή (engagement) και την εξάπλωση τάσεων με ταχείς ρυθμούς (δηλ. υιογενούς περιεχομένου (viral content) στην καθομιλουμένη) συχνά δίνουν προτεραιότητα στο εντυπωσιακό ή διχαστικό περιεχόμενο, ανεξάρτητα από την ακρίβειά του. Επιπλέον, το φαινόμενο του θαλάμου απήχησης (echo chamber), όπου οι χρήστες εκτίθενται σε περιεχόμενο που ευθυγραμμίζεται με τις υπάρχουσες πεποιθήσεις και προτιμήσεις τους, μπορεί να εδραιώσει περαιτέρω ψευδείς αφηγήσεις και να απομονώσει τα άτομα από αντιφατικές πληροφορίες (οι οποίες συχνά ενδέχεται να περιέχουν μεγαλύτερες δόσεις αλήθειας). Ο μαζικός και ταχύρρυθμος χαρακτήρας των μέσων κοινωνικής δικτύωσης επιτρέπει στις ψευδείς ειδήσεις να εξαπλώνονται γρήγορα στα δίκτυα, φτάνοντας σε εκατομμύρια χρήστες μέσα σε λίγα λεπτά και συχνά ξεπερνώντας τις προσπάθειες για διάψευση ή έλεγχο των πληροφοριών.

1.2. Συνεισφορά της εργασίας

Παρατηρήθηκε ότι παλαιότερες επισκοπήσεις δίνουν έμφαση είτε σε τεχνικές ανίχνευσης είτε στην ανάδειξη διαφόρων πτυχών του φαινομένου. Λίγες από αυτές παρείχαν πλήρη κάλυψη του φαινομένου [5, 11, 30], καθώς και τον συνδυασμό τεχνολογικών με κοινωνικές και επικοινωνιακές προσεγγίσεις, ενσωματώνοντας παράλληλα τις τεχνολογικές εξελίξεις που επηρεάζουν το φαινόμενο αυτό [78]. Τέλος, δεν περιέχονται σύνολα δεδομένων ούτε υποστηρίζεται η αξιολόγηση των προσεγγίσεων με συγκεκριμένα κριτήρια.

Με την παρούσα εργασία θέλουμε να συνεισφέρουμε στην κάλυψη των προαναφερθέντων κενών παρέχοντας σωστή και καλά δομημένη βιβλιογραφία που συνδυάζει τεχνολογική, κοινωνική και επικοινωνιακή προσέγγιση, κατηγοριοποίηση μεθόδων ανίχνευσης και πρόληψης, συμπεριλαμβανομένων μεγάλων γλωσσικών μοντέλων, συγκριτική ανάλυσή τους, την ανάδειξη ερευνητικών κενών και τέλος την παρουσίαση συνόλων δεδομένων για την αξιολόγηση και ανάπτυξη μεθόδων ανίχνευσης. Ο τρόπος που προσεγγίζουμε την εργασία θέλουμε να πιστέψουμε ότι προσφέρει ένα σημείο αναφοράς για ερευνητές και επαγγελματίες του χώρου, με σκοπό την ανάπτυξη πιο εξειδικευμένων μεθόδων αντιμετώπισης των ψευδών ειδήσεων.

Με βάση την ευκολία εξάπλωσης ψευδών ειδήσεων λόγω της ευρείας χρήσης του διαδικτύου, των πολλαπλών τρόπων με τους οποίους αυτά μπορούν αβίαστα να φτάσουν στους τελικούς αποδέκτες, της πληθώρας συμφερόντων πίσω από τη διάδοση ψευδών ειδήσεων, όπως επίσης και των πολλαπλών επιδράσεων που οι ψευδείς ειδήσεις έχουν σε κοινωνικό και σε προσωπικό επίπεδο, είναι ιδιαιτέρως χρήσιμη η ανάλυση του εν λόγω πεδίου από ερευνητική σκοπιά. Η έρευνα των ψευδών ειδήσεων συνδυάζει πολλούς επιστημονικούς τομείς συμπεριλαμβανομένων της κοινωνιολογίας, της ψυχολογίας, της επιστήμης μέσων μαζικής ενημέρωσης και διαφήμισης, αλλά και πρόσφατα τομείς που σχετίζονται με την τεχνητή νοημοσύνη, καθώς οι σχετικοί αλγόριθμοι λαμβάνουν ένα κρίσιμο ρόλο στη διάδοση της πληροφορίας (άρα και των ψευδών ειδήσεων), ενώ παράλληλα οι ίδιες τεχνολογίες έχουν την ικανότητα ανίχνευσης και αυτόματου περιορισμού της διάδοσης ψευδούς πληροφορίας.

Λόγω της σημαντικότητας μελέτης των ψευδών ειδήσεων, η συγκεκριμένη εργασία επικεντρώνεται στη βιβλιογραφική ανασκόπηση σημαντικών ερευνητικών άρθρων πάνω στο πεδίο αυτό, με σκοπό να αναδειχθεί η έκταση του προβλήματος, και κυρίως να μελετηθούν οι τρόποι αναγνώρισης/ανίχνευσης, πρόληψης και αντιμετώπισής των ψευδών ειδήσεων. Κατ' αυτόν τον τρόπο, η ανάπτυξη μελλοντικών εφαρμοσμένων ερευνών πάνω στο πεδίο θα μπορεί να βασιστεί σε εγκαθιδρυμένες έρευνες, μέσω των οποίων μπορούμε να αναγνωρίσουμε συγκεκριμένα ερευνητικά κενά προς πλήρωση.

1.3 Δομή αναφοράς

Το υπόλοιπο της παρούσας αναφοράς χωρίζεται στα ακόλουθα κεφάλαια:

1. **Κεφάλαιο 1. Εισαγωγή**
2. **Κεφάλαιο 2. Θεωρητικό Υπόβαθρο:** αναλύεται το θεωρητικό υπόβαθρο, οι τύποι των ψευδών ειδήσεων, τα κύρια χαρακτηριστικά τους, το γιατί είναι πιο επίκαιρες από ποτέ, οι τρόποι και τα μέσα διάδοσής τους, οι υπαίτιοι, τα κίνητρά τους, οι τρόποι περιορισμού τους, το αντίκτυπο που έχουν καθώς και ποιες είναι οι συνέπειές τους.
3. **Κεφάλαιο 3. Μεθοδολογία**: περιγραφή της στρατηγικής αναζήτησης, φιλτραρίσματος και οργάνωσης των επιστημονικών άρθρων άρθρων που εφαρμόστηκε στο πλαίσιο της μεθοδολογίας συστηματικής ανασκόπησης που ακολουθήθηκε. Τα ερευνητικά ερωτήματα οι χρησιμοποιημένες, βλιογραφικές βάσεις, τα μοτίβα αναζήτησης που εφαρμόστηκαν σε αυτές, τα κριτήρια φιλτραρίσματος των εντοπισμένων άρθρων καθώς και τα στατιστικά και η ποσοτική ανάλυση που αφορά τα εν τέλει επιλεγμένα άρθρα και συνοδεύει το περιεχόμενο της επισκόπησης διασφαλίζουν την ποιότητα και την εγκυρότητα της παρούσας εργασίας, επιτρέποντας σε μελλοντικούς αναγνώστες τη χρήση της ως σημείο αναφοράς.
4. **Κεφάλαιο 4. Ανάλυση**: τα συλλεχθέντα επιστημονικά άρθρα ταξινομούνται γύρω από τους δύο κύριους άξονες της ανίχνευσης και της πρόληψης, και στη συνέχεια αναλύονται με χρονολογική σειρά ανά θεματική (κατηγορία). Επιπλέον, παρέχεται συγκριτική αξιολόγηση των άρθρων με βάση την απόδοσή τους σε συγκεκριμένα κριτήρια που έχουμε επινοήσει ή αντλήσει από την βιβλιογραφία με στόχο την ευκολότερη αναφορά και σύγκρισή τους καθώς και την εξαγωγή χρήσιμων και ενδιαφέροντων συμπερασμάτων.
5. **Κεφάλαιο 5. Εύρεση Ερευνητικών Κενών:** σε συνέχεια της προηγούμενης ανάλυσης, αναγνωρίζουμε τα σχετικά ερευνητικά κενά με σκοπό να προτείνουμε κατευθύνσεις και ιδέες για μελλοντική έρευνα γύρω από την ανίχνευση, πρόληψη και αντιμετώπιση ψευδών ειδήσεων.
6. **Κεφάλαιο 6. Σύνολα Δεδομένων:** συλλέγουμε σύνολα δεδομένων (datasets) γύρω από το θέμα των ψευδών ειδήσεων, τα συγκρίνουμε και τα αναλύουμε. Τα σύνολα δεδομένων τοποθετούνται σε πίνακα με σκοπό την κατηγοριοποίησή τους με βάση ορισμένα κριτήρια, διευκολύνοντας την επιλογή τους από ερευνητές με βάση τις απαιτήσεις που αυτοί έχουν αλλά και την προώθηση της αντικειμενικής επιστημονικής επαλήθευσης που στοχεύει στην καθολική και αντικειμενική (συγκριτική) αξιολόγηση των διαφόρων επιστημονικών προσεγγίσεων με βάση τη χρήση κοινών συνόλων δεδομένων και δυνητικά αλγορίθμων ή λογισμικών (υλοποίησης) ανοικτών ως προς τον κώδικά τους.
7. **Κεφάλαιο 7. Συμπεράσματα και Επίλογος:** εκτός από την περιληπτική ανάλυση της βασικής μας συνεισφοράς, καταληκτικά αναφέρουμε τα συμπεράσματα βάσει της προηγούμενης ανάλυσης και τα ερευνητικά κενά που αναδύονται, με στόχο να κατευθύνουμε την επιστημονική κοινότητα προς μελλοντική έρευνα αναφορικά με τις πιο

επιτακτικές προοπτικές. Επίσης, προσδιορίζουμε την προσωπική εμπειρία που αποκομίσαμε από την εκπόνηση της διπλωματικής εργασίας.

2. Θεωρητικό υπόβαθρο

2.1. Ορισμός ψευδών ειδήσεων

Οι ψευδείς ειδήσεις συνήθως αναφέρονται σε ψευδείς ή παραπλανητικές πληροφορίες που παρουσιάζονται ως ειδήσεις ή πραγματικά γεγονότα, συχνά με σκοπό να εξαπατήσουν ή να χειραγωγήσουν το κοινό. Αυτή η παραπληροφόρηση μπορεί να διαδοθεί μέσω παραδοσιακών μέσων ενημέρωσης, πλατφόρμες κοινωνικής δικτύωσης ή άλλων διαδικτυακών καναλιών. Ο όρος κέρδισε σημαντική προσοχή τα τελευταία χρόνια λόγω του αντίκτυπού του στον δημόσιο διάλογο και τις δημοκρατικές διαδικασίες.

Μια επίσημη πηγή που παρέχει έναν καλό ορισμό των ψευδών ειδήσεων είναι ο ιστότοπος της Ευρωπαϊκής Επιτροπής, ιδίως στο πλαίσιο των προσπαθειών της για την καταπολέμηση της παραπληροφόρησης. Η Ευρωπαϊκή Επιτροπή ορίζει τις ψευδείς ειδήσεις ως εξής:

"Ψευδείς, ανακριβείς ή παραπλανητικές πληροφορίες που έχουν σχεδιαστεί, παρουσιαστεί και προωθηθεί με σκοπό να προκαλέσουν σκόπιμα δημόσια ζημία ή να αποφέρουν κέρδος" [4].

Ο ορισμός αυτός αναδεικνύει τον σκόπιμο χαρακτήρα των ψευδών ειδήσεων και τις πιθανές αρνητικές επιπτώσεις τους στην κοινωνία. Επιπλέον, άλλες αξιόπιστες πηγές, όπως ακαδημαϊκές μελέτες, δημοσιογραφικοί οργανισμοί και κυβερνητικοί φορείς, παρέχουν συχνά παρόμοιους ορισμούς και πληροφορίες σχετικά με την έννοια των ψευδών ειδήσεων [5, 10, 25, 33].

2.2. Τύποι ψευδών ειδήσεων

Προτείνουμε μία ταξινόμηση των ψευδών ειδήσεων που μπορεί να πραγματοποιηθεί με βάση 4 διαστάσεις κατηγοριοποίησης: (α) τον τύπο της πρόθεσης (δηλαδή τον σκοπό), (β) το είδος επεξεργασίας υλικού που απαιτείται, (γ) το είδος περιεχομένου προς χρήση και (δ) τις ενδεχόμενες επιπτώσεις μιας ψευδούς είδησης. Αυτό σημαίνει πως μια ψευδή είδηση μπορεί να εντάσσεται σε πολλαπλές κατηγορίες, μία ανά διαφορετική διάσταση. Παραδείγματος χάριν, μια παραποιημένη εικόνα σάτιρας μπορεί να χρησιμοποιείται για λόγους εκούσιας παραπληροφόρησης και ειδικότερα προπαγάνδας. Οπότε, θα ανήκει στο δεύτερο είδος παραπληροφόρησης, θα αφορά αλλοιωμένο περιεχόμενο, σάτιρα (και όχι παρωδία) και θα χρησιμοποιείται για προπαγανδιστικούς λόγους. Παραθέτουμε παρακάτω ένα ταξινομημένο πίνακα με αυτές τις κατηγορίες ψευδών ειδήσεων.

Πίνακας 1. Κατηγορίες ψευδών ειδήσεων.

Τύπος Πρόθεσης	**Επεξεργασία Υλικού**	**Περιεχόμενο**	**Επιπτώσεις**
Κακόβουλη πληροφόρηση	Πλήρης κατασκευή Αλλοίωση	Προπαγάνδα Ψευδείς ειδήσεις Θεωρίες συνωμοσίας	Πόλωση Ζημία φήμης Μείωση εμπιστοσύνης Πανικός
Εκούσια παραπληροφόρηση	Χρήση αυθεντικού υλικού χωρίς πλαίσιο	Φήμες Ανακρίβειες Μερική παρανόηση	Σύγχυση Παραπληροφόρηση
Ακούσια παραπληροφόρηση	Καμία επεξεργασία (λάθος αναμετάδοση ή ερμηνεία)	Παραπλανητικοί τίτλοι Clickbait	Σύγχυση Μείωση εμπιστοσύνης
Χιούμορ / σάτιρα	Πλήρης κατασκευή (σκοπός διασκέδασης)	Σάτιρα Παρωδία	(Ενίοτε) Παρανόηση Σύγχυση αν λείπει η επισήμανση

2.2.1. Παραπληροφόρηση

Στη βιβλιογραφία συναντώνται τρεις τύποι παραπληροφόρησης, οι οποίοι ονομάζονται misinformation (ακούσια παραπληροφόρηση), disinformation (εκούσια παραπληροφόρηση) και malinformation (κακόβουλη παραπληροφόρηση). Στην ελληνική γλώσσα δεν υπάρχει σαφής διάκριση των μεταφρασμένων όρων, καθώς όλες αποτυπώνονται στην έννοια ‘παραπληροφόρηση’ Ωστόσο, ο σκοπός του κάθε τύπου παραπληροφόρησης μας βοηθά στην ένταξη σχετικών ειδήσεων σε αυτόν. Στη συνέχεια θα αναλύσουμε αυτούς τους 3 τύπους παραπληροφόρησης.

2.2.1.1. Ακούσια παραπληροφόρηση

Η ακούσια παραπληροφόρηση (Misinformation**)** αναφέρεται στη διάδοση ψευδών ή ανακριβών πληροφοριών, συχνά ακούσια, χωρίς πρόθεση εξαπάτησης. Μπορεί να προκύψει από διάφορες πηγές, συμπεριλαμβανομένων γνήσιων λαθών ή παρερμηνείας γεγονότων. Αυτός ο τύπος παραπληροφόρησης μπορεί να διαδοθεί με τη βοήθεια κάθε δυνατού μέσου επικοινωνίας. Ένα κοινό χαρακτηριστικό του είναι ότι συνήθως δεν δημιουργείται με κακόβουλη πρόθεση. Αντίθετα, μπορεί να προέρχεται από έλλειψη κατανόησης, ελλιπείς πληροφορίες ή από την ταχεία διάδοση ανεπιβεβαίωτων ισχυρισμών κατά τη διάρκεια έκτακτων ειδήσεων. Στην ψηφιακή εποχή, αυτός ο τύπος παραπληροφόρησης μπορεί να εξαπλωθεί γρήγορα και να φτάσει σε ευρύ κοινό λόγω της δυκολίας διαμοιρασμού του περιεχομένου στο διαδίκτυο.

Για παράδειγμα, ανακριβείς πληροφορίες σχετικά με την υγεία, τις επιστημονικές ανακαλύψεις ή τα πολιτικά γεγονότα μπορούν να κερδίσουν γρήγορα έδαφος, οδηγώντας σε σύγχυση ή παρανοήσεις στο κοινό [5]. Η αντιμετώπιση της ακούσιας παραπληροφόρησης απαιτεί δεξιότητες κριτικής σκέψης, έλεγχο των γεγονότων και υπεύθυνη ενημέρωση από τα μέσα ενημέρωσης. Είναι σημαντικό για τα άτομα να επαληθεύουν τις πληροφορίες από αξιόπιστες πηγές προτού τις μοιραστούν περαιτέρω για να αποτρέψουν την ακούσια διάδοση ψευδών πληροφοριών. Η έρευνα σχετικά με τον τύπο ακούσια παραπληροφόρηση υπογραμμίζει τη σημασία της διάκρισης μεταξύ των ακούσιων σφαλμάτων και των σκόπιμων προσπαθειών εξαπάτησης. Οι προσπάθειες για την καταπολέμηση της ακούσιας παραπληροφόρησης συχνά περιλαμβάνουν εκπαιδευτικές εκστρατείες, την προώθηση της παιδείας στα μέσα ενημέρωσης και την αξιοποίηση της τεχνολογίας για τον εντοπισμό και την επισήμανση ψευδών πληροφοριών στο διαδίκτυο. [5]

Η κατανόηση της φύσης αυτού του τύπου παραπληροφόρησης είναι ζωτικής σημασίας για την ανάπτυξη στρατηγικών για την προώθηση της ακριβούς πληροφόρησης και τον μετριασμό του αντίκτυπου των ψευδών αφηγήσεων στο δημόσιο διάλογο και τις διαδικασίες λήψης αποφάσεων. Οι επιστημονικές μελέτες και η ακαδημαϊκή βιβλιογραφία σχετικά με τον τύπο αυτό παρέχουν πολύτιμες γνώσεις για τα αίτια, τις συνέπειες και τις πιθανές λύσεις σε αυτόν, συμβάλλοντας στις συνεχιζόμενες προσπάθειες για την αντιμετώπιση των προκλήσεων που αυτός θέτει στον ολοένα και πιο διασυνδεδεμένο κόσμο μας.

2.2.1.2. Εκούσια παραπληροφόρηση

Η εκούσια παραπληροφόρηση (Disinformation**)** αναφέρεται σε σκόπιμα ψευδείς ή παραπλανητικές πληροφορίες που διαδίδονται με σκοπό να εξαπατήσουν, να χειραγωγήσουν ή να επηρεάσουν την κοινή γνώμη. Σε αντίθεση με την ακούσια παραπληροφόρηση που περιγράφηκε παραπάνω, η οποία μπορεί να προκύψει από γνήσια λάθη ή παρερμηνείες, ο τύπος εκούσια παραπληροφόρηση δημιουργείται και διαδίδεται σκόπιμα για την επίτευξη συγκεκριμένων στόχων, όπως η υπονόμευση της εμπιστοσύνης στους θεσμούς, η διαμόρφωση πολιτικών αφηγήσεων ή η προώθηση συγκεκριμένων προγραμμάτων ή ατζεντών [5].

Οι δημιουργοί του τύπου εκούσια παραπληροφόρηση χρησιμοποιούν συχνά διάφορες τακτικές για να κάνουν τις ψευδείς πληροφορίες να φαίνονται αξιόπιστες και πιστευτές από τους αποδέκτες. Αυτό μπορεί να περιλαμβάνει τη χρήση ψεύτικων πηγών, τη χειραγώγηση εικόνων ή βίντεο, την κατασκευή αποσπασμάτων ή την εκμετάλλευση συναισθηματικών ερεθισμάτων για την πρόκληση έντονων αντιδράσεων από το κοινό. Οι εκστρατείες παραπληροφόρησης αυτού του τύπου μπορεί να ενορχηστρώνονται από κρατικούς φορείς, πολιτικές οργανώσεις, ομάδες ειδικών συμφερόντων ή άτομα που επιδιώκουν να προωθήσουν τα συμφέροντά τους ή να σπείρουν τη διχόνοια μέσα στις κοινωνίες. Οι εκστρατείες αυτές μπορεί να στοχεύουν σε συγκεκριμένες δημογραφικές ομάδες, να εκμεταλλεύονται τις υπάρχουσες κοινωνικές διαιρέσεις ή να εκμεταλλεύονται τα τρωτά σημεία των οικοσυστημάτων των μέσων ενημέρωσης για να ενισχύουν ψευδείς αφηγήσεις. Η ψηφιακή εποχή έχει διευκολύνει την ταχεία εξάπλωση της παραπληροφόρησης αυτού του τύπου, καθώς οι διαδικτυακές πλατφόρμες παρέχουν πρόσφορο έδαφος για τη διάδοση ψευδών πληροφοριών σε μεγάλα ακροατήρια με ελάχιστη εποπτεία.

Ειδικότερα, οι αλγόριθμοι των μέσων κοινωνικής δικτύωσης μπορούν ακούσια να ενισχύσουν την παραπληροφόρηση δίνοντας προτεραιότητα σε εντυπωσιακό ή αμφιλεγόμενο περιεχόμενο. [5]

Η καταπολέμηση του τύπου εκούσια παραπληροφόρηση απαιτεί μια πολύπλευρη προσέγγιση που περιλαμβάνει την παιδεία στα μέσα ενημέρωσης, πρωτοβουλίες ελέγχου των γεγονότων, τεχνολογικές παρεμβάσεις και ρυθμιστικά μέτρα. Η εκπαίδευση του κοινού σχετικά με τις τακτικές που χρησιμοποιούνται στις εκστρατείες αυτού του τύπου παραπληροφόρησης και η προώθηση των δεξιοτήτων κριτικής σκέψης αποτελούν βασικά βήματα για την ενδυνάμωση των ατόμων ώστε να αναγνωρίζουν και να απορρίπτουν τις ψευδείς πληροφορίες. Η έρευνα για την εκούσια παραπληροφόρηση διερευνά την προέλευση, τις επιπτώσεις και τις στρατηγικές ανίχνευσης και μετριασμού της. Αυτό περιλαμβάνει τη μελέτη της ψυχολογίας της παραπληροφόρησης αυτού του τύπου, την ανάλυση της αποτελεσματικότητας των προσπαθειών αντί-παραπληροφόρησης και την ανάπτυξη εργαλείων για τον εντοπισμό και την αντιμετώπιση ψευδών αφηγήσεων σε πραγματικό χρόνο. Η αντιμετώπιση της απειλής που συνιστά αυτός ο τύπος απαιτεί τη συνεργασία μεταξύ κυβερνήσεων, εταιρειών τεχνολογίας, οργανώσεων της κοινωνίας των πολιτών και ερευνητών για την ανάπτυξη ολοκληρωμένων στρατηγικών που προστατεύουν την ακεραιότητα του δημόσιου λόγου και των δημοκρατικών διαδικασιών [5].

2.2.1.3. Κακόβουλη παραπληροφόρηση

Η κακόβουλη παραπληροφόρηση (Malinformation) αναφέρεται στη διάδοση πληροφοριών με σκοπό να βλάψει, να φέρει σε δύσκολη θέση ή να δυσφημίσει άτομα, οργανισμούς ή οντότητες. Σε αντίθεση με την ακούσια παραπληροφόρηση και την εκούσια παραπληροφόρηση, που περιλαμβάνουν ψευδές ή παραπλανητικό περιεχόμενο, η κακόβουλη παραπληροφόρηση περιλαμβάνει την εσκεμμένη έκθεση αληθινών πληροφοριών που μπορεί να έχουν αποκτηθεί μέσω επιθέσεων κοινωνικής μηχανικής, πειρατείας, παραβίασης δεδομένων, διαρροών ή άλλων παράνομων μέσων [5].

Ο πρωταρχικός στόχος της κακόβουλης παραπληροφόρησης δεν είναι απαραίτητα η παραπλάνηση ή η χειραγώγηση της κοινής γνώμης, αλλά μάλλον η πρόκληση ζημίας στη φήμη ή η υποκίνηση διαμάχης. Αυτός ο τύπος παραπληροφόρησης μπορεί να περιλαμβάνει ιδιωτική αλληλογραφία, προσωπικές φωτογραφίες ή βίντεο, εμπιστευτικά έγγραφα ή άλλες ευαίσθητες πληροφορίες που, όταν εκτεθούν, μπορούν να έχουν επιζήμιες συνέπειες για τα εμπλεκόμενα άτομα ή οντότητες. Ο τύπος αυτός εγείρει σημαντικές ανησυχίες δεοντολογίας και προστασίας της ιδιωτικής ζωής, καθώς μπορεί να παραβιάσει το δικαίωμα των ατόμων στην ιδιωτική ζωή και να τα εκθέσει σε δημόσιο έλεγχο, παρενόχληση ή ακόμη και σωματική βλάβη. Μπορεί επίσης να υπονομεύσει την εμπιστοσύνη στους θεσμούς, να διαβρώσει την εμπιστοσύνη στα ψηφιακά κανάλια επικοινωνίας και να έχει ευρύτερες κοινωνικές επιπτώσεις στην ελευθερία της έκφρασης και της πληροφόρησης [5].

Η διάδοση της κακόβουλης παραπληροφόρησης διευκολύνεται από την ψηφιακή εποχή και την ευκολία με την οποία μπορούν να διαδοθούν οι πληροφορίες στο διαδίκτυο. Οι πλατφόρμες των μέσων κοινωνικής δικτύωσης, ειδικότερα, παρέχουν ένα ισχυρό μέσο στους κακόβουλους φορείς να ενισχύουν και να οπλοποιούν ιδιωτικές πληροφορίες για κακόβουλους σκοπούς. Από την άλλη,

η καταπολέμηση του τύπου αυτού παραπληροφόρησης απαιτεί μια πολύπλευρη προσέγγιση που εξισορροπεί την ανάγκη προστασίας των δικαιωμάτων της ιδιωτικής ζωής με την επιτακτική ανάγκη αποτροπής της κακόβουλης χρήσης ευαίσθητων πληροφοριών.

Αυτό περιλαμβάνει την εφαρμογή ισχυρών μέτρων κυβερνοασφάλειας για την πρόληψη παραβιάσεων δεδομένων και μη εξουσιοδοτημένης πρόσβασης, την επιβολή νόμων και κανονισμών που απαγορεύουν τη διάδοση ιδιωτικών πληροφοριών χωρίς συγκατάθεση και την προώθηση του ψηφιακού αλφαβητισμού και της υπεύθυνης διαδικτυακής συμπεριφοράς των χρηστών [5].

2.2.1.4. Σάτιρα και παρωδία

Η σάτιρα και η παρωδία είναι μορφές δημιουργικής έκφρασης που χρησιμοποιούν το χιούμορ, την υπερβολή ή τη μίμηση για να σχολιάσουν και να επικρίνουν γεγονότα της πραγματικής ζωής, άτομα ή κοινωνικά ζητήματα. Ενώ η σάτιρα έχει ως στόχο να προκαλέσει σκέψη και να εκθέσει την ανοησία ή τη βλακεία μέσω του χιούμορ και της ειρωνείας, η παρωδία περιλαμβάνει την μίμηση ενός συγκεκριμένου στυλ, έργου ή ατόμου για κωμικό αποτέλεσμα.

Η σάτιρα χρησιμοποιεί συχνά σαρκασμό, ευφυΐα και υπερβολή για να αναδείξει παραλογισμούς, αντιφάσεις ή αδικίες στην κοινωνία. Μπορεί να λάβει διάφορες μορφές, όπως γραπτά έργα, γελοιογραφίες, παραστάσεις ή ψηφιακά μέσα, και συνήθως χαρακτηρίζεται από την πρόθεσή της να προκαλέσει κριτικό προβληματισμό ή να αμφισβητήσει τους επικρατούντες κανόνες και συμβάσεις. Η παρωδία, από την άλλη πλευρά, περιλαμβάνει τη μίμηση ή την απομίμηση υφιστάμενων έργων, ειδών ή προσωπικοτήτων με χιουμοριστικό ή υπερβολικό τρόπο. Συχνά περιλαμβάνει παιγνιώδη μίμηση και μπορεί να χρησιμεύσει ως μια μορφή φόρου τιμής ή σχολιασμού της λαϊκής κουλτούρας, της λογοτεχνίας ή της ψυχαγωγίας. [6]

Είναι σημαντικό να σημειωθεί ότι η σάτιρα και η παρωδία διαφέρουν από τις ψευδείς ειδήσεις ή την παραπληροφόρηση. Παρόλο που μπορεί να περιλαμβάνουν φανταστικά στοιχεία ή υπερβολικά σενάρια, ο σκοπός τους δεν είναι να εξαπατήσουν, αλλά μάλλον να διασκεδάσουν, να προκαλέσουν σκέψη ή να προκαλέσουν προβληματισμό για σύγχρονα ζητήματα. Παρόλα αυτά, η σάτιρα και η παρωδία μπορεί μερικές φορές να εμπλακούν με παραπλανητικό περιεχόμενο ή ψευδείς ειδήσεις, έστω και ακούσια. Ένας τρόπος με τον οποίο συμβαίνει αυτή η σύνδεση είναι μέσω της παρερμηνείας ή της παραποίησής τους. Το σατιρικό ή παρωδιακό περιεχόμενο, εάν δεν επισημαίνεται ή δεν αναγνωρίζεται σαφώς ως τέτοιο, μπορεί να εκληφθεί από το κοινό λανθασμένα ως γνήσια είδηση. Άτομα που δεν είναι εξοικειωμένα με το σκοπό ή το πλαίσιο της σάτιρας μπορεί να πάρουν το υπερβολικό ή χιουμοριστικό περιεχόμενο τοις μετρητοίς, πιστεύοντας ότι πρόκειται για πραγματική είδηση. [6, 7]

Επιπλέον, κακόβουλοι φορείς ενδέχεται να χειραγωγήσουν σατιρικό ή παρωδιακό περιεχόμενο για να το κάνουν να φαίνεται πιο πειστικό ή παραπλανητικό. Αλλάζοντας το πλαίσιο, αφαιρώντας ενδείξεις σάτιρας ή επεξεργάζοντας επιλεκτικά το περιεχόμενο, στοχεύουν να εξαπατήσουν το κοινό ώστε να πιστέψει ότι πρόκειται για νόμιμες ειδήσεις. Σε ορισμένες περιπτώσεις, οι δημιουργοί ψευδών ειδήσεων χρησιμοποιούν τη σάτιρα ή την παρωδία ως πρόσχημα για τη

διάδοση παραπληροφόρησης (εκούσια παραπληροφόρηση). Παρουσιάζοντας ψευδείς ή παραπλανητικές πληροφορίες σε σατιρική μορφή, εκμεταλλεύονται τις θολές γραμμές μεταξύ χιούμορ και πραγματικότητας για να αποφύγουν τον εντοπισμό ή τον έλεγχο.

Επιπλέον, η ακούσια ενίσχυση (άρα παραπληροφόρηση τύπου misinformation) μπορεί να συμβεί όταν το σατιρικό ή παρωδιακό περιεχόμενο βρίσκει απήχηση σε ορισμένα κοινά και μοιράζεται ακούσια ως γνήσια είδηση.

Τα άτομα μπορεί να μοιράζονται περιεχόμενο χωρίς να επαληθεύουν τη γνησιότητά του και οι αλγόριθμοι των μέσων κοινωνικής δικτύωσης μπορεί να προωθούν εντυπωσιακό ή προκλητικό περιεχόμενο, ανεξάρτητα από τη σατιρική του φύση. Ορισμένοι δημιουργοί ψευδών ειδήσεων χρησιμοποιούν ακόμη και αξιόπιστες σατιρικές πλατφόρμες για τη διάδοση παραπλανητικού περιεχομένου. Διεισδύοντας σε γνωστά σατιρικά μέσα ή λογαριασμούς παρωδίας, προσεγγίζουν ευρύτερο κοινό και ενδεχομένως εξαπατούν άτομα που εμπιστεύονται αυτές τις πηγές για χιουμοριστικά ή ειρωνικά σχόλια.

Παρόλο που η ίδια η σάτιρα και η παρωδία δεν αποσκοπούν στην παραπλάνηση, η σύνδεσή τους με το παραπλανητικό περιεχόμενο υπογραμμίζει τη σημασία της παιδείας στα μέσα ενημέρωσης και της κριτικής σκέψης. Το κοινό πρέπει να είναι κριτικό και ικανό να διακρίνει μεταξύ των γνήσιων ειδήσεων, της σάτιρας και του απόλυτου ψεύδους. Οι δημιουργοί περιεχομένου και οι πλατφόρμες φέρουν επίσης ευθύνη για τη σαφή επισήμανση της σάτιρας και της παρωδίας, ώστε να ελαχιστοποιείται ο κίνδυνος ακούσιας παραπληροφόρησης (misinformation). Η κατανόηση της διαφοροποιημένης σχέσης μεταξύ της σάτιρας, της παρωδίας και του παραπλανητικού περιεχομένου είναι ζωτικής σημασίας για την πλοήγηση στο σημερινό πολύπλοκο τοπίο των μέσων ενημέρωσης και την καταπολέμηση της εξάπλωσης των ψευδών ειδήσεων.

2.2.2. Παραποιημένο περιεχόμενο

2.2.2.1. Εξ ολοκλήρου κατασκευασμένο περιεχόμενο

Το εξ ολοκλήρου κατασκευασμένο περιεχόμενο (fabricated content) αναφέρεται σε εντελώς ψευδείς ιστορίες, γεγονότα ή πληροφορίες που δημιουργούνται από το μηδέν και παρουσιάζονται ως γνήσιες ειδήσεις. Αυτός ο τύπος ψευδών ειδήσεων περιλαμβάνει την επινόηση αφηγήσεων ή γεγονότων που δεν συνέβησαν ποτέ, συχνά με σκοπό να εξαπατήσουν ή να παραπλανήσουν το κοινό. [8]

Μια συνηθισμένη μορφή κατασκευασμένου περιεχομένου είναι η δημιουργία πλαστών εγγράφων. Αυτό περιλαμβάνει την παραγωγή παραποιημένων εγγράφων, όπως επίσημες εκθέσεις, επιστολές ή δηλώσεις για την υποστήριξη πλασματικών ισχυρισμών ή αφηγήσεων. Μια άλλη συνηθισμένη μέθοδος είναι η δημιουργία ψευδών ειδησεογραφικών ιστοσελίδων ή άρθρων, όπου χρησιμοποιούνται εξολοκλήρου επινοημένες πληροφορίες, τίτλοι, πρόσωπα ή οργανισμοί για να κατασκευαστεί μια αφήγηση που φαίνεται αληθινή. Συχνά, αυτά τα άρθρα μιμούνται τη μορφή και το ύφος αξιόπιστων ειδησεογραφικών πηγών, με στόχο την εξαπάτηση του κοινού. Επιπλέον, το κατασκευασμένο περιεχόμενο μπορεί να περιλαμβάνει την επινόηση εντελώς φανταστικών

γεγονότων ή περιστατικών, που συχνά συνοδεύονται από κατασκευασμένες λεπτομέρειες ή μαρτυρίες αυτοπτών μαρτύρων για να προσδώσουν αξιοπιστία στην ψευδή αφήγηση. [8]

Το παραποιημένο περιεχόμενο μπορεί να διαδίδεται μέσω ιστοτόπων, μέσων κοινωνικής δικτύωσης ή άλλων ψηφιακών πλατφορμών, μερικές φορές μιμούμενο τη μορφή και το ύφος νόμιμων ειδησεογραφικών πηγών για να φαίνεται αξιόπιστο. Ο στόχος της δημιουργίας κατασκευασμένου περιεχομένου είναι η δημιουργία προσοχής, η διάδοση παραπληροφόρησης ή η προώθηση συγκεκριμένων σκοπών εκμεταλλευόμενοι την εμπιστοσύνη του κοινού στα ειδησεογραφικά μέσα. [9, 10]

Η καταπολέμηση του πλήρως κατασκευασμένου περιεχομένου απαιτεί μια πολύπλευρη προσέγγιση που περιλαμβάνει τεχνολογικές, εκπαιδευτικές και ρυθμιστικές στρατηγικές. Μπορούν να αναπτυχθούν προηγμένοι αλγόριθμοι τεχνητής νοημοσύνης και μηχανικής μάθησης για τον εντοπισμό και την επισήμανση κατασκευασμένου περιεχομένου μέσω της ανάλυσης μοτίβων, τόσο αναφορικά με την εικόνα, όσο και με κείμενα. Στην περίπτωση των κειμένων, έχουν αναπτυχθεί μοντέλα για την εύρεση ασυνεπειών των πηγών, η χρήση των οποίων είναι δυνατό να αναδείξει περιπτώσεις πλήρως κατασκευασμένου περιεχομένου.

Συνεπώς, οι πλατφόρμες μέσων κοινωνικής δικτύωσης και οι μηχανές αναζήτησης θα πρέπει να εφαρμόσουν ισχυρούς μηχανισμούς ελέγχου των γεγονότων και να προωθήσουν τη διαφάνεια στις διαδικασίες επιμέλειας του περιεχομένου τους με ευρεία χρήση των προηγμένων αλγορίθμων. Από την άλλη, οι εκστρατείες ευαισθητοποίησης του κοινού και τα εκπαιδευτικά προγράμματα είναι απαραίτητα για τη βελτίωση της παιδείας στα μέσα ενημέρωσης, επιτρέποντας στα άτομα να αξιολογούν κριτικά την αξιοπιστία των πληροφοριών που συναντούν. Επιπλέον, τα ρυθμιστικά μέτρα μπορούν να καθιερώσουν τη λογοδοσία για τη διάδοση ψευδών ειδήσεων, καθιστώντας τους δημιουργούς και τους διανομείς υπεύθυνους. [5]

2.2.2.2. Αλλοιωμένο περιεχόμενο

Το αλλοιωμένο περιεχόμενο (manipulated content) περιλαμβάνει τη διαστρέβλωση ή την αλλοίωση πραγματικών πληροφοριών, γεγονότων ή μέσων ενημέρωσης με σκοπό την παραπλάνηση του κοινού ή την αλλαγή του πλαισίου ή του νοήματος. Σε αντίθεση με το κατασκευασμένο περιεχόμενο, το χειραγωγημένο περιεχόμενο ξεκινά με γνήσιο υλικό που στη συνέχεια τροποποιείται ή παραποιείται.

Μια συνήθης μέθοδος διαστρέβλωσης περιεχομένου είναι η επιλεκτική επεξεργασία βίντεο. Αυτό περιλαμβάνει την επεξεργασία υλικού βίντεο για την αφαίρεση ή την τροποποίηση ορισμένων τμημάτων, του πλαισίου ή του ήχου, ώστε να δημιουργηθεί μια ψευδής εντύπωση ή αφήγηση που διαφέρει από το αρχικό περιεχόμενο. Μια άλλη μορφή χειραγώγησης είναι η τροποποίηση υπαρχόντων φωτογραφιών, όπου οι εικόνες τροποποιούνται μέσω περικοπής, ρετουσαρίσματος ή προσθήκης στοιχείων μέσω λογισμικού τύπου photoshop για να παραποιηθεί η πραγματική σκηνή ή το θέμα που απεικονίζεται στη φωτογραφία. Επιπλέον, το περιεχόμενο μπορεί να πειραχτεί μέσω της διαστρέβλωσης των συμφραζομένων, όπου πραγματικές πληροφορίες ή αποσπάσματα

παρουσιάζονται εκτός συμφραζομένων για να παραπλανήσουν το κοινό σχετικά με το πραγματικό νόημα ή τις συνέπειες του αρχικού περιεχομένου. [11]

Το παραποιημένο περιεχόμενο χρησιμοποιείται συχνά για να υποστηρίξει ψευδείς ισχυρισμούς, να ενισχύσει προκαταλήψεις ή να προκαλέσει διαμάχες διαστρεβλώνοντας την πραγματικότητα. Στην ψηφιακή εποχή, τα εξελιγμένα εργαλεία επεξεργασίας (όπως αυτά που εκμεταλλεύονται την τεχνολογία της τεχνητής νοημοσύνης) και η ευκολία διαμοιρασμού των μέσων στο διαδίκτυο έχουν καταστήσει ευκολότερη τη χειραγώγηση του περιεχομένου και τη διάδοση παραπλανητικών πληροφοριών. [9, 10, 11]

Η αναγνώριση των τακτικών που χρησιμοποιούνται για τη δημιουργία και τη διάδοση κατασκευασμένου και παραποιημένου περιεχομένου είναι απαραίτητη για τα άτομα και τους οργανισμούς ώστε να εντοπίζουν και να καταπολεμούν τις παραπλανητικές αφηγήσεις στο ψηφιακό τοπίο της πληροφόρησης, προωθώντας έτσι τον γραμματισμό στα μέσα ενημέρωσης και την κριτική σκέψη για την πλοήγηση στην πολυπλοκότητα της διαδικτυακής πληροφόρησης [11]. Τακτικές που χρησιμοποιούνται και στην περίπτωση του πλήρως κατασκευασμένου περιεχομένου μπορούν να αξιοποιηθούν, αν και ο ρόλος τους αντιμετωπίζει αυξημένες δυσκολίες, λόγω της ομοιότητας με το πραγματικό περιεχόμενο.

Για παράδειγμα, ένας αλγόριθμος τεχνητής νοημοσύνης ενδέχεται να μπορέσει με μεγαλύτερη ευκολία να αναγνωρίσει τις ασυνέπειες και την έλλειψη ρεαλισμού σε μια πλήρως τεχνητή εικόνα. Ωστόσο, όταν έχουν παραλλαχθεί ελάχιστα στοιχεία, ο εντοπισμός της παραποίησης γίνεται από δύσκολος έως και αδύνατος (όπως στην περίπτωση περικοπής τμήματος μιας εικόνας). Κατά τον ίδιο τρόπο, όσον αφορά τη δημιουργία κειμένου, ένα εργαλείο ανίχνευσης κειμένου παραγμένου από κάποιο γλωσσικό μοντέλο, όπως για παράδειγμα το GPTZero [32], είναι ικανό να εντοπίσει μοτίβα τα οποία συχνά εμφανίζουν τα μεγάλα γλωσσικά μοντέλα (Large Language Models - LLMs) κατά την παραγωγή κειμένου. Αν όμως η παραποίηση συνίσταται από την αλλαγή μερικών λέξεων, διατηρώντας τη δομή, το ύφος και το ρεαλισμό του κειμένου, ακόμα κι ένας άνθρωπος δε θα είναι σε θέση να αντιληφθεί τις αλλαγές.

2.2.2.3. Clickbait

Το δόλωμα κλικ (clickbait) αναφέρεται σε διαδικτυακό περιεχόμενο που έχει σχεδιαστεί για να προσελκύσει την προσοχή και να ενθαρρύνει τους χρήστες να κάνουν κλικ σε έναν σύνδεσμο για να δουν περισσότερο περιεχόμενο. Συχνά χρησιμοποιεί εντυπωσιακή ή υπερβολική γλώσσα, ενδιαφέρουσες εικόνες ή παραπλανητικούς ισχυρισμούς για να κεντρίσει την περιέργεια και να δημιουργήσει κλικ. Ο πρωταρχικός στόχος του clickbait είναι η αύξηση της επισκεψιμότητας σε έναν ιστότοπο ή μια ιστοσελίδα με το να δελεάσει τους χρήστες να κάνουν κλικ σε έναν συγκεκριμένο σύνδεσμο. [12]

Οι τίτλοι clickbait είναι συνήθως εντυπωσιοθηρικοί και χρησιμοποιούν υπερβολική γλώσσα για να προκαλέσουν έντονες συναισθηματικές αντιδράσεις στους αναγνώστες. Μπορεί να υπόσχονται συγκλονιστικές αποκαλύψεις, απίστευτες ιστορίες ή δραματικά γεγονότα για να τραβήξουν την προσοχή. Το Clickbait αξιοποιεί επίσης το "κενό περιέργειας" (curiosity gap), παρουσιάζοντας

ελλιπείς ή ενδιαφέρουσες πληροφορίες για να αφήσει τους αναγνώστες να θέλουν να μάθουν περισσότερα, ενθαρρύνοντάς τους έτσι να κάνουν κλικ στο σύνδεσμο για να ικανοποιήσουν την περιέργειά τους. Εκτός από τους προκλητικούς τίτλους, το clickbait χρησιμοποιεί συχνά ελκυστικές εικόνες ή μικρογραφίες που μπορεί να μην σχετίζονται με το πραγματικό περιεχόμενο ή να παραποιούνται για να τραβήξουν την προσοχή των χρηστών. Αυτές οι εικόνες έχουν ως στόχο να δελεάσουν τους χρήστες να κάνουν κλικ στο σύνδεσμο. Το clickbait μπορεί επίσης να χρησιμοποιεί ασαφείς ή διφορούμενους ισχυρισμούς που κεντρίζουν το ενδιαφέρον χωρίς να παρέχουν σαφείς λεπτομέρειες, προτρέποντας τους χρήστες να κάνουν κλικ για να αποκαλύψουν την πλήρη ιστορία ή τις πληροφορίες. [12, 13]

Παρόλο που το ίδιο το clickbait μπορεί να μην είναι πάντα συνώνυμο των ψευδών ειδήσεων, μπορεί να συμβάλει στη διάδοση της παραπληροφόρησης παραπλανώντας τους χρήστες και διαστρεβλώνοντας την παρουσίαση των πληροφοριών. Το clickbait έχει επικριθεί για την υποβάθμιση της ποιότητας του διαδικτυακού περιεχομένου και την καλλιέργεια μιας κουλτούρας εντυπωσιασμού, η οποία μπορεί να διαβρώσει την εμπιστοσύνη στις πηγές των μέσων ενημέρωσης, όταν οι χρήστες αισθάνονται ότι εξαπατώνται ή παραπλανώνται από υπερβολικούς ή παραπλανητικούς τίτλους. [13]

Για την αντιμετώπιση του clickbait, ορισμένες πλατφόρμες και εκδότες έχουν εφαρμόσει κατευθυντήριες γραμμές ή αλγόριθμους για να δίνουν προτεραιότητα στο αυθεντικό και ενημερωτικό περιεχόμενο έναντι του εντυπωσιοθηρικού ή παραπλανητικού clickbait. Επιπλέον, οι προσπάθειες για την παιδεία στα μέσα ενημέρωσης αποσκοπούν στην εκπαίδευση των χρηστών σχετικά με τον εντοπισμό και την κριτική αξιολόγηση των τίτλων clickbait, ώστε να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με την ενασχόληση με το διαδικτυακό περιεχόμενο.

Η κατανόηση των τακτικών που χρησιμοποιούνται στο clickbait μπορεί να δώσει στα άτομα τη δυνατότητα να περιηγηθούν στο ψηφιακό τοπίο της πληροφόρησης με μεγαλύτερη διάκριση και υπευθυνότητα.

2.2.2.4. Προπαγανδιστικές τεχνικές

Η προπαγάνδα αναφέρεται στη σκόπιμη διάδοση πληροφοριών, ιδεών ή φημών με στόχο τον επηρεασμό της κοινής γνώμης, τη διαμόρφωση αντιλήψεων ή την προώθηση μιας συγκεκριμένης ατζέντας. Σε αντίθεση με τη νόμιμη ειδησεογραφία, η οποία ιδανικά επιδιώκει την αντικειμενικότητα και την ακρίβεια, η προπαγάνδα είναι εγγενώς προκατειλημμένη και συχνά χειραγωγική στη φύση της. Ένα βασικό χαρακτηριστικό της προπαγάνδας είναι η στήριξή της στη συναισθηματική έκκληση και όχι στον ορθολογισμό. Οι προπαγανδιστές χρησιμοποιούν συναισθήματα όπως ο φόβος, ο θυμός, ο πατριωτισμός ή η συμπάθεια για να προκαλέσουν έντονες αντιδράσεις και να επηρεάσουν το κοινό αίσθημα. Έχουν επίσης την τάση να υπεραπλουστεύουν πολύπλοκα ζητήματα και να τα παρουσιάζουν με ακραίους όρους, καλλιεργώντας την πόλωση και αποθαρρύνοντας την κριτική σκέψη. [14, 15]

Η επανάληψη είναι μια άλλη κοινή τακτική που χρησιμοποιείται στην προπαγάνδα για την ενίσχυση των μηνυμάτων και των ιδεών, καθιστώντας τα πιο οικεία και πιστευτά στο κοινό με την

πάροδο του χρόνου. Επιπλέον, η προπαγάνδα παρουσιάζει επιλεκτικά πληροφορίες που υποστηρίζουν μια συγκεκριμένη αφήγηση, ενώ αγνοεί ή διαστρεβλώνει τις αντίθετες απόψεις, με αποτέλεσμα μια μεροληπτική ή μονόπλευρη απεικόνιση της πραγματικότητας. Η προπαγάνδα συχνά χειραγωγεί σύμβολα, συνθήματα ή εικόνες για να προκαλέσει ισχυρούς συνειρμούς και να χειραγωγήσει τις αντιλήψεις. Αυτό μπορεί να περιλαμβάνει τη χρήση πατριωτικών συμβόλων, ηρωικών μορφών ή τη δαιμονοποίηση των αντιπάλων για την ενίσχυση των επιθυμητών αφηγήσεων και τον επηρεασμό της κοινής γνώμης. [14, 15]

Η προπαγάνδα μπορεί να εκδηλωθεί με διάφορες μορφές, συμπεριλαμβανομένης της πολιτικής προπαγάνδας που χρησιμοποιείται από κυβερνήσεις ή πολιτικά κόμματα για να επηρεάσει την κοινή γνώμη, της πολεμικής προπαγάνδας που χρησιμοποιείται κατά τη διάρκεια συγκρούσεων για να κινητοποιήσει την υποστήριξη ή να δαιμονοποιήσει τους εχθρούς, της εταιρικής προπαγάνδας για να διαμορφώσει τις δημόσιες αντιλήψεις ή να προωθήσει προϊόντα και της ιδεολογικής προπαγάνδας που προωθεί συγκεκριμένες πεποιθήσεις ή κοσμοθεωρίες. Η αναγνώριση της προπαγάνδας απαιτεί κριτική σκέψη και δεξιότητες γραμματισμού στα μέσα ενημέρωσης. Τα άτομα θα πρέπει να αμφισβητούν τις πηγές πληροφοριών, να αξιολογούν την αξιοπιστία των πηγών και να εξετάζουν εναλλακτικές προοπτικές προτού αποδεχθούν ή μοιραστούν περιεχόμενο που μπορεί να έχει προπαγανδιστικό χαρακτήρα. [15]

2.3. Κύρια χαρακτηριστικά των ψευδών ειδήσεων

Οι ψευδείς ειδήσεις είναι ένα σύνθετο φαινόμενο που περιλαμβάνει διάφορες μορφές παραπλανητικών ή κατασκευασμένων πληροφοριών που διαδίδονται μέσω των καναλιών των μέσων ενημέρωσης. Αρκετά βασικά χαρακτηριστικά καθορίζουν τις ψευδείς ειδήσεις και συμβάλλουν στον αντίκτυπό τους στην αντίληψη και τον δημόσιο διάλογο.

2.3.1. Κατασκευασμένες ειδήσεις και στρατηγικές παραπληροφόρησης

Στον πυρήνα των ψευδών ειδήσεων βρίσκεται η σκόπιμη εξαπάτηση ή παραπληροφόρηση, με τους υπαίτιους να αξιοποιούν σχετικές τεχνικές μέσα σε αληθοφανές κείμενο ή οπτικοακουστικό υλικό. Οι ψευδείς ειδήσεις περιλαμβάνουν τη διάδοση ψευδών ή παραπλανητικών πληροφοριών που παρουσιάζονται ως πραγματικές ειδήσεις. Αυτό μπορεί να περιλαμβάνει παραπλανητικούς τίτλους και μικρές, στοχευμένες αλλαγές του αρχικού κειμένου που αποσκοπούν στην παραπληροφόρηση ή την εξαπάτηση του κοινού. Επίσης, ο τρόπος γραφής ή σύνθεσης υλικού που διαθέτει την πρόθεση εξαπάτησης συχνά εμποδίζει την αναγνώριση των ψευδών ειδήσεων, καθώς παραπέμπει στο ύφος αληθινών πηγών. [16]

2.3.2. Συναισθηματισμοί και συναισθηματική έκκληση

Οι ψεύτικες ειδήσεις συχνά βασίζονται σε εντυπωσιοθηρικό περιεχόμενο ή συναισθηματικές εκκλήσεις για να προσελκύσουν την προσοχή του κοινούο. Οι τίτλοι και οι ιστορίες μπορούν να διαμορφωθούν έτσι ώστε να προκαλέσουν έντονες συναισθηματικές αντιδράσεις, όπως φόβο,

οργή ή περιέργεια, οδηγώντας σε αυξημένη κοινοποίηση και virality στα μέσα κοινωνικής δικτύωσης. [17]

2.3.3. Έλλειψη διαφάνειας και επαλήθευσης

Οι πηγές ψευδών ειδήσεων συνήθως στερούνται διαφάνειας και αξιοπιστίας. Μπορεί να προέρχονται από ανώνυμες ή μη επαληθευμένες πηγές, καθιστώντας δύσκολο για το κοινό να αξιολογήσει την αξιοπιστία των πληροφοριών που παρουσιάζονται. Ακόμη, ένας πιο προχωρημένος τρόπος διάδοσης ψευδών ειδήσεων είναι ένας κακόβουλος χρήστης να 'χακάρει' ένα σχετικό μέσο, διαδίδοντας από εκεί ψευδείς ειδήσεις. Στο σημείο αυτό, διαφαίνεται η σημασία της διασταύρωσης των διαδιδόμενων γεγονότων, ακόμη κι αν αυτά εμφανίζονται σε αξιόπιστες πηγές. [5]

2.3.4. Μεροληψία επιβεβαίωσης και θάλαμοι απήχησης

Συχνά οι ψεύτικες ειδήσεις είναι δομημένες με τρόπο ώστε να ενισχύουν τις υπάρχουσες πεποιθήσεις και προκαταλήψεις, εξυπηρετώντας συγκεκριμένες ιδεολογικές ή κομματικές προτιμήσεις. Το κοινό μπορεί να είναι πιο πιθανό να αποδεχτεί και να μοιραστεί ψευδείς ειδήσεις που ευθυγραμμίζονται με την κοσμοθεωρία του, συμβάλλοντας στη διάδοση της παραπληροφόρησης μέσα σε θαλάμους απήχησης. Σε αυτό το φαινόμενο συμβάλει η προώθηση συγκεκριμένων ψευδών ειδήσεων στο κατάλληλο κοινό, εφόσον οι δημιουργοί των ψευδών ειδήσεων γνωρίζουν ότι η ενσωμάτωση των απόψεων και η διάδοσή τους είναι πιο εύκολο να συμβεί όταν η είδηση απευθύνεται σε κοινό που ενστερνίζεται κοινές απόψεις. [18]

2.3.5. Ταχεία εξάπλωση και virality

Οι ψευδείς ειδήσεις διαδίδονται γρήγορα και ευρέως στις πλατφόρμες κοινωνικής δικτύωσης λόγω αλγορίθμων που δίνουν προτεραιότητα στη δέσμευση (engagement) των χρηστών και το εντυπωσιακό (viral) περιεχόμενο. Ο χαρακτήρας των ψευδών ειδήσεων μπορεί να οδηγήσει στην ευρεία διάδοσή τους πριν το κοινό ή οι ίδιες οι κοινωνικές πλατφόρμες μπορούν να προβούν σε έλεγχο των γεγονότων, γι' αυτό και οι δημιουργοί τους επενδύουν στο να υιοθετήσουν το απαραίτητο ύφος για να εξασφαλίσουν την ανάλογη προώθηση από τους αυτόματους αλγορίθμους. [17]

2.3.6. Πρόθεση παραπλάνησης ή χειραγώγησης

Συχνά δημιουργούνται και διαδίδονται ψευδείς ειδήσεις με σκοπό την παραπλάνηση, τη χειραγώγηση της κοινής γνώμης ή την επίτευξη συγκεκριμένων πολιτικών, κοινωνικών ή οικονομικών στόχων. Οι κακόβουλοι φορείς μπορούν να εκμεταλλευτούν την παραπληροφόρηση για στρατηγικούς σκοπούς, όπως ο επηρεασμός των εκλογών ή η υπονόμευση της εμπιστοσύνης στους θεσμούς. Και στο χαρακτηριστικό αυτό, οι δημιουργοί ψευδών ειδήσεων εκμεταλλεύονται συνήθεις τεχνικές δόμησης του λόγου ή του υλικού γενικότερα ώστε να χειραγωγήσουν το -συνήθως στοχευμένο- κοινό. [19]

2.3.7. Προκλήσεις για τη γνώση τρόπου χρήσης των μέσων ενημέρωσης

Οι ψεύτικες ειδήσεις θέτουν προκλήσεις για τη γνώση τρόπου χρήσης (literacy) στα μέσα ενημέρωσης και τις δεξιότητες κριτικής σκέψης. Το κοινό μπορεί να δυσκολεύεται να διακρίνει μεταξύ αξιόπιστων πηγών και ψευδών πληροφοριών, ιδίως στο σημερινό ψηφιακό τοπίο πληροφόρησης. Άνθρωποι χωρίς εξοικείωση με το διαδίκτυο είναι εύκολοι στόχοι, και για το λόγο αυτό οι δημιουργοί ψευδών ειδήσεων μπορούν να δομήσουν περιεχόμενο στοχεύοντας σε μη ενημερωμένο κοινό, για παράδειγμα ηλικιωμένους ανθρώπους ή εφήβους. [20]

2.4. Γιατί οι ψευδείς ειδήσεις είναι πιο επίκαιρες από ποτέ;

Πληθώρα παραγόντων έχει καταστήσει τις ψευδείς ειδήσεις ιδιαιτέρως επίκαιρες, και μάλιστα πρόκειται για ένα ζήτημα που τώρα απασχολεί αλλά και θα απασχολήσει διεπιστημονικά την κοινωνία και στο μέλλον. Οι κυριότεροι παράγοντες, οι οποίοι μπορεί να είναι είτε κοινωνικοί είτε τεχνολογικοί αναλύονται εν συνεχεία. [5, 17, 19, 26, 27, 29]

2.4.1. Διάδοση των ψηφιακών μέσων

Η ευρεία υιοθέτηση των ψηφιακών τεχνολογιών και οι πλατφόρμες κοινωνικής δικτύωσης έχουν εκδημοκρατίσει τη δημιουργία και τη διανομή περιεχομένου. Οποιοσδήποτε έχει πρόσβαση στο διαδίκτυο μπορεί να παράγει και να διαδίδει πληροφορίες, οδηγώντας σε πολλαπλασιασμό των ειδησεογραφικών πηγών, ορισμένες από τις οποίες μπορεί να είναι αναξιόπιστες ή σκόπιμα παραπλανητικές.

Ένα παράδειγμα ψευδούς είδησης μέσω ψηφιακών μέσων διαδόθηκε κατά τη διάρκεια της πανδημίας COVID-19, κυκλοφορώντας στα κοινωνικά δίκτυα (Facebook, Twitter, WhatsApp κ.ά.) ότι το πόσιμο νερό ή οι “γαργάρες” με αλάτι σκοτώνει τον ιό και προστατεύει από τη μόλυνση. Δεν υπήρξαν ποτέ επιστημονικά δεδομένα που να υποστηρίζουν αυτή την “πρακτική” αντιμετώπισης του κορονοϊού, ωστόσο αυτό δείχνει πώς τα ψηφιακά μέσα μπορούν να ενισχύσουν την ταχύτητα και την εμβέλεια της παραπληροφόρησης.

2.4.2. Ευκολία ανταλλαγής πληροφοριών

Οι πλατφόρμες των μέσων κοινωνικής δικτύωσης επιτρέπουν την ταχεία και ευρεία ανταλλαγή περιεχομένου, ενισχύοντας την εμβέλεια και τον αντίκτυπο των ψευδών ειδήσεων. Οι πληροφορίες μπορούν να εξαπλωθούν ιογενώς μέσα σε "θαλάμους απήχησης", όπου οι χρήστες εκτίθενται σε ομοϊδεατικές απόψεις και μπορεί να είναι πιο επιρρεπείς στην παραπληροφόρηση.

Κατά τις πυρκαγιές στην Αυστραλία το 2020, κυκλοφόρησαν στα social media εικόνες που υποτίθεται ότι έδειχναν ζώα να καίγονται ζωντανά ή δάση να έχουν καταστραφεί ολοσχερώς. Οι φωτογραφίες αυτές ήταν ψηφιακά επεξεργασμένες ή τελείως ψεύτικες (π.χ. δημιουργημένες από AI ή ψηφιακή τέχνη). Αυτό είναι κλασικό παράδειγμα του πώς η ευκολία με την οποία ανταλλάσσονται πληροφορίες οδηγεί σε ταχύτατη διάδοση ψευδών ειδήσεων.

2.4.3. Αλγοριθμική ενίσχυση

Οι αλγόριθμοι που χρησιμοποιούνται από τις πλατφόρμες κοινωνικής δικτύωσης δίνουν προτεραιότητα στο ελκυστικό και εντυπωσιακό περιεχόμενο για να μεγιστοποιήσουν τη δέσμευση και τη διατήρηση των χρηστών. Αυτή η αλγοριθμική ενίσχυση μπορεί ακούσια να προωθήσει ψευδείς ειδήσεις και εντυπωσιακό αλλά ψευδές περιεχόμενο έναντι των πραγματικών ειδήσεων.

Στις αμερικανικές εκλογές του 2016, ευρέως γνωστή ήταν μια ψεύτικη είδηση ότι ο Πάπας υποστηρίζει τον Ντόναλντ Τραμπ. Αυτή η είδηση ήταν απολύτως ψευδής, αλλά έγινε ευρέως γνωστή, διότι οι αλγόριθμοι το πρότειναν σε πολλούς χρήστες λόγω του μεγάλου ενδιαφέροντος και της αλληλεπίδρασης που προκάλεσε. Έτσι μπορούμε να καταλάβουμε ότι η αλγοριθμική ενίσχυση δεν είναι απλώς ένα τεχνικό θέμα, αλλά επηρεάζει άμεσα το ποια πληροφορία βλέπει ο κόσμος.

2.4.4. Ψυχολογικοί παράγοντες

Τα άτομα είναι πιο επιρρεπή στο να πιστεύουν και να μοιράζονται ψεύτικες ειδήσεις που ευθυγραμμίζονται με τις προ υπάρχουσες πεποιθήσεις τους ή φαίνονται οικείες λόγω επανάληψης. Ακόμη, μπορεί να πιστεύουν σε ψευδείς ειδήσεις που ενισχύουν την ελπίδα τους σε τομείς που το έχουν ανάγκη, όπως συχνά σε θέματα υγείας (για παράδειγμα, τείνουν να πιστέψουν μια είδηση για ένα νέο φάρμακό ή θεραπεία) ή σε οικονομικά ζητήματα (ακολουθώντας αμφισβητήσιμα άρθρα που προωθούν εναλλακτικούς τρόπους απόκτησης εισοδήματος). Οι έντονοι ρυθμοί της σύγχρονης καθημερινότητας και η αστικοποίηση εντείνουν τους ψυχολογικούς παράγοντες που οδηγούν στην ανάγκη υιοθέτησης και διάδοσης των ψευδών ειδήσεων.

Ένα χαρακτηριστικό παράδειγμα είναι ότι πολλοί πίστεψαν τη θεωρία ότι το 5G προκαλεί COVID-19. Προφανώς και δεν υπήρξε καμία αποδεδειγμένη επιστημονικά θεωρία, αλλά αυτή η είδηση εξαπλώθηκε ραγδαία επειδή προσέφερε μια εύκολη εξήγηση για μια δύσκολη κατάσταση εκείνη την περίοδο καθώς και το γεγονός ότι προκάλεσε έντονο φόβο και οργή, άρα μοιράστηκε γρήγορα. Έτσι οι ψυχολογικές ανάγκες και αντιδράσεις των ανθρώπων συμβάλλουν καθοριστικά στη διάδοση ψευδών ειδήσεων.

2.4.5. Κομματική πόλωση

Οι κοινωνικοί διχασμοί και η πολιτική πόλωση συμβάλλουν στη ζήτηση ειδήσεων που ενισχύουν τις υπάρχουσες πεποιθήσεις και προκαταλήψεις. Οι ψεύτικες ειδήσεις μπορούν να εκμεταλλευτούν αυτές τις διαιρέσεις εξυπηρετώντας συγκεκριμένες ιδεολογικές ή κομματικές προτιμήσεις, βαθαίνοντας περαιτέρω τα κοινωνικά ρήγματα. Ιδιαιτέρως σε περιόδους εκλογών ή σε περιπτώσεις πολιτικών αναταραχών και μεταβολών, οι ψευδείς ειδήσεις τείνουν να καταστούν ένα κρίσιμο κοινωνικό πρόβλημα. Η έλλειψη εμπιστοσύνης σε πολιτικά πρόσωπα και ομάδες, η οποία έχει ενισχυθεί μετά την παγκόσμια οικονομική κρίση, ενισχύει την ευκολία διάδοσης ψευδών ειδήσεων, καθώς το κοινό γίνεται ένας πιο εύκολος αποδέκτης, σε μια προσπάθεια εύρεσης διεξόδων από το παρόν πολιτικό, οικονομικό και κοινωνικό σκηνικό.

Χαρακτηριστικό είναι ότι στις εκλογές των ΗΠΑ το 2020, κυκλοφόρησαν ψευδείς ειδήσεις ότι το εκλογικό σύστημα αλλοιώθηκε υπέρ των Δημοκρατικών. Παρότι δεν υπήρχαν αποδείξεις, πολλοί

πολίτες τις πίστεψαν επειδή δεν ήθελαν να πιστέψουν στην ήττα του κόμματός τους. Έτσι, η κομματική πόλωση ενισχύει την παραπληροφόρηση, αφού δημιουργεί ένα περιβάλλον όπου η αντικειμενικότητα παραμερίζεται μπροστά στην πολιτική ταύτιση.

2.4.6. Εκστρατείες επιρροής και τακτικές παραπληροφόρησης

Οι κρατικοί φορείς, οι πολιτικές εκστρατείες και οι ομάδες ειδικών συμφερόντων χρησιμοποιούν εξελιγμένες τακτικές παραπληροφόρησης, συμπεριλαμβανομένων των ψευδών ειδήσεων, για να χειραγωγήσουν την κοινή γνώμη, να επηρεάσουν εκλογές και να υπονομεύσουν την εμπιστοσύνη στους δημοκρατικούς θεσμούς. Κάτι τέτοιο είναι πιο επίκαιρο από ποτέ, δεδομένου ότι οι πολιτικοί μπορούν να απευθυνθούν σε ευρύτερο κοινό λόγω της εξάπλωσης των μέσων ενημέρωσης σε μεγάλη μερίδα του πληθυσμού, όπως επίσης και λόγω της ευρείας χρήσης μέσων κοινωνικής δικτύωσης.

Κατά τη διάρκεια της ρωσικής εισβολής στην Ουκρανία το 2022 [118], υπήρξε τεράστιος όγκος ψευδών ειδήσεων που διακινήθηκαν από κρατικά ελεγχόμενα μέσα και προφίλ στα social media, με σκοπό να δικαιολογήσουν την εισβολή, κατηγορούσαν την Ουκρανία για επιθέσεις που έκανε η Ρωσία (αντιστροφή γεγονότων), αλλά και να προβάλουν ψεύτικες εικόνες για να χειραγωγήσουν τη διεθνή κοινή γνώμη.

2.4.7. Οικονομία της προσοχής

Στην οικονομία της προσοχής (attention economy) όπου η προσοχή είναι σπάνιος πόρος- το εντυπωσιακό και προκλητικό περιεχόμενο τείνει να συγκεντρώνει μεγαλύτερη δέσμευση από πλευράς χρηστών. Οι δημιουργοί ψεύτικων ειδήσεων επωφελούνται από αυτό, παράγοντας δολώματα κλικ (clickbaits) και συναισθηματικά φορτισμένο περιεχόμενο που έχει σχεδιαστεί για να τραβήξει την προσοχή και να δημιουργήσει έσοδα.

Ένα αρκετά πλέον διαδομένο παράδειγμα στις μέρες μας είναι ότι οι ιστοσελίδες παράγουν ψευδείς ειδήσεις με τίτλους όπως «Σοκ! Ο διάσημος ηθοποιός βρέθηκε νεκρός». Με αυτόν τον τρόπο κερδίζουν τεράστιο αριθμό κλικ, και άρα έσοδα από διαφημίσεις, παρόλο που το περιεχόμενο είναι ψευδές.

2.4.8. Παρακμή των παραδοσιακών “φρουρών” των μέσων ενημέρωσης

Η διάβρωση των παραδοσιακών αξιόπιστων μέσων ενημέρωσης, όπως οι επαγγελματίες δημοσιογράφοι και τα συντακτικά πρότυπα, έχει μειώσει τους φραγμούς εισόδου για την παραπληροφόρηση. Χωρίς ισχυρές διαδικασίες ελέγχου και επαλήθευσης των γεγονότων, οι ψευδείς ειδήσεις μπορούν εύκολα να κυκλοφορούν στο διαδίκτυο ανεξέλεγκτα.

Παλαιότερα, μια μεγάλη εφημερίδα όπως η Καθημερινή στην Ελλάδα ή οι New York Times στο εξωτερικό δεν θα δημοσίευε μια είδηση χωρίς επαλήθευση, χωρίς έλεγχο πηγών. Σήμερα, όμως, οποιαδήποτε ιστοσελίδα ή blog μπορεί να διαδώσει μια ψευδή είδηση, χωρίς κανένα έλεγχο που μοιάζει αληθινή για να φτάσει σε εκατομμύρια ανθρώπους. Έτσι καταλαβαίνουμε ότι η αποδυνάμωση των παραδοσιακών “φρουρών” της ενημέρωσης έχει διευκολύνει τη διάδοση ψευδών ειδήσεων.

2.4.9. Παγκόσμιες κρίσεις

Οι παγκόσμιες κρίσεις όπως πανδημίες, πόλεμοι, οικονομικές καταρρεύσεις, φυσικές καταστροφές είναι καταστάσεις που ευνοούν τη διάδοση ψευδών ειδήσεων, γι' αυτό και αποτελούν σημαντικό λόγο για τη σημερινή τους επικαιρότητα. Οι κρίσεις ενισχύουν τις ψευδείς ειδήσεις, διότι ο κόσμος αισθάνεται φόβο, άγχος και αβεβαιότητα. Υπάρχει δίψα για άμεση πληροφόρηση, ακόμα και πριν υπάρξουν επίσημα δεδομένα. Πολλές φορές οι ψευδείς ειδήσεις προσφέρουν καθησυχάζουν τους πολίτες. Τέλος, υπάρχουν και αυτοί οι οποίοι σε τέτοιες περιόδους εκμεταλλεύονται αυτές τις καταστάσεις για προπαγάνδα για να "σπείρουν" τον πανικό ή ακόμα και για κερδοσκοπικούς σκοπούς.

Ένα τρανταχτό παράδειγμα πρόσφατης παγκόσμιας κρίσης αποτελεί η πανδημία του COVID – 19, όπου οι ψευδείς ειδήσεις σχετικά με τα εμβόλια, τις θεραπείες, τις πολιτικές των κρατών και τα στατιστικά που προέκυψαν ως αποτέλεσμα της πανδημίας (κρούσματα, νοσηλείες, θάνατοι) αποτελούσαν πηγές για τη δημιουργία των ψευδών ειδήσεων. Με αφορμή την πανδημία κατάφεραν να διαδοθούν ποικίλες απόψεις, εφόσον κατάφεραν να βρουν το κατάλληλο 'όχημα' για την πόλωση κοινωνικών ομάδων και την ενίσχυση συγκεκριμένων πεποιθήσεων.

2.5. Τρόποι διάδοσης ψευδών ειδήσεων

Οι ψεύτικες ειδήσεις διαδίδονται μέσω διαφόρων μηχανισμών, συχνά αξιοποιώντας τη διασυνδεδεμένη φύση των ψηφιακών πλατφορμών επικοινωνίας και την εκμετάλλευση των γνωστικών προκαταλήψεων και των κοινωνικών δυναμικών.

2.5.1. Κοινοποίηση στα μέσα κοινωνικής δικτύωσης

Οι πλατφόρμες των μέσων κοινωνικής δικτύωσης έχουν φέρει επανάσταση στον τρόπο διάδοσης των πληροφοριών, παρέχοντας μια άνευ προηγουμένου ευκαιρία για την ταχεία διάδοση ψευδών ειδήσεων. Ένας από τους πρωταρχικούς μηχανισμούς μέσω των οποίων διαδίδονται οι ψευδείς ειδήσεις στα μέσα κοινωνικής δικτύωσης είναι ο διαμοιρασμός. Με δισεκατομμύρια χρήστες παγκοσμίως, πλατφόρμες, όπως το Facebook, το Twitter και το Instagram, αποτελούν πρόσφορο έδαφος για τη διάδοση της παραπληροφόρησης. Τα άρθρα ψευδών ειδήσεων, τα μιμίδια (memes) και οι εντυπωσιοθηρικοί τίτλοι μπορούν να κερδίσουν γρήγορα έδαφος, καθώς οι χρήστες τα μοιράζονται με τα δίκτυά τους, συχνά χωρίς να αξιολογούν κριτικά την ακρίβειά τους. Για παράδειγμα, κατά τη διάρκεια των προεδρικών εκλογών του 2016 στις ΗΠΑ, πολλές ψευδείς ιστορίες και θεωρίες συνωμοσίας κυκλοφόρησαν στα μέσα κοινωνικής δικτύωσης, συμπεριλαμβανομένων ισχυρισμών περί απάτης ψηφοφόρων και κατασκευασμένων εγκρίσεων.

Οι ιστορίες αυτές κοινοποιήθηκαν ευρέως και ενισχύθηκαν από τους χρήστες, συμβάλλοντας στη διάβρωση της εμπιστοσύνης στην εκλογική διαδικασία και διαμορφώνοντας την κοινή γνώμη. Ομοίως, στο πλαίσιο της πανδημίας COVID-19, τα μέσα κοινωνικής δικτύωσης κατακλύστηκαν από παραπληροφόρηση σχετικά με τον ιό, την προέλευσή του και τις πιθανές θεραπείες. Οι ψευδείς ισχυρισμοί σχετικά με την αποτελεσματικότητα των μη αποδεδειγμένων θεραπειών ή τη

φύση του ίδιου του ιού εξαπλώθηκαν γρήγορα, τροφοδοτούμενοι από την τάση των χρηστών να μοιράζονται πληροφορίες που ευθυγραμμίζονται με τις πεποιθήσεις ή τα συναισθήματά τους. Ο viral χαρακτήρας των μέσων κοινωνικής δικτύωσης ενισχύει περαιτέρω την εμβέλεια και την ορατότητα των ψευδών ειδήσεων, καθιστώντας δύσκολο τον περιορισμό ή την αντιμετώπισή τους. Ως αποτέλεσμα, η αντιμετώπιση του προβλήματος των ψευδών ειδήσεων στα μέσα κοινωνικής δικτύωσης απαιτεί μια πολύπλευρη προσέγγιση, που περιλαμβάνει αυξημένη παιδεία στα μέσα ενημέρωσης, πρωτοβουλίες ελέγχου των γεγονότων και παρεμβάσεις σε επίπεδο πλατφόρμας για τον περιορισμό της εξάπλωσης της παραπληροφόρησης.

2.5.2. Clickbait τίτλοι

Οι τίτλοι clickbait αντιπροσωπεύουν μια κοινή στρατηγική που χρησιμοποιούν οι προμηθευτές ψευδών ειδήσεων για να προσελκύσουν κλικ και συμμετοχή. Αυτοί οι τίτλοι συχνά δημιουργούνται για να προκαλέσουν περιέργεια, να προκαλέσουν έντονα συναισθήματα ή να υποσχεθούν εντυπωσιακές αποκαλύψεις, ανεξάρτητα από την αλήθεια του περιεχομένου. Για παράδειγμα, κατά τη διάρκεια των αμερικανικών προεδρικών εκλογών του 2016, τίτλοι όπως "BREAKING: Hillary Clinton Linked to Child Trafficking Ring!" και "SHOCKING: Donald Trump Exposed as Russian Agent!" κυκλοφόρησαν ευρέως στα μέσα κοινωνικής δικτύωσης, παρά το γεγονός ότι δεν υπήρχαν αξιόπιστα στοιχεία που να υποστηρίζουν τους ισχυρισμούς τους. Αυτοί οι εντυπωσιοθηρικοί τίτλοι εκμεταλλεύτηκαν την περιέργεια και την επιθυμία των χρηστών για εντυπωσιακές ειδήσεις, ωθώντας τους να κάνουν κλικ στα σχετικά άρθρα (ειδήσεων) και να τα μοιραστούν με άλλους.

Η αποτελεσματικότητα των clickbait τίτλων έγκειται στην ικανότητά τους να τραβούν την προσοχή και να προκαλούν συναισθηματική αντίδραση, παρακάμπτοντας συχνά την κριτική ικανότητα των χρηστών και οδηγώντας τους να καταναλώνουν και να μοιράζονται άκριτα παραπληροφόρηση. Ως εκ τούτου, η καταπολέμηση των ψευδών ειδήσεων απαιτεί όχι μόνο την αντιμετώπιση του ίδιου του περιεχομένου αλλά και την αντιμετώπιση των τακτικών που χρησιμοποιούνται για να παρασύρουν τους χρήστες να ασχοληθούν με αυτό. Αυτό περιλαμβάνει την προώθηση δεξιοτήτων γραμματισμού στα μέσα ενημέρωσης που θα βοηθήσουν τους χρήστες να αναγνωρίζουν και να αντιστέκονται στο clickbait, καθώς και την εφαρμογή μέτρων για τη μείωση της εμβέλειας και της ορατότητας του παραπλανητικού περιεχομένου στις ψηφιακές πλατφόρμες.

2.5.3. Εκμετάλλευση των αλγορίθμων προώθησης περιεχομένου

Η αλγοριθμική ενίσχυση επιδεινώνει την εξάπλωση των ψευδών ειδήσεων, αξιοποιώντας τους αλγόριθμους των μηχανών αναζήτησης μέσων κοινωνικής δικτύωσης που ιεραρχούν το περιεχόμενο, δίνοντας προτεραιότητα σε παράγοντες, όπως η εμπλοκή, η συνάφεια και η συχνότητα αλληλεπίδρασης. Με την εκμετάλλευση αυτών των αλγορίθμων, οι προμηθευτές ψευδών ειδήσεων μπορούν να εξασφαλίσουν ότι το περιεχόμενό τους εμφανίζεται σε περίοπτη θέση στα αποτελέσματα αναζήτησης και στις ροές ειδήσεων, αυξάνοντας την πιθανότητα να το δουν και να το κοινοποιήσουν οι χρήστες.

Για παράδειγμα, πλατφόρμες, όπως το Facebook και το Twitter, χρησιμοποιούν αλγόριθμους που αναλύουν τις αλληλεπιδράσεις των χρηστών, όπως τα likes, οι κοινοποιήσεις και τα σχόλια, για να καθορίσουν την ορατότητα του περιεχομένου στις ροές των χρηστών. Όταν τα άρθρα ψευδών ειδήσεων δημιουργούν υψηλά επίπεδα εμπλοκής, οι αλγόριθμοι αυτοί μπορεί να ερμηνεύουν τις αλληλεπιδράσεις αυτές ως σήματα συνάφειας και δημοτικότητας, ενισχύοντας κατά συνέπεια την εμβέλεια της παραπληροφόρησης. Το φαινόμενο αυτό δημιουργεί έναν βρόχο ανατροφοδότησης, όπου οι ψευδείς ειδήσεις, μόλις αποκτήσουν απήχηση, προωθούνται αλγοριθμικά σε ευρύτερο κοινό, αυξάνοντας περαιτέρω την ορατότητα και τη διάδοσή τους. Για παράδειγμα, κατά τη διάρκεια της πανδημίας COVID-19, οι ψευδείς ισχυρισμοί και οι θεωρίες συνωμοσίας σχετικά με τον ιό και την προέλευσή του πολλαπλασιάστηκαν στις πλατφόρμες κοινωνικής δικτύωσης. Άρθρα που προωθούσαν την παραπληροφόρηση, όπως η ιδέα ότι ο COVID-19 ήταν ένα βιολογικό όπλο που κατασκευάστηκε σε εργαστήριο ή ότι τα εμβόλια περιείχαν μικροτσίπ για τον εντοπισμό ατόμων, συγκέντρωσαν σημαντική συμμετοχή από χρήστες που αναζητούσαν εντυπωσιακό ή συνωμοσιολογικό περιεχόμενο.

Ως αποτέλεσμα, οι αλγόριθμοι των μέσων κοινωνικής δικτύωσης ενίσχυσαν αυτά τα ψεύδη, εξασφαλίζοντας την ευρεία κυκλοφορία τους και συμβάλλοντας στη δημόσια σύγχυση και δυσπιστία. Η αλγοριθμική ενίσχυση υπογραμμίζει έτσι τη σημασία της αντιμετώπισης όχι μόνο του περιεχομένου των ψευδών ειδήσεων αλλά και των υποκείμενων μηχανισμών που διευκολύνουν τη διάδοσή τους στις ψηφιακές πλατφόρμες. Οι προσπάθειες για την καταπολέμηση των ψευδών ειδήσεων πρέπει να περιλαμβάνουν μέτρα για τον μετριασμό της αλγοριθμικής προώθησης της παραπληροφόρησης και την προώθηση της διάδοσης ακριβών και αξιόπιστων πληροφοριών.

Η εμπορευματοποίηση της προσοχής στις ψηφιακές πλατφόρμες έχει δημιουργήσει μια αγορά όπου η παραπληροφόρηση ανταγωνίζεται τις νόμιμες πηγές ειδήσεων, θολώνοντας περαιτέρω τα όρια μεταξύ γεγονότων και μυθοπλασίας. Σε αυτό το πλαίσιο, οι δημιουργοί μπορεί να έχουν ισχυρό κίνητρο να παράγουν ανάλογο περιεχόμενο, συμπεριλαμβανομένων των ψευδών ειδήσεων ή των τίτλων, για να προσελκύσουν την προσοχή και να δημιουργήσουν κοινό, το οποίο αλληλεπιδρά με το περιεχόμενό τους σε τακτά χρονικά διαστήματα. Επιπλέον, ορισμένα άτομα ή ομάδες μπορεί να εκμεταλλεύονται λειτουργίες νομισματοποίησης (monetization) των μέσων κοινωνικής δικτύωσης, όπως προγράμματα διαμοιρασμού εσόδων από διαφημίσεις, για να επωφεληθούν από τη διάδοση περιεχομένου ψευδών ειδήσεων. Εκμεταλλευόμενοι τον viral χαρακτήρα των μέσων κοινωνικής δικτύωσης και τη δυνατότητα του περιεχομένου να παράγει έσοδα, τα συστήματα αυτά μπορεί να ενισχύσουν ακούσια τη διάδοση της παραπληροφόρησης και της παραπληροφόρησης.

Επιπρόσθετα, τα οικονομικά κίνητρα που συνδέονται με την κερδοφορία των μέσων κοινωνικής δικτύωσης μπορεί να δώσουν κίνητρα για τη δημιουργία και την προώθηση εντυπωσιοθηρικού ή πολωτικού περιεχομένου, συμπεριλαμβανομένων των ψευδών ειδήσεων, για τη μεγιστοποίηση της κερδοφορίας. Ως αποτέλεσμα, τα συστήματα νομισματοποίησης στις πλατφόρμες των μέσων

κοινωνικής δικτύωσης μπορούν να επιδεινώσουν τη διάδοση ψευδών ειδήσεων και να συμβάλουν στη διάδοσή τους στο διαδίκτυο.

Ενώ οι τακτικές βελτιστοποίησης μηχανών αναζήτησης (SEO - Search Engine Optimization) μπορούν να συμβάλουν έμμεσα στην προβολή ψευδών ειδήσεων επηρεάζοντας την κατάταξη των ιστότοπων στα αποτελέσματα των μηχανών αναζήτησης, δεν εμπίπτουν εγγενώς στην ίδια κατηγορία με την αλγοριθμική ενίσχυση στις πλατφόρμες κοινωνικής δικτύωσης. Οι τεχνικές SEO επικεντρώνονται κυρίως στη βελτιστοποίηση του περιεχομένου ενός ιστότοπου για τη βελτίωση της κατάταξής του στις σελίδες αποτελεσμάτων μηχανών αναζήτησης (SERPs - search engine results pages), αυξάνοντας έτσι την οργανική επισκεψιμότητα στον ιστότοπο. Ωστόσο, οι ανήθικες πρακτικές SEO, όπως το γέμισμα λέξεων-κλειδιών, η απόκρυψη ή τα συστήματα συνδέσμων, μπορούν ενδεχομένως να χρησιμοποιηθούν για τη χειραγώγηση των αλγορίθμων των μηχανών αναζήτησης από κακόβουλους παράγοντες με σκοπό την προώθηση ιστοσελίδων ή άρθρων με ψευδείς ειδήσεις. Για παράδειγμα, οι εκδότες ψευδών ειδήσεων μπορούν να χρησιμοποιούν στρατηγικές SEO για να αυξήσουν τεχνητά την ορατότητα και την εμβέλεια του περιεχομένου τους στα αποτελέσματα αναζήτησης, καθιστώντας το πιο πιθανό να ανακαλυφθεί από χρήστες που αναζητούν πληροφορίες για ένα συγκεκριμένο θέμα.

2.5.4. Παραποιημένο οπτικοακουστικό υλικό

Οι παραποιημένες εικόνες και τα βίντεο είναι ισχυρά εργαλεία για τη διάδοση ψευδών ειδήσεων, καθώς μπορούν να δημιουργήσουν μια λανθασμένη αντίληψη της πραγματικότητας. Ένα παράδειγμα είναι η ευρεία κοινοποίηση ψηφιακά τροποποιημένων εικόνων κατά τη διάρκεια πολιτικών γεγονότων ή κρίσεων, όπου οι εικόνες επεξεργάζονται για να παραποιήσουν γεγονότα ή καταστάσεις. Για παράδειγμα, η εικόνα ενός πολιτικού ηγέτη μπορεί να τροποποιηθεί ψηφιακά ώστε να φαίνεται ότι επιδεικνύει ανάρμοστη συμπεριφορά, οδηγώντας σε ψευδείς κατηγορίες και δημόσια κατακραυγή. Ένα άλλο παράδειγμα είναι η χρήση των deepfake βίντεο, τα οποία χρησιμοποιούν τεχνητή νοημοσύνη (artifical intelligence - AI) για να τοποθετήσουν το πρόσωπο ενός ατόμου στο σώμα ενός άλλου, επιτρέποντας τη δημιουργία πειστικών αλλά εντελώς κατασκευασμένων βίντεο. Αυτά τα deepfake βίντεο μπορούν να χρησιμοποιηθούν για τη διάδοση ψευδών πληροφοριών, όπως για παράδειγμα να παρουσιάζουν ένα δημόσιο πρόσωπο να πραγματοποιεί έντονες δηλώσεις που δεν αναφέρθηκαν ποτέ εκ μέρους του στην πραγματικότητα ή να εμφανίζεται σε συμβιβαστικές καταστάσεις στις οποίες δεν ενεπλάκη ποτέ. Και στις δύο περιπτώσεις, οι παραποιημένες εικόνες και τα βίντεο χρησιμοποιούνται για να εξαπατήσουν τους θεατές και να χειραγωγήσουν την κοινή γνώμη, αναδεικνύοντας τους κινδύνους των ψευδών ειδήσεων στην ψηφιακή εποχή.

2.5.5. Αυτοματοποιημένα bots κοινωνικής δικτύωσης.

Τα δίκτυα bot είναι αυτοματοποιημένοι λογαριασμοί που δημιουργούνται για να μιμούνται την ανθρώπινη συμπεριφορά στις πλατφόρμες κοινωνικής δικτύωσης. Αυτά τα bots μπορούν να προγραμματιστούν για να κάνουν μαζικά like, να μοιράζονται ή να αναδημοσιεύουν περιεχόμενο, να κάνουν retweet και να σχολιάζουν αναρτήσεις, ενισχύοντας αποτελεσματικά την εμβέλεια και

την ορατότητα των ψευδών ειδήσεων, δημιουργώντας έτσι την ψευδαίσθηση ευρείας υποστήριξης ή ενδιαφέροντος.

Για παράδειγμα, κατά τη διάρκεια πολιτικών εκλογών, τα δίκτυα bot μπορούν να χρησιμοποιηθούν για τη διάδοση παραπληροφόρησης ή προπαγάνδας προς υποστήριξη ενός συγκεκριμένου υποψηφίου ή ατζέντας. Αυτά τα bots μπορούν να κατακλύσουν τις πλατφόρμες κοινωνικής δικτύωσης με ψευδή άρθρα ειδήσεων, memes(μιμίδια) και hashtags, δημιουργώντας την ψευδαίσθηση μιας ευρείας κοινής γνώμης ή υποστήριξης. Επιπλέον, τα δίκτυα bot μπορούν να στοχεύουν σε συγκεκριμένες δημογραφικές ομάδες ή περιοχές για να επηρεάσουν τον δημόσιο διάλογο και την κοινή γνώμη.

Ένα ακόμα παράδειγμα, κατά τη διάρκεια των προεδρικών εκλογών του 2016 στις ΗΠΑ, δίκτυα bot χρησιμοποιήθηκαν για τη διάδοση διχαστικών μηνυμάτων και παραπληροφόρησης, σπέρνοντας διχόνοια και σύγχυση μεταξύ των ψηφοφόρων. Αξιοποιώντας τα δίκτυα bot, οι κακοί φορείς μπορούν να χειραγωγούν τις διαδικτυακές συζητήσεις, να διαστρεβλώνουν την πραγματικότητα και να υπονομεύουν την εμπιστοσύνη στις νόμιμες πηγές ειδήσεων, αναδεικνύοντας τις προκλήσεις που θέτουν οι ψευδείς ειδήσεις στην ψηφιακή εποχή.

2.5.6. Φαινόμενο θαλάμου απήχησης

Οι θάλαμοι απήχησης (echo chamber) αναφέρονται σε διαδικτυακά περιβάλλοντα όπου τα άτομα εκτίθενται κυρίως σε πληροφορίες και απόψεις που ενισχύουν τις υπάρχουσες πεποιθήσεις και προοπτικές τους. Μέσα σε αυτούς τους θαλάμους απήχησης, οι χρήστες είναι λιγότερο πιθανό να συναντήσουν διαφορετικές απόψεις ή κριτικές προοπτικές, οδηγώντας στην ενίσχυση των δικών τους προκαταλήψεων και ιδεολογιών.

Για παράδειγμα, στις πλατφόρμες κοινωνικής δικτύωσης, οι χρήστες μπορεί να ακολουθούν λογαριασμούς ή να συμμετέχουν σε ομάδες που μοιράζονται τις πολιτικές, κοινωνικές ή πολιτισμικές πεποιθήσεις τους, δημιουργώντας έναν αυτοενισχυόμενο κύκλο κατανάλωσης πληροφοριών. Μέσα σε αυτούς τους θαλάμους απήχησης, οι ψευδείς ειδήσεις που ευθυγραμμίζονται με την επικρατούσα ιδεολογία ή αφήγηση μπορούν να διαδοθούν ανεξέλεγκτα, καθώς οι χρήστες είναι λιγότερο πιθανό να αξιολογήσουν κριτικά ή να ελέγξουν τις πληροφορίες που υποστηρίζουν την κοσμοθεωρία τους. Το φαινόμενο αυτό μπορεί να συμβάλει στην πόλωση της κοινωνίας, καθώς τα άτομα οχυρώνονται όλο και περισσότερο στις δικές τους ιδεολογικές φούσκες, απορρίπτοντας τις αντίθετες απόψεις και δαιμονοποιώντας όσους έχουν διαφορετικές πεποιθήσεις. Ως αποτέλεσμα, οι ψευδείς ειδήσεις μπορούν να επιδεινώσουν τις κοινωνικές διαιρέσεις και να υπονομεύσουν τη δυνατότητα εποικοδομητικού διαλόγου και συζήτησης, διαιωνίζοντας την παραπληροφόρηση και τη δυσπιστία στη δημόσια σφαίρα.

2.5.7. Στρατηγικές εκστρατείες παραπληροφόρησης

Οι στρατηγικές εκστρατείες παραπληροφόρησης περιλαμβάνουν τη σκόπιμη διάδοση ψευδών ή παραπλανητικών πληροφοριών στα μέσα μαζικής ενημέρωσης και στα κοινωνικά δίκτυα με σκοπό την επίτευξη συγκεκριμένων πολιτικών, κοινωνικών ή στρατηγικών στόχων. Κρατικοί φορείς, πολιτικές οργανώσεις ή άλλες οντότητες μπορούν να ενορχηστρώσουν αυτές τις

εκστρατείες για να επηρεάσουν την κοινή γνώμη, να υπονομεύσουν την εμπιστοσύνη στους δημοκρατικούς θεσμούς ή να προωθήσουν τις δικές τους γεωπολιτικές ατζέντες. Για παράδειγμα, κατά τη διάρκεια εκλογών, ξένες κυβερνήσεις ή πολιτικές ομάδες μπορεί να συμμετέχουν σε εκστρατείες παραπληροφόρησης για να επηρεάσουν τις αντιλήψεις των ψηφοφόρων, να δυσφημίσουν αντίπαλους υποψηφίους ή να σπείρουν τη διαίρεση εντός των χωρών-στόχων, κάτι το οποίο πλέον είναι πολύ ευκολότερο σε σχέση με το παρελθόν χάρη στα ηλετρονικά μέσα ενημέρωσης. Ένα γνωστό παράδειγμα είναι η ανάμειξη της Ρωσίας στις προεδρικές εκλογές του 2016 στις Ηνωμένες Πολιτείες, όπου οντότητες που συνδέονται με το Κρεμλίνο χρησιμοποίησαν πλατφόρμες κοινωνικής δικτύωσης για τη διάδοση ψευδών ειδήσεων και διχαστικού περιεχομένου, με στόχο την ενίσχυση των κοινωνικών εντάσεων και την υπονόμευση της εμπιστοσύνης στην εκλογική διαδικασία [19, 21, 33].

Ομοίως, τα αυταρχικά καθεστώτα μπορούν να χρησιμοποιούν τακτικές παραπληροφόρησης για να καταστείλουν τις διαφωνίες, να δυσφημίσουν τις ομάδες της αντιπολίτευσης ή να παρουσιάσουν τον εαυτό τους σε ευνοϊκό φως διεθνώς. Για παράδειγμα, τα καθεστώτα μπορεί να διαδίδουν ψευδείς αφηγήσεις για εσωτερική αναταραχή ή εξωτερικές απειλές για να δικαιολογήσουν την καταστολή της διαφωνίας και την εδραίωση της εξουσίας. Σε ορισμένες περιπτώσεις, οι εκστρατείες παραπληροφόρησης μπορεί να στοχεύουν σε συγκεκριμένες εθνοτικές ή θρησκευτικές ομάδες για να τροφοδοτήσουν τις ενδοομαδικές εντάσεις ή να υποκινήσουν τη βία, όπως παρατηρήθηκε σε περιπτώσεις εθνοκάθαρσης ή γενοκτονίας. Μη κρατικοί φορείς, όπως εξτρεμιστικές ομάδες ή τρομοκρατικές οργανώσεις, μπορούν επίσης να συμμετάσχουν σε εκστρατείες παραπληροφόρησης για να στρατολογήσουν οπαδούς, να ριζοσπαστικοποιήσουν συμπαθούντες ή να δικαιολογήσουν πράξεις βίας. Οι ομάδες αυτές μπορεί να διαδίδουν προπαγάνδα και ψευδείς αφηγήσεις μέσω διαδικτυακών καναλιών, εκμεταλλευόμενες τις πλατφόρμες των μέσων κοινωνικής δικτύωσης για τη διάδοση των μηνυμάτων τους και τη στρατολόγηση νέων μελών.

Συνολικά, οι στρατηγικές εκστρατείες παραπληροφόρησης αποτελούν σημαντική πρόκληση για τις δημοκρατικές κοινωνίες, καθώς υπονομεύουν την εμπιστοσύνη στα μέσα ενημέρωσης, διαβρώνουν το δημόσιο διάλογο και θέτουν απειλές για την εθνική ασφάλεια και σταθερότητα. Η αντιμετώπιση αυτών των εκστρατειών απαιτεί συντονισμένες προσπάθειες από τις κυβερνήσεις, τις εταιρείες τεχνολογίας, τις οργανώσεις της κοινωνίας των πολιτών και το κοινό για την καταπολέμηση της παραπληροφόρησης, την προώθηση της παιδείας στα μέσα ενημέρωσης και τη διαφύλαξη των δημοκρατικών θεσμών.

2.6. Μέσα διάδοσης ψευδών ειδήσεων

Τα μέσα ψευδών ειδήσεων είναι εργαλεία, δίκτυα και πλατφόρμες που χρησιμοποιούνται για τη διάδοση, ενίσχυση ή κατασκευή ψευδών πληροφοριών. Δεν ευθύνονται απαραίτητα για τη δημιουργία του περιεχομένου, αλλά λειτουργούν ως μορφή μετάδοσης που επιτρέπουν την μεγάλη και ταχεία εξάπλωση ψευδών ειδήσεων στο κοινό.

Στην εποχή της ψηφιακής πληροφορίας, τα μέσα αυτά αξιοποιούνται συχνά από φορείς παραπληροφόρησης για να επηρεάσουν την κοινή γνώμη, να χειραγωγήσουν εκλογικές διαδικασίες ή να διαδώσουν θεωρίες συνωμοσίας. Στην συνέχεια θα δούμε διάφορα μέσα ψευδών ειδήσεων.

2.6.1. Πλατφόρμες κοινωνικής δικτύωσης

Οι πλατφόρμες κοινωνικής δικτύωσης, όπως το Facebook, το X (πρώην Twitter), το Instagram, το TikTok και το YouTube, αποτελούν σήμερα κυρίαρχο μέσο διάδοσης ψευδούς πληροφορίας. Η ευκολία διαμοιρασμού περιεχομένου και η δυνατότητα δημιουργίας "viral" αναρτήσεων τις καθιστούν ιδιαίτερα αποτελεσματικές στη μαζική εξάπλωση ψευδών ειδήσεων.

Τα κοινωνικά δίκτυα λειτουργούν ως μορφή μετάδοσης παραπληροφόρησης λόγω:

- των αλγορίθμων προτεραιοποίησης περιεχομένου (που ευνοούν το εντυπωσιακό ή συναισθηματικά φορτισμένο περιεχόμενο),
- της ευρείας χρήσης bots και αυτοματοποιημένων λογαριασμών που ενισχύουν την απήχηση ψευδών ειδήσεων,
- της δυσκολίας ελέγχου και επαλήθευσης του αναρτώμενου περιεχομένου από τους χρήστες.

Οι ψευδείς ειδήσεις διαδίδονται συχνά μέσω εικόνων, μιμιδίων (memes), βίντεο, hashtags ή "clickbait" τίτλων, δημιουργώντας την ψευδαίσθηση αυθεντικότητας. Ειδικά σε περιόδους κρίσης ή πολιτικής πόλωσης, οι πλατφόρμες αυτές μετατρέπονται σε περιβάλλοντα όπου η παραπληροφόρηση μπορεί να διαδοθεί ταχύτατα και να επηρεάσει τη δημόσια γνώμη.

Χαρακτηριστικό παράδειγμα για την επικινδυνότητα αλλά και την ταχεία εξάπλωση είναι ότι κατά τη διάρκεια του πολέμου στην Ουκρανία μέσω των πλατφορμών TikTok, Telegram και Twitter έγινε διασπορά παραποιημένων βίντεο, ψευδών ισχυρισμών και πολεμικής προπαγάνδας [5, 28, 116, 117, 118, 119].

2.6.2. Botnets και αυτοματοποιημένοι λογαριασμοί

Τα botnets και οι αυτοματοποιημένοι λογαριασμοί είναι βασικά μέσα διάδοσης ψευδών ειδήσεων. Πρόκειται για λογισμικά ή δικτυωμένους λογαριασμούς που ενεργούν αυτόματα, χωρίς ανθρώπινη επίβλεψη για να παράγουν, αναπαράγουν ή ενισχύουν συγκεκριμένο περιεχόμενο. Τα botnets λειτουργούν ως εργαλεία τεχνητής ενίσχυσης (artificial amplification-AI). Δηλαδή, επαναλαμβάνουν μαζικά την ίδια πληροφορία, για παράδειγμα ένα ψευδές άρθρο, για να δώσουν την εντύπωση ότι έχει ευρεία υποστήριξη ή ενδιαφέρον. Κατά την διάρκεια των εκλογές στις ΗΠΑ το 2016 χιλιάδες bots χρησιμοποίησαν hashtags και ψεύτικες ειδήσεις για να επηρεάσουν τους ψηφοφόρους και να ενισχύσουν διχαστικά αφηγήματα [28, 116, 117, 119].

2.6.3. Ιστοσελίδες παραπληροφόρησης

Οι ιστοσελίδες παραπληροφόρησης αποτελούν ένα από τα πιο διαδεδομένα μέσα παραγωγής και διάδοσης ψευδών ειδήσεων στο διαδίκτυο. Πρόκειται για ιστοσελίδες που μιμούνται την “μορφή”

έγκυρων ειδησεογραφικών μέσων, ενώ στην πραγματικότητα δημοσιεύουν ψευδές, ανακριβές ή παραποιημένο περιεχόμενο. Οι ιστοσελίδες παραπληροφόρησης παρουσιάζονται συχνά ως ανεξάρτητα μέσα που το σύστημα δεν θέλει να ακουστούν. Επίσης χρησιμοποιούν clickbait τίτλους που προκαλούν περιέργεια, αναπαράγουν θεωρίες συνωμοσίας, ψευδή επιστημονικά άρθρα και πολιτικά περιεχόμενα.

Οι ιστοσελίδες αυτές παράγουν ψευδείς ειδήσεις και στη συνέχεια διαμοιράζονται σε πλατφόρμες κοινωνικής δικτύωσης μέσω links και μιμιδίων(memes). Οι ψευδείς ειδήσεις αναπαράγονται από δίκτυα bots ή ψεύτικους λογαριασμούς με στόχο την τεχνική ενίσχυση της απήχησής τους. Παράλληλα, προωθούνται μέσω αλγορίθμων, ιδίως όταν συνοδεύονται από ελκυστικούς τίτλους ή εικόνες [5, 117, 118].

2.6.4. Εφαρμογές μηνυμάτων

Οι εφαρμογές μηνυμάτων, όπως το WhatsApp, το Viber, το Messenger και το Telegram, έχουν εξελιχθεί σε μέσα διάδοσης ψευδών ειδήσεων. Παρότι σχεδιάστηκαν για ιδιωτική και γρήγορη επικοινωνία, έχουν μετατραπεί σε δίκτυα ανεπίσημης ενημέρωσης, όπου η παραπληροφόρηση μπορεί να διαδοθεί αθόρυβα και ανεξέλεγκτα. Οι εφαρμογές αυτές παρέχουν κρυπτογραφημένη επικοινωνία (end-to-end encryption), που καθιστά αδύνατο τον έλεγχο περιεχομένου από τις ίδιες τις πλατφόρμες, ομαδικές συνομιλίες στις οποίες το ίδιο μήνυμα μπορεί να φτάσει ταυτόχρονα σε εκατοντάδες ή χιλιάδες χρήστες, καθώς και γρήγορη προώθηση περιεχομένου (εικόνες, links, ηχητικά μηνύματα). Αυτό καθιστά τις εφαρμογές ιδανικό έδαφος για τη μαζική και ανεξέλεγκτη αναπαραγωγή ψευδών πληροφοριών, χωρίς δυνατότητα άμεσης διάψευσης.

Κατά τη διάρκεια της πανδημίας COVID-19, μέσω του ιδιωτικών μηνυμάτων στο WhatsApp, διαδόθηκαν ψευδείς ειδήσεις για επικίνδυνες «θεραπείες» ή θεωρίες συνωμοσίας για τα εμβόλια, με σοβαρές συνέπειες στη δημόσια υγεία [5, 117, 118].

2.7. Υπαίτιοι ψευδών ειδήσεων

2.7.1. Κακόβουλοι δράστες και κρατικά υποστηριζόμενες εκστρατείες παραπληροφόρησης

Οι εκστρατείες παραπληροφόρησης που οργανώνονται από κακόβουλους φορείς και χρηματοδοτούνται από το κράτος περιλαμβάνουν συντονισμένες προσπάθειες για τη διάδοση ψευδών ή παραπλανητικών πληροφοριών μέσω διαφόρων καναλιών. Οι εκστρατείες αυτές στοχεύουν σε συγκεκριμένα ακροατήρια, εκμεταλλεύονται τις κοινωνικές διαιρέσεις και ενισχύουν το διχαστικό περιεχόμενο για να υπονομεύσουν τους δημοκρατικούς θεσμούς και να επηρεάσουν την κοινή γνώμη.

Οι τακτικές που χρησιμοποιούνται περιλαμβάνουν τη δημιουργία ιστοσελίδων ή σημείων με ψεύτικες ειδήσεις που μιμούνται νόμιμες πηγές, τη χρήση bots και αυτοματοποιημένων εργαλείων για την τεχνητή ενίσχυση του περιεχομένου και τη στρατηγική διάδοση προσαρμοσμένων αφηγήσεων για την επίτευξη γεωπολιτικών στόχων. Οι υποστηριζόμενοι από κράτη φορείς

αξιοποιούν αυτές τις τακτικές για να διαστρεβλώσουν τον δημόσιο διάλογο, να σπείρουν τη σύγχυση και να διαβρώσουν την εμπιστοσύνη στους θεσμούς, θέτοντας σημαντικές προκλήσεις για τις δημοκρατικές κοινωνίες και την παγκόσμια σταθερότητα. Η αντιμετώπιση αυτής της απειλής απαιτεί διεθνή συνεργασία, πρωτοβουλίες για την παιδεία των μέσων ενημέρωσης και προσπάθειες για την προώθηση της διαφάνειας και της λογοδοσίας στα διαδικτυακά οικοσυστήματα πληροφοριών [11, 29].

2.7.2. Οντότητες με γνώμονα το κέρδος και clickbait farms

Οι οντότητες με γνώμονα το κέρδος και οι φάρμες δολωμάτων κλικ (clickbait farms) είναι σημαντικοί δράστες ψευδών ειδήσεων, εκμεταλλευόμενοι το τοπίο των ψηφιακών μέσων ενημέρωσης για οικονομικό κέρδος μέσω παραπλανητικών πρακτικών. Συγκεκριμένα, τα clickbait farms είναι οργανισμοί ή δίκτυα που παράγουν και διανέμουν μεγάλους όγκους χαμηλής ποιότητας, εντυπωσιακού ή παραπλανητικού περιεχομένου κυρίως για να δημιουργήσουν διαφημιστικά έσοδα μέσω υψηλής επισκεψιμότητας. Βασίζονται σε προκλητικούς τίτλους, οι οποίοι είναι συχνά φορτισμένοι συναισθηματικά, καθώς επίσης και υπερβολικούς ισχυρισμούς για να προσελκύσουν κλικ, συχνά εις βάρος ακριβών και πολύτιμων πληροφοριών. Το ίδιο το περιεχόμενο είναι συνήθως εντυπωσιακό, υπερβολικό ή εντελώς κατασκευασμένο για να προκαλέσει έντονες συναισθηματικές αντιδράσεις στους αναγνώστες. Μεγιστοποιώντας την εμπλοκή και τα ποσοστά κλικ, οι οντότητες με γνώμονα το κέρδος αυξάνουν τις εμφανίσεις διαφημίσεων και εκμεταλλεύονται την προσοχή που δημιουργούν οι ψευδείς ειδήσεις [17, 28].

Η καταπολέμηση των ψευδών ειδήσεων με γνώμονα το κέρδος απαιτεί ένα συνδυασμό αυστηρότερων πολιτικών διαφήμισης, ενισχυμένης εκπαίδευσης σε θέματα παιδείας στα μέσα ενημέρωσης, τεχνολογικών λύσεων για τον εντοπισμό και την επισήμανση παραπλανητικού περιεχομένου και προσπαθειών για την προώθηση της διαφάνειας στις διαδικτυακές διαφημιστικές πρακτικές. Με την αντιμετώπιση των κινήτρων και των τακτικών που χρησιμοποιούν οι κερδοσκοπικές οντότητες, οι ενδιαφερόμενοι φορείς μπορούν να μετριάσουν τον αντίκτυπο των ψευδών ειδήσεων στο δημόσιο διάλογο και να διατηρήσουν την ακεραιότητα των διαδικτυακών οικοσυστημάτων πληροφόρησης [17, 28].

2.7.3. Κομματικές και ιδεολογικές ομάδες

Η ακαδημαϊκή βιβλιογραφία συζητά εκτενώς πώς κομματικές ομάδες, ιδεολογικά κινήματα και εξτρεμιστικές οργανώσεις χρησιμοποιούν ψευδείς ειδήσεις για να προωθήσουν τις ατζέντες τους, να πολώσουν το δημόσιο διάλογο και να κινητοποιήσουν τους υποστηρικτές τους.

Οι κομματικές και ιδεολογικές ομάδες είναι σημαντικοί φορείς ψευδών ειδήσεων, αξιοποιώντας την παραπληροφόρηση για να προωθήσουν συγκεκριμένες ατζέντες, να ενισχύσουν ιδεολογικές πεποιθήσεις και να σπείρουν το διχασμό μεταξύ των κοινοτήτων. Οι ομάδες αυτές εκμεταλλεύονται το τοπίο των ψηφιακών μέσων ενημέρωσης για να διαδώσουν μεροληπτικές ή διαστρεβλωμένες πληροφορίες που ευθυγραμμίζονται με τις πολιτικές ή ιδεολογικές τους απόψεις.

Στοχεύοντας στο κοινό με εξατομικευμένες αφηγήσεις και εκμεταλλευόμενες τις κομματικές διαιρέσεις, ενισχύουν την παραπληροφόρηση για να επηρεάσουν την κοινή γνώμη και να προωθήσουν τα συμφέροντά τους. Οι κομματικές ψευδείς ειδήσεις συχνά αξιοποιούν συναισθηματικά ερεθίσματα, ενισχύοντας τις υπάρχουσες πεποιθήσεις και προκαταλήψεις για να δημιουργήσουν υποστήριξη ή αντίθεση σε συγκεκριμένες πολιτικές, υποψηφίους ή αιτίες. Ιδεολογικές ομάδες, συμπεριλαμβανομένων εξτρεμιστικών κινημάτων ή οργανώσεων υποστήριξης, χρησιμοποιούν επίσης ψεύτικες ειδήσεις για να ριζοσπαστικοποιήσουν το κοινό, να στρατολογήσουν υποστηρικτές και να νομιμοποιήσουν περιθωριακές ιδεολογίες. Μπορεί να παράγουν παραπλανητικό περιεχόμενο για να δαιμονοποιήσουν τους αντιπάλους, να τροφοδοτήσουν την κοινωνική αναταραχή ή να υπονομεύσουν την εμπιστοσύνη στους δημοκρατικούς θεσμούς.

Οι τακτικές που χρησιμοποιούν οι κομματικές και ιδεολογικές ομάδες περιλαμβάνουν επιλεκτική αναφορά, επιλογή γεγονότων και ενίσχυση θεωριών συνωμοσίας για τη δημιουργία θαλάμων απήχησης που ενισχύουν την κοσμοθεωρία τους. Ως αποτέλεσμα, ο πολλαπλασιασμός των κομματικών και ιδεολογικών ψευδών ειδήσεων συμβάλλει στην πόλωση και τη δυσπιστία στα μέσα ενημέρωσης, καθιστώντας δύσκολο για τα άτομα να διακρίνουν τις πραγματικές πληροφορίες από το μεροληπτικό ή παραπλανητικό περιεχόμενο [21, 30].

Η καταπολέμηση της χρήσης ψευδών ειδήσεων από κομματικές και ιδεολογικές ομάδες είναι δύσκολη, καθώς προέρχεται από επαγγελματίες του χώρου, και για το λόγο αυτό απαιτεί μια ολοκληρωμένη και πολύπλευρη προσέγγιση. Πρώτον, η ενίσχυση του γραμματισμού στα μέσα ενημέρωσης μέσω της εκπαίδευσης είναι ζωτικής σημασίας, επιτρέποντας στα άτομα να αξιολογούν κριτικά τις πηγές πληροφόρησης και να αναγνωρίζουν την παραπληροφόρηση. Οι πλατφόρμες μέσων κοινωνικής δικτύωσης και οι εταιρείες τεχνολογίας πρέπει να ενισχύσουν τις πολιτικές τους για τη συγκράτηση του περιεχομένου, χρησιμοποιώντας προηγμένους αλγόριθμους και ανθρώπινη εποπτεία για τον εντοπισμό και την αφαίρεση ψευδών ειδήσεων, διατηρώντας παράλληλα την ελευθερία της έκφρασης.

Η συνεργασία με ανεξάρτητους οργανισμούς ελέγχου των γεγονότων μπορεί να βοηθήσει στην επαλήθευση της αυθεντικότητας του viral περιεχομένου. Οι κυβερνήσεις και οι ρυθμιστικοί φορείς θα πρέπει να εφαρμόζουν και να επιβάλλουν αυστηρότερους νόμους κατά της διάδοσης ψευδών πληροφοριών και να καθιστούν τους δράστες υπεύθυνους. Επιπλέον, η προώθηση του ανοικτού και περιεκτικού δημόσιου διαλόγου μπορεί να μετριάσει την πόλωση, ενθαρρύνοντας τις διαφορετικές απόψεις και τον εποικοδομητικό διάλογο. Η οικοδόμηση ανθεκτικών κοινοτήτων που εκτιμούν τις πραγματικές πληροφορίες έναντι του εντυπωσιασμού είναι απαραίτητη για τη μείωση του αντίκτυπου των ψευδών ειδήσεων και τον περιορισμό της επιρροής των εξτρεμιστικών αφηγήσεων [33].

2.7.4. Φάρσες, τρολ και παραπλανημένα άτομα

Ο ρόλος των διαδικτυακών φαρσέρ, των τρολ και των παραπλανημένων ατόμων στην διάδοση ψευδών ειδήσεων μέσω των πλατφορμών κοινωνικής δικτύωσης είναι ιδιαιτέρως καθοριστικός,

συμβάλλοντας σε μεγάλο βαθμό στον συνολικό όγκο της παραπληροφόρησης που κυκλοφορεί στο διαδίκτυο.

Από την άλλη πλευρά, τα παραπλανημένα άτομα μπορεί να μοιράζονται άθελά τους ψευδείς ειδήσεις λόγω έλλειψης παιδείας στα μέσα ενημέρωσης ή δεξιοτήτων κριτικής σκέψης. Τα άτομα αυτά μπορεί να συναντήσουν παραπλανητικές πληροφορίες στο διαδίκτυο και να τις μοιραστούν με τα κοινωνικά τους δίκτυα χωρίς να επαληθεύσουν την ακρίβειά τους. Παράγοντες όπως οι γνωστικές προκαταλήψεις, συμπεριλαμβανομένης της προκατάληψης επιβεβαίωσης και του φαινομένου της ψευδαίσθησης της αλήθειας, μέσω των οποίων τα άτομα τείνουν να αποδέχονται και να διαδίδουν πληροφορίες που ευθυγραμμίζονται με τις προϋπάρχουσες πεποιθήσεις τους ή εκλαμβάνονται ως αξιόπιστες λόγω επανειλημμένης έκθεσης, ακόμη και όταν αυτές είναι ανακριβείς, συμβάλλοντας έτσι στην ακούσια διάδοση ψευδών ειδήσεων από καλοπροαίρετα άτομα [20, 27].

2.7.5. Αλγόριθμοι κοινωνικής δικτύωσης ως ενισχυτές ψευδών ειδήσεων

Οι αλγόριθμοι που υπάρχουν στα μέσα κοινωνικής δικτύωσης παίζουν σημαντικό ρολό στην ενίσχυση και στην ευρεία διάδοση των ψευδών ειδήσεων. Οι τεχνολογικοί ενισχυτές και ο χρηστές των μέσων κοινωνικής δικτύωσης δεν δημιουργούν οι ίδιοι το περιεχόμενο, όμως έχουν μεγάλο ρολό στο ποιο περιεχόμενο θα εμφανίζεται, σε ποιους χρήστες αναφέρεται και ποιο περιεχόμενο έχει μεγαλύτερή απήχηση στο κοινό. Ο τρόπος λειτουργίας τους βασίζεται σε κοινοποίησης της δημοσίευσης, σε αντιδράσεις, σε σχόλια ή ακόμα και στον χρόνο θέασης και έχουν ως στόχο την υποστήριξη ψευδών πληροφοριών, καθώς επίσης συμβάλλουν στη διάδοσή τους και στην παραμονή του χρήστη στην εκάστοτε πλατφόρμα [22, 115].

Η προσέγγιση αυτή φέρνει σαν αποτέλεσμα την προώθηση περιεχομένου, το οποίο προκαλεί έντονα συναισθήματα. Έτσι οι αλγόριθμοι ενισχύουν την διάδοση ψευδών ειδήσεων χωρίς αυτές οι ειδήσεις να έχουν ελεγχθεί για το πόσο έγκυρες είναι [111].

Οι χρήστες των μέσων κοινωνικής δικτύωσης, συμπεριλαμβανομένων των καθημερινών ατόμων, παίζουν επίσης ρόλο στη διαιώνιση των ψευδών ειδήσεων μέσω της διαδικτυακής τους συμπεριφοράς, καθώς επηρεάζουν τους αλγορίθμους που ενισχύουν τη διάδοση ψευδών ειδήσεων. Οι θάλαμοι απήχησης (echo chambers) και οι φυσαλίδες φίλτρων (filter bubbles) είναι αλγοριθμικά επιμελημένα περιβάλλοντα όπου οι χρήστες εκτίθενται κυρίως σε πληροφορίες και απόψεις που ενισχύουν τις υπάρχουσες πεποιθήσεις τους, περιορίζοντας έτσι την έκθεσή τους σε διαφορετικές προοπτικές και την ενίσχυση της παραπληροφόρησής σε κοινότητες ομοϊδεατών. Ο ορός φυσαλίδες φίλτρων (filter bubbles) μπορούμε να πούμε ότι αναφέρεται κυρίως ως μια νοητή σφαίρα πληροφόρησης στην οποία ο χρήστης είναι παγιδευμένος και δεν το έχει συνειδητοποίηση. Μεγάλες πλατφόρμες όπως (όπως Google, Facebook, Instagram, YouTube, TikTok κ.ά.) χρησιμοποιούν τέτοιους αλγορίθμους και δημιουργούν περιεχόμενο με βάση τις αναζητήσεις του εκάστοτε πολίτη, τα likes, τα shares ή τα σχόλιά του, αλλά και τη γεωγραφική τοποθεσία. Έτσι ο πολίτης βλέπει περιεχόμενα που συμφωνούν με τις απόψεις του και στην ουσία αποκλείει διαφορετικές γνώμες. Για παράδειγμα, αν ο πολίτης ψάχνεις πληροφορίες που υποστηρίζουν έναν συγκεκριμένο πολιτικό ή μια θεωρία, οι πλατφόρμες θα του δείχνουν παρόμοιο περιεχόμενο. Έτσι

δημιουργείται μια "φυσαλίδα" γύρω του, δηλαδή μια περιορισμένη οπτική γωνία όπου αυτός αρχίζει και βλέπει μόνο ειδήσεις και απόψεις που ευνοούν το φαινόμενο αυτό, χωρίς να εκτίθεται σε αντίθετα επιχειρήματα. Παρόλο που οι αλγόριθμοι κοινωνικής δικτύωσης δημιουργούνται με σκοπό την ευκολία και την εξατομικευμένη εμπειρία υπάρχουν σοβαροί κίνδυνοι που απορρέουν από αυτούς ως προς την κοινή γνώμη, αλλά και την κοινωνική συνοχή. Ο τρόπος αντιμετώπισής τους απαιτεί οι πολίτες να είναι ενεργοί και να γνωρίζουν πως να διαχειρίζονται τις πληροφορίες [22, 112, 113].

Εν κατακλείδι, οι αλγόριθμοι κοινωνικής δικτύωσης δεν αποτελούν απλώς μηχανισμούς ταξινόμησης περιεχομένου, αλλά λειτουργούν ως δομικοί επιταχυντές της παραπληροφόρησης [22, 114].

2.7.6. Ανεύθυνα μέσα ενημέρωσης

Τα ανεύθυνα μέσα ενημέρωσης συμβάλλουν σημαντικά στη διάδοση ψευδών ειδήσεων και παραπλανητικών πληροφοριών και σε ορισμένες περιπτώσεις, ενδέχεται να δημοσιεύουν ή να μεταδίδουν ακούσια ψευδείς πληροφορίες χωρίς τον κατάλληλο έλεγχο ή επαλήθευση των γεγονότων. Η έλλειψη ελέγχου των πηγών και το γεγονός ότι επιθυμούν να εντυπωσιάσουν το κοινό έναντι της ακρίβειας, συχνά οδηγούν στην εξαπάτησή του. Επιπλέον, η σκόπιμη παραπληροφόρηση από ορισμένα μέσα ενημέρωσης γίνεται για οικονομικούς λόγους και αυτό επιδεινώνει το πρόβλημα. Ωστόσο υπάρχουν περιπτώσεις όπου η διάδοση ψευδών ειδήσεων οφείλεται σε τεχνικά προβλήματα, όπως “χακάρισμα” και έτσι η διάδοσή τους συμβαίνει χωρίς να γίνει αισθητή από τους δημοσιογράφους και συντάκτες. Ακόμη και τότε όμως η ευθύνη για την αδυναμία πρόληψης παραμένει στα ίδια μέσα.

Όταν η παραπληροφόρηση διαδίδεται από αξιόπιστα μέσα ενημέρωσης, μπορεί να προσδώσει ψευδή αξιοπιστία σε ανακριβείς ισχυρισμούς και να συμβάλει στη δημόσια σύγχυση. Για να αντιμετωπιστεί αυτό το φαινόμενο απαιτούνται ενισχυμένα συντακτικά πρότυπα, αυστηροί μηχανισμοί ελέγχου, αλλά και η παρουσία ενός ανεξάρτητου ρυθμιστικού φορέα, ο οποίος θα έχει το καθήκον, αλλά και την αρμοδιότητά να επιβάλει κυρώσεις και να παρεμβαίνει οπότε ο ίδιος το θεωρεί πρέπον. Οι κυρώσεις αυτές μπορεί να είναι είτε κάποιο πρόστιμο, είτε προσωρινή αναστολή λειτουργίας. Η επιβολή τέτοιων κυρώσεων θα λειτουργεί αποτρεπτικά.

Πρωτοβουλίες για την παιδεία στα μέσα ενημέρωσης που αποσκοπούν στην εκπαίδευση του κοινού σχετικά με τη διάκριση αξιόπιστων πηγών πληροφόρησης μπορούν επίσης να ενδυναμώσουν τα άτομα να αξιολογούν κριτικά το περιεχόμενο των μέσων ενημέρωσης και να εντοπίζουν παραπλανητικές αφηγήσεις. Με την προώθηση υπεύθυνων δημοσιογραφικών πρακτικών και την καλλιέργεια μιας κουλτούρας διαφάνειας και λογοδοσίας στον κλάδο των μέσων ενημέρωσης, οι ενδιαφερόμενοι φορείς μπορούν να συμβάλουν στον μετριασμό των επιπτώσεων των ψευδών ειδήσεων και να υποστηρίξουν την ακεραιότητα της διάδοσης πληροφοριών στην ψηφιακή εποχή [25, 26].

2.8. Κίνητρα και σκοποί στη διάδοση ψευδών ειδήσεων

Στη δημιουργία ψευδών ειδήσεων συχνά εμπλέκονται πολλοί κακόβουλοι παράγοντες, δηλαδή άτομα, οργανισμοί ή οντότητες που εμπλέκονται σκόπιμα σε παραπλανητική ή επιβλαβή συμπεριφορά για να επιτύχουν τους στόχους τους. Οι παράγοντες αυτοί μπορεί να έχουν διάφορα κίνητρα, όπως πολιτικά, οικονομικά, ιδεολογικά ή προσωπικά συμφέροντα. Οι εν λόγω παράγοντες εκμεταλλεύονται τα τρωτά σημεία του οικοσυστήματος πληροφοριών για να διαδίδουν ψευδείς ή παραπλανητικές πληροφορίες για διάφορους σκοπούς.

Ο τρόπος δράσης των κακόβουλων παραγόντων περιλαμβάνει την κατασκευή ιστοριών χωρίς αναγκαστικά μια πραγματική βάση, τη χειραγώγηση εικόνων ή βίντεο, και τη διάδοση φημών με σκοπό να επηρεάσουν την κοινή γνώμη, να προκαλέσουν φόβο ή οργή ή να δημιουργήσουν κλικ, εγγεγραμμένους συνδρομητές και συμμετοχή. Μπορεί να θέτουν ως προτεραιότητα το προσωπικό όφελος ή να επιδιώκουν να προωθήσουν μια συγκεκριμένη ατζέντα, όπως η προώθηση μιας πολιτικής ιδεολογίας ή η υποστήριξη ενός συγκεκριμένου σκοπού. Στη συνέχεια, αναλύουμε κίνητρα και σκοπούς που συνεισφέρουν στη διασπορά ψευδών ειδήσεων μέσω του διαδικτύου.

2.8.1. Προώθηση πολιτικών γραμμών και ιδεολογιών

Αρχικά, ένας σημαντικός χώρος διάδοσης ψευδών ειδήσεων συνδέεται με την πολιτική. Στο κομμάτι της προώθησης πολιτικών συμφερόντων, οι προπαγανδιστές και οι κρατικοί φορείς διαδραματίζουν επίσης σημαντικό ρόλο στη διάδοση ψευδών ειδήσεων, καθώς κυβερνήσεις, πολιτικά κόμματα ή οργανώσεις που χρηματοδοτούνται από το κράτος μπορούν να συμμετάσχουν σε εκστρατείες παραπληροφόρησης για να χειραγωγήσουν την κοινή γνώμη, να προωθήσουν μια συγκεκριμένη ιδεολογία, να σπείρουν τη διχόνοια ή να υπονομεύσουν την εμπιστοσύνη στους δημοκρατικούς θεσμούς, αλλά και την εμπιστοσύνη σε πολιτικούς αντιπάλους. Μπορούν να χρησιμοποιούν εξελιγμένες τακτικές, συμπεριλαμβανομένης της προπαγάνδας, των εκστρατειών συκοφάντησης, της στοχευμένης αποστολής μηνυμάτων και της χειραγώγησης των μέσων κοινωνικής δικτύωσης, για να προωθήσουν τους γεωπολιτικούς τους στόχους ή να διατηρήσουν τον πολιτικό τους έλεγχο. Τέτοια φαινόμενα ενδέχεται να ενταθούν σε εποχές πολιτικών αναταραχών, περιόδους εκλογών ή όταν λαμβάνουν χώρα πολιτικά ντιμπέιτ (debate).

Ακόμη περισσότερους κινδύνους ενέχει η κατηγορία κακόβουλων παραγόντων που εμπλέκεται και προωθεί ακραίες ιδεολογικές πεποιθήσεις. Οι παράγοντες αυτοί χρησιμοποιούν συχνά ψευδείς ειδήσεις ως μέσο για να προωθήσουν τα σχέδιά τους, να χειραγωγήσουν την κοινή γνώμη και να σπείρουν τη διαίρεση στην κοινωνία με σκοπό να κάνουν τη δράση τους να φαίνεται ορθολογική και αιτιολογημένη σε βαθμό που να πείσουν ότι η χρήση ακραίων τακτικών είναι αναγκαία και αποτελεσματική. Για παράδειγμα, εξτρεμιστικές ομάδες μπορεί να διαδίδουν ψευδείς πληροφορίες για να δαιμονοποιήσουν ορισμένες εθνοτικές ή θρησκευτικές ομάδες, παρουσιάζοντάς τες ως απειλές για την εθνική ασφάλεια ή την κοινωνική σταθερότητα. Αυτή η τακτική του αποδιοπομπαίου τράγου και της τρομολαγνείας μπορεί να τροφοδοτήσει την προκατάληψη, τις διακρίσεις και τη βία κατά περιθωριοποιημένων κοινοτήτων. Επιπλέον, οι ψευδείς ειδήσεις χρησιμεύουν ως ισχυρό εργαλείο για τη ριζοσπαστικοποίηση και τον εξτρεμισμό,

εκμεταλλευόμενες ευάλωτα άτομα που είναι επιρρεπή σε εξτρεμιστικές ιδεολογίες. Ριζοσπαστικές ομάδες μπορεί να διαδίδουν αφηγήσεις ψευδών ειδήσεων που εκμεταλλεύονται τις ανασφάλειες, τα παράπονα ή τα αισθήματα στέρησης των δικαιωμάτων των ατόμων, προσφέροντας απλοϊκές εξηγήσεις και λύσεις σε πολύπλοκα κοινωνικά ή πολιτικά ζητήματα.

Για παράδειγμα, η εξτρεμιστική προπαγάνδα μπορεί να διαδίδει ψευδείς θεωρίες συνωμοσίας που ισχυρίζονται ότι μεταναστευτικές και προσφυγικές ομάδες κατηγορούνται για "εισβολή", αύξηση της εγκληματικότητας ή "αλλοίωση της εθνικής ταυτότητας. Ακόμη, οι ψεύτικες ειδήσεις μπορούν να συμβάλουν στη διαιώνιση θεωριών συνωμοσίας που ενισχύουν ιδεολογικές αφηγήσεις και υπονομεύουν την εμπιστοσύνη στα κυρίαρχα θεσμικά όργανα και τις αρχές.

Χαρακτηριστικό παράδειγμα, οι συνωμοσιολόγοι μπορεί να διαδίδουν αβάσιμους ισχυρισμούς για κυβερνητικές συγκαλύψεις, μυστικές συμμορίες ή παγκόσμιες συνωμοσίες, ισχυριζόμενοι ότι ισχυρές ελίτ συνωμοτούν για να ελέγξουν την κοινωνία ή να χειραγωγήσουν τα παγκόσμια γεγονότα. Αυτές οι θεωρίες συνωμοσίας μπορούν να διαβρώσουν την εμπιστοσύνη του κοινού στους δημοκρατικούς θεσμούς, την επιστημονική συναίνεση και την αντικειμενική δημοσιογραφία, καλλιεργώντας ένα κλίμα παράνοιας, κυνισμού και πόλωσης στην κοινωνία.

2.8.2. Οικονομικό κέρδος

Εκτός της προώθησης ιδεολογιών, ένα σημαντικό κίνητρο στη διάδοση ψευδών ειδήσεων είναι το οικονομικό όφελος. Για το λόγο αυτό, ορισμένοι κακόβουλοι παράγοντες δίνουν προτεραιότητα στο κέρδος έναντι της ακρίβειας (του περιεχομένου), δημιουργώντας τεχνητά εντυπωσιακό και δημοφιλές περιεχόμενο με σκοπό να προσελκύσουν κλικ και άλλες αντιδράσεις (likes, share κλπ) ώστε μέσω αυτών να δημιουργήσουν διαφημιστικά έσοδα. Μπορεί να χρησιμοποιούν τίτλους δολώματος κλικ (clickbait), παραπλανητικές μικρογραφίες ή προκλητικές εικόνες για να παρασύρουν ανυποψίαστους χρήστες να καταναλώσουν ή να μοιραστούν το περιεχόμενό τους. Δεδομένου ότι τα σχήματα χρηματικής ανταμοιβής που χρησιμοποιούν τα μέσα κοινωνικής δικτύωσης ενισχύουν την αλληλεπίδραση χρηστών και δημιουργών, όπως επίσης και το χρόνο παραμονής στην πλατφόρμα (για παράδειγμα με πλοήγηση στο περιεχόμενο ή διάβασμα/παρακολούθηση συγκεκριμένου περιεχομένου), οι δημιουργοί που αποκτούν ‘πιστό’ κοινό μέσω της αλληλεπίδρασης είναι αυτοί που καταφέρνουν και να επωφεληθούν οικονομικά, συνεπώς στοχεύουν αποκλειστικά στο να αυξήσουν τις μετρικές που τους αποφέρουν όφελος χωρίς να ενδιαφέρονται (εντελώς ή σχεδόν καθόλου) για το γενικότερο αντίκτυπο.

2.8.3. Προσωπική προβολή και κοινωνική επιρροή

Εκτός του οικονομικού κέρδους, πολλοί δημιουργοί επιθυμούν την προβολή στο διαδίκτυο και την απόκτηση κοινωνικής επιρροής. Με τη διάδοση ψευδών, αν και εντυπωσιακών, ειδήσεων καταφέρνουν να ενισχύσουν τη διαδικτυακή τους φήμη ή να αυξήσουν τους ακόλουθούς τους στα μέσα κοινωνικής δικτύωσης. Οι δημιουργοί αυτοί, παράγοντας περιεχόμενο που βρίσκει απήχηση σε συγκεκριμένα κοινά ή ενισχύει τις υπάρχουσες πεποιθήσεις και προκαταλήψεις, μπορούν να προσελκύσουν οπαδούς, να δημιουργήσουν κοινότητες ή να καθιερωθούν ως παράγοντες επιρροής σε συγκεκριμένες θέσεις.

2.8.4. Πρόκληση χάους και σύγχυσης

Ένα ακόμη κίνητρο στη διάδοση ψευδών ειδήσεων είναι απλώς για την ευχαρίστηση της εξαπάτησης ή για την πρόκληση χάους και σύγχυσης. Οι εμπλεκόμενοι κακόβουλοι παράγοντες μπορεί να επιδίδονται σε φάρσες, τρολάρισμα (trolling) ή διασπορά φημών ως μορφή ψυχαγωγίας, κοινωνικού πειραματισμού ή κακόβουλης δραστηριότητας με στόχο τη διατάραξη των κοινωνικών κανόνων ή θεσμών. Για παράδειγμα, τα λεγόμενα 'τρολ του διαδικτύου' (internet trolls) μπορεί να αντλούν ευχαρίστηση από την πρόκληση συναισθηματικών αντιδράσεων και διαμάχης με τη διάδοση ψευδών ειδήσεων ή την εμπλοκή σε 'εκρηκτική' συμπεριφορά στο διαδίκτυο. Αυτά τα τρολ μπορεί να δημιουργούν φανταστικές ειδήσεις για σκάνδαλα διασημοτήτων ή κατασκευασμένες θεωρίες συνωμοσίας για να προκαλέσουν έντονες αντιδράσεις από ανυποψίαστους χρήστες του διαδικτύου. Ομοίως, έχουν υπάρξει περιπτώσεις όπου έχουν επινοήσει περίπλοκες φάρσες ή ψεύτικες ειδήσεις σε πλατφόρμες κοινωνικής δικτύωσης για να εξαπατήσουν το κοινό και να προσελκύσουν την προσοχή του. Η θεωρία συνωμοσίας "Pizzagate" [1] είναι ένα πασίγνωστο παράδειγμα, η οποία ισχυρίζεται ψευδώς ότι μια πιτσαρία στην Ουάσιγκτον εμπλέκεται σε ένα κύκλωμα εμπορίας παιδιών για σεξουαλική εκμετάλλευση που συνδέεται με εξέχοντες πολιτικούς. Αυτός ο αβάσιμος ισχυρισμός οδήγησε σε παρενοχλήσεις και απειλές κατά των εργαζομένων και των πελατών του εστιατορίου.

Εκτός από τις φάρσες στα μέσα κοινωνικής δικτύωσης, ορισμένοι ιστότοποι δημιουργούνται αποκλειστικά με σκοπό τη διάδοση ψευδών ειδήσεων ή τη διάπραξη φάρσας ως μορφή ψυχαγωγίας. Αυτές οι ιστοσελίδες φάρσας συχνά δημοσιεύουν σατιρικό περιεχόμενο που είναι σκόπιμα παραπλανητικό ή χιουμοριστικό. Ιστοσελίδες όπως το "The Onion"[1] και το "ClickHole"[2] παράγουν σατιρικά άρθρα και τίτλους ειδήσεων με σκοπό την παρωδία πραγματικών ειδησεογραφικών γεγονότων και κοινωνικών τάσεων. Αν και η ύπαρξη τέτοιων ιστοσελίδων μπορεί να θεωρηθεί αστεία και καθαρά ψυχαγωγική, μπορεί να πείσει κάποιους χρήστες που δε διαθέτουν την απαραίτητη ικανότητα να κρίνουν και να διαχωρίσουν ένα κατά βάση σατιρικό περιεχόμενο (για παράδειγμα χρήστες χαμηλού μορφωτικού επιπέδου), έτσι ώστε να διαδώσουν την εν λόγω είδηση και τελικά αυτή να λάβει μεγάλες διαστάσεις από στόμα σε στόμα.

Επιπλέον, στην κατηγορία της διάδοσης ψευδών ειδήσεων με σκοπό τη δημιουργία χάους, έχουν εμφανιστεί ψεύτικες προκλήσεις στα μέσα κοινωνικής δικτύωσης ή viral τάσεις, οι οποίες ενθαρρύνουν τους συμμετέχοντες να εμπλακούν σε επικίνδυνες ή παράλογες συμπεριφορές που βασίζονται σε ψευδείς πληροφορίες. Για παράδειγμα, η φάρσα "Tide Pod Challenge"[3] κυκλοφόρησε τα μέσα κοινωνικής δικτύωσης, ισχυριζόμενη ψευδώς ότι η κατάποση κάδων απορρυπαντικού ρούχων ήταν μια δημοφιλής και ακίνδυνη τάση μεταξύ των εφήβων. Στην πραγματικότητα, η κατανάλωση των Tide Pods μπορεί να προκαλέσει σοβαρές επιπλοκές στην υγεία.

---

[1] https://www.theonion.com/

[2] https://clickhole.com/

[3] https://en.wikipedia.org/wiki/Consumption_of_Tide_Pods

2.8.5. Προκατάληψη της επιβεβαίωσης

Οι καταναλωτές ψευδών ειδήσεων συχνά συμβάλλουν άθελά τους στη διάδοσή τους, καθοδηγούμενοι από γνωστικές προκαταλήψεις που διαμορφώνουν τις συνήθειες κατανάλωσης πληροφοριών. Η προκατάληψη επιβεβαίωσης (confirmation bias) ειδικότερα, παίζει σημαντικό ρόλο, καθώς τα άτομα τείνουν να αναζητούν και να αποδέχονται πληροφορίες που επιβεβαιώνουν τις υπάρχουσες πεποιθήσεις ή απόψεις τους, ενώ απορρίπτουν ή αγνοούν τα αντιφατικά στοιχεία. Για παράδειγμα, ένα άτομο που έχει έντονες αντιεμβολιαστικές απόψεις μπορεί να είναι πιο πρόθυμο να πιστέψει και να μοιραστεί ψευδείς ειδήσεις που ισχυρίζονται ψευδώς ότι συγκεκριμένα εμβόλια είναι επιβλαβή, αγνοώντας παράλληλα τα επιστημονικά στοιχεία που αποδεικνύουν το αντίθετο. Ομοίως, τα πολιτικοποιημένα άτομα μπορεί να έλκονται από ψεύτικες ειδήσεις που διασύρουν αντίπαλα πολιτικά πρόσωπα ή επικυρώνουν τις ιδεολογικές τους θέσεις, ακόμη και αν οι πληροφορίες στερούνται αξιοπιστίας ή πραγματικής βάσης.

Επιπλέον, οι αλγόριθμοι των μέσων κοινωνικής δικτύωσης και οι θάλαμοι απήχησης (echo chambers) επιδεινώνουν τον αντίκτυπο της προκατάληψης επιβεβαίωσης με την επιμέλεια των ειδήσεων των χρηστών ώστε να ενισχύουν τις υπάρχουσες πεποιθήσεις και προτιμήσεις τους. Ως αποτέλεσμα, οι ψευδείς ειδήσεις που ευθυγραμμίζονται με τον ιδεολογικό θάλαμο απήχησης ενός ατόμου είναι πιο πιθανό να κυκλοφορήσουν στο κοινωνικό τους δίκτυο, ενισχύοντας περαιτέρω την εμβέλεια και την επιρροή τους. Για παράδειγμα, κατά τη διάρκεια πολιτικών εκστρατειών, τα άρθρα ψευδών ειδήσεων που εξυπηρετούν συγκεκριμένες κομματικές αφηγήσεις μπορεί να γίνουν viral εντός των σχετικών "κομματικών" θαλάμων απήχησης διότι εξαπλώνονται έτσι ώστε να διαδοθούν ταχέως μεταξύ ομοϊδεατών ατόμων που έχουν προδιάθεση να αποδέχονται και να μοιράζονται τέτοιο περιεχόμενο χωρίς να αξιολογούν κριτικά την ακρίβεια ή την αξιοπιστία του.

Η συναισθηματική απήχηση των ψευδών ειδήσεων και η προκατάληψη επιβεβαίωσης είναι δύο διαφορετικές έννοιες. Η συναισθηματική απήχηση με ένα μήνυμα δημιουργεί συναισθήματα όπως φόβο, θυμό, λύπη κ.α που έχει ως στόχο την επιρροή στο κοινό. Αντίθετα η προκατάληψη επιβεβαίωσης (*confirmation bias*) αποτελεί ένα "σφάλμα" σκέψης στο οποίο το άτομο αποδέχεται πληροφορίες που επιβεβαιώνουν τα πιστεύω του, απορρίπτοντας αντίθετα δεδομένα. Με λίγα λόγια, η συναισθηματική απήχηση δεν είναι είδος προκατάληψης επιβεβαίωσης αλλά αποτελεί μέσο που μπορεί να την ενδυναμώσει. Επιπλέον, η συναισθηματική απήχηση συμβάλει στην ταχεία εξάπλωσή τους (virality), καθώς οι εντυπωσιακές ή συναισθηματικά φορτισμένες ιστορίες τείνουν να προκαλούν έντονες αντιδράσεις και εμπλοκή από το κοινό. Για παράδειγμα, τα άρθρα ψευδών ειδήσεων που προκαλούν φόβο, οργή ή αγανάκτηση συχνά κοινοποιούνται ευρέως στις πλατφόρμες κοινωνικής δικτύωσης, ανεξάρτητα από την εγκυρότητά τους, απλώς και μόνο επειδή προκαλούν ισχυρές συναισθηματικές αντιδράσεις από τους χρήστες. Κατά συνέπεια, οι ψευδείς ειδήσεις που εκμεταλλεύονται συναισθηματικά ερεθίσματα ή ενισχύουν βαθιά ριζωμένες πεποιθήσεις μπορούν γρήγορα να κερδίσουν έδαφος και να διεισδύσουν στον διαδικτυακό διάλογο, διαιωνίζοντας την παραπληροφόρηση και υπονομεύοντας την εμπιστοσύνη σε αξιόπιστες πηγές πληροφοριών.

2.9. Έλεγχος και τρόποι περιορισμού διάδοσης ψευδών ειδήσεων

Οι προσπάθειες για τον έλεγχο της εξάπλωσης των ψευδών ειδήσεων είναι πολύπλευρες και περιλαμβάνουν διάφορους ενδιαφερόμενους φορείς, όπως κυβερνήσεις, εταιρείες τεχνολογίας, οργανισμούς μέσων ενημέρωσης και την κοινωνία των πολιτών.

2.9.1. Έλεγχος αληθείας αναφερόμενων γεγονότων

Στόχος των οργανισμών και των πρωτοβουλιών ελέγχου των γεγονότων είναι να επαληθεύουν την ακρίβεια των πληροφοριών που κυκλοφορούν στο διαδίκτυο και να διαψεύδουν ψευδείς ή παραπλανητικούς ισχυρισμούς. Οι υπεύθυνοι για τον έλεγχο των γεγονότων θα πρέπει να αναλύουν ειδήσεις, εικόνες, βίντεο και αναρτήσεις στα μέσα κοινωνικής δικτύωσης για να παρέχουν ακριβείς πληροφορίες στο κοινό και να αντιμετωπίζουν την παραπληροφόρηση [41, 42, 46, 102, 103].

2.9.2. Κριτική αξιολόγηση στα μέσα ενημέρωσης

Η προώθηση των δεξιοτήτων κριτικής σκέψης στα μέσα ενημέρωσης είναι απαραίτητη για την ενδυνάμωση των ατόμων ώστε αυτά να διακρίνουν τα γεγονότα από τη μυθοπλασία και να αξιολογούν κριτικά τις πληροφορίες που συναντούν στο διαδίκτυο. Καθώς οι πολίτες εκτίθενται καθημερινά σε έναν τεράστιο όγκο πληροφοριών, συχνά χωρίς επαρκή φιλτράρισμα ή αξιολόγηση, καθίσταται αναγκαία η ενίσχυση της ικανότητάς τους να αναγνωρίζουν την αξιοπιστία των πηγών και την ύπαρξη πιθανής παραπληροφόρησης. Γι' αυτόν τον λόγο πρέπει να διεξαχθούν εκπαιδευτικά προγράμματα, εργαστήρια και ενημερωτικές εκστρατείες, τα οποία θα στοχεύουν στην εκπαίδευση του κοινού σχετικά με τη σημασία της επαλήθευσης των πηγών, του ελέγχου των γεγονότων και της αναγνώρισης της προκατάληψης στο περιεχόμενο των μέσων ενημέρωσης. Οι ενημερωτικές εκστρατείες είναι στοχευμένες επικοινωνιακές και εκπαιδευτικές δράσεις που αποσκοπούν στην ενημέρωση του κοινού σε κρίσιμα κοινωνικά ζητήματα και ειδικότερα, λειτουργούν προληπτικά και διαφωτιστικά, συμβάλλοντας στην ευαισθητοποίηση του και στην καλλιέργεια μιας συνειδητής και υπεύθυνης στάσης απέναντι στην πληροφόρηση [13, 25, 88, 89].

2.9.3. Αξιοποίηση τεχνολογίας και αυτοματοποιημένων εργαλείων

Παράλληλα με τα εκπαιδευτικά προγράμματα και τις ενημερωτικές εκστρατείες, σημαντικό ρόλο μπορεί να διαδραματίσει και η τεχνολογική υποστήριξη της κριτικής σκέψης μέσω της ανάπτυξης αυτοματοποιημένων εργαλείων. Με τη χρήση σύγχρονων τεχνολογιών, όπως η τεχνητή νοημοσύνη (AI), γίνεται πιο εύκολη η επαλήθευση της αξιοπιστίας των πηγών, η ανίχνευση ψευδών ή παραποιημένων πληροφοριών στις ειδήσεις και ο εντοπισμός προκαταλήψεων στο περιεχόμενο. Τα εργαλεία αυτά δεν αποσκοπούν στο να αντικαταστήσουν την ανθρώπινη κρίση, αλλά στο να την υποστηρίξουν. Επομένως, είναι αρκετά χρήσιμα για τους πολίτες, προσφέροντας ενδείξεις και αναλύσεις που μπορούν να αξιοποιηθούν ως μία περαιτέρω αξιολόγηση της πληροφορίας [23, 28, 53, 55, 56, 57, 76, 77].

2.9.4. Αλγοριθμικές αλλαγές

Οι πλατφόρμες κοινωνικής δικτύωσης και οι μηχανές αναζήτησης εφαρμόζουν συνεχώς αλγοριθμικές αλλαγές για να μειώσουν την ορατότητα των ψευδών ειδήσεων και να προωθήσουν αξιόπιστες πηγές πληροφοριών. Αυτό περιλαμβάνει την προσαρμογή των αλγορίθμων ώστε να δίνουν προτεραιότητα σε περιεχόμενο από αξιόπιστες πηγές, την επισήμανση ή τον έλεγχο των αμφισβητούμενων πληροφοριών και τη μείωση της εξάπλωσης της παραπληροφόρησης μέσω συστημάτων συστάσεων (recommendation systems) [111, 112, 114, 115].

2.9.5. Συνεργασία και συμπράξεις

Η συνεργασία μεταξύ κυβερνήσεων, εταιρειών τεχνολογίας και οργανώσεων της κοινωνίας των πολιτών είναι ζωτικής σημασίας για την αποτελεσματική αντιμετώπιση της εξάπλωσης των ψευδών ειδήσεων. Οι εταιρικές σχέσεις μπορεί να περιλαμβάνουν την ανταλλαγή δεδομένων, πόρων και εμπειρογνωμοσύνης για την ανάπτυξη καινοτόμων λύσεων, τη βελτίωση των πρακτικών ελέγχου περιεχομένου (content control) και τον συντονισμό των αντιδράσεων σε εκστρατείες παραπληροφόρησης [5, 11, 30].

2.9.6. Ρυθμιστικά μέτρα

Ορισμένες κυβερνήσεις έχουν θεσπίσει νομοθεσία και κανονισμούς με στόχο την αντιμετώπιση των ψευδών ειδήσεων και της παραπληροφόρησης. Ωστόσο, η αποτελεσματική εφαρμογή των ρυθμιστικών αυτών παρεμβάσεων εξακολουθεί να αποτελεί πρόκληση, κυρίως λόγω του μεγάλου όγκου διαδικτυακού περιεχομένου και της ταχύτητας με την οποία αυτό διαδίδεται στις πλατφόρμες κοινωνικής δικτύωσης. Τα μέτρα που έχουν υιοθετηθεί περιλαμβάνουν, μεταξύ άλλων, την επιβολή κυρώσεων για τη διάδοση ψευδών πληροφοριών, την απόδοση αυξημένων ευθυνών στις πλατφόρμες κοινωνικής δικτύωσης όσον αφορά τη μετριοπάθεια του περιεχομένου, καθώς και την προώθηση μεγαλύτερης διαφάνειας στη διαδικτυακή διαφήμιση [4, 115, 117].

2.9.7. Εκστρατείες ευαισθητοποίησης του κοινού

Οι εκστρατείες ευαισθητοποίησης του κοινού προμηνύουν τους κινδύνους των ψευδών ειδήσεων, παρέχουν συμβουλές για τον εντοπισμό της παραπληροφόρησης και προωθούν την υπεύθυνη διαδικτυακή συμπεριφορά. Οι εκστρατείες αυτές μπορούν να χρησιμοποιούν διάφορα κανάλια, όπως τα μέσα κοινωνικής δικτύωσης, τα παραδοσιακά μέσα ενημέρωσης και τα εκπαιδευτικά ιδρύματα, για να προσεγγίσουν διαφορετικά ακροατήρια [26, 88, 89].

2.10. Συνέπειες, εμπόδια και αντίκτυπο των ψευδών ειδήσεων

Οι ψευδείς ειδήσεις μπορούν να έχουν εκτεταμένες και σημαντικές επιπτώσεις σε άτομα, οργανισμούς, κοινωνίες, κράτη ακόμη και σε παγκόσμιες υποθέσεις. Σε ατομικό επίπεδο, η έκθεση σε ψευδείς ειδήσεις μπορεί να οδηγήσει σε παραπληροφόρηση, όπου τα άτομα πιστεύουν ότι ψευδείς ή παραπλανητικές πληροφορίες είναι αληθινές. Αυτό μπορεί να επηρεάσει τις πεποιθήσεις, τις στάσεις και τις συμπεριφορές των ατόμων, οδηγώντας σε σύγχυση, πόλωση και

δυσπιστία απέναντι σε αξιόπιστες πηγές πληροφοριών. Η παραπληροφόρηση μπορεί επίσης να έχει συνέπειες στον πραγματικό κόσμο, όπως το να λαμβάνουν τα άτομα κακώς πληροφορημένες αποφάσεις, να διαδίδουν ψευδείς πληροφορίες σε άλλους ή ακόμη και να πέφτουν θύματα απάτης.

Σε κοινωνικό επίπεδο, η εξάπλωση ψευδών ειδήσεων μπορεί να υπονομεύσει τις δημοκρατικές διαδικασίες και τους θεσμούς διαστρεβλώνοντας τον δημόσιο διάλογο, διαμορφώνοντας τις πολιτικές απόψεις και επηρεάζοντας τα εκλογικά αποτελέσματα. Οι ψευδείς ειδήσεις μπορούν να χρησιμοποιηθούν ως εργαλείο πολιτικής χειραγώγησης, όπου άτομα ή ομάδες διαδίδουν ψευδείς πληροφορίες για να επηρεάσουν την κοινή γνώμη, να δυσφημίσουν τους αντιπάλους ή να υπονομεύσουν την εμπιστοσύνη στους δημοκρατικούς θεσμούς. Αυτό διαβρώνει τα θεμέλια της δημοκρατίας, καθώς οι πολίτες μπορεί να δυσκολεύονται να λάβουν τεκμηριωμένες αποφάσεις σε ένα κλίμα παραπληροφόρησης.

Στην ψηφιακή εποχή, οι ψευδείς ειδήσεις έχουν παγκόσμιες επιπτώσεις, καθώς η παραπληροφόρηση μπορεί να εξαπλωθεί γρήγορα πέρα από τα σύνορα και να φτάσει σε τεράστια ακροατήρια στο διαδίκτυο. Αυτό έχει οδηγήσει σε ανησυχίες σχετικά με τις αποσταθεροποιητικές επιπτώσεις των ψευδών ειδήσεων στις διεθνείς σχέσεις, την ασφάλεια στον κυβερνοχώρο και τη δημόσια ασφάλεια. Για παράδειγμα, οι ψευδείς πληροφορίες σχετικά με κρίσεις δημόσιας υγείας, όπως η πανδημία COVID-19, μπορούν να οδηγήσουν σε πανικό, παραπληροφόρηση και δυσπιστία προς τις υγειονομικές αρχές, εμποδίζοντας τις προσπάθειες για τον περιορισμό της εξάπλωσης του ιού και την προστασία της δημόσιας υγείας.

Συνολικά, ο αντίκτυπος των ψευδών ειδήσεων είναι πολύπλευρος και σύνθετος, επηρεάζοντας τα άτομα, τις κοινότητες και τις κοινωνίες με διάφορους τρόπους. Η αντιμετώπιση της εξάπλωσης των ψευδών ειδήσεων απαιτεί μια συντονισμένη προσπάθεια από τις κυβερνήσεις, τις εταιρείες τεχνολογίας, τους οργανισμούς μέσων ενημέρωσης και την κοινωνία των πολιτών για την προώθηση της παιδείας στα μέσα ενημέρωσης, την καταπολέμηση της παραπληροφόρησης και την προάσπιση των αρχών της αλήθειας, της ακρίβειας και της διαφάνειας στο δημόσιο διάλογο.

Η ευρεία διάδοση των ψευδών ειδήσεων συνοδεύεται από ποικίλες κοινωνικές συνέπειες, οι οποίες έχουν μελετηθεί ευρέως στην ακαδημαϊκή βιβλιογραφία.

2.10.1. Παραπληροφόρηση και εξαπάτηση

Οι ψευδείς ειδήσεις διαδίδουν εσφαλμένες ή παραπλανητικές πληροφορίες, οδηγώντας σε ευρεία παραπληροφόρηση του κοινού. Αυτή η παραπληροφόρηση μπορεί να διαστρεβλώσει τις δημόσιες αντιλήψεις, να διαμορφώσει πεποιθήσεις και να επηρεάσει τη λήψη αποφάσεων, υπονομεύοντας τελικά την ποιότητα του δημόσιου λόγου και της συμμετοχής των πολιτών. Η παραπληροφόρηση υπονομεύει την ικανότητα του κοινού να λαμβάνει τεκμηριωμένες αποφάσεις. Όταν τα άτομα εκτίθενται σε ψευδείς ειδήσεις ή παραπλανητικές αφηγήσεις, μπορεί να αναπτύξουν ανακριβείς αντιλήψεις της πραγματικότητας και να βασίσουν τις αποφάσεις τους σε λανθασμένες πληροφορίες. Αυτό μπορεί να είναι ιδιαίτερα προβληματικό κατά τη διάρκεια κρίσιμων γεγονότων, όπως εκλογές, καταστάσεις έκτακτης ανάγκης για τη δημόσια υγεία ή κοινωνικά κινήματα, όπου η ακριβής πληροφόρηση είναι ζωτικής σημασίας για τη λήψη τεκμηριωμένων

επιλογών. Η παραπληροφόρηση επίσης σχετίζεται με την αυξανόμενη δυσπιστία σε προηγουμένως αξιόπιστα μέσα πληροφόρησης, καθώς και συνολικότερα κοινωνικές παθογένειες, που επιφέρουν πόλωση και διχασμό της κοινωνίας [5].

Η παραπληροφόρηση συμβάλλει στη μείωση της εμπιστοσύνης και της αξιοπιστίας των πηγών πληροφόρησης συμπεριλαμβανομένων των κυρίαρχων μέσων ενημέρωσης, των επιστημονικών ιδρυμάτων και των κυβερνητικών υπηρεσιών. Καθώς η εμπιστοσύνη του κοινού διαβρώνεται, γίνεται πιο δύσκολο για τις έγκυρες πηγές να μεταδίδουν ακριβείς πληροφορίες και να αντιμετωπίζουν αποτελεσματικά τα πιεστικά κοινωνικά ζητήματα. Αυτή η διάβρωση της εμπιστοσύνης μπορεί να οδηγήσει σε σκεπτικισμό απέναντι στη νόμιμη εμπειρογνωμοσύνη και να συμβάλει στην ευρύτερη κοινωνική δυσπιστία. Επίσης, τροφοδοτεί τον κοινωνικό διχασμό και την πόλωση ενισχύοντας τις υπάρχουσες προκαταλήψεις και ενισχύοντας τους ιδεολογικούς θαλάμους απήχησης. Όταν τα άτομα εκτίθενται σε ψευδείς ή παραπλανητικές αφηγήσεις που ευθυγραμμίζονται με τις προ υπάρχουσες πεποιθήσεις τους ή τις πολιτικές τους πεποιθήσεις, αυτό μπορεί να βαθύνει τις κομματικές διαιρέσεις και να εμποδίσει τον εποικοδομητικό διάλογο μεταξύ διαφορετικών γνωμών. Αυτή η πόλωση μπορεί να αποδυναμώσει την κοινωνική συνοχή, να θέσει σε κίνδυνο τη δημοκρατική διακυβέρνηση και να επιδεινώσει τις κοινωνικές εντάσεις [5].

2.10.2. Διάβρωση της εμπιστοσύνης στα μέσα ενημέρωσης και τους θεσμούς

Η συνέπεια της διάβρωσης της εμπιστοσύνης στα μέσα ενημέρωσης και τους θεσμούς λόγω της εξάπλωσης των ψευδών ειδήσεων είναι ένα κρίσιμο ζήτημα που υπονομεύει τη δημοκρατική διακυβέρνηση και τον δημόσιο διάλογο. Όταν οι ψευδείς ειδήσεις πολλαπλασιάζονται και η παραπληροφόρηση εξαπλώνεται ανεξέλεγκτα, αυτό συμβάλλει στη μείωση της εμπιστοσύνης και της αξιοπιστίας του κοινού προς τα κυρίαρχα μέσα ενημέρωσης, τις έγκυρες πηγές πληροφόρησης και τους καθιερωμένους θεσμούς. Μία από τις πρωταρχικές επιπτώσεις αυτής της διάβρωσης της εμπιστοσύνης είναι η μειωμένη αποτελεσματικότητα των οργανισμών και των θεσμών των μέσων ενημέρωσης στην ενημέρωση του κοινού και στην προώθηση της ενημερωμένης συμμετοχής των πολιτών. Καθώς η εμπιστοσύνη διαβρώνεται, τα άτομα μπορεί να γίνουν πιο επιφυλακτικά απέναντι στην ειδησεογραφική κάλυψη, αμφισβητώντας την ακρίβεια και την αμεροληψία της πληροφόρησης.

Αυτός ο σκεπτικισμός μπορεί να οδηγήσει σε απομάκρυνση από τις κύριες πηγές των μέσων ενημέρωσης και σε αυξανόμενη εξάρτηση από εναλλακτικές ή κομματικές πηγές ειδήσεων που εξυπηρετούν συγκεκριμένες ιδεολογικές απόψεις. [31]

Επιπλέον, η διάβρωση της εμπιστοσύνης στα μέσα ενημέρωσης και τους θεσμούς υπονομεύει τα θεμέλια της δημοκρατικής διακυβέρνησης. Η εμπιστοσύνη σε θεσμούς, όπως οι κυβερνητικές υπηρεσίες, οι επιστημονικοί οργανισμοί και οι αρχές δημόσιας υγείας είναι απαραίτητη για την αποτελεσματική χάραξη πολιτικής και τη δημόσια διοίκηση. Όταν η εμπιστοσύνη διακυβεύεται, γίνεται πιο δύσκολο για τους θεσμούς να συγκεντρώσουν την υποστήριξη του κοινού για πολιτικές, πρωτοβουλίες ή προσπάθειες αντιμετώπισης κρίσεων, επηρεάζοντας τελικά την κοινωνική συνοχή και τις συλλογικές διαδικασίες λήψης αποφάσεων.

Επιπλέον, η διάβρωση της εμπιστοσύνης στα μέσα ενημέρωσης και τους θεσμούς μπορεί να συμβάλει σε ευρύτερες κοινωνικές διαιρέσεις και πόλωση. Καθώς τα άτομα στρέφονται προς τις πηγές πληροφόρησης που ενισχύουν τις υπάρχουσες πεποιθήσεις ή προκαταλήψεις τους, βαθαίνουν οι ιδεολογικοί θάλαμοι απήχησης και παρεμποδίζεται ο εποικοδομητικός διάλογος μεταξύ διαφορετικών απόψεων. Αυτή η πόλωση υπονομεύει την κοινωνική συνοχή και θέτει σε κίνδυνο την ικανότητα των κοινωνιών να αντιμετωπίζουν συλλογικά πολύπλοκες προκλήσεις. [31]

2.10.3. Κοινωνική πόλωση και διαίρεση

Οι συνέπειες της κοινωνικής πόλωσης και του διχασμού που επιδεινώνονται από τη διάδοση των ψευδών ειδήσεων είναι βαθιές, επηρεάζοντας τον ιστό των κοινωνιών και τον δημοκρατικό διάλογο. Οι ψευδείς ειδήσεις συχνά εκμεταλλεύονται τα κοινωνικά ρήγματα ενισχύοντας διχαστικές αφηγήσεις και ενισχύοντας ιδεολογικούς θαλάμους απήχησης, οι οποίοι με τη σειρά τους συμβάλλουν στην αυξημένη πόλωση μεταξύ ατόμων και κοινοτήτων. Μια σημαντική επίπτωση αυτής της πόλωσης είναι η εμβάθυνση των ιδεολογικών διαχωρισμών και η δημιουργία απομονωμένων θαλάμων απήχησης εντός των διαδικτυακών και μη διαδικτυακών κοινοτήτων. Όταν τα άτομα εκτίθενται επανειλημμένα σε παραπληροφόρηση που ευθυγραμμίζεται με τις υπάρχουσες πεποιθήσεις ή προκαταλήψεις τους, αυτό ενισχύει την προκατάληψη επιβεβαίωσης και περιορίζει την έκθεσή τους σε διαφορετικές προοπτικές. Το φαινόμενο αυτό οδηγεί στο σχηματισμό απομονωμένων ομάδων με παγιωμένες απόψεις, εμποδίζοντας τον ουσιαστικό διάλογο και την αμοιβαία κατανόηση πέρα από ιδεολογικές γραμμές. Ακόμη, η κοινωνική πόλωση που τροφοδοτείται από ψευδείς ειδήσεις μπορεί να δημιουργήσει δυσπιστία, εχθρότητα και μισαλλοδοξία απέναντι σε άτομα ή ομάδες που θεωρούνται διαφορετικές ή αντίθετες. Οι διχαστικές αφηγήσεις που προπαγανδίζουν οι ψευδείς ειδήσεις μπορούν να υποδαυλίσουν τη δυσαρέσκεια και να επιδεινώσουν τις κοινωνικές εντάσεις, διαβρώνοντας την κοινωνική συνοχή και υπονομεύοντας την εμπιστοσύνη μεταξύ ατόμων και κοινοτήτων. [20]

Στο πεδίο της δημοκρατικής διακυβέρνησης, η αυξημένη κοινωνική πόλωση θέτει προκλήσεις για τη δημιουργία συναίνεσης, τον συμβιβασμό και την αποτελεσματική χάραξη πολιτικής. Όταν ο πολιτικός διάλογος πολώνεται, μπορεί να οδηγήσει σε αδιέξοδο, κομματικές συγκρούσεις και κατάρρευση των δημοκρατικών κανόνων. Αυτή η πόλωση υπονομεύει την ικανότητα των δημοκρατικών θεσμών να αντιμετωπίζουν σύνθετα κοινωνικά ζητήματα και κινδυνεύει να υπονομεύσει την πίστη του κοινού στις δημοκρατικές διαδικασίες. [20]

2.10.4. Κίνδυνοι για τη δημόσια υγεία

Οι συνέπειες των κινδύνων για τη δημόσια υγεία που απορρέουν από τη διάδοση ψευδών ειδήσεων είναι βαθιές και εκτεταμένες, ιδίως σε περιόδους κρίσεων υγείας, όπως η πανδημία COVID-19. Οι ψευδείς ειδήσεις και η παραπληροφόρηση γύρω από τις θεραπείες υγείας, τα προληπτικά μέτρα και την ασφάλεια των εμβολίων μπορούν να θέσουν σε σημαντικό κίνδυνο τα αποτελέσματα της δημόσιας υγείας και να υπονομεύσουν τις προσπάθειες για τον μετριασμό της εξάπλωσης των μολυσματικών ασθενειών.

Ένας σημαντικός αντίκτυπος των ψευδών ειδήσεων στη δημόσια υγεία είναι η διάβρωση της εμπιστοσύνης στην επιστημονική εμπειρογνωμοσύνη και στις τεκμηριωμένες παρεμβάσεις. Όταν τα άτομα εκτίθενται σε ψευδείς ισχυρισμούς ή θεωρίες συνωμοσίας για θέματα που σχετίζονται με την υγεία, αυτό μπορεί να οδηγήσει σε σκεπτικισμό απέναντι σε καθιερωμένα μέτρα δημόσιας υγείας, με αποτέλεσμα τη μειωμένη συμμόρφωση με προγράμματα εμβολιασμού ή προληπτικές οδηγίες. Αυτός ο δισταγμός και η απροθυμία υιοθέτησης των συνιστώμενων πρακτικών υγείας μπορεί να εμποδίσει τις προσπάθειες ελέγχου των ασθενειών και να αυξήσει τον κίνδυνο εμφάνισης επιδημιών εντός των κοινοτήτων.

Επιπλέον, η διάδοση ψευδών ειδήσεων σχετικά με τη δημόσια υγεία μπορεί να επιτείνει το φόβο, τη σύγχυση και την ανησυχία του κοινού. Οι παραπλανητικές πληροφορίες σχετικά με τη σοβαρότητα, τη μετάδοση ή τη θεραπεία ασθενειών μπορούν να προκαλέσουν περιττό πανικό και να ενισχύσουν τη συναισθηματική δυσφορία, καθιστώντας πιο δύσκολη την επικοινωνία ακριβών πληροφοριών για την υγεία και τη διαχείριση των αντιλήψεων του κοινού κατά τη διάρκεια καταστάσεων έκτακτης ανάγκης στον τομέα της υγείας. [27]

2.10.5. Πολιτική χειραγώγηση

Οι ψεύτικες ειδήσεις χρησιμοποιούνται συχνά για πολιτικούς σκοπούς, συμπεριλαμβανομένης της χειραγώγησης των εκλογών, της δολοφονίας χαρακτήρων και της υπονόμευσης των δημοκρατικών διαδικασιών. Οι υποστηριζόμενες από το κράτος εκστρατείες παραπληροφόρησης και οι κομματικοί φορείς χρησιμοποιούν τις ψευδείς ειδήσεις για να σπείρουν τη δυσπιστία στους δημοκρατικούς θεσμούς και να επηρεάσουν τα εκλογικά αποτελέσματα. Συγκεκριμένα, όταν τα άτομα εκτίθενται σε ψευδείς αφηγήσεις που αποσκοπούν στην απαξίωση των πολιτικών αντιπάλων ή στην απονομιμοποίηση των εκλογικών διαδικασιών, αυτό μπορεί να ενισχύσει τον κυνισμό και την απάθεια απέναντι στη δημοκρατική διακυβέρνηση. Αυτή η διάβρωση της εμπιστοσύνης υπονομεύει τη νομιμότητα των δημοκρατικών θεσμών και θέτει προκλήσεις για τη λειτουργία της αντιπροσωπευτικής δημοκρατίας. Η πολιτική χειραγώγηση που τροφοδοτείται από ψευδείς ειδήσεις μπορεί να επιδεινώσει τις κοινωνικές διαιρέσεις και να πολώσει τις κοινότητες κατά μήκος κομματικών γραμμών. Όταν η παραπληροφόρηση διαδίδεται στρατηγικά για να εκμεταλλευτεί τις υπάρχουσες ιδεολογικές διαιρέσεις, μπορεί να εμβαθύνει τις κοινωνικές εντάσεις, να τροφοδοτήσει τις κομματικές συγκρούσεις και να εμποδίσει τον εποικοδομητικό διάλογο πέρα από τις πολιτικές διαιρέσεις. Αυτή η πόλωση υπονομεύει την κοινωνική συνοχή και θέτει σε κίνδυνο την ικανότητα των κοινωνιών να αντιμετωπίζουν συλλογικά τις συλλογικές προκλήσεις [30].

3. Μεθοδολογία βιβλιογραφικής ανασκόπησης

Στην παρούσα εργασία χρησιμοποιήσαμε την μεθοδολογία της συστηματικής βιβλιογραφικής ανασκόπησης(Systematic Literature Review – SLR), με στόχο τη μεθοδική μελέτη της εργασίας που σχετίζεται με την ανίχνευση και την πρόληψη των ψευδών ειδήσεων. Η επιλογή αυτής της μεθοδολογίας μας επιτρέπει την οργανωμένη συλλογή, αξιολόγηση και σύνθεση ερευνητικών αποτελεσμάτων.

Η διαδικασία της ανασκόπησης έγινε με βάση τις κατευθυντήριες οδηγίες Kitchenham [121], που συνιστούν ένα ευρέως αποδεκτό πλαίσιο για τη συστηματική ανασκόπηση της βιβλιογραφίας στον χώρο της Πληροφορικής. Οι οδηγίες αυτές παρέχουν ένα σύνολο βημάτων, το οποίο διασφαλίζει τη διαφάνεια της διαδικασίας και τη δυνατότητα αναπαραγωγής της μεθοδολογίας από μελλοντικούς ερευνητές.

Το σύνολο των βημάτων χωρίζεται σε τέσσερα στάδια. Αρχικά καθορίζεται ο ερευνητικός στόχος της εργασίας και διατυπώνονται βασικά ερευνητικά ερωτήματα, τα οποία αποτελούν την βάση για την πορεία της ανασκόπησης. Στην συνέχεια ακολουθεί η διαδικασία αναζήτησης της βιβλιογραφίας, με τον προσδιορισμό των κατάλληλων επιστημονικών πηγών και των όρων αναζήτησης ώστε να εντοπιστούν μελέτες σχετικές με το θέμα μας. Έπειτα, εφαρμόζεται η διαδικασία φιλτραρίσματος των εντοπισμένων άρθρων, βάσει συγκεκριμένων κριτηρίων ένταξης, αποκλεισμού και ποιότητας με σκοπό τη διαμόρφωση ενός αξιόπιστου και αντιπροσωπευτικού συνόλου μελετών. Τέλος, πραγματοποιήθηκε μια αρχική ποσοτική ανάλυση των επιλεγμένων άρθρων, η οποία επιτρέπει τη συνοπτική αποτύπωση βασικών χαρακτηριστικών της βιβλιογραφίας και των κυρίαρχων ερευνητικών τάσεων. Τα επιμέρους αυτά στάδια παρουσιάζονται και αναλύονται διεξοδικά στις επόμενες ενότητες του κεφαλαίου.

3.1. Ερευνητικός στόχος και ερωτήματα

Στην συγκεκριμένη ενότητα περιγράφονται οι στόχοι της παρούσας βιβλιογραφικής επισκόπησης που είναι η ταξινόμηση, ανάλυση και σύγκριση της παρούσας βιβλιογραφίας γύρω από το πεδίο της ανίχνευσης και πρόληψης των ψευδών ειδήσεων, καθώς πρόκειται για ένα σημαντικό σύγχρονο πρόβλημα, ενισχυμένο από την ευρεία διάδοση του διαδικτύου και των μέσων κοινωνικής δικτύωσης και με πολλαπλές συνέπειες, όπως διαφάνηκε από την ανάλυση στα προηγούμενα κεφάλαια. Επιχειρούμε με βάση την παρούσα επισκόπηση να αναλύσουμε τους τρόπους ανίχνευσης και πρόληψης των ψευδών ειδήσεων, αναδεικνύοντας κατ’ αυτόν τον τρόπο τα ερευνητικά κενά που παρουσιάζονται, ώστε να βοηθήσουμε με στοχευμένο τρόπο τη μελλοντική έρευνα εκ μέρους της επιστημονικής κοινότητας.

Με βάση τους στόχους αυτούς, επιχειρούμε να απαντήσουμε στα εξής ερευνητικά ερωτήματα:

1. Πώς επιτυγχάνεται η ανίχνευση και η πρόληψη των ψευδών ειδήσεων;
2. Πώς συμμετέχουν οι σύγχρονες τάσεις τεχνητής νοημοσύνης σε αυτά τα εγχειρήματα;

3. Ποια είναι τα σημαντικότερα ερευνητικά κενά που σχετίζονται με μια πιο ολοκληρωμένη, ακριβή, αποδοτική και αποτελεσματική ανίχνευση και πρόληψη ψευδών ειδήσεων και εξακολουθούν να υπάρχουν;
4. Πως μπορούμε να κατηγοριοποιήσουμε τις μεθόδους ανίχνευσης και πρόληψης ψευδών ειδήσεων;
5. Ποιες μέθοδοι θεωρούνται οι καλύτερες με βάση ποια κριτήρια;
6. Σε ποια είδη άρθρων και από ποιους εκδότες έχουν δημοσιευθεί οι μέθοδοι αυτές;
7. Ποιες είναι οι ερευνητικές κατευθύνσεις που θα πρέπει να ακολουθηθούν στο άμεσο μέλλον;

3.2. Διαδικασία αναζήτησης

Στο τρέχων κεφάλαιο περιγράφεται η διαδικασία αναζήτησης της βιβλιογραφίας. Η διαδικασία αυτή περιλαμβάνει την επιλογή των βιβλιογραφικών βάσεων, τον σχεδιασμό του μοτίβου επερώτησης και τις τεχνικές που χρησιμοποιήθηκαν για την πληρέστερη δυνατή κάλυψη του επιστημονικού πεδίου.

Αρχικά ανατρέξαμε στους ιστότοπους Semantic Scholar[4] και Google Scholar[5], οι οποίοι αρχειοθετούν αυτόματα τα δημοσιευμένα άρθρα συνήθως λίγες μόλις μέρες αφότου αυτά δημοσιευθούν είτε στα πρακτικά συνεδρίων, είτε στον τόμο κάποιου περιοδικού, είτε στο αποθετήριο ArXiV, όπου οι συγγραφείς μπορούν να μεταφορτώσουν το άρθρο τους, ακόμα και προτού ολοκληρωθεί η φάση της έγκριτης αξιολόγησής του. Επιλέξαμε τις δύο αυτές πλατφόρμες ως τους κύριους ιστότοπους αναζήτησης επειδή παρουσιάζουν συμπληρωματικά χαρακτηριστικά. Το Google Scholar καλύπτει ένα πιο ευρύ φάσμα πεδίων, όπως ανθρωπιστικά, νομικά και κοινωνικές επιστήμες και έχει ευρεία κάλυψη, μάλιστα εντάσσει και γκρίζα βιβλιογραφία, ενώ το Semantic Scholar καλύπτει κυρίως επιστημονικά και τεχνολογικά πεδία (STEM) και παρέχει πιο στοχευμένα εργαλεία ανάλυσης και φιλτραρίσματος των επιστημονικών άρθρων. Συνεπώς, θεωρούμε πως η επιλογή των δύο αυτών βιβλιοθηκών μας δίνει την καλύτερη δυνατή κάλυψη της βιβλιογραφίας.

Εμπειρικά, το Google Scholar και το Semantic Scholar έχουν μικρές διαφορές στις αναφορές/παραπομπές ανά άρθρο, συνεπώς αναζητώντας ανεξάρτητα τους ίδιους όρους σε κάθε έναν από τους δύο ιστότοπους μπορούμε με μεγάλη βεβαιότητα να καλύψουμε τη βιβλιογραφία που επιθυμούμε. Επιλέξαμε να χρησιμοποιήσουμε το Semantic Scholar, διότι μας δίνει την δυνατότητα να χρησιμοποιήσουμε εργαλεία υψηλής τεχνολογίας, χρήση AI τεχνολογίας για την ανάλυση επιστημονικών άρθρων, παρέχει το TLDR (too long didn't read) το οποίο αποτελεί μια (σημασιολογική) σύνοψη του άρθρου σε μια γραμμή και διαθέτει ετεροαναφορές (citations). Τέλος, έχουμε την δυνατότητα να αναζητήσουμε επιστημονικά άρθρα με βάση την χρονολογία, το είδος, το μέγεθος, καθώς και την γλώσσα. Έτσι λοιπόν ο συνδυασμός χρήσης αυτών των δυο ιστοτόπων θεωρούμε ότι είναι ο ιδανικός για την ολοκλήρωση της εργασίας.

[4] https://www.semanticscholar.org/

[5] https://scholar.google.com/

Στην συγκεκριμένη βιβλιογραφική διπλωματική εργασία χρησιμοποιήθηκε ένα μοτίβο επερώτησης που υλοποιήθηκε με διαφορετικό τρόπο ανάλογα με τα χαρακτηριστικά και ιδιαιτερότητες αναζήτησης της κάθε βιβλιογραφικής βάσης. Το μοτίβο αυτό χωρίστηκε σε τρεις κατηγορίες:

1. Το αντικείμενο που αναζητάμε που περιλαμβάνει τους όρους: fake news, misinformation, disinformation, malinformation, counter-information, counterinformation, fabricated/manipulated content, propaganda, clickbait.
2. Τις τεχνολογικές τάσεις ή τεχνολογικούς τομείς που περιλαμβάνει τους όρους: artificial intelligence, machine learning, natural language processing, knowledge graph, knowledge base.
3. Τους όρους που σχετίζονται, το είδος των προσεγγίσεων προς ανάλυση ή την ενέργεια στην οποία εξειδικεύονται όπως detection (ανίχνευση), recognition (αναγνώριση), prevention (πρόληψη) και countermeasures (αντίμετρα) καθώς και συνώνυμούς τους, όπως identification (εξακρίβωση), classification (ταξινόμηση) και, mitigation (μετριασμός

Επιπλέον, οι προαναφερόμενοι ιστότοποι παρέχουν βιβλιογραφικές αναφορές (references) (δημοσιεύσεις πάνω στις οποίες έχει αναφερθεί ένα προς μελέτη άρθρο) αλλά και παραπομπές (citations) (μεταγενέστερες δημοσιεύσεις οι οποίες αναφέρονται στο προς μελέτη άρθρο), επιτρέποντας να καλυφθεί η βιβλιογραφία του πεδίου ακολουθώντας αλυσιδωτά τη μια αναφορά μετά την άλλη, έως ότου είτε δεν υπάρχουν μεταγενέστερες αναφορές (άρα συνήθως έχουμε φτάσει σε ένα πολύ πρόσφατο άρθρο), είτε οι δημοσιεύσεις που ανακτούμε ξεφεύγουν από τη θεματολογία μελέτης μας. Αυτό αναφέρεται και ως μέθοδος snowball [122] στις μεθοδολογίες βιβλιογραφικής ανασκόπησης και η οποία εφαρμόστηκε συμπληρωματικά στην περίπτωσή μας σε σχέση με την αναζήτηση στους προαναφερόμενους ιστοτόπους.

3.3. Φιλτράρισμα άρθρων

Για να εξασφαλίσουμε την εγκυρότητα της επισκόπησης αναζητούμε επιστημονικά άρθρα σε δημοφιλείς ιστότοπους/βιβλιοθήκες ακαδημαϊκών άρθρων (είτε αυτά έχουν λάβει αξιολογήσεις από κριτές κι έχουν δημοσιευθεί σε κάποιο επιστημονικό συνέδριο ή περιοδικό, είτε βρίσκονται ακόμη σε στάδιο αξιολόγησης, όπως συνήθως συμβαίνει με πολύ πρόσφατα άρθρα).

Με βάση τις λέξεις κλειδιά που αναφέρθηκαν στο προηγούμενο κεφάλαιο, γίνεται ανάκτηση των σχετικών δημοσιεύσεων. Κάθε άρθρο σε πρώτο στάδιο εξετάστηκε ως προς τα βασικά κριτήρια ένταξης και αποκλεισμού, τα οποία βασίζονται κυρίως στα μεταδεδομένα του, όπως τον τίτλο, την περίληψη (abstract) και το TLDR (too long didn’t read), την χρονολογία, το είδος, το μέγεθος, καθώς και την γλώσσα. Σκοπός του σταδίου αυτού είναι ο περιορισμός των άρθρων σε αυτά που συνάδουν με το αντικείμενο της εργασίας. Στο δεύτερο στάδιο τα άρθρα που έχουν συλλεχθεί υποβάλλονται σε ποιοτική αξιολόγηση, η οποία βασίζεται στο περιεχόμενό τους. Η αξιολόγηση αυτή αφορά αφενός τη σχετικότητά τους με το παρόν θέμα και τις αντίστοιχες ερευνητικές ερωτήσεις που τέθηκαν, αφετέρου τη χρήση τεκμηριωμένης επιστημονικής μεθοδολογίας. Με την

ολοκλήρωση και των δύο σταδίων δεν διασφαλίζεται μόνο ότι τα επιλεγμένα άρθρα συνάδουν με το θέμα μας, αλλά και ότι είναι και επιστημονικά αξιόπιστα.

Εφόσον ένα άρθρο ικανοποιεί το κριτήριο της σχετικότητας, εξετάζουμε αν έχει δημοσιευτεί σε κάποιο επιστημονικό περιοδικό ή συνέδριο και πόσες ετεροαναφορές (citations) διαθέτει, λαμβάνοντας υπόψιν ότι παλαιότερα άρθρα ενδέχεται να περιέχουν περισσότερα citations, ενώ πιο καινούργια ίσως να μη διαθέτουν και καθόλου.

Για άρθρα που είναι διαθέσιμα σε ακαδημαϊκές πλατφόρμες χωρίς να έχουν δημοσιευθεί (προδημοσιεύσεις(preprints)), εφαρμόζεται αυστηρότερο κριτήριο αξιολόγησης. Συγκεκριμένα, προδημοσιεύσεις που παραμένουν αδημοσίευτες για χρονικό διάστημα μεγαλύτερο των δύο ετών απορρίπτονται εφόσον διαθέτουν μικρό αριθμό ετεροαναφορών, ο οποίος στην παρούσα εργασία ορίζεται σε λιγότερες από πέντε ετεροαναφορές. Αντίστοιχα, για άρθρα που έχουν ήδη δημοσιευθεί σε περιοδικά ή συνέδρια και έχουν παρέλθει περισσότερα από τρία έτη από τη δημοσίευσή τους, εξετάζονται ως προς την επιστημονική τους απήχηση με βάση τον αριθμό των ετεροαναφορών. Δημοσιευμένες εργασίες με λιγότερες από πέντε ετεροαναφορές αποκλείονται από το τελικό σύνολο.

Επιπλέον, εξετάζουμε άρθρα που είναι δημοσιευμένα σε workshop, είναι επιστημονικά αναγνωρισμένα και εφαρμόζεται σε αυτά η διαδικασία αξιολόγησης (peer review). Αυτά λοιπόν διατηρούνται, αλλά ίσως με χαμηλότερη βαρύτητα από ένα πλήρες άρθρο σε περιοδικό υψηλής απήχησης. Πέραν των άρθρων αυτών, συμπεριλήφθηκε περιορισμένος αριθμός τεχνικών αναφορών (technical reports), εκθέσεων (reports), καθώς κρίθηκαν απαραίτητα για την ανάλυση της χειροκίνητης ανίχνευσης ψευδών ειδήσεων. Οι διπλωματικές εργασίες προπτυχιακού και μεταπτυχιακού επίπεδου και τα βιβλία που δεν περνούν από διαδικασία αξιολόγησης, αποκλείονται, καθώς δεν μας παρέχουν τις ίδιες/κατάλληλες επιστημονικές εγγυήσεις. Τέλος, μόνο οι διδακτορικές διατριβές, οι οποίες περνάνε από αυστηρό έλεγχο κατά την διαδικασία αξιολόγησης, από επίσημες εξεταστικές επιτροπές είναι οι μόνες που διατηρούνται. Σκοπός της συγκεκριμένης διαδικασίας είναι η διασφάλιση ότι το οι πήγες που διατηρήθηκαν βασίζονται σε εργασίες με αποδεδειγμένη επιστημονική εγκυρότητα και αξιοπιστία.

Σημειώνεται ότι το impact factor είναι ένα υποκειμενικό κριτήριο ποιότητας. Αν το άρθρο είναι σχετικό με το θέμα μας, έχει κάποιες ετεροαναφορές και έχει δημοσιευτεί με αξιολόγηση (peer-reviewed) τότε συνήθως διατηρείται, ακόμα και αν το περιοδικό έχει χαμηλό impact factor. Ενδέχεται ένα τέτοιο άρθρο να βαθμολογηθεί χαμηλότερα σε ποιότητα από εμάς, αλλά αυτό δεν σημαίνει άμεση απόρριψή του.

Εστιάζουμε στην αγγλική γλώσσα, καθώς μας παρέχει πληθώρα ποιοτικών δημοσιεύσεων ως κύρια επιστημονική γλώσσα στα διεθνή περιοδικά και συνέδρια. Στην πράξη, σχεδόν όλα τα άρθρα που ανακτώνται στην παρούσα θεματική στους σχετικούς ιστότοπους είναι στα αγγλικά. Ωστόσο, πολλά άρθρα ενδέχεται να αφορούν μελέτες σε άλλες γλώσσες. Δεδομένου ότι η πολυγλωσσικότητα (multilinguality) περιέχει ιδιάζοντα στοιχεία, τα οποία σχετίζονται με τους πόρους που διαθέτει κάθε γλώσσα για τη δημιουργία συνόλων δεδομένων και την εκπαίδευση μοντέλων, επιλέγουμε να μη συμπεριλάβουμε ενεργά σχετικά άρθρα σε γλώσσα συγγραφής πέρας της αγγλικής και της ελληνικής στην ανάλυσή μας.

Ιδιαίτερη, έμφαση δίνουμε στο έτος δημοσίευσης ενός άρθρου, διότι άρθρα που είναι δημοσιευμένα αρκετά χρόνιά πριν μπορεί να έχουν ξεπερασμένες πληροφορίες, άρα μην βασίζονται σε σύγχρονα δεδομένα. Θέσαμε σαν κατώτερο όριο το έτος 2010, ώστε να διασφαλιστεί η κάλυψη των σύγχρονων εξελίξεων στο πεδίο της ανίχνευσης και πρόληψης ψευδών ειδήσεων, οι οποίες συνδέονται άμεσα με τη ραγδαία ανάπτυξη των μεθόδων μηχανικής μάθησης και τεχνητής νοημοσύνης. Ενώ το χρονικό όριο για τις τεχνικές ανίχνευσης τέθηκε στο 2010, επιλέχθηκε η κατ' εξαίρεση συμπερίληψη 2 θεμελιωδών άρθρων (seminal papers) της περιόδου 2000-2009 στο τελικό δείγμα. Τα άρθρα αυτά, παρότι παλαιότερα, ορίζουν βασικές αρχές της διάδοσης πληροφορίας που παραμένουν επίκαιρες και απαραίτητες για την κατανόηση των σύγχρονων μοντέλων. Ωστόσο, δεδομένου ότι δεν προτείνουν συγκεκριμένες τεχνικές ανίχνευσης ή πρόληψης, δεν συμπεριλαμβάνονται στη συγκριτική ανάλυση, αλλά αξιοποιούνται αποκλειστικά για την κάλυψη του θεωρητικού υποβάθρου.

Επιπρόσθετα, η έκταση ενός άρθρου λαμβάνεται υπόψη ως βοηθητικό κριτήριο αξιολόγησης. Η αυξημένη έκταση δεν εγγυάται κατ’ ανάγκη την πληρότητα ή την καλή δομή ενός άρθρου, καθώς ακόμη και εκτενείς εργασίες ενδέχεται να είναι δυσνόητες ή ανεπαρκώς οργανωμένες. Ωστόσο, πολύ σύντομα άρθρα συχνά δεν παρέχουν επαρκή περιγραφή της μεθοδολογίας, των δεδομένων και των αποτελεσμάτων, γεγονός που καθιστά δυσχερή την ουσιαστική αξιολόγησή τους. Στην παρούσα εργασία έχουμε εφαρμόσει κατώτατο όριο έκτασης, με διατήρηση μόνο άρθρων που εκτείνονται σε τουλάχιστον τέσσερις με πέντε δίστηλες σελίδες. Το συγκεκριμένο όριο επιλέχθηκε ώστε να αποκλειστούν σύντομες δημοσιεύσεις, όπως extended abstracts ή συνοπτικές παρουσιάσεις. Με τον τρόπο αυτό εξασφαλίζουμε ότι τα άρθρα που περιλαμβάνονται στη βιβλιογραφική ανασκόπηση διαθέτουν επαρκές περιεχόμενο για την αξιολόγηση της συνεισφοράς τους, χωρίς η έκταση να λειτουργεί ως υποκατάστατο της επιστημονικής ποιότητας.

Ακόμη, άρθρα που έχουν υψηλό βαθμό ομοιότητας ως προς την προτεινόμενη τεχνική, εξετάζονται συστηματικά για το ποιο έχει πιο ολοκληρωμένη εκδοχή. Συνήθως αυτό παρατηρείται όταν οι ίδιοι συγγραφείς εκδίδουν workshop ή conference εκδόσεις μιας συγκεκριμένης ιδέας. Η επιλογή άρθρου που διατηρείται βασίζεται σε συνδυασμό κριτηρίων, όπως η πληρότητα της μεθοδολογικής περιγραφής, η έκταση και σαφήνεια της πειραματικής αξιολόγησης, το κύρος και η απήχηση του χώρου δημοσίευσης (περιοδικό ή συνέδριο), καθώς και ο αριθμός των ετεροαναφορών. Στις περιπτώσεις όπου μια προτεινόμενη τεχνική αποτελεί σαφές υποσύνολο μιας άλλης, προτιμάται η πιο ολοκληρωμένη προσέγγιση. Αντίθετα, παρόμοιες τεχνικές που προτείνονται ανεξάρτητα από διαφορετικούς ερευνητές δεν υπόκεινται σε διαδικασία αποκλεισμού, καθώς αποτελούν ξεχωριστές επιστημονικές συνεισφορές και συμβάλλουν στη συνολική χαρτογράφηση του ερευνητικού πεδίου. Με τον τρόπο αυτό διασφαλίζεται ότι η βιβλιογραφική ανασκόπηση παραμένει αντιπροσωπευτική, χωρίς να αποκλείονται εργασίες υψηλής ποιότητας λόγω επιφανειακής ομοιότητας.

Τέλος, ιδιαίτερη μεταχείριση δόθηκε στα άρθρα επισκόπησης, τα οποία δεν εντάχθηκαν στο βασικό σύνολο άρθρων που αξιοποιήθηκαν για την ανάλυση της παρούσας εργασίας. Τα άρθρα αυτά δεν προτείνουν συνήθως νέες τεχνικές ανίχνευσης ή αντιμετώπισης ψευδών ειδήσεων, αλλά

συγκεντρώνουν και οργανώνουν υπάρχουσες μελέτες, λειτουργώντας ως δευτερογενείς πηγές γνώσης. Στο πλαίσιο αυτό, χρησιμοποιήθηκαν επικουρικά, κυρίως για την κατανόηση του ερευνητικού πεδίου, την άντληση εννοιολογικών πλαισίων, διαστάσεων κατηγοριοποίησης και γενικών τάσεων, καθώς και για την επιβεβαίωση της συνοχής των ευρημάτων της παρούσας ανάλυσης.

Ωστόσο, η κύρια βιβλιογραφική ανάλυση και η συγκριτική αξιολόγηση που ακολουθεί βασίστηκαν αποκλειστικά σε πρωτότυπες ερευνητικές μελέτες, οι οποίες παρουσιάζουν νέες προσεγγίσεις, μεθοδολογίες και πειραματικά αποτελέσματα. Με τον διαχωρισμό αυτό διασφαλίζεται ότι η παρούσα εργασία στηρίζεται σε πρωτογενές ερευνητικό υλικό, χωρίς να παραβλέπεται η συμβολή των άρθρων επισκόπησης ως υποστηρικτικών και οργανωτικών πηγών.

3.4. Ποσοτική Ανάλυση

Με βάση την επισκόπηση που παρουσιάστηκε στις προηγούμενες ενότητες, το σύνολο των πηγών που επιλέχθηκαν χρησιμοποιήθηκε για την ποσοτική ανάλυση. Η επιλογή των πηγών βασίστηκε στην διαδικασία φιλτραρίσματος που ακολουθήσαμε, η οποία εφαρμόστηκε στο σύνολο των peer-reviewed επιστημονικών πηγών που εντοπίσαμε στο πρώτο στάδιο της αναζήτησης. Συνολικά συλλέχθηκαν 75 δημοσιεύσεις.

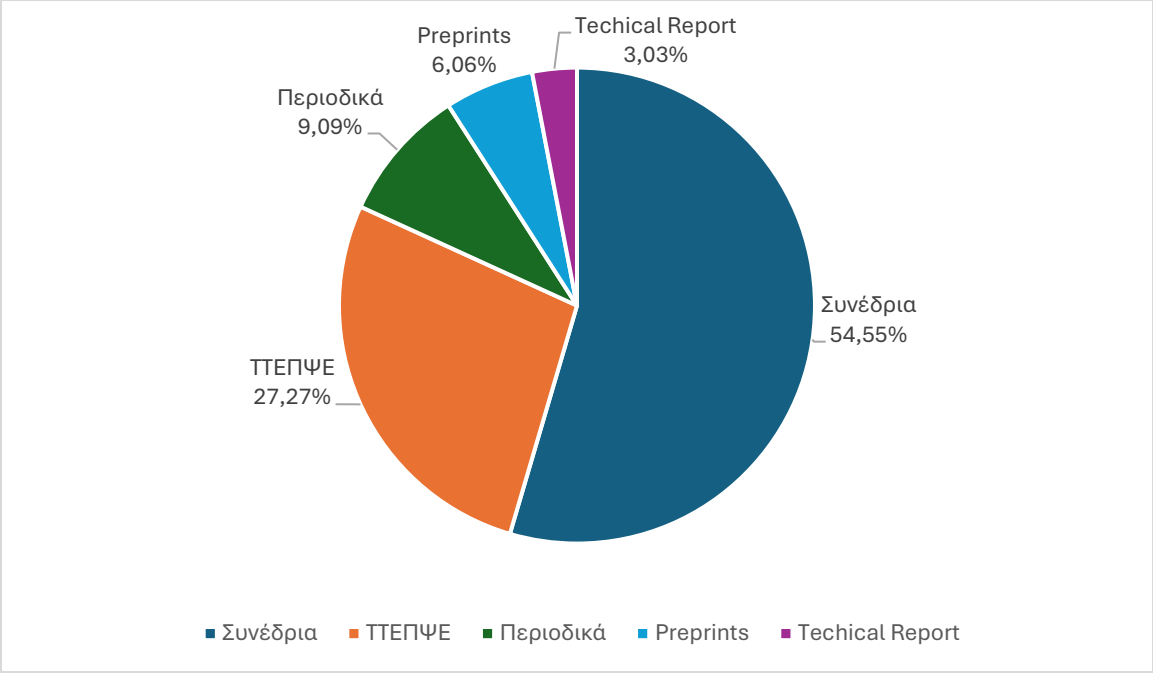
Αρχικά, εξετάστηκαν τα άρθρα ως προς το αντικείμενο της ανίχνευσης και της πρόληψης ψευδών ειδήσεων, τον τύπο της δημοσίευσης και τη γλώσσα συγγραφής. Στη συνέχεια, ακολούθησε ο ποιοτικός έλεγχος, κατά τον οποίο αξιολογήθηκαν η σαφήνεια της προτεινόμενης μεθοδολογίας. Παράλληλα, λάβαμε υπόψιν εργασίες που εμφανίστηκαν αρχικά σε workshops και στη συνέχεια σε εκτενέστερη μορφή σε συνέδρια ή περιοδικά, με την επικάλυψη να αποτελεί έναν από τους παράγοντες απόρριψης. Ο κυριότερος παράγοντας απόρριψης ήταν η επικάλυψη άρθρων μεταξύ workshops, συνεδρίων ή περιοδικών ή η επικάλυψη λόγω ταυτόχρονων συνεισφορών. Με βάση την προαναφερόμενη διαδικασία ενισχύθηκε η συνοχή της μελέτης και διασφαλίστηκε πως τα συμπεράσματά μας προκύπτουν από αξιόπιστες πηγές.

Μετά το πέρας αυτής της διαδικασίας διατηρήθηκαν 33 επιστημονικές πηγές με ποσοστό διατήρησης 44% και τα υπόλοιπα 42 απορρίφθηκαν σε κάποιο από τα βήματα διαδικασίας που αναφέρθηκε παραπάνω με ποσοστό απόρριψης 56%. Από αυτά τα 33, οι 26 επιστημονικές πηγές (74,29%) αποτελούν αμιγώς επιστημονικά άρθρα σχετικά με την αυτοματοποιημένη ανίχνευση, ενώ τα υπόλοιπα 9 (25,71%) αφορούν τεκμηριωμένη γκρίζα βιβλιογραφία, η οποία περιλαμβάνει τεχνικές αναφορές και επίσημες εκθέσεις (reports) ψηφιακών εργαλείων, τα οποία ανιχνεύουν ψευδείς ειδήσεις από ανθρώπους. Οι τελευταίες 9 πηγές αντλήθηκαν από το Google Scholar.

Η επιλογή των τελευταίων αυτών πηγών κρίθηκε απαραίτητη αποκλειστικά και μόνο για την ανάλυση της χειροκίνητης παροχής γνώσης (Ενότητα 4.2.1.1), προκειμένου να αποτυπωθεί η πρακτική εφαρμογή του ελέγχου των ειδήσεων στο σύγχρονο ψηφιακό περιβάλλον.

Αναφορικά με τις 33 επιστημονικές πηγές ανίχνευσης, 18 πηγές (54,55%) προέρχονται από συνέδρια/workshops, 9 πηγές (27,27%) αφορούν τεχνικές εκθέσεις/τεκμηρίωση ψηφιακών

εργαλείων(ΤΤΕΠΨΕ), 1 άρθρο αποτελεί technical report (3,03%), 2 από αυτά αποτελούν preprints (6,06%) (στο arXiv) και 3 πήγες έχουν δημοσιευθεί σε περιοδικά (9,09%).



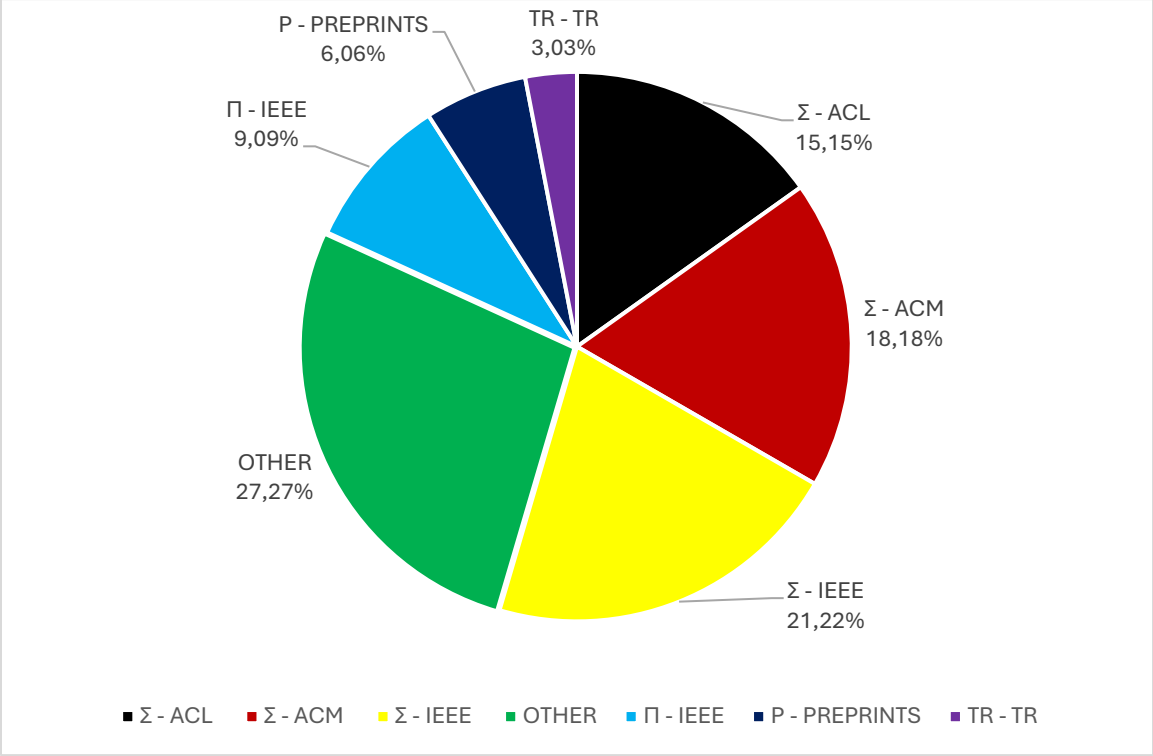
Σχήμα 1. Κατανομή πηγών ως προς το είδος τους (n=33).

Αναφορικά με τα άρθρα σε συνέδρια, η κατανομή των οργανισμών που εκδίδουν τα πρακτικά των συνεδρίων στα οποία δημοσιεύθηκαν είναι:

- Association of Computational Linguistics: 5 άρθρα (15,15%)
- ACM: 6 άρθρα (18,18%)
- Institute of Electrical and Electronics Engineers (IEEE): 7 άρθρα (21,22%)
- Other (CEUR-WS, ICCL, Neural Information Processing Systems Foundation(NIPSF), Association for the Advancement of Artificial Intelligence (AAAI)): 9 άρθρα (27,27%)

Αντίστοιχα, για τα 3 περιοδικά, η κατανομή είναι η ακόλουθη:

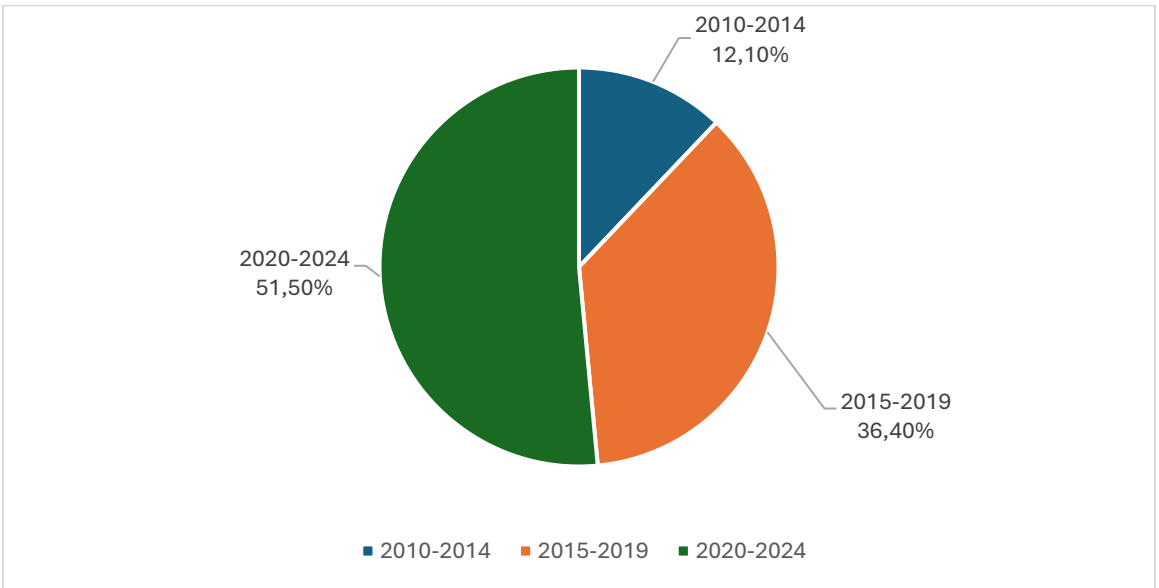
- Institute of Electrical and Electronics Engineers (IEEE): 3 άρθρα (9,09%)



Σχήμα 2. Εστιασμένη κατανομή πηγών (n=33).

Στο συγκεκριμένο γράφημα για δική μας διευκόλυνση έχουμε ορίσει όπου Π ότι είναι περιοδικά, όπου Σ είναι συνέδρια, όπου P είναι preprints και όπου TR είναι technical reports.

Αναφορικά με τις 33 πηγές ανίχνευσης, 17 (51,5%) δημοσιεύθηκαν την χρονολογία 2020 έως 2024, 12 από αυτές (36,4%) το 2015 έως 2019, 4 (12,1%) την τετραετία 2010 έως 2014.



Σχήμα 3. Κατανομή πηγών ως προς το έτος (n=33).

3.5. Απαντήσεις στα ερευνητικά ερωτήματα

Στη συγκεκριμένη ενότητα επίσης παραθέτουμε τα ερευνητικά ερωτήματα που αναφέρθηκαν στην Ενότητα 3.1 καθώς και τα σημεία/κεφάλαια στην αναφορά όπου απαντώνται τα ερωτήματα αυτά. Αρχικά αποτυπώνεται το ερώτημα και έπειτα η απάντηση ή οι απαντήσεις ως προς αυτό.

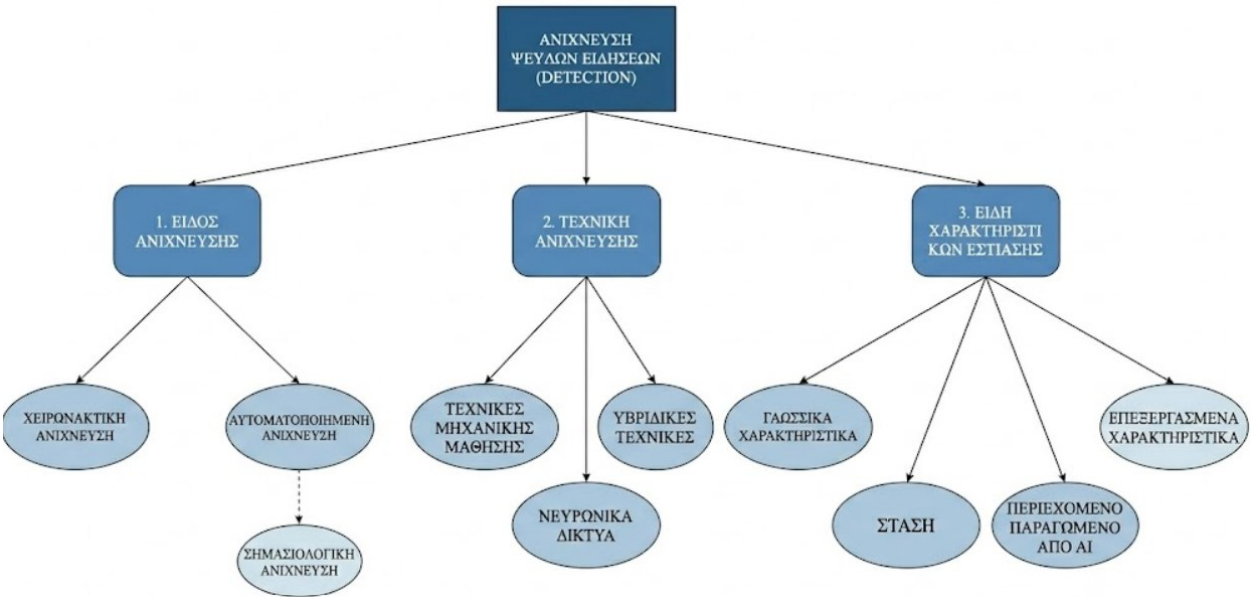
- **Ερώτημα νούμερο 1:**
  Πώς επιτυγχάνεται η ανίχνευση και η πρόληψη των ψευδών ειδήσεων;
  Απάντηση:
  Στην Ενότητα 2.9
  Στην Ενότητα 4.2
  Στην Ενότητα 4.4
- **Ερώτημα νούμερο 2:**
  Πώς συμμετέχουν οι σύγχρονες τάσεις τεχνητής νοημοσύνης σε αυτά τα εγχειρήματα;
  Απάντηση:
  Στην Ενότητα 4.2.1.2
  Στην Ενότητα 4.2.2.4
  Στην Ενότητα 4.4.4
- **Ερώτημα νούμερο 3:**
  Ποια είναι τα σημαντικότερα ερευνητικά κενά που σχετίζονται με μια πιο ολοκληρωμένη, ακριβή, αποδοτική και αποτελεσματική ανίχνευση και πρόληψη ψευδών ειδήσεων και εξακολουθούν να υπάρχουν;
  Απάντηση:
  Στο Κεφάλαιο 5
- **Ερώτημα νούμερο 4:**
  Πως μπορούμε να κατηγοριοποιήσουμε τις μεθόδους ανίχνευσης και πρόληψης ψευδών ειδήσεων;
  Απάντηση:
  Στην Ενότητα 4.1
  Στην Ενότητα 4.4.1
- **Ερώτημα νούμερο 5:**
  Ποιες μέθοδοι θεωρούνται οι καλύτερες με βάση ποια κριτήρια;
  Απάντηση:
  Στην Ενότητα 4.3
  Στην Ενότητα 4.5
- **Ερώτημα νούμερο 6:**
  Σε ποια είδη άρθρων και από ποιος εκδότες έχουν δημοσιευθεί οι μέθοδοι αυτές;
  Απάντηση:
  Στην Ενότητα 3.4
- **Ερώτημα νούμερο 7:**
  Ποιες είναι οι ερευνητικές κατευθύνσεις που θα πρέπει να ακολουθηθούν στο άμεσο μέλλον;
  Απάντηση:
  Στο Κεφάλαιο 5.

4. Ανάλυση

Σε αυτό το κεφάλαιο παρέχεται η ανάλυση των προσεγγίσεων που προτείνονται στα άρθρα που επιλέχθηκαν και μελετήθηκαν με βάση συγκεκριμένες κατηγορίες που έχουμε διαμορφώσει σύμφωνα με την έρευνά μας. Επίσης, πραγματοποιείται μια σύγκριση των προσεγγίσεων αυτών με βάση συγκεκριμένα κριτήρια. Η δομή του κεφαλαίου αυτού έχει ως εξής. Η Ενότητα 4.1 παρέχει την προτεινόμενη κατηγοριοποίηση για τις προσεγγίσεις ανίχνευσης ψευδών. Η Ενότητα 4.2 αναλύει τις προσεγγίσεις ανίχνευσης ψευδών ειδήσεων ενώ η Ενότητα 4.3 συγκρίνει τις προσεγγίσεις αυτές με βάση συγκεκριμένα κριτήρια. Η Ενότητα 4.4 κατηγοριοποιεί και αναλύει τις προσεγγίσεις πρόληψης ψευδών ειδήσεων ενώ η Ενότητα 4.5 τις συγκρίνει με βάση συγκεκριμένα κριτήρια. Τέλος, στην Ενότητα 4.6 παρουσιάζουμε συνοπτικά τα συμπεράσματα της ανάλυσης και την συγκριτική αποτίμηση των βασικών κατηγόριων προσεγγίσεων που έχουμε ως προς το επίπεδο ωριμότητας και απόδοσης.

4.1.Κατηγοριοποίηση προσεγγίσεων ανίχνευσης ψευδών ειδήσεων

Στην παρούσα ενότητα παρουσιάζεται το ιεραρχικό δέντρο ως προς την κατηγοριοποίηση των προσεγγίσεων ανίχνευσης ψευδών ειδήσεων και μια σύντονη περιγραφή για την κάθε κατηγορία του διαγράμματος. Η κατηγοριοποίηση αυτή είναι βασισμένη σε μια πολυδιάστατη λογική, η οποία λαμβάνει υπόψη το είδος της ανίχνευσης, τις τεχνικές που αξιοποιούνται καθώς και τα χαρακτηριστικά στα οποία εστιάζουν.



Εικόνα 1. Ιεραρχικό δέντρο κατηγοριών ανίχνευσης.

Στην παραπάνω εικόνα παρατηρούμε τις τρεις βασικές διαστάσεις. Αρχικά έχουμε το είδος της ανίχνευσης, το οποίο χωρίζεται σε χειρωνακτική και αυτοματοποιημένη ανίχνευση. Η χειρωνακτική ανίχνευση βασίζεται στην ανθρώπινη αξιολόγηση και σε πρακτικές επαλήθευσης, ενώ η αυτοματοποιημένη ανίχνευση βασίζεται σε υπολογιστικούς αλγορίθμους για την αυτόματη αναγνώριση ψευδών ειδήσεων. Η σημασιολογική ανίχνευση εστιάζει στο νόημα και στην γλωσσική δομή του περιεχόμενου και αποτελεί υποκατηγορία της αυτοματοποιημένης ανίχνευσης.

Στην συνέχεια έχουμε την τεχνική ανίχνευσης, η οποία περιλαμβάνει τις τεχνικές μηχανικής μάθησης, τα νευρωνικά δίκτυα και τις υβριδικές τεχνικές. Οι τεχνικές μηχανικής μάθησης/νευρωνικών δικτύων είναι βασισμένες στην εκπαίδευση μοντέλων πάνω σε δεδομένα και έχουν ως στόχο την αυτόματη ταξινόμηση ειδήσεων ως αληθής ή ψευδής.

Τέλος, έχουμε τα είδη χαρακτηριστικών. Τέτοια είναι τα γλωσσικά, τα οποία αφορούν το λεξιλόγιο, τη σύνταξη και τη σημασιολογική συνοχή του κειμένου, τα επεξεργασμένα χαρακτηριστικά που συντελεστούν στην βελτιστοποίηση μοντέλων και την χρήση συνδυαστικών μεθόδων, το περιεχόμενο παραγόμενο από AI που εστιάζει στον εντοπισμό κειμένου που έχει παραχθεί από AI εργαλεία και η στάση (Stance Detection) που εξετάζει τη σχέση του κειμένου με ένα γεγονός ως προς τη στάση που εκφράζει.

4.2. Ανίχνευση ψευδών ειδήσεων

4.2.1. Είδη ανίχνευσης ψευδών ειδήσεων

4.2.1.1. Χειροκίνητη παροχή γνώσης

Με τον έλεγχο των γεγονότων μέσω διαφόρων έγκυρων πηγών γνώσης είναι εφικτή η ανίχνευση ψευδών ειδήσεων με αρκετή ακρίβεια. Ο πιο προφανής τρόπος (ανίχνευσης) αφορά το χειροκίνητο έλεγχο των ειδήσεων, κάτι που συμπεριλαμβάνει την ανθρώπινη κρίση. Σε αυτό το σημείο, είναι σημαντική η ‘αυθεντία’, δηλαδή το πόσο ειδικός είναι κάποιος στο θέμα που εξετάζει ως αληθή ή ψευδή είδηση.

Ο έλεγχος των ειδήσεων παρέχεται στο ευρύ κοινό μέσω ιστοσελίδων. Για παράδειγμα, το PolitiFact [34] αποτελεί μια ιστοσελίδα που αξιοποιεί ειδικούς πάνω σε συγκεκριμένα θέματα πολιτικής φύσεως ώστε να δώσει στατιστικά στοιχεία σχετικά με την ορθότητα κάποιας θεματικής. Επιπλέον, το FactCheck [35] αξιολογεί την ορθότητα ειδήσεων αναφορικά με την πολιτική σκηνή των ΗΠΑ με τη βοήθεια ειδικών. Σε παρόμοιες θεματικές επικεντρώνεται και το Washington post fact-checker [36]. Άλλες ιστοσελίδες καλύπτουν διαφορετικά πεδία, όπως για παράδειγμα το Snopes [37], το οποίο εξετάζει φημολογίες, δίνοντας περισσότερους χαρακτηρισμούς ως προς την ορθότητα μιας πληροφορίας, όπως για παράδειγμα το αν η πληροφορία αυτή είναι πλέον παρωχημένη ή μη αποδεδειγμένη. Παρομοίως, το TruthOrFiction [38] καλύπτει πληθώρα θεματικών, αναθέτοντας χειρωνακτικά ταμπέλες αληθείας ή ψεύδους σε κάθε δείγμα.

Από την άλλη μερία, μεταβαίνοντας από τους ειδικούς στη γνώση του πλήθους, άλλες ιστοσελίδες όπως το Fiskkit [39], επιτρέπουν σε χρήστες το ανέβασμα άρθρων και το λεπτομερή χαρακτηρισμών τους. Κατά τον ίδιο τρόπο, άλλοι χρήστες μπορούν επίσης να χαρακτηρίσουν το ίδιο άρθρο, δίνοντας τελικά τη δυνατότητα στον καθένα να αξιοποιεί την κρίση του κάθε χρήστη προκειμένου να αποφανθεί για την εγκυρότητα κάθε άρθρου. Ωστόσο, η γνώση των ειδικών ή του κοινού είναι συνδεδεμένη με περιορισμούς, όπως η αδυναμία αξιολόγησης μεγάλου όγκου περιεχομένου άρθρων σε μικρό χρόνο και η διαρκής ανάγκη για ολοένα αυξανόμενη ανθρώπινη συνεισφορά, καθώς οι ανάγκες για αξιολόγηση ψευδών ειδήσεων διογκώνονται όσο διευκολύνεται η δημιουργία και διάδοσή τους. Ωστόσο η χρήση του κοινού ενέχει επιπλέον κινδύνους, καθώς δεν είναι πάντοτε άρτια εκπαιδευμένο και μπορεί να αναπαράγει έτοιμο περιεχόμενο. Παράλληλα ελλοχεύει ο κίνδυνος σκόπιμης παραπληροφόρησης από αυτοματοποιημένους πράκτορες (bots) με σκοπό την χειραγώγηση του κοινού σε αυτούς τους ιστότοπους.

4.2.1.2. Αυτοματοποιημένη παροχή γνώσης

Για το λόγο αυτό, η δημιουργία βάσεων γνώσης είναι αναγκαία για να αυτοματοποιηθεί η διαδικασία ελέγχου ψευδών ειδήσεων κατά το δυνατό περισσότερο. Οι βάσεις γνώσης μπορεί να παρέχονται σε διάφορες μορφές. Η εξαγωγή γνώσης μέσω λέξεων-κλειδιών όντως αποτελεί κοινή πρακτική στην επεξεργασία μεγάλου όγκου πληροφορίας, ιδιαίτερα όταν η πηγή είναι αδόμητη, όπως το περιεχόμενο του διαδικτύου. Οι λέξεις κλειδιά χρησιμοποιούνται ως σημείο αναφοράς, οι οποίες έπειτα μετατρέπονται σε δεδομένα. Έτσι, λοιπόν, η γνώση αυτή εξάγεται με λέξεις κλειδιά, τροφοδοτώντας αλγορίθμους μηχανικής μάθησης (δέντρα απόφασης, Naive Bayes ταξινομητές κλπ.) προκειμένου να γίνει εντοπισμός συναισθήματος που εξάγεται από την κάθε είδηση [41].

Για την διασύνδεση της βάσης γνώσης με τους αλγορίθμους μηχανικής μάθησης, αρχικά συλλέγονται αυτόματα δεδομένα από το διαδίκτυο και δομούνται σημασιολογικά ώστε να είναι αναγνώσιμα από τα υπολογιστικά μοντέλα. Η δόμηση των δεδομένων δεν αφορά μονό την οργάνωση του κειμένου, αλλά και την παροχή ετικετών (labels) στα δεδομένα εκπαίδευσης, καθώς οι αλγόριθμοι μάθησης θέλουν συγκεκριμένες ετικέτες για να αναγνωρίσουν ένα πρότυπο. Έπειτα αυτά τα δεδομένα μπορούν να αποθηκευτούν σε μια βάση γνώσης και μετά αυτή η βάση χρησιμοποιείται ως σημείο αναφοράς για τον έλεγχο γεγονότων. Με αυτόν τον τρόπο η βάση γνώσης και οι αλγόριθμοι μηχανικής μάθησης λειτουργούν κυκλικά, δηλαδή η βάση δίνει στους αλγορίθμους επισημειωμένα (labeled) δεδομένα και οι αλγόριθμοι παράγουν νέα γνώση.

Ωστόσο, για την εγκυρότητα του συστήματος η γνώση αυτή υποβάλλεται σε κάποιον έλεγχο πριν ενσωματωθεί μόνιμα στην βάση. Με αυτό τον τρόπο αποφεύγεται η “μόλυνση” των δεδομένων που μπορούν να επιφέρουν την υποβάθμιση της αξιοπιστίας της ανίχνευσης. Έτσι η νέα γνώση εμπλουτίζει τη βάση. Συνεπώς, ο "αντίκτυπος" εδώ δεν είναι μια γενική έννοια, αλλά η ποσοτικοποίηση της έντασης και της κατεύθυνσης των συναισθημάτων που είναι ενσωματωμένα στο κείμενο.

Μια βάση γνώσης συχνά επίσης παρέχεται στη μορφή ενός γράφου γνώσης, οπότε και είναι εφικτή αλγοριθμικά η ανάκτηση στοιχείων και η μετέπειτα ενσωμάτωσή τους σε ένα σύστημα μηχανικής μάθησης, το οποίο λαμβάνοντας τόσο τη γνώση όσο και το προς εξέταση άρθρο, είναι πλέον σε θέση να αποφανθεί με κάποια βεβαιότητα σχετικά με την εγκυρότητά του. Για παράδειγμα, ο γράφος της Wikipedia είναι σε θέση να προσδώσει χρήσιμες πληροφορίες μέσω της σημασιολογικής εγγύτητας των στοιχείων εντός δικτύου και να εξεταστεί η συμφωνία της απόφασης αυτής σε σύγκριση με την ανθρώπινη κρίση [42]. Η τελική απόφαση του συστήματος για μια είδηση (ως αληθούς ή ψευδούς) συγκρίνεται με την ανθρώπινη κρίση, προκειμένου να επικυρωθεί η αξιοπιστία του μοντέλου [42]. Η απόφαση αυτή δεν αφορά την ίδια την εγγύτητα στοιχείων, αλλά το συμπέρασμα σχετικά με την εγκυρότητα ενός ισχυρισμού.

Η εγγύτητα στοιχείων σε σχέση με τον γράφο γνώσης είναι ένα από τα κριτήρια που χρησιμοποιούνται για να παρθεί αυτή η απόφαση. Στον γράφο, η εγγύτητα συνδέεται με τη θέση των κόμβων και τις συνδέσεις τους. Οι αλγόριθμοι εύρεσης συντομότερου μονοπατιού και μοτίβων συνεκτικότητας είναι σε θέση να αναγνωρίσουν πολλαπλές περιπτώσεις αληθών ή ψευδών ισχυρισμών, όταν αυτοί γίνονται σε μορφή τριπλέτας [44], αξιοποιώντας παράλληλα και τη δομή του γράφου γνώσης ώστε τελικά να παρέχουν αποτελεσματικά κάποιες εγγυήσεις στην ανίχνευση ψευδών στοιχείων [43].

Στην κατηγορία των τεχνικών γράφων, ιδιαίτερο ενδιαφέρον παρουσιάζει η προσέγγιση των ροών δικτύου (network flow). Η μέθοδος αυτή δεν αρκεί για την απλή διασύνδεση, αλλά αναδεικνύει συγκεκριμένα στοιχεία του γράφου που λειτουργούν ως αποδεικτικά τεκμήρια. Έτσι, καθίσταται δυνατή η εξαγωγή συμπερασμάτων (inferencing) σχετικά με την εγκυρότητα ενός ισχυρισμού, ιδιαίτερα όταν αυτός έχει τη δομή τριπλέτας (Υποκείμενο-Κατηγόρημα-Αντικείμενο). Στο ίδιο πλαίσιο, μεγάλες βάσεις γνώσης όπως η DBpedia [40] αξιοποιούνται για τη διασταύρωση γεγονότων μέσω εξειδικευμένων αλγοριθμικών τεχνικών. Μεταξύ αυτών, ξεχωρίζει η ανάλυση υπογραφής (signature analysis), η οποία πηγαίνει ένα βήμα πέρα από την απλή εύρεση μονοπατιού. Η τεχνική αυτή εστιάζει στα δομικά μοτίβα και την τοπολογία γύρω από τους κόμβους, δημιουργώντας ουσιαστικά μια μοναδική ταυτότητα για κάθε οντότητα, γεγονός που διευκολύνει την ταυτοποίηση. Τέλος, η σύγχρονη έρευνα στρέφεται και στις διανυσματικές αναπαραστάσεις (embeddings), όπου, σύμφωνα με τη βιβλιογραφία [45], ο έλεγχος της εγκυρότητας μεταφράζεται σε υπολογισμό αποστάσεων στον διανυσματικό χώρο, εντοπίζοντας τα σημασιολογικά κοντινότερα διανύσματα.

Εκτός του γράφου γνώσης, επιπλέον γνώση μπορεί να εξαχθεί από ‘έμμεσες’ πηγές γνώσης, όπως η γλώσσα των λογικών κανόνων. Με εξαγωγή κανόνων από τη γλώσσα (λογικοί κανόνες τύπου Horn), είναι εφικτό το ταίριασμα των στοιχείων του κανόνα με τα στοιχεία του γράφου γνώσης, αφού πρώτα εξεταστούν τα επιμέρους στοιχεία του κανόνα. Ουσιαστικά, τα επιμέρους στοιχεία μπορεί να είναι αληθή ή ψευδή, και προχωρώντας σε μεγαλύτερο τμήμα του κανόνα μπορούμε να αποφανθούμε για την εγκυρότητα ολόκληρου του ισχυρισμού [46]. Αυτοί οι λογικοί κανόνες τύπου Horn περιλαμβάνουν σχέσεις από τους γράφους, οντότητες που αντιστοιχούν σε άτομα τοποθεσίες κλπ., δεδομένα βασισμένα σε κείμενα, δηλαδή φράσεις ή αποσπάσματα που

στηρίζονται στις σχέσεις και την δομή συλλογισμού που καθορίζει ότι οι σχέσεις συνδέονται λογικά μεταξύ τους.

4.2.2. Είδη χαρακτηριστικών εστίασης

4.2.2.1. Γλωσσικά χαρακτηριστικά

Με την αυτόματη εξαγωγή χαρακτηριστικών από δεδομένα ειδήσεων είναι εφικτή η εύρεση μοτίβων για διαχωρισμό των ειδήσεων σε αληθείς και ψευδείς. Γλωσσικά στοιχεία, όπως για παράδειγμα λεξιλογικά και συντακτικά χαρακτηριστικά, όπως επίσης και στυλ γραφής είναι συχνά συνδεδεμένα με την ορθότητα μιας είδησης. Τα λεξιλογικά χαρακτηριστικά συνήθως δεν απαιτούν χρήση αλγορίθμων μηχανικής μάθησης ή νευρωνικών δικτύων, και μπορεί να περιλαμβάνουν στατιστικά δεδομένα αναφορικά με τον αριθμό και το μήκος λέξεων και χαρακτήρων, όπως επίσης και συχνοτικά χαρακτηριστικά (πχ. πόσες φορές εμφανίστηκαν λέξεις που διαθέτουν πάνω από ένα συγκεκριμένο αριθμό χαρακτήρων). Από την άλλη, τα συντακτικά χαρακτηριστικά, όπως η εύρεση μέρους του λόγου (part of speech – POS) και στοιχεία σε επίπεδο πρότασης (πχ εύρεση n-grams) συχνά αξιοποιούν κάποια μικρά νευρωνικά δίκτυα. [47]. Εκτός των γλωσσικών χαρακτηριστικών, η αξιοποίηση των συνοδευόμενων φωτογραφιών που συχνά εμφανίζονται σε άρθρα οδηγεί σε προσθήκη επιπλέον χαρακτηριστικών σε έναν αλγόριθμο ταξινόμησης, ο οποίος τελικά αποφαίνεται για την εγκυρότητα μιας πηγής. Τα επιπλέον χαρακτηριστικά είναι οπτικά και στατιστικά δεδομένα εικόνων από την ανάλυση των συνοδευτικών φωτογραφιών ενός άρθρου. Είναι σημαντικά, διότι όταν συνδυάζονται με λεξιλογικά και συντακτικά χαρακτηριστικά προσφέρουν συμπληρωματική πληροφόρηση για την εγκυρότητα της είδησης. [48]

Δεδομένης της ανόδου των κοινωνικών δικτύων, τα χαρακτηριστικά των χρηστών τόσο του αρχικού δημιουργού/συγγραφέα όσο και των χρηστών αποτελούν κριτήρια διαχωρισμού μιας είδησης, όπως για παράδειγμα ο αριθμός ακολούθων (follower number). Τέτοια στοιχεία μπορούν να αξιοποιηθούν για την εκπαίδευση ενός ταξινομητή μηχανικής μάθησης, όπως για παράδειγμα ενός δέντρου απόφασης [49]. Εξετάζοντας συγκεντρωτικά πολλούς χρήστες, είναι εφικτή η εύρεση των χαρακτηριστικών των χρηστών, οι οποίοι διαβάζουν συγκεκριμένα άρθρα. Η εγκυρότητα ή μη αυτών των άρθρων συνοδεύεται από διαχωρίσιμα χαρακτηριστικά στις αντίστοιχες ομάδες χρηστών. Συγκεκριμένα, έχει μελετηθεί ένα σύνολο χρονικών, δομικών και γλωσσολογικών διαχωρισμών για τις διάφορες ομάδες χρηστών, τα οποία εφαρμόζονται στην εκπαίδευση μοντέλων μηχανικής μάθησης, όπως δέντρα αποφάσεων και support vector machines (SVMs) [50].

Σε πιο συγκεκριμένα σενάρια χρήσης, όπως για παράδειγμα στην εύρεση ψευδών ισχυρισμών στις βουλευτικές εκλογές των ΗΠΑ, γλωσσικά χαρακτηριστικά είναι σε θέση να παρέχουν ικανοποιητική αναγνώριση. Αξιοποιώντας το PolitiFact ως την απόλυτη αλήθεια σχετικά με το αν μια πολιτική είδηση είναι αληθής ή ψευδής, μαζί με το πρόγραμμα LIWC [51] για ανάλυση κειμένου, ήταν εφικτή η εξαγωγή γλωσσικών χαρακτηριστικών και η συσχέτισή τους με το αν πρόκειται για αληθή ή ψευδή είδηση [52]. Με χρήση πάλι του PolitiFact, δεδομένα συλλέγονται και επισημειώνονται, επιτρέποντας έτσι την αξιολόγηση νέων πολιτικών ειδήσεων με χρήση

μοντέλων μηχανικής μάθησης (πχ. SVM, Logistic Regression) και νευρωνικών δικτύων (πχ. LSTM, CNNs) [53].

Δεδομένου ότι το στυλ γραφής μπορεί να συσχετιστεί με το χαρακτηρισμό μιας είδησης ως αληθούς ή ψευδούς, στυλομετρικά (stylometric) χαρακτηριστικά μπορούν να εξαχθούν από τα άρθρα. Για παράδειγμα, πολλά ψευδή άρθρα συχνά προσομοιάζουν τη σάτιρα με βάση το στυλ γραφής, το οποίο μπορεί να προσδιοριστεί με χρήση λεξιλογικών στοιχείων βασισμένων στα μέρη του λόγου και συχνοτικές αναλύσεις παρουσίας συγκεκριμένων λέξεων. Τα χαρακτηριστικά αυτά τροφοδοτούν στη συνέχεια αλγορίθμους μηχανικής μάθησης που εκπαιδεύονται να προβλέπουν την ταξινόμηση νέων ειδήσεων [54]. Η αξιοποίηση περισσότερων χαρακτηριστικών συνδεδεμένων με το στυλ γραφής, όπως για παράδειγμα η χρήση σημείων στίξης, ψυχογλωσσικά χαρακτηριστικά, συντακτικό και αναγνωσιμότητα μπορεί να προσδώσει μεγαλύτερη εγκυρότητα στην ταξινόμηση ειδήσεων, αυξάνοντας τις μετρικές αξιολόγησης της ταξινόμησης, όπως αυτή επιτυγχάνεται από έναν ταξινομητή SVM [55].

Άλλες προσεγγίσεις στοχεύουν στην εξέταση πιο συχνών γλωσσικών χαρακτηριστικών, επιτυγχάνοντας καλύτερη γενίκευση του μοντέλου ανίχνευσης ψευδών ειδήσεων. Με χρήση ταξινομητών τύπου δασών τυχαίων αποφάσεων (random forests) αποδεικνύεται όχι μόνο καλύτερος διαχωρισμός μεταξύ ψευδών και αληθών ειδήσεων βάσει στυλ γραφής, αλλά και η αναγνώριση της σάτιρας [56]. Ψυχολογικά και κοινωνικά φαινόμενα, όπως το confirmation bias (προκατάλειψη επιβεβαίωσης) και echo chamber effect θάλαμος απήχησης) αποτυπώνονται στον τρόπο γραφής, δίνοντας έτσι μια επιπλέον δυνατότητα για τη χρήση σχετικών χαρακτηριστικών στην αναγνώριση ψευδών ειδήσεων με τη βοήθεια ταξινομητών μηχανικής μάθησης [57].

Θεωρώντας πιο γενικά γλωσσικά χαρακτηριστικά σε επίπεδο παραγράφου αντί λέξης απαιτείται ένα πιο ικανό μοντέλο μηχανικής/βαθιάς μάθησης, το οποίο να είναι σε θέση να αξιολογήσει το περιεχόμενο γύρω από κάθε λέξη. Αυτό επετεύχθη μέσω του μηχανισμού προσοχής (attention mechanism), ο οποίος είναι σε θέση να συνυπολογίζει τις λέξεις που υπάρχουν σε μία πρόταση, δίνοντας την ανάλογη ‘προσοχή’ σε αυτές ανάλογα το περιεχόμενο [58]. Με χρήση του μηχανισμού προσοχής, είναι πλέον εφικτή η εξαγωγή χαρακτηριστικών από μεγαλύτερα κομμάτια κειμένου, επιτρέποντας την ανίχνευση ψευδών ειδήσεων με βάση το ευρύτερο πλαίσιο κάθε λέξης [59].

Η επέλαση των μέσων κοινωνικής δικτύωσης έχει επιφέρει αλλαγές στον τρόπο γραφής, ειδικά σε περιβάλλοντα αλληλεπίδρασης όπου οι χρήστες αντιδρούν και αλληλεπιδρούν σε μια δημοσίευση σε κοινωνικό μέσο. Τα νέα αυτά χαρακτηριστικά επικεντρώνονται στη χρήση emoji αλλά και γενικότερα στην ενσωμάτωση φορτισμένων και υβριστικών λέξεων, τα οποία συνδέονται έντονα με δημοσιεύσεις παραπληροφόρησης. [60]

Τα γλωσσικά χαρακτηριστικά που εξάγονται από ένα άρθρο επιτρέπουν πλέον την αξιοποίηση εξωτερικών πηγών μέσω σημασιολογικής αναπαράστασης. Συγκεκριμένα, η χρήση διανυσματικών εμφυτεύσεων (word embeddings) μετατρέπει το κείμενο από απλή ακολουθία χαρακτήρων σε πολυδιάστατους φορείς εννοιών. Κατ’ αυτόν τον τρόπο, οι τεχνικές ανάκτησης

πληροφορίας δεν περιορίζονται στην απλή ταύτιση λέξεων-κλειδιών, αλλά αξιοποιούν το διαδίκτυο ως μια δυναμική βάση γνώσης. Μέσω της ομοιότητας συνημιτόνου (cosine similarity) των embeddings, καθίσταται εφικτή η αναγνώριση σύνθετων εννοιών και η εύρεση σχετικών πηγών με βάση το νοηματικό τους πλαίσιο (context) και όχι μόνο τη λεκτική διατύπωση [61].

Αυτή η προσέγγιση εισάγει προηγμένα χαρακτηριστικά στην ανάλυση, καθώς επιτρέπει στο μοντέλο να αντιλαμβάνεται τη συνάφεια μεταξύ ετερογενών πηγών πληροφορίας, βελτιώνοντας την ακρίβεια στην επαλήθευση των ισχυρισμών [61].

Μια διαφορετική προσέγγιση συστήνει τη χρήση εγκεκριμένων θεωριών από κοινωνικές επιστήμες και ψυχολογία για την ανίχνευση ψευδών ειδήσεων, απομακρύνοντας την εστίαση στα δεδομένα. Η εξέταση των άρθρων πραγματοποιείται σε πολλά επίπεδα, συνυπολογίζοντας λεξικά (πχ. χρήση λέξεων που υποδηλώνουν βεβαιότητα ή έντονο συναίσθημα), συντακτικά (πχ. συχνότητα σημείων στίξης, κεφαλαίων γραμμάτων ή δομή προτάσεων), σημασιολογικά (πχ. ανάλυση θέματος και νοηματική συνοχή) και διαλογικά χαρακτηριστικά (πχ. το ύφος της ρητορικής και η πρόθεση πειθούς), με όλα τα εξαγόμενα μοτίβα να συγκρίνονται με αντίστοιχες θεωρίες. Η συγκεκριμένη προσέγγιση δίνει τη δυνατότητα ανίχνευσης ψευδών ειδήσεων σε οποιοδήποτε στάδιο διάδοσης, δεδομένου ότι δεν αξιοποιεί κοινωνικές αλληλεπιδράσεις και αντιδράσεις, επιτρέποντας την πολύ έγκαιρη ανίχνευση ψευδών ειδήσεων [72].

4.2.2.2. Στάση

Η αναγνώριση στάσης (Stance detection) αφορά τον προσδιορισμό του κατά πόσο ο συγγραφέας μιας είδησης τοποθετείται υποστηρικτικά, αρνητικά ή ουδέτερα απέναντι στο περιεχόμενο της είδησης. Η μέθοδος αυτή αποτελεί ένα εξειδικευμένο χαρακτηριστικό ανώτερου επιπέδου, καθώς υπερβαίνει τη στατική ανάλυση του κειμένου και εξετάζει τη "σχέση" μεταξύ διαφορετικών τμημάτων της πληροφορίας. Η αναγκαιότητα ξεχωριστής ανάλυσης της στάσης έγκειται στο γεγονός ότι επιτρέπει την ανίχνευση ασυνεπειών (contextual incongruity), όπως στην περίπτωση που το σώμα ενός κειμένου διαψεύδει έναν προκλητικό τίτλο (clickbait). Η χρήση πιο προηγμένων τεχνικών σε σχέση με την απλή εξαγωγή λέξεων και εύρεση συχνοτικών μοτίβων είναι αναγκαία στην εύρεση στάσης, καθώς απαιτείται μια πιο συνολική παρά τοπική εξέταση του κειμένου, θέτοντας τη χρήση πιο προηγμένων μεθόδων μηχανικής μάθησης στο προσκήνιο [62].

Με γνώμονα την αξιοποίηση της στάσης για ανίχνευση ψευδών ειδήσεων, ο διαγωνισμός 2017 Fake News Challenge Stage 1 (FNC-1) [123] παρείχε μια βαθιά ανάλυση μοντέλων μηχανικής μάθησης και βαθιάς μάθησης με σκοπό την εξαγωγή χρήσιμων συμπερασμάτων και τη δημιουργία ανταγωνιστικών baselines για την πρόοδο στην αντιμετώπιση του συγκεκριμένου προβλήματος. Η σχετική ανάλυση υποδεικνύει ότι ακόμη και ποιοτικά γλωσσικά χαρακτηριστικά δεν επαρκούν στην παροχή επαρκούς πληροφορίας στο τελικό μοντέλο ώστε να ανιχνεύσει δύσκολες περιπτώσεις [62]. Μετατρέποντας το αρχικό πρόβλημα της στάσης του συγγραφέως στη στάση του τίτλου ως προς το περιεχόμενου του άρθρου, καθώς επίσης και την ενσωμάτωση πολλαπλών γλωσσικών χαρακτηριστικών σε ένα νευρωνικό δίκτυο LSTM είναι εφικτή η βελτίωση της επίδοσης των βασικών μοντέλων του FNC-[63].

Η χρήση νευρωνικών δικτύων συχνά απαιτεί τροποποιήσεις στην αρχιτεκτονική τους παρά εξαντλητική εξέταση των γλωσσικών χαρακτηριστικών, στρέφοντας την προσοχή προς το ίδιο το μοντέλο. Κατ' αυτόν τον τρόπο, ακόμα και με χρήση των βασικών προσεκτικά επιλεγμένων χαρακτηριστικών, είναι εφικτή η βελτιωμένη γενίκευση πολυστρωματικών νευρωνικών δικτύων (MLP) σε νέα δεδομένα [64]. Η παράλληλη χρήση λεξιλογικών χαρακτηριστικών, τα οποία αποτυπώνονται με σημασιολογικό τρόπο μέσω embeddings που ενσωματώνουν το πλαίσιο κάθε λέξης, μαζί με μοντέλα βαθιάς μάθησης πολλών σταδίων είναι σε θέση να βελτιώσει ακόμα περισσότερο την ανίχνευση της στάσης του τίτλου ως προς το περιεχόμενο, δίνοντας έναυσμα στη δυνατότητα ανίχνευσης ψευδών ειδήσεων [65]. Προκειμένου να επιτευχθεί ακόμα καλύτερος διαχωρισμός των ψευδών ειδήσεων στις περιπτώσεις που το ποσοστό αληθείας είναι βαθμονομημένο παρά δυαδικό (αληθές ή ψευδές), η ενσωμάτωση περισσότερων σταδίων σε ένα μοντέλο βαθιάς μάθησης, σε συνδυασμό με γλωσσικά χαρακτηριστικά επιτυγχάνει βελτίωση της ταξινόμησης πολιτικών ειδήσεων[66].

4.2.2.3. Τεχνικές επεξεργασίας χαρακτηριστικών

Ενώ προηγουμένως αναλύσαμε τη σημασία εξαγωγής κι επιλογής χαρακτηριστικών από το κείμενο προς αξιολόγηση, δεν έχουμε ακόμη συμπεριλάβει κάποιες τεχνικές μετατροπής των χαρακτηριστικών προκειμένου αυτά να καταστούν 'καταλληλότερα' αναφορικά με το μοντέλο μηχανικής μάθησης ή βαθιάς μάθησης που θα τροφοδοτήσουν στη συνέχεια.

Η χρήση τεχνικών μείωσης διαστάσεων των αρχικών δεδομένων, όπως για παράδειγμα το Principal Component Analysis (PCA) και το Chi-square, μπορούν να βελτιώσουν ακόμα περισσότερο την απόδοση των νευρωνικών δικτύων ως προς την ανίχνευση ψευδών ειδήσεων, επιτυγχάνοντας πολύ υψηλή ακρίβεια ταξινόμησης [67].

Ένα από τα βασικά προβλήματα στην εκπαίδευση αξιόπιστων μοντέλων ανίχνευσης ψευδών ειδήσεων είναι η ανεπάρκεια και η ανισορροπία των διαθέσιμων δεδομένων (class imbalance). Η συλλογή και η επιβεβαίωση των ειδήσεων από ειδικούς είναι μια εξαιρετικά κοστοβόρα διαδικασία, οπότε τα διαθέσιμα σύνολα δεδομένων συνήθως είναι μικρά. Επιπλέον, οι αληθινές ειδήσεις είναι πολύ περισσότερες από τις ψευδείς, οδηγώντας τα μοντέλα σε μεροληπτική συμπεριφορά.

Για να αντιμετωπίσουμε αυτό το ζήτημα προτείνεται η χρήση της επαύξησης δεδομένων (data augmentation).Η στρατηγική αυτή αποσκοπεί στην αύξηση του συνόλου εκπαίδευσης. Αυτό γίνεται με την δημιουργία τεχνητών νέων παραδειγμάτων από τα ήδη υπάρχοντα. Ωστόσο, η αποτελεσματικότητά της εξαρτάται από το ίδιο το μοντέλο ταξινόμησης. Οι κλασικοί αλγόριθμοι μηχανικής μάθησης (τα SVM ή τα Random Forests), θεωρούνται "ρηχές" προσεγγίσεις και δυσκολεύονται να καταλάβουν την επιπλέον πληροφορία που προσφέρει η επαύξηση και συνήθως δεν βελτιώνονται ιδιαίτερα σε σχέση με το αρχικό τους επίπεδο [68].

Αντιθέτως, η χρήση νευρωνικών δικτύων κρίνεται απαραίτητη, διότι λόγω της μεγάλης τους χωρητικότητας (capacity) και της ιεραρχικής δομής που έχουν, επιτρέπουν την εκμετάλλευση της ποικιλομορφίας των επαυξημένων δεδομένων. Η επαύξηση δεν είναι απλώς μια προ-επεξεργασία,

αλλά συνδέεται άμεσα με τους μηχανισμούς προσοχής (attention mechanisms) των νευρωνικών δικτύων, βοηθώντας το μοντέλο να εντοπίσει βαθύτερα σημασιολογικά μοτίβα που οι κλασικοί αλγόριθμοι αδυνατούν να αντιληφθούν [69]. Τέλος, η επαύξηση πρέπει να ταιριάζει απόλυτα στις ανάγκες του συγκεκριμένου προβλήματος, στοχεύοντας όχι μόνο στην αύξηση τον δεοδμένων, αλλά και στην βελτίωση της ποιότητας και της ποικιλίας των δεδομένων [70].

Στρέφοντας την προσοχή αντί της εισόδου στην έξοδο ενός μοντέλου ανίχνευσης ψευδών ειδήσεων, τεχνικές συνόλου (ensemble) έχουν τη δυνατότητα βελτίωσης της απόδοσης σε σχέση με μεμονωμένους ταξινομητές. Για παράδειγμα, η εφαρμογή της τεχνική της πλειοψηφικής ψηφοφορίας (ensemble voting) σε μια ομάδα βασικών ταξινομητών μηχανικής μάθησης επιτυγχάνει ιδιαιτέρως ανταγωνιστικές επιδόσεις, ακόμα και σε σχέση με τις πιο σύνθετες αρχιτεκτονικές νευρωνικών δικτύων. Αξιοποιώντας μια πληθώρα βασικών μοντέλων η προσέγγιση αυτή καταφέρνει να πλησιάσει ή και να ξεπεράσει τα όρια ακρίβειας (state-of-the-art) για συγκεκριμένα σύνολα δεδομένων. Η στρατηγική σύνθεσης αποτελεσμάτων είναι εξίσου αποτελεσματική με τη χρήση βαθιών μοντέλων [71].

4.2.2.4. Περιεχόμενο παραγόμενο από AI

Στην προηγούμενη ανάλυσή μας θεωρήσαμε ότι οι ειδήσεις που εξετάζουμε έχουν δημιουργηθεί από ανθρώπους με κακόβουλες προθέσεις. Ωστόσο, η εξέλιξη της τεχνητής νοημοσύνης στον τομέα της δημιουργίας ρεαλιστικών κειμένων, οδηγεί σε μία νέα πρόκληση, την αυτόματη παραγωγή ψευδών ειδήσεων (Neural Fake News). Άρα, σε αυτό το πλαίσιο κύριο μέλημα δεν είναι οι ποσότητα των δεδομένων που διαθέτουμε, αλλά η ποιότητά τους. Τα σύγχρονα μοντέλα παράγουν κείμενα με φυσικό τρόπο, σε βαθμό που τα παραδοσιακά κριτήρια, όπως ο εντοπισμός γλωσσικών λαθών, ή το “περίεργο” ύφος δεν μας είναι πλέον χρήσιμα.

Καθώς η αναπαραγωγή γλωσσικών μοτίβων από την τεχνητή νοημοσύνη ολοένα και περισσότερο αυξάνεται και γίνεται καλύτερη, η διάκριση αληθινής και κατασκευασμένης είδησης καθίσταται εξαιρετικά δύσκολη. Πλέον οι αλγόριθμοι ανίχνευσης καλούνται να εντοπίσουν στατιστικά αποτυπώματα που δεν είναι πλέον ορατά μέσω της απλής ανάλυσης των κειμένων.

Μια χαρακτηριστική προσέγγιση που ξεχωρίζει και δίνει την λύση είναι τα ανταγωνιστικά νευρωνικά δίκτυα (Generative Adversarial Networks-GANs). Υπάρχουν δύο μοντέλα ο γεννήτορας που παράγει πειστικές ψευδείς ειδήσεις και έναν ο ταξινομητής που καλείται να τις εντοπίσει. Για να επιτευχθεί αυτό χρησιμοποιούμε μοντέλα Transformer [58], τα οποία εκπαιδεύονται για να αναπαράγουν σύνθετα γλωσσικά μοτίβα. Για να είναι όμως επιτυχημένη η ανίχνευση αυτών των "νευρωνικών" ψευδών ειδήσεων πρέπει ο ταξινομητής να χρησιμοποιεί την ίδια ακριβώς αρχιτεκτονική Transformer με τον γεννήτορα [73]. Αυτό συμβαίνει διότι με την πλούσια σημασιολογική αναπαράσταση που μαθαίνει το μοντέλο κατά τη δημιουργία του κειμένου είναι ικανό πλέον να αναγνωρίσει μικρές στατιστικές λεπτομέρειες (artifacts). Εν κατακλείδι η αρχιτεκτονική Transformer έχει δύο “όψεις”, αρχικά είναι το μέσο παραγωγής της απειλής, αλλά είναι και το κατάλληλο εργαλείο για την αντιμετώπισή της.

Με την κλασική έννοια των ανταγωνιστικών νευρωνικών δικτύων, έχουμε έναν παραγωγό και έναν ταξινομητή που εκπαιδεύονται ταυτόχρονα. Με βάση τους Zellers et al. και Goldstein et al.[74], τονίζουμε πως η εκπαίδευση δεν είναι απαραίτητα ταυτόχρονη, αλλά είναι συχνά σειριακή (sequential). Αρχικά εκπαιδεύεται ένας generative Transformer και στην συνέχεια εκπαιδεύεται το discriminative Transformer. Αυτό όντως μοιάζει ιδεολογικά με την αρχή των GANs, ωστόσο εδώ συχνά μιλάμε για ανταγωνιστική αξιολόγηση (adversarial evaluation) και όχι για ένα κανονικό GAN. Η κύρια διαφορά τους είναι στον τρόπο που τα δύο δίκτυα επικοινωνούν μεταξύ τους.

Στα κλασσικά GANs τα δύο δίκτυα εκπαιδεύονται ταυτόχρονα και ανταγωνίζονται συνέχεια μέσα σε ένα επίπεδο μηδενικού αθροίσματος, ενώ στην συγκεκριμένη περίπτωση το γεννητικό μοντέλο παράγει το υλικό του και μετά ο ταξινομητής εκπαιδεύεται ξεχωρίστηκα πάνω σε αυτό. Η μεθοδολογία αυτή αποδεικνύεται πιο αποδοτική, διότι ο ανιχνευτής χρησιμοποιεί την ίδια δομή με τον γεννήτορα και με αυτό τον τρόπο εντοπίζονται μικρές διαφορές (artifacts) που ένας παραδοσιακός ταξινομητής δεν θα τις εντόπιζε.

Η ανταγωνιστική αξιολόγηση είναι μια πιο εξειδικευμένη διαδικασία, όπου ο ανιχνευτής είναι το τελικό κριτήριο για το παραγόμενο κείμενο εάν αυτό είναι πειστικό ή όχι. Τέλος, να τονίσουμε εδώ το γεγονός ότι η εκπαίδευση ανιχνευτών σε δεδομένα που έχουν παραχθεί από AI είναι σημαντική για να διατηρηθεί η αποτελεσματικότητά τους καθώς τα Large Language Models (LLMs) βελτιώνονται.

Η άνοδος των σχετικών δημιουργικών μοντέλων γλώσσας (generative language models) αναμένεται να δυσχεράνει ακόμη περισσότερο την ανίχνευση ψευδών ειδήσεων, όσο τα σχετικά μοντέλα βελτιώνονται, εναποθέτοντας τις δυνατότητες αποφυγής παραγωγής ψευδών ειδήσεων στους ίδιους τους δημιουργούς αυτών των μοντέλων [74]. Συγκεκριμένα, τα μεγάλα γλωσσικά μοντέλα (Large Language Models – LLMs) διαθέτουν ιδιαιτέρως ανταγωνιστικές ικανότητες παραγωγής γλώσσας στο σημείο που το παραχθέν κείμενο είναι ποιοτικά ισοδύναμο με αυτό που έχει γραφτεί από ανθρώπους. Παρ' όλα αυτά, έχουν γίνει κάποιες ανεξάρτητες προσπάθειες αναγνώρισης χαρακτηριστικών σε άρθρα παραπληροφόρησης που έχουν παραχθεί από γλωσσικά μοντέλα με εξαγωγή κρίσιμων γλωσσικών χαρακτηριστικών που 'προδίδουν' την ταξινόμησή τους. Προκύπτει επίσης ότι τα μοτίβα ψευδών ειδήσεων διαφέρουν σε περιπτώσεις που έχουν παραχθεί από AI σε σχέση με περιπτώσεις παραγωγής τους από ανθρώπους, επιβεβαιώνοντας τη μειωμένη δυνατότητα ανίχνευσης από μοντέλα εκπαιδευμένα σε πραγματικά δεδομένα (δηλ. δεδομένα που αφορούν ψευδές ειδήσεις κατασκευασμένες από ανθρώπους). Η εκπαίδευση νευρωνικών δικτύων με δεδομένα παραγμένα από AI μπορεί να επιλύσει το πρόβλημα αυτό [75].

Από την άλλη μεριά, ένα σύστημα ανίχνευσης εκπαιδευμένο σε συνθετικά δεδομένα ψευδών ειδήσεων παρουσιάζει ιδιαίτερα χαμηλή επίδοση στην περίπτωση ανίχνευσης ειδήσεων γραμμένων από ανθρώπους. Ο περιορισμός αυτός εμφανίζεται σε διαφορετικά μοντέλα ανίχνευσης ψευδών ειδήσεων, τα οποία αξιοποιούν την αρχιτεκτονική Transformer πάνω στην οποία εκπαιδεύονται λίγα μόνο στρώματα νευρώνων πάνω σε συγκεκριμένα δεδομένα (fine – tuning). Λόγω του συγκεκριμένου περιορισμού, συστήνονται στρατηγικές εξισορρόπησης

πραγματικών και AI παραγμένων δεδομένων για την εκπαίδευση ταξινομητών ανίχνευσης ψευδών ειδήσεων, δεδομένου ότι στο μέλλον θα είναι πιθανό να εμφανίζονται ψευδείς ειδήσεις τόσο παραγμένες από ανθρώπους όσο και από LLMs. [76]

Η παραγωγή πιο πειστικών ψευδών ειδήσεων είναι ολοένα πιο εφικτή, θέτοντας ως σημαντικό ερώτημα το κατά πόσο τα μοντέλα ανίχνευσης ψευδών ειδήσεων παραμένουν ανταγωνιστικά σε περιπτώσεις που τα LLM βελτιστοποιούν τη σύνθεση ψευδών άρθρων, φέρνοντάς τα από πλευράς γλωσσικών χαρακτηριστικών όσο το δυνατό πιο κοντά γίνεται στις πραγματικές ειδήσεις. Το fine – tuning ταξινομητών βασισμένων σε αρχιτεκτονικές Transformer, όπως επίσης και το fine – tuning LLMs ως ταξινομητές επιφέρει αρκετά ικανοποιητικά αποτελέσματα ανίχνευσης, υπογραμμίζοντας την ανάγκη χρήσης συνθετικών άρθρων για εκμάθηση των απαραίτητων μοτίβων μεταξύ ψευδών ή μη ειδήσεων. Να επισημάνουμε εδώ τον κίνδυνο της "καταστροφικής λήθης" (catastrophic forgetting) [77], ο οποίος αφορά τις περιπτώσεις όπου η τυχαία μεταβολή των παραμέτρων κατά την επανεκπαίδευση ενδέχεται να αλλοιώσει την προϋπάρχουσα γνώση του μοντέλου.

Για την δημιουργία ανθεκτικών ταξινομητών δεν αρκεί η απλή προσαρμογή, άλλα πρέπει να εκπαιδεύσεις το μοντέλο από την αρχή, πάνω σε υβριδικά σύνολα δεδομένων. Ο συνδυασμός ανθρώπινων και συνθετικών ψευδών ειδήσεων, διασφαλίζει ότι ο αλγόριθμος καταλαβαίνει τα μοτίβα παραπλάνησης χωρίς να περιορίζεται στις ιδιαιτερότητες μιας συγκεκριμένης πηγής. Έτσι το συνθετικό περιεχόμενο [77] γίνεται απαραίτητο συμπλήρωμα επιτρέποντας στους ανιχνευτές να αναγνωρίζουν τα νέα, εξελισσόμενα μοτίβα που εισάγουν τα σύγχρονα γλωσσικά μοντέλα.

4.2.3. Τεχνική ανίχνευσης

Συνεχίζοντας την κατηγοριοποίηση , προχωράμε στην τελευταία διάσταση που είναι η τεχνική ανίχνευσης. Αποτελεί τον πυρήνα των σύγχρονων συστημάτων ταξινόμησης Η σημασιολογική ανίχνευση αναζητάει πληροφορίες για να τις διασταυρώσει με εξωτερικές πηγές γνώσης, ενώ η τεχνική ανίχνευσης εστιάζει στα υπολογιστικά μοντέλα που επεξεργάζονται τα δεδομένα Σύμφωνα με την πολυδιάστατη προσέγγιση που προτείνεται, οι τεχνικές αυτές χωρίζονται σε τρεις βασικές κατηγορίες: μηχανική μάθηση, νευρωνικά δίκτυα και υβριδικές τεχνικές. Οι υβριδικές τεχνικές, συγκρατούν τα θετικά από κάθε μεθοδολογία, για παράδειγμα συνδυάζουν δομημένη γνώση με βαθιά μάθηση και με αυτό τον τρόπο καταφέρνουν να κάνουν τα συστήματα ανίχνευσης πιο ακριβή και πιο ανθεκτικά.

4.2.3.1 Τεχνικές μηχανικής μάθησης

Οι προσεγγίσεις της παραδοσιακής μηχανικής μάθησης βασίζονται στην εκπαίδευση αλγορίθμων πάνω σε επισημειωμένα σύνολα δεδομένων (datasets). Σκοπός τους είναι να καταλαβαίνουν αυτόματα τις αληθινές από τις ψεύτικες ειδήσεις. Βασικό ρόλο εδώ έχει μηχανική χαρακτηριστικών (feature engineering), όπου χρειαζόμαστε την χειροκίνητη εξαγωγή και επιλογή των κατάλληλων μεταβλητών από κείμενο και μεταδεδομένα. [125]

Για την ψηφιακή αναπαράσταση χρησιμοποιούνται συνήθως στατιστικές μέθοδοι όπως το TF-IDF (Term Frequency-Inverse Document Frequency) και τα N-grams, τα οποία επιτρέπουν τη μετατροπή των γλωσσικών δεδομένων σε αριθμητικά διανύσματα. Αυτά τα δεδομένα

τροφοδοτούν αλγορίθμους όπως οι μηχανές διανυσμάτων υποστήριξης(SVM) και η λογιστική παλινδρόμηση(logistic regression) οι οποίες αποτελούν σταθερές επιλογές όταν η ταξινόμηση βασίζεται σε γλωσσικά χαρακτηριστικά. Παράλληλα αλγόριθμοι όπως τα δεντρά απόφασης (decision trees) χρησιμοποιούνται για την ανάλυση μεταδεδομένων χρηστών, για παράδειγμα πόσους ακόλουθους ή πόσες δημοσιεύσεις έχει [126].

Επίσης, τεχνικές όπως τα τυχαία δάση (random forrest) έχουν φανεί ιδιαίτερα αποτελεσματικές όταν χρειάζεται να γίνει διαχωρίσιμος της παραπληροφόρησης και με αυτό τον τρόπο επιτυγχάνεται υψηλή γενίκευση του μοντέλου. Συνολικά, οι τεχνικές της παραδοσιακής μηχανικής μάθησης παραμένουν εξαιρετικά χρήσιμες στην ανίχνευση ψευδών ειδήσεων λόγω της υψηλής ερμηνευσιμότητάς (interpretability) τους και των συγκριτικά χαμηλών υπολογιστικών απαιτήσεων που παρουσιάζουν [124].

4.2.3.2 Νευρωνικά δίκτυα

Η εξέλιξη στο πεδίο οδήγησε στην υιοθέτηση των νευρωνικών δικτύων και της βαθιάς μάθησης. Η υιοθέτησή τους ξεχωρίζει διότι πρέπει να μάθουν αυτόματα πως να αναπαριστούν μη δομημένο κείμενο. Για παράδειγμα τα συνελικτικά νευρωνικά δίκτυα (CNNs) χρησιμοποιούνται για τον εντοπισμό τοπικών μοτίβων μέσα στα κείμενα, ενώ τα δίκτυα μακράς βραχυπρόθεσμης (LTSM) διαχειρίζονται την αλληλουχία και το πλαίσιο των πληροφοριών.

Έμφαση δίνεται στους μηχανισμούς προσοχής (attention mechanisms), καθώς βοηθούν τα μοντέλα να εστιάσουν σε λέξεις ή φράσεις σε κείμενα μεγάλα, σύνθετα και οπότε η ακρίβειά τους βελτιώνεται. Σύγχρονες τεχνικές επεξεργασίας της φυσικής γλώσσας χρησιμοποιούν διανυσματικές αναπαραστάσεις (word embeddings), που η εγκυρότητα ενός ισχυρισμού ελέγχεται μέσω της ομοιότητας συνημίτονου(cosine similarity), συγκρίνοντας το άρθρο με αξιόπιστες πήγες στο διαδίκτυο.

Εν κατακλείδι η επιλογή της τεχνικής ανίχνευσης εξαρτάται άμεσα από την φύση των δεδομένων. Η παραδοσιακή μηχανική μάθησης παραμένει ισχυρή για δομημένα χαρακτηριστικά, ενώ τα νευρωνικά δίκτυα και η βαθιά μάθηση αποτελούν τον καλύτερο τρόπο για την ανάλυση του ευρύτερου πλαισίου (content) και των σύνθετων γλωσσικών δομών που χρησιμοποιούνται στις ψευδείς ειδήσεις.

4.2.3.3 Υβριδικές τεχνικές

Οι υβριδικές τεχνικές αποτελούν την πιο σύγχρονη και ολοκληρωμένη προσέγγιση στο πρόβλημα της ανίχνευσης, καθώς επιδιώκουν να συνδυάσουν τα πλεονεκτήματα διαφορετικών μεθοδολογιών για την επίτευξη καλύτερης ακρίβειας και αξιοπιστίας. Στηρίζονται στο γεγονός ότι οι ψευδείς ειδήσεις είναι πολυδιάστατα φαινόμενα, που είναι δύσκολο να αναλυθούν πλήρως εστιάζοντας μόνο σε μια διάσταση μόνο. Στην πράξη οι τεχνικές αυτές λειτουργούν σε δύο επίπεδα:

- Συνδυασμός χαρακτηριστικών: χρησιμοποιούν συνδυασμό αρχιτεκτονικών βαθιάς μάθησης για την ανάλυση του μη δομημένου περιεχομένου και μηχανικής μάθησης που επεξεργάζονται μεταδεδομένα και προφίλ χρηστών.
- Συνδυαστική μάθηση (ensemble learning): αξιοποίηση πολλαπλών μοντέλων ταξινόμησης

που λειτουργούν παράλληλα. Ο συνδυασμός προβλέψεων όλων των επιμέρους μοντέλων, οδηγεί στην μείωση σφαλμάτων και στην ενίσχυση της ανθεκτικότητας του συστήματος.

Επιπλέον, προχωρημένες υβριδικές προσεγγίσεις συνδυάζουν τη σημασιολογική ανάλυση μέσω γράφων γνώσης (knowledge graphs) με νευρωνικά δίκτυα επεξεργασίας φυσικής γλώσσας. Με αυτόν τον τρόπο, το σύστημα δεν εξετάζει μόνο τη δομή και το ύφος του κειμένου, αλλά επαληθεύει ταυτόχρονα την εγκυρότητα των γεγονότων έναντι αξιόπιστων πηγών [127, 128].

4.3. Συγκριτική ανάλυση ερευνητικών προσεγγίσεων ανίχνευσης

Σε αυτήν την ενότητα παρουσιάζουμε έναν πίνακα συγκέντρωσης και σύγκρισης προσεγγίσεων μέσα από τα επιστημονικά άρθρα που έχουμε μελετήσει. Σκοπός του πίνακα είναι να γίνει σωστά η συγκριτική ανάλυση ερευνητικών προσεγγίσεων ανίχνευσης ψευδών ειδήσεων Ο παρακάτω πίνακας περιέχει τα κριτήρια που χρησιμοποιούμε καθώς και τα άρθρα που μελετήθηκαν.

Τα κριτήρια αυτά είναι η βάση γνώσης, δηλαδή από που συλλέγει τα δεδομένα, το είδος αλγορίθμου μηχανικής μάθησης που ενδεχομένως χρησιμοποιείται, το είδος νευρωνικών δικτύων που ενδεχομένως χρησιμοποιούνται, αν γίνεται χρήση δεδομένων παραγόμενα από AI, το είδος των χαρακτηριστικών που γίνονται εκμεταλλεύσιμα, δηλαδή τι πληροφορίες χρησιμοποιεί το μοντέλο, ο βαθμός/επίπεδο απόδοσης, η πολυπλοκότητα, η κλιμακωσιμότητα και το κόστος.

Ως βαθμό/επίπεδο απόδοσης ορίζουμε την αποτελεσματικότητα της κάθε προσέγγισης. Η αποτίμηση γίνεται κυρίως με βάση την ακρίβεια (accuracy). Οι τιμές που έχουμε είναι η υψηλή με ποσοστό ακρίβειας μεγαλύτερο του 90%, την καλή με ποσοστό 75% έως 90% και η χαμηλή με ποσοστό ακρίβειας μικρότερο του 75%.

Όσον αφορά την πολυπλοκότητα ορίζεται ως η απεικόνιση του υπολογιστικού κόστους της κάθε προσέγγισης. Οι τιμές που έχουμε είναι η χαμηλή που αφορά μεθόδους με ελάχιστες υπολογιστικές απαιτήσεις, όπως απλές γλωσσικές αναλύσεις ή απευθείας αναζητήσεις σε βάσεις γνώσης (π.χ. Wikipedia), η μεσαία που γίνεται χρήση κλασικών αλγορίθμων μηχανικής μάθησης ή μεθόδων συνόλου (π.χ Ensemble Learning) και η υψηλή που γίνεται χρήση βαθιών νευρωνικών δικτύων (Deep Learning), μοντέλων Transformers ή πολυτροπικών προσεγγίσεων.

Η κλιμακωσιμότητα αξιολογεί την ικανότητα ενός μοντέλου να επεξεργάζεται τεράστιους όγκους δεδομένων σε πραγματικό χρόνο. Η χαμηλή τιμή σε αυτό το κριτήριο αφορά πολυτροπικές προσεγγίσεις που αναλύουν ταυτόχρονα εικόνα, κείμενο και βίντεο και απαιτούν την κατασκευή ολόκληρου του γράφου. Η μέτρια τιμή αφορά μοντέλα που απαιτούν ανάλυση δικτύου σε τοπικό επίπεδο ή βαθιά ανάλυση περιεχομένου και χρειάζονται σημαντικούς υπολογιστικούς πόρους και τέλος έχουμε την υψηλή κλιμακωσιμότητα στην οποία τα μοντέλα βασίζονται σε απλά γλωσσικά χαρακτηριστικά και αναλύουν πολλές αναρτήσεις σε μικρό χρονικό διάστημα.

Τέλος το κόστος εστιάζει στο κόστος της ανάλυσης και της επεξεργασίας, ενώ είναι κυρίως υπολογιστικό. Οι τιμές του είναι το χαμηλό, δηλαδή έχουμε γλωσσική ανάλυση και μεταδεδομένα, στο μέτριο όπου έχουμε βαθιά μάθηση και υβριδικά μοντέλα και στο υψηλό όπου έχουμε πολυτροπικές προσεγγίσεις και δυναμικά γραφήματα διάδοσης.

Άρθρο	Βάση Γνώσης	Μηχανική Μάθη	Νευρωνικά Δίκτ	Δεδομένα AI	Χαρακτηριστικά	Απόδοση	Πολυπλοκότητα	Κλιμακωσιμότη	Κόστος
[41]	Web	✓ Supervised an	X	X	Σημασιολογικά/Σ	Καλή	Μεσαία	Μέτρια	Μέτριο
[42]	Wikipedia	X	X	X	Σημασιολογικά	Υψηλή	Χαμηλή	Υψηλή	Χαμηλό
[43]	Wikipedia	X	X	X	Συντακτικά	Καλή	Χαμηλή	Υψηλή	Χαμηλό
[44]	DBpedia	X	X	X	Σημασιολογικά	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[45]	DBpedia	X	X	X	Σημασιολογικά	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[46]	DBpedia/YAGO	X	X	X	Γλωσσικά/Σημασ	Καλή	Μεσαία	Μέτρια	Μέτριο
[47]	Web	✓ Supervised Le	X	X	Γλωσσικά/Συντα	Καλή	Χαμηλή	Υψηλή	Χαμηλό
[49]	Web	✓ Supervised Le	X	X	Διάδοσης	Καλή	Χαμηλή	Υψηλή	Χαμηλό
[50]	Web	✓ Ensemble Lea	X	X	Διάδοσης	Καλή	Μεσαία	Μέτρια	Μέτριο
[52]	Text	X	X	X	Γλωσσικά	Μέτρια	Χαμηλή	Υψηλή	Χαμηλό
[53]	Text	✓ Supervised Le	X	X	Μεταδεδομένα/Γ	Χαμηλή	Χαμηλή	Υψηλή	Χαμηλό
[54]	Text	✓ Supervised Le	X	X	Γλωσσικά	Καλή	Χαμηλή	Υψηλή	Χαμηλό
[55]	Text	✓ Supervised Le	X	X	Γλωσσικά	Μέτρια	Χαμηλή	Υψηλή	Χαμηλό
[56]	Text	✓ Supervised Le	X	X	Γλωσσικά	Μέτρια	Χαμηλή	Υψηλή	Χαμηλό
[57]	Text	✓ Semi-Supervis	X	X	Σημασιολογικά/Γ	Καλή	Υψηλή	Χαμηλή	Υψηλό
[59]	Text	X	✓ Recurrent Neu	X	Σημασιολογικά/Γ	Καλή	Υψηλή	Χαμηλή	Υψηλό
[60]	Web	✓ Ensemble Lea	X	X	Σημασιολογικά/Γ	X	Μεσαία	Μέτρια	Μέτριο
[61]	Web	✓ Supervised Le	X	X	Σημασιολογικά	Καλή	Μεσαία	Μέτρια	Μέτριο
[62]	Text	✓ Ensemble Lea	X	X	Σημασιολογικά	Καλή	Μεσαία	Μέτρια	Μέτριο
[63]	Text	X	X	X	Γλωσσικά/Σημασ	Καλή	Χαμηλή	Υψηλή	Χαμηλό
[64]	Text	X	✓ CNN	X	Γλωσσικά/Σημασ	Καλή	Μεσαία	Μέτρια	Μέτριο
[65]	Text	X	✓ CNN	X	Σημασιολογικά	Καλή	Υψηλή	Χαμηλή	Υψηλό
[66]	Text	X	✓ CNN and Hybı	X	Γλωσσικά/Σημασ	Καλή	Υψηλή	Χαμηλή	Υψηλό
[67]	Text	X	✓ RNN and CNN	X	Σημασιολογικά/Γ	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[68]	Text	✓ Supervised Le	X	X	Γλωσσικά	Υψηλή	Χαμηλή	Υψηλή	Χαμηλό
[69]	Text	X	✓ BERT Transfo	✓	Σημασιολογικά/Γ	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[70]	Text	✓ Supervised Le	X	✓	Σημασιολογικά/Γ	Καλή	Υψηλή	Χαμηλή	Υψηλό
[71]	Text	✓ Ensemble Lea	X	X	Σημασιολογικά	Καλή	Μεσαία	Μέτρια	Μέτριο
[72]	Web	✓ Supervised Le	X	X	Διάδοσης/Μεταδ	Καλή	Μεσαία	Μέτρια	Μέτριο
[73]	Text	X	✓ Transformers	✓	Σημασιολογικά	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[75]	Text	X	✓ Transformers	✓	Σημασιολογικά	Καλή	Υψηλή	Χαμηλή	Υψηλό
[76]	Text	X	✓ Transformers	✓	Σημασιολογικά	Καλή	Υψηλή	Χαμηλή	Υψηλό
[77]	Text	X	✓ Transformers	✓	Σημασιολογικά	Καλή	Υψηλή	Χαμηλή	Υψηλό

Εικόνα 2. Συγκεντρωτική ανάλυση άρθρων για ανίχνευση ψευδών ειδήσεων.

Με βάση τον πίνακα μπορούμε να εξάγουμε κάποια συμπεράσματα. Αρχικά θα τα χωρίσουμε σε δυο μεγάλες κατηγορίες:

1. Εξαγωγή συμπεράσματος για κάθε κριτήριο ξεχωριστά.
2. Εξαγωγή συμπεράσματος καθολικά.

Ας αναλύσουμε λοιπόν τις δυο αυτές κατηγορίες.

Για την εξαγωγή συμπεράσματος για κάθε κριτήριο ξεχωριστά έχουμε:

- **Χρήση βάσης γνώσης:** Παρατηρούμε ότι παλιότερα άρθρα χρησιμοποιούν δομημένες πήγες για την επαλήθευση γεγονότων πχ. Wikipedia και Dbpedia. Ωστόσο οι σύγχρονες πήγες φαίνεται ότι αντλούν δεδομένα από το Web. Με αυτό τον τρόπο αντιμετωπίζουμε την ταχύτητα και την ρευστότητα των ειδήσεων σε πραγματικό χρόνο.

- **Μηχανική μάθησης:** Η επιβλεπόμενη μάθηση κυριαρχεί και οι αλγόριθμοι SVM και random forest λόγω της αξιοπιστίας τους είναι συχνές επιλογές. Η συνδυαστική μάθηση ολοένα και αναδεικνύεται επειδή συνδυάζει διαφορετικούς αλγορίθμους και πετυχαίνει καλύτερη ακρίβεια.
- **Νευρωνικά δίκτυα:** Ακμαία αύξηση χρήσης τους λόγο της αύξησης της ακρίβειας στον εντοπισμό ψευδών ειδήσεων.
- **Δεδομένα παραγόμενα από AI:** Αποτελεί ανερχόμενο πεδίο. Χρησιμοποιείται από τις πιο πρόσφατες έρευνες και εστιάζει στην ανίχνευση κειμένων που γράφονται από μοντέλα όπως LLMs. Πλέον δεν μας απασχολεί μόνο η ποσότητα αλλά και η ποιότητα των δεδομένων.
- **Είδος χαρακτηριστικών:** Οι περισσότερες προσεγγίσεις συνδυάζουν σημασιολογικά και γλωσσικά χαρακτηριστικά. Παρόλα αυτά κρίνεται αναγκαία η ενσωμάτωση νέων χαρακτηριστικών για την ανίχνευση ψευδών ειδήσεων, οι οποίες γλωσσικά πλέον είναι άρτιες. Τέτοια χαρακτηριστικά είναι αυτά της διάδοσης και των μεταδεδομένων, τα οποία χρησιμοποιούνται από κοινωνικά δίκτυα.
- **Βαθμός/Επίπεδο απόδοσης:** Οι περισσότερες μελέτες έχουν "καλή" ή "υψηλή" απόδοση. Όταν συνδυάζονται προσεγγίσεις με βάσεις γνώσης και με νευρωνικά δίκτυα το ποσοστό τείνει να ξεπερνάει το 90%.
- **Πολυπλοκότητα:** Παρατηρείται άμεση σχέση με την τεχνολογία. Οι μέθοδοι που βασίζονται σε γλωσσική ανάλυση έχουν χαμηλή πολυπλοκότητα, ενώ με την χρήση Transformers και GANs έχουν υψηλή πολυπλοκότητα, εξαιτίας των απαιτήσεων που χρειάζονται για την εκπαίδευση και την εκτέλεση των σχετικών μοντέλων.
- **Κλιμακωσιμότητα:** Παρατηρήθηκε ότι παλιές γλωσσικές μέθοδοι έχουν υψηλή κλιμακωσιμότητα, ενώ οι σύγχρονες μέθοδοι βαθιάς μάθησης έχουν χαμηλή κλιμακωσιμότητα. Αυτό σημαίνει στην πράξη ότι για να ελεγχθεί η ροή πληροφοριών σε ένα ολόκληρο κοινωνικό δίκτυο, χρειάζονται τεράστιες υπολογιστικές υποδομές.
- **Κόστος:** Παρατηρήθηκε ότι μέθοδοι όπως BERT ή GANs έχουν υψηλό κόστος. Αντιθέτως μέθοδοι που χρησιμοποιούν μεταδεδομένα ή στατιστικές αναλύσεις κειμένου έχουν χαμηλό κόστος.

Για την εξαγωγή καθολικών συμπερασμάτων έχουμε:

- **Τάσεις:** Υπάρχει ξεκάθαρη μετάβαση από knowledge graph που χρησιμοποιούνται στην αρχή σε machine learning μοντέλα και στη συνέχεια σε deep learning μοντέλα.
- **Μεθοδολογική Σύγκλιση και Υβριδικά Μοντέλα:** Οι ερευνητές "αφήνουν" πίσω τις κλασικές μεμονωμένες μεθόδους και στρέφονται σε υβριδικά μοντέλα. Με λίγα λόγια γίνεται συνδυασμός δομημένης γνώσης καθώς και μηχανικής και βαθιάς μάθησης για την επίτευξη αποτελεσματικότερης ανίχνευσης. Επίσης, η στατική ανάλυση κειμένου δεν είναι αρκετή πλέον καθώς κρίνεται αναγκαία να αντληθούν δεδομένα σε πραγματικό χρόνο μέσω του web για την αντιμετώπιση των ψευδών ειδήσεων. Οπότε, σύνθεση των μοντέλων είναι αναγκαία.

- **Πρόκληση τα δεδομένα παραγόμενα από AI:** Πλέον γίνεται αντιληπτό πως η ανάγκη για ανίχνευση περιεχόμενου παραγόμενη από AI, ολοένα και αυξάνεται, διότι πλέον παράγονται ποιότητα και φυσικά κείμενα που τα παλιά φίλτρα ανίχνευσης δεν μπορούν να τα αντιμετωπίσουν. Έτσι φτάνουμε στο κρίσιμο σημείο που τα μοντέλα ανίχνευσης πρέπει να εκπαιδεύονται όχι μόνο σε ανθρώπινα χαρακτηριστικά, αλλά και σε δεδομένα παραγόμενα από AI.
- **Σχέση απόδοσης-πολυπλοκότητας:** Όσο πιο σύνθετο είναι το μοντέλο τόσο καλύτερα θα είναι τα αποτελέσματα. Κλασικοί αλγόριθμοι (SVM, random forest), διατηρούν "καλή" απόδοση χωρίς πολλές απαιτήσεις, ενώ για να έχουμε "υψηλή" απόδοση γίνεται χρήση αρχιτεκτονικών transformer και GANs. Για την διασφάλιση της αξιοπιστίας δεν αρκεί να έχουμε καλούς αλγορίθμους. Πρέπει να κάνουμε επένδυση σε υπολογιστικούς πόρους.
- **Πολυεπίπεδη Ανάλυση Χαρακτηριστικών:** Τα ψεύτικα κείμενα έχουν βελτιώσει τα γλωσσικά χαρακτηριστικά τους και η αντιμετώπισή τους μόνο από τα γραμματικά ή τα συντακτικά στοιχειά έχει γίνει αρκετά πιο δύσκολη. Έτσι λοιπόν η έρευνα εστιάζει σε σημασιολογικά χαρακτηριστικά, αλλά και σε χαρακτηριστικά διάδοσης και μεταδεδομένα. Η δυνατότητα κατανόησης του πλαισίου και ο τρόπος διαμοιρασμού μιας είδησης στα κοινωνικά δίκτυα έχει πρωταγωνιστικό ρόλο για την επίτευξη της ταξινόμησής της.

Η ανάλυση που πραγματοποιήθηκε σε αυτό το κεφάλαιο απαντά στα ερευνητικά ερωτήματα 1 και 2: υπάρχει πληθώρα αυτοματοποιημένων τεχνικών, οι οποίες είτε αξιοποιούν υπάρχουσες πηγές γνώσης είτε όχι, με τις πιο πρόσφατες τεχνικές να στρέφονται σε τεχνικές που αξιοποιούν νευρωνικά δίκτυα. Ο ρόλος της τεχνητής νοημοσύνης είναι κεντρικός στην ανίχνευση ψευδών ειδήσεων αφού μπορεί να προσφέρει αποδοτικά μοντέλα ανίχνευσης. Ταυτόχρονα, περιπτώσεις όπου παρουσιάζονται συνθετικά δεδομένα ψευδών ειδήσεων μπορούν επίσης να αντιμετωπιστούν πάλι με χρήση μοντέλων ανταγωνιστικής μάθησης όπως τα GAN.

Ωστόσο δεν πρέπει να ξεχάσουμε και το γεγονός ότι παρατηρήσαμε σε αρκετά άρθρα που μελετήσαμε ότι υπάρχει συνδυασμός μοντέλων. Με τον όρο αυτό εννοούμε ότι δεν βασιζόμαστε μόνο σε ένα μοντέλο για ανίχνευση, αλλά χρησιμοποιούνται δύο ή περισσότερα μοντέλα για την επίτευξη καλύτερης ακρίβειας. Τέτοια άρθρα είναι τα εξής:

- Το [64] που χρησιμοποιεί Neural model και external features (λεξικά χαρακτηριστικά και embeddings) και έχει ως στόχο την αντιμετώπιση ανισορροπίας των δεδομένων. Ανήκει στην ευρεία κατηγορία υβριδικών αρχιτεκτονικών (Hybrid Architectures), διότι συνδυάζει διαφορετικά νευρωνικά δίκτυα.
- Το [65] που χρησιμοποιεί Lexical features → Neural model και στόχος του είναι να βελτιώσει την απόδοση σε διάφορα στάδια ταξινόμησης. Εμπεριέχεται στην ευρεία κατηγορία Multi-stage Pipelines (αγωγοί πολλαπλών σταδίων), διότι το πρώτο μοντέλο κάνει ταξινόμηση και στην συνέχεια ένα άλλο μοντέλο επεξεργάζεται το αποτέλεσμα για να έχουμε πιο ακριβή απόφαση.
- Το [66] που χρησιμοποιεί Feature-assisted Neural Modeling (παραδοσιακά χαρακτηριστικά + deep learning) με σκοπό την καλύτερη ακρίβεια με τον συνδυασμό knowledge και νευρωνικών δικτύων. Εντάσσεται στην κατηγορία Hybrid Architectures,

διότι συνδυάζει διαφορετικά νευρωνικά δίκτυα όπως CNN και RNN δίνοντας την δυνατότητα επεξεργασίας τόσο στην τοπική δομή των προτάσεων όσο και στο ευρύτερο πλαίσιο της είδησης. Έτσι η εξωτερική πληροφορία ενσωματώνεται στην λειτουργία των νευρωνικών δικτύων και βελτιώνεται η δυνατότητα της ταξινόμησης.

- Το [67] που χρησιμοποιεί υβριδική αρχιτεκτονική (CNN για χωρικά patterns, LSTM για χρονικές εξαρτήσεις) και έχει ως στόχο την εκμετάλλευση και των δύο τύπων πληροφορίας για να έχουμε καλύτερη απόδοση. Εντάσσετε στην κατηγορία Hybrid Architectures, γιατί συνδυάζει διαφορετικά νευρωνικά δίκτυα για εκμετάλλευση διαφορετικών ειδών πληροφορίας.
- Στο [69] οι ερευνητές συνδυάζουν τεχνικές επαύξησης δεδομένων (data augmentation) με μοντέλα seq2seq και τεχνικές φυσικής γλωσσικής συνεπαγωγής (NLI) και έχουν ως κύριο στόχο τους την ενίσχυση της ανθεκτικότητας των συστημάτων ανίχνευσης για τις ψευδείς ειδήσεις. Ο συνδυασμός αυτών των τεχνικών εντάσσεται στην κατηγορία της μάθησης συνόλου (ensemble learning), καθώς η επαύξηση δεδομένων έχει συμπληρωματικό ρόλο για την ενδυνάμωση του συνόλου εκπαίδευσης. Η αρχιτεκτονική στηρίζεται σε έναν αγωγό πολλαπλών σταδίων (multi-stage pipeline) στον οποίο τα μοντέλα seq2seq και NLI αλληλεπιδρούν διαδοχικά, δηλαδή το πρώτο στάδιο άφορα την δημιουργία επαύξησης δεδομένων, ενώ το δεύτερο στάδιο εστιάζει στο φιλτράρισμα πληροφοριών μεσώ της φυσικής γλωσσικής συνεπαγωγής (Natural Language Inference - NLI).

- Το [70] βασίζεται σε έναν αγωγό πολλαπλών σταδίων (multi-stage pipeline) που συνδυάζει επαύξηση δεδομένων (data augmentation) και μάθηση συνόλου (ensemble learning). Αρχικά εφαρμόζονται τεχνικές επαύξησης για τον εμπλουτισμό των συνόλων δεδομένων διασφαλίζοντας ποικιλομορφία. Στην συνέχεια γίνεται πέρασμα των δεδομένων σε ταξινομητές μηχανικής μάθησης (ML), οι οποίοι λειτουργούν συνδυαστικά μεσώ πλειοψηφικής ψηφοφορίας (ensemble voting). Με αυτό τον τρόπο το μοντέλο ξεπερνά τους περιορισμούς των αλγορίθμων και αυξάνει την γενίκευση και την ακρίβεια στην ανίχνευση των ψευδών ειδήσεων.

- Το [73] που συνδυάζει Generative model (Grover) για παραγωγή και detection μοντέλο (Grover discriminator) στοχεύοντας στην ανίχνευση neural fake news με χρήση της ίδιας γλωσσικής αναπαράστασης. Εντάσσεται στην ευρύτερη κατηγορία Multi-stage Pipelines, καθώς αρχικά το πρώτο μοντέλο κάνει την παραγωγή χαρακτήρων ψευδών ειδήσεων και στη συνέχεια το δεύτερο μοντέλο ανιχνεύει αυτά τα αποτελέσματα για πιο ακριβή απόφαση.

Όπως αναφέραμε και παραπάνω υπάρχει συνδυασμός μοντέλων. Ορισμένες προσεγγίσεις ξεχώρισαν. Αρχικά ξεχωρίσανε τα υβριδικά μοντέλα που συνδυάζουν βάσεις γνώσης (knowledge bases) με βαθία νευρωνικά δίκτυα (deep learning/transformers). Αυτό οφείλεται στο γεγονός πως τα επίπεδα απόδοσής τους είναι αρκετά υψηλά ξεπερνώντας το 90% (πχ [67], [69], [73]), διότι δεν περιορίζονται μόνο στην γλωσσική ανάλυση, αλλά παράλληλα χρησιμοποιούν διασταύρωση γεγονότων για την αντιμετώπιση της σημασιολογικής πολυπλοκότητας των σύγχρονων ψευδών ειδήσεων. Επίσης, ξεχωρίσανε ως οι χειρότερες προσεγγίσεις αυτές που βασίζονται αποκλειστικά σε ταξινομητές μηχανικής μάθησης με χρήση μόνο γλωσσικών χαρακτηριστικών. Τέτοια μοντέλα

που στηρίχθηκαν σε άπλες λεξιλογικές αναλύσεις (πχ. [52], [53]) εμφανίζουν χαμηλότερη απόδοση, διότι έχουν αδυναμία ανίχνευσης ειδήσεων παραγόμενων από AI.

4.4. Πρόληψη ψευδών ειδήσεων

Η παρεμπόδιση της διάδοσης ψευδών ειδήσεων περιλαμβάνει τόσο τεχνολογικά και αλγοριθμικά προϊόντα, όσο και τη μελέτη κοινωνικών φαινομένων που συνεισφέρουν στην ευκολία ή δυσκολία μετάδοσης. Οι προσπάθειες παρεμπόδισης και περιορισμού των ψευδών ειδήσεων χαρακτηρίζονται ως ιδιαιτέρως δύσκολες από τους ερευνητές, οι οποίοι χωρίζουν τις προτεινόμενες δράσεις στις ακόλουθες κατηγορίες [78]:

1. **Παρεμπόδιση της επιρροής:** πρόκειται για μια παράμετρο με αλγοριθμικά στοιχεία από τη γραφοθεωρία, η οποία στοχεύει στον ελάχιστο αριθμό ανθρώπων σε ένα κοινωνικό δίκτυο, οι οποίοι αν εξαιρεθούν από αυτό θα οδηγήσουν στον περιορισμό των ψευδών ειδήσεων στο δίκτυο. Αυτός ο ελάχιστος αριθμός ανθρώπων με βάση το άρθρο δεν χαρακτηρίζονται ως κακόβουλοι, αλλά παρουσιάζονται ως κόμβοι με ισχυρή επιρροή, οι οποίοι λειτουργούν ως μέσα διάδοσης. Δηλαδή έχουν την τάση να πιστεύουν και να διαδίδουν ψευδείς ειδήσεις.
2. **Καμπάνιες για διάδοση αληθινών ειδήσεων:** μία αμιγώς κοινωνική παράμετρος συνιστά την ενημέρωση του κοινού με πραγματικές ειδήσεις, στοχεύοντας έτσι τη μείωση της επιρροής των ψευδών, δεδομένου ότι σε πρότερες μελέτες έχει αποδειχθεί ότι οι χρήστες τείνουν εν τέλει να πιστέψουν την πραγματική είδηση, όταν έχουν να επιλέξουν ανάμεσα σε αυτή και σε μία ψευδή, γεγονός που αυθόρμητα οδηγεί σε μείωση της διάδοσης ψευδών γεγονότων. Ωστόσο για να επιτευχθεί αυτό θα πρέπει αφενός να δούμε αν αυτή η είδηση είναι πραγματική και αφετέρου, ο χρονισμός παίζει σημαντικό ρόλο, διότι η επιλογή ανάμεσα σε δύο εκδοχές έχει νόημα όταν αυτές παρουσιάζονται ταυτόχρονα και όχι η μια μετά την άλλη. Αν φθάσει πρώτα η ψευδής, ενδεχομένως να κάνει άμεση ζημία πριν αναγνωριστεί ως ψευδής έπειτα (με την άφιξη της αληθινής είδησης).
3. **Εργαλεία περιορισμού:** κυρίως αφορά τεχνολογικά προϊόντα βασισμένα σε κοινωνικές επιστήμες που υποβοηθούν την ανίχνευση και τον αποκλεισμό ψευδών ειδήσεων σε ένα κοινωνικό δίκτυο.

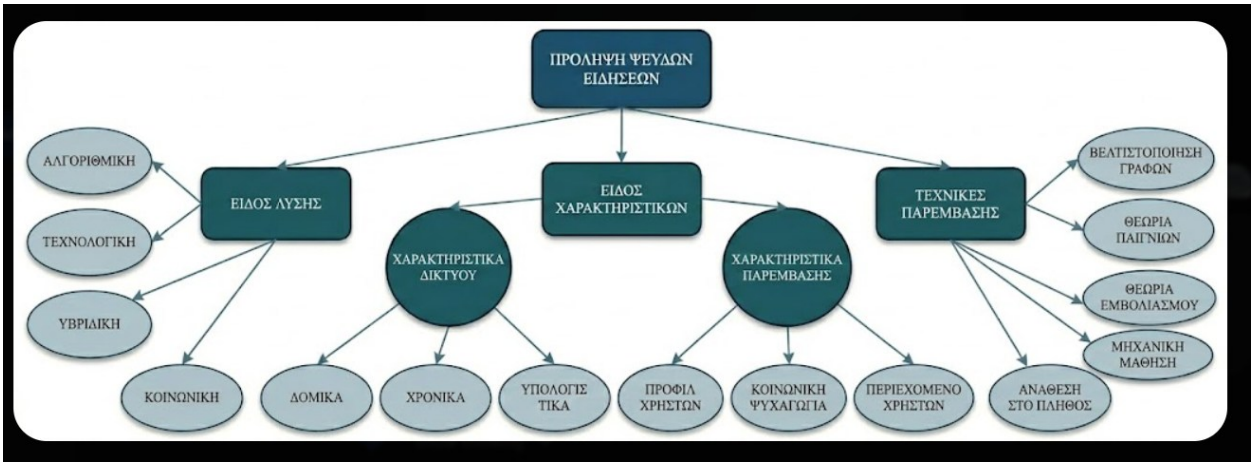
4.4.1 Κατηγοριοποίηση προσεγγίσεων πρόληψης ψευδών ειδήσεων

Στην παρακάτω εικόνα παρατηρούμε τις τρεις βασικές διαστάσεις της πρόληψης κατά των ψευδών ειδήσεων. Αρχικά έχουμε το είδος της λύσης, το οποίο χωρίζεται σε αλγοριθμική, σε κοινωνική, σε τεχνολογική και σε υβριδική. Η αλγοριθμική κάνει χρήση μαθηματικών μοντέλων και βελτιστοποίησης με σκοπό τον περιορισμό της διάδοσης, η κοινωνική εστιάζει στην εκπαίδευση, την ενημέρωση και την αλλαγή της ανθρώπινης συμπεριφοράς, η τεχνολογική στοχεύει στην ανάπτυξη συγκεκριμένων εργαλείων, εφαρμογών ή επεκτάσεων και τέλος η υβριδική λύση είναι ο συνδυασμός δομικής ανάλυσης δικτύου με κοινωνιολογικούς παράγοντες.

Στην συνέχεια έχουμε το είδος των χαρακτηριστικών, το οποίο περιλαμβάνει τα χαρακτηριστικά δικτύου και τα χαρακτηριστικά παρέμβασης. Τα χαρακτηριστικά δικτύου είναι δομικά (τοπολογία

γράφου), χρονικά (ταχύτητα διάδοσης) και υπολογιστικά στοιχεία, ενώ τα χαρακτηριστικά παρέμβασης, ασχολούνται με προφίλ χρηστών, κοινωνική ψυχολογία και το περιεχόμενο της ιδίας της πληροφορίας.

Τέλος, έχουμε τις τεχνικές παρέμβασης. Τέτοιες είναι η βελτιστοποίηση γράφων, η θεωρία παιγνίων, η θεωρία εμβολιασμού, η μηχανική μάθησης και τέλος η ανάθεση στο πλήθος. Στην βελτιστοποίηση γράφων έχουμε την αποκοπή ακμών ή κόμβων, στην θεωρία παιγνίων μοντελοποίηση της σύγκρουσης μεταξύ της αλήθειας και του ψεύδους, στην θεωρία εμβολιασμού έχουμε προληπτική έκθεση σε μικρές δόσεις παραπληροφόρησης για την ενίσχυση της κριτικής σκέψης, στην μηχανική μάθησης έχουμε αυτόματη παραγωγή διορθωτικών απαντήσεων και διασταύρωση γεγονότων και τέλος στην ανάθεση στο πλήθος έχουμε συλλογική επαλήθευση από τους ίδιους τους χρήστες.



Εικόνα 3. Ιεραρχικό δέντρο κατηγοριών πρόληψης.

4.4.2. Παρεμπόδιση επιρροής ατόμων σε κοινωνικά δίκτυα

Θεωρώντας τους χρήστες ενός κοινωνικού δικτύου ως κόμβους ενός γράφου, με τις μεταξύ τους ακμές να εκφράζουν τις αλληλεπιδράσεις των χρηστών, το πρόβλημα τις διάδοσης ψευδών ειδήσεων μεταφράζεται σε πρόβλημα εύρεσης ακμών και μονοπατιών πάνω στο γράφο. Κατ' αυτόν τον τρόπο, είναι εφικτή η αλγοριθμική επίλυση του προβλήματος της παρεμπόδισης μετάδοσης ψευδών ειδήσεων αναγνωρίζοντας τους χρήστες με τη μεγαλύτερη επιρροή στη διάδοσή τους.

Η εύρεση των χρηστών που αποτελούν την 'πηγή' διάδοσης στο γράφο και η τοποθέτηση οθονών (monitors) σε αυτούς είναι ένα αλγοριθμικά δύσκολο πρόβλημα, που όμως αντιμετωπίζει τη μετάδοση ψευδών ειδήσεων από τη ρίζα του. Η αλγοριθμική πολυπλοκότητα έγκειται στο γεγονός ότι δεν είναι γνωστό πόσοι χρήστες αποτελούν την 'πηγή' των ψευδών ειδήσεων, ούτε είναι φανερή η σχέση τους πάνω στο γράφο, καθώς δεν είναι αναγκαίο να σχετίζονται άμεσα, αλλά είναι πιθανό να μοιράζονται κοινές απόψεις ή να πληροφορούνται από τις ίδιες πηγές εκτός του

κοινωνικού δικτύου. Έτσι λοιπόν, ο εντοπισμός τους μπορεί να απαιτεί σχεδόν εξαντλητική αναζήτηση των πιθανών συνδυασμών, εγείροντας την ανάγκη για αλγοριθμικές βελτιστοποιήσεις.

Παρόλα αυτά, η τοποθέτηση monitors σε αυτούς είναι σαφώς ένα πολύ ευκολότερο πρόβλημα υπολογιστικά [79]. Ωστόσο η στρατηγική αυτή έχει κάποιο κόστος. Η τοποθέτηση monitors παντού δεν είναι ούτε αποδοτική ούτε απαραίτητη. Θα πρέπει να περιορίσουμε τους κόμβους σε "σίγουρους", "σχετικά σίγουρους" και "αμφιλεγόμενους" και να τοποθετήσουμε monitors μόνο για τα πρώτα 2 είδη κόμβων. Έτσι, δεν χρειάζεται να μπουν σε χρήστες που δεν πρόκειται να βοηθήσουν σημαντικά στη διάδοση των ψευδών ειδήσεων.

Σε παρόμοια λογική, η αντιμετώπηση διάδοσης φημών στον κοινωνικό γράφο αναγνωρίζεται ως ένα αλγοριθμικά πολύπλοκο πρόβλημα, ιδιαιτέρως σε περιπτώσεις περιορισμένου υπολογιστικού προϋπολογισμού. Για το λόγο αυτό, εισάγεται ένας αλγόριθμος τμηματοποίησης της κοινότητας που επιλέγει ένα υποσύνολο κόμβων προς αποκλεισμό, στοχεύοντας τόσο σε ανενεργούς όσο και σε αρνητικούς κόμβους, με στόχο τη βελτιστοποίηση του περιορισμού των φημών μέσω της αξιοποίησης της δομής της κοινότητας [80].

Ένας άλλος τρόπος του περιορισμού φημών σε μια κοινότητα αφορά την αξιοποίηση του άπληστου αλγορίθμου ελάχιστης κάλυψης κορυφών (minimum vertex cover) από την αλγοριθμική θεωρία, στοχεύοντας στην εύρεση του ελάχιστου αριθμού ατόμων (τα οποία αντιστοιχούν στις κορυφές ενός γράφου κοινότητας), τα οποία θα πρέπει να μπλοκαριστούν για να αποκοπεί η διάδοση της φήμης [81].

Μια πιο δυναμική μέθοδος μπλοκαρίσματος χρηστών υπολογίζει την επιρροή ενός χρήστη πρώτα σε τοπικό επίπεδο και εν συνεχεία επεκτείνεται δυναμικά σε όλο το δίκτυο. Με αυτόν τον επαναληπτικό τρόπο είναι εφικτό να βρεθούν οι πιο επιδραστικοί χρήστες πάνω στο δίκτυο, οι οποίοι και καλούνται να μπλοκαριστούν για τον περιορισμό διάδοσης φήμης [82]. Το πρόβλημα του μπλοκαρίσματος διάδοσης φημών αναφορικά με το πώς επηρεάζονται οι κοινότητες χρηστών συνιστά το αλγοριθμικό πρόβλημα του ελαχίστου κόστους μπλοκαρίσματος φήμης (Least Cost Rumor Blocking - LCRB). Η εύρεση των ακμών του γράφου, οι οποίες προσπίπτουν σε κόμβους που έχουν τουλάχιστον ένα γειτονικό κόμβο εντός μιας κοινότητας που έχει επηρεαστεί από τη φήμη αποτελεί ένα τρόπο επίλυσης του προβλήματος αυτού. [83]

Το πρόβλημα του περιορισμού φημών επιδέχεται επεκτάσεις, συχνότερα με τη μορφή συγκεκριμένων περιορισμών. Θέτοντας τον παράγοντα του χρονικού περιορισμού, το πρόβλημα του μπλοκαρίσματος διάδοσης φημών μπορεί να επεκταθεί συμπεριλαμβάνοντας το χρόνο που χρειάζεται για να αποδυναμωθεί η επιρροή ενός χρήστη. Έτσι, λοιπόν, αν ο αποκλεισμός απαιτεί ένα χρονικό παράθυρο, το οποίο ξεπερνά ένα προκαθορισμένο κατώφλι, η χρηστικότητα του δικτύου μειώνεται συνολικά, απαιτώντας μια δυναμική λύση μπλοκαρίσματος χρηστών [84]. Το συγκεκριμένο πρόβλημα, ειδικά στην περίπτωση που οι φήμες αφορούν πολλαπλές θεματικές, έχει μελετηθεί κατά κύριο λόγο με χρήση εξεζητημένων τεχνικών, αποδεικνύοντας ότι ανήκει στην NP-hard αλγοριθμική κλάση, η οποία συμπεριλαμβάνει δύσκολα και πολύπλοκα προβλήματα, τα οποία αδύνατο να επιλυθούν σε πολυωνυμικό χρόνο [85].

Γενικότερα, ο περιορισμός φήμης όταν υπόκειται σε περιορισμούς αποτελεί ένα πρόβλημα βελτιστοποίησης. Για παράδειγμα, το ακόλουθο είναι ένα πρόβλημα βελτιστοποίησης του ελάχιστου αριθμού χρηστών που διαδίδουν φήμες, όταν τουλάχιστον η επιρροή τους ξεπερνά ένα συγκεκριμένο κατώφλι ανοχής, έτσι ώστε να ελαχιστοποιηθεί το ολικό κόστος ως προϊόν της διάδοσης φήμης, ακόμα και σε περιπτώσεις που η φήμη καταφέρει να φτάσει σε όλους τους τελικούς αποδέκτες. Τέτοια γραμμικά προβλήματα μπορούν να επιλυθούν τόσο με άπληστες (greedy) τεχνικές, όσο και με αλγορίθμους γραμμικού προγραμματισμού (linear programming), οι οποίοι στοχεύουν στη βελτιστοποίηση της λύσης [86].

Ωστόσο στην περίπτωση που το πρόβλημα δεν είναι γραμμικό, θα πρέπει να χρησιμοποιηθούν άλλες τεχνικές, όπως για παράδειγμα ο προγραμματισμός με περιορισμους (constraint programming).

Περαιτέρω τεχνικές μπλοκαρίσματος φήμης αφορούν εξεζητημένες αλγοριθμικές μεθόδους, εστιάζοντας στη βελτιστοποίηση του αλγορίθμου γράφου από τη σκοπιά της θεωρητικής επιστήμης υπολογιστών παρά της επιστήμης των ψευδών ειδήσεων. Συνεπώς, κρίνουμε ότι ξεφεύγουν από το σκοπό της παρούσας εργασίας και αφορούν συγκεκριμένες θεωρητικές περιπτώσεις του προβλήματος. Για τέτοιου σκοπού εξεζητημένες λύσεις, παραπέμπουμε στη μελέτη των Saxena et al. [78].

4.4.3. Καμπάνιες διάδοσης της αλήθειας

Στις μεθόδους μπλοκαρίσματος επιρροής και διάδοσης φημών οι κόμβοι ή οι ακμές αποκόπτονται στρατηγικά για να αποτραπεί η περαιτέρω εξάπλωση των ψευδών ειδήσεων, μετριάζοντας έτσι τον αντίκτυπό τους. Ωστόσο, μια άλλη αποτελεσματική προσέγγιση είναι η αντιμετώπιση της επιρροής των ψευδών ειδήσεων με τη διάδοση αληθινών πληροφοριών στους χρήστες. Με την παροχή αληθινών πληροφοριών, οι χρήστες μπορούν να αποκτήσουν μια ολοκληρωμένη κατανόηση των ειδησεογραφικών θεμάτων από πολλαπλές οπτικές γωνίες, προωθώντας τον σχηματισμό αμερόληπτων απόψεων. Αυτή η ενημερωμένη προοπτική μπορεί να επηρεάσει τις αποφάσεις των χρηστών σχετικά με την κοινοποίηση περιεχομένου, ενθαρρύνοντάς τους να διαδίδουν επαληθευμένες πληροφορίες αντί για φήμες. Η μέθοδος αυτή δεν είναι μόνο πιο πρακτική σε σενάρια του πραγματικού κόσμου, αλλά μειώνει επίσης σημαντικά τον συνολικό αντίκτυπο των ψευδών ειδήσεων σε όλο το δίκτυο, προωθώντας τη διάδοση της αλήθειας.

Μια ενδιαφέρουσα οπτική αναδεικνύει ότι πολλοί χρήστες ενδιαφέρονται για άρθρα τα οποία αντικρούουν τις απόψεις τους και συχνά ξοδεύουν περισσότερο χρόνο πληροφόρησης σε αυτά, σε σχέση με άρθρα που συμφωνούν με την υπάρχουσα οπτική τους [87]. Κάτι τέτοιο θα μπορούσε να επανορθώσει την οπτική τους σε περίπτωση που έχει επηρεαστεί από ψευδείς ειδήσεις. Επιπλέον, η ανάδειξη των συγκεκριμένων άρθρων παραπληροφόρησης είναι αποτελεσματική όταν συνοδεύονται από τις αληθινές απόψεις που τα αντικρούουν, όπως αυτές υποστηρίζονται από επιστημονικά αποδεδειγμένες θέσεις [88], είτε όταν υπογραμμίζονται οι ελαττωματικές σκοπιές των επιχειρημάτων [89]. Όσον αφορά τη χρονική διάσταση της παρέμβασης για την αντιμετώπιση ψευδών ειδήσεων, η παροχή της πραγματικής πληροφορίας πριν την έκθεση του χρήστη στην ψευδή είδηση αποδεικνύεται πολύ αποτελεσματικότερη σε σύγκριση με την

αντίστροφη περίπτωση, όπου η πραγματική πληροφορία αποδίδεται αφότου ο χρήστης έχει εκτεθεί στην ψευδή είδηση [90].

Μία άλλη προσέγγιση εξετάζει το σκοπό διάδοσης μιας φήμης και της αντίστοιχης πραγματικής είδησης αντί του αριθμού των ανθρώπων που ενστερνίζονται την κάθε άποψη, προτείνοντας τη συγκεκριμένη μέθοδο σαν μια τεχνική αντιμετώπισης ψευδών ειδήσεων πιο κοντά στην πραγματικότητα [91]. Στην συγκεκριμένη προσέγγιση μελετήθηκε το πόσο αποτελεσματική είναι η διάψευση και η προειδοποίηση ως τρόποι περιορισμού των ψευδών ειδήσεων. Με αυτόν τον τρόπο εάν οι χρήστες λάβουν μια ξεκάθαρη διάψευση ή ένα ενημερωτικό μήνυμα, μειώνεται η πιθανότητα αναπαραγωγής περαιτέρω ψευδών ειδήσεων.

Πλατφόρμες και οργανισμοί μπορούν να σχεδιάζουν στρατηγικές που δεν στοχεύουν μόνο στη διόρθωση του περιεχομένου, αλλά και στην μείωση των στοιχείων που κάνουν μια είδηση ελκυστική προς κοινοποίηση. Έτσι λοιπόν το να κατανοήσουμε τον σκοπό διάδοσης είναι εργαλείο για να αναπτυχθούν στρατηγικές επικοινωνίας απέναντι στις ψευδείς ειδήσεις.

Πιο συγκεκριμένες τακτικές έχουν αναπτυχθεί σχετικά με καμπάνιες διάδοσης αληθών ειδήσεων. Όσον αφορά διαδόσεις μέσω του διαδικτύου, και θεωρώντας το διαδίκτυο ως ένα γράφο χρηστών, η άμυνα ενάντια στις ψευδείς ειδήσεις συνιστά την αύξηση του αριθμού ‘αληθινών’ κόμβων σε σχέση με τους ψευδείς [92]. Θεωρώντας τη σχέση αληθινών και ψευδών ειδήσεων ως ένα ανταγωνιστικό παίγνιο, η εύρεση της βέλτιστης στρατηγικής και για τους δύο είναι ένα NP-complete πρόβλημα, δηλαδή ανήκει στην κλάση των πολύ δύσκολων αλγοριθμικών προβλημάτων για τα οποία δε μπορεί να βρεθεί μια λύση πολυωνυμικού χρόνου. Ακόμη και η εύρεση πλεονεκτήματος δεν είναι προφανής, γιατί ακόμη κι αν ο ένας ‘παίκτης’ έχει το πλεονέκτημα του πρώτου, η επέκταση του δεύτερου σε περισσότερους κόμβους αναιρεί το προαναφερθέν πλεονέκτημα [93].

Με παρόμοιο τρόπο, η εύρεση του ελάχιστου αριθμού χρηστών, οι οποίοι διαθέτουν την απαραίτητη προθυμία να διαδώσουν μια αληθινή είδηση ελαχιστοποιώντας κατά το δυνατό την επιρροή όσων διαδίδουν ψευδείς φήμες, ανήκει επίσης στα NP-complete προβλήματα [94]. Σε αυτό το σημείο επανέρχεται η θεωρία των Yang et al. [86], οι οποίοι υποστηρίζουν ότι δοθέντων δύο επιλογών (αληθής - ψευδής είδηση), η αληθής τελικά θα επικρατήσει, δίνοντας έτσι μια κατεύθυνση στην επίλυση του ανταγωνιστικού προβλήματος που τίθεται στο [94].

Πέρα της μελέτης μεμονωμένων γράφων, αντιλαμβανόμαστε πως λειτουργεί η διάδοση πληροφοριών σε πολλά και επικαλυπτόμενα δίκτυα. Σύμφωνα με την θεωρία των παιγνίων οι λύσεις διαφοροποιούνται ανάλογα την δομή της εκάστοτε πληροφορίας. Κάποιες φορές οι ανταγωνιστές δρουν σε ξεχωριστά δίκτυα χωρίς να επηρεάζει ο ένας τον άλλο, ενώ άλλες φορές έχουμε επικαλυπτόμενα δίκτυα όπου ο χρήστης δέχεται μηνύματα από διάφορες πήγες ταυτόχρονα. Σε τέτοιες περιπτώσεις η στρατηγική μοντελοποιείται ως ένα παίγνιο μηδενικού αθροίσματος (zero-sum game). Στο σενάριο αυτό ο “αμυνόμενος” προσπαθεί να περιορίσει την διασπορά με την εφαρμογή στρατηγικών ελαχιστοποίησης (minimax), ενώ ο “επιτιθέμενος” προσπαθεί να μεγιστοποίηση την επιρροή του. Από κάθε οπτική γωνία μπορούμε να καταλάβουμε

τα κρίσιμα σημεία της ισορροπίας (Nash Equilibrium) στο δίκτυο, αναδεικνύοντας πως η δομή επικάλυψης δικτύων μπορεί είτε να βοηθήσει στην καταπολέμηση την παραπληροφόρησης είτε να ενισχύσει την ανθεκτικότητά της [98, 99].

Σε πραγματικές εφαρμογές παρατηρείται η σταδιακή μείωση επιρροής των ψευδών ειδήσεων με την πάροδο του χρόνου. Παρά το γεγονός ότι ο πρωτεύων αντίκτυπος της παραπληροφόρησης παραμένει σημαντικός αν καταλάβουμε ότι το φαινόμενο αυτό ασθενεί με τον καιρό μπορούμε να έχουμε καλύτερο σχεδιασμό αντιμετώπισης. Αν θέσουμε συγκεκριμένα χρονικά όρια που η επιρροή μιας είδησης εξασθενεί τότε οι αρμόδιοι κατανέμουν τους πόρους πιο αποδοτικά. Με βάση αυτά είναι ευκολότερο να προβλέψουμε ποιοι κόμβοι/χρήστες επηρεάζονται πιο εύκολα από μια ψευδή είδηση.

Η πρόβλεψη αυτή λειτουργεί ως εργαλείο στόχευσης, επιτρέποντας την προληπτική εστίαση των κόμβων του δικτύου που είναι πιθανότερο αρχικά να επηρεαστούν και στην συνέχεια να μεταδώσουν την ψευδή είδηση. Με αυτό τον τρόπο η παρέμβαση λειτουργεί ως μηχανισμός αναχαίτισης μειώνοντας έτσι τον αριθμό των χρηστών που έχουν επηρεαστεί [95].

Μια άλλη τακτική αφορά την επιλογή χρηστών οι αφενός είναι πρόθυμοι στην συμμετοχή της διάδοσης μιας αληθινής είδησης, αφετέρου είναι διατιθέμενοι να έχουν συμμετοχή και στην αναχαίτιση ψευδών ειδήσεων. Οι καλόβουλοι χρήστες παρακολουθούν το δίκτυο με σκοπό τον εντοπισμό των ενεργών χρηστών στοχεύοντας στην ανίχνευση παραπληροφόρησης. Μόλις εντοπίσουν κάποια παραπληροφόρηση αρχικά μπλοκάρουν τους χρήστες που διαδίδουν ψευδείς ειδήσεις για να σταματήσει η διασπορά. Στην συνέχεια ξεκινάει η διάδοση της αληθινής είδησης εκμεταλλευόμενοι την απομόνωση των πηγών της παραπληροφόρησης. Ωστόσο η απαλοιφή των χρηστών που συμμετέχουν στην ψευδή πληροφόρηση είναι ένα NP-complete πρόβλημα, στρέφοντας έτσι την προσοχή σε ευριστικές (heuristic) ή άπληστες (greedy) μεθόδους παρακολούθησης της ροής πληροφορίας και μπλοκαρίσματος κακόβουλων χρηστών. [97]

Συνολικά, οι καμπάνιες διάδοσης της αλήθειας βασίζονται σε κοινωνικές συμπεριφορές και τάσεις, ενώ αξιοποιούν αλγόριθμους, όπως στην περίπτωση της παρεμπόδισης επιρροής συγκεκριμένων χρηστών, αναδεικνύοντας την αλληλένδετη φύση κοινωνικών και αλγοριθμικών λύσεων για την αντιμετώπιση διάδοσης ψευδών ειδήσεων σε κοινωνικά δίκτυα.

4.4.4. Τεχνολογικές λύσεις

Ο σκοπός των τεχνολογικών λύσεων είναι να αναπτυχθούν εφαρμογές και εργαλεία, τα οποία θα είναι σε θέση να αναγνωρίζουν και να σταματούν τη διάδοση ψευδών ειδήσεων ή να παρέχουν πραγματικές πληροφορίες ως αντίμετρο.

Η παροχή ποικίλων απόψεων έχει αποδειχθεί ωφέλιμη, μιας και επιτρέπει στους ανθρώπους να ενημερωθούν σχετικά με απόψεις που αποκλίνουν από τις υπάρχουσες απόψεις τους [87]. Σε αυτή την κατεύθυνση, υπηρεσίες όπως η NewsCube[6] παρέχουν πολλαπλές οπτικές ως προς μια θεματική στους αναγνώστες, επιτρέποντάς τους να σχηματίσουν τη δική τους προσωπική άποψη

[6] http://newscube.io/

απαλλαγμένη από προκαταλήψεις [100]. Με σκοπό την εύρεση αντιφατικών πληροφοριών έχει σχεδιαστεί η εφαρμογή Dispute Finder[7], η οποία διατίθεται ως επέκταση στο φυλλομετρητή, επιτρέποντας την πρόσβαση σε ποικίλες ιστοσελίδες. Παράλληλα, παραθέτει άρθρα τα οποία υποστηρίζουν την αντίθετη άποψη, ώστε να μπορέσει να τα συμβουλευθεί ο αναγνώστης προκειμένου να αποκτήσει μια ποιο ποικιλόμορφη πληροφόρηση [101].

Ένα σύστημα παροχής πραγματικών γεγονότων αναφορικά με υπάρχουσες ειδήσεις είναι το FactWatcher[8], το οποίο επιτρέπει τη διασταύρωση ειδήσεων με πραγματικά δεδομένα [102]. Ποικίλα στάδια της διασταύρωσης των ειδήσεων με αληθή γεγονότα παρέχονται από το ClaimBuster[9], το οποίο διαθέτει στους χρήστες του ολοκληρωμένες αναφορές για το βαθμό αληθείας ενός άρθρου με βάση συλλεγμένα δεδομένα, την αναγνώριση παρόμοιων άρθρων με βάση το πόσο αληθή είναι αυτά και την ανάκτηση πληροφοριών που επαληθεύουν την εγκυρότητα των δεδομένων αυτών [103].

Ακόμη, μια υπηρεσία επαλήθευσης της αλήθειας από tweets, η Truthy[10], αναγνωρίζει πολλούς τύπους ψευδών ειδήσεων με βάση ποικίλα χαρακτηριστικά των tweets, τα οποία αναφέρονται σε πολιτικά τεκταινόμενα [104].

4.5. Συγκριτική ανάλυση ερευνητικών προσεγγίσεων πρόληψης

Σε αυτήν την ενότητα παρουσιάζουμε έναν πίνακα συγκέντρωσης και σύγκρισης προσεγγίσεων μέσα από τα επιστημονικά άρθρα που έχουμε μελετήσει. Σκοπός του πίνακα είναι να γίνει σωστά η συγκριτική ανάλυση ερευνητικών προσεγγίσεων πρόληψης. Ο παρακάτω πίνακας περιέχει τα κριτήρια που χρησιμοποιούμε καθώς και τα άρθρα που μελετήθηκαν.

Τα κριτήρια αυτά είναι το είδος λύσης, δηλαδή η μεθοδολογία που χρησιμοποιεί μια ερευνητική εργασία για την αντιμετώπιση της διάδοσης των ψευδών ειδήσεων. Το είδος λύσης μπορεί να είναι είτε αλγοριθμική λύση, δηλαδή ο περιορισμός της παραπληροφόρησης με την χρήση αλγορίθμων, θεωρητικών μοντέλων και μαθηματικής βελτιστοποίησης, είτε η κοινωνική λύση, δηλαδή οι τρόποι πρόληψης παραπληροφόρησης που εστιάζουν στην ενημέρωση, στην εκπαίδευση και στην ενεργή συμμετοχή των χρηστών ή και της κοινωνίας, είτε η τεχνολογική λύση, δηλαδή εργαλεία, σύστημα ή τεχνική υλοποίηση που βοηθάει στην ανίχνευση, τον περιορισμό ή την πρόληψη της παραπληροφόρησης σε online περιβάλλοντα. Επίσης, έχουμε ως κριτήρια την χρήση δεδομένων παραγόμενα από AI, δηλαδή χρησιμοποιούμε AI τεχνολογία για την αντιμετώπιση της παραπληροφόρησης.

---

[7] http://disputefinder.icsi.berkeley.edu/

[8] https://idir.uta.edu/factwatcher/

[9] https://factcheck.uta.edu/

[10] https://osome.iu.edu/

Ένα ακόμα κριτήριο είναι το είδος των χαρακτηριστικών, δηλαδή η περιγραφή των βασικών μεθοδολογικών στοιχείων κάθε ερευνητικής προσέγγισης (μοντέλα, αλγόριθμοι, δεδομένα, θεωρίες, εργαλεία) που τη διαφοροποιούν από άλλες. Υπάρχουν δύο μεγάλες κατηγορίες χαρακτηριστικών: του δικτύου, δηλαδή πως εξαπλώνεται η ψευδής είδηση, και της παρέμβασης, δηλαδή πως μπορούμε να αναχαιτίσουμε την ψευδή είδηση. Ωστόσο για κάθε μία κατηγορία χαρακτηριστικού έχουμε και κάποιες υποκατηγορίες. Αρχικά στα χαρακτηριστικά δικτύου έχουμε τα δομικά χαρακτηριστικά, τα οποία αφορούν τη δομή του γράφου, τα χρονικά χαρακτηριστικά, που έχουν να κάνουν με την ταχύτητα και τον χρόνο και τέλος τα υπολογιστικά που αφορούν τα μαθηματικά μοντέλα και τους αλγορίθμους. Ενώ στα χαρακτηριστικά παρέμβασης έχουμε τα χαρακτηριστικά των χρηστών, δηλαδή αυτά που αφορούν το προφίλ, την αξιοπιστία και τη συμπεριφορά των χρηστών-πηγών, τα κοινωνικά χαρακτηριστικά, τα οποία έχουν να κάνουν με την ψυχολογία της μάζας και τις κοινωνικές αλληλεπιδράσεις και τα χαρακτηριστικά περιεχομένου που αφορούν τη χρήση αληθών πληροφοριών ως αντίμετρο.

Στην συνέχεια έχουμε τον βαθμό/επίπεδο απόδοσης, δηλαδή πόσο αποτελεσματική είναι η παρέμβαση ώστε να περιορίσει την διάδοση της ψευδούς είδησης. Αρχικά έχουμε την χαμηλή απόδοση, η οποία αφορά μικρή επίδραση στην διασπορά μιας ψευδής είδησης (η μείωση διασποράς είναι κάτω του 40%), ακολουθεί η μέτρια απόδοση, όπου η παρέμβαση επηρεάζει την διάδοση (η μείωση διασποράς είναι περίπου 40%-70%) και τέλος έχουμε την υψηλή στην οποία καταφέρνουμε περιορισμό την διάδοσης σε ένα μικρό τμήμα του δικτύου (η μείωση διασποράς είναι 70% και άνω).

Έπειτα, έχουμε την πολυπλοκότητα, δηλαδή τους υπολογιστικούς πόρους (χρόνο και μνήμη) που απαιτείται από έναν αλγόριθμο για να παρθεί η απόφαση. Έχουμε την χαμηλή πολυπλοκότητα, όπου ο αλγόριθμος είναι γραμμικός ως προς το μέγεθος του δικτύου (πχ. O(n)) και βασίζεται σε απλά στατιστικά στοιχεία χρηστών, την μέτρια στην οποία είναι πολυωνυμικός (πχ. O(n^2)) και βασίζεται σε μεθόδους μηχανικής μάθησης και ανάλυσης κοινοτήτων.

Τέλος, έχουμε την υψηλή στην οποία η πολυπλοκότητα των αλγορίθμων είναι αρκετά υψηλή (είναι δύσκολη μαθηματικά η επίλυση (πχ. NP-hard) οπότε βασίζονται σε ευρετικούς (heuristics) αλγορίθμους για την μείωσή της (αλλά πολλές φορές αυτή ξεπερνά την πολυωνυμική).

Επί πρόσθετα σαν κριτήριο ορίζουμε και την κλιμακωσιμότητα, δηλαδή πόσο εύκολο είναι να επεκταθεί μια προσέγγιση αλλά και να αντιμετωπίσει ένα μεγάλο όγκο από ειδήσεις και κόμβους/δίκτυα. Αρχικά έχουμε την χαμηλή στην οποία η προσέγγιση γίνεται πολύ αργά σε μεγάλα δίκτυα λόγω της εκθετικής πολυπλοκότητας και περιλαμβάνει μεθόδους που απαιτούν ανθρώπινη παρέμβαση, την μέτρια στην οποία η προσέγγιση λειτουργεί ικανοποιητικά, αλλά παρουσιάζει καθυστέρηση σε καθολική εφαρμογή και τέλος την υψηλή στην οποία η προσέγγιση βασίζεται σε γραμμικούς αλγορίθμους ή κατανεμημένη υπολογιστική και επεξεργάζεται πάρα πολλές αναρτήσεις.

Ένα ακόμα κριτήριο είναι το κόστος, το οποίο δεν αφορά μόνο τα χρήματα, αλλά κυρίως τους πόρους (ανθρώπινους και τεχνικούς) και τις κοινωνικές επιπτώσεις της εφαρμογής μιας λύσης.

Υπάρχει το χαμηλό κόστος, το οποίο αφορά πλήρως αυτοματοποιημένες αλγοριθμικές λύσεις και απαιτείται μικρή ανάγκη για ανθρώπινη παρέμβαση, το μεσαίο κόστος στο οποίο η προσέγγιση είναι ημι-αυτοματοποιημένη και απαιτείται εκπαίδευση των μοντέλων και συντήρηση των βάσεων δεδομένων και τέλος το υψηλό κόστος όπου απαιτείται χειροκίνητος έλεγχος από ανθρώπους και χρειάζεται ακριβό και εξειδικευμένο υλικό για την επεξεργασία των μεγάλων γράφων.

Τέλος, έχουμε και την τεχνολογία παρέμβασης ως κριτήριο, το οποίο προσδιορίζει το υπόβαθρο που χρησιμοποιείται για την επίλυση του προβλήματος της πρόληψης. Τέτοιες τεχνικές είναι η θεωρία παιγνίων (game theory), γίνεται αναπαράσταση της σύγκρουσης αυτού που διαδίδει το ψέμα και αυτού που προσπαθεί να προωθήσει την αλήθεια, η βελτιστοποίηση γράφων (graph optimization) όπου χρησιμοποιούνται τεχνικές όπως η ελαχιστοποίηση επιρροής (influence minimization) ή η αποκοπή ακμών/κόμβων(link/node removal) για τον περιορισμό της διάδοσης της παραπληροφόρησης, η μηχανική μάθηση/LLMs με χαρακτηριστικά αυτόματης παραγωγής για την διόρθωση απαντήσεων, την θεωρία εμβολιασμού (inoculation theory), δηλαδή την έκθεση του χρήστη σε μικρές ψευδείς ειδήσεις, για την ανάπτυξη κριτικής σκέψης και τέλος ανάθεση στο πλήθος (Crowdsourcing) για την γρήγορη επαλήθευση των δεδομένων της είδησης.

Άρθρα	Είδος Λύσης	Δεδομένα από AI	Είδος Χαρακτηριστικών	Τεχνική Παρέμβασης	Απόδοση	Πολυπλοκότητα	Κλιμακωσιμότητα	Κόστος
[79]	Αλγοριθμική	✓	Δομικά, Υπολογιστικά	Βελτιστοποίηση Γράφων	Υψηλή	Μέτρια	Μέτρια	Μέτριο
[80]	Αλγοριθμική	✓	Δομικά, Χρονικά	Βελτιστοποίηση Γράφων	Υψηλή	Μέτρια	Μέτρια	Μέτριο
[81]	Αλγοριθμική	✓	Δομικά, Υπολογιστικά	Βελτιστοποίηση Γράφων	Υψηλή	Υψηλή	Χαμηλή	Μέτριο
[82]	Αλγοριθμική	✓	Δομικά, Χρονικά	Βελτιστοποίηση Γράφων	Μέτρια	Μέτρια	Μέτρια	Μέτριο
[83]	Αλγοριθμική	✓	Διάδοσης, Δομικά	Βελτιστοποίηση Γράφων	Υψηλή	Υψηλή	Χαμηλή	Μέτριο
[84]	Αλγοριθμική	✓	Χρονικά, Δομικά	Βελτιστοποίηση Γράφων	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[85]	Αλγοριθμική	X	Υπολογιστικά, Δομικά	Βελτιστοποίηση Γράφων	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[86]	Αλγοριθμική	X	Υπολογιστικά	Βελτιστοποίηση Γράφων	Υψηλή	Μέτρια	Μέτρια	Μέτριο
[87]	Κοινωνική	X	Κοινωνικά, Περιεχομένου	Θεωρία Εμβολιασμού	Υψηλή	Χαμηλή	Χαμηλή	Υψηλό
[88]	Κοινωνική	X	Κοινωνικά, Χρήστη	Θεωρία Εμβολιασμού	Υψηλή	Χαμηλή	Χαμηλή	Υψηλό
[89]	Κοινωνική	X	Κοινωνικά, Περιεχομένου	Θεωρία Εμβολιασμού	Υψηλή	Χαμηλή	Χαμηλή	Υψηλό
[90]	Κοινωνική	X	Χρονικά, Κοινωνικά	Βελτιστοποίηση Γράφων	Υψηλή	Χαμηλή	Χαμηλή	Υψηλό
[91]	Κοινωνική	X	Κοινωνικά	Θεωρία Εμβολιασμού	Υψηλή	Χαμηλή	Χαμηλή	Υψηλό
[92]	Κοινωνική	X	Διάδοσης, Περιεχομένου	Θεωρία Παιγνίων	Υψηλή	Χαμηλή	Μέτρια	Μέτριο
[93]	Υβριδική	X	Δομικά, Υπολογιστικά	Θεωρία Παιγνίων	Μέτρια	Υψηλή	Χαμηλή	Υψηλό
[94]	Υβριδική	X	Υπολογιστικά	Θεωρία Παιγνίων	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[95]	Κοινωνική	X	Ψυχολογικά, Κοινωνικά	Θεωρία Εμβολιασμού	Υψηλή	Χαμηλή	Χαμηλή	Υψηλό
[96]	Υβριδική	X	Δομικά, Υπολογιστικά	Ελαχιστοποίηση Επιρροής	Υψηλή	Υψηλή	Μέτρια	Μέτριο
[97]	Υβριδική	X	Υπολογιστικά, Δομικά	Βελτιστοποίηση Γράφων	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[98]	Υβριδική	X	Υπολογιστικά	Θεωρία Παιγνίων	Υψηλή	Υψηλή	Χαμηλή	Υψηλό
[99]	Υβριδική	X	Υπολογιστικά	Θεωρία Παιγνίων	Υψηλή	Υψηλή	Μέτρια	Μέτριο
[100]	Τεχνολογική	X	Περιεχομένου, Χρήστη	Crowdsourcing	Υψηλή	Μέτρια	Υψηλή	Χαμηλό
[101]	Τεχνολογική	X	Περιεχομένου	Μηχανική Μάθηση / LLMs	Μέτρια	Χαμηλή	Υψηλή	Χαμηλό
[102]	Τεχνολογική	X	Περιεχομένου, Χρήστη	Μηχανική Μάθηση / LLMs	Υψηλή	Χαμηλή	Υψηλή	Χαμηλό
[103]	Τεχνολογική	✓	Περιεχομένου, Χρήστη	Μηχανική Μάθηση / LLMs	Μέτρια	Μέτρια	Υψηλή	Μέτριο
[104]	Τεχνολογική	✓	Διάδοσης, Χρήστη	Μηχανική Μάθηση / LLMs	Υψηλή	Μέτρια	Υψηλή	Μέτριο

Εικόνα 4. Συγκεντρωτική ανάλυση άρθρων για αντιμετώπιση ψευδών ειδήσεων.

Με βάση τον πίνακα μπορούμε να εξάγουμε κάποια συμπεράσματα. Αρχικά θα τα χωρίσουμε σε δυο μεγάλες κατηγορίες:

1. Εξαγωγή συμπεράσματος για κάθε κριτήριο ξεχωριστά.
2. Εξαγωγή συμπεράσματος καθολικά.

Ας αναλύσουμε λοιπόν τις δυο αυτές κατηγορίες.

Για την Εξαγωγή συμπεράσματος για κάθε κριτήριο ξεχωριστά έχουμε:

- **Είδος λύσης:** Παρατηρούμε μετάβαση από προσεγγίσεις που αντιμετωπίζουν το πρόβλημα των ψευδών ειδήσεων ως ένα καθαρά μαθηματικό πρόβλημα βελτιστοποίησης γράφων, αγνοώντας τον ανθρωπινό παράγοντα, σε υβριδικά και τεχνολογικά μοντέλα. Τα μοντέλα αυτά, επειδή αντιλαμβάνονται ότι η παραπληροφόρηση είναι ένα κοινωνικό-τεχνικό φαινόμενο και δεν αντιμετωπίζεται μόνο με εξισώσεις, αλλά απαιτεί και την κατανόηση την ανθρώπινης αντιμετώπισης, αποκτούν περισσότερο έδαφος.

- **Χρήση δεδομένων παραγόμενα από AI:** Παρατηρούμε ότι είναι ακόμα σε πρώιμα στάδια, αλλά αποτελεί το μέλλον της πρόληψης. Τα συστήματα εκπαιδεύονται σε σύνθετα σενάρια επιθέσεων, έχοντας την δυνατότητα προετοιμασίας για την αντιμετώπιση των εξελιγμένων bots πριν αυτά εμφανιστούν στα κοινωνικά δίκτυα.

- **Είδη χαρακτηριστικών:** Παρατηρούμε ότι τα δομικά και τα χρονικά χαρακτηριστικά υπερτερούν των γλωσσικών. Έτσι για να σταματήσει την διασπορά μιας ψευδούς είδησης χρειάζεται κάποιος να ξέρει ποιος την διαδίδει και πως συνδέεται με τους άλλους, παρά τι αναφέρει η ίδια η είδηση. Η ανάλυση της τοπολογίας του δικτύου είναι ο καλύτερος τρόπος που έχουμε για την πρόληψη.

- **Απόδοση:** Παρατηρούμε ότι η απόδοση παραμένει σε υψηλά επίπεδα στις προσομοιώσεις, ειδικά όταν χρησιμοποιούμε αλγοριθμικές μεθόδους. Ωστόσο εμφανίζεται κυρίως σε ελεγχόμενα περιβάλλοντα. Το δύσκολο όμως είναι να κρατήσουμε την απόδοση σε υψηλό στάδιο σε πραγματικές συνθήκες θορύβου και μεγάλου όγκου δεδομένων.

- **Πολυπλοκότητα:** Παρατηρούμε ότι οι περισσότερες αλγοριθμικές λύσεις έχουν υψηλή πολυπλοκότητα (NP-hard) και αυτό τις καθιστά αργές για τις παρεμβάσεις. Για αυτό τον λόγο η χρήση άπληστων/ευρετικών αλγορίθμων είναι αναγκαία. Αυτό οδηγεί στην μείωση της ακρίβειας, αλλά κερδίζουμε σε υπολογιστικό χρόνο.

- **Κλιμακωσιμότητα:** Παρατηρούμε την σχέση πολυπλοκότητας και κλιμακωσιμότητας. Από την μία πλευρά οι τεχνολογίες αποδεικνύουν την υψηλή κλιμακωσιμότητα εξυπηρετώντας πάρα πολλούς χρήστες. Αντιθέτως, οι κοινωνικές και αλγοριθμικές λύσεις δυσκολεύονται στην επέκτασή τους σε παγκόσμια κλίμακα, για αυτό τον λόγο λειτουργούν καλύτερα σε περιορισμένα δίκτυα.

- **Κόστος:** Παρατηρούμε ότι οι κοινωνικές προσεγγίσεις είναι κοστοβόρες επειδή χρειάζονται ανθρώπους για να ελέγχουν τα γεγονότα, ενώ στις αυτοματοποιημένες τεχνολογίες είναι πιο χαμηλό. Γι αυτό τον λόγο οι πλατφόρμες κοινωνικής δικτύωσης διαλέγουν τις τεχνολογικές λύσεις ακόμα και αν αυτές δεν είναι αποτελεσματικές.

- **Τεχνική παρέμβασης:** Παρατηρούμε ότι η βελτιστοποίηση γράφων και η θεωρία παιγνίων είναι οι καλύτερες για τον περιορισμό της εξάπλωσης, ενώ η θεωρία εμβολιασμού προστατεύει μακροπρόθεσμα τους χρήστες. Έρευνες έδειξαν ότι ο συνδυασμός αλγοριθμικών τεχνικών για τον εντοπισμό με αλγορίθμους εμβολιασμού για την ενημέρωση αποτελεί μια ολοκληρωμένη προστασία.

Τα ακόλουθα καθολικά συμπεράσματα μπορούν να εξαχθούν:

- **Η μετατόπιση σε υβριδικό:** Στην αρχή οι ερευνητές βασίστηκαν σε προσεγγίσεις που αντιμετωπίζουν το πρόβλημα των ψευδών ειδήσεων ως ένα καθαρά μαθηματικό πρόβλημα βελτιστοποίησης γράφων, αγνοώντας τον ανθρωπινό παράγοντα οπότε αντιμετώπιζαν το δίκτυο ως ένα παθητικό γράφο και τους χρήστες σαν απλούς κόμβους. Η βελτιστοποίηση γράφων είναι μια τέτοια μέθοδος που επιτυγχάνει υψηλή θεωρητική απόδοση αλλά υστερεί στην πράξη επειδή έχουν υψηλή πολυπλοκότητα και δεν μπορεί να υπολογίσει την ανθρώπινη συμπεριφορά. Ωστόσο οι ερευνητές πλέον στρέφονται σε υβριδικές λύσεις, διότι αυτές συνδυάζουν δομική ανάλυση με την κοινωνική ψυχολογία, προσφέροντας μια πιο αποτελεσματική προστασία.

- **Η κλιμακωσιμότητα ως ο κρίσιμος παράγοντας:** Όσο πιο λεπτομερής και πολύπλοκος είναι ένας αλγόριθμος, τόσο πιο δύσκολα λειτουργεί σε περιβάλλοντα μεγάλων δεδομένων. Οι τεχνολογικές λύσεις αποδεικνύονται πιο βιώσιμες για μεγάλες πλατφόρμες, επειδή θυσιάζουν την ακρίβεια για την ταχύτητα και την δυνατότητα διαχείρισης πολλών κόμβων ταυτόχρονα.

- **Ο ρόλος της AI και του κόστους:** Η ενσωμάτωση δεδομένων παραγόμενα από AI και LLMs, μετατοπίζει το τοπίο στην πρόληψη, επιτρέποντας την αυτόματη επαλήθευση των πληροφοριών και μείωση του ανθρώπινου κόστους. Από την άλλη πλευρά, οι κοινωνικές λύσεις εξακολουθούν και παρέχουν βαθύτερη και ποιοτικά καλύτερη ανάλυση, αλλά είναι πολύ κοστοβόρες για αυτό τον λόγο λειτουργούν συμπληρωματικά, ενώ οι αυτοματοποιημένες τεχνολογικές λύσεις είναι το πρώτο στάδιο της άμυνας.

- **Τελική Διαπίστωση:** Παρατηρούμε ότι η αποτελεσματική πρόληψη δεν μπορεί να βασιστεί σε μία μόνο τεχνική, για αυτό χρειαζόμαστε συστήματα που χρησιμοποιούν βελτιστοποίηση γράφων για τον εντοπισμό κρίσιμων σημείων, που εφαρμόζουν μηχανική μάθηση για την αυτόματη παρέμβαση χαμηλού κόστους και ενισχύουν την κριτική σκέψη των χρηστών με τη θεωρία εμβολιασμού έτσι ώστε το δίκτυο να αντέχει στην παραπληροφόρηση.

Τέλος, πρέπει να αναφέρουμε ότι μέσα από τον πίνακα διακρίνουμε τις προσεγγίσεις που πιστεύουμε ότι είναι καλύτερες αλλά και αυτές που είναι χειρότερες. Οι τεχνολογικές λύσεις είναι οι καλύτερες για άμεση εφαρμογή σε κοινωνικά δίκτυα μεγάλης κλίμακας, επειδή συνδυάζουν καλά την κλιμακωσιμότητα και το χαμηλό κόστος. Στην αντίπερα όχθη, οι αλγοριθμικές λύσεις

με NP-hard πολυπλοκότητα δυσκολεύονται πολύ στην πράξη και αυτό γιατί δεν αντέχουν τον ρυθμό που απαιτεί η ταχύτητα διάδοσης της πληροφορίας στο διαδίκτυο. Η θεωρία του εμβολιασμού παραμένει η καλύτερη ποιοτικά μέθοδος για μακροπρόθεσμη αντιμετώπιση, παρά τις δυσκολίες στην κλιμάκωσή της.

4.6. Σύνοψη συμπερασμάτων και συγκριτική αποτίμηση.

Κοιτάζοντας τις διάφορες ερευνητικές προσεγγίσεις για την ανίχνευση και την πρόληψη των ψευδών ειδήσεων, παρατηρούνται κάποιες βασικές διαφορές. Ορισμένες μέθοδοι είναι πιο ώριμες, αξιόπιστες ή πιο εύκολες να εφαρμοστούν στην πράξη, ενώ κάποιες βρίσκονται ακόμα στα αρχικά τους βήματα. Στο κεφάλαιο αυτό αναλύθηκαν δύο μεγάλες κατηγορίες: α) τεχνικές εντοπισμού ψευδών ειδήσεων και β) τεχνικές περιορισμού της εξάπλωσή τους.

Στην ανίχνευση των ψευδών ειδήσεων η έρευνα έχει αναπτυχθεί σε μεγάλο βαθμό. Μεγάλο μέρος των ερευνών ασχολείται κυρίως με αυτοματοποιημένα μοντέλα μηχανικής μάθησης καθώς και βαθιάς μάθησης, εφαρμόζοντας γλωσσικά χαρακτηριστικά, πληροφορίες για τις πηγές, τρόπος διάδοσης των ειδήσεων στα κοινωνικά δίκτυα και πολυτροπικά δεδομένα. Τέτοιου είδους μοντέλα χρησιμοποιούνται κυρίως σε σύνολα δεδομένων και παρουσιάζουν υψηλά επίπεδα ακριβείας. Επίσης οι έρευνες γύρω από την ανίχνευση περιεχομένου παραγόμενα από AI αυξάνονται δραματικά.

Αντιθέτως στην πρόληψη των ψευδών ειδήσεων τα πράγματα δεν είναι το ίδιο εξελιγμένα. Οι έρευνες εστιάζουν κυρίως σε θεωρητικά μοντέλα διάδοσης, παρέμβασης στα social media, ενημερωτικές εκστρατείες. Αυτές οι ιδέες μένουν κυρίως σε θεωρητικό επίπεδο και χωρίς πολλά πειραματικά δεδομένα ή βασίζονται σε προσομοιώσεις και υποθετικά σενάρια, με αποτέλεσμα να δυσκολεύει η σύγκριση της αποτελεσματικότητά τους.

Επιπλέον η επιτυχία της πρόληψης βασίζεται σε κοινωνικούς, ψυχολογικούς και θεσμικούς παράγοντες, άρα είναι δύσκολη η μοντελοποίησή τους.

Συγκρίνοντας τις δύο αυτές κατηγορίες παρατηρούμε ότι οι τεχνικές ανίχνευσης είναι πιο ώριμες και σωστά τεκμηριωμένες., ενώ οι τεχνικές πρόληψης βρίσκονται σε στάδιο ανάπτυξης με έμφαση στη θεωρία και όχι σε καθολικές αξιολογήσεις. Ωστόσο μπορούμε να συμπεράνουμε ότι οι τεχνικές ανίχνευσης από μόνες τους δεν είναι αρκετές, καθώς η αργοπορημένη αναγνώριση μια ψευδούς είδησης δεν αλλάζει τον αντίκτυπό της.

Εν κατακλείδι, με βάση την ανάλυσή μας συμπεραίνουμε ότι χρειαζόμαστε ολοκληρωμένες λύσεις που συνδυάζουν ώριμα συστήματα ανίχνευσης με μηχανισμούς πρόληψης και περιορισμού της διάδοσης. Απαιτείται ενσωμάτωση τεχνολογικών κοινωνικών και επικοινωνιακών παραγόντων. Τέλος, όλα τα παραπάνω μας εισάγουν στην συστηματική μελέτη των ερευνητικών κενών και τις κατευθύνσεις τους στο πεδίο της ανίχνευσης και της πρόληψης.

5. Ερευνητικά κενά

Στις επόμενες ενότητες αναφέρουμε κάποια ερευνητικά κενά, που βρήκαμε μέσα από την μελέτη και την ανάλυση της βιβλιογραφίας, ενώ παράλληλα προτείνουμε κάποιες ερευνητικές κατευθύνσεις για την αντιμετώπισή τους. Δεν παραμένουμε μόνο στην καταγραφή των ελλείψεων, αλλά προχωράμε και σε προτάσεις που αφορούν την τεχνολογική και μεθοδολογική εξέλιξη του πεδίου.

5.1. Ψευδές ειδήσεις παραγμένες από LLMs.

Στις επόμενες ενότητες με βάση τις πρόσφατες εξελίξεις διαφαίνεται πως η ερευνητική κοινότητα επικεντρώνεται στα μεγάλα γλωσσικά μοντέλα (LLMs) για διάφορα στάδια της ανίχνευσης ψευδών ειδήσεων [73]. Επίσης, να τονίσουμε εδώ ότι η ανακρίβεια στην ανίχνευση αποτελεί σοβαρό κίνδυνο για τον εντοπισμό ψευδών ειδήσεων που έχουν παραχθεί από AI [74]. Η έρευνα μας ανέδειξε μικρό αριθμό εργασιών που είναι υπεύθυνες για την ανίχνευση ψευδών ειδήσεων στην εποχή των LLMs. Οι περιορισμοί αυτοί χωρίζονται σε τρεις βασικές κατηγορίες:

1. Χρονικοί, δηλαδή το τεχνολογικό χάσμα: πολλές μέθοδοι βασίζονται σε παλιότερα μοντέλα (πχ. GPT-2), τα οποία δεν μπορούν να εντοπίσουν τα πιο σύνθετα γλωσσικά μοτίβα των νέων γεννητριών (πχ. GPT-4).

2. Μεθοδολογικοί, δηλαδή ελλιπής πρακτική εφαρμογή: πολλές προσεγγίσεις μένουν θεωρητικές, χωρίς κάποια δοκιμή ή υλοποίηση σε πραγματικές συνθήκες. Αυτό δημιουργεί προβλήματα πρακτικότητας και δυσκολεύει τη διεύρυνση της χρήσης τους.

3. Εμπειρικοί, δηλαδή υπάρχουν περιορισμένα δεδομένα: παρατηρήθηκε έλλειψη ενημερωμένων δεδομένων με αρκετά δείγματα ψευδών ειδήσεων που έχουν παραχθεί μέσω LLMs [75].

Στο επίπεδο δημιουργίας δεδομένων εκπαίδευσης, τα LLMs αξιοποιούνται για την δημιουργία πληθώρας συνθετικών δειγμάτων ψευδών και αληθών ειδήσεων, επιτρέποντας την τροφοδότηση των ταξινομητών ανίχνευσης με πολλαπλά μοτίβα ψευδών και πραγματικών/αληθινών ειδήσεων. Ωστόσο το ζητούμενο δεν είναι η δημιουργία μεγάλου όγκου δεδομένων, αλλά η ανάπτυξη πιο αξιόπιστων μεθόδων παραγωγής. Κατ' αυτόν τον τρόπο τα παραγόμενα δείγματα θα πρέπει να αποτυπώνουν με μεγαλύτερη ακρίβεια τις μορφές ψευδών ειδήσεων που συναντώνται τόσο σε περιεχόμενο χρηστών όσο και σε κείμενα που δημιουργούνται από LLMs, ώστε να ενισχύεται η αποτελεσματικότητα της ανίχνευσης στην πράξη.

Παράλληλα, τα LLMs δεν είναι μόνο για παραγωγή κειμένου, αλλά δρουν και ως μοντέλα ανίχνευσης [77]. Αν τα εκπαιδεύσεις (fine–tuning) με πληθώρα συνθετικών και μη δεδομένων, είναι σε θέση να μάθουν να αναγνωρίζουν μοτίβα προς αναγνώριση. Εκτός όμως ότι αυτό είναι κοστοβόρο, είναι επίσης χρονοβόρο, θέλει ισχυρό υπολογιστικά υλικό και προσεκτική προετοιμασία των δεδομένων. Ωστόσο η δημιουργία και εξέταση προτροπών (prompting) είναι πιο φθηνή, αλλά δεν είναι το ίδιο ακριβής και αξιόπιστη στην ανίχνευση.

Για να αποφευχθεί αυτό το αδιέξοδο, οδηγούμαστε στην χρήση των τεχνικών chain of thought (COT). Με αυτές το μοντέλο σκέφτεται βήμα βήμα και αναλύει το περιεχόμενο, εντοπίζοντας αντιφάσεις και ασυνέπειες που ένα απλό prompt θα αγνοούσε. Επίσης επιτρέπουν σε σκεπτόμενα LLMs να ανιχνεύουν με περισσότερη ακρίβεια ψευδείς ειδήσεις με βάση την ισχυρή ικανότητα λογισμού (reasoning) που έχουν. Με αυτό τον τρόπο τα LLMs απαντούν με τεκμηριωμένο τρόπο αν μια είδηση είναι έγκυρη ή όχι και γεφυρώνεται το χάσμα μεταξύ υπολογιστικής απόδοσης και αξιοπιστίας [75, 76, 77].

Επιπλέον, νεότερες τεχνικές που ενσωματώνουν τις δυνατότητες των LLMs με συστήματα ανάκτησης οδηγούν σε νέες μεθόδους, όπως το Retrieval Augmented Generation (RAG), φέροντας κοντά τις γλωσσικές και γνωστικές ικανότητες των LLMs με την εγκυρότητα των πηγών γνώσης [74, 75, 76]. Το Retrieval-Augmented Generation (RAG) είναι μια αρχιτεκτονική που συνδυάζει τις δημιουργικές ικανότητες των LLMs με την δυνατότητα να βρίσκει ακριβείς πληροφορίες από αξιόπιστες πήγες. Δεν στηρίζεται αποκλειστικά στη μνήμη (δηλαδή στην εκπαίδευση του μοντέλου), άλλα έχει μια βιβλιοθήκη που εξετάζει τα γεγονότα αν είναι αληθής ή ψευδή και τα παρουσιάζει μετά.

5.2. Επεξηγηματικότητα και διαφάνεια

Η επεξηγηματικότητα και η διαφάνεια στην ανίχνευση ψευδών ειδήσεων αναφέρονται στην ικανότητα των μοντέλων και των αλγορίθμων να παρέχουν σαφείς και κατανοητούς λόγους για τις αποφάσεις τους, πράγμα που είναι ζωτικής σημασίας για την οικοδόμηση εμπιστοσύνης και τη διασφάλιση της ηθικής χρήσης [5]. Πολλά τρέχοντα μοντέλα, ιδίως εκείνα που βασίζονται στη βαθιά μάθηση, λειτουργούν συχνά ως "μαύρα κουτιά" με ελάχιστη εικόνα του τρόπου με τον οποίο καταλήγουν στα συμπεράσματά τους [46]. Αυτή η αδιαφάνεια μπορεί να οδηγήσει σε σκεπτικισμό και αντίσταση από τους χρήστες, τους υπεύθυνους χάραξης πολιτικής και άλλους ενδιαφερόμενους, αναδεικνύοντας την ανάγκη για πιο διαφανή και ερμηνεύσιμα συστήματα [57].

Μια σημαντική πρόκληση είναι η εγγενής πολυπλοκότητα των προηγμένων μοντέλων μάθησης, και ιδιαιτέρως πλέον των LLMs, η οποία καθιστά δύσκολη την παρακολούθηση της διαδικασίας λήψης αποφάσεων. Η ανάπτυξη τεχνικών που μπορούν να απλοποιήσουν και να διαφωτίσουν αυτές τις διαδικασίες είναι απαραίτητη. Για παράδειγμα, μέθοδοι όπως οι μηχανισμοί προσοχής μπορούν να αναδείξουν σε ποια μέρη των δεδομένων εισόδου επικεντρώθηκε το μοντέλο κατά τη λήψη μιας απόφασης [72, 76].

Η διαφάνεια δεν περιλαμβάνει μόνο τη διασφάλιση ότι τα σύνολα δεδομένων και οι διαδικασίες κατάρτισης που χρησιμοποιούνται για την ανάπτυξη μοντέλων ανίχνευσης ψευδών ειδήσεων είναι ανοικτά και προσβάσιμα για έλεγχο, αλλά και την διασφάλιση της ποιότητάς τους. Αυτή η διαφάνεια βοηθά στον εντοπισμό πιθανών προκαταλήψεων και σφαλμάτων στα δεδομένα που θα μπορούσαν να οδηγήσουν σε άδικα ή ανακριβή αποτελέσματα. Γι’ αυτόν τον λόγο απαιτείται λοιπόν προ επεξεργασία ή και μηχανισμοί καθαρισμού δεδομένων κατά τη διάρκεια της εκπαίδευσης [5, 10, 16, 25, 31].

Να επισημάνουμε εδώ ότι η αντιπροσωπευτικότητα των δεδομένων είναι κρίσιμη διότι, αν ένα μοντέλο εκπαιδεύεται κυρίως σε δεδομένα από μια συγκεκριμένη περιοχή ή γλώσσα, ενδέχεται να μην αποδίδει καλά σε διαφορετικά περιοχές ή γλώσσες [46, 105, 109]. Ως εκ τούτου, η δημιουργία αξιόπιστων μοντέλων εξαρτάται τόσο από το πόσο ανοιχτά και προσβάσιμα είναι όσο και την καθαρότητα, την ποικιλία και την αντιπροσωπευτικότητα των συνόλων δεδομένων.

Μια άλλη πτυχή της διαφάνειας αφορά όχι μόνο την σαφή κοινοποίηση των περιορισμών, αλλά την αξιόπιστη εκτίμηση και παρουσίαση των αβεβαιοτήτων του μοντέλου. Κανένα μοντέλο δεν είναι αλάνθαστο οπότε η απλή αναφορά στους χρήστες για πιθανή εμφάνιση ψευδώς θετικών ή αρνητικών αποτελεσμάτων δεν είναι επαρκής. Η ουσιαστική πρόκληση έγκειται στον ακριβή υπολογισμό και την κατάλληλη επικοινωνία αυτών των αβεβαιοτήτων [5, 17, 25, 31]. Πρόσφατες μελέτες έχουν δείξει ότι η παραπληροφόρηση που παράγεται είτε από χρήστες είτε από LLMs συχνά εμφανίζει χαρακτηριστικά που δυσκολεύουν την εκτίμηση αξιοπιστίας [54, 55, 56, 67, 76, 77]. Αυτό μας δείχνει την ανάγκη για νέες μεθόδους που να υπολογίζουν την αβεβαιότητα με συστηματικό τρόπο και να την αποδίδουν με μορφές κατανοητές στους τελικούς χρήστες, ώστε να μην δημιουργείται ψευδής αίσθηση βεβαιότητας. Τέτοιες προσεγγίσεις συνδέονται άμεσα με την εμπιστοσύνη και τη μείωση της επιρροής της παραπληροφόρησης [114, 117].

Επιπλέον η ενσωμάτωση της επεξηγηματικότητας στο σχεδιασμό συστημάτων ανίχνευσης ψευδών ειδήσεων δεν περιορίζεται μόνο στη βελτίωση της κατανόησης από τους χρήστες, αλλά μπορεί να αποτελέσει και μηχανισμό ανατροφοδότησης για τα ίδια τα μοντέλα. Για παράδειγμα, οι διεπαφές χρήστη που οπτικοποιούν το σκεπτικό πίσω από μια απόφαση διευκολύνουν δημοσιογράφους και ελεγκτές γεγονότων να να κατανοήσουν και να επαληθεύσουν τα αποτελέσματα, ενώ ταυτόχρονα παρέχουν πολύτιμη πληροφόρηση για τη βελτίωση της ακρίβειας του συστήματος [5, 55, 56]. Η αξιοποίηση τεχνικών επεξήγησης που βασίζονται σε γράφους γνώσης (knowledge graphs) και δίκτυα γνώσης [42, 44, 46] προσφέρει πιο στοχευμένες ενδείξεις, δηλαδή που τα μοντέλα αποτυγχάνουν, δημιουργώντας έναν κύκλο συνεχούς προσαρμογής και βελτίωσης. Με αυτόν τον τρόπο, η επεξηγηματικότητα δεν λειτουργεί μόνο ως εργαλείο οικοδόμησης εμπιστοσύνης, αλλά και ως πρακτικός μηχανισμός ενίσχυσης της απόδοσης, ιδιαίτερα σε περιβάλλοντα όπου τα LLMs δημιουργούν νέες προκλήσεις ανίχνευσης [62, 73, 76, 117].

Υπάρχει επίσης ανάγκη διεπιστημονικής συνεργασίας για τη βελτίωση της επεξηγηματικότητας και της διαφάνειας στην ανίχνευση ψευδών ειδήσεων. Η συγκέντρωση εμπειρογνωμόνων από την επιστήμη των υπολογιστών, τη γλωσσολογία, την ψυχολογία και άλλους τομείς μπορεί να βοηθήσει στην ανάπτυξη μοντέλων που κατανοούν και αντικατοπτρίζουν καλύτερα την πολυπλοκότητα της ανθρώπινης επικοινωνίας και της παραπληροφόρησης. Επιπλέον, η συμμετοχή των τελικών χρηστών στο σχεδιασμό και τη δοκιμή αυτών των συστημάτων διασφαλίζει ότι τα εργαλεία που αναπτύσσονται ανταποκρίνονται στις πρακτικές ανάγκες αυτών που θα τα χρησιμοποιούν.

5.3. Ανίχνευση σε πραγματικό χρόνο

Η ανίχνευση ψευδών ειδήσεων σε πραγματικό χρόνο περιλαμβάνει τον εντοπισμό και την επισήμανση ψευδών ή παραπλανητικών πληροφοριών κατά τη δημιουργία και τη διάδοσή τους και όχι αφού έχουν ήδη διαδοθεί ευρέως [22]. Η επίτευξη ανίχνευσης σε πραγματικό χρόνο είναι ζωτικής σημασίας για τον μετριασμό της ταχείας διάδοσης της παραπληροφόρησης και την ελαχιστοποίηση των επιπτώσεών της [49]. Ωστόσο, πρέπει να αντιμετωπιστούν αρκετές σημαντικές προκλήσεις και να υιοθετηθούν κατάλληλες ερευνητικές κατευθύνσεις για να καταστεί αυτό εφικτό [50].

Μια σημαντική πρόκληση είναι η κλιμακωσιμότητα δεδομένου του τεράστιου όγκου περιεχομένου που παράγεται κάθε λεπτό στα μέσα κοινωνικής δικτύωσης και στις ειδησεογραφικές πλατφόρμες. Αυτό απαιτεί την ανάπτυξη αλγορίθμων και συστημάτων που μπορούν να επεξεργάζονται και να αναλύουν αποτελεσματικά ροές δεδομένων μεγάλης κλίμακας σε πραγματικό χρόνο.

Με βάση τα παραπάνω, θα πρέπει να καταγράφουν κάποιες ερευνητικές κατευθύνσεις που μπορούν να αντιμετωπίσουν τις προαναφερόμενες προκλήσεις:

- Μία κρίσιμη ερευνητική κατεύθυνση αποτελεί **ο Real-time εντοπισμός bots και αυτοματοποιημένα προφίλ** δεδομένου του αυξημένου ρόλου των social bots στη διαδικτυακή επικοινωνία και ειδικότερα στη διάδοση παραπληροφόρησης. Περιγράφεται πως τα αυτοματοποιημένα προφίλ μπορούν να μιμούνται ανθρώπινες αλληλεπιδράσεις, να δημιουργούν περιεχόμενο σε μεγάλη κλίμακα και να ενισχύουν την διάδοση ψευδών ειδήσεων. Το κεντρικό ζήτημα είναι ότι, λόγω του όγκου και της ταχύτητας με την οποία δραστηριοποιούνται τα bots, η ανίχνευσή τους δεν μπορεί να βασίζεται αποκλειστικά σε εκ των υστέρων αναλύσεις (ex-post). Αντίθετα, απαιτούνται μηχανισμοί παρακολούθησης σε πραγματικό χρόνο, οι οποίοι αξιοποιούν την συμπεριφορά και τα δικτυακά χαρακτηριστικά των χρηστών προκειμένου να αποτρέπεται η μαζική και η συντονισμένη διάδοση ψευδών πληροφοριών. Με αυτόν τον τρόπο, παρουσιάζονται τα θεμέλια για την ανάπτυξη συστημάτων ανίχνευσης και φιλτραρίσματος της δραστηριότητας των bots (bot activity) σε real-time περιβάλλοντα, ως κρίσιμη στρατηγική για τον περιορισμό της επιρροής τους στο δημόσιο λόγο.[23]
- Διακριτή ερευνητική κατεύθυνση αποτελεί **η κλιμάκωση της ανάλυσης (Scalability) της διάδοσης περιεχόμενου χαμηλής αξιοπιστίας.** Ενώ η προηγούμενη κατεύθυνση εστιάζει στον εντοπισμό των μεμονωμένων bots, εδώ εστιάζουμε στη συνολική δυναμική του δικτύου. Η παραπληροφόρηση δεν εξαπλώνεται με μεμονωμένες αναρτήσεις, αλλά λειτουργεί μέσα από μαζικά συντονισμένα κύματα που απαιτούν την ανάπτυξη κλιμακούμενων συστημάτων ανίχνευσης (scalable detection pipelines). Τα συστήματα ανίχνευσης ολοένα και περισσότερο πρέπει να διαχειρίζονται τεράστιες ποσότητες πληροφοριών από πολλές πηγές. Επίσης, παρακολουθούν πως εξελίσσονται τα φαινόμενα σε ολόκληρη την πλατφόρμα, όχι μόνο σε επίπεδο λογαριασμών [28].

- Μία συμπληρωματική ερευνητική κατεύθυνση επικεντρώνεται στην **αξιοποίηση χαρακτηριστικών αξιοπιστίας και διάδοσης για ανίχνευση σε ροή δεδομένων.** Αυτή η προσέγγιση ξεχωρίζει από κάποιες πρωτοποριακές μελέτες στο άρθρο [49] και μετατοπίζει το ενδιαφέρον από την στατική ανάλυση προς την εξέταση της συμπεριφορικής δυναμικής των πληροφοριών. Η έρευνα συνδυάζει δείκτες, όπως τη γλωσσική ποιότητα των μηνυμάτων και τη δομή των διαδικτυακών αλληλεπιδράσεων, ώστε η αξιοπιστία να εκτιμάται ως μία εξελισσόμενη ιδιότητα. Η ανίχνευση σε πραγματικό χρόνο μας δείχνει πόσο χρειαζόμαστε συστήματα που δεν μένουν απλώς σε ανενεργές αναλύσεις, άλλα εντοπίζουν μοτίβα παραπληροφόρησης την ώρα που δημιουργούνται. Έτσι καταφέρνουμε να έχουμε έγκαιρη παρέμβαση σε μεγάλη κλίμακα.
- Η **ανάπτυξη early detection μοντέλων** σχετίζεται με την πρόκληση της έγκαιρης ανίχνευσης ψευδών ειδήσεων, προτού αυτές αποκτήσουν εκτεταμένη διάδοση. Στο πλαίσιο αυτό εστιάζουμε σε μοντέλα που στηρίζονται στην θεωρία και προσπαθούν να αξιοποιήσουν στοιχεία που εμφανίζονται στα πρώτα στάδια διάδοσης μιας είδησης. Αντί να βασίζονται σε ιστορικά δεδομένα, τα οποία είναι διαθέσιμα μόνο αφού η είδηση έχει ήδη εξαπλωθεί, οι ερευνητές δείχνουν πως τα αρχικά σήματα, όπως μοτίβα αλληλεπίδρασης χρηστών, μας οδηγούν σε στοιχεία για το αν μια είδηση είναι αληθής ή όχι. Έτσι, μπορούμε να φτάσουμε σε μοντέλα που λειτουργούν σε streaming περιβάλλοντα επιτρέποντας την έγκαιρη παρέμβαση πριν τη καθιέρωση της παραπληροφόρηση. Αυτός ο τρόπος προσέγγισης δημιουργεί τις προϋποθέσεις για αλγορίθμους ανίχνευσης ψευδών ειδήσεων σε πραγματικό χρόνο και ταυτόχρονα περιορίζουν την εξάπλωσή τους [72].
- **Χρήση μεγάλων πολυδιάστατων datasets για εκπαίδευση επεκτάσιμων συστημάτων:** Προτείνεται η δημιουργία ενός αποθετηρίου που υπερβαίνει τα απλά δεδομένα και συνδυάζει τρεις κρίσιμες διαστάσεις: το ίδιο το περιεχόμενο των ειδήσεων, το κοινωνικό πλαίσιο που προκύπτει από τις αντιδράσεις των χρηστών και τα χρονικά δεδομένα που αποτυπώνουν την πορεία της διάδοσης. Αυτό το σχήμα επιτρέπει την ρεαλιστική μελέτη της παραπληροφόρησης, καθώς η αξιοπιστία μιας είδησης δεν εξαρτάται αποκλειστικά από το κείμενο αλλά και από τον τρόπο που αυτή διαμοιράζεται και γίνεται αντιληπτή στο κοινωνικό δίκτυο. Η ενσωμάτωση αυτών των παραμέτρων να υποστηρίζει την ανάπτυξη μοντέλων ανίχνευσης, τα οποία θα είναι σε θέση να γενικεύουν καλύτερα σε πραγματικές συνθήκες και ροές δεδομένων μεγάλης κλίμακας. Έτσι, λοιπόν ανοίγει ο δρόμος για μελέτες που δεν περιορίζονται σε απλοποιημένα σενάρια, αλλά ανταποκρίνονται στις πολυπλοκότητες του οικοσυστήματος της παραπληροφόρησης [105].

- Μια άλλη κατεύθυνση είναι η εξισορρόπηση της ακρίβειας και της ταχύτητας. Συχνά υπάρχει ένας συμβιβασμός μεταξύ των δύο, καθώς οι ταχύτεροι αλγόριθμοι ενδέχεται να χάσουν ανεπαίσθητους δείκτες ψευδών ειδήσεων, ενώ οι ακριβέστεροι αλγόριθμοι ενδέχεται να απαιτούν περισσότερο χρόνο επεξεργασίας. Ο σχεδιασμός μοντέλων που επιτυγχάνουν ισορροπία μεταξύ ταχύτητας και ακρίβειας, ενδεχομένως μέσω ιεραρχικών συστημάτων ανίχνευσης ή συστημάτων ανίχνευσης πολλαπλών σταδίων που φιλτράρουν

γρήγορα τις προφανείς περιπτώσεις και στη συνέχεια αναλύουν βαθύτερα τις οριακές περιπτώσεις, είναι ουσιαστικής σημασίας [49, 50].

Ο δυναμικός χαρακτήρας των ψευδών ειδήσεων, οι οποίες μπορούν να εξελίσσονται ταχύτατα με νέες ιστορίες, τροποποιήσεις και παραλλαγές που εμφανίζονται συχνά, περιπλέκει περαιτέρω την ανίχνευση σε πραγματικό χρόνο. Η δημιουργία προσαρμοστικών μοντέλων που μπορούν να ενημερώνονται και να μαθαίνουν από νέα δεδομένα σε πραγματικό χρόνο, ώστε να παραμένουν αποτελεσματικά απέναντι στις αναδυόμενες τάσεις παραπληροφόρησης, είναι απαραίτητη [28, 72]. Επιπλέον, οι ψευδείς ειδήσεις συχνά περιλαμβάνουν κείμενο, εικόνες, βίντεο και συνδυασμούς αυτών, γεγονός που καθιστά δύσκολη την ανάλυσή τους μόνο με παραδοσιακές μεθόδους που βασίζονται στο κείμενο. Η ανάπτυξη πολυτροπικών συστημάτων ανίχνευσης που μπορούν να αναλύουν ταυτόχρονα διάφορους τύπους περιεχομένου και τις μεταξύ τους σχέσεις είναι ζωτικής σημασίας [47, 48].

5.4. Πολυγλωσσικότητα

Αν και στην παρούσα επισκόπηση δεν εξετάσαμε πολυγλωσσικά δεδομένα, ωστόσο υπήρχε λόγος γύρω από αυτή την επιλογή. Κυριότερα, η εξέταση πολυγλωσσικών δεδομένων σχετίζεται με την ύπαρξη αρκετών δεδομένων για τη συγκεκριμένη γλώσσα εξ αρχής [53, 55]. Το γεγονός αυτό καθιστά αναγκαία τη δημιουργία και την επιμέλεια πολύγλωσσων συνόλων δεδομένων που καλύπτουν διαφορετικές περιοχές, πολιτισμούς και περιβάλλοντα [69, 70]. Κάθε γλώσσα συνοδεύεται από μοναδικό συντακτικό, σημασιολογία, ιδιωματισμούς και πολιτισμικές αναφορές, γεγονός που απαιτεί την ανάπτυξη μοντέλων που μπορούν να κατανοήσουν και να επεξεργαστούν με ακρίβεια αυτές τις αποχρώσεις [59, 60]. Τα μεταφραστικά σφάλματα που εισάγονται από τα εργαλεία αυτόματης μετάφρασης μπορούν επίσης να αποτελέσουν πρόβλημα, οδηγώντας ενδεχομένως σε εσφαλμένες ταξινομήσεις.

Εδώ να τονίσουμε ότι η αξιοποίηση μεθόδων μηχανικής μάθησης και κυρίως των πρόσφατων γλωσσικών μοντέλων μεγάλης κλίμακας (LLMs) αποδεικνύεται κρίσιμη, διότι έχουν την δυνατότητα επεξεργασίας πολλών γλωσσών και αξιοποιούν κοινές σημασιολογικές αναπαραστάσεις. Επίσης, τεχνικές όπως η επαύξηση δεδομένων(data augmentation) [69 70] ενισχύσουν την εκπαίδευση σε γλώσσες με περιορισμένους πόρους, βελτιώνοντας την ακρίβεια και την ανθεκτικότητα των συστημάτων. Αυτό αναδεικνύει την ανάγκη βελτίωσης της μεταφραστικής ακρίβειας για εργασίες ανίχνευσης ψευδών ειδήσεων ή ανάπτυξης μοντέλων που λειτουργούν εγγενώς σε πολλές γλώσσες χωρίς να βασίζονται στη μετάφραση [23, 33, 78].

Ένα άλλο σημαντικό ζήτημα είναι ότι τα μοντέλα που εκπαιδεύονται σε μια γλώσσα συχνά δεν αποδίδουν καλά σε μια άλλη γλώσσα λόγω γλωσσικών διαφορών. Ως εκ τούτου, η διερεύνηση τεχνικών μάθησης μεταφοράς (transfer learning) που επιτρέπουν στα μοντέλα που εκπαιδεύονται σε μια γλώσσα να προσαρμόζονται αποτελεσματικά σε άλλες γλώσσες είναι ζωτικής σημασίας [63, 67].

Τεχνικές όπως η επαύξηση δεδομένων, όπως η οπίσθια μετάφραση (back-translation) και η αντικατάσταση συνωνύμων, ενισχύουν σε μεγάλο βαθμό την ανθεκτικότητα των μοντέλων. Με

αυτές τις τεχνικές δημιουργούνται συνθετικά παραδείγματα που εμπλουτίζουν τα περιορισμένα σύνολα δεδομένων. Παράλληλα η χρήση των παραγωγικών ανταγωνιστικών δικτύων (Generative Adversarial Networks-GANs) είναι μία πιο υποσχόμενη λύση, καθώς μας αφήνουν την δημιουργία ρεαλιστικών, συνθετικών δειγμάτων ψευδών ειδήσεων.

Τα GANs βοηθούν στην ανισορροπία των δεδομένων και εκπαιδεύουν τους ταξινομητές σε ένα ευρύτερο φάσμα γλωσσικών παραλλαγών και μοτίβων παραπληροφόρησης. Αυτό δυναμώνει τη γενικευμένη ικανότητά τους.

Επιπλέον, πολλές γλώσσες, ιδίως αυτές που ομιλούνται από μικρότερους πληθυσμούς, δεν διαθέτουν τους υπολογιστικούς πόρους και την ερευνητική εστίαση που απαιτούνται για την αποτελεσματική ανίχνευση ψευδών ειδήσεων [33]. Η διάθεση πόρων και η ανάπτυξη κλιμακούμενων μεθόδων που μπορούν να εφαρμοστούν σε γλώσσες με χαμηλούς πόρους είναι ουσιαστικής σημασίας [69, 70].

Η έρευνα στην πολύγλωσση ανίχνευση ψευδών ειδήσεων περιλαμβάνει διάφορες βασικές κατευθύνσεις. Η δημιουργία ολοκληρωμένων, υψηλής ποιότητας συνόλων δεδομένων σε πολλές γλώσσες με ποικίλο περιεχόμενο που αντιπροσωπεύει διαφορετικά πολιτισμικά πλαίσια είναι θεμελιώδους σημασίας [53, 55]. Η χρήση διαγλωσσικών embeddings μπορεί να αντιστοιχεί λέξεις από διαφορετικές γλώσσες σε έναν κοινό σημασιολογικό χώρο, βελτιώνοντας την κατανόηση και την ανίχνευση σε όλες τις γλώσσες [59, 60]. Τέλος, η δημιουργία γλωσσικών μοντέλων ικανών να επεξεργάζονται και να κατανοούν κείμενα ανεξάρτητα από τη γλώσσα, ενδεχομένως μέσω τεχνικών μετα-μάθησης (meta learning) είναι μια άλλη πολλά υποσχόμενη προσέγγιση [63, 67].

5.5. Περιορισμοί των νευρωνικών δικτύων γράφων σε μεγάλα/δυναμικά κοινωνικά δίκτυα

Τα νευρωνικά δίκτυα γράφων (Graph Neural Networks – GNNs) προσφέρουν μια πολλά υποσχόμενη προσέγγιση για την καταπολέμηση της διάδοσης ψευδών ειδήσεων στα κοινωνικά δίκτυα, αξιοποιώντας την εγγενή δομή γραφημάτων αυτών των πλατφορμών [57].

Τα κοινωνικά δίκτυα μπορούν να αναπαρασταθούν ως γράφοι, όπου οι κόμβοι αντιπροσωπεύουν τους χρήστες και οι ακμές τις συνδέσεις ή τις αλληλεπιδράσεις μεταξύ τους 78]. Αυτή η αναπαράσταση με βάση τον γράφο επιτρέπει στα GNNs να αναλύουν και να επεξεργάζονται αποτελεσματικά τη δυναμική της διάδοσης των πληροφοριών, κάτι που στην πρότερη βιβλιογραφία αντιμετωπίζεται με κλασικούς αλγορίθμους γράφων, οι οποίοι όμως απαιτούν υπολογιστικά πολύπλοκες λύσεις, ενδεχομένως αδύνατες σε πραγματικό χρόνο ή όσο ο κοινωνικός γράφος μεγαλώνει [79]. Η ευελιξία των GNNs είναι ικανή να προσφέρει λύσεις σε πολλαπλά επίπεδα όσον αφορά τον εντοπισμό και την αντιμετώπιση χρηστών που συμβάλλουν στη διάδοση [80].

Τα GNNs μπορούν να χρησιμοποιηθούν για τον εντοπισμό κόμβων (χρηστών) που αποτελούν πηγές ψευδών ειδήσεων. Με την εξέταση της δομής του δικτύου και του περιεχομένου που μοιράζεται κάθε κόμβος, τα GNNs μπορούν να μάθουν να διακρίνουν μεταξύ αξιόπιστων και αναξιόπιστων πηγών.

Οι σχεσιακές πληροφορίες σε έναν γράφο βοηθούν τα GNNs να κατανοήσουν τον τρόπο με τον οποίο οι πληροφορίες ρέουν μέσω του δικτύου, καθιστώντας δυνατό τον εντοπισμό ανωμαλιών ή μοτίβων που είναι τυπικά για τη διάδοση ψευδών ειδήσεων [78]. Σε επέκταση αυτού, τα GNNs μπορούν να μοντελοποιήσουν τα μονοπάτια διάδοσης της πληροφορίας στο δίκτυο. Παρακολουθώντας τον τρόπο με τον οποίο μια είδηση διαδίδεται από τον ένα χρήστη στον άλλο, τα GNNs μπορούν να εντοπίσουν χαρακτηριστικά μοτίβα διάδοσης ψευδών ειδήσεων, όπως η ταχεία διάδοση μέσω νεοσύστατων λογαριασμών ή ομάδων bots [84].

Η κατανόηση αυτών των διαδρομών διάδοσης βοηθά στην έγκαιρη ανίχνευση ψευδών ειδήσεων και την σχετική παρέμβαση για τον εμποδισμό διάδοσής τους [78, 79, 85, 86].

Επιπλέον, τα GNNs μπορούν να αναλύουν τη συμπεριφορά των χρηστών εντός του δικτύου. Αυτό περιλαμβάνει την κατανόηση του τρόπου με τον οποίο οι διάφοροι χρήστες αλληλεπιδρούν με το περιεχόμενο και μεταξύ τους. Τα GNNs μπορούν να εντοπίσουν τους χρήστες που μοιράζονται συχνά ανεπιβεβαίωτες ή ψευδείς πληροφορίες και εκείνους που λειτουργούν ως κόμβοι για τη διάδοση πληροφοριών. Με τη σκιαγράφηση αυτών των συμπεριφορών, τα GNNs μπορούν να βοηθήσουν στη δημιουργία πιο στοχευμένων στρατηγικών για την πρόληψη της διάδοσης ψευδών ειδήσεων [78, 79].

Μία ακόμη δυνατότητα των GNNs είναι ότι υπερέχουν στον εντοπισμό κοινοτήτων μέσα σε γραφήματα. Οι ψεύτικες ειδήσεις συχνά εξαπλώνονται ταχύτερα σε ορισμένες κοινότητες ή θαλάμους απήχησης. Ανιχνεύοντας αυτές τις κοινότητες, τα GNNs μπορούν να βοηθήσουν στην κατανόηση και την πρόβλεψη της εξάπλωσης των ψευδών ειδήσεων σε συγκεκριμένες ομάδες. Στη συνέχεια, μπορούν να αναπτυχθούν στοχευμένες παρεμβάσεις σε αυτές τις κοινότητες για τον περιορισμό της εξάπλωσης της παραπληροφόρησης. Ακόμη, βάσει της τρέχουσας κατάστασης ενός δικτύου, τα GNNs μπορούν να χρησιμοποιηθούν για την πρόβλεψη της μελλοντικής διάδοσης των πληροφοριών. Με την κατανόηση των πιθανών διαδρομών που θα ακολουθήσουν οι πληροφορίες, μπορούν να εφαρμοστούν προληπτικά μέτρα για να σταματήσουν οι ψευδείς ειδήσεις πριν φτάσουν σε μεγάλο κοινό. Τα μοντέλα πρόβλεψης μπορούν να εντοπίσουν κόμβους υψηλού κινδύνου που είναι πιθανό να γίνουν μεγάλοι διακινητές ψευδών ειδήσεων [82, 83, 84, 85, 86].

Ο συνδυασμός των GNNs με τεχνικές επεξεργασίας φυσικής γλώσσας (NLP) μπορεί να οδηγήσει στην ανάλυση του περιεχομένου των μηνυμάτων των χρηστών, δίνοντας περιεχόμενο στις συνομιλίες που καλείται να επεξεργαστεί ένα νευρωνικό δίκτυο φυσικής γλώσσας [57]. Με την ενσωμάτωση των χαρακτηριστικών περιεχομένου με τα χαρακτηριστικά δικτύου, τα GNNs μπορούν να παρέχουν μια πιο ολοκληρωμένη ανάλυση. Για παράδειγμα, ένα GNN θα μπορούσε να βοηθήσει στον εντοπισμό ψεύτικων ειδήσεων αναγνωρίζοντας γλωσσικά μοτίβα που είναι κοινά σε παραπλανητικό περιεχόμενο, λαμβάνοντας παράλληλα υπόψη τον τρόπο με τον οποίο το περιεχόμενο διαμοιράζεται εντός του δικτύου [78].

6. Σύνολα Δεδομένων

Το κεφάλαιο αυτό έχει διττό σκοπό. Πρώτον, να αναδείψει όλα τα σύνολα δεδομένων (datasets) σχετικά με ψευδείς ειδήσεις. Δεύτερον, να τα συγκρίνει. Τα σύνολα δεδομένων προς ανάλυση εντός του κεφαλαίου παρέχουν ποικίλους πόρους για τη μελέτη διαφορετικών πτυχών των ψευδών ειδήσεων, συμπεριλαμβανομένου του περιεχομένου κειμένου, του κοινωνικού πλαισίου και των αλληλεπιδράσεων των χρηστών.

6.1 Κατηγορίες συνόλων δεδομένων

6.1.1. LIAR dataset

Το σύνολο δεδομένων LIAR περιέχει 12,8 χιλιάδες σύντομες δηλώσεις με χειροκίνητη επισήμανση από διάφορες πηγές, συμπεριλαμβανομένων ειδησεογραφικών άρθρων, τηλεοπτικών ειδήσεων, ομιλιών, αναρτήσεων στα μέσα κοινωνικής δικτύωσης και συνεντεύξεων. Κάθε δήλωση επισημαίνεται με μία από τις έξι αξιολογήσεις ειλικρίνειας, που κυμαίνονται από “Pants on Fire” (απολύτως αναληθές) ως "αλήθεια". Στο άρθρο αυτό με την έκφραση “Pants on Fire” (απολύτως αναληθές) χαρακτηρίζονται όλες οι δηλώσεις που είναι ψευδείς και αναληθείς. [53]

6.1.2. FakeNewsNet dataset

Το FakeNewsNet είναι ένα ολοκληρωμένο σύνολο δεδομένων που περιέχει περίπου 23,2 χιλιάδες άρθρα ειδήσεων για την έρευνα ανίχνευσης ψευδών ειδήσεων. Περιλαμβάνει περιεχόμενο ειδήσεων, κοινωνικό πλαίσιο και χωροχρονικές πληροφορίες για ψεύτικα και πραγματικά άρθρα ειδήσεων. Συγκεντρώνει δεδομένα από δύο δικτυακούς τόπους ελέγχου γεγονότων: PolitiFact και GossipCop [105].

6.1.3. BuzzFeed dataset

Αυτό το σύνολο δεδομένων περιέχει περίπου 1,6 χιλιάδες άρθρα ειδήσεων από εννέα ειδησεογραφικούς ιστότοπους που ταξινομήθηκαν χειροκίνητα από δημοσιογράφους του BuzzFeed. Τα άρθρα ταξινομούνται ως "ως επί το πλείστον αληθή", "ως επί το πλείστον ψευδή" ή "μείγμα αληθούς και ψευδούς". [106]

6.1.4. BS Detector

Το σύνολο δεδομένων BS Detector αφορά περίπου 4 χιλιάδες domains και χρησιμοποιείται για τον εντοπισμό ψευδών ειδήσεων. Περιλαμβάνει διευθύνσεις URL που χαρακτηρίζονται ως ψευδείς, σατιρικές, ακραία προκατειλημμένες, θεωρίες συνωμοσίας, φήμες, ειδήσεις που χρηματοδοτούνται από το κράτος και επιστημονικά σκουπίδια. Τα δεδομένα δεν αποτελούνται από συνδέσμους προς επιμέρους άρθρα αλλά από λίστα domains. Δηλαδή, κάθε καταχώριση αντιστοιχεί σε ολόκληρο ιστότοπο και όχι σε συγκεκριμένη είδηση μέσα σε έναν ιστότοπο. Άρα, λοιπόν, αναφέρονται σε sites στο σύνολό τους και όχι σε επιμέρους άρθρα. Επίσης, να τονιστεί ότι το συγκεκριμένο σύνολο δεδομένων αναφέρουμε ότι παράγεται αυτόματα, καθώς η συλλογή των δεδομένων και η απόδοση των ετικετών γίνονται με αυτοματοποιημένες μεθόδους και όχι χειροκίνητα από ανθρώπους [107].

6.1.5. PHEME dataset

Το σύνολο δεδομένων PHEME αποτελείται από συνομιλίες Twitter πλούσιες σε φήμες, παρέχοντας έναν πλούσιο πόρο για τη μελέτη της εξάπλωσης και του αντίκτυπου των φημών και των ψευδών ειδήσεων στα μέσα κοινωνικής δικτύωσης. Περιέχει περίπου 5,8 χιλιάδες tweets, τα οποία επισημαίνονται σύμφωνα με δυαδική κατηγοριοποίηση ως εξής: 0 για όχι φήμη (non-rumor) και 1 για φήμη (rumor). Το συγκεκριμένο dataset δεν περιλαμβάνει βαθμό βεβαιότητας της φήμης, αλλά βασίζεται σε δυαδική επισήμανση [108].

6.1.6. ISOT dataset

Το σύνολο δεδομένων ISOT συλλέχθηκε από πηγές του πραγματικού κόσμου και περιέχει επισημασμένα ψευδή και πραγματικά άρθρα ειδήσεων. Περιλαμβάνει περίπου 44,9 χιλιάδες άρθρα ειδήσεων. Έχει σχεδιαστεί για να διευκολύνει την έρευνα σχετικά με την ανίχνευση ψευδών ειδήσεων και περιλαμβάνει άρθρα από διάφορους τομείς. [109]

6.1.7. Fake News Corpus dataset

Το σύνολο δεδομένων Fake News Corpus dataset είναι ένα από τα πρώτα δημοσίως διαθέσιμα σύνολα δεδομένων για ψευδείς ειδήσεις. Συλλέχθηκαν ειδησεογραφικά άρθρα από διάφορες διαδικτυακές πηγές και κάθε άρθρο χαρακτηρίζεται ως ψευδές ή πραγματικό άρθρο ειδήσεων. Χρησιμοποιείται ως σημείο αναφοράς (benchmark**)** για μεθόδους μηχανικής μάθησης και NLP στην ανίχνευση ψευδών ειδήσεων. Ο όγκος του είναι περίπου 3 χιλιάδες άρθρα [55].

6.1.8. Hyperpartisan News dataset

Το σύνολο δεδομένων Hyperpartisan News Dataset είναι ένα σύνολο δεδομένων που μελετάει την υπερκομματική (hyperpartisan) αρθρογραφία και τη σχέση της με τις ψευδείς ειδήσεις. Συλλέγει τα άρθρα από διάφορες διαδικτυακές ειδησεογραφικές πηγές και ο όγκος του είναι περίπου 1,6 χιλιάδες άρθρα. Χρησιμοποιήθηκε για την διερεύνηση αν το "ύφος γραφής" (writing style) μπορεί να διακρίνει fake/hyperpartisan από mainstream ειδήσεις [56].

6.2 Συγκριτική ανάλυση συνόλων δεδομένων

Σε αυτήν την ενότητα παρουσιάζουμε έναν πίνακα συγκέντρωσης και σύγκρισης συνόλων δεδομένων με βάση τα επιστημονικά άρθρα που έχουμε μελετήσει. Σκοπός του πίνακα είναι να γίνει σωστά η συγκριτική ανάλυση των διαθέσιμων συνόλων δεδομένων. Ο παρακάτω πίνακας περιέχει τα κριτήρια που χρησιμοποιούμε καθώς και τα σύνολα δεδομένων που μελετήθηκαν (μαζί με τις παραπομπές τους).

Τα κριτήρια αυτά είναι ο όγκος, δηλαδή πόσο μεγάλο είναι το σύνολο δεδομένων. Αρχικά, έχουμε το μικρής κλίμακας που είναι κάτω από 5000 δείγματα, μετά το μεσαίας που είναι από 5000 έως 50000 δείγματα και τέλος έχουμε τον υψηλής κλίμακας που είναι πάνω από 100000 δείγματα. Επίσης, έχουμε την κατηγορία ανίχνευσης, δηλαδή ποιο χαρακτηριστικό του συνόλου δεδομένων αξιοποιείται ως κύριο κριτήριο για την ανίχνευση.

Ένα άλλο κριτήριο είναι το εύρος κατάταξης, δηλαδή το πόσες και ποιες κατηγορίες χρησιμοποιούνται για να κατατάξουμε το περιεχόμενο. Οι κατηγορίες αυτές είναι η δυαδική κατάταξη, δηλαδή η κατηγοριοποίηση παίρνει τη πιο απλή μορφή (πχ. αληθές ή ψευδές), η διαβαθμισμένη κατάταξη, όπου εδώ το εύρος έχει πιο περίπλοκη μορφή (πχ. κυρίως αληθές, παραπλανητικό, εν μέρει αληθές κλπ.) και η ποιοτική κατάταξη όπου εστιάζουμε στο είδος και τον σκοπό (πχ. σάτιρα, προπαγάνδα κλπ). Έχουμε ακόμη τον τρόπο κατάταξης, δηλαδή ποια είναι η μεθοδολογία με την οποία ταξινομήθηκαν τα σύνολα δεδομένων.

Τέλος, ένα ακόμη κριτήριο είναι το είδος της πηγής από όπου αντλούμαι πληροφορίες, δηλαδή η βασική κατηγορία πηγής από την οποία συγκεντρώθηκαν τα σύνολα δεδομένων.

Πίνακας 2. Συγκριτική ανάλυση συνόλων δεδομένων.

Σύνολο Δεδομένων	**Όγκος**	**Κατηγορία ανίχνευσης**	**Εύρος κατάταξής**	**Τρόπος κατάταξης**	**Είδος πηγής άντλησης πληροφορίας**
LIAR Dataset [53]	Μεσαίος	Με βάση το κείμενο	Διαβαθμισμένη κατάταξη 6 κατηγορίες	Χειρωνακτικός	Fact-checking
Fake News Corpus dataset [55]	Μικρός	Με βάση το κείμενο	Δυαδική κατάταξη 2 κατηγορίες	Χειρωνακτικός	Ειδησεογραφικοί
Hyperpartisan News Dataset [56]	Μικρός	Με βάση το κείμενο	Δυαδική κατάταξη 2 κατηγορίες	Χειρωνακτικός	Ειδησεογραφικοί
FakeNewsNet dataset [105]	Μεσαίος	Με βάση δεδομένα κοινωνικών δικτύων	Δυαδική κατάταξη 2 κατηγορίες	Χειρωνακτικός	Fact-checking
BuzzFeed dataset [106]	Μικρός	Με βάση διαδικτυακές πηγές / domains	Διαβαθμισμένη κατάταξη 3 κατηγορίες	Χειρωνακτικός	Ειδησεογραφικοί

BS Detector [107]	Μικρός	Με βάση διαδικτυακές πηγές / domains	Διαβαθμισμένη και ποιοτική κατάταξη 7 κατηγορίες	Αυτόματος	Λίστες domains
PHEME dataset [108]	Μεσαίος	Με βάση δεδομένα κοινωνικών δικτύων	Δυαδική κατάταξη 2 κατηγορίες	Χειρωνακτικός	Κοινωνικά δίκτυα
ISOT dataset [109]	Μεσαίος	Με βάση το κείμενο	Δυαδική κατάταξη 2 κατηγορίες	Χειρωνακτικός	Ειδησεογραφικοί

Με βάση τον πίνακα μπορούμε να εξάγουμε κάποια συμπεράσματα. Αρχικά θα τα χωρίσουμε σε δυο μεγάλες κατηγορίες:

1. Εξαγωγή συμπεράσματος για κάθε κριτήριο ξεχωριστά
2. Εξαγωγή συμπεράσματος καθολικά.

Ας αναλύσουμε λοιπόν τις δυο αυτές κατηγορίες.

Για την εξαγωγή συμπεράσματος για κάθε κριτήριο ξεχωριστά έχουμε:

- **Όγκος:** Παρατηρήθηκε σαφής πόλωση μεταξύ ποιότητας και ποσότητας. Μικρά σε όγκο σύνολα δεδομένων χρησιμοποιούνται για την αξιολόγηση της ακρίβειας συγκεκριμένων αλγορίθμων σε ελεγχόμενα περιβάλλοντα.

- **Εύρος κατάταξης:** Παρατηρήθηκε ότι για την σωστή αντιμετώπιση του προβλήματος, τα μοντέλα πρέπει να μάθουν να ξεχωρίζουν τα ποιοτικά χαρακτηριστικά, όπως η σάτιρα ή η προπαγάνδα και όχι μόνο να επαληθεύουν τα γεγονότα.

- **Κατηγορία Ανίχνευσης:** Παρατηρούμε ότι το κείμενο παραμένει ως η κύρια πηγή πληροφορίας και ότι τα περισσότερα σύνολα εστιάζουν στο περιεχόμενο. Ωστόσο η κατανόηση του τρόπου διάδοσης είναι εξίσου σημαντική για τον εντοπισμό κακόβουλου λογισμικού περιεχομένου σε πραγματικό χρόνο.

- **Τρόπος κατάταξης:** Παρατηρήθηκε μια ισορροπία ανάμεσα στην κλιμακωσιμότητα και την εγκυρότητα. Οι ειδικοί ελέγχουν χειροκίνητα τα δεδομένα εξασφαλίζοντας πραγματική εγκυρότητα, ωστόσο υπάρχει το πρόβλημα του μεγάλου όγκου. Στην αντίπερα όχθη, αν χρησιμοποιήσουν αυτοματοποιημένες λίστες δύναται η δυνατότητα δημιουργίας

τεράστιων βάσεων δεδομένων, αλλά το σύστημα ορίζει τους ιστότοπους ως αναξιόπιστους διότι δεν εξετάζει τα άρθρα ξεχωριστά.

- **Είδος πηγής:** Παρατηρούμε ότι μοντέλα που έχουν εκπαιδευτεί σε πολιτικές δηλώσεις δυσκολεύονται στο να αναγνωρίζουν ψευδές ειδήσεις. Με την μεγάλη ποικιλία σε πηγές είναι αναγκαία η χρήση πολλαπλών συνόλων δεδομένων για την δημιουργία ενός καθολικού συστήματος ανίχνευσης.

Για την εξαγωγή συμπεράσματος καθολικά έχουμε:

- **Η ισορροπία μεταξύ όγκου και βάθους:** Υπάρχει μια σχέση ανάμεσα στην ποσότητα και την ποιότητα των δεδομένων. Τα δεδομένα δίνουν δύναμη στα σύγχρονα νευρωνικά δίκτυα, αλλά το πρόβλημα της δυαδικής κατάταξης δεν αποδίδει τις γκρίζες ζώνες της πραγματικότητας, ενώ τα σύνολα δεδομένων με διαβαθμισμένη κατάταξη εξηγούν καλύτερα τις αποφάσεις τους.

- **Η εξάρτηση από τον τομέα:** Η ανάλυση μας δείχνει ότι δεν υπάρχει μόνο ένα σύνολο δεδομένων που καλύπτει τα πάντα. Η αποτελεσματικότητα ενός συστήματος ανίχνευσης εξαρτάται άμεσα από το πόσο σχετικό είναι το εκπαιδευτικό του σύνολο με το πεδίο που χρησιμοποιείς.

- **Η στροφή προς το κοινωνικό πλαίσιο:** Η ανάλυση δεν μένει πλέον μόνο στο κείμενο. Ολοένα και περισσότερο εντάσσονται τα κοινωνικά συμφραζόμενα, δηλαδή ποιος, τι, πως διαδίδει την είδηση και τι μεταδεδομένα υπάρχουν για τους χρήστες. Τα σύνολα δεδομένων που έχουν τέτοια μεταδεδομένα οδηγούν σε καλύτερη απόδοση, διότι επιτρέπουν την κατανόηση πως αλλάζει και διαδίδεται μια ψευδής είδηση.

- **Η ανάγκη για υβριδικές προσεγγίσεις:** Το βασικό συμπέρασμα είναι πως η έρευνα εξελίσσεται προς τα υβριδικά μοντέλα. Πρέπει να δουλεύουμε με πολλά και διάφορα σύνολα δεδομένων, αν θέλουμε να αποφύγουμε την υπερπροσαρμογή και να μπορέσουμε να εγγυηθούμε ότι οι αλγόριθμοι μαθαίνουν να ανταποκρίνονται σε διάφορες γλώσσες, πλατφόρμες και είδη παραπληροφόρησης. Με λίγα λόγια δεν αρκεί ένα σύνολο δεδομένων και μία προσέγγιση για να πετύχεις πραγματικά καλό αποτέλεσμα, χρειάζεται ο συνδυασμός.

7. Συμπεράσματα και επίλογος

7.1 Συνεισφορά και συμπεράσματα από την εργασία

Στην παρούσα διπλωματική εργασία, ασχοληθήκαμε με το φαινόμενο των ψευδών ειδήσεων, καλύπτοντας κενά που δημιουργήθηκαν από παλιότερες ανασκοπήσεις. Δεν εργαστήκαμε μόνο από την τεχνολογική πλευρά, αλλά εμβαθύναμε στο κοινωνικό και επικοινωνιακό τμήμα. Εφόσον κάποιος μελετήσει την βιβλιογραφία, παρατηρεί πόσο πολύπλοκες είναι στις μέρες μας οι μέθοδοι παραπληροφόρησης, καθώς και την μεγάλη ανάγκη για εξίσου πολύπλοκα και εξεζητημένα εργαλεία για την αντιμετώπισής της. Με την κατηγοριοποίηση των μεθόδων ανίχνευσης και πρόληψης, καθώς επίσης και την ανάλυση σύγχρονων εργαλείων, όπως είναι τα μεγάλα γλωσσικά μοντέλα (LLMs), αλλά και τη συγκριτική αξιολόγηση των ερευνητικών άρθρων και βάσεων δεδομένων, η εργασία μας έχει ως στόχο να γίνει σημείο αναφοράς. Σε αυτό συντελεί επίσης το γεγονός πως επισημαίνονται επισημαίνουμε ερευνητικά κενά μαζί με κατευθύνσεις αντιμετώπισής τους που οδηγούν στην ανάπτυξη νέων και πιο εξειδικευμένων μεθόδων αντιμετώπισης της παραπληροφόρησης.

Μελετώντας καταλήξαμε στα κυριότερα συμπεράσματα, τα οποία και αναφέρουμε παρακάτω.

1. Η υπεροχή των υβριδικών μοντέλων. Ο συνδυασμός βάσεις γνώσης με βαθιά νευρωνικά δίκτυα (πχ. Transformers και GANs) αποδίδουν τα μέγιστα για να έχουμε επίπεδο ακρίβειας πάνω από 90%. Αυτό γίνεται, διότι τα μοντέλα αυτά συγκρίνουν τις πληροφορίες με τα πραγματικά δεδομένα.

2. Η ωριμότητα ανίχνευσης απέναντι της πρόληψης. Οι τεχνικές ανίχνευσης έχουν αναπτυχθεί και τεκμηριωθεί αρκετά, ενώ οι τεχνικές πρόληψης παραμένουν κυρίως σε θεωρητικό επίπεδο ή βασίζονται σε υποτιθέμενες προσεγγίσεις.

3. Ο ρόλος της δομής του δικτύου. Η εξέταση για το πως διαδίδονται οι πληροφορίες μέσα από το δίκτυο, μαζί με τα μεταδεδομένα πετυχαίνουν καλύτερη πρόληψη σε σχέση με την ανάλυση κειμένου.

4. Οι περιορισμοί των παραδοσιακών μεθόδων. Οι κλασικοί αλγόριθμοι μηχανικής μάθησης που βασίζονται σε λεξιλόγιο δεν αντιμετωπίζουν σωστά τις νευρωνικές ψευδές ειδήσεις.

7.2 Σημαντικότερα ερευνητικά κενά και ερευνητικές κατευθύνσεις

Δεν μπορούμε όμως να μην προσδιορίσουμε και τα προβλήματα που εντοπίστηκαν σε αυτή την έρευνα, καθώς και τους τους άξονες για την περαιτέρω διερεύνησή τους.

1. Ανίχνευση περιεχομένου από LLMs. Είναι αναγκαίο να δημιουργήσουμε ανιχνευτές που είναι εκπαιδευμένοι όχι μόνο σε ανθρώπινα, αλλά και σε AI δεδομένα. Με αυτό τον τρόπο μπορούμε να αντιμετωπίσουμε κείμενα που είναι πανομοιότυπα με τα ανθρώπινα.

2. Επεξηγηματικότατα Είναι αναγκαία η μετάβαση από παλιά μοντέλα σε μοντέλα με διαφανή συστήματα που εξηγούν τις αποφάσεις τους, έτσι ώστε να ενισχύσουν την εμπιστοσύνη των χρηστών και των δημοσιογράφων.
3. Λειτουργία σε πραγματικό χρόνο. Η έρευνα στρέφεται σε λύσεις που αναπτύσσονται εύκολα και μπορούν να εντοπίζουν και να περιορίζουν bots όσο δημιουργούν την παραπληροφόρηση πριν αυτή αποκτήσει μεγάλη διάσταση.

4. Πολυγλωσσικότητα. Έλλειψη πόρων σε γλώσσες πέραν της αγγλικής, γεγονός που οδηγεί στην ανάγκη χρήσης τεχνικών Transfer Learning και συλλογής πολυγλωσσικών συνόλων δεδομένων.

7.3 Προσωπική εμπειρία που αποκομίσθηκε

Η συγγραφή αυτής της διπλωματικής αποτέλεσε μια ευχάριστα μεγάλη πρόκληση. Μου δόθηκε η δυνατότητα να εμβαθύνω σε ένα από τα πιο μείζον κοινωνικά-τεχνικά θέματα των σύγχρονων εποχών μας. Η διαδικασία μελέτης και ανάλυσης των βιβλιογραφιών, με βοήθησε να μάθω να ξεχωρίζω τα κυριότερα επιστημονικά ζητήματα μέσα σε ένα τοπίο που αλλάζει συνεχώς.

Ωστόσο πέρα του τεχνικού ζητήματος, η μελέτη προσέφερε μια νέα οπτική γωνιά σχετικά με το θέμα της ευθύνης που έχουμε ως χρήστες και ως επιστήμονες. Αντιλαμβάνομαι πλέον πως η τεχνολογία δεν είναι μόνο το μέσο διάδοσης των ψευδών ειδήσεων, αλλά χρησιμοποιείται και ως εργαλείο για την αντιμετωπίσιμη της διασποράς της, κάτι το οποίο βοήθησε και την κριτική μου σκέψη. Επίσης, να τονίσω εδώ ότι με την βοήθεια και την συνεργασία του επιβλέποντα καθηγητή ανέλυσα δύσκολους όρους, το οποίο οδήγησε σε ένα πιο ολοκληρωμένο επιστημονικό αποτέλεσμα.

Βιβλιογραφία

[1] Pizzagate conspiracy theory https://www.britannica.com/topic/Pizzagate

[2] The Onion America's Finest News Source https://www.theonion.com/

[3] Clickhole https://clickhole.com/

[4] ‘Tackling online disinformation’, European Commission https://digital-strategy.ec.europa.eu/en/policies/online-disinformation

[5] David M. J. Lazer et al., ‘The science of fake news’. Science 359, 1094-1096 (2018). DOI:10.1126/science.aao2998

[6] Gray, Jonathan, et al. "Satire TV: Politics and Comedy in the Post-Network Era." NYU Press, 2009.

[7] D. Das and A. J. Clark, "Satire vs Fake News: You Can Tell by the Way They Say It," 2019 First International Conference on Transdisciplinary AI (TransAI), Laguna Hills, CA, USA, 2019, pp. 22-26, doi: 10.1109/TransAI46475.2019.00012.

[8] Jennifer Kavanagh and Michael D. Rich, "Truth Decay: An Initial Exploration of the Diminishing Role of Facts and Analysis in American Public Life" (RAND Corporation, 2018) https://www.rand.org/pubs/research_reports/RR2314.html

[9] "The Disinformation Age: Politics, Technology, and Disruptive Communication in the United States", Cambridge University Press, 2019, edited by W. Lance Bennett and Steven Livingston, https://www.ssrc.org/publications/the-disinformation-age-politics-technology-and-disruptive-communication-in-the-united-states/

[10] Brian L. Ott and Greg Dickinson, "Fake News: Understanding Media and Misinformation in the Digital Age" (Routledge, 2019)

[11] "Media Manipulation and Disinformation Online" edited by Alice Marwick and Rebecca Lewis, Data & Society Research Institute, 2017.

[12] Tim Wu , "The Attention Merchants: The Epic Scramble to Get Inside Our Heads" , 2016.

[13] Eiri Elvestad and Angela Phillips, "Misunderstanding News Audiences: Seven Myths of the Social Media Era", Routledge, 2018.

[14] Garth S. Jowett and Victoria O'Donnell, "Propaganda and Persuasion", Sage Publications, 2020.

[15] Nicholas J. Cull, "Media and Propaganda: Understanding the Basics" John Wiley & Sons, 2018.

[16] Pennycook, G., & Rand, D. G. (2019). Fighting misinformation on social media using crowdsourced judgments of news source quality. Proceedings of the National Academy of Sciences, 116(7), 2521-2526.

[17] Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. Science, 359(6380), 1146-1151.

[18] Pennycook, G., Epstein, Z., Mosleh, M., Arechar, A. A., Eckles, D., & Rand, D. G. (2021). Shifting attention to accuracy can reduce misinformation online. Nature, 592(7853), 590-595.

[19] Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 US presidential election. Science, 363(6425), 374-378.

[20] Guess, A., Nyhan, B., & Reifler, J. (2019). Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. European Research Council

[21] Bradshaw, S., & Howard, P. N. (2018). Troops, trolls and troublemakers: A global inventory of organized social media manipulation. Oxford, UK: Computational Propaganda Research Project, University of Oxford.

[22] Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., & Quattrociocchi, W. (2016). The spreading of misinformation online. Proceedings of the National Academy of Sciences, 113(3), 554-559.

[23] Ferrara, E., Varol, O., Davis, C. A., Menczer, F., & Flammini, A. (2016). The rise of social bots. Communications of the ACM, 59(7), 96-104.

[24] Guess, A. M., Nagler, J., & Tucker, J. (2019). Less than you think: Prevalence and predictors of fake news dissemination on Facebook. Science Advances, 5(1).

[25] Lewandowsky, S., Ecker, U. K., & Cook, J. (2012). Beyond misinformation: Understanding and coping with the "post-truth" era. Journal of Applied Research in Memory and Cognition, 6(4), 353-369.

[26] Mitchell, A., Gottfried, J., Stocking, G., & Walker, M. (2018). Media and misinformation in the digital age. Pew Research Center.

[27] Pennycook, G., McPhetres, J., Zhang, Y., Lu, J. G., & Rand, D. G. (2020). Fighting COVID-19 misinformation on social media: Experimental evidence for a scalable accuracy-nudge intervention. Psychological Science, 31(7), 770-780.

[28] Shao, C., Ciampaglia, G. L., Varol, O., Yang, K. C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. Nature Communications, 9(1), 4787.

[29] Stier, S., Berman, M. G., & Bettencourt, L. M. (2020). COVID-19 attack rate increases with city size. PLoS ONE, 15(9), e0239801.

[30] Woolley, S. C., & Howard, P. N. (2016). Automation, algorithms, and politics| political communication, computational propaganda, and autonomous agents—introduction. International Journal of Communication, 10, 17.

[31] Edelman Trust Barometer. (2021). Edelman Trust Barometer 2021: Trust in crisis. Retrieved from https://www.edelman.com/trust/2021-trust-barometer

[32] GPTZero https://gptzero.me/

[33] Tucker, J. A., Guess, A., Barbera, P., Vaccari, C., Siegel, A., Sanovich, S., ... & Nyhan, B. (2018). Social media, political polarization, and political disinformation: A review of the scientific literature. Political Science Research and Methods, 8(2), 1-22.

[34] PolitiFact https://www.politifact.com/

[35] FactCheck https://www.factcheck.org/

[36] Washington post fact-checker https://www.washingtonpost.com/politics/fact-checker/

[37] Snopes https://www.snopes.com/

[38] TruthOrFiction https://www.truthorfiction.com/

[39] Fiskkit https://fiskkit.com/

[40] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary G. Ives. 2007. DBpedia: A Nucleus for a Web of Open Data. In Proceedings of ISWC. 722–735.

[41] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. 2017. Toward automated fact-checking: Detecting check-worthy factual claims by claimbuster. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1803–1812.

[42] Giovanni Luca Ciampaglia, Prashant Shiralkar, Luis Mateus Rocha, Johan Bollen, Filippo Menczer, and Alessandro Flammini. 2015. Computational fact checking from knowledge networks. CoRR abs/1501.03471 (2015).

[43] Baoxu Shi and Tim Weninger. 2016. Discriminative predicate path mining for fact checking in knowledge graphs. Knowl.-Based Syst. 104 (2016), 123–133.

[44] Prashant Shiralkar, Alessandro Flammini, Filippo Menczer, and Giovanni Luca Ciampaglia. 2017. Finding Streams in Knowledge Graphs to Support Fact Checking. In Proceedings of ICDM 2017. 859–864.

[45] Huynh, Viet-Phi & Papotti, Paolo. (2018). Towards a Benchmark for Fact Checking with Knowledge Bases. 1595-1598. Proceedings of The Web Conference 2018, Lyon, France, April 2018.

[46] Mohamed H. Gad-Elrab, Daria Stepanova, Jacopo Urbani, and Gerhard Weikum. 2019. ExFaKT: A Framework for Explaining Facts over Knowledge Graphs and Text. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM '19). Association for Computing Machinery, New York, NY, USA, 87–95. https://doi.org/10.1145/3289600.3290996

[47] Chen Y, Conroy N.J, Rubin VL (2015) Misleading online content: recognizing clickbait as false news. In: Proceedings of the 2015 ACM on workshop on multimodal deception detection, pp 15–19

[48] Jin Z, Cao J, Zhang Y, Zhou J, Tian Q (2016) Novel visual and statistical image features for microblogs news verification. IEEE Trans Multimedia 19(3):598–608

[49] Castillo C, Mendoza M, Poblete B (2011) Information credibility on twitter. In: Proceedings of the 20th international conference on world wide web pp 675–684

[50] Kwon S, Cha M, Jung K, Chen W, Wang Y (2013) Prominent features of rumor propagation in online social media. In: 2013 IEEE 13th international conference on data mining. IEEE, pp 1103–1108

[51] Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). The de-velopment and psychometric properties of LIWC2015. Austin, TX:University of Texas at Austin.

[52] Gary D. Bond, Rebecka D. Holman, Jamie-Ann L. Eggert, Lassiter F. Speller, Olivia N. Garcia, Sasha C. Mejia, Kohlby W. Mcinnes, Eleny C. Ceniceros, and Rebecca Rustige. 2017. ‘Lyin’ Ted’,‘Crooked Hillary’, and ‘Deceptive Donald’: Language of lies in the 2016 US presidential debates. Applied Cognitive Psychology 31, 6 (2017), 668–677.

[53] William Yang Wang. 2017. "liar, liar pants on fire": A new benchmark dataset for fake news detection. In Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL 2017, Vancouver, Canada, July 30 - August 4, Volume 2: Short Papers, pages 422–426.

[54] Benjamin D. Horne and Sibel Adali. 2017. This just in: Fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news. The Workshops of the Eleventh International AAAI Conference on Web and Social MediaAAAI Technical Report WS-17-17: News and Public Opinion

[55] Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. COLING 2017.

[56] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A stylometric inquiry into hyperpartisan and fake news. Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers).

[57] Shu, K., Wang, S., & Liu, H. (2017). Exploiting Tri-Relationship for Fake News Detection. ArXiv, abs/1712.07709.

[58] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł. & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems (p./pp. 5998--6008).

[59] Yang, F., Mukherjee, A., & Dragut, E.C. (2017). Satirical News Detection and Analysis using Attention Mechanism and Linguistic Features. Conference on Empirical Methods in Natural Language Processing.

[60] Jiang, S., & Wilson, C. (2018). Linguistic Signals under Misinformation and Fact-Checking. Proceedings of the ACM on Human-Computer Interaction, 2, 1 - 23.

[61] Bilal Ghanem, Manuel Montes-y Gomez, Francisco `Rangel, and Paolo Rosso. 2018b. Upv-inaoeautoritas - check that: An approach based on external sources to detect claims credibility. In CLEF 2018 Working Notes. Working Notes of CLEF 2018 - Conference and Labs of the Evaluation Forum, CEUR Workshop Proceedings, Avignon, France.

[62] Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M Meyer, and Iryna Gurevych. 2018. A retrospective analysis of the fake news challenge stance detection tas, 2018.k. Proceedings of the 27th International Conference on Computational Linguistics. https://arxiv.org/pdf/1806.05180

[63] Ghanem, B., Rosso, P., & Rangel, F. (2018). Stance Detection in Fake News A Combined Feature Representation. FEVER@EMNLP.

[64] Hassan, F.M., & Lee, M.G. (2019). Imbalanced Stance Detection by Combining Neural and External Features. International Conference on Statistical Language and Speech Processing.

[65] Hassan, F.M., & Lee, M.G. (2019). Multi-stage News-Stance Classification Based on Lexical and Neural Features. Computational Intelligence in Security for Information Systems.

[66] Hassan, F.M., & Lee, M.G. (2020). Political Fake Statement Detection via Multistage Feature-assisted Neural Modeling. 2020 IEEE International Conference on Intelligence and Security Informatics (ISI), 1-6.

[67] Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G.S., & On, B. (2020). Fake News Stance Detection Using Deep Learning Architecture (CNN-LSTM). IEEE Access, 8, 156695-156706.

[68] Ashraf, N., Butt, S., Sidorov, G., & Gelbukh, A.F. (2021). CIC at CheckThat! 2021: Fake News detection Using Machine Learning And Data Augmentation. Conference and Labs of the Evaluation Forum.

[69] Glazkova, A. (2023). Data Augmentation for Fake News Detection by Combining Seq2seq and NLI. Recent Advances in Natural Language Processing.

[70] Salah, I., Jouini, K., & Korbaa, O. (2023). On the use of text augmentation for stance and fake news detection. Journal of Information and Telecommunication, 7, 359 - 375.

[71] Patil, D.R. (2022). Fake News Detection Using Majority Voting Technique. ArXiv, abs/2203.09936.

[72] Zhou, Xinyi & Jain, Atishay & Phoha, Vir & Zafarani, Reza. (2019). Fake News Early Detection: A Theory-driven Model.

[73] Rowan Zellers, Ari Holtzman, Hannah Rashkin, Yonatan Bisk, Ali Farhadi, Franziska Roesner, and Yejin Choi. 2019. Defending against neural fake news. In Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, pages 9051–9062, Vancouver, BC, Canada.

[74] Josh A Goldstein, Girish Sastry, Micah Musser, Renee DiResta, Matthew Gentzel, and Katerina Sedova. 2023. Generative language models and automated influence operations: Emerging threats and potential mitigations. ArXiv preprint, abs/2301.04246.

[75] Jiawei Zhou, Yixuan Zhang, Qianni Luo, Andrea G Parker, and Munmun De Choudhury. 2023. Synthetic lies: Understanding AI-generated misinformation and evaluating algorithmic and human solutions. In Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI ’23, Hamburg, Germany. Association for Computing Machinery.

[76] Jinyan Su, Claire Cardie, and Preslav Nakov. 2024. Adapting Fake News Detection to the Era of Large Language Models. In Findings of the Association for Computational Linguistics: NAACL 2024, pages 1473–1490, Mexico City, Mexico. Association for Computational Linguistics.

[77] Sun, Y., He, J., Cui, L., Lei, S., & Lu, C. (2024). Exploring the Deceptive Power of LLM-Generated Fake News: A Study of Real-World Detection Challenges. ArXiv, abs/2403.18249.

[78] Saxena, A., Saxena, P., Reddy, H. (2022). Fake News Propagation and Mitigation Techniques: A Survey. In: Biswas, A., Patgiri, R., Biswas, B. (eds) Principles of Social Networking. Smart Innovation, Systems and Technologies, vol 246. Springer, Singapore. https://doi.org/10.1007/978-981-16-3398-0_16

[79] Amoruso, M., Anello, D., Auletta, V., Ferraioli, D.: Contrasting the spread of misinformation in online social networks. In: Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems. International Foundation for Autonomous Agents and Multiagent Systems, pp. 1323–1331 (2017)

[80] Cao, C., Yu, L., Hu, Y.: Containment of rumors underlimitcost budget in social network. In: Fourteenth Wuhan International Conference on E-Business, pp. 341–348 (2015).

[81] Zheng, J., Pan, L.: Least cost rumor community blocking optimization in social networks. In: 2018 Third International Conference on Security of Smart Cities, Industrial Control System and Communications (SSIC), pp. 1–5. IEEE (2018)

[82] Wu, Q., Zhao, X., Zhou, L., Wang, Y., Yang, Y.: Minimizing the influence of dynamic rumors based on community structure. Int. J. Crowd Sci. (2019)

[83] Fan, L., Lu, Z., Wu, W., Thuraisingham, B., Ma, H., Bi, Y.: Least cost rumor blocking in social networks. In: 2013 IEEE 33rd International Conference on Distributed Computing Systems, pp. 540–549. IEEE (2013)

[84] Wang, B., Chen, G., Luoyi, F., Song, L., Wang, X.: DRIMUX: dynamic rumor influence minimization with user experience in social networks. IEEE Trans. Knowl. Data Eng. 29(10), 2168–2181 (2017)

[85] Pham, D.V., Duong, H.V., Pham, C.V., Bao, B.Q., Nguyen, A.V.: Multiple topics misinformation blocking in online social networks. In: 2019 11th International Conference on Knowledge and Systems Engineering (KSE), pp. 1–6. IEEE (2019)

[86] Yang, L., Li, Z., Giua, A.: Influence minimization in linear threshold networks. Automatica 100, 10–16 (2019)

[87] Garrett, R.K.: Echo chambers online?: Politically motivated selective exposure among internet news users. J. Comput.-Mediat. Commun. 14(2), 265–285 (2009)

[88] Van der Linden, S., Leiserowitz, A., Rosenthal, S., Maibach, E.: Inoculating the public against misinformation about climate change. Glob. Chall. 1(2), 1600008 (2017)

[89] Cook, J., Lewandowsky, S., Ecker, U.K.H.: Neutralizing misinformation through inoculation: exposing misleading argumentation techniques reduces their influence. PloS One 12(5), e0175799 (2017)

[90] Tanaka, Y., Sakamoto, Y., Matsuka, T.: Toward a social-technological system that inactivates false rumors through the critical thinking of crowds. In: 2013 46th Hawaii International Conference on System Sciences (HICSS), pp. 649–658. IEEE (2013)

[91] Ozturk, P., Li, H., Sakamoto, Y.: Combating rumor spread on social media: The effectiveness of refutation and warning. In: 2015 48th Hawaii International Conference on System Sciences (HICSS), pp. 2406–2414. IEEE (2015)

[92] Nicol, D.M., Liljenstam, M.: Models and analysis of active worm defense. In: International Workshop on Mathematical Methods, Models, and Architectures for Computer Network Security, pp. 38–53. Springer (2005)

[93] Kostka, J., Oswald, Y.A., Wattenhofer, R.: Word of mouth: rumor dissemination in social networks. In: International Colloquium on Structural Information and Communication Complexity, pp. 185–196. Springer (2008)

[94] Budak, C., Agrawal, D., El Abbadi, A.: Limiting the spread of misinformation in social networks. In: Proceedings of the 20th international conference on World wide web, pp. 665–674. ACM (2011)

[95] Song, C., Hsu, W., Lee, M.-L.: Temporal influence blocking: minimizing the effect of misinformation in social networks. In: 33rd IEEE International Conference on Data Engineering, ICDE 2017, San Diego, CA, USA, 19–22 April 2017, pp. 847–858 (2017)

[96] Saxena, A., Hsu, W., Lee, M.L., Chieu, H.L., Ng, L., Teow, L.N.: Mitigating misinformation in online social network with top-k debunkers and evolving user opinions. In: Companion Proceedings of the Web Conference 2020, pp. 363–370 (2020)

[97] Xu, W., Zhang, F., Zhu, S.: Toward worm detection in online social networks. In: Proceedings of the 26th Annual Computer Security Applications Conference, pp. 11–20. ACM (2010)

[98] Wilder, B., Vorobeychik, Y.: Defending elections against malicious spread of misinformation (2018). The Thirty-Third AAAI Conference on Artificial Intelligence

[99] Tanınmış, K., Aras, N., Altınel, İK., Güney, E.: Minimizing the misinformation spread in social networks. IISE Trans. 52(8), 850–863 (2020)

[100] Park, S., Kang, S., Chung, S., Song, J. Newscube: delivering multiple aspects of news to mitigate media bias. In: Proceedings of the SIGCHI Conference on H:uman Factors in Computing Systems, pp. 443–452. ACM (2009)

[101] Ennals, R., Trushkowsky, B., Agosta, J.M.: Highlighting disputed claims on the web. In: Proceedings of the 19th International Conference on World Wide Web, pp. 341–350. ACM (2010)

[102] Hassan, N., Sultana, A., You, W., Zhang, G., Li, C., Yang, J., Cong, Yu.: Data in, fact out: automated monitoring of facts by factwatcher. Proc. VLDB Endow. 7(13), 1557–1560 (2014)

[103] Hassan, N., Arslan, F., Li, C., Tremayne, M.: Toward automated fact-checking: detecting check-worthy factual claims by claimbuster. In: Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1803–1812. ACM (2017)

[104] Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., Menczer, F.: Truthy: mapping the spread of astroturf in microblog streams. In: Proceedings of the 20th International Conference Companion on World Wide Web, pp. 249–252. ACM (2011)

[105] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu, "FakeNewsNet: A Data Repository with News Content, Social Context and Spatiotemporal Information for Studying Fake News on Social Media." Journal of Big Data, 2018.

[106] Craig Silverman, Jeremy Singer-Vine, "Most Americans Who See Fake News Believe It, New Survey Says," BuzzFeed News, 2016.

[107] BS detector dataset. GitHub. https://github.com/bs-detector/bs-detector

[108] Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., & Tolmie, P. (2016). "PHEME dataset of rumors and non-rumors." Security and Communicaiton networks, Wiley online.

[109] Ahmed H. Ahmed, Liza Potgieter, and Rossouw von Solms, "ISOT Fake News Dataset. https://onlineacademiccommunity.uvic.ca/isot/2022/11/27/fake-news-detection-datasets/

[110] Tide Pod Challenge https://en.wikipedia.org/wiki/Consumption_of_Tide_Pods

[111] Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opinion on Facebook. Science, 348(6239), 1130-1132.

[112] Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. Public Opinion Quarterly, 80(S1), 298-320.

[113] Pariser, E. (2011). The filter bubble: What the Internet is hiding from you.

[114] Pennycook, G., & Rand, D. G. (2018). The Implied Truth Effect: Attaching Warnings to a Subset of Fake News Stories Increases Perceived Accuracy of Stories Without Warnings. Management Science, 66(11), 4944–4957.

[115] Tufekci, Z. (2015). Algorithmic harms beyond Facebook and Google: Emergent challenges of computational agency. Colorado Technology Law Journal, 13(203), 203–218.

[116] Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2018). Bots increase exposure to negative and inflammatory content in online social systems. arXiv preprint arXiv:1802.07292.

[117] Howard, P. N. (2020). Lie machines: How to save democracy from troll armies, deceitful robots, junk news operations, and political operatives. Yale University Press.

[118] EU DisinfoLab. (2023). Disinformation landscape in Greece. https://www.disinfo.eu/publications/disinformation-landscape-in-greece/

[119] BotNet Detection on Social Media. (2021). Botnets as media for misinformation and propaganda. arXiv preprint arXiv:2110.05661. https://arxiv.org/pdf/2110.05661

[120] https://chatgpt.com/

[121]Kitchenham, B., & Charters, S. (2007). Guidelines for performing Systematic Literature Reviews in Software Engineering.

[122] Wohlin, C. (2014). Guidelines for snowballing in systematic literature studies and a replication in software engineering

[123] https://arxiv.org/pdf/1806.05180

[124] Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.

[125] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1), 22-36.

[126] Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. arXiv preprint arXiv:1811.00770.

[127] Zhang, X., & Ghorbani, A. A. (2020). An overview of online fake news: Characterization, detection, and discussion. Information Processing & Management, 57(2), 102025.

[128] Zhou, X., & Zafarani, R. (2020). A survey of fake news: Fundamental theories, propagation, and detection. ACM Computing Surveys (CSUR), 53(5), 1-40.