

costaid

Coping Strategies Against Information Disorder

Κατευθυντήριες γραμμές για τα άτομα στην πρώτη γραμμή



Co-funded by
the European Union

Με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης. Οι απόψεις και οι γνώμες που διατυπώνονται εκφράζουν αποκλειστικά τις απόψεις των συντακτών και δεν αντιπροσωπεύουν κατ' ανάγκη τις απόψεις της Ευρωπαϊκής Ένωσης ή του Ευρωπαϊκού Εκτελεστικού Οργανισμού Εκπαίδευσης και Πολιτισμού (ΕΑΕΕΑ). Η Ευρωπαϊκή Ένωση και ο ΕΑΕΕΑ δεν μπορούν να θεωρηθούν υπεύθυνοι για τις εκφραζόμενες απόψεις.

Συγγραφείς:

*Eliane Smits van Waesberghe & Tim Paulusse — Verwey-Jonker Instituut
(Κύριοι συντάκτες)*

Leen D'Haenens & Joyce Vissenberg — KU Leuven

Tzvetalina Genova – International Management Institute

Wolfgang Eisenreich – Wissenschaftsinitiative Niederösterreich

Sonja Bercko Eisenreich – Integra Institute

Alenka Valjašková — QUALED

Παντελής Μπαλαούρας — Connexions

Δήλωση σχετικά με τα δικαιώματα πνευματικής ιδιοκτησίας:



Το έργο αυτό έχει αδειοδοτηθεί με άδεια [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

Μπορείτε να:

- Μοιραστείτε — αντιγράψετε και αναδιανέμετε το υλικό με κάθε μέσο και τρόπο
- Προσαρμόστε — αναμείξτε, τροποποιήστε και δημιουργήστε πάνω στο υλικό

Υπό τους ακόλουθους όρους:

- Αναφορά Δημιουργού - Θα πρέπει να καταχωρίσετε αναφορά στο δημιουργό, με σύνδεσμο της άδειας, και με αναφορά αν έχουν γίνει αλλαγές. Μπορείτε να το κάνετε αυτό με οποιονδήποτε εύλογο τρόπο, αλλά όχι με τρόπο που να υπονοεί ότι ο δημιουργός αποδέχεται το έργο σας ή τη χρήση που εσείς κάνετε.
- Μη Εμπορική Χρήση — Δε μπορείτε να χρησιμοποιήσετε το υλικό για εμπορικούς σκοπούς.
- Παρόμοια Διανομή — Αν αναμείξετε, τροποποιήσετε, ή δημιουργήσετε πάνω στο υλικό, πρέπει να διανείμετε τις δικές σας συνεισφορές υπό την ίδια άδεια όπως και το πρωτότυπο.

Κεφάλαιο 3

Τεχνολογία & Εργαλεία

Ομάδα-στόχος

Οι εν λόγω κατευθυντήριες γραμμές απευθύνονται στους λεγόμενους «πρώτους τομείς». Ο όρος «πρώτης γραμμής» είναι ένας γενικός όρος για όλους τους ανθρώπους που έρχονται σε άμεση επαφή με άτομα που είναι ευάλωτα σε διαταραχές της πληροφορίας, εστιάζοντας σε ομάδες στην επαγγελματική εκπαίδευση. Παραδείγματα ανθρώπων που εμπίπτουν στον γενικό όρο είναι: εκπαιδευτικοί, εκπαιδευτικοί, εκπαιδευτές, σύμβουλοι και σύμβουλοι νεολαίας, κοινωνικοί λειτουργοί και εργαζόμενοι στον τομέα της νεολαίας. Ωστόσο, αυτή είναι μια μη εξαντλητική λίστα. Το πεδίο εφαρμογής αυτού του έργου περιλαμβάνει επίσης και άλλα άτομα που εργάζονται στον τομέα της εκπαίδευσης, της κοινωνικής πρόνοιας ή της υγειονομικής περίθαλψης.

Ομάδα-στόχος.....	3
3.1 Εισαγωγή.....	1
3.2 Μηχανές αναζήτησης & αλγόριθμοι.....	3
3.3 Διαδικτυακές στρατηγικές παραπληροφόρησης και εξτρεμιστικών οργανώσεων.....	5
Οι πιο συνηθισμένες μέθοδοι.....	5
Τρολάρισμα & doxing.....	6
Ενσωμάτωση της διάστασης.....	7
Επηρεαστές.....	7
Ειρωνεία, σάτιρα & μιμίδια.....	7
3.4 Παραποιημένο περιεχόμενο.....	9
Κατανόηση deepfakes: χειρισμός συνθετικών μέσων.....	10
Ανίχνευση και μετριασμός των deepfakes: τεχνολογικές προσεγγίσεις.....	11
3.5 Αναφορές.....	12

3.1 Εισαγωγή

Η έννοια της παραπληροφόρησης (κατασκευασμένες ειδήσεις) έχει συζητηθεί σε προηγούμενα κεφάλαια. Οι ειδήσεις διαδίδονται συνήθως από επαγγελματίες παρόχους ειδήσεων, συμπεριλαμβανομένων δημόσιων μέσων ενημέρωσης, εμπορικών ειδησεογραφικών μέσων, ανεξάρτητης επαγγελματικής δημοσιογραφίας ή ερασιτέχνες χρήστες (στην περίπτωση πλατφορμών μέσων κοινωνικής δικτύωσης). Τα νέα είναι προσβάσιμα και διαθέσιμα σε διάφορες μορφές:

- **Ψηφιακά Νέα:** Ειδήσεις που διανέμονται μέσω διαδικτυακών καναλιών σε μορφή ψηφιακών μέσων (κείμενο, εικόνες, ήχος, βίντεο).
- **Μέσα εκτύπωσης:** Εφημερίδες και περιοδικά που περιέχουν περιεχόμενο κειμένου και εικόνας.
- **Ραδιοηλεκτρονικές μεταδόσεις:** Τηλεόραση και ραδιόφωνο με περιεχόμενο βίντεο και ήχου.

Σε αυτή την ενότητα, η εστίασή μας περιστρέφεται κυρίως γύρω από την πρώτη μορφή, τις ψηφιακές ειδήσεις. Ωστόσο, οι συζητήσεις που παρουσιάζονται μπορεί επίσης να ισχύουν και για άλλες μορφές που χρησιμοποιούν βίντεο, ήχο και εικόνες.

Οι ψηφιακές ειδήσεις είναι προσβάσιμες κυρίως μέσω ειδησεογραφικών πλατφορμών τις οποίες διαχειρίζονται επαγγελματίες πάροχοι ειδήσεων. Οι χρήστες μπορούν να έχουν πρόσβαση σε αυτές τις πλατφόρμες ειδήσεων είτε δωρεάν είτε μέσω συνδρομητικών υπηρεσιών. Αυτές οι πλατφόρμες προσφέρουν web feeds μέσα σε δημοφιλή προγράμματα περιήγησης και ροές ειδήσεων σε πλατφόρμες κοινωνικών μέσων, επιτρέποντας στους χρήστες να λαμβάνουν εξατομικευμένες ενημερώσεις ειδήσεων.

Μια ροή ειδήσεων είναι μια ιστοσελίδα ή μια οθόνη που ενημερώνεται συχνά για να εμφανίσει τα τελευταία νέα ή πληροφορίες. Οι εξατομικευμένες ροές ειδήσεων είναι υπηρεσίες ενσωματωμένες σε προγράμματα περιήγησης ιστού (web feeds) και πλατφόρμες μέσων κοινωνικής δικτύωσης που παρέχουν ειδήσεις στους χρήστες με βάση τις προσωπικές τους προτιμήσεις. Αυτές οι προτιμήσεις καθορίζονται από διάφορους παράγοντες, όπως η εγγραφή σε κανάλια τροφοδοσίας ιστού, η επίσκεψη σε συγκεκριμένες ιστοσελίδες και πολλά άλλα.

Στις πλατφόρμες μέσων κοινωνικής δικτύωσης, οι χρήστες λαμβάνουν επίσης κοινά μέσα από άλλους χρήστες, τα οποία μπορεί να περιλαμβάνουν περιπτώσεις παραπληροφόρησης.

Εξατομικευμένες ειδήσεις εμφανίζονται στο πρόγραμμα περιήγησης ή την πλατφόρμα μέσων κοινωνικής δικτύωσης ενός χρήστη μέσω υπηρεσιών και αλγορίθμων μέσων κοινωνικής δικτύωσης. Οι χρήστες ενδέχεται να μην έχουν πλήρη επίγνωση του τρόπου λειτουργίας αυτών των αλγορίθμων, καθώς παραμένει ασαφές εάν η επιλογή των ειδήσεων βασίζεται αποκλειστικά στις προτιμήσεις των χρηστών ή σε άλλα κριτήρια. Για παράδειγμα, οι πλατφόρμες κοινωνικών μέσων παρουσιάζουν ειδήσεις με βάση τους παρόχους ειδήσεων που ακολουθεί ένας χρήστης, τις συνήθειες ανάγνωσης των φίλων τους ή πρόσφατα κλικ σε

άρθρα, τα οποία μπορεί να αποτελούν ένα είδος προτίμησης. Ως εκ τούτου, μια λίστα των ειδήσεων φιλτράρεται, πράγμα που σημαίνει ότι δεν εμφανίζονται όλες οι ειδήσεις, αλλά μάλλον εκείνες που ο αλγόριθμος θεωρεί πιο ενδιαφέρον για το χρήστη. Πολλοί υποστηρίζουν ότι αυτό δημιουργεί μια φούσκα πληροφοριών, όπου οι ειδήσεις επιλέγονται από έναν αλγόριθμο και μπορεί να μην περιλαμβάνουν ειδήσεις που ένας χρήστης ενδιαφέρεται, αλλά ο αλγόριθμος δεν περιλαμβάνει. Ως εκ τούτου, είναι σημαντικό να μην βασίζονται αποκλειστικά σε προσωπικές ροές ειδήσεων, αλλά μάλλον να επισκέπτονται επαγγελματικές πλατφόρμες παρόχων ειδήσεων που είναι αξιόπιστες. Είναι πάντα σημαντικό να αξιολογούμε κριτικά την αξιοπιστία και την ακρίβεια των ειδήσεων, ανεξάρτητα από τη μορφή ή τη μέθοδο παράδοσης.

3.2 Μηχανές αναζήτησης & αλγόριθμοι

Πολλοί ιστότοποι στο διαδίκτυο έχουν ως στόχο να κρατήσουν τους χρήστες τους δεσμευμένους και να μεγιστοποιήσουν το χρόνο που αφιερώνουν στην πλατφόρμα. Αυτό ισχύει ιδιαίτερα για τις πλατφόρμες κοινωνικών μέσων, οι οποίες χρησιμοποιούν διάφορες στρατηγικές για να βελτιώσουν την εμπειρία του χρήστη. Μια τέτοια στρατηγική είναι να δείξει στους χρήστες περιεχόμενο που ευθυγραμμίζεται με τα ενδιαφέροντά τους. Ωστόσο, οι ιστότοποι δεν μπορούν να διαβάσουν τα μυαλά για να γνωρίζουν τις προτιμήσεις των χρηστών. Για να αντιμετωπιστεί αυτό, οι αλγόριθμοι χρησιμοποιούνται για την ανάλυση των δεδομένων των χρηστών και την παροχή εξατομικευμένου περιεχομένου.

Οι αλγόριθμοι είναι σύνθετοι τύποι που παρατηρούν, μετρούν και υπολογίζουν τις προτιμήσεις περιεχομένου ενός ατόμου. Αυτό μπορεί να περιλαμβάνει παράγοντες όπως ο χρόνος παρακολούθησης του χρήστη σε συγκεκριμένους τύπους βίντεο, ο χρόνος που δαπανάται σε μια συγκεκριμένη ανάρτηση ή ενέργειες δέσμευσης, όπως η αποχώρηση από συμπάθειες ή σχόλια. Αναλύοντας αυτά τα δεδομένα, οι αλγόριθμοι καθορίζουν τον τύπο του περιεχομένου που κρατά τους χρήστες δεσμευμένους για μεγαλύτερο χρονικό διάστημα. Αυτές οι πληροφορίες χρησιμοποιούνται στη συνέχεια για να επιμεληθούν και να συστήσουν παρόμοιο περιεχόμενο με το χρήστη.

Ενώ αυτή η προσέγγιση μπορεί να φαίνεται λογική και αβλαβής, υπάρχουν πιθανά μειονεκτήματα στην αλγοριθμική χρήση. Η προτίμηση του αλγορίθμου για το περιεχόμενο που ο χρήστης βρίσκει ενδιαφέρον μπορεί να δημιουργήσει φυσαλίδες φίλτρου, όπου οι χρήστες εκτίθενται μόνο σε συγκεκριμένες απόψεις. Οι φυσαλίδες φίλτρων περιορίζουν την ποικιλομορφία του περιεχομένου, οδηγώντας ενδεχομένως σε θαλάμους ηχώ ή ενισχύοντας τους υπάρχοντες.

Εκτός από τις φυσαλίδες φίλτρου, οι αλγόριθμοι μπορούν να ενισχύσουν το άκρο του περιεχομένου συνιστώντας όλο και πιο εξειδικευμένες θέσεις, περιθωριακές και ακραίες θέσεις. Ο στόχος είναι να κρατήσει τους χρήστες δεσμευμένους, αλλά αυτό μπορεί να οδηγήσει τους χρήστες να διοχετεύονται σε διαδικτυακούς χώρους χωρίς διαφορετικές οπτικές γωνίες, με αποτέλεσμα αυτό που ονομάζεται «λαγούμι».

Οι "φούσκες φίλτρων" και τα «λαγούμια» εκθέτουν τους χρήστες σε ριζοσπαστικό περιεχόμενο και τις κοινότητες που συνδέονται με αυτό. Αυτοί οι διαδικτυακοί χώροι παρέχουν γόνιμο έδαφος για την ανάπτυξη, την ανάπτυξη, τη στρέβλωση και τη διάδοση της παραπληροφόρησης και της παραπληροφόρησης.

Καθώς οι χρήστες προχωρούν μέσα από την τρύπα του κουνελιού, ακραία σημεία ομιλίας και ψευδείς πληροφορίες ομαλοποιούνται. Η εξομάλυνση αυτή διευκολύνει περαιτέρω τη διάδοση και την αποδοχή της παραπληροφόρησης και της παραπληροφόρησης από ριζοσπαστικές πηγές.

Κατανοώντας τον αντίκτυπο των αλγορίθμων στις εμπειρίες των χρηστών, μπορούμε να κατανοήσουμε καλύτερα τους κινδύνους που σχετίζονται με τις "φούσκες φίλτρων", τους

θαλάμους αντήχησης και τις «κουνελότρυπες-λαγούμια». Οι γνώσεις αυτές είναι απαραίτητες για την πλοήγηση στο διαδικτυακό τοπίο και την αντιμετώπιση των προκλήσεων που θέτει η διαταραχή της πληροφόρησης.

3.3 Διαδικτυακές στρατηγικές παραπληροφόρησης και εξτρεμιστικών οργανώσεων

Στις προηγούμενες ενότητες, διερευνήσαμε πώς δημιουργείται και διαιωνίζεται η διαταραχή της πληροφορίας μέσα σε θαλάμους ηχώ και φυσαλίδες φίλτρων. Αλλά πώς οι ψευδείς πληροφορίες φτάνουν σε ανθρώπους έξω από αυτούς τους χώρους;

Οι πιο συνηθισμένες μέθοδοι

Η παραπληροφόρηση και η παραπληροφόρηση μπορούν να διαδοθούν μέσω αναρίθμητων μορφών επικοινωνίας. Ωστόσο, συνήθως διαδίδονται μέσω διαφόρων καναλιών, όπως πλατφόρμες κοινωνικών μέσων, ιστοσελίδες, email και από στόμα σε στόμα. Οι πιο κοινές μέθοδοι με τις οποίες διαδίδεται η διαταραχή της πληροφορίας συνδέονται στενά με τις επτά κατηγορίες προβληματικού περιεχομένου που εξετάζονται στο *Κεφάλαιο 1: Κατανόηση των «ψευδών ειδήσεων»*:

- **Σάτιρα ή παρωδία:** Ορισμένες λανθασμένες και παραπληροφόρησης δημιουργούνται για ψυχαγωγικούς σκοπούς ή σάτιρα, αλλά μπορεί να παρερμηνευθούν ως γνήσια νέα. Οι σατιρικοί ιστότοποι ή οι λογαριασμοί στα μέσα κοινωνικής δικτύωσης μπορεί να δημοσιεύουν χιουμοριστικές ή υπερβολικές ιστορίες, αλλά οι αναγνώστες που δεν γνωρίζουν τη σατιρική τους φύση μπορούν να τους μπερδέψουν για τεκμηριωμένες πληροφορίες.
- **Clickbait:** τα άρθρα που περιλαμβάνουν εσφαλμένη ή παραπληροφόρηση συχνά χρησιμοποιούν εντυπωσιακούς ή παραπλανητικούς τίτλους για να τραβήξουν την προσοχή και να δημιουργήσουν περισσότερα κλικ ή προβολές. Στόχος τους είναι να εκμεταλλευτούν την περιέργεια ή τα συναισθήματα των ανθρώπων για να οδηγήσουν την κίνηση σε μια ιστοσελίδα και να δημιουργήσουν έσοδα μέσω της διαφήμισης.
- **Διαστρέβλωση:** Αυτό συνεπάγεται παραμόρφωση ή παρερμηνεία πραγματικών ειδήσεων, παρουσιάζοντας επιλεκτικά γεγονότα ή παραλείποντας κρίσιμες πληροφορίες. Μπορεί να περιλαμβάνει τη λήψη δηλώσεων από το πλαίσιο, την αλλαγή εικόνων ή βίντεο ή τη στρέβλωση της έννοιας μιας ιστορίας ώστε να ταιριάζει σε μια συγκεκριμένη αφήγηση.
- **Απομίμηση:** η παραπληροφόρηση και η παραπληροφόρηση μπορεί επίσης να περιλαμβάνουν τη μίμηση έγκριτων πηγών ειδήσεων ή δημόσιων προσώπων για να προσδώσουν αξιοπιστία σε ψευδείς πληροφορίες. Αυτό μπορεί να γίνει μέσω της δημιουργίας ψεύτικων ιστοσελίδων ή λογαριασμών κοινωνικών μέσων που μιμούνται νόμιμες πηγές, εξαπατώντας τους αναγνώστες να πιστέψουν ότι οι πληροφορίες είναι αξιόπιστες.
- **Πολιτική χειραγώγηση:** Η διαταραχή της πληροφορίας μερικές φορές δημιουργείται ή εξαπλώνεται με σκοπό να επηρεάσει την κοινή γνώμη ή τις εκλογές. Αυτό μπορεί να περιλαμβάνει τη διάδοση ψευδών πληροφοριών σχετικά

με τους πολιτικούς υποψηφίους, τη χειραγώγηση του δημόσιου αισθήματος ή την εκμετάλλευση των υφιστάμενων προκαταλήψεων και διαιρέσεων εντός της κοινωνίας.

- **Επεξεργασία:** η παραπληροφόρηση και η παραπληροφόρηση μπορούν να κατασκευαστούν εξ ολοκλήρου, χωρίς καμία βάση στην πραγματικότητα. Περιλαμβάνει τη δημιουργία ψευδών ιστοριών, αποσπάσματα ή γεγονότα για να παραπλανήσει τους αναγνώστες ή τους θεατές.
- **Ενίσχυση μέσω των μέσων κοινωνικής δικτύωσης:** Οι πλατφόρμες μέσων κοινωνικής δικτύωσης διαδραματίζουν σημαντικό ρόλο στη διάδοση της παραπληροφόρησης και της παραπληροφόρησης. Οι ψευδείς ιστορίες μπορούν γρήγορα να γίνουν viral καθώς οι χρήστες τις μοιράζονται και τις αναδημοσιεύουν, συχνά χωρίς να επαληθεύουν την ακρίβεια των πληροφοριών. Οι αλγόριθμοι που χρησιμοποιούνται από τις πλατφόρμες μέσων κοινωνικής δικτύωσης μπορούν επίσης να συμβάλουν στην ενίσχυση με την προώθηση περιεχομένου που βασίζεται στην εμπλοκή και όχι στην ακρίβεια.

Υπάρχουν επίσης άλλες κοινές μέθοδοι που χρησιμοποιούνται για τη διάδοση της παραπληροφόρησης και της παραπληροφόρησης, οι οποίες θα πρέπει να συζητηθούν λεπτομερέστερα: τρολινγκ, doxxing και mainstreaming. Αυτό γίνεται στα επόμενα δύο υποκεφάλαια.

Τρολάρισμα & doxxing

Μια στρατηγική που χρησιμοποιείται συχνά είναι το trolling. Το trolling ορίζεται ως η σκόπιμη χρήση της αγένειας, της επιθετικότητας, της εξαπάτησης και της χειραγώγησης στην ηλεκτρονική επικοινωνία για να προκαλέσει συγκρούσεις ή διασκέδαση. Τα trolls υποκινούν διαδικτυακές συγκρούσεις εξαπατώντας, χειραγωγώντας ή όντας επιθετικοί. Εκτροχιάζουν τις συνομιλίες για τη δική τους διασκέδαση, εμπλέκοντας ουσιαστικά στον ψηφιακό εκφοβισμό.

Σε μικρή κλίμακα, το trolling μπορεί να φαίνεται σχετικά ακίνδυνο, που εμφανίζεται ως απλή ενόχληση. Ωστόσο, όταν οι οργανωμένες ομάδες τρολ μοιράζονται έναν συγκεκριμένο στόχο, αυτή η ενόχληση μπορεί να μετατραπεί γρήγορα σε επιδημία παραπληροφόρησης. Ένα παράδειγμα αυτού είναι η χρήση των τρολ των μέσων κοινωνικής δικτύωσης από τη Ρωσία ως «όπλο». Η Ρωσία χρησιμοποίησε ένα μεγάλο δίκτυο τρολ για τη διάδοση της παραπληροφόρησης σε πολλές γλώσσες παγκοσμίως, με στόχο τον έλεγχο του διαδικτυακού διαλόγου γύρω από τη Ρωσία. Αυτά τα trolls όχι μόνο διαδίδουν ψευδείς πληροφορίες, αλλά στοχεύουν και χρήστες των μέσων κοινωνικής δικτύωσης με αναρτήσεις που παρεκκλίνουν από την αφήγηση που οργανώθηκαν για να προωθήσουν. Κατά συνέπεια, πολλοί χρήστες των μέσων κοινωνικής δικτύωσης απέφυγαν να συζητήσουν για τη Ρωσία, επιτρέποντας αποτελεσματικά στα τρολ να ελέγχουν την αφήγηση με την παραπληροφόρηση τους.

Το doxing, μια άλλη μορφή εκφοβισμού στο διαδίκτυο, περιλαμβάνει την αποκάλυψη προσωπικών πληροφοριών ή ταυτοτήτων ατόμων στο διαδίκτυο χωρίς τη συγκατάθεσή τους. Αν και αυτή η τακτική δεν περιλαμβάνει συγκεκριμένα την εξάπλωση της παραπληροφόρησης, είναι μια άλλη στρατηγική που χρησιμοποιούν τα τρολ για να ελέγξουν την αφήγηση σε ένα συγκεκριμένο θέμα, παρόμοια με την τρολ. Το doxing μπορεί να χρησιμοποιηθεί για να εκφοβίσει τους χρήστες των μέσων κοινωνικής δικτύωσης, καταστέλλοντας την προθυμία τους να δημοσιεύσουν περιεχόμενο που έρχεται σε αντίθεση με την προτιμώμενη αφήγηση του troll.

Η κατανόηση του αντίκτυπου της trolling και του doxing είναι ζωτικής σημασίας για την αναγνώριση των διαφόρων τακτικών που χρησιμοποιούνται για τη χειραγώγηση και τον έλεγχο των διαδικτυακών αφηγήσεων. Οι στρατηγικές αυτές όχι μόνο συμβάλλουν στη διάδοση ψευδών πληροφοριών, αλλά δημιουργούν επίσης προκλήσεις για την προώθηση ενός ανοικτού και ενημερωμένου ψηφιακού περιβάλλοντος.

Ενσωμάτωση της διάστασης

Μια σημαντική στρατηγική για τη διάδοση της παραπληροφόρησης και του εξτρεμιστικού περιεχομένου είναι μέσω της ομαλοποίησης ή της «ενσωμάτωσής της». Η έκθεση παίζει καθοριστικό ρόλο σε αυτή τη διαδικασία. Η έκθεση σε εσφαλμένη και παραπληροφόρηση μπορεί να οδηγήσει σε επίμονες παρανοήσεις των ανθρώπων γύρω από τα συγκεκριμένα θέματα, γεγονός που ομαλοποιεί τις ψευδείς ιδέες στο μυαλό τους. Αυτή η έκθεση μπορεί να συμβεί σε διάφορες μορφές.

Επηρεαστές

Μια κοινή μορφή έκθεσης είναι η διάχυση μεταξύ προσώπων, όπου τα άτομα ανταλλάσσουν πληροφορίες με άλλους. Αυτό μπορεί να συμβεί μέσω προσωπικών αλληλεπιδράσεων ή σε μεγαλύτερη κλίμακα με τους influencers στα μέσα κοινωνικής δικτύωσης. Οι influencers, οι οποίοι έχουν σημαντική εμβέλεια μεταξύ διαφορετικών ομάδων, μπορούν εν αγνοία ή σκόπιμα να μοιράζονται ψευδείς πληροφορίες, επηρεάζοντας μεγάλο αριθμό ατόμων. Αυτή η εκτεταμένη έκθεση οδηγεί στην εξομάλυνση της παραπληροφόρησης μεταξύ διαφορετικών ακροατηρίων.

Ειρωνεία, σάτιρα & μιμίδια

Εξτρεμιστικά άτομα και οργανώσεις συχνά ενσωματώνουν χιούμορ, σάτιρα και ειρωνεία για να διαδώσουν τις ιδέες τους.

Η σάτιρα μπορεί να είναι ένα ισχυρό εργαλείο για να αμφισβητήσει καταπιεστικές ιδεολογίες, να μετατοπίσει αφηγήσεις ή να εξομαλύνει τις εξειδικευμένες απόψεις μέσα στο mainstream. Στον τομέα της παραπληροφόρησης, η σάτιρα χρησιμοποιείται σε ένα φάσμα. Ιστοσελίδες παρωδίας όπως το The Onion ή De Speld δημοσιεύουν μη πραγματικό περιεχόμενο για

λόγους χιούμορ, χωρίς να σκοπεύουν να εξαπατήσουν το κοινό. Ωστόσο, ορισμένα άτομα και ομάδες χρησιμοποιούν τη σάτιρα και την ειρωνεία με κακόβουλη πρόθεση να δυσφημήσουν την επικρατούσα δημοσιογραφία, την επιστήμη ή να προωθήσουν εξτρεμιστικές ιδέες και παραπληροφόρηση. Αξιοποιώντας τη σάτιρα και το χιούμορ, ένα τέτοιο περιεχόμενο γίνεται πιο προσιτό και αποδεκτό στον πολιτικό λόγο, εκθέτοντας το σε ένα ευρύτερο κοινό.

Το εξτρεμιστικό περιεχόμενο συχνά απευθύνεται στους νέους ως μορφή ψυχαγωγίας ή αναζήτησης αισθήσεων. Τα νεαρά άτομα, που καθοδηγούνται από μια αναζήτηση νοήματος, τείνουν να στρέφονται προς έντονες και νέες εμπειρίες, καθιστώντας τους πιο επιρρεπείς σε εξτρεμιστικές ιδέες και στη σχετική παραπληροφόρηση.

Τα μιμίδια, τα οποία είναι ευρέως κοινά κομμάτια χιουμοριστικού πολιτιστικού περιεχομένου, χρησιμεύουν ως μια άλλη οδός διάδοσης εξτρεμιστικών ιδεολογιών. Τα μιμίδια έρχονται σε διάφορες μορφές, συμπεριλαμβανομένων εικόνων, βίντεο, ηχητικών κλιπ, emojis και συμβόλων. Ενώ τα ίδια τα μιμίδια δεν είναι εγγενώς επιβλαβή, οι εξτρεμιστές τα χρησιμοποιούν για να ομαλοποιήσουν τις ιδέες τους. Η παιχνιδιάρικη φύση των μιμιδίων επιτρέπει στους εξτρεμιστές να μεταμφιέζονται, να απομιμούνται ή να αρνούνται τη βλαβερότητα των μηνυμάτων τους. Αυτό το «νευρικό» ή προκλητικό περιεχόμενο γίνεται πιο αποδεκτό, και όταν έρχονται αντιμέτωποι με κατηγορίες για σεξισμό, ρατσισμό ή ξενοφοβία, οι δημιουργοί μπορούν εύκολα να το απορρίψουν ως «απλά ένα αστείο». Αυτή η ασάφεια των ορίων μεταξύ παιχνιδιάρικης κακουχίας και προβληματικού περιεχομένου δημιουργεί ασάφεια, καθιστώντας δύσκολο να διακρίνουμε αθώα αστεία από τα εξτρεμιστικά μηνύματα. Ο Πέπτε ο Βάτραχος, ένας χαρακτήρας κινουμένων σχεδίων στο διαδίκτυο που είχε αρχικά δημιουργηθεί για να είναι ένα αβλαβές αστείο, είχε ιδιοποιηθεί από τους σε απευθείας σύνδεση λευκούς ρατσιστές. Αυτό προκάλεσε σύγχυση με τους χρήστες του διαδικτύου, επειδή οι εξτρεμιστικές επαναλήψεις αυτού του meme βρέθηκαν αναμειγμένες με τις ακίνδυνες. Η εξομάλυνση του εξτρεμιστικού περιεχομένου συμβαίνει καθώς περισσότερα άτομα εκτίθενται σε αυτά τα μηνύματα, θολώνοντας τις γραμμές μεταξύ του τι είναι αποδεκτό και τι όχι.

3.4 Παραποιημένο περιεχόμενο

Από τεχνική άποψη, όλες οι πληροφορίες ή «ειδήσεις» στα μέσα ενημέρωσης είναι ένας συνδυασμός κειμένου, εικόνας, ήχου και βίντεο. Ωστόσο, η ανησυχία έγκειται στο να καθοριστεί αν οι πληροφορίες είναι αυθεντικές ή όχι. Αξίζει να σημειωθεί ότι η παραπληροφόρηση μπορεί να χρησιμοποιεί αυθεντικές εικόνες αλλά να χειραγωγεί την ιστορία, διαστρεβλώνοντας τα πραγματικά γεγονότα.

Στο παρελθόν, ήταν ευρέως κατανοητό ότι ο καθένας θα μπορούσε να γράψει ένα κείμενο, ενώ οι εικόνες, ο ήχος και το βίντεο υποτίθεται ότι είναι περισσότερο ή λιγότερο αυθεντικές, απαιτώντας επαγγελματικές δεξιότητες για τροποποιήσεις. Ωστόσο, με τις πρόσφατες τεχνολογικές εξελίξεις, ακόμη και οι εικόνες, ο ήχος και το βίντεο μπορεί να υπόκεινται σε χειραγώγηση. Αυτό μπορεί να επιτευχθεί από επαγγελματίες ή μέσω εφαρμογών που χρησιμοποιούν συστήματα τεχνητής νοημοσύνης, όπως η τεχνολογία deepfake. Κατά συνέπεια, είναι αναγκαίο να γίνει διάκριση μεταξύ του κατά πόσον ένας ήχος ή βίντεο έχει πράγματι ληφθεί από μικρόφωνο ή βιντεοκάμερα, ή εάν είναι αποτέλεσμα συστημάτων επεξεργασίας εμπειρογνομόνων ή τεχνητής νοημοσύνης (TN) (γεννητική TN και συνθετικά μέσα: κλώνοι φωνής, **deepfake** videos). Επιπλέον, θα πρέπει να είναι τεχνικά εφικτό να προσδιοριστεί η αρχική πηγή, ο παραγωγός ή ο εκδότης μιας πηγής εικόνας, ήχου ή βίντεο. Αυτό συμβαίνει επειδή οι πόροι μπορούν να μοιραστούν, να αντιγραφούν ή να αναδιανεμηθούν πολλές φορές σε όλο τον Παγκόσμιο Ιστό και στα μέσα κοινωνικής δικτύωσης. Ως εκ τούτου, για τους τακτικούς χρήστες, μπορεί να είναι δύσκολο να προσδιοριστεί η αρχική πηγή και παραγωγός, ακόμη και αν υποψιάζονται ότι οι ειδήσεις μπορεί να είναι παραπληροφόρηση.

Η ενδυνάμωση των χρηστών να διακρίνουν μεταξύ πραγματικών και κατασκευαζόμενων ειδήσεων απαιτεί διάφορα βήματα. Περισσότερες πληροφορίες σχετικά με αυτό μπορείτε να βρείτε στο *κεφάλαιο 2: Δράσεις & Δεξιότητες*. Ενθαρρύνεται να διαβάσει αυτό το κεφάλαιο για να μάθει τις περίπλοκες λεπτομέρειες πίσω από την αναγνώριση ψευδών πληροφοριών. Ωστόσο, εδώ είναι μια σύντομη, πολύ απλοποιημένη περίληψη:

- **Βήμα 1: Ενίσχυση της ευαισθητοποίησης των χρηστών:** Οι χρήστες πρέπει να γνωρίζουν ότι τα νέα μπορούν να κατασκευαστούν. Η διεξαγωγή δραστηριοτήτων ευαισθητοποίησης είναι ζωτικής σημασίας για την ενημέρωση των χρηστών σχετικά με το τι συνιστά κατασκευασμένη είδηση και πώς μπορούν να προστατευτούν από τις συνέπειές τους.
- **Βήμα 2: Επαλήθευση αξιοπιστίας εκδοτών:** Η αυξημένη ευαισθητοποίηση σχετικά με τις κατασκευασμένες ειδήσεις ωθεί τους χρήστες να αμφισβητήσουν την αξιοπιστία των πηγών ειδήσεων και των εκδοτών. Είναι σημαντικό να εξεταστεί η μορφή των μέσων ενημέρωσης, είτε πρόκειται για τηλεοπτικό κανάλι, περιοδικό, εφημερίδα (online ή έντυπη), είτε για πλατφόρμα μέσων κοινωνικής δικτύωσης. Τα κανάλια μέσων ενημέρωσης που επιτρέπουν την εύκολη ανταλλαγή ή αναδιανομή ειδήσεων μπορεί να είναι λιγότερο αξιόπιστα. Αντίθετα, οι δίαυλοι μέσων ενημέρωσης που διευκολύνουν τον εντοπισμό και την επαλήθευση της πηγής τείνουν να είναι πιο αξιόπιστοι.

Όσον αφορά τις ειδήσεις που διανέμονται μέσω του διαδικτύου, οι πάροχοι υπηρεσιών όπως τα News Feeds και τα Social Media Networks θα πρέπει να αξιοποιήσουν τις αναδυόμενες τεχνολογίες για την επαλήθευση της αξιοπιστίας των πηγών και την παρακολούθηση του αρχικού εκδότη και πηγής. Η τεχνολογία blockchain είναι μια τέτοια τεχνολογία που μπορεί να διευκολύνει αυτές τις προσπάθειες.

Ακολουθώντας αυτά τα βήματα και αξιοποιώντας την τεχνολογία, οι χρήστες μπορούν να έχουν τη δυνατότητα να περιηγούνται στο ψηφιακό τοπίο, να διακρίνουν τα πραγματικά νέα από την παραπληροφόρηση και να λαμβάνουν τεκμηριωμένες αποφάσεις σχετικά με τις πληροφορίες που συναντούν.

Κατανόηση deepfakes: χειρισμός συνθετικών μέσων

Τα deepfakes, όπως ορίζονται από το Cambridge Dictionary, είναι «*ταβίντεο ή οι ηχογραφήσεις που αντικαθιστούν το πρόσωπο ή τη φωνή κάποιου με αυτό κάποιου άλλου, με τρόπο που φαίνεται πραγματικός*».

Στο άρθρο «Deepfake εξήγησε» από το 2020, η συγγραφέας Meredith Somers αναφέρει ότι «*α) το deepfake αναφέρεται σε ένα συγκεκριμένο είδος συνθετικών μέσων όπου ένα άτομο σε μια εικόνα ή βίντεο ανταλλάσσεται με την ομοιότητα ενός άλλου ατόμου. Επιπλέον, εξηγείται ότι «ο όρος «deepfake» επινοήθηκε για πρώτη φορά στα τέλη του 2017 από έναν χρήστη του Reddit με το ίδιο όνομα. Αυτός ο χρήστης δημιούργησε ένα χώρο στον διαδικτυακό ιστότοπο ειδήσεων και συγκέντρωσης, όπου μοιράστηκε πορνογραφικά βίντεο που χρησιμοποιούσαν τεχνολογία ανταλλαγής προσώπου ανοικτού κώδικα.*»

Τα deepfakes έχουν βρει εφαρμογές σε διάφορους τομείς και έχουν χρησιμοποιηθεί για διαφορετικούς σκοπούς. Μερικά αξιοσημείωτα παραδείγματα περιλαμβάνουν:

- **Εκβιασμός:** Τα deepfakes μπορούν να χρησιμοποιηθούν για την παραγωγή ψευδούς ενοχοποιητικού υλικού, οδηγώντας ενδεχομένως σε εκβιασμό. Επιπλέον, καθώς γίνεται όλο και πιο δύσκολο να διακριθεί το deepfakes από το γνήσιο περιεχόμενο, τα θύματα πραγματικού εκβιασμού μπορούν να ισχυριστούν ότι τα αποδεικτικά στοιχεία είναι ψευδή, παρέχοντάς τους εύλογη δυνατότητα άρνησης.
- **Πορνογραφία:** Η βαθιά ψεύτικη πορνογραφία έχει αποκτήσει σημαντική σημασία στο διαδίκτυο. Μια έκθεση της ολλανδικής εκκίνησης στον κυβερνοχώρο Deeptrace εκτιμά ότι περίπου το 96 % όλων των online deepfakes ήταν πορνογραφικό.
- **Πολιτική:** Τα deepfakes έχουν χρησιμοποιηθεί για να παραποιήσουν γνωστούς πολιτικούς σε βίντεο, διαδίδοντας παραπληροφόρηση και χειραγωγώντας την αντίληψη του κοινού. Παραδείγματα περιλαμβάνουν τους Barack Obama, Donald Trump, Volodymyr Zelenskyy και Vladimir Putin.
- **Υποκριτική/φίλτρα:** Υπάρχουν εικασίες σχετικά με τη χρήση deepfakes για τη δημιουργία ψηφιακών ηθοποιών σε μελλοντικές ταινίες. Ενώ οι ψηφιακά

κατασκευασμένοι ή αλλοιωμένοι άνθρωποι έχουν παρουσιαστεί σε ταινίες στο παρελθόν, τα deepfakes θα μπορούσαν να συμβάλουν σε νέες εξελίξεις σε αυτόν τον τομέα.

- **Μέσα κοινωνικής δικτύωσης:** Τα deepfakes έχουν χρησιμοποιηθεί από τους χρήστες σε διάφορες πλατφόρμες κοινωνικών μέσων. Τα άτομα αντικαθιστούν τα πρόσωπα σε δημοφιλείς σκηνές ταινιών ή σειρών με τα δικά τους, δημιουργώντας εξατομικευμένα βίντεο. Πλατφόρμες όπως το Facebook έχουν λάβει μέτρα για τον εντοπισμό και την επισήμανση των deepfakes ως ψεύτικων, μειώνοντας την προτεραιότητά τους στις ροές των χρηστών.

Ανίχνευση και μετριασμός των deepfakes: τεχνολογικές προσεγγίσεις

Οι ερευνητές διερευνούν ενεργά μεθόδους για τον εντοπισμό και την αντιμετώπιση του ζητήματος του deepfake ήχου και βίντεο. Ακολουθούνται διάφορες προσεγγίσεις:

- **Αλγοριθμική ανίχνευση:** Μια προσέγγιση περιλαμβάνει την ανάπτυξη αλγορίθμων που μπορούν να αναγνωρίσουν το επεξεργασμένο περιεχόμενο. Αυτοί οι αλγόριθμοι αναλύουν διάφορες οπτικές και ακουστικές ενδείξεις για να ανιχνεύσουν ασυνέπειες ή ανωμαλίες που υποδεικνύουν την παρουσία ενός deepfake. Αξιοποιώντας τις τεχνικές μηχανικής μάθησης και τεχνητής νοημοσύνης, αυτοί οι αλγόριθμοι μπορούν να βελτιώσουν τις δυνατότητες ανίχνευσης τους με την πάροδο του χρόνου.
- **Τεχνολογία blockchain:** Μια άλλη τεχνική προτείνει τη χρήση τεχνολογίας blockchain για την επαλήθευση της πηγής των μέσων ενημέρωσης. Το blockchain είναι ένα ψηφιακό καθολικό που καταγράφει τις συναλλαγές σε ένα δίκτυο υπολογιστών με ασφαλή, διαφανή και ανθεκτικό στην παραποίηση τρόπο. Χρησιμοποιεί την αποκέντρωση και την κρυπτογραφία για να εξασφαλίσει την εμπιστοσύνη χωρίς να χρειάζεται μια κεντρική αρχή. Σε αυτό το σενάριο, τα βίντεο θα πρέπει να υποβληθούν σε επαλήθευση μέσω ενός βιβλίου blockchain πριν εμφανιστούν σε πλατφόρμες κοινωνικών μέσων. Με τη διασφάλιση ότι εγκρίνονται μόνο βίντεο από αξιόπιστες πηγές, η εξάπλωση δυνητικά επιβλαβών deepfake media θα μπορούσε να μειωθεί.
- **Ψηφιακές υπογραφές:** Ορισμένοι προτείνουν την ψηφιακή υπογραφή όλων των βίντεο και εικόνων που λαμβάνονται από κάμερες, συμπεριλαμβανομένων των καμερών smartphone, ως μέσο για την καταπολέμηση των deepfakes. Αυτό θα περιλαμβάνει την απόδοση μοναδικών ψηφιακών υπογραφών σε κάθε κομμάτι των μέσων, επιτρέποντας την ιχνηλάτηση κάθε φωτογραφίας ή βίντεο πίσω στον αρχικό ιδιοκτήτη της. Ενώ αυτή η προσέγγιση μπορεί να βοηθήσει στην παρακολούθηση της προέλευσης του περιεχομένου, υπάρχουν ανησυχίες σχετικά με πιθανή κατάχρηση, όπως η στόχευση αντιφρονούντων ή η παραβίαση της ιδιωτικής ζωής.

3.5 Αναφορές

- Aro, J. (2016). The Cyberspace War: Propaganda and Trolling as Warfare Tools. *European View*, 15(1), 121–132. <https://doi.org/10.1007/s12290-016-0395-5>
- Cambridge English Dictionary: Meanings & Definitions*. (2023). <https://dictionary.cambridge.org/dictionary/english>
- Daniels, J. (2018). The Algorithmic Rise of the “Alt-Right.” *Contexts*, 17(1), 60–65. <https://doi.org/10.1177/1536504218766547>
- Egelhofer, J. L., & Lecheler, S. (2019). Fake news as a two-dimensional phenomenon: a framework and research agenda. *Annals of the International Communication Association*, 43(2), 97–116. <https://doi.org/10.1080/23808985.2019.1602782>
- Greene. (2019). “Deplorable” Satire: Alt-Right Memes, White Genocide Tweets, and Redpilling Normies. *Studies in American Humor*, 5(1), 31–69. <https://doi.org/10.5325/studamerhumor.5.1.0031>
- Hardaker, C. (2013). “Uh. . . . not to be nitpicky,,,,,but. . .the past tense of drag is dragged, not drug.” *Journal of Language Aggression and Conflict*, 1(1), 58–86. <https://doi.org/10.1075/jlac.1.1.04har>
- IED. (2022, August 23). *How Do Social Media Algorithms Work*. Institute of Entrepreneurship Development. <https://ied.eu/blog/technology-blog/how-do-social-media-algorithms-work/>
- Johnson, D., & Johnson, A. (2023, June 15). What are deepfakes? How fake AI-powered audio and video warps our perception of reality. *Business Insider*. <https://www.businessinsider.com/guides/tech/what-is-deepfake?international=true&r=US&IR=T>
- Levy, G., & Razin, R. (2019). Echo Chambers and Their Effects on Economic and Political Outcomes. *Annual Review of Economics*, 11, 303–328. <https://doi.org/10.1146/annurev-economics-080218-030343>
- Lewis, B., & Marwick, A. E. (2017). Media Manipulation and Disinformation Online. *Data & Society Research Institute*. <https://www.posiel.com/wp-content/uploads/2016/08/Media-Manipulation-and-Disinformation-Online-1.pdf>

- McNealy, J. (2015). Readers react negatively to disclosure of poster's identity. *Newspaper Research Journal*, 38(3).
<https://doi.org/10.1177/0739532917722977>
- Munn, L. (2019). Alt-right pipeline: Individual journeys to extremism online. *First Monday*. <https://doi.org/10.5210/fm.v24i6.10108>
- Sample, I. (2020, January 13). What are deepfakes – and how can you spot them? *The Guardian*. <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>
- Schumpe, B. M., Bélanger, J. J., Moyano, M., & Nisa, C. F. (2020). The role of sensation seeking in political violence: An extension of the Significance Quest Theory. *Journal of Personality and Social Psychology*, 118(4), 743–761.
<https://doi.org/10.1037/pspp0000223>
- Seth, S. (2023, September 11). The World's Top 10 News Media Companies. *Investopedia*. <https://www.investopedia.com/stock-analysis/021815/worlds-top-ten-news-companies-nws-gci-trco-nyt.aspx>
- Somers, M. (2020, July 21). Deepfakes, explained. *MIT Sloan*.
<https://mitsloan.mit.edu/ideas-made-to-matter/deepfakes-explained>
- Tandoc, E. C., Lim, Z. W., & Ling, R. (2017). Defining “Fake news”: A Typology of Scholarly Definitions. *Digital Journalism*, 6(2), 137–153.
<https://doi.org/10.1080/21670811.2017.1360143>
- Van Puffelen, M. (2021). Rechtsextremisme: Geweld met een rechtsextremistisch motie. In *DSP-groep*. DSP-groep. <https://www.dsp-groep.nl/wp-content/uploads/18MP-Rechtsextremisme-DSP-2021.pdf>
- Van Wonderen, R. (2023). *Rechts-extremistische Radicalisering op Sociale Media Platformen*. Verwey-Jonker Instituut.
- Van Wonderen, R. (2023). *Richtlijn / onderbouwing Radicalisering*. Verwey-Jonker Instituut.
- Van Wonderen, R. & Peeters, M. (2021). *Werken aan weerbaarheid tegen desinformatie en eenzijdige meningsvorming. Evaluatie lesprogramma Under Pressure*. Utrecht: Verwey-Jonker Instituut. https://www.verwey-jonker.nl/wp-content/uploads/2022/04/120550_Werken-aan-weerbaarheid-tegen-desinformatie-eenzijdige-meningsvorming.pdf.

Wasike, B. (2022). When the influencer says jump! How influencer signaling affects engagement with COVID-19 misinformation. *Social Science & Medicine*, 315, 115497. <https://doi.org/10.1016/j.socscimed.2022.115497>

Wolfowicz, M., Weisburd, D., & Hasisi, B. (2021). Examining the interactive effects of the filter bubble and the echo chamber on radicalization. *Journal of Experimental Criminology*, 19(1), 119–141. <https://doi.org/10.1007/s11292-021-09471-0>