EFCSN

Πως να εντοπίσεις και να καταλάβεις περιεχόμενο δημιουργημένο με ΤΝ και ψηφιακά τροποποιημένο περιεχόμενο

Τα προσβάσιμα και εύχρηστα μοντέλα τεχνητής νοημοσύνης μπορούν να βοηθήσουν τους ανθρώπους να μάθουν και να δημιουργούν περιεχόμενο, αλλά μπορούν επίσης να **ενισχύσουν τους κινδύνους που η σκόπιμη ή μη παραπληροφόρηση θέτει** για τις ανοιχτές κοινωνίες και τον δημοκρατικό διάλογο. Είναι σημαντικό να αποτρέψουμε το ενδεχόμενο να γεμίσουν οι κοινοί χώροι πληροφόρησής μας με παραπληροφόρηση που πηγάζει από υλικό που παράγεται χάρη στη χρήση τεχνητής νοημοσύνης και τροποποιείται ψηφιακά.

Μέρος της λύσης είναι οι νέες τεχνολογίες, όπως οι πρωτοβουλίες για τον εντοπισμό της προέλευσης περιεχομένου και τα λογισμικά ανίχνευσης. **Ωστόσο, οι τεχνολογικές λύσεις απέχουν πολύ από το να είναι τέλειες και χρειαζόμαστε το έργο των ανεξάρτητων ελεγκτών γεγονότων για να παρέχουμε στην κοινωνία ένα κοινό σύνολο επαληθευμένων πληροφοριών.**

Ακολουθεί/Εδώ είναι μια σύντομη επισκόπηση του τι κάνουν οι επαγγελματίες, ανεξάρτητοι fact-checkers για να εντοπίσουν και να καταρρίψουν παραπληροφόρηση που δημιουργείται με τη χρήση τεχνητής νοημοσύνης και τι μπορείτε να μάθετε από αυτούς.

Το περιεχόμενο που δημιουργείται με τη χρήση Τεχνητής Νοημοσύνης αυξάνεται

Σήμερα, η παραπληροφόρηση - σκόπιμη ή μη - που δημιουργείται με τη χρήση τεχνητής νοημοσύνης αποτελεί ένα μικρό ποσοστό των ισχυρισμών που ελέγχονται από ανεξάρτητους, επαγγελματίες ελεγκτές γεγονότων. Είναι το ψηφιακά τροποποιημένο περιεχόμενο αυτό με το οποίο οι fact-checkers έρχονται πιο συχνά αντιμέτωποι στη δουλειά τους.

Ωστόσο, σε μια εσωτερική έρευνα των μελών του EFCSN, **οι περισσότεροι ελεγκτές γεγονότων συμφώνησαν ότι το περιεχόμενο που δημιουργείται από τεχνητή νοημοσύνη και το ψηφιακά τροποποιημένο περιεχόμενο θα είναι πιο συχνά και σχετικά με τη δουλειά τους στο μέλλον**. Πρόσφατα παραδείγματα στο πλαίσιο των Ευρωπαϊκών Εκλογών υποστηρίζουν αυτήν την πρόγνωση.

ΚΑΤΑΝΟΗΣΕ: Το *ψηφιακά τροποποιημένο περιεχόμενο* αναφέρεται σε οποιαδήποτε μορφή περιεχομένου που έχει αλλοιωθεί σημαντικά για να χειραγωγήσει ή να αλλάξει το μήνυμα που αρχικά μετέφερε. Αυτό μπορεί να περιλαμβάνει και τις επεξεργασίες που μπορεί να έχουν γίνει από εργαλεία τεχνητής νοημοσύνης, ωστόσο δεν αφορά τροποποιήσεις που μπορεί να έχουν γίνει για λόγους σαφήνειας ή βελτίωσης της ποιότητας.

Ο όρος *AI-generated* αναφέρεται σε κάθε μορφή περιεχομένου που έχει δημιουργηθεί από κάποιο σύστημα τεχνητής νοημοσύνης.

Οι πληροφορίες παρέχονται από το **AI@EUElections**, έργο το οποίο χρηματοδοτείται από τη Meta και οργανώνεται από το **European Fact-Checking Standards Network** σε συνεργασία με τα μέλη του και εξωτερικούς ειδικούς.

Η τεχνολογία εξελίσσεται γρήγορα, αλλά δεν μπορούμε να βασιζόμαστε μόνο σε αυτήν.

Ειδικοί στο πεδίο της τεχνητής νοημοσύνης και επαγγελματίες ελεγκτές γεγονότων συμφωνούν: **Από μόνα τους τα εργαλεία εντοπισμού υλικού που έχει δημιουργηθεί με τεχνητή νοημοσύνη δεν είναι αρκετά για να εντοπίσουν ή να καταρρίψουν τέτοιο περιεχόμενο ή περιεχόμενο που έχει τροποποιηθεί ψηφιακά.**

Πριν από τη χρήση ενός εργαλείου εντοπισμού περιεχομένου που έχει παραχθεί με ΤΝ, οι ειδικοί συνιστούν την εξοικείωση με τα συστήματα που δημιουργούν και εντοπίζουν τέτοιο περιεχόμενο. Με την κατανόηση του τρόπου εκπαίδευσης των μοντέλων που χρησιμοποιούν τέτοια εργαλεία και με λίγη στατιστική, οι ελεγκτές γεγονότων μπορούν να αρχίσουν να αναγνωρίζουν τα δυνατά και αδύνατα σημεία τους καθώς και την πιθανότητα επιτυχίας τους. **Ακόμα και έτσι, τα εργαλεία αυτά μπορούν να αποτελέσουν ένα χρήσιμο σημείο εκκίνησης.**

Πρωτοβουλίες που επικεντρώνονται στην εξακρίβωση της **προέλευσης του περιεχομένου**, όπως οι προδιαγραφές του Συνασπισμού για την Προέλευση και Αυθεντικότητα Περιεχομένου (C2PA), μπορούν να βοηθήσουν στην πιστοποίηση της πηγής και του ιστορικού επεξεργασίας ψηφιακού υλικού. Σε κάθε περίπτωση η χρήση υδατογραφημάτων και οι συμβατικές μέθοδοι επαλήθευσης δεν αποτελούν πανάκεια.

Πώς η παραγόμενη από την τεχνητή νοημοσύνη παραπληροφόρηση επηρεάζει τους ανθρώπους

«Κάθε φορά που αντιδράς με το ένστικτό σου, παρακάμπτεται η σκέψη σου.»
- Christine Dugoin*

ΣΤΟΧΟΙ: Γιατί οι κακόβουλοι παράγοντες μπορεί να βασίζονται στην τεχνητή νοημοσύνη για να δημιουργήσουν ή να διαδώσουν παραπληροφόρηση; Ποια είναι η επιδιωκόμενη επίδρασή τους στον πραγματικό κόσμο;

ΨΥΧΟΛΟΓΙΑ: Οι εκστρατείες επιρροής της κοινής γνώμης συχνά σχεδιάζονται για να εκμεταλλευτούν ασυνείδητες προκαταλήψεις.

Η κατανόηση των δικών σας, και του κοινού σας, μπορεί να βοηθήσει στην αντιμετώπιση της παραπληροφόρησης.

- ➔ Να επεκτείνουν την εμβέλειά τους σε μια άλλη χώρα ή κοινότητα;
- ➔ Να αποφύγουν τον εντοπισμό ή να κατακλύσουν τους ελεγκτές γεγονότων δημιουργώντας πολλές παραλλαγές παρόμοιων ισχυρισμών;
- ➔ Να επηρεάσουν σκέψεις ή πεποιθήσεις κατασκευάζοντας αξιοπιστία μέσω δικτύων ψεύτικων λογαριασμών;

** Η Christine Dugoin είναι ερευνήτρια στον τομέα της πληροφοριακής επιρροής στο Πανεπιστήμιο Παντεόν-Σορμπόν .*

Η αποδόμηση απαιτεί μια πολύπλευρη προσέγγιση και λεπτομερή κατανόηση

Αν λοιπόν τα εργαλεία εντοπισμού δεν λειτουργούν, τι λειτουργεί; Είναι σημαντικό να κατανοούμε το πλαίσιο ενός ισχυρισμού όσο και το περιεχόμενό του. Οι επαγγελματίες ελεγκτές γεγονότων είναι ειδικοί στις απαραίτητες πρακτικές έρευνας. Εδώ είναι μερικές συμβουλές.

Τα εργαλεία εντοπισμού δεν θα λειτουργήσουν ποτέ στο 100% - δεν περιμένω να το κάνουν ποτέ.»
– Henk van Ess**

ΛΑΒΕ ΥΠΟΨΗ ΤΗΝ ΠΗΓΗ: Μπορείς να επιβεβαιώσεις την ταυτότητά τους; Τι συζητούν και τι μοιράζονται; Ποιοι αλληλεπιδρούν με το περιεχόμενό τους; Ποια επίδραση μπορεί να έχει αυτό το περιεχόμενο στους αναγνώστες;

ΕΛΕΓΞΕ ΤΗΝ ΑΞΙΟΠΙΣΤΙΑ: Επαλήθευσε ανεξάρτητα τις πληροφορίες με αξιόπιστες πηγές, όπως ειδικούς με πρακτική εμπειρία στον τομέα. Βγάζει βάσει των γνώσεών σου νόημα αυτό που απεικονίζεται;

Χρησιμοποίησε **ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ ΨΗΦΙΑΚΩΝ ΜΕΣΩΝ** για να συμπληρώσεις το παραδοσιακό ερευνητικό ρεπορτάζ και την αρχειακή έρευνα. Μερικές τεχνικές περιλαμβάνουν: τη συλλογή δεδομένων, τον γεωεντοπισμό, τη βιομετρική αναγνώριση, την ανάλυση μοτίβων και άλλα.

ΜΑΘΕ & ΠΡΟΣΑΡΜΟΣΟΥ: Όσοι δημιουργούν παραπληροφόρηση με τη χρήση τεχνητής νοημοσύνης προσαρμόζονται συνεχώς. Προσάρμοσε την προσέγγισή σου στο μεταβαλλόμενο αυτό τοπίο.

ΜΟΙΡΑΣΟΥ ΤΗ ΔΟΥΛΕΙΑ ΣΟΥ

Μαζί με έναν καταρριπτόμενο ισχυρισμό, οι ειδικοί συνιστούν την παροχή μιας διαφανούς ανάλυσης και συνδέσμων προς τις πηγές. Αυτό μπορεί να βοηθήσει τους αναγνώστες να ακολουθήσουν την έρευνα και να κατανοήσουν ένα περίπλοκο αφήγημα.

*** Ο Henk van Ess είναι ειδικός στις τεχνικές OSINT και στον έλεγχο γεγονότων.*

Γρήγορος Οδηγός Αναφοράς: Τι να Κάνεις, Τι να Μην Κάνεις και Χρήσιμα Στοιχεία

Ακολουθούν κάποιες ενδείξεις που μπορεί να υποδηλώνουν ότι ένα περιεχόμενο έχει παραχθεί με τη χρήση τεχνητής νοημοσύνης ή έχει τροποποιηθεί ψηφιακά. Μαζί με τις άλλες συμβουλές που αναφέρονται σε αυτόν τον οδηγό (πλαίσιο, τεχνικές διερεύνησης και εργαλεία εντοπισμού), μπορούν να σε βοηθήσουν να καταλάβεις την αλήθεια πίσω από αυτό που βλέπεις.

Κείμενο

- Συχνά (αλλά όχι πάντα) έχει **καλύτερη γραμματική** σε σχέση κείμενα που έχουν γραφτεί από ανθρώπους.
- Πιθανόν να χρησιμοποιεί **υπερβολικά επίσημη ή δομημένη γλώσσα**, ειδικά για το πλαίσιο των μέσων κοινωνικής δικτύωσης.
- **Υπερβολική χρήση επιρρημάτων ή επιθέτων.**
- Έλλειψη ανθρώπινου συναισθήματος, χιούμορ, σαρκασμού και ιδιωματικών εκφράσεων.
- Μπορεί να **στερείται συγκεκριμένων λεπτομερειών** (ονόματα, ημερομηνίες, τοποθεσίες) ή πρωτότυπων ιδεών.
- Το πιο σημαντικό: είναι σωστά τα γεγονότα που αναφέρονται στο κείμενο;

Βίντεο

- Μην χρησιμοποιείς εργαλεία που προορίζονται για ανίχνευση εικόνων που έχουν δημιουργηθεί από ΤΝ για στατικές εικόνες που έχουν απομονωθεί από βίντεο.
- Παρατήρησε **τις εκφράσεις και την κίνηση του προσώπου**, όπως το ανοιγοκλείσιμο των ματιών και το κατά πόσο η κίνηση του στόματος ταιριάζει με τον ήχο.
- **Απότομες μεταβάσεις ή κοψίματα** μπορεί να είναι μερικά από τα χαρακτηριστικά τους.

Ήχος

- **Σύγκρινε ύποπτα ηχητικά αποσπάσματα με ένα αυθεντικό δείγμα** χρησιμοποιώντας εργαλεία που μπορούν να ανιχνεύσουν διαφορές στα μοτίβα ομιλίας και αναπνοής, στον τονισμό...
- Όταν χρησιμοποιείς εργαλεία ανίχνευσης περιεχομένου που έχει παραχθεί με τη χρήση ΤΝ, απέφυγε δείγματα ήχου χαμηλής ποιότητας με στατικό θόρυβο ή θόρυβο στο παρασκήνιο.
- Μπορεί να χαρακτηρίζονται από **αφύσικα ή μηχανικά μοτίβα ομιλίας**, έλλειψη παύσεων ή φυσικής αναπνοής.

Εικόνες

- Αναζήτησε περιοχές με **αφύσικες λεπτομέρειες**: τέλειο δέρμα, θολωμένα φόντα, αφύσικη ομορφιά ή φως, και παράδοξα όπως επιπλέον δάχτυλα.
- **Αναζήτησε υδατογράφημα** από δημοφιλή εργαλεία δημιουργίας εικόνων με ΤΝ.
- Δώσε προσοχή στις λεπτομέρειες του τι απεικονίζεται: είναι λογικό; Είναι αναμενόμενο;
- Όταν χρησιμοποιείς εργαλεία για τον εντοπισμό εικόνων που έχουν δημιουργηθεί με ΤΝ, **επίλεξε μια εκδοχή της εικόνας σε υψηλή ανάλυση** ή μια εκδοχή της εικόνας από τις πρώτες μεταφορτώσεις αυτής αντί για μια εικόνα που έχει κοινοποιηθεί και επαναχρησιμοποιηθεί.