

# Αναζήτηση μεθόδων διάκρισης γραπτών από λογισμικό τεχνητής νοημοσύνης

**Σ**ύντομα, οι χρήστες του Διαδικτύου θα διαβάζουν ένα μήνυμα στο Twitter, ένα δοκίμιο ή μία είδηση και θα αναρωτιούνται αν γράφτηκε από λογισμικό τεχνητής νοημοσύνης. Για να αντιμετωπισθεί το φαινόμενο αυτό, νέα εργαλεία για την αναγνώριση κειμένων γραμμένων από μηχανές τεχνητής νοημοσύνης έχουν δημιουργηθεί τους τελευταίους μήνες, ένα από αυτά σχεδιασμένο μάλιστα από την εταιρεία OpenAI, δημιουργό του λογισμικού ChatGPT.

Το εργαλείο της OpenAI αξιοποιεί την τεχνητή νοημοσύνη προκειμένου να εντοπίζει διαφορές μεταξύ κειμένων γραμμένων από ανθρώπους και αυτών που παρήχθησαν από μηχανές. Όταν η εταιρεία δοκίμασε το νέο σύστημα, αυτό κατάφερε να αναγνωρίσει με ακρίβεια εμπλοκή τεχνητής νοημοσύνης σε μόλις 50% των κειμένων που ανέλυσε, το σύνολο των οποίων είχε προέλθει από λογισμικό τεχνητής νοημοσύνης.

Ο εντοπισμός κειμένου παραγόμενου από μηχανές τεχνητής νοημοσύνης έχει καταστεί ολοένα και πιο δύσκολος, σύμφωνα με ειδικούς, καθώς λογισμικό όπως το ChatGPT συνεχίζει να προοδεύει, δημιουργώντας κείμενα ικανά να πείσουν ότι γράφτηκαν από άνθρωπο μυαλό. Η OpenAI πειραματίζεται τώρα με τεχνολογία η οποία θα εισάγει ειδικές λέξεις σε κείμενα του ChatGPT, καθιστώντας ευκολότερο τον εντοπισμό τους. Η τεχνική αυτή ονομάζεται υδατογράφημα, όπως αυτό που υπάρχει στα χαρτονομίσματα.

Η μέθοδος του υδατογραφήματος εγγυάται ότι οι «λέξεις-κλειδιά» στο κείμενο



**Ο εντοπισμός κειμένου παραγόμενου από μηχανές τεχνητής νοημοσύνης έχει καταστεί ολοένα και πιο δύσκολος, σύμφωνα με ειδικούς, καθώς λογισμικό όπως το ChatGPT συνεχίζει να προοδεύει**

τεχνητής νοημοσύνης δεν θα μπορούν να εντοπισθούν από άνθρωπο, οπότε δεν θα μπορεί να γίνει παρέμβαση. Ακόμη και αν κάποιος πετύχει να αλλάξει κάποιες από τις ειδικές λέξεις, το κείμενο της τεχνητής νοημοσύνης θα συνεχίσει να αναγνωρίζεται ως τέτοιο από το λογισμικό του υδατογραφήματος.

Ο καθηγητής του Πανεπιστημίου του Μέριλαντ και συντάκτης της μελέτης για τη μέθοδο αυτή, Τομ Γκόλντσταϊν, λέει ότι το υδατογράφημα μπορεί να εντοπισθεί ακόμη και χάρη σε «πολύ σύντομο θραύσμα κειμένου», όπως μέρος ενός tweet. Το εργαλείο εντοπισμού της OpenAI, αντίθετα, απαιτεί το λιγότερο 1.000 χαρακτήρες (150 λέξεις) για να αποφασίσει εάν το κείμενο γράφτηκε από άνθρωπο.

Όπως κάθε εργαλείο εντοπισμού, όμως, έτσι και η μέθοδος του υδατογραφήματος δεν είναι τέλεια, παραδέχεται ο Γκόλ-

ντσταϊν. Οι περισσότεροι ειδικοί προειδοποιούν και αυτοί από τη μεριά τους ότι θα χρειασθεί πολύς χρόνος έως ότου βρεθεί απολύτως αξιόπιστο εργαλείο εντοπισμού κειμένων παραγόμενων από λογισμικό τεχνητής νοημοσύνης.

«Η ιδέα ότι σύντομα θα υπάρξει κάποιο μαγικό ραβδί, δημιουργία του σχεδιαστή του λογισμικού τεχνητής νοημοσύνης ή κάποιου τρίτου, το οποίο θα διαλύει κάθε αμφιβολία, αποτελεί φαντασίωση. Θα έχουμε πεθάνει προ πολλού όταν θα εφευρεθεί ένα τέτοιο σύστημα», σχολιάζει ο Ντέιβιντ Κοξ, επικεφαλής του εργαστηρίου Ουότσον του MIT στη Βοστώνη. Ο διευθύνων σύμβουλος της OpenAI, Σαμ Αλτμαν, είπε τον περασμένο μήνα: «Μέρος του προβλήματος με τα κείμενα τεχνητής νοημοσύνης είναι ότι η τεχνολογία εντοπισμού τους θα βρίσκεται νομοτελειακά ένα βήμα πίσω από τα νέα γλωσσικά μοντέλα τεχνητής νοημοσύνης, ικανά να διαχειρίζονται περισσότερες γλώσσες και να δίνουν ακόμη και λογοτεχνική χροιά στα κείμενα που παράγουν».

Ορισμένοι ειδικοί πιστεύουν ότι η OpenAI και άλλες εταιρείες του χώρου πρέπει να βρουν λύσεις στο θέμα της αναγνώρισης της γνησιότητας κειμένων πριν από την κυκλοφορία του λογισμικού τους στην αγορά. Η OpenAI κυκλοφόρησε την εφαρμογή ChatGPT τον Νοέμβριο, ενώ το λογισμικό εντοπισμού έγινε διαθέσιμο δύο ολόκληρους μήνες αργότερα. Στο διάστημα αυτό, εκπαιδευτικοί και ερευνητές είχαν ήδη συνειδητοποιήσει την ανάγκη ύπαρξης αξιόπιστων μεθόδων εντοπισμού κειμένου που έχει γραφτεί από μηχανές.

## Έχει τρεις συζύγους να δουλεύουν ενώ κάθεται σπίτι: «Ο βασιλιάς δεν κινείται πολύ»

**Ο** Νικ Ντέιβις, που αποκαλεί τον εαυτό του «σύζυγο τρόπαιο», όχι μόνο είναι παντρεμένος με τρεις γυναίκες, αλλά δεν εργάζεται καν. Συγκρίνοντας τον ρόλο του σε μια πολυγαμική οικογένεια με το σκάκι, ο πατέρας δύο παιδιών λέει ότι δεν έχει δουλειά επειδή «ο βασιλιάς δεν κινείται πολύ», ενώ οι βασίλισσες έχουν όλη τη δύναμη. Στο μεταξύ, οι τρεις σύζυγοί του - Έιπριλ, Ντανιέλ και Τζένιφερ- κάνουν όλες δουλειές πλήρους απασχόλησης.

Ο Νικ γνώρισε την πρώτη του σύζυγο, την Έιπριλ, στο πανεπιστήμιο πριν από 15 χρόνια, πριν έρθει η Τζένιφερ στο γάμο τους εννέα χρόνια αργότερα. Αν και μπορεί να ακούγεται ασυνήθιστο, η 38χρονη Έιπριλ δηλώνει χαρούμενη που έχει επιπλέον γυναίκες στη σχέση της για να καλύπτει τις «ανάγκες» του Νικ. «Ο Νικ είναι δύσκολος για να τον χειριστείς με την προσωπικότητά του. Είναι ωραίο να μπο-

**Για την ακρίβεια, η Έιπριλ ήταν εκείνη που σύστησε τον 39χρονο σύζυγό της στην 25χρονη Τζένιφερ, η οποία ήταν 19 ετών τότε, πιστεύοντας ότι θα ταίριαζε στην οικογένειά τους.**



ρείς να έχεις χέρια βοήθειας», είπε στην εκπομπή Seeking Sister Wife του TLC. Για την ακρίβεια, η Έιπριλ ήταν εκείνη που σύστησε τον 39χρονο σύζυγό της στην 25χρονη Τζένιφερ, η οποία ήταν 19 ετών τότε, πιστεύοντας ότι θα ταίριαζε

στην οικογένειά τους.

«Απλώς μου φάνηκε ότι ήταν ένας σπουδαίος άνθρωπος, υπήρχε υψηλή νοημοσύνη σε αυτό το μυαλό και ήξερα ότι ο Νικ θα την έβρισκε ελκυστική», σημειώνει.