Μεγάλα γλωσσικά μοντέλα και δικανική γλωσσολογία: Ευκαιρίες και απειλές στην εποχή της παραγωγικής τεχνητής νοημοσύνης

Γιώργος Μικρός

Καθηγητής Υπολογιστικής και Ποσοτικής Γλωσσολογίας,
Κολέγιο Ανθρωπιστικών και Κοινωνικών Επιστημών,
Πανεπιστήμιο Hamad Bin Khalifa, Κατάρ

1. Εισαγωγή

Η δικανική γλωσσολογία (forensic linguistics), οριζόμενη ευρέως ως η εφαρμογή της γλωσσικής γνώσης και των γλωσσολογικών μεθόδων σε νομικά προβλήματα, έχει εξελιχθεί σε αναγνωρισμένο επιστημονικό κλάδο από τα τέλη του εικοστού αιώνα (Coulthard & Johnson 2007, McMenamin 2002). Το πεδίο αυτό περιλαμβάνει τρεις αλληλεπικαλυπτόμενους τομείς (Coulthard 2010):

1. Τη γλώσσα του νόμου (language of the law). Ασχολείται με την ανάλυση της γλώσσας των νομικών κειμένων (π.χ. νόμων, συμβολαίων, δικαστικών αποφάσεων) και εξετάζει ζητήματα όπως η σαφήνεια, η ερμηνεία και η προσβασιμότητα του νομικού λόγου για τον απλό πολίτη.

2. η γλώσσα στη δικαστική διαδικασία (language in the legal process). Μελετά τη γλωσσική αλληλεπίδραση στο δικαστήριο και σε άλλα νομικά περιβάλλοντα: εξετάσεις μαρτύρων, ανακρίσεις, διερμηνεία, δικαιώματα κατανόησης του κατηγορουμένου, καθώς και θέματα ισχύος και ασυμμετρίας στον δικαστικό λόγο.

3. Τη γλώσσα ως μαρτυρικό στοιχείο (language as evidence). Αφορά τη χρήση γλωσσολογικής ανάλυσης για αποδεικτικούς σκοπούς, και κυρίως την ταυτοποίηση συγγραφέα (authorship attribution), την ανάλυση απειλητικών μηνυμάτων, πλαστών κειμένων, ομολογιών, καθώς και τη φωνητική ταυτοποίηση ομιλητών.

Κεντρική θέση σε μεγάλο μέρος της δικανικής πρακτικής κατέχει η έννοια της ιδιολέκτου, η ιδέα δηλαδή ότι κάθε άτομο διαθέτει ένα σχετικά σταθερό και διακριτό γλωσσικό «υφομετρικό αποτύπωμα» που εκδηλώνεται στον προφορικό και γραπτό του λόγο (Coulthard 2004).

Αυτή η παραδοχή έχει στηρίξει πολυάριθμες υποθέσεις στις οποίες έχουν προσαχθεί γλωσσικά στοιχεία για να συνδέσουν ή να αποσυνδέσουν έναν ύποπτο και ένα αμφισβητούμενο κείμενο.

Η ραγδαία ανάπτυξη των μεγάλων γλωσσικών μοντέλων (ΜΓΜ - LLMs)[1] αποτελεί τη σοβαρότερη πρόκληση μέχρι σήμερα γι' αυτό το πλαίσιο. Συστήματα βασισμένα στην αρχιτεκτονική μετασχηματιστών[2] (transformer architecture), εκπαιδευμένα σε τεράστια σώματα κειμένων, παράγουν πλέον ρέοντα και κατάλληλα για το εκάστοτε πλαίσιο λόγο, ο οποίος μπορεί να μιμηθεί ένα ευρύ φάσμα γλωσσικών επιπέδων (registers) και υφών (Kumarage & Liu 2023). Η δημόσια διάθεση του ChatGPT τον Νοέμβριο του 2022 και η ταχεία υιοθέτηση παρόμοιων συστημάτων έχουν αυξήσει δραματικά τον όγκο του συνθετικού κειμένου στην καθημερινή επικοινωνία. Μέσα σε δύο μήνες, το ChatGPT προσέλκυσε πάνω από 100 εκατομμύρια ενεργούς μηνιαίους χρήστες, αλλάζοντας την οικολογία της ψηφιακής παραγωγής κειμένου σε παγκόσμια κλίμακα (Liang et al. 2023). Σε αυτό το πλαίσιο, οι παλαιότερες εκκλήσεις για αυστηρή επικύρωση και προσαρμογή της ανάλυσης πατρότητας ως κλάδου των δικανικών επιστημών (forensic science) (Ainsworth & Juola 2019) έχουν αποκτήσει νέα επιτακτικότητα.

Οι επιπτώσεις δεν είναι απλώς θεωρητικές. Η ακριβής απόδοση πατρότητας και η στιβαρή αξιολόγηση των γλωσσικών τεκμηρίων παραμένουν κρίσιμες για τις ποινικές έρευνες, τις αστικές διαφορές, τις διαδικασίες ακαδημαϊκής ακεραιότητας και τον μετριασμό της παραπληροφόρησης (Huang et al. 2024). Ωστόσο, οι ίδιες τεχνολογίες που μπορούν να βοηθήσουν τη δικανική ανάλυση μπορούν επίσης να χρησιμοποιηθούν για την αποφυγή της. Οι συγγραφείς μπορούν πλέον να εξωπορίσουν μέρη της συγγραφικής τους δουλειάς, να συγκαλύψουν την υφολογική τους ιδιόλεκτο ή να μιμηθούν το ύφος άλλων με σχετική ευκολία. Ως αποτέλεσμα, είναι ολοένα πιο δύσκολο να υποτεθεί ότι ένα δεδομένο έγγραφο αποτελεί προϊόν ενός και μόνο ανθρώπινου συγγραφέα που λειτουργεί χωρίς τη βοήθεια μηχανής.

1. Μεγάλα Γλωσσικά Μοντέλα (Large Language Models - LLMs): Συστήματα τεχνητής νοημοσύνης εκπαιδευμένα σε τεράστιο όγκο κειμένων, ικανά να κατανοούν και να παράγουν κείμενο σε φυσική γλώσσα. Γνωστά παραδείγματα: ChatGPT, Claude, Gemini.

2. Αρχιτεκτονική Μετασχηματιστών (Transformer Architecture): Είναι ένας τύπος τεχνητού νευρωνικού δικτύου που παρουσιάστηκε το 2017 από ερευνητές της Google (Vaswani et al. 2017). Σε αντίθεση με προηγούμενες προσεγγίσεις που επεξεργάζονταν το κείμενο λέξη προς λέξη με σειριακό τρόπο, οι μετασχηματιστές αξιοποιούν έναν μηχανισμό που ονομάζεται «προσοχή» (attention mechanism), ο οποίος επιτρέπει στο σύστημα να εξετάζει ταυτόχρονα όλες τις λέξεις μιας πρότασης και να αναγνωρίζει τις μεταξύ τους σχέσεις, ανεξαρτήτως της απόστασής τους στο κείμενο. Αυτή η καινοτομία επέτρεψε την αποτελεσματικότερη κατανόηση του γλωσσικού πλαισίου και αποτέλεσε τη βάση για τη δημιουργία των σύγχρονων μεγάλων γλωσσικών μοντέλων, όπως τα GPT (OpenAI), BERT (Google) και Claude (Anthropic).

Το παρόν άρθρο, επομένως, εξετάζει τα ακόλουθα τρία αλληλένδετα ερευνητικά ερωτήματα:

1. Με ποιους τρόπους μπορούν τα ΜΓΜ να ενισχύσουν τις μεθοδολογίες της δικανικής γλωσσολογίας και να επεκτείνουν τις αναλυτικές ικανότητες των επαγγελματιών του χώρου;

2. Ποιες απειλές θέτουν τα ΜΓΜ στις παραδοσιακές δικανικές πρακτικές, και ιδιαίτερα στην απόδοση πατρότητας βάσει ιδιολέκτου και στην αξιολόγηση κειμενικών τεκμηρίων;

3. Πώς πρέπει να προσαρμοστεί το πεδίο μεθοδολογικά, θεσμικά και νομικά ώστε να διατηρήσει την αξιοπιστία και την αποδεκτότητά του βάσει προτύπων όπως το Daubert στην εποχή της παραγωγικής τεχνητής νοημοσύνης (generative AI);

Για να απαντήσουμε σε αυτά τα ερωτήματα, προσφέρουμε μια κριτική αφηγηματική επισκόπηση της έρευνας για τα ΜΓΜ και τη δικανική γλωσσολογία που δημοσιεύθηκε κυρίως μεταξύ 2023 και 2025, συμπληρωμένη από θεμελιώδη έργα στην ανάλυση της συγγραφικής πατρότητας και τη νομική αποδεκτότητα. Η βιβλιογραφική επισκόπηση δίνει προτεραιότητα σε μελέτες που σχετίζονται άμεσα με εργασίες δικανικής προέλευσης, συμπεριλαμβανομένης της απόδοσης ανθρώπινης πατρότητας, της ανίχνευσης κειμένου ΤΝ, της απόδοσης της πηγής-μοντέλου[3] (model source attribution) και των υβριδικών κειμένων ανθρώπου-ΜΓΜ. Τα άρθρα που αναλύονται εντοπίστηκαν μέσω στοχευμένων αναζητήσεων σε πηγές υπολογιστικής γλωσσολογίας, ψηφιακής εγκληματολογίας και νομικών περιοδικών, με έμφαση σε εργασίες που αναφέρουν εμπειρικές αξιολογήσεις, ποσοστά σφάλματος των μοντέλων πρόβλεψης, προκατάληψη (bias) ή ανθεκτικότητα έναντι αντιθετικών επιθέσεων (adversarial robustness). Δεδομένου του ρυθμού αλλαγής των μοντέλων και της επικράτησης πρόσφατων προδημοσιεύσεων (preprints), τα ευρήματα αντιμετωπίζονται ως ενδεικτικά των τρεχουσών τάσεων και όχι ως αποδείξεις μιας ευρύτερης συναίνεσης της επιστημονικής κοινότητας σε όλα τα θέματα που θίγονται. Η σύνθεση οργανώνεται μέσω μιας διπλής οπτικής: α) χαρτογραφούμε πώς τα ΜΓΜ επεκτείνουν τη δικανική ικανότητα και ταυτόχρονα αναθεωρούν τις παραδοχές που βασίζονται στην ιδιόλεκτο, και β) η νομική συζήτηση βασίζεται στα πρότυπα Daubert/Kumho[4] για τη μετάφραση των τεχνικών περιορισμών σε αποδεικτικές συνέπειες.

3. Απόδοση της πηγής-μοντέλου (model source attribution): Η διαδικασία αναγνώρισης του συγκεκριμένου γλωσσικού μοντέλου τεχνητής νοημοσύνης που παρήγαγε ένα κείμενο. Σε αντίθεση με την απλή ανίχνευση κειμένου ΤΝ, η οποία διακρίνει μόνο αν ένα κείμενο είναι ανθρώπινο ή μηχανικά παραγόμενο, η απόδοση πηγής-μοντέλου στοχεύει στον εντοπισμό του εργαλείου προέλευσης (π.χ. GPT-4, Claude, Gemini, LLaMA). Η εργασία αυτή βασίζεται στην υπόθεση ότι κάθε μοντέλο αφήνει χαρακτηριστικά γλωσσικά «ίχνη» (π.χ. λεξιλογικές προτιμήσεις, συντακτικά μοτίβα ή στατιστικές κανονικότητες) που μπορούν να λειτουργήσουν ως διακριτικά στοιχεία ταυτοποίησης, κατ› αναλογία με την παραδοσιακή απόδοση συγγραφικής πατρότητας σε ανθρώπινα κείμενα.

4. Πρότυπα Daubert/Kumho: Νομικά κριτήρια που εφαρμόζονται στα δικαστήρια των ΗΠΑ για την αξιολόγηση της αποδεκτότητας επιστημονικής και τεχνικής εμπειρογνωμοσύνης ως αποδεικτικού μέσου. Το πρότυπο Daubert (1993) καθόρισε

Το άρθρο διαρθρώνεται ως εξής. Ξεκινάμε περιγράφοντας τα θεωρητικά και μεθοδολογικά θεμέλια της δικανικής γλωσσολογίας, παράλληλα με τα τεχνικά χαρακτηριστικά των ΜΓΜ που σχετίζονται με δικανικές εφαρμογές. Οι επόμενες ενότητες εξετάζουν τις ευκαιρίες που δημιουργούν τα ΜΓΜ ως αναλυτικά εργαλεία, τις απειλές που θέτουν στις υπάρχουσες πρακτικές και στις τρέχουσες στρατηγικές ανίχνευσης και αντίμετρων. Μια ειδική ενότητα συνθέτει αυτές τις πτυχές, παρουσιάζοντας τις συνέπειες για τους επαγγελματίες και τα νομικά συστήματα, τους περιορισμούς της παρούσας ανάλυσης και τις προτεραιότητες για μελλοντική έρευνα, πριν από την παράθεση των τελικών συμπερασμάτων.

2. Θεωρητικό πλαίσιο

2.1 Δικανική γλωσσολογία: Βασικές αρχές και μέθοδοι

Η απόδοση συγγραφικής πατρότητας κειμένου και οι συναφείς πρακτικές της δικανικής γλωσσολογίας στηρίζονται σε δύο βασικές παραδοχές. Η πρώτη είναι ότι κάθε άτομο διαθέτει μια διακριτή ιδιόλεκτο. Η δεύτερη είναι ότι τα χαρακτηριστικά αυτής της ιδιολέκτου επανεμφανίζονται με επαρκή σταθερότητα στα κείμενα ενός ατόμου ώστε να επιτρέπουν ουσιαστική σύγκριση (Coulthard 2004, Grant 2007). Οι παραδοχές αυτές ακολουθούν στενά εκείνες άλλων εγκληματολογικών επιστημών σύγκρισης προτύπων (pattern comparison forensic sciences), στις οποίες οι ιδιότητες ενός υπόπτου συνάγονται από τις κανονικότητες που παρατηρούνται στα προϊόντα τους (Ainsworth & Juola 2019).

Τα θεωρητικά θεμέλια της ιδιολέκτου έχουν πρόσφατα κωδικοποιηθεί μέσα από τη θεωρία της γλωσσικής ατομικότητας (Theory of Linguistic Individuality) του Nini (2023), η οποία παρέχει ένα τυπικό πλαίσιο βασισμένο στη γνωσιακή γλωσσολογία και τη γλωσσική επεξεργασία βάσει χρήσης (usage-based language processing). Ο Nini υποστηρίζει ότι οι προηγούμενες υπολογιστικές προσεγγίσεις στην ανάλυση της συγγραφικής πατρότητας υιοθετούσαν σιωπηρά ένα μη βιώσιμο μοντέλο γλωσσικής παραγωγής, σύμφωνα με το οποίο οι συγγραφείς επιλέγουν λέξεις μία προς μία βάσει γραμματικών κανόνων. Αντλώντας από τη θεωρία Now-or-Never Bottleneck των Christiansen και Chater (2016) και τη γνωσιακή γραμματική του Langacker (1987), ο Nini προτείνει αντ' αυτού ότι η επεξεργασία της γλώσσας γίνεται σε τεμά-

ότι ο δικαστής οφείλει να αξιολογεί αν η προσκομιζόμενη επιστημονική μέθοδος είναι: (α) ελέγξιμη και διαψεύσιμη, (β) δημοσιευμένη και αξιολογημένη από ομοτίμους, (γ) με γνωστό ποσοστό σφάλματος, και (δ) γενικά αποδεκτή στον οικείο επιστημονικό κλάδο. Η απόφαση Kumho (1999) επέκτεινε τα κριτήρια αυτά σε κάθε μορφή τεχνικής εμπειρογνωμοσύνης, πέραν της αυστηρά επιστημονικής. Στο πλαίσιο της δικανικής γλωσσολογίας, τα πρότυπα αυτά καθορίζουν κατά πόσο μια γλωσσολογική ανάλυση, όπως η απόδοση συγγραφικής πατρότητας ή η ανίχνευση κειμένου ΤΝ, μπορεί να γίνει αποδεκτή ως μαρτυρικό στοιχείο στο δικαστήριο.

χια (chunks): αυτόματα παραγόμενες μονάδες αποθηκευμένες στη μακρόχρονη μνήμη, οι οποίες κυμαίνονται από συστάδες χαρακτήρων και μορφήματα έως ακολουθίες πολλών λέξεων και σχηματικές κατασκευές. Αυτές οι μονάδες υπάρχουν σε ένα συνεχές παγίωσης (entrenchment), όπου οι συχνότερα χρησιμοποιούμενες δομές γίνονται βαθύτερα αυτοματοποιημένες και ευκολότερα προσβάσιμες. Η γραμματική ενός ατόμου σε οποιαδήποτε χρονική στιγμή είναι, επομένως, το σύνολο των μονάδων των οποίων η παγίωση υπερβαίνει ένα κατώφλι που επιτρέπει την αυτόματη παραγωγή. Έτσι, η ιδιόλεκτος ενός ατόμου ορίζεται τυπικά ως η «δεικτοδοτημένη οικογένεια» αυτών των γραμματικών κατά τη διάρκεια της ζωής του, αποτυπώνοντας τη δυναμική και μεταβλητή φύση της γλωσσικής γνώσης (Nini 2023).

Το θεωρητικό αυτό σχήμα εξηγεί γιατί η συνδυαστική της γλώσσας οδηγεί στην ατομικότητα. Όπως πρότεινε ο Coulthard (2004), η μοναδικότητα δεν προκύπτει από μεμονωμένα διακριτικά χαρακτηριστικά, αλλά από την ιδιολεκτική συνεπιλογή (idiolectal co-selection): τη σύζευξη πολλών επιλογών που μεμονωμένα μπορεί να είναι κοινές, αλλά σε συνδυασμό σχηματίζουν πρότυπα δυνάμει μοναδικά για ένα άτομο. Η υπόθεση του Unabomber αποτελεί παράδειγμα αυτής της αρχής: δώδεκα γλωσσικά στοιχεία, συμπεριλαμβανομένων φράσεων όπως at any rate, clearly και presumably, συνέδεσαν μοναδικά το μανιφέστο με γραπτά του Ted Kaczynski (Coulthard et al. 2017). Το πλαίσιο του Nini παρέχει τη γνωσιακή-γλωσσολογική εξήγηση γι' αυτό το φαινόμενο: τα άτομα διαφέρουν στα ρεπερτόρια των παγιωμένων τεμαχίων τους επειδή η γλωσσική εμπειρία κάθε ατόμου είναι μοναδική, οδηγώντας σε ιδιοσυγκρασιακά πρότυπα αυτοματοποίησης που εκδηλώνονται ως διακριτά υφολογικά προφίλ.

Ιστορικά, αυτές οι παραδοχές οδήγησαν σε δύο ευρείες μεθοδολογικές παραδόσεις. Η μία είναι η ποιοτική δικανική υφολογία, που σχετίζεται με την εκ του σύνεγγυς ανάγνωση (close reading) κειμένων για τον εντοπισμό διακριτών λεξιλογικών, συντακτικών και πραγματολογικών χαρακτηριστικών (McMenamin 2002). Η άλλη είναι η υφομετρία (stylometry), η οποία προέκυψε από τις λογοτεχνικές σπουδές και την υπολογιστική γλωσσολογία, και εστιάζει σε στατιστικές μετρήσεις χαρακτηριστικών όπως οι συχνότητες λειτουργικών λέξεων και τα ν-γράμματα χαρακτήρων.[5] Το έργο του Burrows (2002) είναι ιδιαίτερα επιδραστικό στη δεύτερη παράδοση. Το μέτρο Delta που ανέπτυξε υπολογίζει τη μέση απόλυτη διαφορά μεταξύ των τιμών z (z-scores) συχνών τύπων λέξεων στα κείμενα[6], επιτρέποντας την ταυτόχρονη σύγκριση

5. Ν-γράμματα χαρακτήρων (character n-grams): Συνεχόμενες ακολουθίες n χαρακτήρων που εξάγονται από ένα κείμενο και χρησιμοποιούνται ως χαρακτηριστικά για την υπολογιστική ανάλυσή του. Για παράδειγμα, η λέξη «γλώσσα» αποδίδει τα εξής τριγράμματα (n=3): «γλώ», «λώσ», «ώσσ», «σσα». Η μέθοδος αυτή αποτυπώνει υπογλωσσικά μοτίβα, δηλαδή μορφολογικές καταλήξεις, συχνές συλλαβικές δομές, ορθογραφικές συνήθειες που συχνά διαφεύγουν της συνειδητής προσοχής του γράφοντος, καθιστώντας τα ν-γράμματα χαρακτήρων ιδιαίτερα αποτελεσματικά στην απόδοση συγγραφικής πατρότητας και σε άλλες εργασίες υφολογικής ανάλυσης. Επιπλέον, η μέθοδος είναι ανθεκτική σε ορθογραφικά λάθη και λειτουργεί ανεξάρτητα από γλώσσα, χωρίς να απαιτεί λεξικά ή γραμματική ανάλυση.

6. Τιμές z (z-scores) συχνών τύπων λέξεων: Στατιστικό μέτρο που εκφράζει πόσο αποκλίνει η συχνότητα μιας λέξης σε

ενός αμφισβητούμενου κειμένου με πολλαπλούς υποψήφιους συγγραφείς. Αυτή η προσέγγιση βασίζεται στη διαπίστωση ότι οι ασυνείδητες, υψηλής συχνότητας γλωσσικές επιλογές, και ιδιαίτερα οι λειτουργικές λέξεις, είναι πιο ανθεκτικές στη σκόπιμη χειραγώγηση και επομένως χρησιμεύουν ως ισχυροί δείκτες του συγγραφικού ύφους. Οι Evert et al. (2017) κατέδειξαν ότι η επιτυχία μέτρων τύπου Delta εξαρτάται πρωτίστως από την κανονικοποίηση των διανυσμάτων και πρότειναν την υπόθεση των βασικών προφίλ (Key Profiles Hypothesis): οι συγγραφείς διακρίνονται από το συνολικό πρότυπο διακύμανσης στα χαρακτηριστικά σε σχέση με έναν κανόνα, και όχι από το μέγεθος επιμέρους αποκλίσεων. Αυτή η διαπίστωση ευθυγραμμίζεται με τη θεωρία του Nini (2023), υποδηλώνοντας ότι η διαμόρφωση (configuration) των παγιωμένων γλωσσικών μονάδων ενός ατόμου και η θέση τους σε έναν πολυδιάστατο χώρο γλωσσικών επιλογών συνιστούν την υφομετρική του υπογραφή.

Ταυτόχρονα, η σύγχρονη δικανική υφολογία εκτείνεται πολύ πέρα από την απλή ανάλυση της συχνότητας λέξεων. Οι επαγγελματίες εξετάζουν πλέον συστηματικά χαρακτηριστικά σε πολλαπλά επίπεδα γλωσσικής οργάνωσης, συμπεριλαμβανομένης της φρασεολογίας, των πρακτικών στίξης, των ορθογραφικών συμβάσεων, καθώς και μέτρων λεξιλογικής ποικιλότητας και συντακτικής πολυπλοκότητας (Berriche & Larabi-Marie-Sainte 2024). Οι Mikros και Perifanos (2013) τυποποίησαν αυτή τη μεθοδολογία με την αναπαράσταση των πολυεπίπεδων προφίλ ν-γραμμάτων συγγραφέα (Author's Multilevel N-gram Profiles - AMNP), η οποία συνδυάζει ν-γράμματα χαρακτήρων ποικίλου μήκους σε ένα ενιαίο μοντέλο συγγραφέα. Εφαρμοζόμενη αρχικά σε δεδομένα του ελληνικού Twitter, η μέθοδος AMNP κατέδειξε ότι η αποτύπωση ορθογραφικών και υπολεξιλογικών προτύπων[7] σε πολλαπλά επίπεδα λεπτομέρειας (granularities) προσφέρει ισχυρή απόδοση συγγραφικής πατρότητας ακόμη και στις δύσκολες συνθήκες των κειμένων μέσων κοινωνικής δικτύωσης, όπου τα δείγματα είναι σύντομα και το ύφος άτυπο. Εργαλεία όπως το StyloMetrix (Okulska et al. 2023) συνεχίζουν αυτή την πολυδιάστατη προσέγγιση εξάγοντας υφομετρικά διανύσματα σε γραμματικές, λε-

---

ένα κείμενο από τη μέση συχνότητά της σε ένα σύνολο κειμένων (corpus), μετρημένη σε μονάδες τυπικής απόκλισης. Η τυποποίηση αυτή επιτρέπει τη σύγκριση λέξεων με πολύ διαφορετικές απόλυτες συχνότητες σε κοινή κλίμακα. Για παράδειγμα, αν ένας συγγραφέας χρησιμοποιεί το άρθρο «ο» συχνότερα από τον μέσο όρο, η τιμή z για τη λέξη αυτή θα είναι θετική· αν το χρησιμοποιεί σπανιότερα, θα είναι αρνητική. Στη μέθοδο Delta του Burrows, υπολογίζονται οι τιμές z για τις πιο συχνές λέξεις (συνήθως λειτουργικές λέξεις όπως άρθρα, προθέσεις, σύνδεσμοι) και συγκρίνονται μεταξύ κειμένων: όσο μικρότερη η συνολική απόσταση των τιμών z, τόσο μεγαλύτερη η υφολογική ομοιότητα και πιθανότερη η κοινή συγγραφική προέλευση.

7. Υπολεξιλογικά πρότυπα (sublexical patterns): Γλωσσικά μοτίβα που εντοπίζονται σε επίπεδο μικρότερο της λέξης, όπως ακολουθίες χαρακτήρων, συλλαβικές δομές, μορφολογικές καταλήξεις ή συνδυασμοί γραμμάτων. Σε αντίθεση με την ανάλυση σε επίπεδο λέξεων (lexical), η υπολεξιλογική ανάλυση «τεμαχίζει» το κείμενο σε μικρότερες μονάδες. Χαρακτηριστικό παράδειγμα είναι τα ν-γράμματα χαρακτήρων, τα οποία αποτελούν τυπική μορφή υπολεξιλογικής αναπαράστασης. Τα πρότυπα αυτά αντανακλούν ασυνείδητες συνήθειες του γράφοντος, όπως προτιμήσεις σε ορισμένους συνδυασμούς γραμμάτων ή μορφολογικές επιλογές, και παραμένουν σχετικά σταθερά ακόμη και όταν ο συγγραφέας αλλάζει θεματολογία ή λεξιλόγιο. Στο πλαίσιο της απόδοσης συγγραφικής πατρότητας, η υπολεξιλογική προσέγγιση είναι ιδιαίτερα χρήσιμη σε σύντομα ή άτυπα κείμενα, όπου το διαθέσιμο λεξιλόγιο είναι περιορισμένο.

ξιλογικές και συντακτικές διαστάσεις που μπορούν να χρησιμοποιηθούν τόσο στην έρευνα όσο και σε πραγματικές υποθέσεις.

Σε οποιαδήποτε δικανική εφαρμογή, ωστόσο, η μεθοδολογική αρτιότητα δεν επαρκεί από μόνη της. Στις Ηνωμένες Πολιτείες, το παραδεκτό των αποδεικτικών στοιχείων εμπειρογνωμόνων, συμπεριλαμβανομένων των γλωσσικών τεκμηρίων, διέπεται στα ομοσπονδιακά δικαστήρια από το πρότυπο Daubert, όπως καθορίστηκε στην υπόθεση *Daubert v. Merrell Dow Pharmaceuticals* (1993) και επεκτάθηκε από την *Kumho Tire Co. v. Carmichael* (1999). Σύμφωνα με αυτό το πλαίσιο, οι δικαστές της δίκης αξιολογούν εάν η κατάθεση του εμπειρογνώμονα στηρίζεται σε επαρκώς αξιόπιστη επιστημονική βάση. Σχετικοί παράγοντες περιλαμβάνουν την ελεγξιμότητα (testability) της τεχνικής, την έκθεσή της σε ομότιμη αξιολόγηση και δημοσίευση, το γνωστό ή δυνητικό ποσοστό σφάλματός της, την ύπαρξη προτύπων που ελέγχουν την εφαρμογή της και τον βαθμό αποδοχής της στη σχετική επιστημονική κοινότητα (National Institute of Justice, n.d.). Για τη δικανική γλωσσολογία, αυτά τα κριτήρια συνεπάγονται ότι η απόδοση πατρότητας και οι συναφείς μέθοδοι πρέπει να είναι εμπειρικά επικυρωμένες, με σαφώς προσδιορισμένα ποσοστά σφάλματος και διαφανή αναφορά, προκειμένου να παραμείνουν παραδεκτές και πειστικές στο δικαστήριο.

2.2 Μεγάλα γλωσσικά μοντέλα: Τεχνικά θεμέλια και δικανική συνάφεια

Τα μεγάλα γλωσσικά μοντέλα (ΜΓΜ) αποτελούν την πιο προηγμένη τεχνολογία για την επεξεργασία κειμένου από υπολογιστές. Βασίζονται σε μια αρχιτεκτονική που ονομάζεται «μετασχηματιστής» (Transformer), η οποία επιτρέπει στο μοντέλο να λαμβάνει υπόψη ολόκληρη την πρόταση ταυτόχρονα και έτσι μπορεί να συνδέει λέξεις που βρίσκονται μακριά η μία από την άλλη και να κατανοεί σύνθετες προτάσεις. Τα μοντέλα αυτά εκπαιδεύονται διαβάζοντας τεράστιες ποσότητες κειμένων και προσπαθώντας να μαντέψουν την επόμενη λέξη σε κάθε σημείο. Αυτή είναι μια διαδικασία που δεν απαιτεί ανθρώπινη καθοδήγηση, καθώς η ίδια η γλώσσα «διδάσκει» το μοντέλο. Μέσα από αυτή τη διαδικασία, τα μοντέλα αποκτούν εσωτερική γνώση της γραμματικής, του νοήματος και του τρόπου με τον οποίο οι άνθρωποι χρησιμοποιούν τη γλώσσα. Σύγχρονα συστήματα όπως τα GPT, Claude, LLaMA και Gemini περιέχουν δισεκατομμύρια ή ακόμα και τρισεκατομμύρια παραμέτρους και επιδεικνύουν αναδυόμενες ικανότητες, συμπεριλαμβανομένης της μάθησης εντός πλαισίου (incontext learning), της ακολουθίας οδηγιών και του ευέλικτου ελέγχου ύφους (Kumarage & Liu 2023).

Αρκετές ιδιότητες των ΜΓΜ είναι ιδιαίτερα σημαντικές για τη δικανική γλωσσολογία. Πρώτον, είναι εξαιρετικά ευέλικτα ως προς το ύφος. Μέσω κατάλληλων εντολών (prompting) ή μικρορύθμισης (fine-tuning), μπορούν να παράγουν κείμενο σε συγκεκριμένα γλωσσικά επί-

πεδα (registers) και κειμενικά γένη. Έρευνες σχετικά με τη μεταφορά ύφους αξιοποιώντας ένα μόνο δείγμα (one-shot style transfer)[8] καταδεικνύει ότι ακόμη και περιορισμένα συγγραφικά δείγματα μπορούν να χρησιμοποιηθούν για να κατευθύνουν τις εξόδους των ΜΓΜ προς συγκεκριμένα υφολογικά προφίλ (MirallesGonzález et al. 2025). Δεύτερον, στην προεπιλεγμένη λειτουργία τους, τα ΜΓΜ τείνουν προς στατιστικά «μέσα» πρότυπα που παρατηρούνται στα δεδομένα εκπαίδευσης. Όπως δείχνουν οι Przystalski et al. (2025), τα κείμενα που παράγονται από ΜΓΜ επιδεικνύουν τυπικά μεγαλύτερη γραμματική τυποποίηση σε σχέση με τα ανθρώπινα κείμενα. Αυτή η τυποποίηση δημιουργεί ανιχνεύσιμες διαφορές που μπορούν να αξιοποιηθούν για τον εντοπισμό τους. Ωστόσο, αυτό σημαίνει επίσης ότι η γλωσσική παραγωγή των ΜΓΜ στερείται συχνά των ιδιοσυγκρασιακών ανωμαλιών στις οποίες βασίζεται η απόδοση ανθρώπινης συγγραφικής πατρότητας.

Οι υφομετρικές μελέτες υποδηλώνουν ότι, επί του παρόντος, τα ΜΓΜ δεν έχουν συγκλίνει πλήρως με τους ανθρώπινους συγγραφείς. Ο O'Sullivan (2025) διαπιστώνει ότι τα δημιουργικά κείμενα που έχουν γραφτεί από ανθρώπους και οι μιμήσεις που παράγονται από ΜΓΜ καταλαμβάνουν διακριτές περιοχές του υφομετρικού χώρου όταν αναλύονται με τη μέθοδο Burrows' Delta. Ο Mikros (2025) χρησιμοποίησε την ιεραρχική ανάλυση συστάδων για να αναλύσει την ικανότητα του GPT4o να μιμείται λογοτεχνικό ύφος. Διαπίστωσε ότι ενώ το μοντέλο μπορεί να ομαδοποιήσει αποτελεσματικά τις μιμήσεις ενός συγκεκριμένου συγγραφέα, αυτές οι μιμήσεις παραμένουν διαχωρίσιμες από τα αυθεντικά κείμενα του εν λόγω συγγραφέα. Οι Wang et al. (2025) δείχνουν ότι τα ΜΓΜ εξακολουθούν να δυσκολεύονται να αναπαραγάγουν τις λεπτές υφολογικές αποχρώσεις των συγγραφέων όταν στηρίζονται σε περιορισμένα δείγματα. Ταυτόχρονα, υπάρχουν ενδείξεις ότι το χάσμα μεταξύ ΜΓΜ και ανθρώπινων κείμενων μικραίνει. Οι Zaitsu et al. (2025) αναφέρουν ότι τα γλωσσικά εξαγόμενα από πιο πρόσφατα συστήματα, όπως το GPTo1, είναι υφομετρικά πιο κοντά στην ανθρώπινη γραφή συγκριτικά με προγενέστερα μοντέλα, όπως το GPT4o, σε ένα ιαπωνικό σώμα κειμένων.[9]

8. Μεταφορά ύφους με ένα μόνο δείγμα (one-shot style transfer): Τεχνική κατά την οποία ένα γλωσσικό μοντέλο τεχνητής νοημοσύνης μπορεί να προσαρμόσει το ύφος γραφής του ώστε να μιμηθεί έναν συγκεκριμένο συγγραφέα, έχοντας στη διάθεσή του μόνο ένα κείμενο-δείγμα του συγγραφέα αυτού. Ο όρος «one-shot» (μία προσπάθεια) αντιδιαστέλλεται προς τις παραδοσιακές μεθόδους μηχανικής μάθησης, οι οποίες απαιτούν πολλά παραδείγματα για την εκπαίδευση του μοντέλου. Στο πλαίσιο της δικανικής γλωσσολογίας, η ικανότητα αυτή εγείρει σημαντικά ερωτήματα: αν ένα μοντέλο μπορεί να αναπαράγει πειστικά το ύφος ενός ατόμου από ελάχιστο δείγμα, τότε η παραδοσιακή απόδοση συγγραφικής πατρότητας, η οποία βασίζεται στην υπόθεση ότι κάθε άτομο έχει μοναδικό υφολογικό αποτύπωμα, αντιμετωπίζει νέες προκλήσεις.

9. GPT-4o και GPT-o1: Γλωσσικά μοντέλα τεχνητής νοημοσύνης που αναπτύχθηκαν από την εταιρεία OpenAI. Το GPT-4o (Μάιος 2024) είναι ένα πολυτροπικό μοντέλο (το «ο» σημαίνει «omni») και είναι σχεδιασμένο να επεξεργάζεται κείμενο, εικόνα και ήχο ταυτόχρονα, με έμφαση στην ταχύτητα και την αποδοτικότητα. Το GPT-o1 (Σεπτέμβριος 2024) ανήκει σε νεότερη γενιά μοντέλων που εστιάζουν στη σύνθετη συλλογιστική: πριν δώσουν απάντηση, αναπτύσσουν εσωτερικά μια αλυσίδα σκέψης (chain of thought), γεγονός που βελτιώνει την απόδοσή τους σε εργασίες που απαιτούν λογική ανάλυση, όπως μαθημα-

Για τη δικανική γλωσσολογία, το αποτέλεσμα είναι ένα δυναμικό και ασταθές τοπίο. Τα ΜΓΜ παρέχουν νέα εργαλεία για την ανάλυση κειμένου σε μεγάλη κλίμακα και για την αναπαράσταση υφολογικών προτύπων σε πολυδιάστατους χώρους διανυσματικών αναπαραστάσεων (embedding spaces)[10]. Ταυτόχρονα, ωστόσο, απειλούν τις υπάρχουσες παραδοχές σχετικά με το τι συνιστά ανθρώπινη συγγραφική ταυτότητα και αμφισβητούν την ευρωστία μεθόδων που αναπτύχθηκαν αποκλειστικά για ανθρώπινο κείμενο.

3. Ευκαιρίες: Τα ΜΓΜ ως εργαλεία για τη δικανική γλωσσολογία

Το πρώτο ερευνητικό ερώτημα αφορά τους τρόπους με τους οποίους τα ΜΓΜ μπορούν να λειτουργήσουν ως εργαλεία που επεκτείνουν, αντί απλώς να υπονομεύουν, την πράξη της δικανικής γλωσσογίας. Πρόσφατες εργασίες προτείνουν τρεις κύριους τομείς ευκαιριών: ενισχυμένη απόδοση συγγραφικής πατρότητας, κλιμακωτή πολυδύναμη και πολύγλωσση ανάλυση, και νέες μορφές λεπτομερούς και επεξηγήσιμης υφολογικής ανάλυσης.

3.1 Ενίσχυση της απόδοσης συγγραφικής πατρότητας

Μια ενδιαφέρουσα ερευνητική κατεύθυνση χρησιμοποιεί τα ίδια τα μεγάλα γλωσσικά μοντέλα ως εργαλεία για τον εντοπισμό του συγγραφέα ενός κειμένου. Οι Hu et al. (2024) ανέπτυξαν μια μέθοδο που αξιοποιεί τις ικανότητες των γλωσσικών μοντέλων για να απαντήσει στο ερώτημα: «Πόσο πιθανό είναι ένα αμφισβητούμενο κείμενο να έχει γραφτεί από συγκεκριμένο άτομο;». Αντί να μετρά απλώς εξωτερικά χαρακτηριστικά του κειμένου (όπως μήκος προτάσεων ή συχνότητα λέξεων), η μέθοδος ζητά από το μοντέλο να κρίνει αν το ύφος του αμφισβητούμενου κειμένου «ταιριάζει» με δείγματα γνωστών συγγραφέων. Χρησιμοποιώντας ένα σύγχρονο γλωσσικό μοντέλο (Llama 3) και δίνοντάς του μόνο ένα κείμενο-δείγμα από κάθε υποψήφιο συγγραφέα, οι ερευνητές πέτυχαν να αναγνωρίσουν

τικά προβλήματα ή νομική επιχειρηματολογία. Η παρατήρηση των Zaitsu et al. ότι τα κείμενα του GPT-o1 μοιάζουν υφολογικά περισσότερο με ανθρώπινη γραφή υποδηλώνει ότι η εξέλιξη των μοντέλων τείνει να μειώσει τα διακριτικά χαρακτηριστικά που επιτρέπουν την αναγνώριση κειμένου ως μηχανικά παραγόμενου.

10. Πολυδιάστατοι χώροι διανυσματικών αναπαραστάσεων (embedding spaces): Μαθηματικός τρόπος αναπαράστασης λέξεων ή κειμένων ως σημείων σε έναν νοητό χώρο πολλών διαστάσεων. Κάθε λέξη ή κείμενο μετατρέπεται σε μια σειρά αριθμών (διάνυσμα, εκατοντάδες ή χιλιάδες τιμές) που καθορίζουν τη «θέση» του σε αυτόν τον χώρο. Η βασική αρχή είναι ότι στοιχεία με παρόμοια σημασία ή υφολογικά χαρακτηριστικά τοποθετούνται κοντά μεταξύ τους: οι λέξεις «δικαστής» και «δικαστήριο» θα βρίσκονται εγγύτερα από τις λέξεις «δικαστής» και «ποδήλατο». Αντίστοιχα, κείμενα με παρόμοιο ύφος, π.χ. του ίδιου συγγραφέα, θα ομαδοποιούνται στον ίδιο χώρο. Για τη δικανική γλωσσολογία, η τεχνική αυτή επιτρέπει την ποσοτική σύγκριση υφολογικών προφίλ: η «απόσταση» μεταξύ δύο κειμένων στον χώρο αυτό μπορεί να αποτελέσει ένδειξη κοινής ή διαφορετικής συγγραφικής προέλευσης.

σωστά τον συγγραφέα σε περίπου 85% των περιπτώσεων, σε δοκιμές με κριτικές ταινιών και αναρτήσεις ιστολογίων. Η σημασία αυτής της εργασίας έγκειται στο ότι παρέχει αριθμητικές εκτιμήσεις πιθανότητας και όχι απλώς μια κατηγορηματική απάντηση «ναι» ή «όχι», κάτι που ανταποκρίνεται καλύτερα στον τρόπο με τον οποίο τα δικαστήρια αξιολογούν αποδεικτικά στοιχεία..

Οι Choi et al. (2025) εφάρμοσαν τη χρήση γλωσσικών μοντέλων για την αναγνώριση συγγραφέα στον κώδικα προγραμμάτων H/Y, έναν τομέα που παρουσιάζει αρκετές ομοιότητες με την ανάλυση κειμένου: όπως κάθε συγγραφέας έχει το δικό του ύφος γραφής, έτσι και κάθε προγραμματιστής έχει τον δικό του τρόπο να γράφει κώδικα. Η μελέτη τους έδειξε ότι τα γλωσσικά μοντέλα μπορούν να κρίνουν αν δύο κομμάτια κώδικα γράφτηκαν από το ίδιο άτομο, χωρίς να χρειαστεί καμία ειδική εκπαίδευση, απλώς με την κατάλληλη διατύπωση της ερώτησης στο ΜΓΜ. Τα αποτελέσματα ήταν αρκετά αξιόπιστα (συντελεστής συσχέτισης 0,78 σε κλίμακα που φτάνει το 1). Για περιπτώσεις με πολλούς υποψήφιους συγγραφείς, οι ερευνητές σχεδίασαν μια μέθοδο διαδοχικών συγκρίσεων, σαν αγώνες τουρνουά, όπου οι υποψήφιοι συγκρίνονται σε μικρές ομάδες μέχρι να αναδειχθεί ο επικρατέστερος. Με αυτή τη μέθοδο πέτυχαν να αναγνωρίσουν σωστά τον συγγραφέα σε περίπου 75% των περιπτώσεων, ακόμη και με εκατοντάδες υποψήφιους και διαφορετικές γλώσσες προγραμματισμού, έχοντας μόνο ένα δείγμα κώδικα από κάθε προγραμματιστή. Αν και η έρευνα επικεντρώθηκε στον κώδικα, η λογική της μεθόδου (συγκρίσεις ανά ζεύγη με περιορισμένο υλικό) είναι άμεσα εφαρμόσιμη σε δικανικές υποθέσεις, όπου συχνά υπάρχουν πολλοί ύποπτοι συγγραφείς και λίγα διαθέσιμα δείγματα γραφής. Μια σχετική αλλά διακριτή εξέλιξη αφορά την απόδοση κειμένου που παράγεται από ΜΓΜ σε συγκεκριμένα μοντέλα. Οι Bisztray et al. (2025) παρουσιάζουν το CodeT5 Authorship, ένα μοντέλο εκπαιδευμένο σε κώδικα παραγόμενο από ΜΓΜ που επιτυγχάνει ακρίβεια άνω του 97% στη διάκριση μεταξύ στενά συγγενικών συστημάτων όπως το GPT 4.1 και το GPT 4o. Το σύνολο δεδομένων τους LLM AuthorBench περιλαμβάνει 32.000 προγράμματα υπολογιστών γραμμένα από ΤΝ και παρέχει ένα σημείο αναφοράς για την απόδοση πηγής-μοντέλου. Ενώ αυτή η εργασία εστιάζει επίσης στον κώδικα, καταδεικνύει ότι τα συνθετικά κείμενα μπορούν να φέρουν συγκεκριμένες υπογραφές μοντέλου που είναι ανιχνεύσιμες σε κλίμακα.[11] Η δικα-

11. Υπογραφές μοντέλου ανιχνεύσιμες σε κλίμακα (model signatures detectable at scale): Χαρακτηριστικά γλωσσικά μοτίβα που αφήνει ένα συγκεκριμένο γλωσσικό μοντέλο τεχνητής νοημοσύνης στα κείμενα που παράγει, κάτι σαν «δακτυλικό αποτύπωμα» του μοντέλου. Αυτά μπορεί να περιλαμβάνουν προτιμήσεις σε ορισμένες λέξεις, συντακτικές δομές ή στατιστικές κανονικότητες που ο άνθρωπος δύσκολα αντιλαμβάνεται, αλλά μπορούν να εντοπιστούν με υπολογιστική ανάλυση. Η φράση «σε κλίμακα» υποδηλώνει ότι η ανίχνευση αυτών των υπογραφών είναι εφικτή όχι μόνο σε μεμονωμένα κείμενα, αλλά και όταν εξετάζονται χιλιάδες ή εκατομμύρια κείμενα ταυτόχρονα – για παράδειγμα, για τον εντοπισμό μαζικά παραγόμενου συνθετικού περιεχομένου στο διαδίκτυο. Η ύπαρξη τέτοιων υπογραφών είναι κρίσιμη για τη δικανική γλωσσολογία, καθώς υποδηλώνει ότι η αναγνώριση του μοντέλου προέλευσης ενός κειμένου παραμένει θεωρητικά εφικτή, παρά την ολοένα βελτιούμενη ποιότητα των συνθετικών κειμένων.

νική γλωσσολογία, η οποία ολοένα περισσότερο καλείται να απαντήσει σε ερωτήματα σχετικά με το αν ένα δεδομένο κείμενο παρήχθη από ένα συγκεκριμένο ΜΓΜ, έχει να ωφεληθεί άμεσα από τέτοιες τεχνικές.

Από κοινού, οι εν λόγω μελέτες υποδεικνύουν ότι τα ΜΓΜ μπορούν να χρησιμοποιηθούν για την ενίσχυση της ανάλυσης της συγγραφικής πατρότητας με δύο συμπληρωματικούς τρόπους: παρέχοντας ισχυρότερα εργαλεία για την απόδοση ανθρώπινης πατρότητας και επιτρέποντας την ισχυρή απόδοση κειμένων παραγόμενων από ΜΓΜ σε συγκεκριμένα μοντέλα και οικογένειες μοντέλων παραγωγικής τεχνητής νοημοσύνης.

3.2 Κλιμακωτή, πολυδύναμη και πολύγλωσση ανάλυση

Μια δεύτερη ευκαιρία έγκειται στην ικανότητα των συστημάτων που βασίζονται σε ΜΓΜ να εκτελούν από κοινού πολλαπλές εργασίες σχετικές με την προέλευση και να λειτουργούν διαγλωσσικά. Οι Rao et al. (2025) εισάγουν το DA MTL, ένα πλαίσιο πολυδύναμης μάθησης (multitask learning framework) που συνδυάζει την ανίχνευση κειμένου ΤΝ με την απόδοση πηγής ΜΓΜ.[12] Αξιολογημένο σε εννέα σύνολα δεδομένων και τέσσερα μοντέλα κορμού, το DA MTL μαθαίνει κοινές αναπαραστάσεις που βελτιώνουν την απόδοση και στις δύο εργασίες, διατηρώντας παράλληλα τις διακρίσεις που αφορούν κάθε συγκεκριμένη εργασία. Αυτός ο σχεδιασμός αντανακλά την πραγματικότητα πολλών δικανικών προβλημάτων στα οποία το ερώτημα δεν είναι απλώς αν ένα κείμενο είναι «παραγόμενο από ΤΝ», αλλά και ποιο σύστημα το παρήγαγε και πώς αυτή η πληροφορία αλληλεπιδρά με υποθέσεις σχετικά με την εμπλοκή ανθρώπινου παράγοντα.

Οι La Cava et al. (2025) εξετάζουν την πολύγλωσση διάσταση διατυπώνοντας το πρόβλημα της απόδοσης πατρότητας για κείμενα που έχουν παραχθεί από μηχανές σε 18 γλώσσες. Τα αποτελέσματά τους δείχνουν ότι η διαγλωσσική μεταφορά είναι δύσκολη, με την απόδοση να υποβαθμίζεται κατά τη μετακίνηση μεταξύ γλωσσικών οικογενειών και συστημάτων γραφής. Τα ευρήματα αυτά απηχούν προγενέστερες εμπειρικές εργασίες στη διαγλωσσική υφομετρία. Οι Juola και Mikros (2016) κατέδειξαν ότι ορισμένα υφομετρικά χαρακτηριστικά, συμπεριλαμβανομένων των μέτρων λεξιλογικού πλούτου, του μέσου μήκους λέξης και συμβάσεων ειδικών για την πλατφόρμα, συσχετίζονται ισχυρά μεταξύ των γλωσσών για τον ίδιο συγγραφέα, υποδηλώνοντας ότι υπάρχουν σήματα συγγραφικής πατρότητας ανεξάρτητα από τη γλώσσα.

Ωστόσο, η μεταγενέστερη συγκριτική μελέτη τους (Juola, Mikros, & Vinsick 2019) αποκάλυψε

12. Βλ. υποσημείωση 2.

ότι η ακρίβεια απόδοσης ποικίλλει σημαντικά μεταξύ των γλωσσών ακόμη και υπό αυστηρά ελεγχόμενες συνθήκες: χρησιμοποιώντας τους ίδιους συγγραφείς που έγραφαν για τα ίδια θέματα, η απόδοση στα αγγλικά κείμενα ήταν σημαντικά υψηλότερη απ' ό,τι στα ελληνικά, μια ανισότητα που αποδόθηκε σε διαφορές στη μορφολογική πολυπλοκότητα και την ευελιξία της σειράς των λέξεων. Ωστόσο, αυτό το σώμα εργασιών αποδεικνύει ότι οι προσεγγίσεις που βασίζονται σε ΜΓΜ μπορούν να εφαρμοστούν πέρα από τα αγγλικά, υπό την προϋπόθεση τα μοντέλα να αξιολογούνται προσεκτικά σε κάθε συγκεκριμένο γλωσσικό πλαίσιο. Για τους επαγγελματίες της δικανικής γλωσσολογίας που ασχολούνται με πολύγλωσσα τεκμήρια, η δυνατότητα απόδοσης χωρίς εκτεταμένη μηχανική χαρακτηριστικών (feature engineering)[13] ειδική για κάθε γλώσσα είναι ιδιαίτερα ελκυστική, αλλά υπογραμμίζει επίσης την ανάγκη για προσεκτική ερμηνεία και τοπικά επικυρωμένα μοντέλα.

3.3 Νέες αναλυτικές δυνατότητες και επεξηγήσιμες ροές εργασίας

Ο τρίτος τομέας ευκαιριών αφορά μορφές ανάλυσης που ήταν προηγουμένως δύσκολες ή αδύνατες. Οι Römisch et al. (2025) εξετάζουν εάν τα ΜΓΜ μπορούν να ανιχνεύσουν αλλαγή ύφους σε επίπεδο πρότασης, μια εργασία σχετική με ζητήματα παρεμβολής, αφανούς συγγραφής (ghostwriting) και παραποίησης εγγράφων. Τα αποτελέσματά τους δείχνουν ότι τα ΜΓΜ αιχμής μπορούν, ακόμη και χωρίς ειδική για την εργασία μικρορύθμιση (fine tuning), να εντοπίσουν λεπτές υφολογικές μετατοπίσεις εντός ενός εγγράφου με μεγαλύτερη ακρίβεια από τις παραδοσιακές μεθόδους αναφοράς (baselines).[14] Σε δικανικά περιβάλλοντα, τέτοιες δυνατότητες θα μπορούσαν να υποστηρίξουν τη συστηματική εξέταση μακροσκελών κειμένων για σημεία στα οποία ενδέχεται να έχει αλλάξει η συγγραφική πατρότητα.

13. Μηχανική χαρακτηριστικών (feature engineering): Η διαδικασία κατά την οποία ο ερευνητής επιλέγει και σχεδιάζει χειροκίνητα τα γλωσσικά στοιχεία που θα χρησιμοποιήσει ένας αλγόριθμος για να αναλύσει ένα κείμενο. Στην παραδοσιακή απόδοση συγγραφικής πατρότητας, αυτό σημαίνει ότι ο αναλυτής πρέπει να αποφασίσει εκ των προτέρων ποια χαρακτηριστικά θα μετρήσει, π.χ. μέσο μήκος πρότασης, συχνότητα επιρρημάτων, χρήση σημείων στίξης, αναλογία ρημάτων προς ουσιαστικά, κ.ο.κ. Η διαδικασία αυτή απαιτεί τόσο γλωσσολογική εξειδίκευση όσο και γνώση της συγκεκριμένης γλώσσας: τα χαρακτηριστικά που λειτουργούν καλά στα αγγλικά μπορεί να μην είναι κατάλληλα για τα ελληνικά ή τα αραβικά, λόγω διαφορετικής μορφολογίας και σύνταξης. Τα μεγάλα γλωσσικά μοντέλα παρακάμπτουν εν μέρει αυτή την ανάγκη, καθώς μαθαίνουν αυτόματα εσωτερικές αναπαραστάσεις της γλώσσας χωρίς να απαιτείται ρητός σχεδιασμός χαρακτηριστικών από τον ερευνητή, γεγονός που διευκολύνει την εφαρμογή τους σε πολλές γλώσσες με μικρότερη προσπάθεια προσαρμογής.

14. Μέθοδοι αναφοράς (baselines): Καθιερωμένες, συνήθως απλούστερες τεχνικές που χρησιμοποιούνται ως σημείο σύγκρισης για την αξιολόγηση νέων μεθόδων. Όταν οι ερευνητές προτείνουν μια καινοτόμο προσέγγιση, τη συγκρίνουν με τις μεθόδους αναφοράς για να αποδείξουν ότι προσφέρει πραγματική βελτίωση και όχι απλώς διαφορετικά αποτελέσματα. Στο πλαίσιο της απόδοσης συγγραφικής πατρότητας, τυπικές μέθοδοι αναφοράς είναι οι παραδοσιακές υφομετρικές τεχνικές όπως η μέτρηση συχνοτήτων λέξεων ή τα ν-γράμματα χαρακτήρων σε συνδυασμό με αλγορίθμους ταξινόμησης. Εδώ, η διαπίστωση ότι τα μεγάλα γλωσσικά μοντέλα ξεπερνούν τις μεθόδους αναφοράς σημαίνει ότι αποδίδουν καλύτερα από τις τεχνικές που μέχρι τώρα θεωρούνταν το πρότυπο στον τομέα.

Ο Abbas (2025) συγκρίνει δύο προσεγγίσεις στην απόδοση πατρότητας σε ένα παράλληλο σώμα κειμένων Ανθρώπου–ΤΝ που καλύπτει έξι τομείς: διανυσματικές αναπαραστάσεις σταθερού ύφους και ένα ΜΓΜ (GPT-4o) στο ρόλο του κριτή συντονισμένο με οδηγίες (instruction-tuned). Αναφέρει ότι το ΜΓΜ ως κριτής αποδίδει καλύτερα στη λογοτεχνία και τον ακαδημαϊκό λόγο, όπου τα σημασιολογικά χαρακτηριστικά και τα χαρακτηριστικά επιπέδου λόγου είναι κρίσιμα, ενώ οι διανυσματικές αναπαραστάσεις σταθερού ύφους αποδίδουν καλύτερα σε δεδομένα διαλογικού τύπου με περισσότερες δομικές κανονικότητες. Αυτά τα ευρήματα υποδηλώνουν ότι καμία μεμονωμένη μέθοδος δεν είναι βέλτιστη σε όλους τους τομείς. Αντ› αυτού, οι υβριδικές ροές εργασίας που συνδυάζουν πολλαπλά γλωσσικά χαρακτηριστικά, καθένα ευθυγραμμισμένο με συγκεκριμένους τύπους κειμένου, είναι πιθανό να αποφέρουν τα πιο αξιόπιστα δικανικά συμπεράσματα.

Η επεξηγησιμότητα είναι μια επιπλέον κρίσιμη παράμετρος. Οι Roemling et al. (2024) διερευνούν τη χρήση προσεγγίσεων επεξηγήσιμης μηχανικής μάθησης σε μια μελέτη περίπτωσης όπου επιχειρείται η γεωγλωσσική σκιαγράφηση του προφίλ συγγραφικής πατρότητας. Χρησιμοποιώντας τεχνικές όπως το SHAP για τον εντοπισμό χαρακτηριστικών που οδηγούν τις αποφάσεις του μοντέλου, δείχνουν ότι τα μοντέλα υψηλής απόδοσης μπορούν να γίνουν ερμηνεύσιμα με τρόπους που έχουν νόημα για νομικά ακροατήρια. Για τη δικανική γλωσσολογία, αυτό είναι ιδιαίτερα σημαντικό: οποιοδήποτε αναλυτικό όφελος επιτυγχάνεται με μεθόδους βασισμένες σε ΜΓΜ θα έχει περιορισμένη αξία εάν οι εμπειρογνώμονες δεν μπορούν να εξηγήσουν, με γλωσσικούς και νομικά προσβάσιμους όρους, γιατί ένα σύστημα κατέληξε σε ένα συγκεκριμένο συμπέρασμα.

Οι προαναφερθείσες μελέτες καταδεικνύουν ότι τα ΜΓΜ μπορούν τόσο να επαυξήσουν τις καθιερωμένες μεθόδους όσο και να ανοίξουν νέους δρόμους για ανάλυση. Παρέχουν πλουσιότερες αναπαραστάσεις υφολογικής ομοιότητας, υποστηρίζουν την από κοινού ανίχνευση και απόδοση μεταξύ γλωσσών και εργασιών, και επιτρέπουν λεπτομερείς και επεξηγήσιμες αναλύσεις που είναι καλύτερα ευθυγραμμισμένες με τις αποδεικτικές απαιτήσεις των δικαστηρίων.

4. Απειλές: Πώς τα ΜΓΜ διαταράσσουν τη δικανική γλωσσολογία

Το δεύτερο ερευνητικό ερώτημα αφορά τους τρόπους με τους οποίους τα ΜΓΜ απειλούν τις υπάρχουσες δικανικές πρακτικές. Τρεις δέσμες προκλήσεων ξεχωρίζουν: η υπονόμευση των παραδοχών που βασίζονται στην ιδιόλεκτο μέσω της μίμησης ύφους και της συγκάλυψης, η διάδοση συνθετικού κειμένου και οι αδυναμίες των υπαρχόντων εργαλείων ανίχνευσης, και η επακόλουθη πίεση στα νομικά και μεθοδολογικά θεμέλια του κλάδου.

4.1 Μίμηση ύφους και η πρόκληση της ιδιολέκτου

Τα ΜΓΜ καθιστούν ουσιαστικά ευκολότερο για τους συγγραφείς να χειραγωγήσουν το ύφος, είτε για να αποκρύψουν τη δική τους ταυτότητα είτε για να υποδυθούν άλλους. Οι Alperin et al. (2025) εξετάζουν τέτοιες επιθέσεις «μάσκας και μίμησης» (masks and mimicry) σε συστήματα επαλήθευσης συγγραφικής πατρότητας. Χρησιμοποιώντας ΜΓΜ για την παράφραση και τον υφολογικό μετασχηματισμό κειμένων, δείχνουν ότι οι κακόβουλοι χρήστες μπορούν να υποβαθμίσουν σημαντικά την απόδοση των μοντέλων επαλήθευσης συγγραφικής ταυτότητας, αποδυναμώνοντας έτσι την αποδεικτική αξία της υφολογικής συνέπειας ή ασυνέπειας.

Οι Huang et al. (2024), σε μια περιεκτική επισκόπηση, υποστηρίζουν ότι η παραδοσιακή έννοια της ιδιολέκτου ως σταθερής, ατομικής υφολογικής υπογραφής καθίσταται ολοένα πιο προβληματική στην εποχή των ΜΓΜ. Διακρίνουν τέσσερις βασικές εργασίες: απόδοση κειμένου γραμμένου από άνθρωπο, ανίχνευση κειμένου που παράγεται από ΤΝ, απόδοση κειμένου ΤΝ σε μοντέλα και απόδοση σε περιβάλλοντα συν-συγγραφής ανθρώπου–ΜΓΜ. Τα υβριδικά κείμενα, όπου οι προτάσεις της ΤΝ ενσωματώνονται, επεξεργάζονται και επαναπλαισιώνονται από ανθρώπινους συγγραφείς, είναι ιδιαίτερα απαιτητικά, καθώς δεν αντιστοιχούν πλέον στις σχετικά καθαρές κατηγορίες τις οποίες υπέθεταν πολλές παλαιότερες μέθοδοι απόδοσης.

Παρ› όλα αυτά, οι υφομετρικές μελέτες υποδεικνύουν ότι, τουλάχιστον προς το παρόν, τα ΜΓΜ δεν έχουν εξαλείψει τη δικανική αξία του ύφους. Ο O'Sullivan (2025) καταδεικνύει ότι τα δημιουργικά κείμενα που παράγονται από ανθρώπους και από ΜΓΜ σχηματίζουν διακριτές συστάδες υπό ανάλυση βάσει Delta, ακόμη και όταν τα μοντέλα επιχειρούν να μιμηθούν συγκεκριμένους συγγραφείς. Οι Wang et al. (2025) διαπιστώνουν ομοίως ότι τα ΜΓΜ δυσκολεύονται να αναπαράγουν τα υπόρρητα στυλ γραφής των καθημερινών συγγραφέων από περιορισμένες εντολές, υποδηλώνοντας ότι τα βαθύτερα ιδιοσυγκρασιακά πρότυπα παραμένουν σχετικά ανθεκτικά στη μίμηση. Ο Mikros (2025) παρέχει περαιτέρω λεπτομέρειες στην ανάλυσή του για τις λογοτεχνικές μιμήσεις του GPT4o. Αναφέρει ότι ενώ το GPT4o μπορεί να παράγει εσωτερικά συνεπείς μιμήσεις του ύφους ενός συγγραφέα-στόχου, αυτές οι μιμήσεις παραμένουν υφομετρικά διαχωρίσιμες από τα αυθεντικά κείμενα του εν λόγω συγγραφέα, και οι γενικές γλωσσικές παραγωγές του GPT είναι ακόμα πιο διακριτές.

Αυτά τα συγκλίνοντα ευρήματα υποδηλώνουν μια σύνθετη εικόνα. Τα ΜΓΜ χαμηλώνουν σημαντικά το εμπόδιο για τη χειραγώγηση του ύφους και έτσι υπονομεύουν τις άστοχες εφαρμογές της έννοιας της ιδιολέκτου. Ωστόσο, ταυτόχρονα, τα κείμενα που υποβοηθούνται από ΜΓΜ φαίνεται να φέρουν τις δικές τους ανιχνεύσιμες υπογραφές, οι οποίες, εάν μοντελοποιηθούν κατάλληλα, μπορούν οι ίδιες να υποστηρίξουν δικανικά συμπεράσματα σχετικά με την πατρότητα και την εμπλοκή της ΤΝ.

4.2 Συνθετικό κείμενο, αποτυχίες ανίχνευσης και προκατάληψη

Μια δεύτερη δέσμη απειλών προκύπτει από τη δυσκολία αξιόπιστης ταυτοποίησης κειμένου που παράγεται από ΤΝ. Τα σφάλματα ανίχνευσης δεν κατανέμονται τυχαία και έχουν σοβαρές επιπτώσεις για την ισοτιμία και τη δέουσα διαδικασία (due process).

Οι Dalalah και Dalalah (2023) αναλύουν την απόδοση εργαλείων ανίχνευσης ΤΝ σε ακαδημαϊκά πλαίσια και διαπιστώνουν ότι οι ενότητες επισκόπησης βιβλιογραφίας είναι ιδιαίτερα επιρρεπείς σε ψευδώς θετικά αποτελέσματα, πιθανώς επειδή το τυπικό και τυποποιημένο ύφος τους μοιάζει με τη γλωσσική παραγωγή των ΜΓΜ. Οι Rashidi et al. (2023) τεκμηριώνουν παρόμοια προβλήματα στην ιατρική πληροφορική, όπου χειρόγραφα γραμμένα από ανθρώπους ταξινομούνται μερικές φορές εσφαλμένα ως παραγόμενα από ΤΝ από λογισμικό ανίχνευσης, με ποσοστά ψευδώς θετικών που υπερβαίνουν το 8% για ορισμένους τύπους περιοδικών.

Πιο ανησυχητικά είναι τα ευρήματα των Liang et al. (2023), που καταδεικνύουν ότι αρκετοί ευρέως χρησιμοποιούμενοι ανιχνευτές είναι έντονα προκατειλημμένοι εις βάρος μη φυσικών ομιλητών της αγγλικής. Στη μελέτη τους σε επτά ανιχνευτές, ένα μεγάλο ποσοστό κειμένων TOEFL γραμμένων από Κινέζους φοιτητές χαρακτηρίστηκε εσφαλμένα ως παραγόμενο από ΤΝ, με έναν ανιχνευτή να επισημαίνει σχεδόν το 98% αυτών των δοκιμίων ως συνθετικά, ενώ κείμενα από φυσικούς ομιλητές της αγγλικής σπάνια ταξινομήθηκαν εσφαλμένα. Οι συγγραφείς αποδίδουν αυτό το αποτέλεσμα στη χρήση ευρετικών κανόνων που βασίζονται στον στατιστικό προσδιορισμό της έκπληξης του μοντέλου (perplexity): κείμενα με απλούστερο, πιο προβλέψιμο λεξιλόγιο τείνουν να ταξινομούνται ως παραγόμενα από ΤΝ, ενώ κείμενα πιο πολύπλοκα από λεξιλογικής άποψης ταξινομούνται ως γραμμένα από άνθρωπο. Αυτή η σχεδιαστική επιλογή τιμωρεί συστηματικά τους συγγραφείς των οποίων οι γλωσσικοί πόροι είναι περιορισμένοι, ενσωματώνοντας έτσι μια προκατάληψη για την γλωσσική επάρκεια των συγγραφέων στα αποτελέσματα ανίχνευσης.

Τα συστήματα ανίχνευσης είναι επίσης ευάλωτα στην αντιθετική χειραγώγηση (adversarial manipulation). Οι Creo και Pudasaini (2024) δείχνουν ότι η υποκατάσταση ομογλύφων, δηλαδή η αντικατάσταση χαρακτήρων με οπτικά παρόμοιους χαρακτήρες από άλλα συστήματα γραφής, μπορεί να υποβαθμίσει δραματικά την απόδοση του ανιχνευτή. Σε επτά συστήματα ανίχνευσης, αναφέρουν μέση πτώση του συντελεστή συσχέτισης Matthews από 0,64 σε –0,01, μειώνοντας ουσιαστικά την απόδοση ταξινόμησης κάτω από το επίπεδο της τύχης, διατηρώντας παράλληλα πλήρως την αναγνωσιμότητα από τον άνθρωπο. Οι Zeng et al. (2024) εξετάζουν την ανίχνευση σε επίπεδο πρότασης σε συνεργατικά κείμενα ανθρώπου–ΤΝ από το σώμα κειμένων CoAuthor και διαπιστώνουν ότι οι ανιχνευτές δεν αποδίδουν με επάρκεια όταν τα τμήματα είναι σύντομα, η πατρότητα εναλλάσσεται συχνά ή οι προτάσεις που παρά-

γονται από ΤΝ έχουν υποστεί επεξεργασία από ανθρώπους. Υπό τέτοιες συνθήκες, η ακριβής ποσοτικοποίηση της συνεισφοράς της ΤΝ σε ένα έγγραφο καθίσταται εξαιρετικά δύσκολη.

Αυτά τα ευρήματα υποδεικνύουν ότι τα τρέχοντα εργαλεία ανίχνευσης είναι τόσο προκατειλημμένα όσο και εύθραυστα. Ταξινομούν εσφαλμένα την ανθρώπινη γραφή με τρόπους που επηρεάζουν δυσανάλογα τους μη φυσικούς ομιλητές και τους χρήστες συγκεκριμένων ειδών λόγου, και παρακάμπτονται εύκολα με σχετικά απλές χειραγωγήσεις. Ως εκ τούτου, δεν μπορούν ακόμη να χρησιμεύσουν ως αξιόπιστα δικανικά εργαλεία σε περιβάλλοντα υψηλού ρίσκου.

4.3 Νομική πίεση

Οι τεχνικές ευπάθειες που περιγράφηκαν παραπάνω μεταφράζονται άμεσα σε νομικές προκλήσεις. Σύμφωνα με τα *Daubert* και *Kumho Tire*, οι μέθοδοι των εμπειρογνωμόνων πρέπει να είναι αποδεδειγμένα αξιόπιστες, με γνωστά ποσοστά σφάλματος που αντέχουν σε αντιθετικό έλεγχο (adversarial scrutiny) (*Daubert v. Merrell Dow Pharmaceuticals*, 1993; *Kumho Tire Co. v. Carmichael*, 1999; National Institute of Justice, n.d.). Οι Ainsworth και Juola (2019) υποστηρίζουν ότι η ανάλυση της συγγραφικής πατρότητας μπορεί να λειτουργήσει ως μοντέλο για την εγκληματολογική επιστήμη μόνο εάν ενστερνιστεί την αυστηρή επικύρωση των μεθόδων της. Δεδομένου του είδους των σφαλμάτων που αναφέρθηκαν από τους Liang et al. (2023) και άλλους, είναι δύσκολο να υποστηριχθεί ότι οι σύγχρονοι ανιχνευτές κειμένου ΤΝ πληρούν αυτές τις απαιτήσεις, ιδιαίτερα σε περιβάλλοντα όπου οι αποφάσεις επηρεάζουν την ελευθερία, την οικονομική κατάσταση ή τη φήμη των ανθρώπων.

Επιπλέον, η διαδεδομένη διαθεσιμότητα εξελιγμένων γεννητριών κειμένου αλλάζει το στρατηγικό τοπίο της δικαστικής διαμάχης. Ένας συνήγορος υπεράσπισης μπορεί να επικαλεστεί την πιθανότητα ένα αμφισβητούμενο έγγραφο να έχει παραχθεί ή τροποποιηθεί από ένα ΜΓΜ, εισάγοντας έτσι αμφιβολία σχετικά με τη συγγραφική πατρότητα ή την πρόθεση. Οι εισαγγελείς, με τη σειρά τους, ενδέχεται να δυσκολευτούν περισσότερο να ισχυριστούν ότι η υφολογική ευθυγράμμιση μεταξύ της γραφής ενός ατόμου και ενός αμφισβητούμενου κειμένου υποστηρίζει ισχυρά την συγγραφική πατρότητα όταν οι εναλλακτικές λύσεις με υποβοήθηση ΤΝ είναι εύλογες. Η Giray (2024) δείχνει, στον ακαδημαϊκό τομέα, πώς οι ψευδείς κατηγορίες για συγγραφή με υποβοήθηση ΤΝ επηρεάζουν δυσανάλογα τους συγγραφείς που γράφουν σε δεύτερη γλώσσα ή με διακριτά υφολογικά προφίλ, εγείροντας ανησυχίες σχετικά με τις διακρίσεις και τη δέουσα διαδικασία που είναι άμεσα σχετικές με τη δικανική πρακτική.

Δικανική εργασία	Τυπικό ερώτημα υπόθεσης	Ευκαιρία παρεχόμενη από ΜΓΜ	Κύρια απειλή στην εποχή των ΜΓΜ	Τρέχουσες προσεγγίσεις (ενδεικτικές)	Γνωστές αδυναμίες	Συνιστώμενη δικανική προσαρμογή
Απόδοση ανθρώπινης πατρότητας (παραδοσιακή)	«Έγραψε ο Χ αυτό το κείμενο;»	Η ομοιότητα βάσει διανυσματικών αναπαραστάσεων (embeddings) και τα παραδείγματα πιθανοκρατίας/ ΜΓΜ στο ρόλο κριτή μπορούν να συμπληρώσουν την υφομετρία και την ποιοτική ανάλυση	Υφολογική συγκάλυψη, επιθέσεις παράφρασης, συγγραφή με υποβοήθηση ΤΝ θολώνουν τα όρια του συγγραφέα	Υβριδική υφομετρία + Μετασχηματιστές/ ΜΓΜ στο ρόλο των κριτών	Πτώση απόδοσης σε σύντομα, θορυβώδη ή υβριδικά κείμενα, ευαισθησία στο θεματικό πεδίο	Αντιμετώπιση της εμπλοκής ΤΝ ως ανταγωνιστικής υπόθεσης, αναφορά αβεβαιότητας, απαίτηση επικύρωσης ταιριαστής με το είδος κειμένου
Επαλήθευση πατρότητας	«Είναι αυτά τα κείμενα γραμμένα από το ίδιο άτομο;»	Τα ΜΓΜ μπορούν να βοηθήσουν την κατά ζεύγη σύγκριση και στη διαλογή συνόλων υποψηφίων	Η μίμηση που καθοδηγείται από ΜΓΜ υποβαθμίζει τα ποσοστά αναφοράς (baselines) της επαλήθευσης	Κατά ζεύγη αξιολόγηση με ΜΓΜ, υφομετρική επαλήθευση	Ευάλωτη σε αντιθετική επανασυγγραφή, περιορισμένη ερμηνευσιμότητα	Χρήση σύγκλισης πολλαπλών μεθόδων, ρητός έλεγχος ανθεκτικότητας στη συγκάλυψη
Ανίχνευση κειμένου ΤΝ (δυαδική)	«Έχει γραφτεί αυτό το κείμενο από ΤΝ;»	Υψηλή ακρίβεια σε συγκεκριμένα Σώματα Κειμένων Αναφοράς (benchmarks)	Ψευδώς θετικά (ιδίως σε συγγραφείς δεύτερης γλώσσας), αντιθετική διαφυγή (π.χ. ομόγλυφα), υβριδικά κείμενα	Ανίχνευση βασισμένη σε ταξινομητές, σύνολα υφομετρικών χαρακτηριστικών	Ζητήματα προκατάληψης και δικαιοσύνης, φτωχή αξιοπιστία σε επεξεργασίες ή σύντομα τμήματα κειμένου	Αποφυγή αποκλειστικής εξάρτησης για αποφάσεις υψηλού ρίσκου, παρουσίαση ποσοστών σφάλματος ανά δημογραφική/ γλωσσική ομάδα
Απόδοση πηγής μοντέλου (ΤΝ-σε-μοντέλο)	«Ποιο μοντέλο παρήγαγε πιθανώς αυτό;»	Τα συνθετικά κείμενα ενδέχεται να φέρουν υπογραφές ειδικές για κάθε μοντέλο	Η ραγδαία εξέλιξη των μοντέλων μπορεί να διαγράψει τις υπογραφές, η μικρορύθμιση ανοιχτού κώδικα περιπλέκει τους ισχυρισμούς πηγής	Πολυκατηγορικοί ανιχνευτές, υφομετρικά αποτυπώματα/ υφομετρικές υπογραφές μοντέλων	Απαιτεί επικαιροποιημένα επισημειωμένα δεδομένα, ασταθής μεταξύ εκδόσεων	Χρήση ισχυρισμών με επίγνωση της έκδοσης, διατήρηση κυλιόμενων Σωμάτων Κειμένων Αναφοράς
Ανάλυση συν-συγγραφής Ανθρώπου–ΜΓΜ	«Ποιος είναι ο βαθμός/ρόλος της εμπλοκής της ΤΝ στο εξεταζόμενο κείμενο;»	Η αναγνώριση ρόλου και η εκτίμηση επιρροής ευθυγραμμίζονται με πραγματικά δικανικά ερωτήματα	Η ασάφεια ρόλων και η επεξεργασία συσκοτίζουν τα ίχνη	Πλαίσια λεπτομερούς (fine-grained) ρόλου & εμπλοκής, ανάλυση σε επίπεδο κειμενικού αποσπάσματος	Περιορισμένη απόδοση σε εναλλασσόμενα ή έντονα επεξεργασμένα κείμενα	Μετατόπιση της αναφοράς από δυαδικές ετικέτες σε διαβαθμισμένη εμπλοκή με όρια εμπιστοσύνης
Ανίχνευση αλλαγής ύφους / παρεμβολής	«Έχει γίνει συρραφή ή αφανής συγγραφή (ghostwriting) στο έγγραφο;»	Η ανίχνευση αλλαγής ύφους σε επίπεδο πρότασης μπορεί να επισημάνει εσωτερικές μετατοπίσεις	Η εξομάλυνση (smoothing) από ΜΓΜ μπορεί να καλύψει τις ανθρώπινες μεταβάσεις, μίξεις πολλών συγγραφέων + ΤΝ περιπλέκουν τα σήματα	Ανιχνευτές αλλαγής ύφους υποβοηθούμενοι από ΜΓΜ, κυλιόμενη υφομετρία	Ψευδείς συναγερμοί σε θεματικές αλλαγές/αλλαγές γένους, απαιτεί προσεκτική βαθμονόμηση	Συνδυασμός στοιχείων λόγου/ πλαισίου με υφομετρικά σήματα
Υδατογράφηση (όπου είναι διαθέσιμη)	«Μπορούμε να επαληθεύσουμε τη χρήση ΜΓΜ μέσω ενσωματωμένων δεικτών;»	Δυνητικά ισχυρή, επαληθεύσιμη από μηχανή προέλευση	Εξάρτηση από την υιοθέτηση, αφαίρεση μέσω παράφρασης/ μετάφρασης, συμβιβασμοί ποιότητας/ ευθυγράμμισης	Στατιστικοί ανιχνευτές υδατογραφήματος	Όχι καθολική, μπορεί να αφαιρεθεί ή να αποφευχθεί	Να αντιμετωπίζεται ως υποστηρικτικό, όχι ως καθοριστικό στοιχείο, τεκμηρίωση δοκιμών αντοχής στην αφαίρεση
Κρυπτογραφική προέλευση / έλεγχοι οικοσυστήματος	«Υπάρχει αλυσίδα επιτήρησης για τη δημιουργία του κειμένου;»	Μετατοπίζει την ανίχνευση → επαλήθευση, μειώνοντας το βάρος της συναγωγής συμπερασμάτων	Περιορισμένη ανάπτυξη, κατακερματισμός πλατφορμών	Πλαίσια προέλευσης βασισμένα σε πλατφόρμες	Κενά κάλυψης	Προώθηση πολιτικών/ βιομηχανικών προτύπων, ενσωμάτωση στη δικανική αναφορά όπου είναι δυνατόν

5. Ανίχνευση και αντίμετρα

Το τρίτο ερευνητικό ερώτημα αφορά το πώς η δικανική γλωσσολογία μπορεί να προσαρμοστεί στην παρουσία των ΜΓΜ. Ένα συστατικό αυτής της προσαρμογής περιλαμβάνει τη βελτίωση της ανίχνευσης κειμένου ΤΝ και την ανάπτυξη συμπληρωματικών μηχανισμών προσδιορισμού της κειμενικής ταυτότητας. Η πρόσφατη έρευνα για την ανίχνευση κειμένων τεχνητής νοημοσύνης ακολουθεί τρεις κατευθύνσεις. Η πρώτη επιδιώκει τη βελτίωση των υφολογικών μεθόδων και των αυτόματων ταξινομητών που διακρίνουν αν ένα κείμενο είναι ανθρώπινο ή μηχανικά παραγόμενο. Η δεύτερη αναπτύσσει συστήματα «υδατογράφησης», δηλαδή τρόπους να ενσωματώνονται αόρατα σημάδια στα κείμενα που παράγει η τεχνητή νοημοσύνη, ώστε να μπορεί αργότερα να επαληθευτεί η προέλευσή τους. Η τρίτη επαναπροσδιορίζει το πρόβλημα ως κάτι πιο σύνθετο από ένα απλό «ναι ή όχι»: αντί να ρωτάμε μόνο αν ένα κείμενο είναι τεχνητό, προσπαθούμε να εντοπίσουμε ποια συγκεκριμένα σημεία του είναι τεχνητά, αξιοποιώντας μεθόδους που μαθαίνουν τόσο από κείμενα με γνωστή προέλευση όσο και από κείμενα χωρίς αυτή την πληροφορία.

Ο Πίνακας 1 συγκεντρώνει τους βασικούς ισχυρισμούς του άρθρου στις προηγούμενες δύο ενότητες, ευθυγραμμίζοντας τα πραγματικά δικανικά ερωτήματα με τις δυνατότητες διπλής χρήσης των ΜΓΜ, τους συγκεκριμένους τρόπους αστοχίας που απειλούν την αποδεικτική αξιοπιστία και τις πιο υπερασπίσιμες στρατηγικές προσαρμογής. Ο πίνακας δείχνει επίσης γιατί το πεδίο πρέπει να απομακρυνθεί από τη δυαδική ανίχνευση ως το βασικό προεπιλεγμένο δικανικό στόχο και να κινηθεί προς προσεγγίσεις πολλαπλών μεθόδων, ευαίσθητες στους διαφορετικούς ρόλους που παίζουν τα ΜΓΜ στην παραγωγή κειμένου και δεκτικές σε πολλαπλές μεθόδους επικύρωσης, οι οποίες είναι πιο συμβατές με τις απαιτήσεις της αποδοχής πειστηρίων στο δικαστικό σύστημα.

Πίνακας 1. Χαρτογράφηση των εργασιών δικανικής προέλευσης στην εποχή των ΜΓΜ σε βασικά ερωτήματα υποθέσεων, ευκαιρίες διπλής χρήσης, αναδυόμενες απειλές, τρέχουσες μεθοδολογικές προσεγγίσεις και συνιστώμενες προσαρμογές για εύρωστη και νομικά υπερασπίσιμη πρακτική.

6. Συζήτηση

6.1 Σύνθεση ευκαιριών και απειλών

Οι προηγούμενες ενότητες κατέδειξαν ότι τα ΜΓΜ επεκτείνουν και ταυτόχρονα περιορίζουν το τι μπορεί να επιτύχει η δικανική γλωσσολογία. Ως εργαλεία, προσφέρουν πιο ισχυρές και ευέλικτες μεθόδους για την απόδοση πατρότητας, υποστηρίζουν την από κοινού ανίχνευση και απόδοση μεταξύ γλωσσών και εργασιών, και επιτρέπουν λεπτομερείς, επεξηγήσιμες αναλύσεις σε κλίμακα. Ως πηγές κινδύνου, καθιστούν το ύφος ευκολότερο στη χειραγώγηση, πλημμυρί-

ζουν τους επικοινωνιακούς χώρους με συνθετικό κείμενο και εκθέτουν τους περιορισμούς των τρεχόντων εργαλείων ανίχνευσης, ιδιαίτερα όσον αφορά την προκατάληψη και την ανθεκτικότητα έναντι αντιθετικών επιθέσεων (adversarial robustness).

Η βιβλιογραφία σχετικά με την υφομετρία και την ανίχνευση των ΜΓΜ σε κείμενα υποδηλώνει ότι τα κείμενα που παράγονται από ΜΓΜ εξακολουθούν να διαφέρουν συστηματικά από την ανθρώπινη γραφή και ότι αυτές οι διαφορές μπορούν συχνά να αξιοποιηθούν για ταξινόμηση και απόδοση υπό ελεγχόμενες συνθήκες. Ταυτόχρονα, μελέτες για την προκατάληψη των ανιχνευτών, τις αντιθετικές επιθέσεις και τα υβριδικά έγγραφα ανθρώπου–ΤΝ αποδεικνύουν ότι φαινομενικά ισχυρές μέθοδοι μπορούν να αποτύχουν ακριβώς στα περιβάλλοντα που μοιάζουν περισσότερο με πραγματικά δικανικά προβλήματα. Το αποτέλεσμα είναι ένα τοπίο στο οποίο οι ευκαιρίες και οι απειλές είναι στενά αλληλένδετες: οι ίδιες αρχιτεκτονικές που επιτρέπουν βελτιωμένη απόδοση και ανίχνευση παράγουν επίσης τα κείμενα που προκαλούν προβλήματα αξιοπιστίας σε αυτά ακριβώς τα συστήματα.

6.2 Απάντηση στα ερευνητικά ερωτήματα

Ιδωμένα μέσα από το πρίσμα του πρώτου ερευνητικού ερωτήματος, που αφορά την ενίσχυση της δικανικής μεθοδολογίας, τα στοιχεία δείχνουν ότι τα ΜΓΜ μπορούν να επεκτείνουν σημαντικά τις αναλυτικές ικανότητες της δικανικής γλωσσολογίας. Οι προσεγγίσεις απόδοσης που βασίζονται σε μπεϋζιανά μοντέλα και διανυσματικές αναπαραστάσεις, τα πλαίσια πολυδύναμης ανίχνευσης και απόδοσης πηγής μοντέλου, καθώς και η ανίχνευση αλλαγής ύφους σε επίπεδο πρότασης, καταδεικνύουν ότι τα ΜΓΜ μπορούν να αξιοποιηθούν για την εκτέλεση εργασιών που προηγουμένως ήταν είτε ανέφικτες είτε σημαντικά πιο περιορισμένες σε εύρος (Hu et al. 2024, Choi et al. 2025, Rao et al. 2025, Römisch et al. 2025). Αυτές οι εξελίξεις λειτουργούν παράλληλα και δεν αντικαθιστούν τις παραδοσιακές υφομετρικές και ποιοτικές μεθόδους, ανοίγοντας δυνατότητες για υβριδικές ροές εργασίας ανθρώπου–ΤΝ, στις οποίες κάθε σύστημα αντισταθμίζει τις αδυναμίες του άλλου.

Σε σχέση με το δεύτερο ερευνητικό ερώτημα, που αφορά τις απειλές για την παραδοσιακή πρακτική, οι μελέτες που εξετάστηκαν καθιστούν σαφές ότι τα ΜΓΜ θέτουν μια σοβαρή πρόκληση στην απόδοση πατρότητας που βασίζεται στην ιδιόλεκτο και στην αξιοπιστία της ανίχνευσης κειμένου ΤΝ. Η ευκολία μίμησης ύφους και συγκάλυψης, τα υψηλά και άνισα ποσοστά ψευδώς θετικών αποτελεσμάτων των υπαρχόντων ανιχνευτών και η ευπάθειά τους σε απλές χειραγωγήσεις, καθώς και η δυσκολία ανάλυσης υβριδικών κειμένων ανθρώπου–ΤΝ, υπονομεύουν την απλή εφαρμογή των υπαρχουσών μεθόδων (Alperin et al. 2025, Liang et al. 2023, Creo & Pudasaini 2024, Zeng et al. 2024). Ταυτόχρονα, τα πορίσματα από τις σχετικές υφομετρικές μελέτες τα οποία δείχνουν ότι τα κείμενα που παράγονται από ΜΓΜ διατηρούν

διακριτές υπογραφές παρέχουν κάποιο έδαφος για συγκρατημένη αισιοδοξία ότι μπορούν να αναπτυχθούν νέες μέθοδοι για την αντιμετώπιση αυτών των προκλήσεων.

Το τρίτο ερευνητικό ερώτημα θέτει το ζήτημα του πώς ο κλάδος πρέπει να προσαρμοστεί ώστε να διατηρήσει την αξιοπιστία και τη νομική αποδεκτότητα υπό το πρίσμα αυτών των εξελίξεων. Η βιβλιογραφία για την ανίχνευση και τα αντίμετρα υποδεικνύει αναδυόμενες στρατηγικές, συμπεριλαμβανομένων πιο ισχυρών προσεγγίσεων βασισμένων σε ταξινομητές[15] νέα υφομετρικά χαρακτηριστικά, μηχανισμών υδατογράφησης και προέλευσης, καθώς και λεπτομερών (fine-grained), συχνά ημιεπιβλεπόμενων παραδειγμάτων ανίχνευσης που υπερβαίνουν τη δυαδική ταξινόμηση ανθρώπου-TN (Macko 2025, Przystalski et al. 2025, Kirchenbauer et al. 2023, Cheng et al. 2024, Qazi et al. 2024). Ωστόσο, οι υπάρχουσες μελέτες υπογραμμίζουν επίσης σημαντικά κενά. Πολλές μέθοδοι δεν έχουν ακόμη επικυρωθεί στα είδη σύντομων, άτακτων και υβριδικών κειμένων που είναι τυπικά στις δικανικές υποθέσεις, ούτε έχουν αξιολογηθεί συστηματικά για την ισότιμη αντιμετώπιση (fairness) που επιφυλάσσουν σε διάφορες γλωσσικές και δημογραφικές ομάδες. Οι ακόλουθες ενότητες σχετικά με τις συνέπειες, τους περιορισμούς και τις μελλοντικές κατευθύνσεις συζητούν αυτά τα ευρήματα και τα αναπτύσσουν σε μια πιο ρητά κανονιστική περιγραφή για το πώς πρέπει να ανταποκριθεί η δικανική γλωσσολογία.

7. Συνέπειες για την πρακτική και το δίκαιο

Οι δικανικοί γλωσσολόγοι θα χρειαστεί να επεκτείνουν το μεθοδολογικό τους ρεπερτόριο προκειμένου να εργαστούν αποτελεσματικά σε ένα περιβάλλον κορεσμένο από την TN. Η επάρκεια στην ποιοτική ανάλυση και την παραδοσιακή υφομετρία παραμένει απαραίτητη, αλλά πλέον δεν φτάνει. Οι επαγγελματίες πρέπει επίσης να κατανοήσουν τη βασική λειτουργία, τα δυνατά σημεία και τους περιορισμούς των ΜΓΜ και των μοντέλων ανίχνευσης, συμπεριλαμβανομένων των προφίλ προκατάληψής τους και της επιδεκτικότητάς τους σε αντιθετική χειραγώγηση. Οι υβριδικές αναλυτικές στρατηγικές, στις οποίες εργαλεία βασισμένα σε ΜΓΜ χρησιμοποιούνται για τη δημιουργία υποθέσεων, τον εντοπισμό προτύπων ή την εκτέλεση αρχικού ελέγχου,

15. Προσεγγίσεις βασισμένες σε ταξινομητές (classifier-based approaches): Είναι μέθοδοι μηχανικής μάθησης που στοχεύουν στην αυτόματη κατάταξη δεδομένων σε προκαθορισμένες κατηγορίες. Ένας ταξινομητής είναι ένας αλγόριθμος ο οποίος «εκπαιδεύεται» σε ένα σύνολο παραδειγμάτων με γνωστή ταυτότητα (π.χ. κείμενα γραμμένα από ανθρώπους και κείμενα παραγόμενα από τεχνητή νοημοσύνη) και μαθαίνει να αναγνωρίζει τα διακριτικά χαρακτηριστικά κάθε κατηγορίας. Στη συνέχεια, μπορεί να εφαρμοστεί σε νέα, άγνωστα κείμενα για να προβλέψει την προέλευσή τους. Στο πλαίσιο της δικανικής γλωσσολογίας, τέτοιοι ταξινομητές αξιοποιούν συνήθως γλωσσικά χαρακτηριστικά, όπως λεξιλογική ποικιλία, συντακτικά μοτίβα ή στατιστικά μεγέθη της υφομετρίας, για να διακρίνουν αν ένα κείμενο είναι ανθρώπινης ή μηχανικής προέλευσης. Γνωστοί τύποι ταξινομητών περιλαμβάνουν τα δέντρα απόφασης, τις μηχανές διανυσμάτων υποστήριξης (Support Vector Machines) και τα νευρωνικά δίκτυα.

και οι ανθρώπινοι εμπειρογνώμονες που διεξάγουν την πλαισιωμένη ερμηνεία και την τελική αξιολόγηση, είναι πιθανό να αποτελέσουν ιδιαίτερα παραγωγικές λύσεις.

Τα προγράμματα κατάρτισης στη δικανική γλωσσολογία πρέπει επομένως να ενσωματώσουν τον αλφαβητισμό στην ΤΝ (AI literacy) παράλληλα με τη γλωσσολογία και τη νομική. Αυτό περιλαμβάνει εκπαίδευση σε μεθόδους απόδοσης βασισμένες σε διανυσματικές αναπαραστάσεις και μπεϋζιανή λογική (Hu et al. 2024, Abbas 2025), πλαίσια πολυδύναμης ανίχνευσης και απόδοσης (Rao et al. 2025), τεχνικές υφομετρικής ανίχνευσης (Przystalski et al. 2025, Berriche & LarabiMarieSainte 2024) και τις βασικές αρχές της επεξηγήσιμης ΤΝ (Roemling et al. 2024). Η συνεχιζόμενη επαγγελματική ανάπτυξη θα είναι απαραίτητη για τους καθιερωμένους επαγγελματίες, δεδομένου του γρήγορου ρυθμού της τεχνολογικής αλλαγής. Επαγγελματικοί οργανισμοί όπως η Διεθνής Ένωση για τη Δικανική και Νομική Γλωσσολογία (IAFLL) μπορούν να διαδραματίσουν κεντρικό ρόλο στη διατύπωση βέλτιστων πρακτικών, τη διοργάνωση εκπαίδευσης και την επικαιροποίηση των κατευθυντήριων γραμμών δεοντολογίας για την αντιμετώπιση ζητημάτων που σχετίζονται με την ΤΝ.

Η διεπιστημονική συνεργασία θα καταστεί επίσης ολοένα και πιο σημαντική. Πολλές από τις πιο υποσχόμενες εξελίξεις, όπως η πολυδύναμη ανίχνευση και απόδοση, η λεπτομερής εκτίμηση επιρροής και η υδατογράφηση, προκύπτουν από συνεργασίες μεταξύ επιστημόνων της πληροφορικής, γλωσσολόγων και νομικών. Τα προγράμματα και τα εργαστήρια δικανικής γλωσσολογίας θα ωφελούνταν από τη δημιουργία θεσμικών δεσμών με τμήματα πληροφορικής, επιστήμης της πληροφορίας και νομικής για τη διευκόλυνση τέτοιων συνεργασιών.

Οι παρεμβάσεις πολιτικής ενδέχεται να περιλαμβάνουν απαιτήσεις για αποκάλυψη της βοήθειας ΤΝ σε ορισμένους τομείς, όπως η νομική σύνταξη, οι ακαδημαϊκές υποβολές ή οι κανονιστικές καταθέσεις. Εάν εφαρμοστούν και επιβληθούν τέτοιες αποκαλύψεις, η δικανική εργασία ενδέχεται να μετατοπιστεί από την ανίχνευση στην επαλήθευση, η οποία είναι συνήθως πιο διαχειρίσιμη. Τα κρυπτογραφικά συστήματα προέλευσης και τα σχήματα υδατογράφησης, εάν υιοθετηθούν ευρέως, θα μπορούσαν να διευκολύνουν αυτή τη μετατόπιση. Ελλείψει αυτών, ωστόσο, η δικανική γλωσσολογία θα επωμιστεί σημαντικό βάρος στην προσπάθεια διάκρισης του ανθρώπινου περιεχομένου από το περιεχόμενο που υποβοηθείται από ΤΝ εκ των υστέρων (post hoc).

Τα αναδυόμενα παραδείγματα λεπτομερούς ανίχνευσης που συζητήθηκαν παραπάνω έχουν επίσης νομικές προεκτάσεις. Τα δικαστήρια ίσως χρειαστεί να απομακρυνθούν από την αντιμετώπιση της εμπλοκής της ΤΝ ως απλής δυαδικής συνθήκης και αντ' αυτού να εξετάσουν βαθμούς και τρόπους εμπλοκής κατά την αξιολόγηση της πατρότητας, της πρωτοτυπίας και της ευθύνης. Οι καταθέσεις εμπειρογνωμόνων θα μπορούσαν τότε να επικεντρωθούν στην ανακατασκευή της πιθανής συμβολής ανθρώπινων και μηχανικών παραγόντων σε μια συνεργατική διαδικασία συγγραφής, αντί να πιστοποιούν απλώς ότι ένα κείμενο είναι ή δεν είναι «παραγόμενο από ΤΝ».

8. Περιορισμοί

Η ανάλυση που παρουσιάζεται εδώ υπόκειται σε αρκετούς περιορισμούς. Πρώτον, η εμπειρική βιβλιογραφία σχετικά με τα ΜΓΜ και τη δικανική γλωσσολογία εξελίσσεται ραγδαία. Πολλές από τις μελέτες που αναφέρονται είναι πρόσφατες προδημοσιεύσεις (preprints) και τα ευρήματά τους ενδέχεται να ξεπεραστούν από μεταγενέστερες εργασίες ή από την κυκλοφορία νέων ΜΓΜ. Η παρατήρηση των Zaitsu et al. (2025) ότι τα νεότερα μοντέλα παράγουν κείμενα που βρίσκονται εγγύτερα στην ανθρώπινη γραφή σε σχέση με τα παλαιότερα μοντέλα καταδεικνύει τη χρονική αστάθεια των συζητούμενων φαινομένων.

Δεύτερον, μεγάλο μέρος της διαθέσιμης έρευνας εστιάζει στην αγγλική ή σε μικρό αριθμό γλωσσών υψηλών πόρων (highresource languages) και σε περιορισμένα είδη λόγου, όπως ο ακαδημαϊκός πεζός λόγος, τα σύντομα δοκίμια και η δημιουργική γραφή. Αντίθετα, οι πραγματικές δικανικές υποθέσεις συχνά περιλαμβάνουν κείμενα άτυπα, πολύγλωσσα, με εναλλαγή κώδικα (codeswitched) ή άλλως αποκλίνοντα. Ως εκ τούτου, απαιτείται προσοχή κατά την παρέκταση συμπερασμάτων από τα τρέχοντα σώματα κειμένων αναφοράς (benchmarks) σε πραγματικά δικανικά περιβάλλοντα.

Τρίτον, η νομική ανάλυση στο παρόν άρθρο πλαισιώνεται πρωτίστως σε σχέση με το ομοσπονδιακό δίκαιο αποδείξεων των ΗΠΑ. Άλλες δικαιοδοσίες εφαρμόζουν διαφορετικά πρότυπα για την εμπειρογνωμοσύνη, και οι νομικές κουλτούρες διαφέρουν ως προς τη δεκτικότητά τους σε πιθανολογικές και υπολογιστικές μεθόδους. Οι επιπτώσεις των ΜΓΜ για τη δικανική γλωσσολογία ενδέχεται επομένως να διαφέρουν μεταξύ των νομικών συστημάτων. Η συγκριτική εμπειρική έρευνα σχετικά με τις δικαστικές αντιδράσεις σε γλωσσικά τεκμήρια σχετιζόμενα με την ΤΝ παραμένει σπάνια.

Τέλος, το άρθρο εστιάζει σε κειμενικά τεκμήρια και δεν εξετάζει πολυτροπικά ΜΓΜ που επεξεργάζονται ομιλία, εικόνες ή άλλες τροπικότητες,[16] παρόλο που τέτοια συστήματα ενδέχεται σύντομα να γίνουν σχετικά με τη δικανική πρακτική.

16. Τροπικότητα (modality): Στο πλαίσιο της τεχνητής νοημοσύνης, ο όρος αναφέρεται στους διαφορετικούς τύπους δεδομένων ή «καναλιών» επικοινωνίας που μπορεί να επεξεργαστεί ένα σύστημα. Κάθε τροπικότητα αντιστοιχεί σε διαφορετική μορφή πληροφορίας: το κείμενο, η ομιλία, η εικόνα, το βίντεο και ο ήχος αποτελούν διακριτές τροπικότητες. Τα «πολυτροπικά» (multimodal) μοντέλα είναι συστήματα που μπορούν να δέχονται και να παράγουν περισσότερες από μία τροπικότητες, π.χ. να αναλύουν ταυτόχρονα μια εικόνα και να απαντούν με κείμενο, ή να μεταγράφουν ομιλία σε γραπτό λόγο.

9. Μελλοντικές κατευθύνσεις

9.1 Ερευνητικές προτεραιότητες

Κάποιες ερευνητικές κατευθύνσεις φαίνονται ιδιαίτερα επείγουσες. Μία από αυτές είναι η ανάπτυξη μεθόδων ανίχνευσης με σαφώς χαρακτηρισμένα ποσοστά σφάλματος σε δημογραφικές και γλωσσικές ομάδες. Δεδομένης της τεκμηριωμένης προκατάληψης κατά των μη φυσικών ομιλητών της αγγλικής στους τρέχοντες ανιχνευτές (Liang et al. 2023), ο σχεδιασμός και η επικύρωση με επίγνωση της απαίτησης για ισοτιμία θα πρέπει να αντιμετωπίζονται ως βασική απαίτηση και όχι ως δευτερεύουσα σκέψη.

Μια δεύτερη προτεραιότητα είναι η διαχρονική έρευνα που παρακολουθεί πώς εξελίσσονται οι υφολογικές υπογραφές των ΜΓΜ με την πάροδο του χρόνου. Καθώς τα μοντέλα ενημερώνονται και οι πρακτικές εκπαίδευσης αλλάζουν, τα χαρακτηριστικά που διακρίνουν την ανθρώπινη από τη μηχανική γραφή σήμερα ενδέχεται να καταστούν παρωχημένα. Τα τακτικά ενημερωμένα σώματα κειμένων αναφοράς που περιλαμβάνουν κείμενα από νέα μοντέλα και ποικίλους ανθρώπινους συγγραφείς θα βοηθούσαν στη διατήρηση της συνάφειας των μεθόδων ανίχνευσης και απόδοσης.

Μια τρίτη προτεραιότητα είναι η συστηματική μελέτη των πρακτικών συνεργατικής συγγραφής ανθρώπου–ΤΝ. Απαιτείται εμπειρική εργασία σχετικά με το πώς οι συγγραφείς χρησιμοποιούν τα ΜΓΜ σε διαφορετικά πλαίσια, πώς αυτή η χρήση επηρεάζει τις ιδιολέκτους τους και ποιοι υφολογικοί δείκτες παραμένουν ανιχνεύσιμοι στα υβριδικά κείμενα. Το έργο των Zeng et al. (2024) σχετικά με την ανίχνευση σε επίπεδο πρότασης σε συνεργατικά κείμενα παρέχει ένα σημείο εκκίνησης, αλλά απομένουν πολλά να κατανοηθούν σχετικά με τη δυναμική της συνεργασίας, τα πρότυπα επεξεργασίας και τα προκύπτοντα δικανικά ίχνη.

9.2 Πολιτικές και διακυβέρνηση της τεχνητής νοημοσύνης

Τέλος, η δικανική γλωσσολογία θα πρέπει να συμμετάσχει ενεργά στις συζητήσεις πολιτικής για τη διακυβέρνηση της ΤΝ. Ερωτήματα σχετικά με την υποχρεωτική γνωστοποίηση της βοήθειας ΤΝ, την τυποποίηση των μηχανισμών προέλευσης, τη ρύθμιση της υδατογράφησης και την προστασία των ατόμων από αδικαιολόγητες κατηγορίες χρήσης ΤΝ δεν είναι καθαρά τεχνικά. Εμπλέκουν τα ατομικά δικαιώματα, την ακαδημαϊκή ελευθερία και τη δέουσα διαδικασία. Οι επαγγελματικοί φορείς στη δικανική γλωσσολογία, σε συνεργασία με ενδιαφερόμενα μέρη από τον νομικό, τεχνικό και κοινωνικό χώρο, μπορούν να συνεισφέρουν εμπειρογνωμοσύνη στον σχεδιασμό πολιτικών που υποστηρίζουν τόσο την αποτελεσματική διερεύνηση αδικημάτων όσο και την προστασία των θεμιτών πρακτικών συγγραφής με υποβοήθηση ΤΝ.

Ο διεθνής συντονισμός θα αποβεί ιδιαίτερα σημαντικός. Τα ψηφιακά κείμενα διασχίζουν συστηματικά τα όρια των δικαιοδοσιών και τα συστήματα ΤΝ αναπτύσσονται και χρησιμοποιούνται διακρατικά. Κοινά πρότυπα για τις γνωστοποιήσεις που σχετίζονται με την ΤΝ, τους μηχανισμούς προέλευσης και την αξιολόγηση των τεκμηρίων που διαμεσολαβούνται από την ΤΝ θα διευκόλυναν τη διασυνοριακή συνεργασία και θα μείωναν τον κίνδυνο ασυνεπών ή άδικων αποτελεσμάτων.

10. Συμπεράσματα

Τα ΜΓΜ θέτουν στη δικανική γλωσσολογία ένα βαθύ σύνολο προκλήσεων και ευκαιριών. Ως αναλυτικά εργαλεία, επιτρέπουν εξελιγμένες μεθόδους απόδοσης πατρότητας, πολυδύναμη και πολύγλωσση ανάλυση σε κλίμακα, καθώς και λεπτομερή υφολογική ανίχνευση και εξήγηση. Ως παραγωγικά συστήματα, διευκολύνουν τη χειραγώγηση του ύφους, παράγουν συνθετικά κείμενα που είναι δύσκολο να ανιχνευθούν και υπονομεύουν τις παραδοσιακές παραδοχές για την ιδιόλεκτο και την πατρότητα.

Τα στοιχεία που εξετάστηκαν σε αυτό το άρθρο υποδηλώνουν ότι τα τρέχοντα εργαλεία ανίχνευσης ΤΝ στα κείμενα δεν είναι ακόμη κατάλληλα για δικανικές εφαρμογές υψηλού ρίσκου. Τα υψηλά ποσοστά ψευδώς θετικών αποτελεσμάτων για ορισμένους πληθυσμούς, η ευπάθειά τους σε απλές αντιθετικές επιθέσεις και η δυσκολία τους στον χειρισμό υβριδικών κειμένων σημαίνουν ότι συχνά αποτυγχάνουν να ικανοποιήσουν τις απαιτήσεις τύπου Daubert για αξιοπιστία. Ταυτόχρονα, η υφομετρική και πολυδύναμη έρευνα δείχνει ότι το κείμενο που παράγεται από ΜΓΜ εξακολουθεί να διαφέρει με συστηματικούς τρόπους από την ανθρώπινη γραφή και ότι πιο ολοκληρωμένα παραδείγματα ανίχνευσης μπορούν να συλλάβουν πτυχές της εμπλοκής της ΤΝ πιο αποτελεσματικά από τη δυαδική ταξινόμηση.

Το μέλλον της δικανικής γλωσσολογίας στην εποχή των ΜΓΜ θα εξαρτηθεί από την προθυμία του κλάδου να προσαρμοστεί. Οι υβριδικές μεθοδολογίες που συνδυάζουν την παραδοσιακή υφολογική ανάλυση, εργαλεία βασισμένα σε ΜΓΜ, υφομετρική μοντελοποίηση και επεξηγήσιμη ΤΝ προσφέρουν υποσχέσεις. Η ισχυρή επικύρωση, με ρητή προσοχή στην προκατάληψη και την απαίτηση για ισοτιμία, πρέπει να καταστεί αδιαπραγμάτευτη. Η διεπιστημονική συνεργασία και η εμπλοκή με νομικά και πολιτικά πλαίσια θα είναι απαραίτητες για να διασφαλιστεί ότι τα γλωσσικά τεκμήρια παραμένουν επιστημονικώς αξιόπιστα όσο και νομικώς αποδεκτά.

Η βασική προϋπόθεση της δικανικής γλωσσολογίας, δηλαδή ότι η χρήση της γλώσσας παρέχει αποδείξεις για τον δημιουργό της, παραμένει ορθή. Αυτό που έχει αλλάξει είναι η πολυπλοκότητα της διαδικασίας παραγωγής και το εύρος των πιθανών ανθρώπινων και υπο-

λογιστικών συντελεστών σε κάθε δεδομένο κείμενο. Η αντιμετώπιση αυτής της πολυπλοκότητας θα απαιτήσει διαρκείς προσπάθειες από ερευνητές, επαγγελματίες, δικαστήρια και φορείς χάραξης πολιτικής. Το διακύβευμα, όσον αφορά τα ατομικά δικαιώματα, τη θεσμική νομιμοποίηση και την ακεραιότητα της νομικής λήψης αποφάσεων, είναι ουσιαστικό. Ο κλάδος αντιμετωπίζει τώρα μια επιλογή ανάμεσα στην αντίσταση στην αλλαγή με κίνδυνο την περιθωριοποίηση, ή την υιοθέτηση μιας μεθοδολογικής και θεσμικής μεταρρύθμισης, βοηθώντας στη διαμόρφωση μιας πιο ισχυρής και δίκαιης δικανικής απάντησης στην εποχή της παραγωγικής τεχνητής νοημοσύνης.

Βιβλιογραφικές αναφορές

Abbas, M. (2025). Attribution quality in AIgenerated content: Benchmarking style embeddings and LLM judges. *arXiv preprint.* https://arxiv.org/abs/2510.13898

Ainsworth, J., & Juola, P. (2019). Who wrote this: Modern forensic authorship analysis as a model for valid forensic science. *Washington University Law Review, 96*(5), 1161–1189.

Alperin, K., Leekha, R., Uchendu, A., Nguyen, T., Medarametla, S., Capote, C. L., Aycock, S., & Dagli, C. (2025). Masks and mimicry: Strategic obfuscation and impersonation attacks on authorship verification. *arXiv preprint.* https://arxiv.org/abs/2503.19099

Berriche, L., & LarabiMarieSainte, S. (2024). Unveiling ChatGPT text using writing style. *Heliyon, 10*(12), e32976. https://doi.org/10.1016/j.heliyon.2024.e32976

Bisztray, T., Cherif, B., Dubniczky, R. A., Gruschka, N., Borsos, B., Ferrag, M. A., Kovacs, A., Mavroeidis, V., & Tihanyi, N. (2025). I know which LLM wrote your code last summer: LLM generated code stylometry for authorship attribution. *arXiv preprint.* https://arxiv.org/abs/2506.17323

Burrows, J. F. (2002). 'Delta': A measure of stylistic difference and a guide to likely authorship. *Literary and Linguistic Computing, 17*(3), 267–287. https://doi.org/10.1093/llc/17.3.267

Cheng, Z., Zhou, L., Jiang, F., Wang, B., & Li, H. (2024). Beyond binary: Towards finegrained LLMgenerated text detection via role recognition and involvement measurement. *arXiv preprint.* https://arxiv.org/abs/2410.14259

Choi, S., Tan, Y. K., Meng, M. H., Ragab, M., Mondal, S., Mohaisen, D., & Aung, K. M. M. (2025). I can find you in seconds! Leveraging large language models for code authorship attribution. *arXiv preprint.* https://arxiv.org/abs/2501.08165

Christiansen, M. H., & Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, *39*, e62. https://doi.org/10.1017/S0140525X1500031X

Coulthard, M. (2004). Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics, 25*(4), 431–447. https://doi.org/10.1093/applin/25.4.431

Coulthard, M. (2010). Forensic linguistics: The application of language description in legal contexts. *Langage et Société, 132*(2), 15–33. https://doi.org/10.3917/ls.132.0015

Coulthard, M., & Johnson, A. (2007). *An introduction to forensic linguistics: Language in evidence.* Routledge.

Coulthard, M., Johnson, A., & Wright, D. (2017). *An introduction to forensic linguistics: Language in evidence* (2nd ed.). Routledge.

Creo, A., & Pudasaini, S. (2024). SilverSpeak: Evading AIgenerated text detectors using homoglyphs. *arXiv preprint.* https://arxiv.org/abs/2406.11239

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Dalalah, D., & Dalalah, O. M. A. (2023). The false positives and the false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *The International Journal of Management Education, 21*(2), 100822. https://doi.org/10.1016/j.ijme.2023.100822

Eder, M., Kestemont, M., & Rybicki, J. (2016). Stylometry with R: A Package for Computational Text Analysis. *The R Journal*, *8*(1), 107–121.

Evert, S., Proisl, T., Jannidis, F., Reger, I., Pielström, S., Schöch, C., & Vitt, T. (2017). Understanding and explaining Delta measures for authorship attribution. *Digital Scholarship in the Humanities, 32*(suppl_2), ii4–ii16. https://doi.org/10.1093/llc/fqx023

Giray, L. (2024). The problem with false positives: AI detection unfairly accuses scholars of AI plagiarism. *The Serials Librarian, 85*(5–6). https://doi.org/10.1080/0361526X.2024.2433256

Grant, T. (2007). Quantifying evidence in forensic authorship analysis. *International Journal of Speech, Language and the Law, 14*(1), 1–25. https://doi.org/10.1558/ijsll.v14i1.1

Hu, Z., Zheng, T., & Huang, H. (2024). A Bayesian approach to harnessing the power of LLMs in authorship attribution. *arXiv preprint.* https://arxiv.org/abs/2410.21716

Huang, B., Chen, C., & Shu, K. (2024). Authorship attribution in the era of LLMs: Problems, methodologies, and challenges. *arXiv preprint.* https://arxiv.org/abs/2408.08946

Juola, P., & Mikros, G. K. (2016). Cross-linguistic stylometric features: A preliminary investigation. In *Actes des 13èmes Journées internationales d'Analyse statistique des Données Textuelles (JADT 2016)* (pp. 787–794). Nice, France. https://jadt2016.sciencesconf.org

Juola, P., Mikros, G. K., & Vinsick, S. (2019). A comparative assessment of the difficulty of authorship attribution in Greek and in English. *Journal of the Association for Information Science and Technology*, *70*(1), 61–70. https://doi.org/10.1002/asi.24073

Kirchenbauer, J., Geiping, J., Wen, Y., Shu, M., Saifullah, K., Kong, K., Fernando, K., Saha, A., Goldblum, M., & Goldstein, T. (2023). On the reliability of watermarks for large language models. *arXiv preprint.* https://arxiv.org/abs/2306.04634

Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

Kumarage, T., & Liu, H. (2023). Neural authorship attribution: Stylometric analysis on large language models. *arXiv preprint.* https://arxiv.org/abs/2308.07305

La Cava, L., Macko, D., Móro, R., Srba, I., & Tagarelli, A. (2025). Authorship attribution in multilingual machinegenerated texts. *arXiv preprint.* https://arxiv.org/abs/2508.01656

Langacker, R. W. (1987). *Foundations of cognitive grammar: Vol. 1. Theoretical prerequisites*. Stanford University Press.

Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., & Zou, J. (2023). GPT detectors are biased against nonnative English writers. *Patterns, 4*(7), 100779. https://doi.org/10.1016/j.patter.2023.100779

Macko, D. (2025). Robustly finetuned LLM for binary and multiclass AIgenerated text detection. *arXiv preprint.* https://arxiv.org/abs/2506.01702

McMenamin, G. R. (2002). *Forensic linguistics: Advances in forensic stylistics.* CRC Press.

Mikros, G., & Perifanos, K. (2013). Authorship attribution in Greek tweets using multilevel author's n-gram profiles. In E. Hovy, V. Markman, C. H. Martell, & D. Uthus (Eds.), *Papers from the 2013 AAAI Spring Symposium "Analyzing Microtext", 25-27 March 2013, Stanford, California* (pp. 17–23). AAAI Press.

Mikros, G. (2025). Beyond the surface: Stylometric analysis of GPT4o's capacity for literary style imitation. *Digital Scholarship in the Humanities, 40*(2), 587–601. https://doi.org/10.1093/dsh/fqaf035

MirallesGonzález, P., HuertasTato, J., Martín, A., & Camacho, D. (2025). LLM oneshot style transfer for authorship attribution and verification. *arXiv preprint.* https://arxiv.org/abs/2510.13302

National Institute of Justice. (n.d.). *Law 101: Legal guide for the forensic expert – Daubert and Kumho decisions.* https://nij.ojp.gov/nij-hosted-online-training-courses/law-101-legal-guide-forensic-expert/pretrial/pretrial-rules-evidence/daubert-and-kumho-decisions

Nini, A. (2023). *A theory of linguistic individuality for authorship analysis*. Cambridge University Press. https://doi.org/10.1017/9781108974851

O'Sullivan, J. (2025). Stylometric comparisons of human versus AIgenerated creative writing. *Humanities and Social Sciences Communications, 12,* 1708. https://doi.org/10.1057/s41599-025-05986-3

Okulska, I., Stetsenko, D., Kołos, A., Karlińska, A., Głąbińska, K., & Nowakowski, A. (2023). StyloMetrix: An opensource multilingual tool for representing stylometric vectors. *arXiv preprint.* https://arxiv.org/abs/2309.12810

Przystalski, K., Argasiński, J. K., GrabskaGradzińska, I., & Ochab, J. K. (2025). Stylometry recognizes human and LLMgenerated texts in short samples. *Expert*

Systems with Applications, 296, 129001. https://doi.org/10.1016/j.eswa.2025.129001

Qazi, Z., Shiao, W., & Papalexakis, E. E. (2024). GPTgenerated text detection: Benchmark dataset and tensorbased detection method. *arXiv preprint.* https://arxiv.org/abs/2403.07321

Rao, Z., Mohamed, Y., Liu, S., & Liu, Z. (2025). Two birds with one stone: Multitask detection and attribution of LLMgenerated text. *arXiv preprint.* https://arxiv.org/abs/2508.14190

Rashidi, H. H., Fennell, B. D., Albahra, S., Hu, B., & Gorbett, T. (2023). The ChatGPT conundrum: Humangenerated scientific manuscripts misidentified as AI creations by AI text detection tool. *Journal of Pathology Informatics, 14,* 100342. https://doi.org/10.1016/j.jpi.2023.100342

Roemling, D., Scherrer, Y., & Miletic, A. (2024). Explainability of machine learning approaches in forensic linguistics: A case study in geolinguistic authorship profiling. *arXiv preprint.* https://arxiv.org/abs/2404.18510

Römisch, J., Gorovaia, S., Halchynska, M., Schmidt, G., & Yamshchikov, I. P. (2025). Better call Claude: Can LLMs detect changes of writing style? *arXiv preprint.* https://arxiv.org/abs/2508.00680

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems 30* (pp. 5998–6008). Curran Associates.

Wang, Z., Tripto, N. I., Park, S., Li, Z., & Zhou, J. (2025). Catch me if you can? Not yet: LLMs still struggle to imitate the implicit writing styles of everyday authors. *arXiv preprint.* https://arxiv.org/abs/2509.14543

Zaitsu, W., Jin, M., Ishihara, S., Tsuge, S., & Inaba, M. (2025). Stylometry can reveal artificial intelligence authorship, but humans struggle: A comparison of human and seven large language models in Japanese. *PLOS ONE, 20*(10), e0335369. https://doi.org/10.1371/journal.pone.0335369

Zeng, Z., Liu, S., Sha, L., Li, Z., Yang, K., Liu, S., Gašević, D., & Chen, G. (2024). Detecting AIgenerated sentences in human–AI collaborative hybrid texts: Challenges, strategies, and insights. *arXiv preprint.* https://arxiv.org/abs/2403.03506