ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ

Πρόγραμμα Μεταπτυχιακών Σπουδών

«Κυβερνοασφάλεια και Επιστήμη Δεδομένων »

Μεταπτυχιακή Διατριβή

Τίτλος Διατριβής	**Οπτική Αναγνώριση και Γλωσσική Αποκατάσταση Ελληνικών Παπύρων με χρήση Τεχνητής Νοημοσύνης** **Optical Character Recognition and Textual Restoration of Greek Papyri using Artificial Intelligence**
Ονοματεπώνυμο Φοιτητή	**Ευθύμιος Ανδρικάκης**
Πατρώνυμο	**Δημήτριος**
Αριθμός Μητρώου	**ΜΠΚΕΔ 24002**
Επιβλέπων	**Δρ. Γεώργιος Παπαστεφανάτος, Διδάσκων ΠΜΣ**

Ημερομηνία Παράδοσης **Ιούνιος 2026**

Τριμελής Εξεταστική Επιτροπή

Δρ. Γεώργιος Παπαστεφανάτος Διδάσκων ΠΜΣ	Δημήτριος Αποστόλου Καθηγητής	Δρ. Σταύρος Μαρούλης Διδάσκων ΠΜΣ

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά για την αποδοχή, τη στήριξη και την ουσιαστική επίβλεψη τον κ. Γεώργιο Παπαστεφανάτο, Διευθυντή Ερευνών του Ερευνητικού Κέντρου «Αθηνά», καθώς και τον κ. Βασίλειο Σταματόπουλο, ερευνητικό συνεργάτη στο Ε.Κ. «Αθηνά», για την πολύτιμη καθοδήγησή του. Επίσης, οφείλω θερμές ευχαριστίες στα δύο μέλη της εξεταστικής επιτροπής, κ. Δημήτριο Αποστόλου και κ. Σταύρο Μαρούλη, για την προσοχή τους, τον χρόνο τους και την αξιολόγηση της παρούσας εργασίας. Επιπλέον, θα ήθελα να ευχαριστήσω την οικογένειά μου και τους γονείς μου για την αμέριστη βοήθειά τους κατά την διάρκεια των σπουδών μου. Ιδιαίτερα, θα ήθελα να ευχαριστήσω την κοπέλα μου για την συμπαράσταση και την δύναμη που μου έδινε καθ΄όλη την διάρκεια της εκπόνησης της διπλωματικής μου εργασίας.

Περίληψη

Η ψηφιοποίηση και ανάγνωση αρχαίων ελληνικών παπύρων αποτελεί μια από τις μεγαλύτερες προκλήσεις της Υπολογιστικής Αρχαιογνωσίας, εξαιτίας της έντονης φθοράς του υλικού και της χρήσης συνεχούς γραφής (*scriptio continua*). Η παρούσα διπλωματική εργασία προτείνει ένα καινοτόμο, υβριδικό pipeline δύο σταδίων για την οπτική αναγνώριση και τη σημασιολογική αποκατάσταση παπυρικών κειμένων. Στο πρώτο στάδιο (Υπολογιστική Όραση), αξιοποιείται ο αλγόριθμος YOLO11s σε συνδυασμό με την τεχνική τεμαχισμού SAHI για τον εντοπισμό χαρακτήρων σε σαρώσεις υπερ-υψηλής ανάλυσης, ενώ η ταξινόμηση εκτελείται από ένα δίκτυο Vision Transformer (DeiT). Λόγω του αναπόφευκτου θορύβου του ακατέργαστου OCR (υψηλό CER και WER), το δεύτερο στάδιο εισάγει Μεγάλα Γλωσσικά Μοντέλα (LLMs) για τη γλωσσική αποκατάσταση. Για την αντιμετώπιση των παραισθήσεων (hallucinations) και της αποτυχίας των παραδοσιακών πυκνών διανυσμάτων (dense embeddings) σε ανορθόγραφη συνεχή γραφή, αναπτύχθηκε μια προσαρμοσμένη αρχιτεκτονική Ανάκτησης Επαυξημένης Παραγωγής (RAG), ικανή να λειτουργεί αξιόπιστα στα μορφολογικά μοτίβα της αρχαίας ελληνικής. Η συγκριτική αξιολόγηση σε τοπικά και εμπορικά μοντέλα αιχμής έδειξε ότι το ενσωματωμένο σύστημα RAG λειτουργεί ως απαραίτητος μηχανισμός γείωσης (grounding), βελτιώνοντας μετρήσιμα το Σφάλμα Χαρακτήρων (CER). Τέλος, αν και η αυστηρότητα της φιλολογικής μετάφρασης οδήγησε σε κατάρρευση των παραδοσιακών μετρικών (BLEU, METEOR), η αξιολόγηση μέσω BERTScore απέδειξε την ισχυρή σημασιολογική ανθεκτικότητα των μοντέλων, αναδεικνύοντας το σύστημα ως ένα πολύτιμο εργαλείο εξόρυξης νοήματος (sense-making) για τους ερευνητές της παπυρολογίας.

Λέξεις Κλειδιά:

Ψηφιακή Παπυρολογία, Υπολογιστική Όραση, Μεγάλα Γλωσσικά Μοντέλα (LLM), Ανάκτηση Επαυξημένης Παραγωγής (RAG), Vision Transformers, Οπτική Αναγνώριση Χαρακτήρων (OCR), Αποκατάσταση Κειμένου.

Abstract

The digitization and reading of ancient Greek papyri is one of the most formidable challenges in Computational Humanities, primarily due to severe material degradation and the historical use of continuous script (*scriptio continua*). This diploma thesis proposes an innovative, two-stage hybrid pipeline for the optical recognition and semantic restoration of papyrological texts. In the first stage (Computer Vision), the YOLO11s algorithm, augmented with the SAHI slicing technique, is employed for character bounding box detection on ultra-high-resolution scans, followed by a Vision Transformer (DeiT) for classification. Given the inherently noisy raw OCR output (characterized by high CER and WER), the second stage incorporates Large Language Models (LLMs) for textual restoration. To mitigate model hallucinations and overcome the tokenization failures of traditional dense embeddings on misspelled continuous text, a custom Retrieval-Augmented Generation (RAG) architecture was developed, tailored to the morphological patterns of ancient Greek. Comprehensive benchmarking across local and commercial state-of-the-art models demonstrated that this specialized RAG approach acts as a crucial grounding mechanism, measurably improving the Character Error Rate (CER). Finally, while the rigid syntactical requirements of translation caused traditional evaluation metrics (BLEU, METEOR) to collapse, evaluation via BERTScore revealed strong semantic resilience. The findings establish that while LLMs may struggle with publication-ready syntax, they excel at extracting the core conceptual meaning, positioning the proposed pipeline as a highly valuable sense-making tool for papyrologists.

Key Words:

Digital Papyrology, Computer Vision, Large Language Models (LLM), Retrieval-Augmented Generation (RAG), Vision Transformers, Optical Character Recognition (OCR), Textual Restoration.

Δήλωση Χρήσης Τεχνητής Νοημοσύνης:

Κατά την εκπόνηση της παρούσας διπλωματικής εργασίας, έγινε επικουρική χρήση εργαλείων Παραγωγικής Τεχνητής Νοημοσύνης (Generative AI) με γνώμονα την ακαδημαϊκή δεοντολογία:

- Συγγραφή Κειμένου: Χρησιμοποιήθηκε το γλωσσικό μοντέλο Gemini αποκλειστικά για τη γλωσσική εξομάλυνση, τη βελτίωση της ροής και τη διόρθωση συντακτικών σφαλμάτων.
- Ανάπτυξη Κώδικα: Αξιοποιήθηκε ο AI Agent του IDE Antigravity της Google υποβοηθητικά για την παραγωγή ορισμένων τμημάτων του κώδικα και τον εντοπισμό σφαλμάτων (debugging).

Ο ερευνητικός σχεδιασμός, η μεθοδολογία, η εξαγωγή συμπερασμάτων, καθώς και ο τελικός έλεγχος της ορθότητας τόσο του κειμένου όσο και του λογισμικού, αποτελούν αποκλειστικά πρωτότυπο έργο και ευθύνη του συγγραφέα.

Περιεχόμενα

Κατάλογος Εικόνων

Κατάλογος Πινάκων

1.Εισαγωγή στην Επιστήμη της Παπυρολογίας και το Αρχείο της Οξυρρύγχου

1.1 Ορισμός, Εύρος και Επιστημολογική Ταυτότητα του Κλάδου

Η επιστήμη της παπυρολογίας αποτελεί έναν από τους θεμελιώδεις πυλώνες των κλασικών σπουδών, λειτουργώντας ως το σημείο τομής μεταξύ της φιλολογίας, της ιστορίας, της αρχαιολογίας και της παλαιογραφίας. Παραδοσιακά, ορίζεται ως η συστηματική μελέτη, ανάγνωση, ερμηνεία και έκδοση αρχαίων κειμένων που έχουν διασωθεί, με κυρίαρχο υλικό τον πάπυρο, αλλά και συμπεριλαμβανομένων των οστράκων, των ξύλινων πινακίδων, καθώς και της περγαμηνής [1]. Το χρονικό εύρος της επιστήμης εκτείνεται συμβατικά από την εμφάνιση των πρώτων ελληνικών παπύρων στην Αίγυπτο, περί τον 4ο αιώνα π.Χ., έως και τον 8ο αιώνα μ.Χ., καλύπτοντας την Ελληνιστική, Ρωμαϊκή και Βυζαντινή περίοδο, αν και η αραβική παπυρολογία επεκτείνει αυτό το όριο μέχρι τον 10ο αιώνα μ.Χ. [2]

Η ιδιαιτερότητα της παπυρολογίας έγκειται στη φύση του υλικού της. Σε αντίθεση με την επιγραφική, η οποία εξετάζει κείμενα χαραγμένα σε σκληρά υλικά (λίθος, μέταλλο) που προορίζονταν συνήθως για δημόσια έκθεση ("monumental writing"), η παπυρολογία εστιάζει στο εφήμερο, το καθημερινό και το ιδιωτικό. Τα κείμενα αυτά διασώζουν τη "φωνή" των απλών ανθρώπων, προσφέροντας μια ματιά στις κοινωνικές δομές, την οικονομική ζωή, τη διοικητική μηχανή και τις θρησκευτικές πρακτικές, χωρίς το φίλτρο της επίσημης ιστοριογραφίας. Ενώ η χειρόγραφη παράδοση (manuscript tradition) μας παρέδωσε τα λογοτεχνικά έργα όπως αντιγράφηκαν και επιλέχθηκαν από τους βυζαντινούς λόγιους, οι πάπυροι αποκαλύπτουν τα έργα αυτά στη μορφή που κυκλοφορούσαν στην αρχαιότητα, συχνά φέρνοντας στο φως παραλλαγές ή και ολόκληρα έργα που η μεσαιωνική παράδοση άφησε να χαθούν. [3]

Στη σύγχρονη εποχή, ο κλάδος υφίσταται μια ριζική μεταμόρφωση. Η παραδοσιακή μεθοδολογία, η οποία βασιζόταν στην εγκυκλοπαιδική γνώση του μεμονωμένου ερευνητή για την "αποκρυπτογράφηση" των δυσανάγνωστων και αποσπασματικών κειμένων, ενισχύεται και ενίοτε αντικαθίσταται από υπολογιστικές μεθόδους. Η μετάβαση από την "ανάγνωση" στην "επεξεργασία δεδομένων" και από τη φιλολογική "εικασία" στην "πιθανοτική πρόβλεψη" μέσω νευρωνικών δικτύων, σηματοδοτεί την είσοδο της παπυρολογίας στην εποχή της Τεχνητής Νοημοσύνης (AI). [4] Το κείμενο δεν αντιμετωπίζεται πλέον μόνο ως φορέας νοήματος, αλλά και ως σύνολο δεδομένων που μπορεί να αναλυθεί ποσοτικά, να συσχετιστεί και να αποκατασταθεί με αλγοριθμική ακρίβεια.

1.2 Υλικά Γραφής και Τεχνολογία Παραγωγής

Η κατανόηση της υλικής υπόστασης παπύρων είναι προαπαιτούμενο για την ορθή παπυρολογική μελέτη, καθώς η φυσική δομή του υλικού επηρεάζει τη γραφή, τη φθορά και τις δυνατότητες ψηφιακής αποκατάστασης. [5]

1.2.1 Ο Πάπυρος: Κατασκευή και Μορφολογία

Ο πάπυρος ήταν το βασικό υλικό και έβγαινε από το φυτό Cyperus papyrus, ένα υδρόβιο καλάμι που φύτρωνε άφθονα στα έλη του Νείλου. Ο Πλίνιος ο Πρεσβύτερος στο έργο του «Φυσική Ιστορία» μας δίνει μια λεπτομερή περιγραφή της διαδικασίας. Έκοβαν το μίσχο του φυτού σε λεπτές λωρίδες, που λέγονταν philyrae. Απ' αυτές τις λωρίδες, έφτιαχναν δυο στρώματα: το ένα με τις λωρίδες να πηγαίνουν κάθετα, το άλλο οριζόντια, και τα ακουμπούσαν πάνω σε μια σκληρή επιφάνεια. Μετά, πίεζαν γερά ή χτυπούσαν τα δυο στρώματα μεταξύ τους, κι έτσι απελευθερώνονταν οι φυσικοί χυμοί του φυτού, που ένωναν τις λωρίδες μεταξύ τους. Το αποτέλεσμα ήταν ένα ενιαίο φύλλο.[6]

Η μορφή του φύλλου αποφάσιζε και πώς το χρησιμοποιούσαν. Η πλευρά με τις ίνες οριζόντιες λεγόταν recto. Αυτή ήταν η κύρια επιφάνεια για γράψιμο, αφού οι ίνες βοηθούσαν το καλάμι να γλιστράει πιο εύκολα. Η άλλη πλευρά, με τις ίνες κάθετες (verso), έμενε συνήθως άδεια ή τη χρησιμοποιούσαν αργότερα για πρόχειρες σημειώσεις ή δευτερεύον κείμενο. Όταν το υλικό έλειπε, έγραφαν και στις δύο όψεις. Αυτοί οι πάπυροι ονομάζονται οπισθογραφημένοι. Ένωναν αυτά τα φύλλα μεταξύ τους και έφτιαχναν μεγάλους κυλίνδρους, που πολλές φορές έφταναν αρκετά μέτρα σε μήκος. Αυτοί οι κύλινδροι ήταν η κλασική μορφή του αρχαίου "βιβλίου", δηλαδή το volumen. [7]

1.2.2 Περγαμηνή και Όστρακα

Πέραν του παπύρου, η περγαμηνή (από δέρμα ζώου, συνήθως προβάτου ή κατσίκας) άρχισε να κερδίζει έδαφος κατά την ύστερη αρχαιότητα. Η περγαμηνή ήταν πιο ανθεκτική και επέτρεψε τη μετάβαση από τον κύλινδρο στον κώδικα (codex), τη μορφή δηλαδή του σημερινού βιβλίου με σελίδες, η οποία υιοθετήθηκε ευρέως από τους Χριστιανούς για τα ιερά κείμενα. [8] Μια ειδική κατηγορία αποτελούν τα παλίμψηστα, περγαμηνές όπου το αρχικό κείμενο ξύστηκε για να γραφτεί ένα νέο, τα οποία σήμερα ανακτώνται μέσω φασματοσκοπικών μεθόδων. [9]

Τα όστρακα, δηλαδή τα σπασμένα κομμάτια από αγγεία ή πέτρες, ήταν αυτό που σήμερα θα λέγαμε "χαρτί για όσους δεν είχαν λεφτά". Ήταν παντού, άφθονα και δωρεάν,

οπότε οι άνθρωποι τα χρησιμοποιούσαν για καθημερινά σημειώματα—αποδείξεις φόρων, σχολικές εργασίες, ακόμα και σύντομες επιστολές. Αν και φαίνονται ταπεινά, για τους ερευνητές είναι θησαυρός. Μέσα σε αυτά διασώζονται ιστορίες και στοιχεία από ανθρώπους και κοινωνικές ομάδες που δύσκολα εμφανίζονται στα επίσημα κείμενα ή στα μεγάλα έργα της γραμματείας. [10]

1.3 Ιστορική Αναδρομή της Επιστήμης: Από τα «Περίεργα» στη Συστηματική Έρευνα

Η ιστορία της παπυρολογίας ως επιστημονικού κλάδου είναι άρρηκτα συνδεδεμένη με την ιστορία των ανακαλύψεων στην Αίγυπτο. Αν και μεμονωμένοι πάπυροι ήταν γνωστοί στην Ευρώπη νωρίτερα, η επιστημονική "προϊστορία" ξεκινά το 1778 με την ανακάλυψη της Charta Borgiana, ενός παπύρου που αγόρασε ένας Ευρωπαίος έμπορος βλέποντας ντόπιους να καίνε παπύρους για να απολαύσουν το άρωμά τους. Η δημοσίευσή του από τον Niels Iversen Schow το 1788 (Charta Papyracea Graece Scripta) σηματοδοτεί την πρώτη παπυρολογική έκδοση. [1]

Μια δεύτερη καθοριστική στιγμή ήταν η ανακάλυψη της Βίλας των Παπύρων στο Herculaneum το 1752, η οποία έφερε στο φως μια ολόκληρη βιβλιοθήκη απανθρακωμένων κυλίνδρων, κυρίως επικούρειας φιλοσοφίας. [11] Ωστόσο, η τεχνική δυσκολία στο ξετύλιγμα αυτών των κυλίνδρων καθυστέρησε την πλήρη αξιοποίησή τους μέχρι την εποχή των σύγχρονων ψηφιακών μεθόδων.

Ο 19ος αιώνας χαρακτηρίστηκε ως "ο αιώνας της παπυρολογίας". Η μεγάλη τομή έγινε τη δεκαετία του 1870 με τις μαζικές ευρέσεις στο Φαγιούμ, οι οποίες κατέκλυσαν τις ευρωπαϊκές αγορές αρχαιοτήτων. Η συλλογή του Αρχιδούκα Ράινερ στη Βιέννη, υπό την επιμέλεια του Joseph von Karabacek, και η ίδρυση του τμήματος παπύρων στα Μουσεία του Βερολίνου έθεσαν τις βάσεις για τη συστηματική μελέτη [12]. Ο Ulrich Wilcken, σε συνεργασία με τον Ludwig Mitteis, δημοσίευσε το 1912 το μνημειώδες έργο Grundzüge und Chrestomathie der Papyruskunde, το οποίο παραμένει μέχρι σήμερα το θεμελιώδες εγχειρίδιο του κλάδου, καθιερώνοντας τη μεθοδολογία της ιστορικής και νομικής ανάλυσης των κειμένων. [13]

1.4 Η Ανασκαφή της Οξυρρύγχου: Μεθοδολογία και Σημασία

Η πιο σημαντική στιγμή στην ιστορία του κλάδου, και το κεντρικό σημείο αναφοράς της παρούσας μελέτης, είναι οι ανασκαφές στην Οξυρρύγχο, σημερινή el-Bahnasa, περίπου 160 χιλιόμετρα νότια του Καΐρου. Μεταξύ 1896 και 1907, οι Βρετανοί παπυρολόγοι Bernard P. Grenfell και Arthur S. Hunt, χρηματοδοτούμενοι από το Egypt Exploration Fund, διενήργησαν ανασκαφές που άλλαξαν για πάντα την εικόνα μας για την αρχαιότητα. [14]

1.4.1 Το Πλαίσιο της Ανακάλυψης

Η Οξυρρύγχος δεν επελέγη τυχαία. Ως πρωτεύουσα νομού και ακμάζον κέντρο κατά την ελληνιστική και ρωμαϊκή περίοδο, με έντονη παρουσία του Χριστιανισμού αργότερα, προσέφερε μεγάλες πιθανότητες εύρεσης κειμένων. Ωστόσο, η καινοτομία των Grenfell και Hunt ήταν ότι δεν έψαξαν σε τάφους ή ναούς, αλλά στους αρχαίους σκουπιδότοπους της πόλης. Το ξηρό κλίμα της περιοχής και η έλλειψη υγρασίας επέτρεψαν τη διατήρηση χιλιάδων τόνων παπύρων που είχαν πεταχτεί ως άχρηστα "χαρτιά". [14]

1.4.2 Μεθοδολογία Ανασκαφής και Προβλήματα

Η μεθοδολογία που εφάρμοσαν ήταν, με τα σύγχρονα αρχαιολογικά πρότυπα, προβληματική. Εστιάζοντας στην ταχύτητα και τον όγκο, χρησιμοποίησαν εκατοντάδες ντόπιους εργάτες για να ανασκάψουν τους σωρούς, χωρίς να τηρούν λεπτομερή στρωματογραφικά δεδομένα. Τα ευρήματα συσκευάζονταν βιαστικά σε παλιά κουτιά μπισκότων και κιβώτια για να αποσταλούν στην Οξφόρδη. [14] Αυτή η διαδικασία οδήγησε στη δημιουργία μιας συλλογής που ξεπερνά τα 500.000 θραύσματα, αλλά ταυτόχρονα απέκοψε τα κείμενα από το αρχαιολογικό τους περιβάλλον. Το γεγονός ότι οι πάπυροι βρέθηκαν σε σκουπίδια και όχι π.χ. σε μια βιβλιοθήκη, σε συνδυασμό με την ταχεία ανασκαφή, σημαίνει ότι η χρονολόγηση και η ταύτιση των κειμένων πρέπει να βασίζεται σχεδόν αποκλειστικά σε εσωτερικά τεκμήρια και όχι στη στρωματογραφία. [15]

1.5 Σύνθεση της Συλλογής και το Πρόβλημα του "Backlog"

Η συλλογή της Οξυρρύγχου είναι το απόλυτο “Big Data” της αρχαιότητας. Μέσα της βρίσκεις τα πάντα. Πρώτα απ’ όλα, το 90% αφορά μη λογοτεχνικά κείμενα. Μιλάμε για ακατέργαστα κομμάτια καθημερινής ζωής: φορολογικές δηλώσεις, απογραφές, συμβόλαια γάμου ή μίσθωσης, ιδιωτικές επιστολές, δικαστικά έγγραφα, εντάλματα, ακόμα και προσκλήσεις για δείπνο. Μοιάζει με ένα τεράστιο αρχείο δημόσιας διοίκησης και προσωπικής αλληλογραφίας. Χάρη σ’ αυτό, οι ιστορικοί μπορούν να μελετήσουν δημογραφικά στοιχεία του αρχαίου κόσμου με ακρίβεια.

Μόνο το 10% είναι λογοτεχνικά κείμενα, αλλά η αξία τους είναι ανεκτίμητη. Υπάρχουν αποσπάσματα από χαμένα έργα του Μενάνδρου, ποιήματα της Σαπφούς, του Αλκαίου, του Πινδάρου, ίχνη από τραγωδίες του Σοφοκλή και του Ευριπίδη, ακόμη και σπάνια επιστημονικά έργα, όπως τα “Στοιχεία” του Ευκλείδη με τα διαγράμματά τους.

Υπάρχουν και τα χριστιανικά κείμενα. Η Οξυρρύγχος έφερε στο φως μερικά από τα αρχαιότερα χριστιανικά χειρόγραφα κομμάτια από κανονικά και απόκρυφα ευαγγέλια, όπως το Ευαγγέλιο του Θωμά. Αυτά τα ευρήματα άνοιξαν νέα παράθυρα για να καταλάβουμε καλύτερα τον πρώιμο Χριστιανισμό.

Ωστόσο, ο τεράστιος όγκος του υλικού δημιούργησε ένα ανυπέρβλητο πρόβλημα. Από το 1898, όταν εκδόθηκε ο πρώτος τόμος των The Oxyrhynchus Papyri (P.Oxy.), μέχρι σήμερα, έχουν δημοσιευθεί περίπου 88 τόμοι που περιέχουν περίπου 5.500 με 6.000 κείμενα. Αυτό αντιπροσωπεύει μόλις το 1-2% του συνολικού υλικού που φυλάσσεται στη Βιβλιοθήκη Sackler της Οξφόρδης. Με τους σημερινούς ρυθμούς της παραδοσιακής φιλολογικής μεθόδου, η πλήρης δημοσίευση της συλλογής θα απαιτούσε εκατοντάδες χρόνια. Αυτό το "backlog" δεν είναι απλώς μια καθυστέρηση, αλλά ένα μείζον επιστημονικό αδιέξοδο που επιτάσσει την εξεύρεση νέων, τεχνολογικά υποβοηθούμενων λύσεων. [16]

1.6 Παραδοσιακή Μεθοδολογία: Παλαιογραφία, Έκδοση και Συντήρηση

Για να κατανοήσουμε πώς η Τεχνητή Νοημοσύνη μπορεί να βοηθήσει, πρέπει πρώτα να αναλύσουμε την παραδοσιακή μεθοδολογία την οποία καλείται να μιμηθεί ή να επαυξήσει.

1.6.1 Παλαιογραφική Ανάλυση

Η χρονολόγηση και ταύτιση των κειμένων βασίζεται στην παλαιογραφία, τη μελέτη δηλαδή της εξέλιξης της γραφής. Οι παπυρολόγοι διακρίνουν μεταξύ διαφορετικών στυλ γραφής που εξελίσσονται στο χρόνο. [1]

- **Πτολεμαϊκή Περίοδος**: Χαρακτηρίζεται από αυστηρές, συχνά "επιγραφικές" μορφές στα πρώιμα στάδια, οι οποίες εξελίσσονται σε πιο γρήγορες και συνδεδεμένες γραφές.
- **Ρωμαϊκή Περίοδος**: Εμφανίζεται η "στρογγυλόσχημη" γραφή και η "καγκελαριακή" γραφή, με χαρακτηριστικά στενά και ψηλά γράμματα. Καθιερώνονται σταδιακά τα σημεία στίξης και οι τόνοι στα φιλολογικά κείμενα.
- **Βυζαντινή Περίοδος**: Κυριαρχούν οι περίτεχνες γραφές στα έγγραφα, ενώ στα βιβλία επικρατεί η μεγαλογράμματη (uncial) γραφή.

Η διαδικασία της ανάγνωσης είναι μια πολύπλοκη γνωστική λειτουργία που απαιτεί αναγνώριση προτύπων σε συνθήκες έντονου θορύβου (φθορές, τρύπες, ξεθωριασμένο μελάνι). Ο παπυρολόγος πρέπει να διακρίνει γράμματα που συχνά μοιάζουν μεταξύ τους ή έχουν αλλοιωθεί, βασιζόμενος στα συμφραζόμενα και στη γνώση της γλώσσας. [17]

1.6.2 Το Σύστημα Leiden

Η επιστημονική δημοσίευση των παπύρων ακολουθεί αυστηρούς κανόνες που τυποποιήθηκαν το 1931 στο Συνέδριο του Leiden. Το Σύστημα Leiden (Leiden Conventions) αποτελεί τη διεθνή "γλώσσα" μεταγραφής, επιτρέποντας την ακριβή αποτύπωση της φυσικής κατάστασης του κειμένου [5]. Τα κυριότερα σύμβολα είναι:

- [...] (Αγκύλες): Δηλώνουν χάσμα στο κείμενο λόγω φυσικής φθοράς. Ο αριθμός των κουκκίδων ή ο κενός χώρος υποδηλώνει το μέγεθος του χάσματος.
- (...) (Παρένθεση): Ανάλυση μιας συντομογραφίας από τον εκδότη.
- <...> (Γωνιώδεις αγκύλες): Προσθήκη γραμμάτων που παρέλειψε ο αρχαίος γραφέας εκ παραδρομής.
- {...} (Άγκιστρα): Αθέτηση γραμμάτων που έγραψε ο γραφέας λανθασμένα (π.χ. διπλογραφία).
- α̣β̣: Γράμματα των οποίων η ανάγνωση είναι αβέβαιη ή αμφίβολη λόγω φθοράς.

Αυτό το σύστημα κωδικοποίησης είναι κρίσιμο για την ψηφιακή παπυρολογία, καθώς αποτελεί τη βάση για τη δομή των δεδομένων εκπαίδευσης των αλγορίθμων AI.

1.6.3 Συντήρηση

Η φυσική διατήρηση των παπύρων είναι εξίσου σημαντική. Παραδοσιακά, οι πάπυροι φυλάσσονταν ανάμεσα σε δύο πλάκες γυαλιού, σφραγισμένες με ταινία. Ωστόσο, αυτή η μέθοδος αποδείχθηκε μακροπρόθεσμα επιζήμια, καθώς η πίεση και η έλλειψη αερισμού προκαλούσαν τη συγκέντρωση υγρασίας και τη συσσώρευση αλάτων στην επιφάνεια, καταστρέφοντας το μελάνι [1]. Οι σύγχρονες μέθοδοι, όπως αυτές που εφαρμόζονται στο Princeton, χρησιμοποιούν ειδικά αδρανή υλικά (όπως το Stabiltex και το Melinex) και συστήματα ανάρτησης που επιτρέπουν στον πάπυρο να "αναπνέει" χωρίς να δέχεται πίεση, ενώ προστατεύεται από την υπεριώδη ακτινοβολία. [18] Η κατανόηση αυτών των φθορών είναι απαραίτητη για την ανάπτυξη αλγορίθμων που μπορούν να διαχωρίσουν ψηφιακά το μελάνι από τους λεκέδες ή τη φθορά του υλικού.

1.7 Το πρόβλημα και η πρόταση της έρευνας

Η τεράστια ποσότητα υλικού από την Οξυρρύγχο έφερε αμέσως ένα σοβαρό πρόβλημα: πώς να τα δημοσιεύσουν όλα; Από το 1898, όταν κυκλοφόρησε ο πρώτος τόμος των Oxyrhynchus Papyri, οι ερευνητές δεν σταμάτησαν να δουλεύουν για να συντηρήσουν, να μεταγράψουν και να δημοσιεύσουν τα κομμάτια. Κι όμως, μέχρι το 2024, έχουν

δημοσιευτεί μόνο 5.500 με 6.000 αντικείμενα στη βασική σειρά (P.Oxy.), που αντιστοιχεί περίπου στο 1-2% του συνολικού υλικού που φυλάσσεται στη Βιβλιοθήκη Sackler (τώρα Βιβλιοθήκη Τέχνης, Αρχαιολογίας και Αρχαίου Κόσμου) στην Οξφόρδη και σε άλλα ιδρύματα. [16].

Το κύριο ερευνητικό πρόβλημα που πραγματεύεται αυτή η εργασία είναι ότι αυτό το συσσωρευμένο υλικό αποτελεί ένα θεμελιώδες σημείο συμφόρησης στην παραγωγή ιστορικής γνώσης. Οι παραδοσιακές εκδοτικές μέθοδοι είναι χειρωνακτικές και απαγορευτικά αργές, καθώς εξαρτώνται από τη σπάνια και εξαιρετικά εξειδικευμένη γνώση της παλαιογραφίας, της φιλολογίας και της κωδικολογίας. Ένα μόνο, έντονα φθαρμένο θραύσμα μπορεί να απαιτήσει μήνες έρευνας για να αποκρυπτογραφηθεί, να ταυτοποιηθεί και να συμπληρωθούν τα κενά του. [7]

Για να μειωθεί η μεγάλη διαφορά ανάμεσα στην ταχύτητα με την οποία βρίσκονται νέα ευρήματα στις ανασκαφές και στην καθυστέρηση της δημοσίευσής τους, αυτή η διπλωματική εργασία προτείνει τη δημιουργία ενός κλιμακούμενου, αυτοματοποιημένου pipeline, από την αρχή ως το τέλος. Εδώ, η Τεχνητή Νοημοσύνη δεν έρχεται να πάρει τη θέση του παπυρολόγου. Στόχος είναι να λειτουργεί σαν ένας εξελιγμένος ψηφιακός βοηθός που σηκώνει το μεγαλύτερο βάρος της ανάλυσης και της επεξεργασίας, ώστε η όλη διαδικασία να γίνεται πολύ πιο γρήγορα.

Πιο συγκεκριμένα, η προτεινόμενη λύση αξιοποιεί την Τεχνητή Νοημοσύνη σε δύο διακριτά στάδια, τα οποία αντανακλούν τα βήματα ενός ανθρώπου μελετητή:

- **Ανάγνωση (Υπολογιστική Όραση)**: Μοντέλα ανίχνευσης αντικειμένων και ταξινόμησης (YOLO και DeiT) σαρώνουν την ψηφιακή εικόνα του παπύρου, αναγνωρίζουν τους χαρακτήρες παρά τις φθορές στο μελάνι και εντοπίζουν αυτόματα τα φυσικά κενά (Gap Detection), εξάγοντας το κείμενο με την καθιερωμένη παπυρολογική μορφοποίηση Leiden.
- **Αποκατάσταση και Κατανόηση (Επεξεργασία Φυσικής Γλώσσας)**: Σύγχρονα Μεγάλα Γλωσσικά Μοντέλα (LLMs) αναλαμβάνουν το ακατέργαστο αποτέλεσμα της όρασης. Βασιζόμενα στη βαθιά τους γνώση για την αρχαία ελληνική, διορθώνουν τα σφάλματα ανάγνωσης, συμπληρώνουν τα κενά του παπύρου και παράγουν την τελική αγγλική μετάφραση.

Αν και η χρήση Νευρωνικών Δικτύων στην ψηφιακή επιγραφική έχει γνωρίσει σημαντικά ορόσημα (όπως τα συστήματα *Ithaca* και *Aeneas* της DeepMind), τα υπάρχοντα μοντέλα παρουσιάζουν περιορισμούς όταν εφαρμόζονται στην ακατέργαστη, οπτική παπυρολογία. Η παρούσα εργασία διαφοροποιείται και υπερτερεί της υφιστάμενης βιβλιογραφίας σε τρεις κομβικούς άξονες:

- **Τεμαχισμός της εικόνας (SAHI)**: Τα παραδοσιακά συστήματα Οπτικής Αναγνώρισης (OCR) καταστρέφουν την οπτική πληροφορία όταν συμπιέζουν μακροσκελείς παπύρους για να τροφοδοτήσουν τα δίκτυά τους, δημιουργώντας «ψευδώς αρνητικά» (false negatives). Η παρούσα αρχιτεκτονική ενσωματώνει τον αλγόριθμο Slicing Aided Hyper Inference (SAHI), επιτρέποντας στα

μοντέλα να σαρώνουν εικόνες τεράστιας ανάλυσης μέσω μικρών, επικαλυπτόμενων πλαισίων, χωρίς καμία απώλεια χωρικής λεπτομέρειας.

- **Δυναμική Διαχείριση των Κενών:** Μοντέλα όπως το Ithaca βασίζονται σε masked language models (MLM), απαιτώντας από τον χρήστη να γνωρίζει εξαρχής τον ακριβή αριθμό των χαμένων χαρακτήρων. Αντίθετα, η χρήση LLMs στην παρούσα εργασία επιτρέπει τη δυναμική συμπλήρωση χασμάτων ακαθόριστου μήκους, προσφέροντας μεγαλύτερη ευελιξία στην πραγματική, φθαρμένη εικόνα.
- **Υβριδικό Σύστημα RAG για Ιστορική Ακρίβεια**: Το μεγαλύτερο μειονέκτημα των σύγχρονων LLMs είναι η τάση για παραγωγή ανύπαρκτων ιστορικών γεγονότων («παραισθήσεις»). Για την επίλυση αυτού του προβλήματος, η εργασία δεν βασίζεται στην εσωτερική μνήμη του μοντέλου, αλλά υλοποιεί ένα καινοτόμο Υβριδικό Σύστημα Ανάκτησης Επαυξημένης Παραγωγής (Hybrid RAG). Συνδυάζοντας σημασιολογικά διανύσματα (Vector Embeddings) και αντιστοίχιση χαρακτήρων (N-Gram matching), το σύστημα ανακτά πραγματικά, επαληθευμένα ιστορικά παράλληλα από μια εξωτερική βάση δεδομένων και τα τροφοδοτεί στο LLM. Έτσι, κάθε προτεινόμενη αποκατάσταση διαθέτει απόλυτη ακαδημαϊκή ιχνηλασιμότητα (traceability), καθιστώντας το εργαλείο έμπιστο για επιστημονική χρήση.

2. Υπόβαθρο και Σχετική Βιβλιογραφία

Η μετάβαση από την παραδοσιακή στην ψηφιακή παπυρολογία δεν ήταν ένα ξαφνικό γεγονός, αλλά μια σταδιακή εξέλιξη που ξεκίνησε με την ανάγκη για ψηφιοποίηση και οργάνωση του υλικού και κατέληξε στη σημερινή εποχή της υπολογιστικής επεξεργασίας και της Τεχνητής Νοημοσύνης. Το κεφάλαιο αυτό χαρτογραφεί αυτή την πορεία, εξετάζοντας τις υποδομές, τις θεωρητικές προσεγγίσεις και τα υπολογιστικά μοντέλα που εφαρμόζονται σήμερα.

2.1 Θεωρητικό Πλαίσιο: Από την Ψηφιοποίηση στην "Εικονική Παπυρολογία"

Η εφαρμογή ψηφιακών μέσων στην παπυρολογία ξεκίνησε νωρίς, καθιστώντας τον κλάδο πρωτοπόρο στις Ψηφιακές Ανθρωπιστικές Επιστήμες. Ωστόσο, όπως επισημαίνει ο θεωρητικός της ψηφιακής παπυρολογίας Nicola Reggiani [5], πρέπει να διακρίνουμε μεταξύ της απλής "Ψηφιακής Παπυρολογίας" και της "Εικονικής Παπυρολογίας" (Virtual Papyrology).

Στην πρώτη φάση, η έμφαση δόθηκε στη δημιουργία ψηφιακών αντιγράφων. Έργα όπως το Oxyrhynchus Online και το Imaging Papyri Project στόχευαν στην παραγωγή υψηλής ανάλυσης φωτογραφιών για να προστατεύσουν τα πρωτότυπα από τη φθορά της φυσικής επαφής και να καταστήσουν το υλικό προσβάσιμο στην παγκόσμια κοινότητα. [6] Στον τομέα της "Εικονικής Παπυρολογίας", η ψηφιακή εικόνα δεν αντιμετωπίζεται πλέον ως απλό αντίγραφο, αλλά ως "ψηφιακό υποκατάστατο". Το ψηφιακό αυτό αντικείμενο διαθέτει ιδιότητες που το φυσικό αντικείμενο στερείται: μπορεί να ενισχυθεί χρωματικά, να αναλυθεί φασματικά για να αποκαλυφθεί κρυμμένο κείμενο, ή να "ξεδιπλωθεί" εικονικά (virtual unwrapping) σε τρισδιάστατο χώρο, όπως έγινε με τους κυλίνδρους του En-Gedi και του Herculaneum. Αυτή η "αποϋλοποίηση" του αντικειμένου είναι που επιτρέπει την εισαγωγή προηγμένων αλγορίθμων, οι οποίοι επεξεργάζονται το κείμενο ως δεδομένο και όχι απλώς ως εικόνα. [19]

Η σύγχρονη υπολογιστική έρευνα βασίζεται σε ένα οικοσύστημα διασυνδεδεμένων βάσεων δεδομένων, οι οποίες οργανώνουν τον χαώδη όγκο των παπυρολογικών δεδομένων.

Η πλατφόρμα Papyri.info, που πολλοί ξέρουν και ως Integrated Data for Papyrology (IDP), είναι το βασικό μέρος όπου συγκεντρώνονται παπυρικά κείμενα. Το εργαλείο Papyrological Editor επιτρέπει στους χρήστες να συνεργάζονται και να επεξεργάζονται τα κείμενα μαζί, αλλάζοντας ουσιαστικά τη λογική της κριτικής έκδοσης. Κάθε νέα διόρθωση, ανάγνωση ή σχόλιο μπορεί να ενσωματώνεται άμεσα, ενώ όλα γίνονται με πλήρη καταγραφή των αλλαγών. Αυτή η ζωντάνια είναι ακριβώς ό,τι χρειάζεται για να μπουν στο

παιχνίδι αλγόριθμοι AI που προτείνουν διορθώσεις με βάση στατιστικές πιθανότητες, αφήνοντας την τελική αξιολόγηση στους ίδιους τους ερευνητές.[5]

Το Trismegistos λύνει το πρόβλημα της διασύνδεσης ετερογενών δεδομένων. Κάθε αρχαίο κείμενο, όποια κι αν είναι η συλλογή ή η γλώσσα του, παίρνει έναν μοναδικό αριθμό TM. Αυτή η ταυτότητα ακολουθεί το κείμενο παντού και το συνδέει με όλες τις σχετικές πληροφορίες σε διαφορετικές βάσεις δεδομένων, όπως φωτογραφίες, μεταγραφές, βιβλιογραφία ή αρχαιολογικά δεδομένα. Το Trismegistos δίνει ακόμα βασικά μεταδεδομένα: χρονολόγηση, τόπο εύρεσης, είδος κειμένου. Αυτά είναι απαραίτητα για να εκπαιδεύονται και να ελέγχονται τα μοντέλα μηχανικής μάθησης. Χωρίς τη δομή που προσφέρει το Trismegistos, η εκπαίδευση μοντέλων όπως το Ithaca δεν γίνεται, γιατί δεν υπάρχει ένας σίγουρος τρόπος να συνδέσεις εικόνα, κείμενο και μεταδεδομένα.[20]

2.2 Η Εποχή της Πολυτροπικής Αρχαιογνωσίας

Η πρώτη μεγάλη πρόκληση για την T.N. στην παπυρολογία είναι η οπτική αναγνώριση, ο εντοπισμός και η ταυτοποίηση χαρακτήρων σε φθαρμένες επιφάνειες.

Η περίοδος 2024–2026 σηματοδοτεί μια θεμελιώδη τομή στην ψηφιακή παπυρολογία και την υπολογιστική αρχαιολογία. Η έρευνα έχει μετακινηθεί από την αποσπασματική αυτοματοποίηση μεμονωμένων εργασιών, όπως η απλή οπτική αναγνώριση χαρακτήρων (OCR), σε ολιστικά συστήματα. Οι σύγχρονες προσεγγίσεις συνθέτουν πολυτροπικά δεδομένα (εικόνα, κείμενο, υλικότητα) και αξιοποιούν αρχιτεκτονικές Μεγάλων Γλωσσικών Μοντέλων (LLMs) και Ανάκτησης Επαυξημένης Παραγωγής (RAG) [21] για να επιλύσουν σύνθετα προβλήματα, όπως η ανάγνωση μη-ανελιγμένων κυλίνδρων και η ερμηνεία διαπολιτισμικών ιστορικών δεδομένων. [22]

2.2.1 Υπολογιστική Όραση

Η πιο εμβληματική εξέλιξη στον τομέα της υπολογιστικής όρασης προήλθε από την ανάγκη ανάγνωσης των απανθρακωμένων παπύρων του Herculaneum. Οι Nader, Farritor και Schilliger [23] , στο πλαίσιο του Vesuvius Challenge, ανέπτυξαν μια καινοτόμο μεθοδολογία ογκομετρικής ανάλυσης που ξεπέρασε τους περιορισμούς της παραδοσιακής αξονικής τομογραφίας. Η προσέγγισή τους βασίστηκε στην εκπαίδευση μοντέλων TimeSformer και 3D ResNets για τον εντοπισμό της «μικρο-υφής» που αφήνει το μελάνι άνθρακα πάνω στις ίνες του παπύρου. Αυτή η τεχνική επέτρεψε την πρώτη επιβεβαιωμένη ανάγνωση εκτενών ελληνικών κειμένων από το εσωτερικό ενός κλειστού κυλίνδρου.

Στον τομέα της ανάλυσης δισδιάστατων χειρογράφων με πολύπλοκη δομή, οι Vadlamudi et al. [24] παρουσίασαν το μοντέλο SeamFormer. Αντιμετωπίζοντας το πρόβλημα της αλληλοεπικάλυψης γραμμών σε πυκνά χειρόγραφα (όπως τα φύλλα φοίνικα

ή οι πυκνογραμμένοι πάπυροι), το SeamFormer χρησιμοποιεί έναν μηχανισμό «παραγωγής ραφής» καθοδηγούμενο από scribbles. Το μοντέλο προβλέπει τον μέσο άξονα της γραμμής κειμένου και, μέσω ενός χάρτη χαρακτηριστικών, χαράσσει δυναμικά πολύγωνα που διαχωρίζουν τις γραμμές χωρίς να καταστρέφουν τους χαρακτήρες που εκτείνονται κάθετα, επιτυγχάνοντας ακρίβεια αιχμής σε ιστορικά έγγραφα.

Για την ανίχνευση και αναγνώριση χαρακτήρων σε ελληνικούς παπύρους, η εργασία των Turnbull και Mannix [25] έθεσε νέα πρότυπα συνδυάζοντας την ταχύτητα του YOLOv8 με την ακρίβεια των Vision Transformers. Η μεθοδολογία τους βασίστηκε σε μια στρατηγική συνόλου (ensemble), όπου οι ανιχνεύσεις του YOLOv8 βελτιώθηκαν από ένα δίκτυο ResNet-50 (εκπαιδευμένο με SimCLR) και ένα μοντέλο DeiT. Κρίσιμο στοιχείο της επιτυχίας τους ήταν η χρήση ημι-επιβλεπόμενης μάθησης μέσω «ψευδο-ετικετών» (pseudo-labeling) από αταξινόμητα θραύσματα της συλλογής της Οξυρρύγχου, διαδικασία που αύξησε τη μέση ακρίβεια (mAP) κατά 8-11%.

Στο πεδίο της ταυτοποίησης γραφέων, οι De Gregorio et al. [26] εισήγαγαν το NeuroPapyri, ένα δίκτυο βαθιάς μάθησης που χρησιμοποιεί ενσωματώσεις προσοχής (attention embeddings). Το μοντέλο εκπαιδεύεται με συνάρτηση απώλειας τριπλέτας (triplet loss) για να χαρτογραφεί θραύσματα του ίδιου γραφέα σε κοντινά σημεία ενός διανυσματικού χώρου. Η καινοτομία του έγκειται στην παραγωγή heat maps που υποδεικνύουν ποιες παλαιογραφικές λεπτομέρειες (όπως η καμπυλότητα ενός γράμματος) οδήγησαν στην ταυτοποίηση, προσφέροντας έτσι εξηγήσιμα αποτελέσματα για τους παπυρολόγους.

Παρά τις εντυπωσιακές αυτές προσεγγίσεις, η βιβλιογραφία δεν προσφέρει μια ολιστική, "out-of-the-box" λύση για υπερ-υψηλής ανάλυσης παπύρους χωρίς τεράστιες απώλειες πληροφορίας λόγω downscaling. Η παρούσα διπλωματική εργασία γεφυρώνει αυτό το κενό, επεκτείνοντας τη μεθοδολογία των Turnbull και Mannix [25]. Συγκεκριμένα, υιοθετεί την αρχιτεκτονική YOLO+DeiT, αλλά την εξελίσσει ριζικά ενσωματώνοντας τον αλγόριθμο **SAHI (Slicing Aided Hyper Inference)**. Αυτό επιτρέπει στο μοντέλο μας να εκτελεί sliding-window ανίχνευση στην αρχική, πλήρη ανάλυση του παπύρου, αποφεύγοντας το πρόβλημα των "ψευδώς αρνητικών" που μαστίζει τα υπάρχοντα συστήματα στα δυσανάγνωστα μικρά γράμματα. Επιπλέον, ενώ το NeuroPapyri [26] εστιάζει αποκλειστικά στον γραφέα, το δικό μας σύστημα εστιάζει στην εξαγωγή περιεχομένου, προσθέτοντας έναν πρωτότυπο αλγόριθμο Gap Detection, ο οποίος χαρτογραφεί τη χωρική πληροφορία και παράγει έτοιμη παπυρολογική μορφοποίηση Leiden, τροφοδοτώντας απευθείας τα γλωσσικά μοντέλα.

2.2.2 Η Δημιουργία Δεδομένων Εκπαίδευσης

Οι αλγόριθμοι βαθιάς μάθησης χρειάζονται τεράστιες ποσότητες δεδομένων με σήμανση για να μάθουν να κάνουν σωστά τη δουλειά τους. Τέτοια δεδομένα δεν υπήρχαν στην παπυρολογία, οπότε το έργο Ancient Lives ήρθε να καλύψει αυτό το κενό.[16] Το

Πανεπιστήμιο της Οξφόρδης μαζί με την πλατφόρμα Zooniverse άνοιξαν το παιχνίδι στο ευρύ κοινό, ζητώντας από ανθρώπους να βοηθήσουν στη μεταγραφή των παπύρων της Οξυρρύγχου. Ουσιαστικά, ο καθένας μπορούσε να μαρκάρει και να ταυτοποιήσει γράμματα πάνω στις ψηφιακές εικόνες των παπύρων. Σίγουρα έγιναν λάθη, δεν ήταν όλοι ειδικοί. Παρ' όλα αυτά, η "σοφία του πλήθους" και η σωστή στατιστική επεξεργασία έφεραν αποτελέσματα: δημιουργήθηκαν σύνολα δεδομένων με "συναίνεση" (consensus data). Έτσι προέκυψαν τα AL-ALL και AL-PUB, datasets που μαζεύουν εκατοντάδες χιλιάδες επισημειωμένα πλαίσια οριοθέτησης χαρακτήρων. Αυτά αποτέλεσαν τη βάση για να εκπαιδευτούν τα πρώτα Συνελικτικά Νευρωνικά Δίκτυα στην παπυρολογία. Πρακτικά, κάπως έτσι οι αλγόριθμοι έμαθαν τα σχήματα των αρχαίων ελληνικών γραμμάτων, μέσα από όλα τα διαφορετικά γραφικά στυλ που εμφανίζονται στους παπύρους.[27]

2.2.3 Ταυτοποίηση Γραφέα και Ανάκτηση

Πέρα από την ανάγνωση κειμένου, η Υπολογιστική Όραση χρησιμοποιείται για την ταυτοποίηση και την ένωση διαχωρισμένων θραυσμάτων. Το μοντέλο NeuroPapyri αντιπροσωπεύει μια εξειδικευμένη αρχιτεκτονική για αυτήν την εργασία, χρησιμοποιώντας Δίκτυα Ενσωμάτωσης Βαθιάς Προσοχής (Deep Attention Embedding Networks). [26]

Ενώ η Υπολογιστική Όραση ασχολείται με την εικόνα, η Επεξεργασία Φυσικής Γλώσσας (Natural Language Processing - NLP) ασχολείται με το κείμενο. Οι πιο προβεβλημένες εξελίξεις στην ψηφιακή παπυρολογία έχουν προέλθει από την εφαρμογή της βαθιάς μάθησης στις εργασίες της αποκατάστασης (συμπλήρωση κενών), της χρονολόγησης και της εντοπιότητας (γεωγραφική απόδοση).

Το 2019, ερευνητές της DeepMind και του Πανεπιστημίου της Οξφόρδης παρουσίασαν την Pythia, το πρώτο μοντέλο βαθιάς μάθησης σχεδιασμένο για την αποκατάσταση χαμένου κειμένου σε αρχαίες ελληνικές επιγραφές (οι οποίες μοιράζονται γλωσσικά χαρακτηριστικά με τους παπύρους). [28]

- **Αρχιτεκτονική**: Η Pythia χρησιμοποίησε μια αρχιτεκτονική βασισμένη σε Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης (LSTM) σε επίπεδο χαρακτήρα και λέξης. Αντιμετώπισε την αποκατάσταση κειμένου ως πρόβλημα πρόβλεψης ακολουθίας-σε-ακολουθία (sequence-to-sequence), παρόμοιο με την αυτόματη συμπλήρωση.
- **Αποτελέσματα**: Στο σύνολο δεδομένων PHI-ML (μηχανικά επεξεργάσιμες επιγραφές), η Pythia πέτυχε ποσοστό σφάλματος χαρακτήρων (CER) 30,1%, ξεπερνώντας σημαντικά τους ανθρώπινους εμπειρογνώμονες που είχαν ποσοστό σφάλματος 57,3% στις ίδιες απομονωμένες εργασίες. Σημαντικό είναι ότι στο 73,5% των περιπτώσεων, η σωστή αποκατάσταση βρισκόταν εντός των κορυφαίων 20 υποθέσεων της Pythia.
- **Αντίκτυπος**: Η Pythia απέδειξε ότι τα νευρωνικά δίκτυα μπορούσαν να συλλάβουν το μακροπρόθεσμο πλαίσιο και τη τυπολατρική φύση των αρχαίων κειμένων

καλύτερα από τα μοντέλα n-gram ή την καθαρή ανθρώπινη διαίσθηση σε απομόνωση.

Χτίζοντας πάνω στην Pythia, το μοντέλο Ithaca αντιπροσώπευσε ένα άλμα προς τα εμπρός, εισάγοντας ένα εργαλείο ιστορικής ανάλυσης «πλήρους φάσματος» (full stack) [29]

- **Αρχιτεκτονική**: Το Ithaca χρησιμοποιεί μια αρχιτεκτονική Transformer (συγκεκριμένα έναν μηχανισμό αραιής αυτο-προσοχής), η οποία είναι πιο αποτελεσματική από τα LSTMs στο χειρισμό εξαρτήσεων μεγάλου εύρους στο κείμενο. Εκπαιδεύτηκε σε ένα τεράστιο σύνολο επιγραφών (I.PHI).
- **Πολυεργασία (Multitasking)**: Σε αντίθεση με την Pythia, το Ithaca εκτελεί τρεις εργασίες ταυτόχρονα:
  1. Αποκατάσταση: Πρόβλεψη χαμένων χαρακτήρων.
  2. Γεωγραφική Απόδοση: Πρόβλεψη της περιοχής προέλευσης (71% ακρίβεια).
  3. Χρονολογική Απόδοση: Πρόβλεψη της ημερομηνίας του κειμένου (εντός ~30 ετών από την αλήθεια αναφοράς).
- **Εξηγησιμότητα**: Το Ithaca εισήγαγε «χάρτες σημαντικότητας» (saliency maps) — οπτικούς θερμικούς χάρτες — για να δείξει ποιες λέξεις στο κείμενο επηρέασαν τις προβλέψεις του, επιτρέποντας στους μελετητές να κρίνουν τη συλλογιστική του μοντέλου.

2.2.4 Βαθιά Μάθηση για Αποκατάσταση και Ιστορική Πλαισίωση

Η μετάβαση από την απλή συμπλήρωση κενών στην πλήρη ιστορική πλαισίωση επιτεύχθηκε με το μοντέλο Aeneas των Assael et al. [30]. Ως διάδοχος του Ithaca, το Aeneas εισάγει μια πολυτροπική αρχιτεκτονική που επεξεργάζεται ταυτόχρονα το κείμενο και την εικόνα της επιγραφής. Το μοντέλο μπορεί να αποκαταστήσει κείμενα με κενά άγνωστου μήκους, να χρονολογήσει επιγραφές με ακρίβεια ±13 ετών και να εντοπίσει τη γεωγραφική τους προέλευση. Σημαντικότερη είναι η δυνατότητά του να ανακτά «παράλληλα» από μια βάση δεδομένων 176.000 επιγραφών, λειτουργώντας ως συνεργατικό εργαλείο που αυξάνει την ακρίβεια των ειδικών από 25% σε 72% όταν εργάζονται συνεργατικά.

Παράλληλα, το έργο Logion από τους Graziosi και Haubold [31] εστιάζει στη λογοτεχνική παπυρολογία και συγκεκριμένα στα έργα του Μιχαήλ Ψελλού. Σε αντίθεση με τα γενικά γλωσσικά μοντέλα, το Logion έχει σχεδιαστεί για να μοντελοποιεί τη συμπεριφορά των αντιγραφέων, εντοπίζοντας σφάλματα που προκύπτουν κατά τη χειρόγραφη παράδοση. Το σύστημα δεν παράγει απλώς ένα «σωστό» κείμενο, αλλά μια αναφορά πιθανοτήτων, ενθαρρύνοντας την κριτική εμπλοκή του φιλολόγου με τις προτάσεις της μηχανής.

Μια διαφορετική προσέγγιση ακολουθεί ο Cullhed [32], ο οποίος εφάρμοσε τεχνικές Instruction Tuning στο μοντέλο Llama 3.1 για την αποκατάσταση αρχαίων ελληνικών κειμένων. Εκπαιδεύοντας το μοντέλο να αγνοεί τα κενά μεταξύ των λέξεων (προσομοιώνοντας τη scriptio continua), πέτυχε ποσοστό σφάλματος χαρακτήρων (CER) 14.9% σε παπύρους. Η μελέτη αυτή αποδεικνύει ότι τα προ-εκπαιδευμένα LLMs γενικού σκοπού μπορούν, με τον κατάλληλο συντονισμό, να ξεπεράσουν εξειδικευμένες αρχιτεκτονικές σε συγκεκριμένες εργασίες αποκατάστασης.

Ενώ τα μοντέλα Ithaca και Aeneas είναι πρωτοποριακά, στηρίζονται σε προσεγγίσεις Masked Language Modeling (MLM), γεγονός που τα καθιστά άκαμπτα όταν πρόκειται για μεγάλα, ακαθόριστα κενά στον πάπυρο. Ο ερευνητής πρέπει να γνωρίζει ακριβώς πόσα γράμματα λείπουν. Σε αντίθεση με αυτά τα συστήματα, η δική μας πρόταση βασίζεται σε γενικά Μεγάλα Γλωσσικά Μοντέλα, τα οποία, όπως απέδειξε και ο Cullhed [32], μπορούν να αναπληρώσουν φράσεις ελεύθερου μήκους. Ωστόσο, η δική μας εργασία υπερβαίνει την προσέγγιση του Cullhed. Αντί απλώς να εκπαιδεύσουμε ένα Llama να «μαντεύει» καλύτερα, περιορίζουμε τη φαντασία του ενσωματώνοντας μια αρχιτεκτονική RAG, η οποία προσομοιώνει τη λειτουργία του Aeneas αλλά χωρίς να απαιτεί το χτίσιμο ενός τεράστιου custom νευρωνικού δικτύου.

Εκτός από την ψηφιακή ανάλυση του κειμένου, η σύγχρονη παπυρολογία ενσωματώνει προηγμένες τεχνικές υλικής ανάλυσης για την ανάδειξη κειμένων που είναι δυσανάγνωστα ή αόρατα στο γυμνό μάτι. Μια σημαντική πτυχή αυτής της «Εικονικής Παπυρολογίας» είναι η μελέτη των χρωστικών, ιδιαίτερα του κιννάβαρι (vermilion), το οποίο χρησιμοποιούνταν για τίτλους και διακοσμήσεις σε πολυτελή χειρόγραφα.

- **Κιννάβαρι (Vermilion)**: Η παρουσία κιννάβαρι (θειούχος υδράργυρος) σε παπύρους και περγαμηνές αποτελεί δείκτη κύρους και συχνά βοηθά στη χρονολόγηση ή την ταυτοποίηση. Ωστόσο, το υλικό αυτό υπόκειται σε χημική αλλοίωση (μαύρισμα) με την πάροδο του χρόνου, καθιστώντας το κείμενο δυσανάγνωστο. [33]
- **Φασματοσκοπία**: Τεχνικές όπως η φασματοσκοπία και ο φθορισμός ακτίνων Χ (XRF) επιτρέπουν την ταυτοποίηση αυτών των χρωστικών χωρίς τη λήψη δείγματος, αποκαλύπτοντας «κρυμμένα» κείμενα κάτω από επιχρίσματα ή φθορές. Αυτή η υλική ανάλυση συνδέεται άμεσα με την ψηφιακή αποκατάσταση, καθώς οι φασματικές υπογραφές μπορούν να χρησιμοποιηθούν για τον ψηφιακό διαχωρισμό του μελανιού από το υπόβαθρο (segmentation), τροφοδοτώντας στη συνέχεια τα μοντέλα OCR και αποκατάστασης. [34]

Η χρονολόγηση εγγράφων είναι ένα από τα πιο επίμαχα ζητήματα της παπυρολογίας, βασιζόμενη παραδοσιακά στην «γνωσιολογία» , την υποκειμενική αίσθηση του ειδικού για το ύφος μιας γραφής. Η βαθιά μάθηση προσφέρει έναν τρόπο ποσοτικοποίησης αυτής της διαδικασίας.

Μια θεμελιώδης μελέτη των West, Swindall et al. [35] καθιέρωσε ένα pipeline βαθιάς μάθησης για τη χρονολόγηση εγγράφων παπύρων. Το pipeline απομονώνει μεμονωμένους χαρακτήρες χρησιμοποιώντας το σύνολο δεδομένων AL-PUB και εκπαιδεύει έναν ταξινομητή ResNet για να προβλέψει την ημερομηνία κάθε γράμματος. Αυτές οι μεμονωμένες προβλέψεις συγκεντρώνονται στη συνέχεια χρησιμοποιώντας μια Διαδικασία Gauss (Gaussian Process) για να παραχθεί ένα εύρος ημερομηνιών για ολόκληρο το θραύσμα. Το μοντέλο επιτυγχάνει ακρίβεια ~79% στην τοποθέτηση θραυσμάτων εντός ενός παραθύρου δύο αιώνων.32 Αν και αυτή η ανάλυση είναι χονδροειδής σε σύγκριση με τη δυνητική ακρίβεια ενός εμπειρογνώμονα, είναι αντικειμενική, αναπαραγώγιμη και κλιμακώσιμη σε χιλιάδες ανέκδοτα θραύσματα. Αυτό μετατοπίζει τη χρονολόγηση από μια υποκειμενική τέχνη σε μια στατιστική επιστήμη, εισάγοντας «διαστήματα εμπιστοσύνης» και «κατανομές πιθανοτήτων».

2.2.5 Generative AI, RAG και το Πρόβλημα της Παραίσθησης (Hallucination)

Η αξιοπιστία των Μεγάλων Γλωσσικών Μοντέλων (LLMs) σε εξειδικευμένους τομείς ενισχύεται πλέον μέσω της Ανάκτησης Επαυξημένης Παραγωγής (RAG). Ο Miyagawa [36] παρουσίασε το THOTH AI, ένα σύστημα μετάφρασης για την Αρχαία Αιγυπτιακή και την Κοπτική. Το σύστημα ενσωματώνει ένα διανυσματικό λεξικό και μορφολογική βάση δεδομένων, τα οποία ανακτώνται δυναμικά κατά τη διάρκεια της μετάφρασης. Αυτή η «γείωση» (grounding) στη φιλολογική γνώση μείωσε δραστικά τις παραισθήσεις του μοντέλου σε σπάνιους όρους και επέτρεψε την ακριβή μετάφραση κειμένων όπως το Διάταγμα Στέψης του Τούθμωση Α'.

Στον τομέα του ιστορικού συλλογισμού, οι ερευνητές παρουσίασαν το 2026 τον agent HistAgent και το σημείο αναφοράς HistBench. [37] Το HistBench περιλαμβάνει 414 ερωτήσεις που απαιτούν σύνθεση πληροφοριών από εικόνες, χειρόγραφα και κείμενα. Το HistAgent, βασισμένο στο GPT-4o και εξοπλισμένο με εργαλεία όπως OCR και ακαδημαϊκή αναζήτηση, πέτυχε ακρίβεια 27.54%, υπερβαίνοντας σημαντικά τα γενικά μοντέλα. Αυτό καταδεικνύει ότι η χρήση εξειδικευμένων agents που μπορούν να αναζητούν και να επαληθεύουν πληροφορίες είναι απαραίτητη για την επίλυση σύνθετων ιστορικών προβλημάτων.

Τέλος, οι Zhang et al. [38] εισήγαγαν το UniDIT (Unified Document Image Translation), ένα πλαίσιο για την απευθείας μετάφραση εικόνων εγγράφων. Σε αντίθεση με τα διαδοχικά συστήματα (OCR ακολουθούμενο από μετάφραση), το UniDIT χρησιμοποιεί έναν ενιαίο μηχανισμό που λαμβάνει υπόψη τόσο τη γεωμετρική διάταξη όσο και τη λογική σειρά του κειμένου. Η προσέγγιση αυτή είναι ιδιαίτερα κρίσιμη για παπύρους και ιστορικά έγγραφα όπου η διάταξη φέρει σημαντικό σημασιολογικό φορτίο, επιτρέποντας τη διατήρηση της δομής του πρωτοτύπου στο μεταφρασμένο αποτέλεσμα.

Η εφαρμογή του RAG έχει αποδειχθεί επιτυχής στο THOTH AI [36] για δομημένα λεξικά, όμως δεν έχει αξιολογηθεί η συμπεριφορά του στις ακατέργαστες εξόδους OCR

από αρχαιοελληνικούς παπύρους. Όπως θα δειχθεί στην παρούσα εργασία, τα κλασικά πυκνά διανύσματα (dense embeddings) καταρρέουν και αποτυγχάνουν να βρουν την Ομοιότητα Συνημιτόνου όταν τροφοδοτούνται με ανορθόγραφα αρχαία ελληνικά χωρίς κενά. Η κύρια συνεισφορά της εργασίας μας στο πεδίο αυτό είναι η ανάπτυξη μιας Υβριδικής Στρατηγικής RAG. Ενώ η βιομηχανία βασίζεται αποκλειστικά στη σημασιολογική αναζήτηση διανυσμάτων (όπως στο THOTH), το δικό μας σύστημα αναγνωρίζει την αστοχία της Vector DB στην παπυρολογία και εφαρμόζει δυναμικά το Character-Level N-Gram Matching. Αυτή η τοπική αναζήτηση αλληλεπικάλυψης γραμμάτων επιτρέπει την επιτυχή ανάκτηση ιστορικών παραλλήλων (ground truth) ακόμη και όταν το LLM τροφοδοτείται με ακατέργαστα δεδομένα OCR, προσφέροντας μια επιστημονικά έγκυρη διέξοδο στο πρόβλημα των παραισθήσεων.

2.2.6 Ανάκτηση Επαυξημένης Παραγωγής (RAG) ως Λύση

Για τον περιορισμό αυτού του κινδύνου, οι ερευνητές εφαρμόζουν αρχιτεκτονικές Ανάκτησης Επαυξημένης Παραγωγής (Retrieval Augmented Generation - RAG).

- **Μηχανισμός**: Αντί να επιτρέπεται στο LLM να παράγει κείμενο μόνο από τα εσωτερικά του βάρη, ένα σύστημα RAG αναζητά πρώτα σε μια επαληθευμένη βάση δεδομένων (όπως η DDbDP) για σχετικές παραλλήλους. Στη συνέχεια, τροφοδοτεί αυτά τα ανακτημένα δεδομένα στο LLM.
- **Εφαρμογή**: Για ένα κατεστραμμένο θραύσμα της Οξυρρύγχου, ένα σύστημα RAG θα ανακτούσε όλους τους γνωστούς παπύρους με παρόμοιες φράσεις. Το LLM θα χρησιμοποιούσε στη συνέχεια μόνο αυτά τα ανακτημένα στοιχεία για να προτείνει μια αποκατάσταση, γειώνοντας τη δημιουργικότητα της Τεχνητής Νοημοσύνης σε ιστορικά τεκμήρια. Αυτό μετατρέπει ουσιαστικά την Τεχνητή Νοημοσύνη σε μια «σημασιολογική μηχανή αναζήτησης» με δυνατότητες συλλογιστικής υψηλού επιπέδου.

2.2.7 Επιστημονικές και Ηθικές Επιπτώσεις

Όταν φέρνουμε αυτές τις τεχνολογίες στις ανθρωπιστικές επιστήμες, δεν γίνεται να μην ξαναδούμε από την αρχή τις βασικές αξίες του πεδίου. Ένα μεγάλο αγκάθι με τη βαθιά μάθηση εδώ είναι το λεγόμενο "μαύρο κουτί" των νευρωνικών δικτύων. Για παράδειγμα, αν ένα μοντέλο όπως το Ithaca υπολογίζει ότι ένας πάπυρος γράφτηκε το 150 μ.Χ., αλλά ο ερευνητής δεν μπορεί να καταλάβει από τα παλαιογραφικά στοιχεία γιατί το μοντέλο κατέληξε σε αυτή τη χρονολογία, τότε αξίζει όντως να πάρουμε αυτή την πληροφορία στα σοβαρά;[41] Τεχνικές όπως οι χάρτες σημαντικότητας και οι ενσωματώσεις προσοχής (που χρησιμοποιούνται στο NeuroPapyri και το Ithaca) επιχειρούν να καταστήσουν τη συλλογιστική της μηχανής διαφανή.

Η ροή εργασίας «Άνθρωπος-στον-Βρόχο» (HITL) [42] αλλάζει τον ρόλο του παπυρολόγου. Ο μελετητής δεν γράφει πια μόνος του το κείμενο. Τώρα αξιολογεί τις προτάσεις που φτιάχνει η μηχανή. Η τεχνητή νοημοσύνη ασχολείται με τις βασικές εργασίες, να διαβάζει, να αποκαθιστά, να χρονολογεί τα κείμενα. Κι αυτό αφήνει τον άνθρωπο να επικεντρώνεται εκεί που έχει τη μεγαλύτερη αξία. Στην ιστορική ερμηνεία και στη λογοτεχνική ανάλυση.

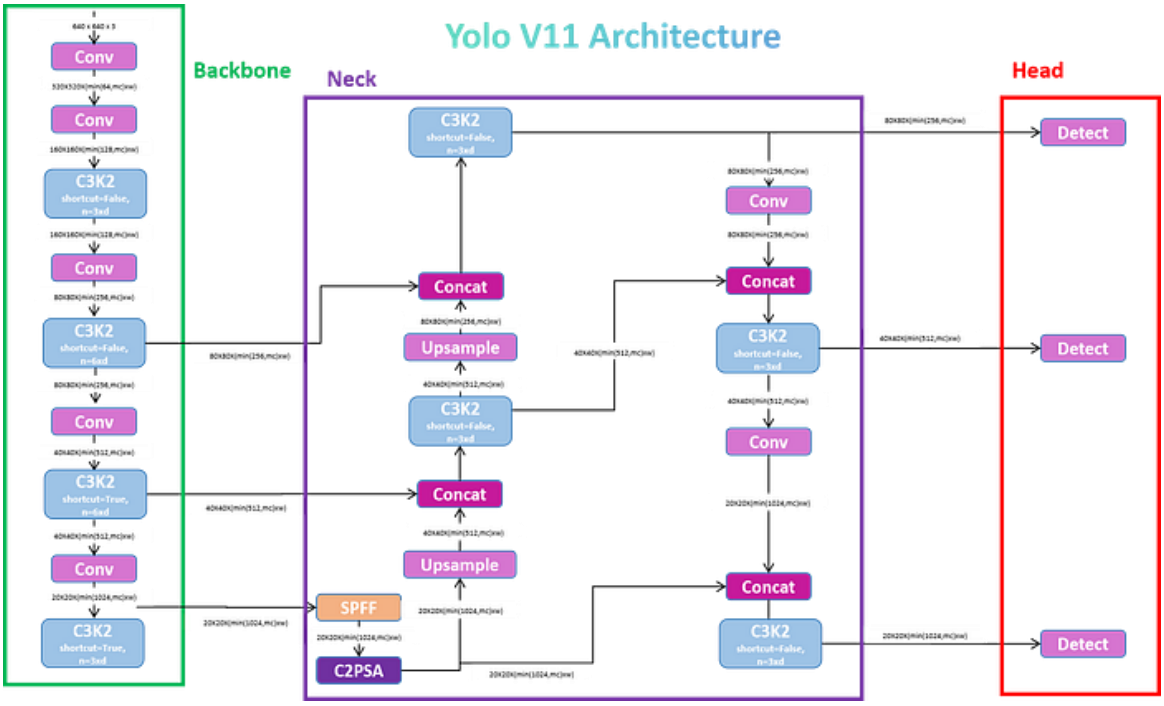
2.2.8 Αλγόριθμοι Ανίχνευσης Αντικειμένων: Οικογένεια YOLO και η Αρχιτεκτονική YOLO11

Η οικογένεια μοντέλων YOLO (You Only Look Once) έφερε επανάσταση στην ανίχνευση αντικειμένων σε πραγματικό χρόνο. Αντίθετα με τα μοντέλα δύο σταδίων που πρώτα προτείνουν περιοχές ενδιαφέροντος και έπειτα τις ταξινομούν, το YOLO μετατρέπει το πρόβλημα του εντοπισμού σε ένα ενιαίο πρόβλημα παλινδρόμησης. Το δίκτυο βλέπει την εικόνα μόνο μία φορά, διαιρώντας την σε ένα πλέγμα και προβλέποντας ταυτόχρονα τα bounding boxes και τις πιθανότητες των κλάσεων για κάθε κελί του πλέγματος.

Η έκδοση YOLO11, η οποία αξιοποιείται στην παρούσα εργασία, αποτελεί την αιχμή αυτής της εξέλιξης, [44] εισάγοντας κρίσιμες αρχιτεκτονικές καινοτομίες που βελτιώνουν δραστικά την εξαγωγή χαρακτηριστικών χωρίς να αυξάνουν υπερβολικά το υπολογιστικό κόστος:

1. **<u>Δομικά Μπλοκ C3k2</u>**: Στο backbone και στο neck του μοντέλου, τα παραδοσιακά μπλοκ C2f έχουν αντικατασταθεί από τα προηγμένα μπλοκ C3k2. Αυτή η δομή ενισχύει τη ροή της πληροφορίας και τη χωρική επεξεργασία των χαρακτηριστικών μέσω επαναλαμβανόμενων συνελίξεων (convolutions), επιτρέποντας στο δίκτυο να αναγνωρίζει πιο περίπλοκα μοτίβα χωρίς να διογκώνεται ο αριθμός των παραμέτρων.
2. **<u>Μονάδα Χωρικής Προσοχής C2PSA</u>**: Το YOLO11 ενσωματώνει τη μονάδα C2PSA (Cross-Stage Partial Spatial Attention / Convolutional block with Parallel Spatial Attention). Η μονάδα αυτή, τοποθετημένη στρατηγικά μετά τη μονάδα SPPF (Spatial Pyramid Pooling - Fast), εισάγει μηχανισμούς προσοχής που επιτρέπουν στο δίκτυο να επικεντρώνεται στις κρίσιμες περιοχές ενδιαφέροντος μιας εικόνας, φιλτράροντας τις άσχετες πληροφορίες του υποβάθρου. Στην παπυρολογία, αυτό είναι ζωτικής σημασίας, καθώς ο αλγόριθμος πρέπει να εστιάσει στα αχνά ίχνη μελανιού και να αγνοήσει τον «θόρυβο» (π.χ. τρύπες, λεκέδες, διασταυρούμενες ίνες του παπύρου).
3. **<u>Κεφαλή Ανίχνευσης χωρίς Αγκυρώσεις (Anchor-Free Head):</u>** Συνεχίζοντας την παράδοση των νεότερων εκδόσεων, η κεφαλή του YOLO11 είναι "anchor-free", απαλλάσσοντας τον ερευνητή από τη χειροκίνητη ρύθμιση προκαθορισμένων πλαισίων (anchor boxes) και απλοποιώντας τον υπολογισμό

της συνάρτησης απώλειας (loss function). Εξάγει προβλέψεις σε πολλαπλές κλίμακες χαρακτηριστικών (multi-scale features), επιτρέποντας τον εξίσου ακριβή εντοπισμό ενός δυσανάγνωστου μικροσκοπικού γράμματος ή ενός μεγάλου συμπλέγματος χαρακτήρων.



Εικόνα 1. Αρχιτεκτονική YOLOv11

Πηγή: https://medium.com/@nikhil-rao-20/yolov11-explained-next-level-object-detection-with-enhanced-speed-and-accuracy-2dbe2d376f71

2.2.9 Συνελικτικά Νευρωνικά Δίκτυα (CNNs): Η Αρχιτεκτονική ResNet-50

Η αρχιτεκτονική ResNet (Residual Network), η οποία δημιουργήθηκε το 2015, σχεδιάστηκε με σκοπό να επιλύσει ένα από τα πιο επίμονα προβλήματα της βαθιάς μάθησης: το πρόβλημα της εξαφάνισης της κλίσης (vanishing gradient) και της υποβάθμισης της απόδοσης (degradation problem). Θεωρητικά, η προσθήκη επιπλέον επιπέδων σε ένα νευρωνικό δίκτυο θα έπρεπε να οδηγεί σε καλύτερη ή τουλάχιστον ίση απόδοση. Στην πράξη, ωστόσο, παρατηρήθηκε ότι καθώς το δίκτυο γινόταν πολύ βαθύ (π.χ. 50 ή 100 επίπεδα), το σφάλμα εκπαίδευσης αυξανόταν, καθώς η πληροφορία "χανόταν" κατά την οπισθοδιάδοση (backpropagation). [45]

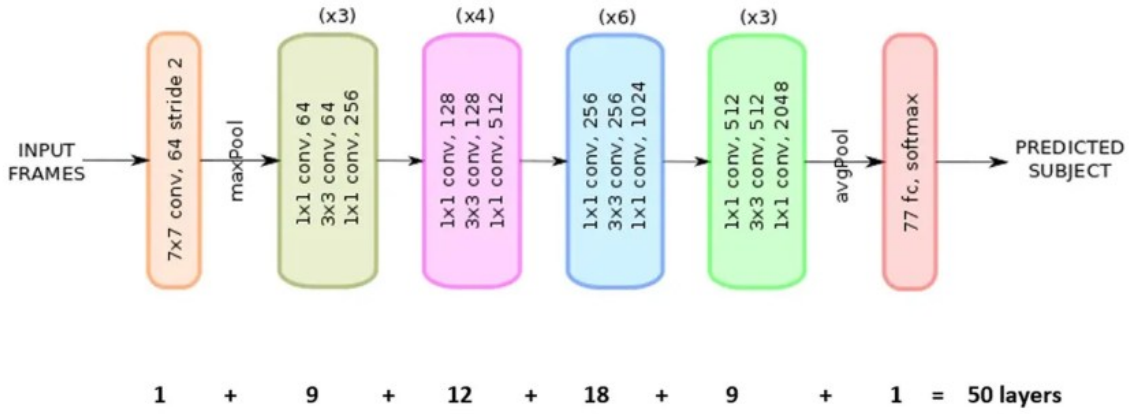
Η επαναστατική λύση του ResNet ήταν η εισαγωγή των "υπολειπόμενων συνδέσεων" (skip connections ή shortcut connections). Αντί να αναγκάζει ένα επίπεδο να μάθει απευθείας την επιθυμητή υποκείμενη χαρτογράφηση $H(x)$, το ResNet επαναπροσδιορίζει το δίκτυο ώστε να μαθαίνει μια υπολειπόμενη συνάρτηση (residual function) $F(x) =$

$H(x) - x$. Η τελική έξοδος του μπλοκ προκύπτει από την πρόσθεση της αρχικής εισόδου (που προσπερνά τα ενδιάμεσα επίπεδα) με την έξοδο της συνέλιξης, δηλαδή $F(x) = H(x) + x$. Αυτή η αρχιτεκτονική επιτρέπει στο δίκτυο να "παρακάμπτει" επίπεδα που δεν προσφέρουν χρήσιμη πληροφορία (θέτοντας τα βάρη τους στο μηδέν), διευκολύνοντας δραστικά την εκπαίδευση εξαιρετικά βαθιών δικτύων.

Το μοντέλο ResNet-50 που χρησιμοποιείται στην παρούσα εργασία αποτελείται από 50 επίπεδα και χρησιμοποιεί ειδικά bottleneck blocks. Κάθε τέτοιο μπλοκ αποτελείται από μια ακολουθία τριών συνελικτικών επιπέδων:

- Ένα επίπεδο 1 X 1 που μειώνει τις διαστάσεις του feature map.
- Ένα επίπεδο 3 X 3 στο οποίο πραγματοποιείται η κύρια εκμάθηση των χαρακτηριστικών.
- Ένα τελικό επίπεδο 1 X 1 που επαναφέρει τις διαστάσεις πριν την πρόσθεση της υπολειπόμενης σύνδεσης.

Αυτός ο σχεδιασμός καθιστά το ResNet-50 εξαιρετικά ικανό στο να εντοπίζει ιεραρχικά χαρακτηριστικά. Τα πρώτα επίπεδα αντιλαμβάνονται απλές ακμές και καμπύλες του μελανιού, ενώ τα βαθύτερα επίπεδα αναγνωρίζουν τα σύνθετα γεωμετρικά σχήματα της παλαιογραφίας, όπως τις ιδιαιτερότητες του μηνοειδούς σίγμα (C) ή ενός φθαρμένου ωμέγα (Ω).



Εικόνα 2. Αρχιτεκτονική ResNet-50

Πηγή: https://www.ultralytics.com/blog/what-is-resnet-50-and-what-is-its-relevance-in-computer-vision

2.2.10 Μετασχηματιστές Όρασης: DeiT (Data-efficient Image Transformers)

Ενώ τα CNNs χρησιμοποιούν local filters, μια νέα γενιά μοντέλων αναδύθηκε με τους Μετασχηματιστές Όρασης (Vision Transformers - ViT). [46] Τα μοντέλα ViT αντιμετωπίζουν την εικόνα ως μια ακολουθία από μικρά επίπεδα τμήματα, με τον ίδιο τρόπο που ένα γλωσσικό μοντέλο αντιμετωπίζει τις λέξεις μιας πρότασης. Ωστόσο, σε

αντίθεση με τα CNNs, τα ViTs στερούνται εγγενούς επαγωγικής προκατάληψης (inductive bias). Εξαιτίας αυτού, για να κατανοήσουν τη δομή μιας εικόνας, απαιτούν τεράστιους όγκους δεδομένων εκπαίδευσης, καθιστώντας τα μη πρακτικά για εξειδικευμένους τομείς με περιορισμένα δεδομένα, όπως η παπυρολογία.

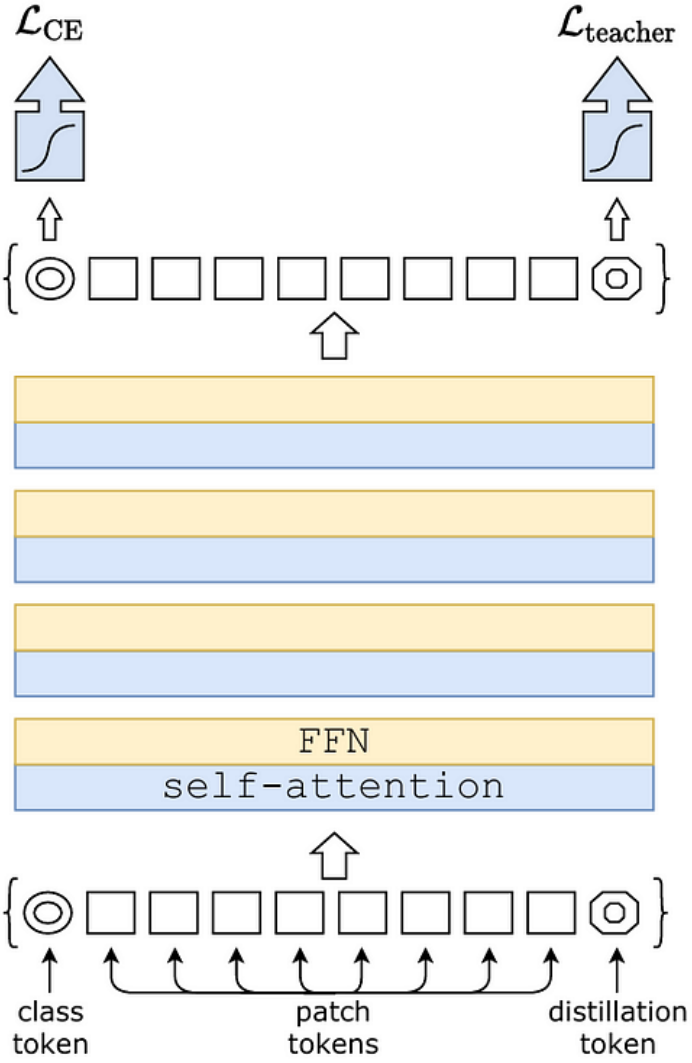
Η αρχιτεκτονική DeiT (Data-efficient Image Transformers) σχεδιάστηκε ακριβώς για να επιλύσει αυτόν τον περιορισμό, επιτρέποντας την εκπαίδευση ισχυρών μετασχηματιστών με πολύ λιγότερα δεδομένα. Η θεμελιώδης καινοτομία του DeiT είναι η εισαγωγή ενός distillation token. [47]

Σε ένα τυπικό ViT, υπάρχει μόνο ένα class token που συλλέγει την πληροφορία από όλα τα patches για να δώσει την τελική πρόβλεψη. Το DeiT προσθέτει ένα δεύτερο token, το οποίο περνάει μέσα από όλα τα επίπεδα αυτο-προσοχής παράλληλα με το class token. Μέσω μιας στρατηγικής εκπαίδευσης «δασκάλου-μαθητή», το DeiT (ως μαθητής) προσπαθεί να ελαχιστοποιήσει την απόκλιση όχι μόνο ως προς τις πραγματικές ετικέτες, αλλά και ως προς τις προβλέψεις ενός ισχυρού CNN δασκάλου (teacher network).

Μαθηματικά, η συνάρτηση απώλειας στην εκπαίδευση με hard distillation συνδυάζει το σφάλμα ως προς την πραγματική ετικέτα L_{CE} και το σφάλμα ως προς την πρόβλεψη του δασκάλου $L_{distill}$

$$L_{global} = \frac{1}{2}\left(L_{CE} + L_{distill}\right)$$

Αποτέλεσμα αυτής της αρχιτεκτονικής είναι ότι το distillation token μαθαίνει διακριτές αναπαραστάσεις σε σχέση με το class token, απορροφώντας την προκατάληψη του CNN δασκάλου, ενώ διατηρεί την ικανότητα καθολικής προσοχής των transformers. Κατά το inference, οι έξοδοι και των δύο tokens υπολογίζονται και λαμβάνεται ο μέσος όρος τους για την τελική πρόβλεψη. Αυτός ο συνδυασμός τοπικότητας και καθολικότητας καθιστά το DeiT εξαιρετικά ισχυρό στο να ξεχωρίζει τον πραγματικό χαρακτήρα από τον περιβάλλοντα φθαρμένο πάπυρο.



Εικόνα 3. Αρχιτεκτονική DeiT

Πηγή: https://towardsdatascience.com/vision-transformer-on-a-budget/

2.2.11 Η Αρχιτεκτονική Transformer και ο Μηχανισμός Αυτο-Προσοχής

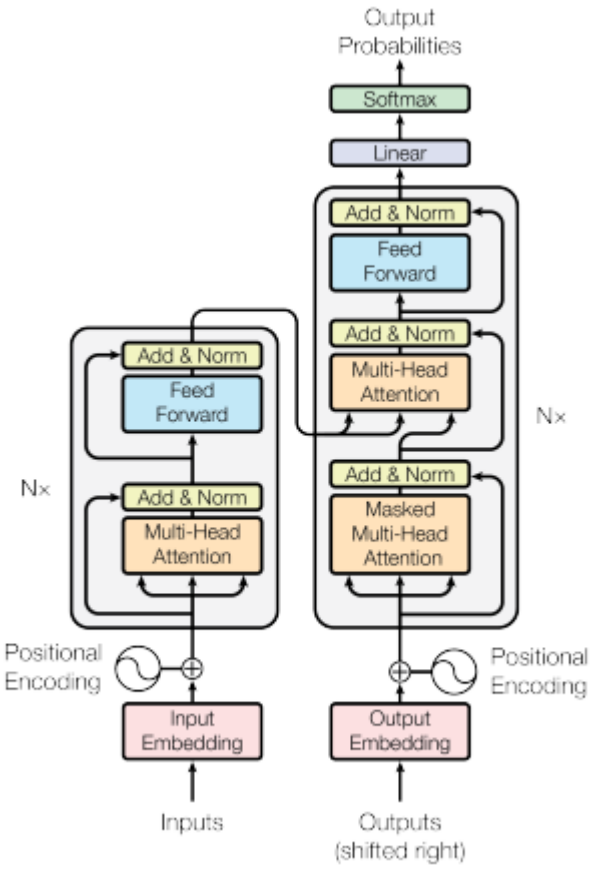
Στο δεύτερο μισό της προτεινόμενης αρχιτεκτονικής, το βάρος μετατοπίζεται από την υπολογιστική όραση στην Επεξεργασία Φυσικής Γλώσσας (NLP). Τα σύγχρονα Μεγάλα Γλωσσικά Μοντέλα, όπως το Gemini της Google ή το Llama της Meta, στηρίζονται εξ ολοκλήρου στην επαναστατική αρχιτεκτονική Transformer που παρουσιάστηκε το 2017. [48]

Πριν τους Transformers, η επεξεργασία κειμένου γινόταν μέσω Αναδρομικών Νευρωνικών Δικτύων (RNNs), τα οποία διάβαζαν το κείμενο σειριακά (λέξη-λέξη). Αυτό τα καθιστούσε αργά και ανίκανα να διατηρήσουν τη μνήμη τους σε μακροσκελή κείμενα.

Οι Transformers εγκατέλειψαν πλήρως τη σειριακή επεξεργασία, εισάγοντας τον μηχανισμό Αυτο-Προσοχής. [48]

Ο μηχανισμός αυτο-προσοχής επιτρέπει στο μοντέλο, κατά την ανάλυση μιας λέξης, να υπολογίσει δυναμικά πόση σημασία πρέπει να δώσει σε κάθε άλλη λέξη της ίδιας πρότασης, ανεξάρτητα από τη φυσική τους απόσταση. Η διαδικασία αυτή υλοποιείται μέσω τριών πινάκων που μαθαίνονται κατά την εκπαίδευση. Του Ερωτήματος (Query - Q), του Κλειδιού (Key - K) και της Τιμής (Value - V). Το "Query" αναζητά σχετικές πληροφορίες, το "Key" περιέχει τα χαρακτηριστικά που αντιστοιχούν σε αυτήν την αναζήτηση, και το "Value" περιέχει την πραγματική πληροφορία του token. Το αποτέλεσμα του πολλαπλασιασμού αυτών των πινάκων ορίζει το attention score. Η χρήση "Πολλαπλών Κεφαλών Προσοχής" (Multi-Head Attention) επιτρέπει στο μοντέλο να εστιάζει ταυτόχρονα σε διαφορετικές σχέσεις.

Θεμελιωδώς, τα παραγωγικά μοντέλα κειμένου εκπαιδεύονται βάσει της αρχής του next-token prediction. [49] Αναλύοντας τεράστια σώματα κειμένων, μαθαίνουν τις στατιστικές κατανομές πιθανοτήτων της γλώσσας. Στο πλαίσιο της παπυρολογίας, όταν το LLM συναντά έναν δείκτη κενού (π.χ. [UNK]), δεν «μαντεύει» τυχαία. Χρησιμοποιεί τον μηχανισμό προσοχής για να σαρώσει όλο το διασωθέν κείμενο πριν και μετά το κενό, κατανοώντας τους συντακτικούς, γραμματικούς και ιστορικούς κανόνες, ώστε να παράγει την πιο πιθανή λέξη που ταιριάζει στο αρχαιοελληνικό λεξιλόγιο.



Εικόνα 4. Αρχιτεκτονική Transformer

Πηγή https://arxiv.org/abs/1706.03762

Για την κατανόηση της εφαρμογής των LLMs στην αποκατάσταση κειμένων, είναι κρίσιμη η διάκριση μεταξύ των διαφορετικών αρχιτεκτονικών Transformer. Ιστορικά, τα μοντέλα που εφαρμόστηκαν στην επιγραφική και την παπυρολογία βασίστηκαν κυρίως στην αρχιτεκτονική Encoder-only, όπως το BERT (Bidirectional Encoder Representations from Transformers). [50] Αυτά τα μοντέλα εκπαιδεύονται μέσω του Masked Language Modeling . Στο MLM, το μοντέλο βλέπει ολόκληρη την πρόταση ταυτόχρονα, με κάποιους χαρακτήρες ή λέξεις να έχουν "κρυφτεί", και καλείται να προβλέψει αποκλειστικά τα κρυμμένα στοιχεία χρησιμοποιώντας bidirectional πλαίσιο.

Αντιθέτως, τα σύγχρονα παραγωγικά LLMs βασίζονται συνήθως σε αρχιτεκτονικές Decoder-only. [51] Εκπαιδεύονται autoregressively μέσω του Causal Language Modeling. Αυτό σημαίνει ότι το μοντέλο δεν μπορεί να "δει" το μέλλον (δεξιά του κενού) με τον ίδιο δομικό τρόπο που το κάνει ένας Encoder. Διαβάζει το κείμενο αυστηρά από αριστερά προς τα δεξιά και προβλέπει το επόμενο token. Ωστόσο, λόγω του τεράστιου μεγέθους τους (δισεκατομμύρια παράμετροι) και του Μηχανισμού Προσοχής, τα LLMs έχουν αναπτύξει βαθιά ικανότητα σημασιολογικής συλλογιστικής, επιτρέποντάς τους να υπερβούν τον περιορισμό της μονόδρομης ανάγνωσης.

Η εφαρμογή της Τεχνητής Νοημοσύνης στην αποκατάσταση αρχαίων κειμένων σημαδεύτηκε από ορόσημα όπως τα συστήματα Pythia, [28] Ithaca [29] και πρόσφατα το Aeneas [30] της DeepMind. Αν και πρωτοποριακά, τα εξειδικευμένα αυτά μοντέλα παρουσιάζουν συγκεκριμένους αρχιτεκτονικούς περιορισμούς, τους οποίους η χρήση γενικών LLMs (σε συνδυασμό με συστήματα RAG) έρχεται να επιλύσει. Η σύγκριση των δύο προσεγγίσεων αναδεικνύει τρεις βασικούς άξονες διαφοροποίησης:

Η πιο σημαντική αδυναμία του Ithaca έγκειται στη φύση της εκπαίδευσής του. Ως μοντέλο MLM, απαιτεί από τον φιλόλογο να γνωρίζει τον ακριβή αριθμό των χαμένων γραμμάτων σε μια φθορά (π.χ. πρέπει να του δοθεί η είσοδος "ἔδ[---]εν", υποδηλώνοντας ακριβώς τρία χαμένα γράμματα για να προτείνει το "ἔδ[ωκ]εν"). Στην πραγματικότητα της παπυρολογίας, όταν λείπει ένα ολόκληρο τμήμα του παπύρου, ο ακριβής αριθμός των χαρακτήρων είναι συχνά αδύνατο να προσδιοριστεί.

Τα παραγωγικά LLMs ουσιαστικά "προλαβαίνουν" το πρόβλημα πριν καν εμφανιστεί. Επειδή δημιουργούν κείμενο, μπορούν να συμπληρώσουν όποιο κενό συνθέτοντας λέξεις και φράσεις που ξαναδίνουν νόημα, είτε ιστορικά είτε γραμματικά, χωρίς να νοιάζονται για το πόσο υλικό λείπει ακριβώς. Μοντέλα όπως το Ithaca βλέπουν την αποκατάσταση απλά σαν ένα πρόβλημα συμπλήρωσης μοτίβων, βασιζόμενα σε αυτά που έχουν ήδη δει στις επιγραφές που διδάχτηκαν. Δεν καταλαβαίνουν το βαθύτερο νόημα ή τη μετάφραση ενός κειμένου. Από την άλλη, τα νεότερα LLMs ξέρουν μέσα τους την αρχαία ελληνική γραμματική, τη σύνταξη, την ιστορία. Ελέγχουν αν αυτό που προτείνουν ταιριάζει απόλυτα στο νόημα της πρότασης, κι επιπλέον, μπορούν να δουλέψουν σε πολλά επίπεδα ταυτόχρονα. Δηλαδή, όσο αποκαθιστούν το κείμενο, παράγουν και την αγγλική μετάφραση δίπλα. Τα LLMs είναι πιθανοτικές μηχανές σχεδιασμένες να παράγουν κείμενο που ακούγεται αληθοφανές, όχι απαραίτητα αληθές. [39]

2.3 Αρχιτεκτονική Ανάκτησης Επαυξημένης Παραγωγής (RAG)

Παρά το πόσο καλά τα Μεγάλα Γλωσσικά Μοντέλα μπορούν να καταλάβουν και να συνθέσουν φυσική γλώσσα, έχουν ένα σημαντικό επιστημονικό πρόβλημα όταν μπαίνουν σε αυστηρά ακαδημαϊκά πεδία, όπως η ιστορική και φιλολογική έρευνα. Το βασικό τους πρόβλημα είναι η τάση τους για «παραισθήσεις», δηλαδή να παράγουν πληροφορίες που φαίνονται αληθινές, αλλά δεν είναι.

Αυτό διότι ό,τι απαντούν βασίζεται στα δεδομένα που έχουν αποθηκευτεί ως βάρη στο νευρωνικό τους δίκτυο από την αρχική εκπαίδευση, και όχι σε πραγματική, επαληθευμένη ιστορική γνώση. Στόχος τους είναι να δώσουν απαντήσεις που ταιριάζουν στατιστικά, όχι πάντα τις σωστές από την ιστορική άποψη.

Στην παπυρολογία, για παράδειγμα, αν ζητήσεις από ένα LLM να συμπληρώσει ένα κενό σε έναν πάπυρο, μπορεί να φτιάξει λέξεις που ακούγονται σωστά μορφολογικά και γραμματικά στην αρχαία ελληνική. Αυτές όμως μπορεί να μην έχουν υπάρξει ποτέ

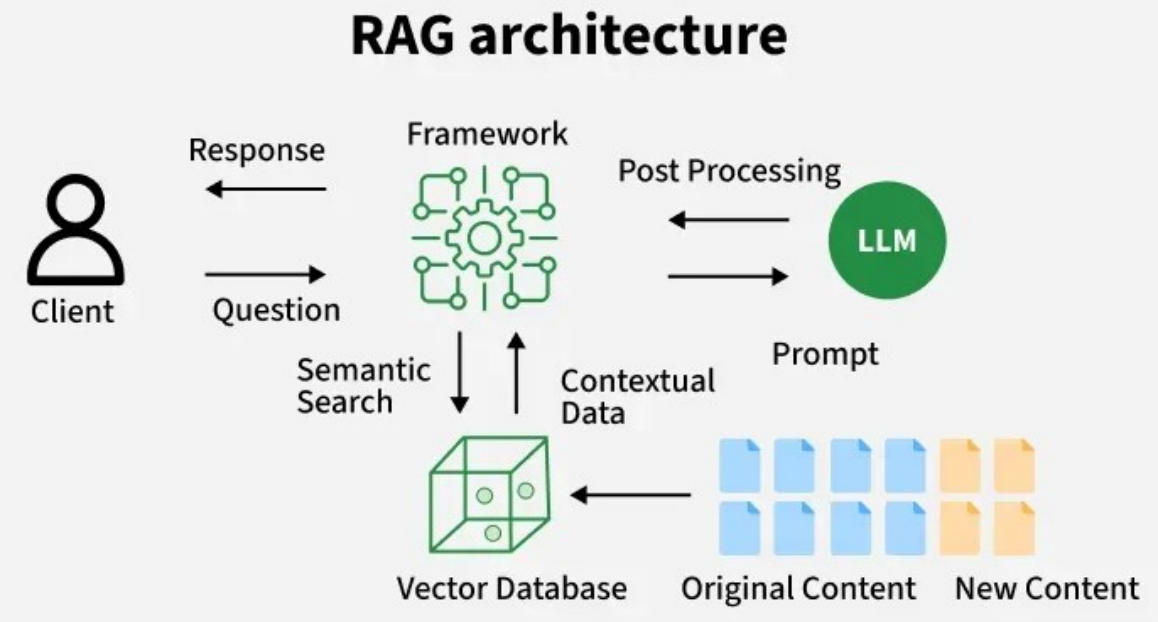
ιστορικά, π.χ. να εφευρίσκει ονόματα αξιωματούχων ή διοικητικούς όρους που δεν υπήρχαν στην Πτολεμαϊκή εποχή.

Η αρχιτεκτονική Ανάκτησης Επαυξημένης Παραγωγής (Retrieval-Augmented Generation - RAG) εισήχθη το 2020 ως το απόλυτο θεωρητικό και πρακτικό αντίδοτο σε αυτό το πρόβλημα. [52] Το RAG γεφυρώνει το μοντέλο με μια εξωτερική, δυναμική και «μη παραμετρική» βάση επαληθευμένων γνώσεων (π.χ. βάση δεδομένων με ήδη δημοσιευμένους παπύρους). Η αρχιτεκτονική αυτή αποτελείται από τρία κύρια και αλληλεξαρτώμενα στάδια:

1. **Ανάκτηση (Retrieval):** Όταν τίθεται ένα ερώτημα (π.χ. η αποκατάσταση μιας συγκεκριμένης φράσης), το σύστημα δεν απευθύνεται αμέσως στο LLM. Αντίθετα, μετατρέπει το ερώτημα σε ένα αριθμητικό διάνυσμα (embedding) χρησιμοποιώντας ένα μοντέλο ενσωμάτωσης. Στη συνέχεια, αναζητά σε μια διανυσματική βάση δεδομένων, χρησιμοποιώντας μετρικές όπως η Ομοιότητα Συνημιτόνου, τα ιστορικά έγγραφα που βρίσκονται σημασιολογικά πιο κοντά στο ερώτημα.
2. **Επαύξηση (Augmentation):** Τα έγγραφα που ανακτήθηκαν (στην περίπτωσή μας, αληθινές μεταγραφές αρχαίων παπύρων) λειτουργούν ως ιστορικό πλαίσιο (context). Το αρχικό prompt του χρήστη "επαυξάνεται", καθώς τα ανακτημένα δεδομένα εισάγονται δυναμικά μέσα σε αυτό.
3. **Παραγωγή (Generation):** Το LLM λαμβάνει πλέον το επαυξημένο prompt. Οι οδηγίες το αναγκάζουν να αντλήσει πληροφορίες από το παρέχον ιστορικό πλαίσιο για να απαντήσει ή να συμπληρώσει το κείμενο.

Με την αρχιτεκτονική RAG, το LLM σταματά να «φαντάζεται» χωρίς όρια και μένει προσγειωμένο στην πραγματική ιστορία. Αυτό κάνει μεγάλη διαφορά για τις Ανθρωπιστικές Επιστήμες γιατί τώρα έχουμε ιχνηλασιμότητα. Δηλαδή, αν το σύστημα προτείνει μια συγκεκριμένη συμπλήρωση σε έναν φθαρμένο πάπυρο, ο ερευνητής βλέπει ακριβώς ποια έγγραφα από τη διανυσματική βάση το οδήγησαν εκεί. Έτσι, το AI φεύγει

από τη λογική του αδιαφανούς «μαύρου κουτιού» και γίνεται ένας διαφανής, ελέγξιμος, επιστημονικά τεκμηριωμένος βοηθός για τη φιλολογική αποκατάσταση.



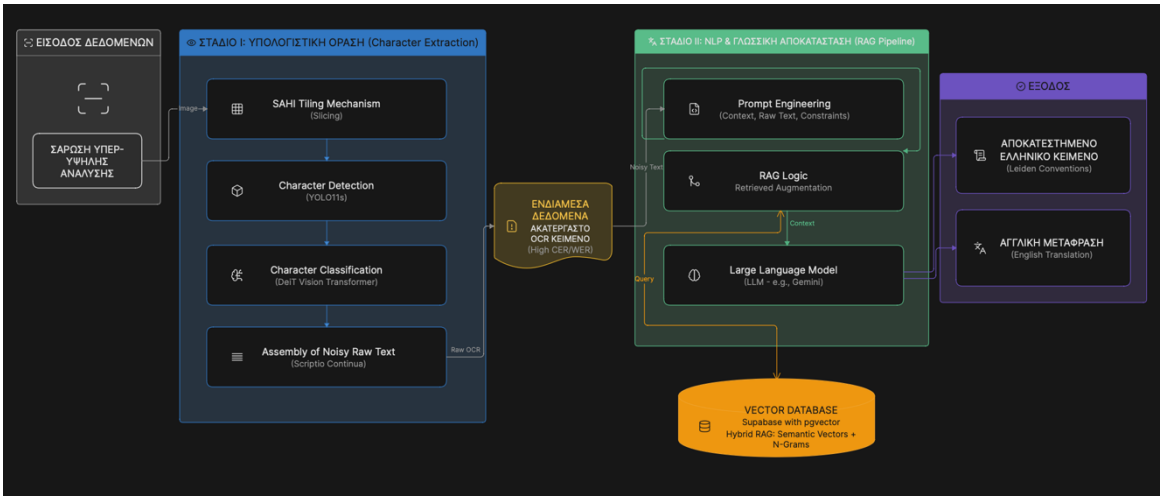
Εικόνα 5. Αρχιτεκτονική RAG

Πηγή https://www.geeksforgeeks.org/nlp/rag-architecture/

3. Μεθοδολογία Οπτικής Ανίχνευσης και Σημασιολογικής Αποκατάστασης

Μεταβαίνοντας από την παραδοσιακή παπυρολογία στην εποχή της υπολογιστικής ανάλυσης, χρειάζεται να χτίσουμε μια στιβαρή, πολυεπίπεδη αρχιτεκτονική. Μια αρχιτεκτονική που δεν κολλάει μπροστά στην αβεβαιότητα ή στη φθορά των αρχαίων κειμένων, αλλά τις διαχειρίζεται έξυπνα. Αυτό το κεφάλαιο ξεδιπλώνει βήμα-βήμα τη μεθοδολογία της διπλωματικής. Εδώ, βρίσκονται δίπλα δίπλα η Υπολογιστική Όραση (Computer Vision) και η Επεξεργασία Φυσικής Γλώσσας (NLP), και συνεργάζονται στενά.

Η υλοποίηση του συστήματος ακολουθεί ένα ενιαίο pipeline. Ξεκινάει με τη λήψη της ψηφιακής εικόνας του παπύρου, βρίσκει και ταξινομεί χαρακτήρες, εντοπίζει φυσικές φθορές μέσω ειδικού αλγορίθμου ανίχνευσης κενών και οδηγεί τελικά στην αποκατάσταση του κειμένου μέσω των Μεγάλων Γλωσσικών Μοντέλων και της αρχιτεκτονικής RAG.



Εικόνα 6. Διάγραμμα αρχιτεκτονικής συστήματος

3.1 Συλλογή, Προεπεξεργασία και Διαχείριση Δεδομένων

Η θεμελιώδης προϋπόθεση για την επιτυχή εκπαίδευση των μοντέλων βαθιάς μάθησης είναι η ποιότητα, η ποικιλομορφία και η ορθή δόμηση των δεδομένων εκπαίδευσης. Η πολυπλοκότητα της αρχαίας ελληνικής γραφής, καθιστά αναγκαία τη χρήση εξειδικευμένων συνόλων δεδομένων για κάθε διακριτό στάδιο του συστήματος. Στην παρούσα υλοποίηση χρησιμοποιήθηκαν τρεις διαφορετικές πηγές δεδομένων: το ICDAR 2023 για τον εντοπισμό [25], το AL-PUB [27] για την κατηγοριοποίηση και το αποθετήριο του Papyri.info [53] για την κατασκευή της σημασιολογικής βάσης γνώσης.

Το σύνολο δεδομένων "ICDAR 2023 Competition on Detection and Recognition of Greek Letters on Papyri" επιλέχθηκε ειδικά για την εκπαίδευση του μηχανισμού εντοπισμού χαρακτήρων. Το συγκεκριμένο σύνολο περιλαμβάνει εικόνες υψηλής ανάλυσης από παπύρους που διασώζουν τμήματα της Ιλιάδας του Ομήρου, καλύπτοντας μια χρονική περίοδο πολλών αιώνων και αντιπροσωπεύοντας ποικίλα γραφικά στυλ και καταστάσεις διατήρησης. Δεδομένου ότι η στρατηγική της παρούσας εργασίας διαχωρίζει ρητά τον εντοπισμό από την αναγνώριση, τα δεδομένα του ICDAR υπέστησαν μια ριζική μετατροπή. Αντί να διατηρηθούν οι 227 διαφορετικές κλάσεις χαρακτήρων που προέβλεπε η αρχική μορφή COCO (συμπεριλαμβανομένων διακριτικών και πνευμάτων), υλοποιήθηκε ένα script σε Python το οποίο μετέτρεψε όλες τις ετικέτες σε μορφή YOLO, ορίζοντας τον δείκτη κλάσης σταθερά στο 0. Αυτή η προσέγγιση της «μονής κλάσης» εκπαιδεύει το μοντέλο να αναγνωρίζει αποκλειστικά το μοτίβο «χαρακτήρας-μελάνι έναντι υποβάθρου-παπύρου», μειώνοντας δραματικά το γνωστικό φορτίο του δικτύου κατά τη φάση της εκμάθησης των χωρικών χαρακτηριστικών και αυξάνοντας την ικανότητα γενίκευσης του μοντέλου σε άγνωστους γραφικούς χαρακτήρες.

Για το δεύτερο στάδιο της κατηγοριοποίησης, δηλαδή την ταυτοποίηση του ακριβούς γράμματος εντός του εντοπισμένου πλαισίου, ενσωματώθηκε το σύνολο δεδομένων AL-PUB (Ancient Lives Public). Το σύνολο αυτό προέκυψε από τη συλλογική προσπάθεια

χιλιάδων εθελοντών και αποτελείται από περισσότερες από 205.000 περικομμένες εικόνες μεμονωμένων χαρακτήρων, οι οποίες προέρχονται από το αρχείο της Οξυρρύγχου. Το AL-PUB καλύπτει το σύνολο του ελληνικού αλφαβήτου, με την κρίσιμη ενσωμάτωση του μηνοειδούς σίγμα (Lunate Sigma - C), το οποίο αντικαθιστά σχεδόν καθολικά το τυπικό κεφαλαίο Σ στα αρχαία κείμενα. Λόγω του τεράστιου όγκου των δεδομένων και της εγγενούς ανισοκατανομής τους, αναπτύχθηκε ένας μηχανισμός downsampling. Συγκεκριμένα, το σύστημα περιορίζει τυχαία το δείγμα σε ένα αυστηρό μέγιστο όριο 2.000 εικόνων ανά κλάση. Η επιλογή της συγκεκριμένης υπερπαραμέτρου τεκμηριώνεται εμπειρικά και υπολογιστικά βάσει τριών αξόνων:

- **Αντιμετώπιση Class Imbalance**: Στη φυσική ροή της ελληνικής γλώσσας, φωνήεντα όπως το 'Α' ή το 'Ο' εμφανίζονται συντριπτικά συχνότερα από σύμφωνα όπως το 'Ψ' ή το 'Ξ'. Το όριο των 2.000 εικόνων επιλέχθηκε στρατηγικά, καθώς προσεγγίζει τη μέγιστη φυσική διαθεσιμότητα των σπανιότερων κλάσεων στο σύνολο δεδομένων. Υποβαθμίζοντας (downsampling) τις κυρίαρχες κλάσεις σε αυτό το κατώφλι, δημιουργείται μια απολύτως ισοσταθμισμένη κατανομή, διασφαλίζοντας ότι τα μοντέλα δεν θα αναπτύξουν στατιστική μεροληψία υπέρ των συχνών γραμμάτων κατά την πρόβλεψη.
- **Υπολογιστική Αποδοτικότητα**: Το συγκεκριμένο όριο περιορίζει το τελικό σύνολο εκπαίδευσης σε περίπου 48.000 εικόνες (24 κλάσεις × 2.000). Το μέγεθος αυτό επιτρέπει την ταχύτατη εκπαίδευση των μοντέλων εντός λογικών χρονικών πλαισίων χρησιμοποιώντας συμβατικές μονάδες GPU, διευκολύνοντας τον γρήγορο πειραματισμό και τη βελτιστοποίηση των υπερπαραμέτρων, χωρίς να θυσιάζεται η ικανότητα γενίκευσης του δικτύου.

Τέλος, για την τροφοδότηση του συστήματος Ανάκτησης Επαυξημένης Παραγωγής (RAG), αντλήθηκαν ιστορικά κείμενα από το αποθετήριο ανοικτού κώδικα papyri/idp.data (Integrating Digital Papyrology), το οποίο ενοποιεί τις βάσεις DDB (Documentary Papyri), DCLP (Literary Papyri) και HGV (Heidelberger Gesamtverzeichnis).[5] Η προεπεξεργασία αυτών των δεδομένων ήταν, χωρίς αμφιβολία, το πιο δύσκολο κομμάτι όταν μιλάμε για τη διαχείριση των μεταδεδομένων. Τα αρχεία του αποθετηρίου είναι φτιαγμένα με το πρότυπο EpiDoc XML, μια ειδική παραλλαγή του Text Encoding Initiative (TEI). Για να μπορέσουν τα κείμενα να "διαβαστούν" από τα Μεγάλα Γλωσσικά Μοντέλα (LLMs), φτιάξαμε έναν converter που παίρνει τη δομή της XML και τη μετατρέπει στη μορφή καταγραφής Leiden. Αυτό το βήμα είναι κρίσιμο, γιατί πρέπει να διατηρηθούν όλοι οι δείκτες των εκδοτών: αγκύλες για τα αποκατεστημένα σημεία, υποστιγμές όταν η ανάγνωση είναι αβέβαιη, ειδικά σύμβολα για διαγραφές ή κενά. Έτσι, η διανυσματική βάση δεδομένων καταγράφει όχι μόνο τις λέξεις, αλλά και όλη την υπόλοιπη διαθέσιμη πληροφορία. Με αυτά τα δεδομένα, το LLM μπορεί μετά να ερμηνεύει σωστά τα κενά στις επιγραφές. Ο παρακάτω πίνακας συνοψίζει τα χαρακτηριστικά των τριών βασικών συνόλων δεδομένων που χρησιμοποιήθηκαν στην υλοποίηση της αρχιτεκτονικής.

Σύνολο Δεδομένων	Κύρια Χρήση στο Σύστημα	Δομή και Μορφή Δεδομένων	Εφαρμοσθείσα Προεπεξεργασία
ICDAR 2023	Εντοπισμός Χαρακτήρων (Localization)	Υψηλής ανάλυσης εικόνες παπύρων με ετικέτες μορφής COCO.	Μετατροπή σε μορφή YOLO, κανονικοποίηση συντεταγμένων, συγχώνευση σε μία (1) κλάση.
AL-PUB	Κατηγοριοποίηση (Classification)	~205.000 περικομμένες εικόνες μεμονωμένων χαρακτήρων.	Στρωματοποιημένη δειγματοληψία (max 2.000 ανά κλάση), διαχωρισμός Train/Val (80/20).
Papyri.info (IDP)	Βάση Γνώσης RAG (Semantic Search)	Χιλιάδες έγγραφα κωδικοποιημένα σε πρότυπο EpiDoc XML.	Εξαγωγή σε JSONL, μετατροπή κειμένου στο σύστημα Leiden+ για διατήρηση μεταδεδομένων.

Πίνακας 1. Σύνολα Δεδομένων

3.2 Υποσύστημα Υπολογιστικής Όρασης: Εντοπισμός Χαρακτήρων με YOLO

Η πρώτη φάση του οπτικού pipeline εστιάζει στην εξαγωγή των χωρικών συντεταγμένων των χαρακτήρων από την ακατέργαστη εικόνα του παπύρου. Αντί της χρήσης παραδοσιακών τεχνικών οπτικής αναγνώρισης χαρακτήρων, οι οποίες αποτυγχάνουν συστηματικά σε θορυβώδη υποστρώματα και φθαρμένα μελάνια, η παρούσα υλοποίηση ενσωματώνει προηγμένα μοντέλα Object Detection της οικογένειας YOLO (You Only Look Once).

Στην παρούσα εργασία εκπαιδεύτηκαν δύο διακριτά μοντέλα YOLO, με το πρωτεύον μοντέλο να εξειδικεύεται αυστηρά στον εντοπισμό, διαχωρίζοντας τη διαδικασία ανίχνευσης από την κατηγοριοποίηση. Η επιλογή της αρχιτεκτονικής YOLO11s (Small) προσφέρει τη βέλτιστη αναλογία μεταξύ ταχύτητας εξαγωγής συμπερασμάτων (inference speed) και αναλυτικής ακρίβειας. Η αρχιτεκτονική του μοντέλου απαρτίζεται από 182 επίπεδα, συνθέτοντας ένα δίκτυο 9.428.179 παραμέτρων που απαιτεί 21,5 GFLOPs υπολογιστικής ισχύος.

Ο συνολικός υπολογισμός της απώλειας γίνεται ως εξής: Χρησιμοποιούμε κυρίως το CIoU Loss για να εντοπίσουμε το πλαίσιο, επειδή αυτό δεν μετράει μόνο το πόσο επικαλύπτονται τα πλαίσια μεταξύ τους, αλλά κοιτάζει και πόσο κοντά βρίσκονται τα κέντρα τους και τις αναλογίες των διαστάσεών τους. Ταυτόχρονα, εφαρμόζουμε το Distribution Focal Loss (DFL) με αυξημένο βάρος 2.0. Έτσι το μοντέλο μαθαίνει να

χειρίζεται καλύτερα την αβεβαιότητα στα όρια του πλαισίου. Κι επειδή έχουμε μοντέλο μιας μόνο κλάσης, η απώλεια ταξινόμησης πέφτει αρκετά, στο 0.1, με αποτέλεσμα ο βελτιστοποιητής να ασχολείται σχεδόν αποκλειστικά με τη γεωμετρία των πλαισίων.

Παρά τη στιβαρή αρχιτεκτονική του YOLO11s, υπάρχει ένα βασικό πρόβλημα όταν δουλεύουμε με ψηφιοποιημένους παπύρους. Οι πρωτότυπες εικόνες έχουν τεράστια ανάλυση και συχνά πολύ παράξενες διαστάσεις. Συνήθως, για να τα χωρέσουμε στην είσοδο του δικτύου, τα συμπιέζουμε σε μικρότερη ανάλυση. Αυτό, όμως, καταστρέφει σημαντικές χωρικές πληροφορίες. Πολλές φορές, μικροσκοπικοί χαρακτήρες εξαφανίζονται εντελώς, κι έτσι το δίκτυο βγάζει ψευδώς αρνητικά: χάνει πράγματα που υπάρχουν μόνο επειδή έγιναν πολύ μικρά για να τα δει.

Για την οριστική επίλυση αυτού του προβλήματος, η διαδικασία εξαγωγής κειμένου ενισχύθηκε με την ενσωμάτωση του αλγορίθμου Slicing Aided Hyper Inference (SAHI). Ο μηχανισμός αυτός επιτρέπει στο μοντέλο YOLO να σαρώνει την πλήρη, ασυμπίεστη εικόνα του παπύρου μέσω ενός συστήματος επικαλυπτόμενων πλαισιων. Ειδικότερα, η εικόνα τεμαχίζεται δυναμικά σε τμήματα διαστάσεων 640x640 pixels, με μια προκαθορισμένη χωρική επικάλυψη (overlap) της τάξης του 20%. Αν και το μοντέλο YOLO11s εκπαιδεύτηκε σε εικόνες 1024x1024 pixels, η επιλογή της ελαφρώς μικρότερης ανάλυσης τεμαχισμού κατά το inference αποτελεί μια στρατηγική επιλογή βελτιστοποίησης, λειτουργώντας ως μηχανισμός εστιασμένης μεγέθυνσης για τους εξαιρετικά μικρούς χαρακτήρες του παπύρου.

Το παράθυρο 640x640 επιλέχθηκε κυρίως για να αποφύγουμε πολύ μεγάλα τεμάχια εικόνας. Αν διαλέγαμε μεγαλύτερα πλακίδια, όπως 1024x1024 ή 1280x1280 pixels, τα μικροσκοπικά και αχνά γράμματα θα πνίγονταν στον όγκο του δεδομένου, πιάνοντας ελάχιστο χώρο μέσα στον τανυστή. Αντιθέτως, με την αποκοπή στα 640x640, οι τοπικοί χαρακτήρες εμφανίζονται σε πιο μεγάλη κλίμακα στο κάθε κομμάτι. Έτσι, το δίκτυο, που έχει ήδη εκπαιδευτεί να αναγνωρίζει χωρικές λεπτομέρειες σε ανάλυση 1024px, μπορεί να "δει" πιο καθαρά και να εντοπίσει ακριβέστερα τις λεπτές γραμμές του μελανιού, χωρίς να χάνει καμία λεπτομέρεια.

Στον αντίποδα, κρίνεται εξίσου αναγκαία η αποφυγή σημαντικά μικρότερων διαστάσεων τεμαχισμού. Στην περίπτωση που επιλέγονταν πλακίδια μικρότερης κλίμακας, όπως 320x320 pixels, θα αυξανόταν εκθετικά το πλήθος των παραγόμενων τμημάτων ανά εικόνα, επιβαρύνοντας δραματικά τον συνολικό υπολογιστικό χρόνο. Παράλληλα, τα υπερβολικά μικρά πλακίδια αποκόπτουν το αναγκαίο οπτικό πλαίσιο, εισάγοντας τον κίνδυνο μεγαλύτεροι χαρακτήρες, εκτενείς φθορές ή παπυρολογικά διαχωριστικά σύμβολα να τεμαχιστούν βίαια στα όρια και να μην χωράνε εξ ολοκλήρου στο οπτικό πεδίο του δικτύου.

Μετά το πέρας της τοπικής ανίχνευσης, ο αλγόριθμος SAHI αναλαμβάνει τη σύνθετη μαθηματική χαρτογράφηση και συγχώνευση των επιμέρους προβλέψεων. Εφαρμόζοντας τεχνικές Non-Maximum Suppression - NMS στα όρια των πλαισίων, ο αλγόριθμος επιλύει

τις συγκρούσεις και απορρίπτει τις διπλοεγγραφές χαρακτήρων που ανιχνεύθηκαν στις ζώνες επικάλυψης, ανασυνθέτοντας με απόλυτη ακρίβεια το κείμενο της αρχικής εικόνας.

Εικόνα 7. SAHI example

3.3 Υποσύστημα Υπολογιστικής Όρασης: Κατηγοριοποίηση Χαρακτήρων

Αφού εντοπίσουμε και κόψουμε τα πλαίσια από την αρχική εικόνα, περνάμε στο επόμενο βήμα: να κατηγοριοποιήσουμε με ακρίβεια το περιεχόμενό τους. Σε αυτό το στάδιο, η διπλωματική εξετάζει μια συγκριτική μέθοδο, εκπαιδεύοντας και δοκιμάζοντας δύο εντελώς διαφορετικές αρχιτεκτονικές νευρωνικών δικτύων. Από τη μία έχουμε το ResNet-50, ένα κλασικό βαθύ συνελικτικό δίκτυο, και από την άλλη το DeiT (Data-efficient Image Transformer).

3.3.1 Αρχιτεκτονική Συνελικτικού Δικτύου: ResNet-50

Η αρχιτεκτονική Residual Network (ResNet), στην έκδοση των 50 επιπέδων, επιλέχθηκε ως το βασικό μοντέλο αναφοράς λόγω της αποδεδειγμένης στιβαρότητάς της στην εξαγωγή ιεραρχικών χαρακτηριστικών.

Για την προσαρμογή του μοντέλου στις ανάγκες της ελληνικής παπυρολογίας, εφαρμόστηκε η τεχνική της μεταφοράς μάθησης (Transfer Learning). Το μοντέλο microsoft/resnet-50, αρχικά εκπαιδευμένο σε εκατομμύρια εικόνες του ImageNet, φορτώθηκε στο περιβάλλον PyTorch. Μέσω της βιβλιοθήκης Hugging Face, το αρχικό classification head των 1.000 κλάσεων αποκόπηκε δυναμικά χρησιμοποιώντας την παράμετρο ignore_mismatched_sizes=True. Στη θέση της, προσαρτήθηκε ένα νέο fully connected layer προσαρμοσμένο στον ακριβή αριθμό των κλάσεων του αλφαβήτου που εξήχθησαν από το AL-PUB. Με τη διατήρηση των βαρών των πρώτων συνελικτικών επιπέδων, τα οποία εντοπίζουν βασικά γεωμετρικά σχήματα (όπως καμπύλες και γωνίες), το μοντέλο χρειάστηκε να προσαρμόσει μόνο τα βαθύτερα επίπεδά του για να αναγνωρίσει τα πολύπλοκα σχήματα γραμμάτων όπως το μηνοειδές σίγμα ή το ωμέγα. Οι εικόνες τροφοδοτούνται στο δίκτυο αφού πρώτα περάσουν από έναν επεξεργαστή εικόνας (AutoImageProcessor), ο οποίος κανονικοποιεί τις τιμές των pixel και αφαιρεί τις περιττές διαστάσεις.

3.3.2 Αρχιτεκτονική Μετασχηματιστή Όρασης: DeiT (Data-efficient Image Transformers)

Παράλληλα δοκιμάστηκε και η αρχιτεκτονική DeiT. Σε αντίθεση με τα CNNs που χρησιμοποιούν τοπικά συνελικτικά φίλτρα, οι μετασχηματιστές όρασης αντιμετωπίζουν την εικόνα ως μια ακολουθία από επίπεδα τμήματα, εφαρμόζοντας μηχανισμούς αυτο-

προσοχής που επιτρέπουν σε κάθε τμήμα της εικόνας να συσχετιστεί με όλα τα υπόλοιπα, ανεξαρτήτως της φυσικής τους απόστασης.

Το κύριο πρόβλημα των τυπικών μοντέλων ViT είναι η ανάγκη τους για μεγάλο όγκο δεδομένων εκπαίδευσης προκειμένου να κατανοήσουν τη δομή της εικόνας. Το DeiT επιλύει αυτό το πρόβλημα εισάγοντας ένα ειδικό distillation token παράλληλα με το κλασικό class token. Κατά την εκπαίδευση, το DeiT λειτουργεί ως μαθητής (student) που προσπαθεί να ελαχιστοποιήσει την απόκλιση όχι μόνο ως προς τις πραγματικές ετικέτες (hard labels) του AL-PUB, αλλά και ως προς τις προβλέψεις ενός ισχυρού CNN που λειτουργεί ως δάσκαλος (teacher network). Η αρχιτεκτονική αυτή είναι εξαιρετικά πλεονεκτική για την ανάγνωση παπύρων, καθώς ο μηχανισμός global attention μπορεί να εστιάσει ταυτόχρονα στα ίχνη μελανιού και στη δομή των ινών του παπύρου γύρω από αυτό, διαχωρίζοντας την πραγματική γραφή από τις φθορές του υλικού καλύτερα από ένα τυπικό συνελικτικό φίλτρο.

3.4 Μηχανισμός Εντοπισμού Κενών και Φθορών

Στην κλασική παπυρολογία, η απουσία κειμένου λόγω φθοράς είναι εξίσου σημαντική με το σωζόμενο κείμενο, καθώς το μέγεθος ενός χάσματος ορίζει τον χωρικό και σημασιολογικό περιορισμό εντός του οποίου ο φιλόλογος μπορεί να προτείνει αποκαταστάσεις. Προκειμένου το αυτοματοποιημένο pipeline να υπερβεί την απλή εξαγωγή χαρακτήρων και να λειτουργήσει ως παραγωγικό εργαλείο, υλοποιήθηκε ένας εξειδικευμένος αλγοριθμικός μηχανισμός εντοπισμού κενών.

Ο μηχανισμός λειτουργεί ως ενδιάμεσο επίπεδο μεταξύ της εξόδου των μοντέλων οπτικής αναγνώρισης (YOLO/DeiT) και της εισόδου στα Γλωσσικά Μοντέλα. Αρχικά, οι ταξινομημένοι χαρακτήρες οργανώνονται σε γραμμές βάσει των γεωμετρικών τους συντεταγμένων. Στη συνέχεια, το σύστημα υπολογίζει το Μέσο Πλάτος Χαρακτήρα για τη συγκεκριμένη γραμμή ή για το σύνολο του εγγράφου. Αυτή η δυναμική μετρική επιτρέπει στον αλγόριθμο να είναι ανθεκτικός σε διαφορετικές κλίμακες ψηφιοποίησης και μεγέθη γραμματοσειρών. Το σύστημα αναλύει την οριζόντια απόσταση μεταξύ διαδοχικών χαρακτήρων και εισάγει ειδικούς δείκτες με βάση μαθηματικά ρυθμιζόμενα κατώφλια.

Πέρα από τα ενδιάμεσα κενά, ο αλγόριθμος ενσωματώνει και έναν μηχανισμό Margin Detection. Χρησιμοποιώντας στατιστική ανάλυση (το 10ο και 90ο εκατοστημόριο των συντεταγμένων έναρξης και λήξης των γραμμών), το σύστημα αντιλαμβάνεται τα γενικά όρια του εγγράφου. Εάν μια γραμμή ξεκινά πιο δεξιά ή τελειώνει πιο αριστερά από τα υπολογισμένα περιθώρια, το σύστημα συμπεραίνει αυτόματα την απουσία χαρακτήρων στα άκρα και προσθέτει τους αντίστοιχους δείκτες φθοράς στην αρχή ή στο τέλος της γραμμής.

Ο παρακάτω πίνακας περιγράφει τη μαθηματική λογική του αλγορίθμου για την εισαγωγή κενών και δεικτών αβεβαιότητας, τα οποία ακολουθούν το σύστημα Leiden.

Συνθήκη Χωρικής Απόστασης ή Βεβαιότητας	**Ερμηνεία Συστήματος**	**Αποδιδόμενος Δείκτης / Μορφοποίηση (Leiden)**
$Gap \leq 1.2 \times ACW$	Κανονική Συνέχεια Λέξης	Κανένας δείκτης (συνεχής γραφή)
$1.2 \times ACW < Gap \leq 2.5\ x\ AXW$	Όριο Λέξης	Κενό διάστημα
$2.5 \times ACW < Gap \leq 4.0 \times ACW$	Απώλεια Ενός Χαρακτήρα	[·]
$4.0 \times ACW < Gap \leq 8.0 \times ACW$	Απώλεια Πολλαπλών Χαρακτηρων	[⋯] ή εκτίμηση πλήθους π.χ. [~5]
$Gap > 8.0 \times ACW$	Εκτεταμένη Φθορά Υλικού	[---]
Ταξινόμηση με $Confidence < 0.3$	Θόρυβος ή καταστροφή	Παράλειψη χαρακτήρα
Ταξινόμηση με $0.3 \leq Confidence < 0.5$	Αβέβαιη Ανάγνωση	Ερωτηματικό (α?)
$0.3 \leq$ Conf < 0.5 & 2η πρόβλεψη > 0.15	Διφορούμενη Ανάγνωση	Ένδειξη εναλλακτικών (π.χ. [α/δ])

Πίνακας 2. Σύστημα Leiden

Ιδιαίτερη μνεία πρέπει να γίνει στον μηχανισμό διφορούμενης ανάγνωσης: εάν το μοντέλο (DeiT) είναι αβέβαιο για την ταξινόμηση, αλλά η δεύτερη πιο πιθανή επιλογή ξεπερνά το 15% σε βεβαιότητα (π.χ. σύγχυση μεταξύ Γάμμα και Ταυ λόγω φθοράς), το σύστημα διατηρεί και τις δύο εκδοχές (π.χ. [Γ/Τ]).

Μέσω αυτής της διαδικασίας, η ακατέργαστη ροή χαρακτήρων μετασχηματίζεται σε μια δομημένη παπυρολογική ακολουθία (π.χ. ΑΒΓΔΕ[·]ΖΗΘΙΚ[⋯]ΛΜΝΟΠ), η οποία χαρτογραφεί με ακρίβεια την υλική πραγματικότητα του παπύρου στον ψηφιακό χώρο. Αυτή η δομημένη αναπαράσταση αποτελεί τον ακρογωνιαίο λίθο για την επιτυχή εκτέλεση του επόμενου σταδίου, καθώς παρέχει στο Μεγάλο Γλωσσικό Μοντέλο (LLM) τους απαραίτητους χωρικούς και οπτικούς περιορισμούς για την παραγωγή ιστορικά και φιλολογικά λογικών εικασιών αποκατάστασης.

3.5 Το Σύστημα RAG και η Διανυσματική Βάση Δεδομένων

Αν και τα σύγχρονα Μεγάλα Γλωσσικά Μοντέλα (LLMs) διαθέτουν εξαιρετική ικανότητα στην παραγωγή φυσικής γλώσσας, υστερούν σημαντικά όταν καλούνται να λειτουργήσουν σε αυστηρά εξειδικευμένους τομείς, συχνά κατασκευάζοντας μη πραγματικά δεδομένα. Για τη θωράκιση της παρούσας αρχιτεκτονικής έναντι αυτού του κινδύνου, σχεδιάστηκε ένα σύστημα Ανάκτησης Επαυξημένης Παραγωγής (RAG), το οποίο γεφυρώνει τις παραγωγικές ικανότητες του LLM με μια επαληθευμένη ιστορική βάση γνώσης.

3.5.1 Υλοποίηση Υποδομής με PostgreSQL και pgvector

Ως backend του συστήματος RAG επιλέχθηκε η πλατφόρμα Supabase, η οποία αξιοποιεί μια σχεσιακή βάση δεδομένων PostgreSQL, επαυξημένη με την επέκταση pgvector. Η αρχιτεκτονική του σχήματος ορίζεται μέσω του αρχείου setup_supabase_schema.sql και επικεντρώνεται στον πίνακα papyri_transcriptions. Σε αυτόν τον πίνακα, κάθε ιστορικό έγγραφο αποθηκεύεται ως μια εγγραφή που περιλαμβάνει: το μοναδικό αναγνωριστικό (id), τον τίτλο, το κείμενο με μορφοποίηση Leiden+ και, το πιο σημαντικό, την αριθμητική διανυσματική του αναπαράσταση (embedding).

3.5.2 Μοντέλα Διανυσματικής Ενσωμάτωσης (Embeddings)

Η μετατροπή των αρχαίων κειμένων σε πολυδιάστατα διανύσματα επιτρέπει την υπολογιστική σύγκριση της σημασιολογικής τους εγγύτητας. Το pipeline ενσωματώνει και υποστηρίζει δύο εναλλακτικά μοντέλα ενσωμάτωσης, το καθένα με συγκεκριμένα πλεονεκτήματα:

- **Μοντέλο Τοπικής Εκτέλεσης (Local Model - all-MiniLM-L6-v2):** Το μοντέλο αυτό μετατρέπει το κείμενο σε διανύσματα 384 διαστάσεων. Το κύριο πλεονέκτημά του είναι η ταχύτητα και η ανεξαρτησία, καθώς εκτελείται εξ ολοκλήρου τοπικά στους πόρους της CPU. Χρησιμοποιήθηκε κυρίως κατά τη φάση της ανάπτυξης και της δοκιμής της διανυσματικής υποδομής.
- **Μοντέλο Gemini (text-embedding-004):** Αυτή η προηγμένη προσέγγιση αντιστοιχεί το κείμενο σε έναν ευρύτερο διανυσματικό χώρο 768 διαστάσεων, προσφέροντας σημαντικά υψηλότερη ποιότητα και λεπτομέρεια στη σημασιολογική αναπαράσταση. Λόγω της πυκνότερης πληροφορίας που φέρει, απαιτεί μεγαλύτερο αποθηκευτικό χώρο στο Supabase. Είναι απαραίτητο το ίδιο μοντέλο να χρησιμοποιηθεί τόσο κατά την εισαγωγή όσο και κατά την αναζήτηση προκειμένου τα διανύσματα να ανήκουν στον ίδιο μαθηματικό χώρο.

3.5.3 Αλγόριθμοι Σημασιολογικής Αναζήτησης

Κατά τη διάρκεια μιας επερώτησης (π.χ. όταν το σύστημα αναζητά έγγραφα που ταιριάζουν στο φθαρμένο απόσπασμα "παρακαλω σε αδελφε"), το κείμενο μετατρέπεται σε διάνυσμα και συγκρίνεται με όλα τα έγγραφα της βάσης υπολογίζοντας την Ομοιότητα

Συνημιτόνου. Η εξίσωση της ομοιότητας μετρά τη γωνία μεταξύ των δύο διανυσμάτων, εξασφαλίζοντας ότι το μέγεθος του εγγράφου δεν θα αλλοιώσει τα αποτελέσματα. Προκειμένου το LLM να τροφοδοτηθεί μόνο με δεδομένα υψηλής αξιοπιστίας, ορίστηκε ένα αυστηρό όριο φιλτραρίσματος (threshold) στο 0.7. Η επιλογή της συγκεκριμένης υπερπαραμέτρου προέκυψε κατόπιν εμπειρικής αξιολόγησης, με στόχο τη βέλτιστη εξισορρόπηση μεταξύ Precision και Recall του συστήματος ανάκτησης:

- **Κίνδυνος Χαμηλού Κατωφλίου (π.χ. < 0.5)**: Εάν το όριο οριζόταν πολύ χαμηλά, το σύστημα θα παρουσίαζε υψηλό Recall αλλά κακό Precision. Θα ανακτούσε έγγραφα με εξαιρετικά χαλαρή ή και ανύπαρκτη σημασιολογική συγγένεια με το ερώτημα (false positives). Η τροφοδότηση τέτοιου άσχετου "θορύβου" στο LLM αυξάνει εκθετικά τον κίνδυνο παραγωγής "παραισθήσεων", ακυρώνοντας τον πρωταρχικό σκοπό της αρχιτεκτονικής RAG.

- **Κίνδυνος Υπερβολικά Υψηλού Κατωφλίου (π.χ. > 0.85)**: Στον αντίποδα, ένα εξαιρετικά αυστηρό όριο θα απαιτούσε σχεδόν απόλυτη, λεκτική ταύτιση μεταξύ του ερωτήματος και της βάσης δεδομένων. Δεδομένης της φύσης του προβλήματος της παπυρολογίας, ένα τέτοιο κατώφλι θα ήταν απαγορευτικό. Θα απέρριπτε πολύτιμα ιστορικά παράλληλα μόνο και μόνο εξαιτίας μερικών ανορθόγραφων χαρακτήρων, οδηγώντας σε αποτυχία ανάκτησης (false negatives).

Συνεπώς, το όριο του 0.7 λειτουργεί ως η ιδανική "χρυσή τομή" για τον συγκεκριμένο διανυσματικό χώρο. Εξασφαλίζει ότι τα ανακτημένα κείμενα ανήκουν στο ίδιο αυστηρό σημασιολογικό και ιστορικό πλαίσιο με τον προς εξέταση πάπυρο, ενώ ταυτόχρονα προσφέρει στο σύστημα την απαραίτητη "ανοχή" για να προσπερνά τα μορφολογικά λάθη του ακατέργαστου κειμένου εισόδου. Μόνο τα έγγραφα που υπερβαίνουν αυτό το όριο ομοιότητας ανακτώνται τελικά και ενσωματώνονται στο prompt του Γλωσσικού Μοντέλου.

id	transcription	embedding
p.oxy.43.3105	15.[ἐπέδωκεν γραφῆς λειτο]υργῶν τῷ ἰσιόντι(*) 16.[- ca.? -][(ἔτει)][- ca.? -][τῇ δι]ελθούσῃ ἡμέρᾳ ἀνέ-	[-0.008235425,0.03207301,-0.010013755,-0.026928369,0.036284227,-0.00
p.oxy.43.3139	8.περὶ μισενης α[πασαιλως ενγραφου η 9.ἀγράφου πράγματ[ος τὸ σ]ύνολον τῶν τῇ 10.αὐτῇ συμβάσει [- ca.? -][- ca.? -][- ca.? -]ν συντελούντων	[-0.0072825537,0.04201165,-0.0002142866,-0.017964084,0.03895175,-0.0
p.oxy.43.3115	1.π(αρὰ) τοῦ στρ(ατηγοῦ) ἐπιμελητοῖς κριθῆς. 2.δότε Παλιουρνοῖς ὑπὲρ ἡμερῶν	[-0.013681304,0.05632033,0.00044667785,-0.032826796,0.018064274,0.01
p.oxy.43.3101	12.[ἀναγράφ]εσθαι ἐπ’ ἀμφόδ(ου) Κρη- 13.[τικοῦ καὶ] ἐπικεκρίσθαι τῇ πρὸς	[-0.013618451,0.03613506,-0.0104595255,-0.020586396,0.03633425,-0.00
p.oxy.43.3129	4.ἔχειν χρεώστας ὁμολογο[υ]- 5.μένους, ὡς γνώσει ἐκ τοῦ.	[-0.03468638,0.043964747,-0.005492106,0.018693402,0.030044597,-0.000
p.oxy.43.3128	1.[ὑπατεί]ας Παπίου Πακατιανοῦ τ[οῦ λαμπροτάτου ἐπάρχου τοῦ ἱεροῦ πραιτ[ω 2.[τοῦ](προτάτου) μην[ὶ] Ἐπ[εὶ]φ ε [- ca.? -]	[-0.013317675,0.0464702,-0.015450603,-0.016058665,0.034277298,0.0110

Εικόνα 8. Vector Database

3.5.4 Γλωσσική Εναρμόνιση και Προεπεξεργασία Κειμένου (Text Normalization)

Κατά τη φάση των πειραματικών δοκιμών του συστήματος RAG, παρατηρήθηκε μια κρίσιμη αδυναμία η οποία, αν και δεν αποτελούσε τεχνικό σφάλμα της διανυσματικής βάσης, υποβάθμιζε δραματικά τη διαδικασία της σημασιολογικής αναζήτησης. Τα σύγχρονα μοντέλα ενσωμάτωσης εκπαιδεύονται ώστε να κωδικοποιούν τη σημασιολογική έννοια των λέξεων και των εννοιών, και όχι τη στενή, κυριολεκτική αλληλουχία των χαρακτήρων.

Το πρόβλημα εντοπίστηκε στη ριζική μορφολογική διαφορά μεταξύ της «Αλήθειας Αναφοράς» και της εξόδου του μοντέλου οπτικής αναγνώρισης. Το ιστορικό κείμενο που βρίσκεται αποθηκευμένο στη βάση δεδομένων είναι "καθαρό", κανονικοποιημένο, με πεζά γράμματα, τόνους, πνεύματα και σημεία στίξης (π.χ. *σωι. ἀσπάζεται ὑμ[ᾶς]*). Στον αντίποδα, το ακατέργαστο κείμενο που εξάγεται από το σύστημα Computer Vision είναι μια συνεχής ακολουθία κεφαλαίων γραμμάτων, στερείται κενών διαστημάτων, κάνει εκτενή χρήση του μηνοειδούς σίγμα (lunate sigma, 'C' αντί για 'Σ') και περιέχει πλήθος αγκυλών που υποδηλώνουν παπυρολογικές φθορές (π.χ. *CΩΙCCΠΑΖΘΤΑΙΙΜ*).

Λόγω αυτής της μορφολογικής απόκλισης, ο αλγόριθμος ενσωμάτωσης αντιλαμβανόταν το ακατέργαστο αποτέλεσμα του OCR ως "γλωσσικό θόρυβο", εντελώς ασύνδετο με το καθαρό λεξιλόγιο της βάσης. Κατά συνέπεια, η μαθηματική ομοιότητα κατέρρεε σε τιμές πολύ κάτω από το ορισμένο κατώφλι, οδηγώντας σε αποτυχία ανάκτησης ακόμα και όταν το αντίστοιχο έγγραφο υπήρχε στη βάση.

Για την επίλυση αυτής της γλωσσικής ασυμβατότητας, αναπτύχθηκε και ενσωματώθηκε στη ροή ένας ειδικός μηχανισμός προεπεξεργασίας αποκλειστικά για την αναζήτηση RAG, μέσω της συνάρτησης preprocess_for_rag(). Αυτή η διαδικασία παρεμβάλλεται ακριβώς πριν το κείμενο μετατραπεί σε διάνυσμα και εφαρμόζει τρεις βασικές γλωσσικές κανονικοποιήσεις:

1. **<u>Μετατροπή σε πεζά (Lowercasing):</u>**
   Ολόκληρη η συμβολοσειρά μετατρέπεται από κεφαλαία σε πεζά γράμματα, μειώνοντας την πολυπλοκότητα του λεξιλογίου.
2. **<u>Αντικατάσταση Χαρακτήρων (Character Substitution):</u>**
   Το μηνοειδές σίγμα ('C' ή 'c'), το οποίο συχνά προκαλεί σύγχυση στα μοντέλα λόγω της οπτικής του ταύτισης με το λατινικό 'C', αντικαθίσταται μαζικά με το τυπικό ελληνικό σίγμα ('σ').
3. **<u>Αφαίρεση Θορύβου και Συμβόλων Φθοράς (Noise Stripping):</u>**
   Διαγράφονται πλήρως όλες οι παπυρολογικές σημάνσεις, οι αγκύλες και τα σύμβολα κενού (π.χ. [---], [·], [X/Y]), τα οποία κατακερματίζουν τεχνητά τις λέξεις και καταστρέφουν τη σημασιολογική τους συνέχεια.

Μέσω αυτής, το παραγόμενο κείμενο μετασχηματίζεται σε μια μορφή (π.χ. *σωισσπαζθταιιμ*) η οποία προσομοιάζει σε τεράστιο βαθμό το λεξιλόγιο του κειμένου αναφοράς. Η διανυσματική του αναπαράσταση τοποθετείται πλέον στον ορθό

σημασιολογικό χώρο, βελτιώνοντας κατακόρυφα τα ποσοστά επιτυχούς ανάκτησης των ιστορικών παραλλήλων. Τονίζεται, τέλος, ότι αυτή η αλλοίωση εφαρμόζεται αυστηρά και μόνο για την παραγωγή του διανύσματος αναζήτησης. Το Μεγάλο Γλωσσικό Μοντέλο εξακολουθεί να τροφοδοτείται με την αρχική, ακατέργαστη έξοδο του OCR προκειμένου να έχει πλήρη εικόνα των φθορών και των κενών που καλείται να αποκαταστήσει.

3.5.5 Υβριδική Στρατηγική Ανάκτησης

Παρά την κανονικοποίηση του κειμένου που προηγείται της αναζήτησης, η πρακτική αξιολόγηση του RAG αποκάλυψε έναν θεμελιώδη δομικό περιορισμό των πυκνών διανυσματικών αναπαραστάσεων όταν έρχονται αντιμέτωπες με ακατέργαστες εξόδους OCR της συνεχομένης αρχαίας γραφής.

Τα σύγχρονα μοντέλα ενσωμάτωσης βασίζονται στο tokenization. Ο αλγόριθμος αναμένει δομημένες λέξεις προκειμένου να τις χαρτογραφήσει σε σημασιολογικές έννοιες εντός του διανυσματικού χώρου. Όταν το σύστημα αναζήτησης τροφοδοτείται με μια αδιάκοπη, ανορθόγραφη ακολουθία γραμμάτων (π.χ. «μελησησεκτιν»), το μοντέλο αδυνατεί να την τεμαχίσει σωστά. Αντιλαμβάνεται τη συμβολοσειρά ως θόρυβο, με αποτέλεσμα η μαθηματική ομοιότητα με το καθαρό κείμενο-στόχο (π.χ. «[ἀ]μελήσῃς ἐκτιν») να κατακρημνίζεται, λαμβάνοντας ακόμη και αρνητικές τιμές, ανεξαρτήτως του επιλεγμένου κατωφλίου.

Για την αντιμετώπιση αυτής της κατάρρευσης, χωρίς ωστόσο να εγκαταλειφθεί η πολύτιμη υποδομή της βάσης δεδομένων Supabase, η παρούσα διπλωματική υιοθέτησε μια Υβριδική Στρατηγική Ανάκτησης, διαχωρίζοντας τις ροές αναζήτησης ανάλογα με τη φύση του ερωτήματος:

1. **Σημασιολογική Αναζήτηση μέσω Διανυσμάτων (Semantic Vector Search):**

   Η διανυσματική προσέγγιση διατηρείται ακέραιη ως η βέλτιστη μέθοδος για εννοιολογικές ερωτήσεις. Όταν ο χρήστης ή το LLM επιχειρεί να αναζητήσει έγγραφα με βάση το νόημα ή τα μεταδεδομένα (π.χ. "Βρες φορολογικές αποδείξεις από την Οξύρρυγχο που αφορούν πωλήσεις σίτου"), τα διανύσματα υπερέχουν συντριπτικά, καθώς μπορούν να συσχετίσουν λέξεις-κλειδιά με συνώνυμα και σημασιολογικά πεδία που ένα απλό σύστημα αναζήτησης κειμένου θα αγνοούσε.

2. **Τοπική Αντιστοίχιση N-Γραμμάτων σε Επίπεδο Χαρακτήρα (Character-Level N-Gram Matching):**

   Όταν ο στόχος είναι η ταύτιση ενός ακατέργαστου, φθαρμένου αποσπάσματος OCR με ένα γνωστό αρχαίο κείμενο, το RAG μεταβαίνει σε μια προσέγγιση παραδοσιακής παπυρολογικής αναζήτησης. Υλοποιήθηκε ένας αλγόριθμος local n-gram search, [54] ο οποίος:

- Κατεβάζει in-memory τον κατάλογο των διαθέσιμων μεταγραφών από το Supabase.
- Συγκρίνει τις αλληλουχίες γραμμάτων μεταξύ της εξόδου του OCR και των κειμένων της βάσης.
- Αγνοεί πλήρως τα κενά, τους τόνους, τα πνεύματα και τις αγκύλες φθοράς.

Η μέθοδος αυτή συγκρίνει raw αρχαία ελληνικά γράμματα απευθείας με αρχαία ελληνικά γράμματα. Το έγγραφο με τον υψηλότερο δείκτη αλληλεπικάλυψης χαρακτήρων κατατάσσεται πρώτο και ενσωματώνεται στο prompt του LLM. Αυτή η υβριδική αρχιτεκτονική εξασφαλίζει ότι το σύστημα διαθέτει τόσο την «επεξεργαστική ευφυΐα» να κατανοεί ιστορικές έννοιες, όσο και τη «μηχανική ακρίβεια» να εντοπίζει σπασμένα θραύσματα κειμένου στον τεράστιο όγκο δεδομένων, επιτυγχάνοντας τα βέλτιστα αποτελέσματα για την τελική αποκατάσταση του παπύρου.

3.6 Διασύνδεση Pipeline και Γλωσσικό Μοντέλο

Το τελικό στάδιο του συστήματος αφορά τη σύνδεση των προηγούμενων υποσυστημάτων σε ένα end-to-end pipeline. Το pipeline ενορχηστρώνει μια ακολουθία τριών διακριτών βημάτων, όπου χρησιμοποιούνται τεχνικές Prompt Engineering. Σε κάθε βήμα, το μοντέλο καλείται να ενσαρκώσει μια συγκεκριμένη ακαδημαϊκή περσόνα:

1. **Φιλολογική Αποκατάσταση (Textual Restoration)**: Η ακατέργαστη συμβολοσειρά, εμπλουτισμένη πλέον με τη σημασιολογική γνώση που ανακτήθηκε μέσω του RAG, προωθείται σε μια νέα, αμιγώς κειμενική διεργασία. Το μοντέλο υιοθετεί τώρα τον ρόλο του "ειδικού στην αποκατάσταση παπύρων". Η διεργασία εστιάζει στην αντικατάσταση των ετικετών [UNK] ή των συμβόλων [···], λαμβάνοντας υπόψη τη γραμματική, το συντακτικό και το μορφολογικό περιβάλλον του αρχαίου κειμένου. Για την αποτροπή των παραισθήσεων, το σύστημα απαιτεί από το μοντέλο, σε περίπτωση αμφιβολίας, να περιβάλλει τη λέξη με την ειδική ετικέτα ``, διατηρώντας έτσι την επιστημονική διαφάνεια για τον τελικό αναγνώστη.
2. **Τελική Μετάφραση (Translation):** Στο τελευταίο βήμα, το πλήρως αποκατεστημένο αρχαιοελληνικό κείμενο μεταφράζεται στη γλώσσα στόχο (Αγγλικά). Ο ρόλος του μοντέλου μεταπίπτει σε αυτόν του "έμπειρου μεταφραστή", με εντολή τη διατήρηση του εννοιολογικού πλαισίου και τη χρήση σύγχρονου λεξιλογίου, ενώ τα σημεία φιλολογικής αβεβαιότητας μεταφέρονται ως ``. Η τελική έξοδος, που συμπεριλαμβάνει το αρχικό αρχείο, την ακατέργαστη μεταγραφή, την αποκατάσταση και τη μετάφραση, δύναται να αποθηκευτεί για περαιτέρω ακαδημαϊκή μελέτη.

Προκειμένου να καταστεί εφικτή η ποσοτική αξιολόγηση του συστήματος μέσω των μετρικών CER, WER και BLEU, ήταν απολύτως κρίσιμο οι έξοδοι των Μεγάλων Γλωσσικών Μοντέλων (LLMs) να είναι δομικά αυστηρές και απαλλαγμένες από τον «θόρυβο» (π.χ. φράσεις όπως *"Ακολουθεί το αποκατεστημένο κείμενο"*).

Για την επίλυση αυτού του προβλήματος, αναπτύχθηκε μια στρατηγική Prompt Engineering, η οποία καθοδηγεί το LLM μέσω αυστηρών περιορισμών και μεθοδολογιών. Η στρατηγική αυτή εδράζεται στους ακόλουθους πυλώνες

Αρχικά ,το ακατέργαστο κείμενο που παράγεται από τα μοντέλα Υπολογιστικής Όρασης (YOLO/DeiT) είναι γεμάτο ειδικούς δείκτες. Αν το LLM δεν κατανοήσει αυτούς τους δείκτες, θα τους αντιμετωπίσει ως τυπογραφικά λάθη. Στο αρχικό στάδιο του prompt, ενσωματώθηκε ένα αναλυτικό «λεξικό κανόνων». Το μοντέλο εκπαιδεύεται επί τόπου στο τι σημαίνουν τα παραγόμενα κενά ([·], [···], [---]), τι σημαίνει η υποστιγμή αβεβαιότητας (α̣) και πώς να διαχειριστεί διφορούμενες οπτικές αναγνώσεις του τύπου [X/Y]. Παράλληλα, του υπενθυμίζονται οι κλασικές εκδοτικές συμβάσεις Leiden (π.χ. [abc] για δική του συμπλήρωση, (*) για διόρθωση σφάλματος του γραφέα ή του OCR).

Επιπλέον ένα συχνό φαινόμενο στα παραγωγικά μοντέλα είναι η τάση τους να αγνοούν πλήρως τα ακατάληπτα γράμματα του OCR και να επινοούν εντελώς νέες, άσχετες λέξεις που απλώς ταιριάζουν γραμματικά. Για να αποτραπεί αυτό, το prompt περιλαμβάνει ειδική οδηγία («Fix OCR errors safely»): επιτρέπεται στο μοντέλο να διορθώσει μια λέξη *μόνο* εάν η διόρθωση αφορά χαρακτηριστικά οπτικά σφάλματα του αλγορίθμου (όπως η σύγχυση παρόμοιων γραμμάτων Η/Ν, Δ/Α, Θ/Ε). Οποιαδήποτε ουσιώδης αλλοίωση υψηλής βεβαιότητας χαρακτήρων απαιτείται να επισημαίνεται ρητά με το σύμβολο (*), διασφαλίζοντας την ιχνηλασιμότητα της επέμβασης.

Το τελικό και σημαντικότερο βήμα για τη διευκόλυνση της αυτοματοποιημένης αξιολόγησης είναι ο ρητός τερματικός περιορισμός: *"Output only the restored text, preserving line structure"*. Η οδηγία αυτή απαγορεύει στο μοντέλο την παραγωγή προλόγων, επιλόγων ή επεξηγηματικών σχολίων. Το παραγόμενο αποτέλεσμα είναι μια καθαρή συμβολοσειρά κειμένου, απολύτως έτοιμη να τροφοδοτήσει τη συνάρτηση normalize_greek() και κατ' επέκταση τους αλγορίθμους Levenshtein (CER/WER) και τα N-Grams, χωρίς κίνδυνο στατιστικών αποκλίσεων.

Μέσω αυτής της στοχευμένης αρχιτεκτονικής του Prompt Engineering, το Γλωσσικό Μοντέλο δεν λειτουργεί απλώς ως γεννήτρια κειμένου, αλλά συμπεριφέρεται ως ένας ντετερμινιστικός μεταφραστής κανόνων, ο οποίος διαχειρίζεται τα ιστορικά και παπυρολογικά δεδομένα εντός ενός αυστηρά ελεγχόμενου πλαισίου μορφοποίησης.

Προς πληρέστερη τεκμηρίωση της αρχιτεκτονικής του συστήματος, παρατίθενται αυτούσιες οι βασικές υποδείξεις που τροφοδοτούνται στο Γλωσσικό Μοντέλο κατά την εκτέλεση του pipeline. Τα prompts είναι γραμμένα στην αγγλική γλώσσα, καθώς τα σύγχρονα LLMs παρουσιάζουν βέλτιστη συμμόρφωση όταν καθοδηγούνται στα αγγλικά, ακόμη και όταν το κείμενο επεξεργασίας είναι αρχαιοελληνικό.

Το βασικό prompt αποκατάστασης λειτουργεί ως ένα δυναμικό πρότυπο. Ενσωματώνει αυτόματα το ακατέργαστο κείμενο του OCR ({transcription}) και, εφόσον είναι ενεργοποιημένη η αρχιτεκτονική RAG, προσθέτει δυναμικά τη σχετική ιστορική γνώση μέσω της μεταβλητής {rag_section}.

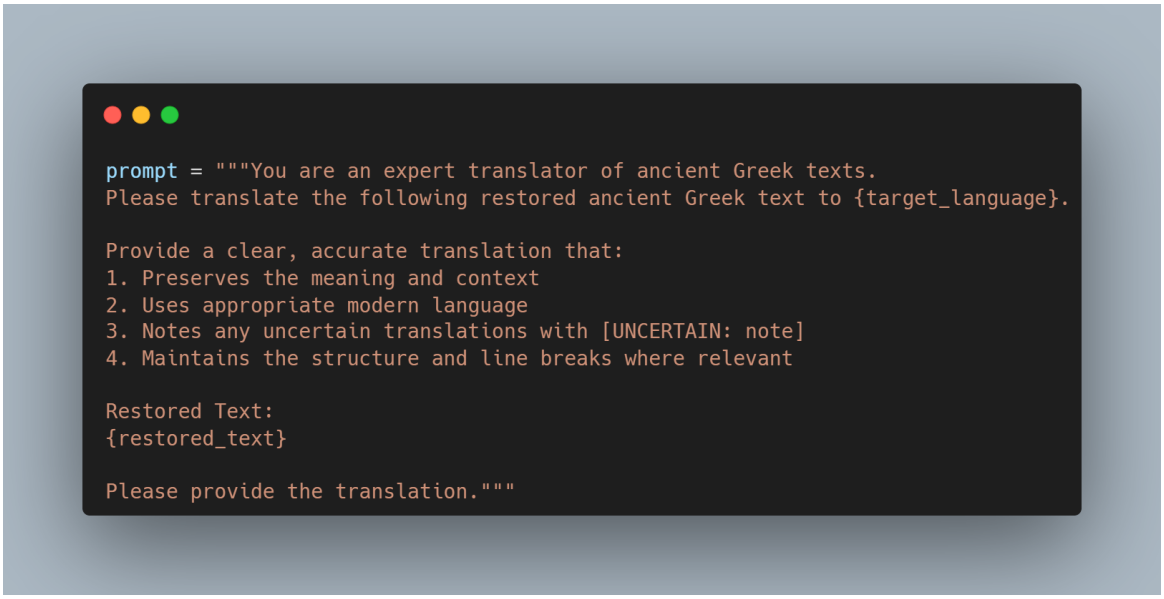
(System Prompt):

```
prompt = """You are an expert in ancient Greek papyrology and text restoration.

The transcription below contains editorial markers following papyrological conventions. You must
understand these markers to properly restore the text.

## EDITORIAL CONVENTIONS

### Gap Detection Markers (from OCR/automated detection):
- `[·]` = Single missing character (small gap detected)
- `[··]` = Two missing characters
- `[···]` = Three missing characters
- `[~N]` = Approximately N missing characters (e.g., `[~5]` means about 5 characters)
- `[---]` = Damaged region with unknown extent (large gap)
- `α̣` = Uncertain character with underdot (low OCR confidence, character visible but unclear)
- `[X/Y]` = Alternative OCR readings (e.g. `[H/N]` means the character is either H or N based on visual
similarity. Pick one.)

### Leiden+ Conventions (from scholarly editions):
- `[abc]` = Supplied/restored text (editor's restoration of missing text)
- `(abc)` = Abbreviation expansion (e.g., `κ(αὶ)` = "και" abbreviated as "κ")
- `(*)` = Correction marker (indicates the scribe made an error, corrected version shown)
- `{{abc}}` = Superfluous text (should be deleted)
- `⟦abc⟧` = Erased text (deliberately deleted by ancient scribe)
- Line numbers at start of lines (e.g., `1`, `2`, `3`)
- Hyphens `-` at end of line indicate word continues on next line

{rag_section}

## YOUR TASK

Given the transcription below, please:

1. **Restore missing characters**: Replace gap markers (`[·]`, `[···]`, `[---]`, etc.) with your best
guess of the missing text
2. **Confirm or correct uncertain characters**: Characters with underdots (`α̣`) should be verified or
corrected
3. **Resolve alternative readings**: For `[X/Y]`, the OCR is unsure whether the letter is X or Y. Pick
the correct one.
4. **Convert to standard Greek**: The input transcription is in majuscule/uncial script (all letters
capitalized) without spaces, accents, or breathings. You MUST convert all text to standard lower-case
Greek, add appropriate accents and breathings, and separate it into proper Greek words.
5. **Fix OCR errors safely**: We want to avoid hallucinations, so do not invent entirely new words that
share no letters with the transcription. However, OCR often misreads similar-looking letters (like H/N,
Δ/Α, Θ/Ε, Γ/C, etc). If a string of letters (like `ΓΠΡΑΦΕΜΡ`) does not make sense, but changing 1-2
letters forms a valid Greek phrase (like `γράφει μοι`), you SHOULD make those corrections.
6. **Annotate major changes**: If you have to change a high-confidence character (not marked with `̣` or
`[X/Y]`) to fix an OCR error as described above, please use the `(*)` marker immediately after the
corrected word so we know you corrected the OCR.
7. **Respect editorial markers**: Keep abbreviation expansions `()` and correction markers `(*)` as they
provide valuable context
8. **Use context**: Consider:
   - Common ancient Greek vocabulary and grammar
   - Document type (letter, contract, literary text, etc.)
   - Adjacent words and phrases
   - Similar papyri from the reference examples above (if provided)
9. **Indicate uncertainty**: If you're unsure about a restoration, mark it as `[abc?]` where `?` shows
uncertainty
10. **Preserve structure**: Maintain line numbers and line breaks

## EXAMPLE

Input:
```
1Διογενὶς [··] τῶι τιμιωτάτῳ
2χαίρ[·]ιν πρὸ παντ[---]
3εὔχομαί σε ὑγιαίνειν
```

Output:
```
1Διογενὶς [Διδύμῳ?] τῶι τιμιωτάτῳ
2χαίρ[ε]ιν πρὸ παντ[ὸς εὔχομαί σοι?]
3εὔχομαί σε ὑγιαίνειν
```

Note: `[Διδύμῳ?]` shows uncertainty, line 3's `ὑγιαίνειν` is corrected to `ὑγιαίνειν`.

## INPUT TRANSCRIPTION

```
{transcription}
```

OUTPUT
Provide the restored text with your restorations. Use `[text?]` for uncertain restorations. Output only
the restored text, preserving line structure."""
```

Εικόνα 9. prompt αποκατάστασης

Μετά την επιτυχή ολοκλήρωση της αποκατάστασης, το αποτέλεσμα ({restored_text}) προωθείται στο δεύτερο στάδιο του pipeline. Εδώ, το prompt είναι πιο λιτό, εστιάζοντας αποκλειστικά στη σημασιολογική μεταφορά του αρχαίου κειμένου στη επιθυμητή σύγχρονη γλώσσα ({target_language}, συνήθως Αγγλικά), διατηρώντας το ύφος και τη δομή του πρωτοτύπου:

(System Prompt):



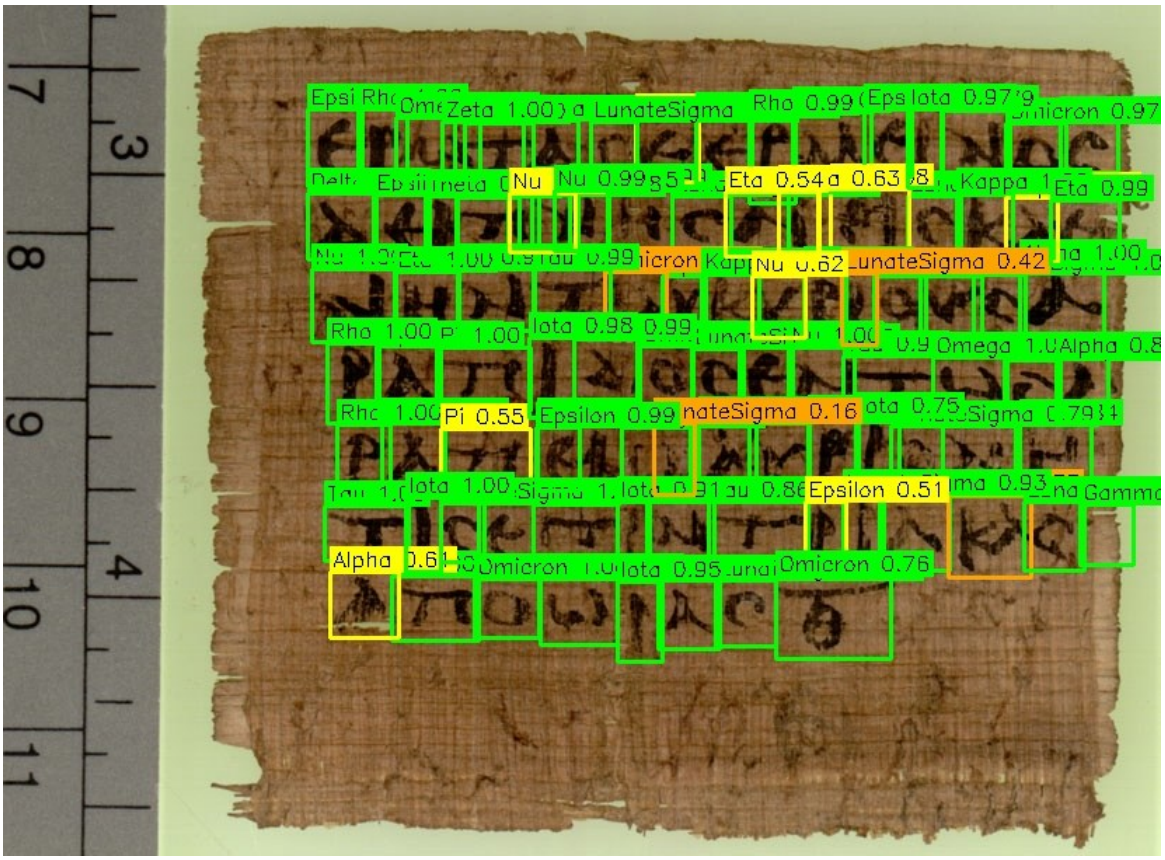
```
prompt = """You are an expert translator of ancient Greek texts.
Please translate the following restored ancient Greek text to {target_language}.

Provide a clear, accurate translation that:
1. Preserves the meaning and context
2. Uses appropriate modern language
3. Notes any uncertain translations with [UNCERTAIN: note]
4. Maintains the structure and line breaks where relevant

Restored Text:
{restored_text}

Please provide the translation."""
```

Εικόνα 10. prompt μετάφρασης

3.7 Παράδειγμα Εκτέλεσης Ροής

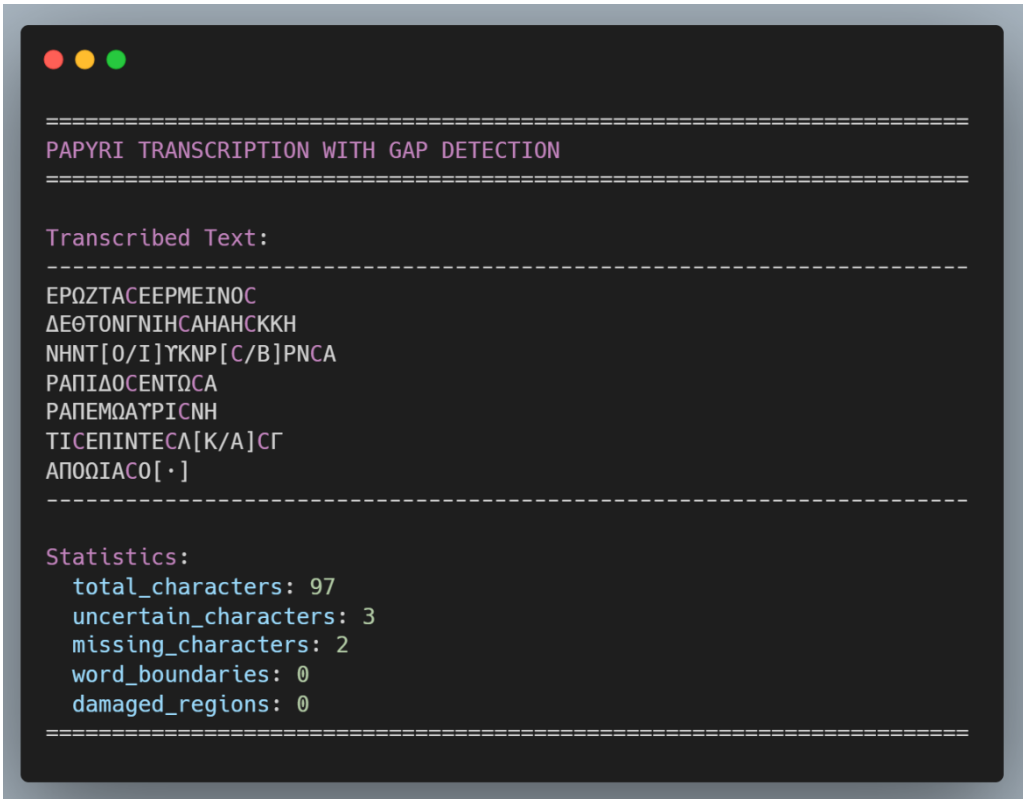
Για την αξιολόγηση του pipeline σε πραγματικές συνθήκες εκτελέστηκε ολόκληρο το σύστημα πάνω σε μια σάρωση υψηλής ανάλυσης του παπύρου P.Oxy. 52 3693. [55] Πρόκειται για μια τυπική ιδιωτική πρόσκληση σε ιεροτελεστικό δείπνο προς τιμήν του θεού Σαράπιδος, η οποία χρονολογείται στη ρωμαϊκή περίοδο. Κατά το πρώτο στάδιο της Υπολογιστικής Όρασης, ο αλγόριθμος YOLO11s (με ενσωμάτωση SAHI) εντόπισε τους χαρακτήρες και το δίκτυο DeiT ανέλαβε την ταξινόμησή τους.

```
1. python scripts/infer_yolo_deit.py
2. --image images\p.oxy.52.3693.jpg
3. --yolo detection_only_m\weights\best.pt
4. --deit alpub_greek_specialist_deit/alpub_deit_model/
5. --output transcription.txt
6. --visualize
7.  --use  sahi
8.
```



Εικόνα 11. Ανίχνευση χαρακτήρων στον πάπυρο P.Oxy. 52 3693

Η εκτέλεση του σχετικού σεναρίου παρήγαγε μια ακατέργαστη μεταγραφή 97 χαρακτήρων η οποία εμφάνισε έντονο οπτικό θόρυβο, με χαρακτηριστικότερο παράδειγμα τη δεύτερη γραμμή, όπου η φράση "δειπνῆσαι εἰς κλίνην" αποδόθηκε με την αλληλουχία "ΔΕΘΤΟΝΓΝΙΗCΑΗΑΗCΚΚΗ". Επιπλέον, το σύστημα όρασης κατέγραψε 3 χαρακτήρες με χαμηλή βεβαιότητα και 2 ως αγνοούμενους, αντανακλώντας τη φθορά του υποστρώματος και τις προκλήσεις της συνεχούς γραφής.



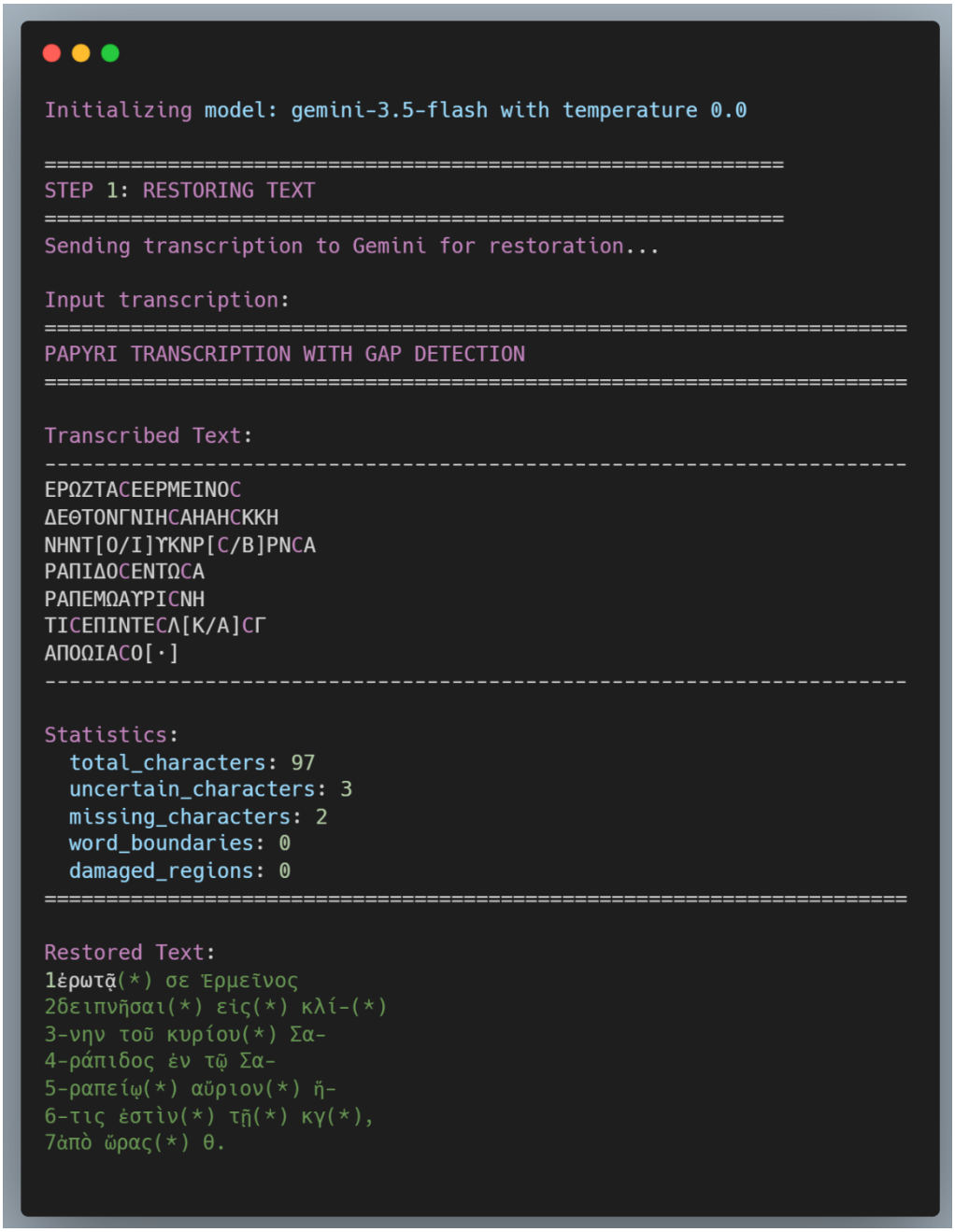
Εικόνα 12. Ανίχνευση κενών στον πάπυρο P.Oxy. 52 3693

Ακολούθως, η ακατέργαστη αυτή μεταγραφή διοχετεύθηκε στο γλωσσικό μοντέλο gemini-3.5-flash, με την παράμετρο της θερμοκρασίας να διατηρείται αυστηρά στο μηδέν (ντετερμινιστική συμπεριφορά έναντι παραισθήσεων) και χωρίς την ενεργοποίηση του υποσυστήματος RAG.

```
python scripts/experiment  transcription  to  text.py
--transcription_file C:\Users\themi\Desktop\ancient-yolo\transcription.txt\p.oxy.52.3693_detailed.txt
--output  file restored.txt
```

Το γλωσσικό μοντέλο επέδειξε εντυπωσιακή ικανότητα σημασιολογικής συλλογιστικής. Λαμβάνοντας τα ακατάληπτα θραύσματα της μεταγραφής, κατανόησε άμεσα το πλαίσιο του εγγράφου ως πρόσκληση σε συμπόσιο και αποκατέστησε με επιτυχία σχεδόν ολόκληρο το κείμενο, μετατρέποντάς το σε κανονική, τονισμένη πεζή ελληνική γραφή. Το μοντέλο επενέβη αποφασιστικά για να εξομαλύνει τα οπτικά σφάλματα του OCR (όπως τη

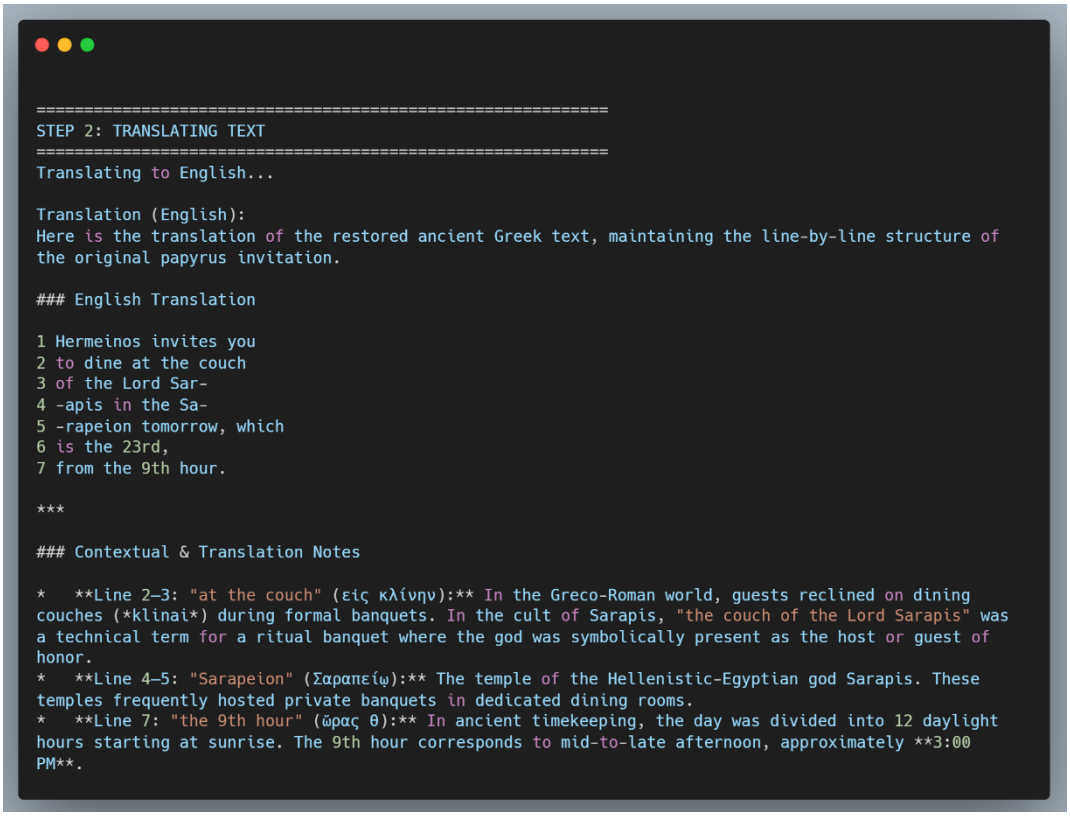
μετατροπή του "ΕΡΩΖΤΑCE" στο ορθό "ἐρωτᾷ σε" και την επιτυχή ανακατασκευή του ρήματος "δειπνῆσαι").



Εικόνα 13 . Αποκατάσταση κειμένου με Gemini 3.5 Flash

Αμέσως μετά, το σύστημα προχώρησε στη μετάφραση και παρήγαγε αυτόματα ένα συνοδευτικό κείμενο ιστορικών και φιλολογικών παρατηρήσεων. Εκεί, επεξήγησε αυτόματα τον τεχνικό όρο της "κλίνης του Κυρίου Σαράπιδος" ως ιεροτελεστικό συμπόσιο στο Σαραπεῖο, καθώς και τη μετατροπή της 9ης ώρας σε σύγχρονο χρόνο (περίπου 15:00). Η παραβολή του τελικού αποτελέσματος με την επίσημη ακαδημαϊκή έκδοση του παπύρου καταδεικνύει την υψηλή ακρίβεια του pipeline. Η μοναδική αξιοσημείωτη απόκλιση εντοπίζεται στη γραμμή 6, όπου η πραγματική ανάγνωση "ἥτις ἐστὶν τριακὰς" αποδόθηκε

από το μοντέλο ως "ἥτις ἐστὶν τῇ κγ(*)". Η απόκλιση αυτή είναι απολύτως δικαιολογημένη, καθώς η ακατέργαστη είσοδος του OCR ("ΤΙCΕΠΙΝΤΕCΛ[Κ/Α]CΓ") διέθετε το γράμμα "Γ" στο τέλος, οδηγώντας το LLM να υποθέσει λογικά την ημερομηνία "κγ" (23η) αντί του ολογράφου "τριακάς" (30ή). Η συμπεριφορά αυτή αποδεικνύει περίτρανα ότι τα σύγχρονα LLMs διαθέτουν βαθιά εγγενή γνώση των παπυρολογικών φορμών και μπορούν να λειτουργήσουν αυτόνομα ως ισχυροί μηχανισμοί αποθορύβοποίησης.

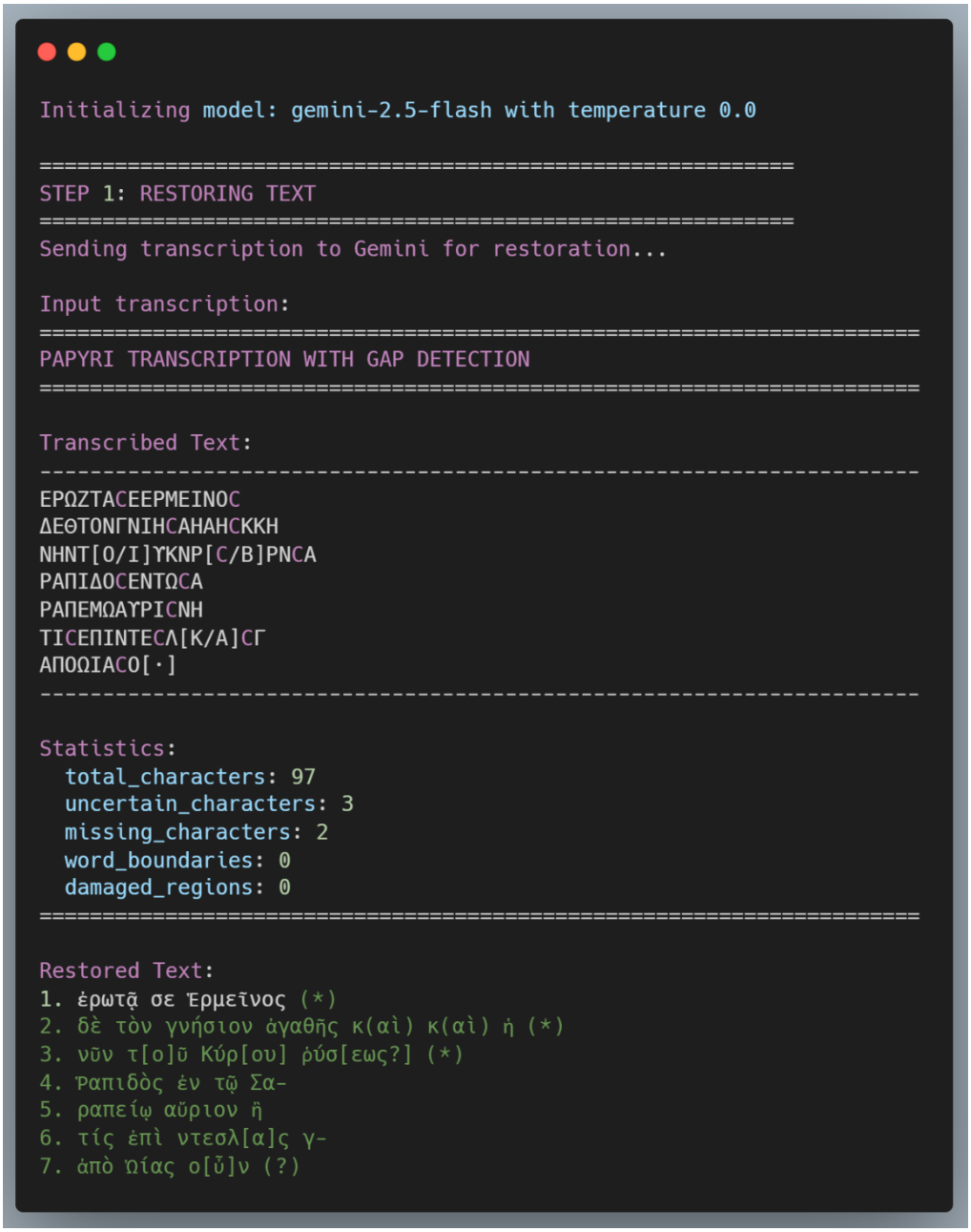


```
============================================================
STEP 2: TRANSLATING TEXT
============================================================
Translating to English...

Translation (English):
Here is the translation of the restored ancient Greek text, maintaining the line-by-line structure of
the original papyrus invitation.

### English Translation

1 Hermeinos invites you
2 to dine at the couch
3 of the Lord Sar-
4 -apis in the Sa-
5 -rapeion tomorrow, which
6 is the 23rd,
7 from the 9th hour.

***

### Contextual & Translation Notes

*   **Line 2–3: "at the couch" (εἰς κλίνην):** In the Greco-Roman world, guests reclined on dining
couches (*klinai*) during formal banquets. In the cult of Sarapis, "the couch of the Lord Sarapis" was
a technical term for a ritual banquet where the god was symbolically present as the host or guest of
honor.
*   **Line 4–5: "Sarapeion" (Σαραπείῳ):** The temple of the Hellenistic-Egyptian god Sarapis. These
temples frequently hosted private banquets in dedicated dining rooms.
*   **Line 7: "the 9th hour" (ὥρας θ):** In ancient timekeeping, the day was divided into 12 daylight
hours starting at sunrise. The 9th hour corresponds to mid-to-late afternoon, approximately **3:00
PM**.
```

Εικόνα 14. Μετάφραση κειμένου με Gemini 3.5 Flash

Για τη διερεύνηση της εξέλιξης των γλωσσικών μοντέλων στην ερμηνεία θορυβωδών παπυρικών μεταγραφών απουσία εξωτερικής βάσης γνώσης, το πείραμα στον πάπυρο P.Oxy. 52 3693 επαναλήφθηκε υπό ακριβώς τις ίδιες συνθήκες με τη χρήση του μοντέλου Gemini 2.5 Flash. Ενώ προηγουμένως το προηγμένο Gemini 3.5 Flash είχε αναγνωρίσει επιτυχώς το πλαίσιο της πρόσκλησης σε δείπνο («δειπνῆσαι εἰς κλίνην του κυρίου Σαράπιδος»), το Gemini 2.5 Flash υπέστη πλήρη σημασιολογική κατάρρευση. Αδυνατώντας να φιλτράρει τον έντονο οπτικό θόρυβο της δεύτερης και τρίτης γραμμής,

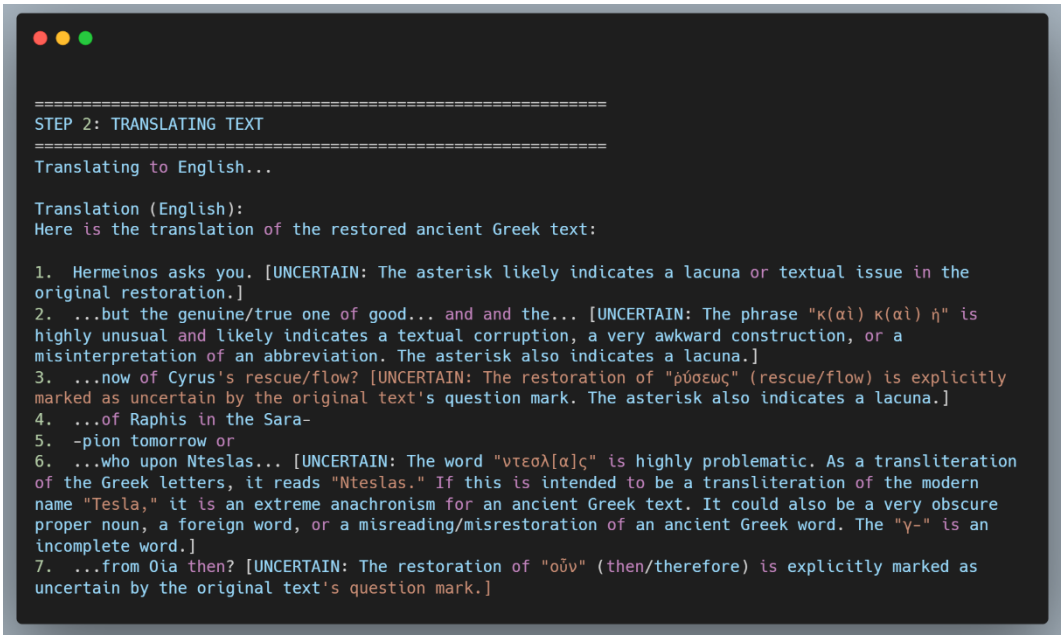
παρήγαγε μια σειρά από ασύνδετες και λανθασμένες λέξεις, μετατρέποντας το συμπόσιο σε μια ασυναρτησία περί «γνησίου αγαθῆς» και «Κύρου ῥύσεως».



Εικόνα 15. Αποκατάσταση κειμένου με Gemini 2.5 Flash

Η συγκριτική αυτή αποτυχία γίνεται κρισιμότερη κατά το στάδιο της μετάφρασης και του σχολιασμού, όπου αποκαλύπτεται η έντονη τάση του παλαιότερου μοντέλου για παραισθήσεις. Ενδεικτική είναι η επεξεργασία της έκτης γραμμής, όπου η ακατέργαστη έξοδος του OCR («ΤΙCΕΠΙΝΤΕCΛ[Κ/Α]CΓ») οδήγησε το μοντέλο στην αυθαίρετη δημιουργία της ανύπαρκτης λέξης «ντεσλ[α]ς». Στις μεταφραστικές του σημειώσεις, το ίδιο το μοντέλο παραδέχεται την αδυναμία του, επισημαίνοντας ότι ο όρος μοιάζει με αναχρονιστική μεταγραφή του σύγχρονου ονόματος "Tesla", επιβεβαιώνοντας πως

ελλείψει RAG τα παλαιότερα μοντέλα τείνουν να προβάλλουν σύγχρονες έννοιες πάνω στο αρχαίο υπόστρωμα.



Εικόνα 16. Μετάφραση κειμένου με Gemini 2.5 Flash

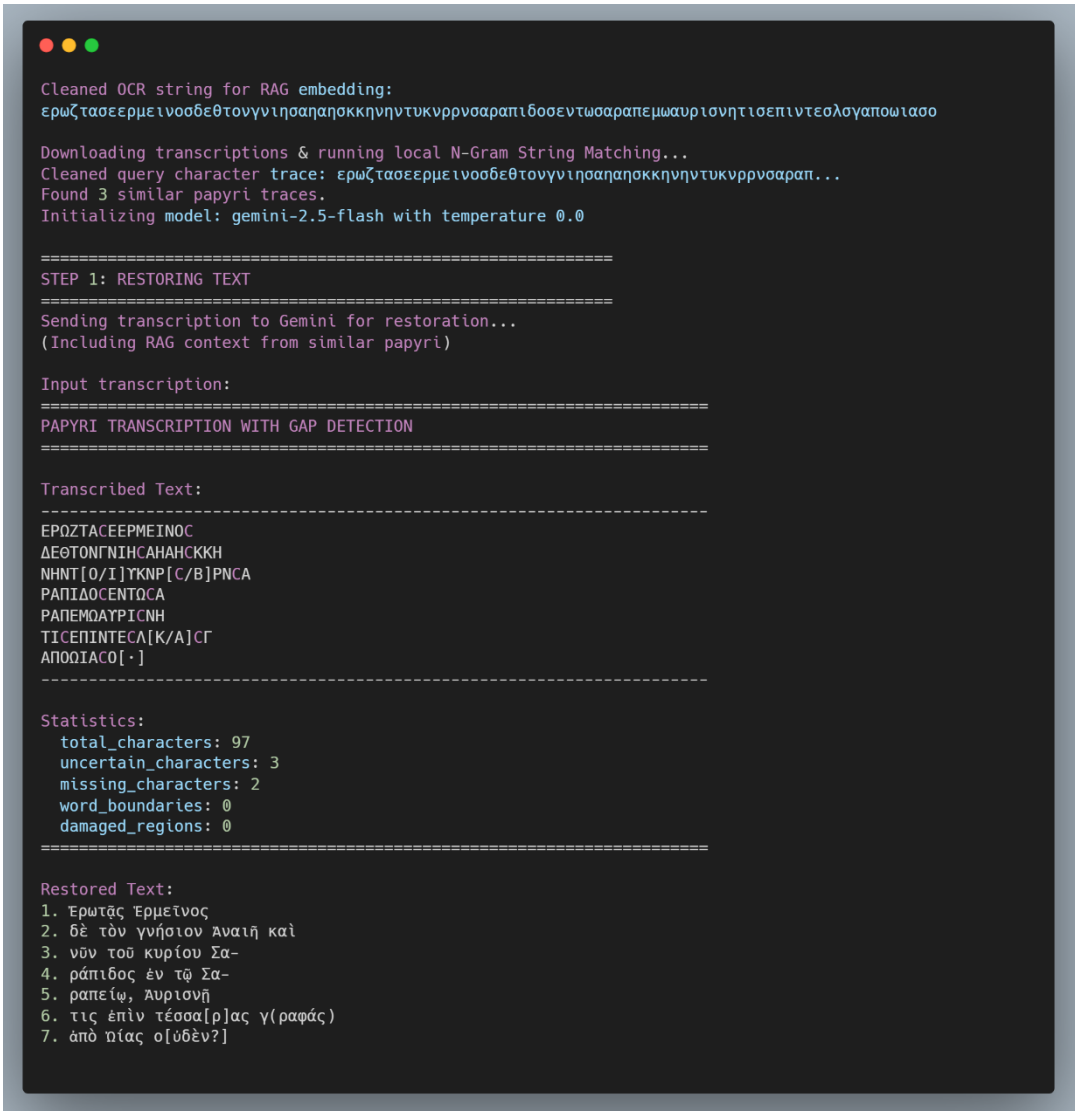
Η αντιπαραβολή των δύο πειραμάτων αποδεικνύει περίτρανα τη ραγδαία πρόοδο των αρχιτεκτονικών LLM στην αποθορυβοποίηση δεδομένων. Ενώ το Gemini 2.5 Flash παρασύρεται από τα σφάλματα του αλγορίθμου υπολογιστικής όρασης και καταφεύγει σε κατά λέξη ερμηνείες που καταστρέφουν το νόημα, η νεότερη έκδοση 3.5 Flash επιδεικνύει βαθιά, εγγενή γνώση των αρχαιοελληνικών δομών, λειτουργώντας σχεδόν σε επίπεδο ανθρώπινου παπυρολόγου ακόμα και χωρίς εξωτερική ανάκτηση δεδομένων.

Για τη διερεύνηση του ρόλου της εξωτερικής γνώσης στην αποκατάσταση κειμένου, το πείραμα με το Gemini 2.5 Flash επαναλήφθηκε με την ενεργοποίηση της αρχιτεκτονικής RAG.

```
python scripts/experiment_transcription_to_text.py
--transcription  file C:\Users\themi\Desktop\ancient-yolo\transcription.txt\p.oxy.52.3693  detailed.txt
--output  file restored.txt
--use_rag
```

Η συμβολή του RAG ήταν άμεση και κρίσιμη στο επίπεδο της τοπικής διόρθωσης των δεδομένων. Μέσω της αντιστοίχισης N-γραμμάτων, το σύστημα εντόπισε παρόμοιες παπυρικές εγγραφές στη διανυσματική βάση και τροφοδότησε το μοντέλο με έγκυρα ιστορικά παράλληλα. Αυτή η διαδικασία βοήθησε το γλωσσικό μοντέλο, καθώς του επέτρεψε να διορθώσει τη δομική αστοχία της προηγούμενης εκτέλεσης, αποκαθιστώντας επιτυχώς τη φράση "τοῦ κυρίου Σαράπιδος ἐν τῷ Σαραπείῳ". Είναι σαφές ότι, χωρίς το

RAG, το μοντέλο δεν είχε κανένα σημείο αναφοράς για να διακρίνει τις σωστές ιερατικές εκφράσεις μέσα στον έντονο θόρυβο του OCR.



Εικόνα 17. Αποκατάσταση κειμένου με Gemini 2.5 Flash με RAG

Ωστόσο, η ανάλυση ανέδειξε και τα όρια του RAG ως υποβοηθητικού μηχανισμού. Το μοντέλο χρησιμοποίησε το πλαίσιο του RAG για να διορθώσει το όνομα του θεού, αλλά αδυνατούσε να συνθέσει αυτά τα σωστά αποσπάσματα σε ένα ενιαίο, νοηματικά συνεκτικό πλαίσιο, καταφεύγοντας τελικά σε μια αυθαίρετη ερμηνεία του υπόλοιπου κειμένου. Αυτό το εύρημα επιβεβαιώνει ότι το RAG λειτουργεί ιδανικά ως μηχανισμός «γειώματος» για συγκεκριμένους όρους και σταθερές φράσεις, μετατρέποντας μια ακατάληπτη είσοδο σε

κάτι πιο οικείο για το μοντέλο, πλην όμως δεν μπορεί να διδάξει "in-context" το γενικότερο νόημα.

Εν κατακλείδι, η συμβολή του RAG στο pipeline αποδείχθηκε καθοριστική για την επίτευξη ακρίβειας σε συγκεκριμένα παπυρολογικά "κλειδιά". Ενώ στο προηγούμενο πείραμα το μοντέλο παραπατούσε σε όλη την έκταση του κειμένου, με την προσθήκη του RAG πέτυχε μια σημαντική βελτίωση στην αποκατάσταση της ορολογίας, επιβεβαιώνοντας πως η υβριδική στρατηγική είναι απαραίτητη για την εξασφάλιση της ιστορικής ορθότητας σε συστήματα αυτόματης αποκατάστασης, ακόμη και όταν το γλωσσικό μοντέλο βρίσκεται στα όρια των δυνατοτήτων του.

4. Πειραματική Διαδικασία, Παραμετροποίηση, Εκπαίδευση και Αξιολόγηση Μοντέλων

Η επιβεβαίωση της λειτουργικότητας της προτεινόμενης αρχιτεκτονικής απαιτεί μια συστηματική και αυστηρή πειραματική διαδικασία. Στο παρόν κεφάλαιο αναλύονται οι τεχνικές παράμετροι του περιβάλλοντος εκτέλεσης, προσδιορίζονται οι υπερπαράμετροι που χρησιμοποιήθηκαν για την εκπαίδευση των μοντέλων YOLO, ResNet και DeiT, και καθορίζονται τα επιστημονικά κριτήρια αξιολόγησης. Στο τέλος του κεφαλαίου παρατίθενται πίνακες σύγκρισης, οι οποίοι προορίζονται για την καταγραφή των εμπειρικών μετρήσεων και την απρόσκοπτη σύγκριση των τρεχόντων μοντέλων με μελλοντικές αρχιτεκτονικές.

4.1 Περιβάλλον Εκπαίδευσης και hardware

Η υπολογιστική πολυπλοκότητα των αλγορίθμων βαθιάς μάθησης επιβάλλει τη χρήση προηγμένου hardware και βελτιστοποιημένων βιβλιοθηκών. Για την εκπαίδευση των συνελικτικών μοντέλων και των μοντέλων μετασχηματιστών της παρούσας διπλωματικής, δημιουργήθηκε ένα ειδικό περιβάλλον με βάση τη γλώσσα Python 3.12.

Όσον αφορά το hardware, τα πειράματα προσαρμόστηκαν ανάλογα με τις απαιτήσεις του εκάστοτε μοντέλου. Για το μοντέλο εντοπισμού YOLO, η εκπαίδευση εκτελέστηκε σε περιβάλλον CPU, αξιοποιώντας έναν πολυπύρηνο επεξεργαστή 13ης γενιάς Intel Core i5-13400F. Αντίθετα, για τα μοντέλα κατηγοριοποίησης (ResNet/DeiT), όπου απαιτείται η επεξεργασία χιλιάδων εικόνων από το σύνολο AL-PUB, το σύστημα ρυθμίστηκε ώστε να χρησιμοποιεί την πλατφόρμα CUDA, μεταφέροντας το φορτίο εκμάθησης σε Μονάδα

Επεξεργασίας Γραφικών (GPU), και συγκεκριμένα στην κάρτα NVIDIA GeForce RTX 4060, ελαχιστοποιώντας δραστικά τον χρόνο εκτέλεσης.

4.2 Διαδικασία Εκπαίδευσης Μοντέλων Υπολογιστικής Όρασης

4.2.1 Εκπαίδευση Μοντέλου YOLO11s για Εντοπισμό Χαρακτήρων

Το μοντέλο ανίχνευσης αντικειμένων YOLO11s (έργο "greek_papyri_detection", εκτέλεση "detection_only_m") υποβλήθηκε σε εντατική εκπαίδευση για 200 εποχές (epochs). Η επιλογή της συγκεκριμένης αρχιτεκτονικής κλίμακας (Small) έναντι μικρότερων (YOLO11n - Nano) ή μεγαλύτερων παραλλαγών (YOLO11m - Medium) προέκυψε ως η βέλτιστη λύση στον άξονα της υπολογιστικής απόδοσης έναντι της ακρίβειας. Ειδικότερα, το μικρότερο μοντέλο Nano υστερεί σε αρχιτεκτονική πολυπλοκότητα, με αποτέλεσμα να αδυνατεί να αναγνωρίσει αξιόπιστα τα λεπτά όρια των φθαρμένων, μικροσκοπικών γραμμάτων, οδηγώντας σε αυξημένα «ψευδώς αρνητικά» αποτελέσματα. Στον αντίποδα, η χρήση μεγαλύτερων μοντέλων θεωρήθηκε πλεονασμός για ένα πρόβλημα ανίχνευσης μονής κλάσης, ενώ θα απαιτούσε περισσότερη μνήμη. Το YOLO11s αποτέλεσε τον ιδανικό συμβιβασμό, επιτρέποντας την αποτελεσματική εκπαίδευση σε εικόνες υψηλής ανάλυσης (1024 X 1024 pixel). Εξαιτίας ακριβώς αυτής της υψηλής ανάλυσης εισόδου, το batch size διατηρήθηκε υποχρεωτικά χαμηλό (4 εικόνες), ενώ προκειμένου να αποτραπεί το overfitting, ενεργοποιήθηκε ο μηχανισμός early stopping με ανοχή στις 50 εποχές.

Για τη βελτιστοποίηση των βαρών επιλέχθηκε ο αλγόριθμος AdamW (Adam with Weight Decay), με αρχικό ρυθμό εκμάθησης 0.005 και momentum 0.937. Εφαρμόστηκε ένα warmup 10 εποχών, ενώ η συνέχεια της μείωσης του ρυθμού εκμάθησης ρυθμίστηκε μέσω της τεχνικής Cosine Annealing (cos_lr=True).

Ο παρακάτω πίνακας περιγράφει τις μεθόδους επαύξησης δεδομένων (Data Augmentation) που εφαρμόστηκαν.

Τεχνική Επαύξησης (Augmentation)	Τιμή / Πιθανότητα Εφαρμογής	Περιγραφή / Παπυρολογική Σκοπιμότητα
Mosaic	0.3	Σύνθεση 4 εικόνων. Βελτιώνει τον εντοπισμό χαρακτήρων. Απενεργοποιήθηκε τις τελευταίες 50 εποχές.
Degrees (Περιστροφή)	3.0°	Μικρή περιστροφή για προσομοίωση της διακύμανσης στην κλίση γραφής.
Translate (Μετατόπιση)	0.1	Μετατόπιση στον άξονα για ανθεκτικότητα στη θέση του χαρακτήρα.
Scale (Αλλαγή Κλίμακας)	0.5	Αυξομείωση μεγέθους για έγγραφα με διαφορετική εστίαση ψηφιοποίησης.
Erasing (Τυχαία Διαγραφή)	0.4	Απόκρυψη τμημάτων της εικόνας, προσομοιώνοντας φθορές και ξεθώριασμα.

Πίνακας 3. Παράμετροι Επαύξησης Δεδομένων Εκπαίδευσης YOLO11s

4.2.2 Εκπαίδευση Μοντέλου YOLO11s με Εστίαση στον Εντοπισμό Χαρακτήρων (Detection-Only)

Το μοντέλο YOLO11s "detection_only" διαμορφώθηκε με αποκλειστικό στόχο τον ακριβή εντοπισμό των χαρακτήρων στον πάπυρο, παρακάμπτοντας στο στάδιο αυτό την ταξινόμησή τους. Για τον σκοπό αυτό, ενεργοποιήθηκε η παράμετρος single_cls=True, η οποία αντιμετωπίζει όλα τα γράμματα ως μία ενιαία κλάση.

Η εκπαίδευση διήρκεσε 200 εποχές με batch size 4 εικόνων, προκειμένου να διατηρηθεί η υψηλή ανάλυση εισόδου των 1024x1024 pixel, η οποία είναι κρίσιμη για τη διατήρηση της οπτικής λεπτομέρειας των μικρών γραμμάτων. Ο μηχανισμός early stopping patience ρυθμίστηκε στις 50 εποχές. Λόγω της συνήθους πυκνότητας του κειμένου στα παπυρικά έγγραφα, το μέγιστο όριο ανιχνεύσεων ανά εικόνα αυξήθηκε στους 500 χαρακτήρες (max_det=500).

Προκειμένου το δίκτυο να βελτιστοποιηθεί γεωμετρικά για την εξαγωγή των bounding boxes), τροποποιήθηκαν σημαντικά τα βάρη της συνάρτησης απώλειας: δόθηκε εξαιρετικά υψηλή βαρύτητα στην απώλεια πλαισίου (box=10.0) και στην απώλεια εστίασης κατανομής (dfl=2.0), ενώ η απώλεια ταξινόμησης ελαχιστοποιήθηκε (cls=0.1).

Ο παρακάτω πίνακας αναλύει τις μεθόδους Data Augmentation που επιλέχθηκαν, προσαρμοσμένες αυστηρά στις φυσικές και οπτικές ιδιαιτερότητες των παπύρων.

Τεχνική Επαύξησης (Augmentation)	Τιμή / Πιθανότητα Εφαρμογής	Περιγραφή / Παπυρολογική Σκοπιμότητα
Mosaic	0.3	Σύνθεση 4 εικόνων. Διατηρήθηκε σε χαμηλό ποσοστό, καθώς η χωρική διάταξη του κειμένου έχει σημασία. Απενεργοποιήθηκε τις τελευταίες 50 εποχές (close_mosaic=50).
Flip (Οριζόντια & Κάθετη)	0.0	Πλήρης απενεργοποίηση. Οι αρχαίοι ελληνικοί χαρακτήρες έχουν αυστηρή κατεύθυνση και δεν μπορούν να αναγνωριστούν ανεστραμμένοι ή καθρεφτισμένοι.
Degrees (Περιστροφή)	3.0°	Πολύ μικρή περιστροφή για προσομοίωση της φυσικής κλίσης των γραμμών του εκάστοτε γραφέα ή μικρών αποκλίσεων κατά την ψηφιοποίηση.
Scale (Αλλαγή Κλίμακας)	0.5	Αυξομείωση μεγέθους για να καταστεί το μοντέλο ανθεκτικό σε έγγραφα διαφορετικής ανάλυσης και εστίασης φακού.
Translate (Μετατόπιση)	0.1	Μικρή μετατόπιση στον άξονα για την ενίσχυση της ανθεκτικότητας ως προς τη θέση του χαρακτήρα στο πλαίσιο.
Erasing (Τυχαία Διαγραφή)	0.4	Απόκρυψη τμημάτων της εικόνας, προσομοιώνοντας με ρεαλιστικό τρόπο τις φθορές του υλικού και το ξεθώριασμα του μελανιού.
HSV (Χρώμα, Κορεσμός, Φωτεινότητα)	H: 0.01, S: 0.5, V: 0.4	Ελάχιστη αλλαγή απόχρωσης, αλλά μέτρια διακύμανση κορεσμού και φωτεινότητας για προσομοίωση διαφορετικών συνθηκών φωτισμού, παλαιότητας του παπύρου και πυκνότητας μελανιού.

Πίνακας 4. Παράμετροι Επαύξησης Δεδομένων Εκπαίδευσης για Εντοπισμό

Η επιλογή της εκπαίδευσης του μοντέλου YOLO11s αγνοώντας την ταξινόμηση των επιμέρους γραμμάτων, αποτελεί μια στρατηγική αρχιτεκτονική απόφαση. Αντί για μια προσέγγιση «ενός σταδίου», όπου ένα και μόνο δίκτυο αναλαμβάνει τόσο τον εντοπισμό όσο και την αναγνώριση, η παρούσα έρευνα υιοθετεί μια αρχιτεκτονική δύο διακριτών σταδίων.

Αυτή η μεθοδολογική προσέγγιση επιλέχθηκε για να αντιμετωπίσει τρία θεμελιώδη προβλήματα της υπολογιστικής παπυρολογίας:

1. **<u>Στατιστική Ανισορροπία Κλάσεων:</u>**
   Στο ελληνικό αλφάβητο, γράμματα όπως το Άλφα (Α), το Όμικρον (Ο) και το Πι (Π) εμφανίζονται με συντριπτικά μεγαλύτερη συχνότητα σε σχέση με

γράμματα όπως το Ξι (Ξ) ή το Ψι (Ψ). Όταν ένα μοντέλο ανίχνευσης (όπως το YOLO) προσπαθεί να μάθει ταυτόχρονα τις συντεταγμένες και την κλάση σε ένα τόσο ανισόρροπο σύνολο δεδομένων, τείνει να δημιουργεί μεροληψία (bias) υπέρ των συχνών χαρακτήρων, υποβαθμίζοντας την ακρίβεια στα σπάνια γράμματα.

2. **Λεπτομερής Εξαγωγή Χαρακτηριστικών:**
   Η διάκριση μεταξύ οπτικά παρόμοιων αρχαίων χαρακτήρων (π.χ. μεταξύ ενός φθαρμένου Γάμμα και ενός Ταυ) απαιτεί ανάλυση υφής υψηλής ανάλυσης. Η αποσύνδεση επιτρέπει τη χρήση ενός εξειδικευμένου, ξεχωριστού Συνελικτικού Νευρωνικού Δικτύου ή Transformer στο δεύτερο στάδιο, το οποίο θα εστιάσει αποκλειστικά στην εξαγωγή μορφολογικών χαρακτηριστικών χωρίς το υπολογιστικό βάρος της χωρικής αναζήτησης σε όλη την εικόνα του παπύρου.
3. **Αρθρωτή Αρχιτεκτονική (Modularity):**

   Ο διαχωρισμός του pipeline σημαίνει ότι τα μοντέλα μπορούν να βελτιστοποιηθούν ανεξάρτητα. Αν στο μέλλον χρειαστεί να προστεθεί μια νέα γραμματοσειρά ή να εισαχθεί ένα νέο μοντέλο για τα Κοπτικά, το μοντέλο ανίχνευσης (YOLO) παραμένει άθικτο, και απαιτείται επανεκπαίδευση μόνο του ταξινομητή.

4.2.3 Εκπαίδευση Μοντέλων ResNet-50 και DeiT για Κατηγοριοποίηση

Η εκπαίδευση των μοντέλων ταξινόμησης σχεδιάστηκε για να λειτουργεί συμπληρωματικά με το μοντέλο εντοπισμού, αναλαμβάνοντας την αναγνώριση των μεμονωμένων αποκομμάτων στο δεύτερο στάδιο του pipeline. Η υλοποίηση βασίστηκε στη βιβλιοθήκη transformers της Hugging Face και στο οικοσύστημα του PyTorch. Για λόγους συγκριτικής αξιολόγησης, επιλέχθηκαν δύο θεμελιωδώς διαφορετικές αρχιτεκτονικές: ένα βαθύ Συνελικτικό Νευρωνικό Δίκτυο (ResNet-50) και ένας Μηχανισμός Προσοχής (DeiT - Data-efficient Image Transformers).

Για την εκπαίδευση χρησιμοποιήθηκε το σύνολο δεδομένων AL-PUB. Κατά την ανάλυση του dataset, εντοπίστηκαν 24 διακριτές κλάσεις (αρχαίοι ελληνικοί χαρακτήρες) με συνολικό όγκο άνω των 205.000 εικόνων. Προκειμένου να διασφαλιστεί η ταχεία σύγκλιση των μοντέλων και η αποτροπή στατιστικής μεροληψίας υπέρ των συχνότερα εμφανιζόμενων χαρακτήρων, εφαρμόστηκε stratified sampling. Εξήχθη έτσι ένα ισορροπημένο υποσύνολο, περιορίζοντας τον μέγιστο αριθμό δειγμάτων σε 2.000 ανά κλάση. Το τελικό σύνολο 43.151 εικόνων υποβλήθηκε σε αυστηρό διαχωρισμό 80% για Εκπαίδευση και 20% για Επαλήθευση.

Ο AutoImageProcessor κανονικοποίησε τις εικόνες στη βέλτιστη ανάλυση εισόδου, μετατρέποντάς τες παράλληλα σε χρωματικό χώρο RGB. Η επαύξηση των δεδομένων

εκπαίδευσης υλοποιήθηκε μέσω της κλάσης torchvision.transforms.Compose και διαφοροποιήθηκε στρατηγικά ανάλογα με την αρχιτεκτονική:

- **Μοντέλο ResNet-50:** Η επαύξηση εστίασε στις φυσικές ανωμαλίες του φθαρμένου υλικού. Εφαρμόστηκε τυχαία περιστροφή έως 15 μοίρες, τυχαία περικοπή με αλλαγή κλίμακας (RandomResizedCrop από 0.8 έως 1.0), χρωματική ταλάντωση (ColorJitter) με παράγοντα 0.2 για φωτεινότητα/αντίθεση, και οριζόντια αναστροφή (RandomHorizontalFlip) με πιθανότητα p=0.2.
- **Μοντέλο DeiT:** Δεδομένου ότι τα δίκτυα Transformer απαιτούν μεγαλύτερη ποικιλομορφία δεδομένων για την αποφυγή υπερεκπαίδευσης, εφαρμόστηκε πιο επιθετική επαύξηση. Το εύρος αλλαγής κλίμακας (RandomResizedCrop) διευρύνθηκε (0.7 έως 1.0) και η χρωματική ταλάντωση ενισχύθηκε (brightness=0.3, contrast=0.3, saturation=0.2).

Ως μοντέλα βάσης χρησιμοποιήθηκαν τα προεκπαιδευμένα microsoft/resnet-50 και facebook/deit-base-patch16-224. Για την προσαρμογή τους στο ειδικό πεδίο της παπυρολογίας, το τελικό επίπεδο ταξινόμησης των 1000 κλάσεων του ImageNet αφαιρέθηκε και αντικαταστάθηκε δυναμικά από ένα νέο πλήρως συνδεδεμένο επίπεδο 24 κλάσεων, επιτρέποντας στα δίκτυα να αξιοποιήσουν την προϋπάρχουσα γνώση εξαγωγής χαρακτηριστικών.

Τα δίκτυα βελτιστοποιήθηκαν ως προς τη συνάρτηση κόστους Cross-Entropy Loss. Ως βελτιστοποιητής επιλέχθηκε ο AdamW με σταθερό ρυθμό μάθησης και Weight Decay 0.01. Για την εξασφάλιση ομαλής σύγκλισης στα τελικά στάδια, εφαρμόστηκε Cosine Annealing, ενώ προς αποφυγή exploding gradients επιβλήθηκε αποκοπή κλίσης (Gradient Clipping, max_norm=1.0).

Παράμετρος	Τιμή
Σύνολο Δεδομένων (Dataset)	AL-PUB (Balanced Υποσύνολο)
Πλήθος Εικόνων (Train / Val)	34.520 / 8.631 (Διαχωρισμός 80/20)
Αριθμός Κλάσεων (Classes)	24
Ανάλυση Εισόδου (Image Size)	224 x 224 pixels (RGB)
Συνάρτηση Απώλειας (Loss)	Cross-Entropy Loss
Βελτιστοποιητής (Optimizer)	AdamW (Weight Decay: 0.01)
Χρονοδιάγραμμα LR (Scheduler)	Cosine Annealing LR

Μέγεθος Δέσμης (Batch Size)	16 (Train) / 32 (Validation)
Gradient Clipping (max_norm)	1.0
Εποχές Εκπαίδευσης (Epochs)	10 (ResNet) / 8 (DeiT)

Πίνακας 5. Παράμετροι Εκπαίδευσης Ταξινομητών (ResNet-50 & DeiT)

4.3 Μετρικές Αξιολόγησης

Δεδομένης της σύνθετης αρχιτεκτονικής του προτεινόμενου pipeline, η αξιολόγηση της απόδοσης του συστήματος δεν μπορεί να βασιστεί σε μία ενιαία μετρική. Αντιθέτως, απαιτείται η χρήση εξειδικευμένων δεικτών που ποσοτικοποιούν την επιτυχία κάθε επιμέρους σταδίου ανεξάρτητα: από τη γεωμετρική ακρίβεια του εντοπισμού, στην ορθότητα της ταξινόμησης, έως τη σημασιολογική συνοχή του παραγόμενου κειμένου.

Στο στάδιο της εξαγωγής των bounding boxes, η αξιολόγηση εστιάζει αποκλειστικά στην ικανότητα του μοντέλου να εντοπίζει τα ίχνη μελανιού και να τα διαχωρίζει από τον οπτικό θόρυβο. Βασικός πυλώνας αυτών των μετρικών είναι η Επικάλυψη προς Ένωση (Intersection over Union - IoU), η οποία μετρά τον βαθμό γεωμετρικής ταύτισης μεταξύ του προβλεπόμενου πλαισίου και του πραγματικού πλαισίου.

- **Precision:** Ποσοτικοποιεί την ορθότητα των ανιχνεύσεων (δηλαδή, πόσα από τα πλαίσια που πρότεινε το μοντέλο περιείχαν πράγματι χαρακτήρες, ελαχιστοποιώντας τα Ψευδώς Θετικά - False Positives).

- **Recall:** Μετρά την ικανότητα του μοντέλου να εντοπίσει όλους τους υπαρκτούς χαρακτήρες του εγγράφου, ελαχιστοποιώντας τα Ψευδώς Αρνητικά (False Negatives). Στα φθαρμένα παπυρικά κείμενα, υψηλό Recall σημαίνει ότι το μοντέλο δεν «χάνει» δυσανάγνωστα γράμματα.

- **Μέση Ακρίβεια (mAP - Mean Average Precision):** Αποτελεί την πιο ολοκληρωμένη μετρική, καθώς υπολογίζει το εμβαδόν κάτω από την καμπύλη Precision-Recall Curve.

- **mAP@50:** Η μέση ακρίβεια όταν απαιτείται ελάχιστο IoU 50% για να θεωρηθεί μια ανίχνευση επιτυχής. Αποτελεί τη βασική μετρική του συστήματος, καθώς ένα πλαίσιο που περιλαμβάνει το 50% του γράμματος είναι συχνά επαρκές για να τροφοδοτήσει επιτυχώς το μοντέλο ταξινόμησης (Στάδιο 2).

- **mAP@50-95:** Πιο αυστηρή μετρική που υπολογίζει τον μέσο όρο του mAP σε πολλαπλά κατώφλια IoU (από 0.50 έως 0.95 με βήμα 0.05). Δείχνει το πόσο

"σφιχτά" και απόλυτα ευθυγραμμισμένα είναι τα πλαίσια του YOLO γύρω από τους χαρακτήρες.

$$Precision = \frac{True\ Positives}{True\ Positives + False\ Positives}$$

$$Recall = \frac{True\ Positives}{True\ Positives + False\ Negatives}$$

Το mAP υπολογίζεται στο όριο 50% (mAP@50) και στον μέσο όρο των ορίων 50-95% (mAP@50-95).

Κατά το δεύτερο στάδιο, το βάρος μετατοπίζεται από τη γεωμετρία στην οπτική αναγνώριση προτύπων. Η απόδοση των μοντέλων ResNet-50 και DeiT αξιολογείται με βάση τον πίνακα confusion matrix και τις κάτωθι μετρικές:

- **Ολική Ακρίβεια (Accuracy):** Το ποσοστό των συνολικά σωστών προβλέψεων επί του συνόλου των δεδομένων αξιολόγησης. Αν και ενδεικτική της γενικής απόδοσης, η μετρική αυτή είναι συχνά παραπλανητική σε ανισόρροπα σύνολα δεδομένων, όπως αυτά της αρχαίας ελληνικής γραμματείας, όπου φωνήεντα (π.χ. Άλφα, Όμικρον) κυριαρχούν συντριπτικά έναντι ορισμένων συμφώνων (π.χ. Ψι, Ξι).
- **Μακρο-σταθμισμένο F1-Score (Macro F1-Score):** Για την αντιμετώπιση του παραπάνω προβλήματος, η αξιολόγηση βασίζεται κυρίως στο Macro F1-Score. Το F1-Score αποτελεί τον αρμονικό μέσο του Precision και του Recall. Η "Macro" προσέγγιση υπολογίζει το F1-Score ανεξάρτητα για κάθε μία από τις 24 κλάσεις γραμμάτων και στη συνέχεια εξάγει τον μη σταθμισμένο μέσο όρο τους. Αυτό σημαίνει ότι όλες οι κλάσεις αντιμετωπίζονται ισότιμα, εμποδίζοντας τα συχνά γράμματα να "καλύψουν" τη χαμηλή απόδοση του μοντέλου στα σπάνια γράμματα.

Στο τελικό στάδιο, όπου τα Μεγάλα Γλωσσικά Μοντέλα (LLMs) και η αρχιτεκτονική RAG αναλαμβάνουν να γεφυρώσουν τα κενά και να εξάγουν το τελικό κείμενο, οι παραδοσιακές μετρικές μηχανικής όρασης δεν επαρκούν. Για τον λόγο αυτόν χρησιμοποιούνται:

- **Ποσοστό Σφάλματος Χαρακτήρων (CER - Character Error Rate):** Αποτελεί την κατεξοχήν μετρική αξιολόγησης συστημάτων OCR. Βασίζεται στην Απόσταση Levenshtein και υπολογίζει τον ελάχιστο αριθμό παρεμβολών, διαγραφών και αντικαταστάσεων που απαιτούνται για να μετατραπεί το παραγόμενο κείμενο στο αρχικό, πραγματικό κείμενο. Χαμηλό CER υποδηλώνει υψηλή πιστότητα της μεταγραφής του παπύρου.

- **Ποσοστό Σφάλματος Λέξεων (WER - Word Error Rate):** Λειτουργώντας συμπληρωματικά με το CER, το WER εφαρμόζει τον αλγόριθμο της Απόστασης Levenshtein σε επίπεδο ολόκληρων λέξεων αντί για μεμονωμένους χαρακτήρες. Στο πλαίσιο της αποκατάστασης κειμένου από LLMs, το WER είναι κρίσιμο διότι αξιολογεί τη λεξιλογική και σημασιολογική ορθότητα της πρόβλεψης. Ένα μοντέλο μπορεί να παρουσιάζει σχετικά χαμηλό CER αλλά υψηλό WER, εάν τα λιγοστά σφάλματα χαρακτήρων αλλοιώνουν ολοκληρωτικά τις παραγόμενες λέξεις. Ειδικά για τα αρχαία ελληνικά παπυρικά κείμενα, τα οποία συχνά ακολουθούν τη συνεχή γραφή, το WER πιστοποιεί την ικανότητα του RAG και του LLM να διαχωρίζουν, να ανακτούν και να συνθέτουν έγκυρο και ιστορικά ακριβές λεξιλόγιο.
- **Μετρική BLEU (Bilingual Evaluation Understudy):** Αν και παραδοσιακά χρησιμοποιείται στη Μηχανική Μετάφραση, στο πλαίσιο αυτής της έρευνας αξιολογεί την ικανότητα του LLM να παράγει γραμματικά και συντακτικά ορθές μεταφράσεις. Μετρά την επικάλυψη των n-RAMS μεταξύ της παραγόμενης πρότασης του AI και του κειμένου αναφοράς που έχει μεταγραφεί από φιλολόγους. Υψηλό σκορ BLEU επιβεβαιώνει ότι το σύστημα RAG ανάκτησε επιτυχώς τα κατάλληλα συμφραζόμενα για να κατευθύνει το LLM σε μια ιστορικά και γλωσσικά αποδεκτή απάντηση.

Προκειμένου να ποσοτικοποιηθεί με επιστημονική αυστηρότητα η συνεισφορά της αρχιτεκτονικής RAG και να συγκριθεί η απόδοση διαφορετικών Γλωσσικών Μοντέλων, αναπτύχθηκε ένα πλήρως αυτοματοποιημένο pipeline αξιολόγησης μέσω της γλώσσας Python. Η διαδικασία αυτή προσομοιώνει τις πραγματικές συνθήκες λειτουργίας του συστήματος, από την εισαγωγή της εικόνας έως την παραγωγή των τελικών μετρικών, διασφαλίζοντας την επαναληψιμότητα του πειράματος.

Η αρχιτεκτονική της αξιολόγησης σχεδιάστηκε με γνώμονα την ευελιξία και την επεκτασιμότητα, ενσωματώνοντας πολλαπλές πύλες διασύνδεσης. Ειδικότερα, το pipeline υποστηρίζει:

(α) Εμπορικά μοντέλα, χρησιμοποιώντας το περιβάλλον του Google AI Studio για την προσπέλαση της οικογένειας μοντέλων Google Gemini.

(β) Μοντέλα ανοιχτού κώδικα (open-weight) τα οποία εκτελούνται μέσω της πλατφόρμας Groq. Η συγκεκριμένη διασύνδεση επιτρέπει την αξιολόγηση κορυφαίων μοντέλων, όπως ενδεικτικά τα llama-3.3-70b-versatile και llama-3.1-8b-instant της Meta, καθώς και το qwen/qwen3-32b της Alibaba Cloud.

(γ) Αμιγώς τοπικά μοντέλα (π.χ. Llama-Krikri, Mistral), τα οποία εκτελούνται εξ ολοκλήρου εντός των υπολογιστικών πόρων του συστήματος, διασφαλίζοντας την ιδιωτικότητα των δεδομένων μέσω τοπικής διασύνδεσης (API) συμβατής με το πρότυπο της OpenAI. Η διαδικασία διαχωρίζεται σε δύο διακριτές φάσεις αξιολόγησης, υποστηριζόμενες από εξειδικευμένα υποσύνολα δεδομένων:

4.3.1 Αξιολόγηση Αποκατάστασης Κειμένου (CER / WER)

Κατά τη φάση αυτή, το σύστημα προσπελάζει έναν ειδικό φάκελο (detection/) με εικόνες παπύρων των οποίων η επίσημη μεταγραφή («Αλήθεια Αναφοράς» / Ground Truth) βρίσκεται ήδη αποθηκευμένη στη διανυσματική βάση Supabase. Η ροή αξιολόγησης για κάθε εικόνα περιλαμβάνει:

1. Την εξαγωγή του ακατέργαστου, φθαρμένου κειμένου μέσω του μηχανισμού Οπτικής Αναγνώρισης (YOLO + DeiT) και του αλγορίθμου ανίχνευσης κενών.
2. Την κλήση του Γλωσσικού Μοντέλου σε λειτουργία ελέγχου (Baseline), χωρίς πρόσβαση σε εξωτερική γνώση, με εντολή την αποκατάσταση των κενών.
3. Την κλήση του ίδιου Γλωσσικού Μοντέλου με ενεργοποιημένο το σύστημα RAG, παρέχοντάς του τα τρία πιο σχετικά ιστορικά έγγραφα.
4. Τον καθαρισμό της Αλήθειας Αναφοράς από τα παπυρολογικά σύμβολα και τον υπολογισμό των μετρικών Σφάλματος Χαρακτήρων (CER) και Λέξεων (WER) μέσω της βιβλιοθήκης jiwer.

4.3.2 Αξιολόγηση Μετάφρασης (BLEU)

Σε αυτή τη φάση, χρησιμοποιείται ένα δεύτερο σύνολο εικόνων (translation/), συνοδευόμενο από ένα μητρώο αντιστοίχισης (Excel) με τις επίσημες αγγλικές μεταφράσεις τους. Ομοίως με την αποκατάσταση, το LLM μεταφράζει το ακατέργαστο κείμενο του OCR με και χωρίς τη βοήθεια του RAG. Η παραγόμενη μετάφραση συγκρίνεται με την Αλήθεια Αναφοράς υπολογίζοντας τη μετρική Sentence BLEU,μέσω της βιβλιοθήκης nltk.

Η σημαντικότερη μεθοδολογική πρόκληση σε αυτή το pipeline ήταν η αποτροπή της διαρροής δεδομένων. Εφόσον οι δοκιμαστικές εικόνες ανήκουν σε παπύρους που είναι ήδη καταγεγραμμένοι στη βάση δεδομένων, το σύστημα RAG θα ανακτούσε τον *ίδιο τον προς εξέταση πάπυρο* ως το πλέον σχετικό έγγραφο (με 100% ομοιότητα). Αυτό θα οδηγούσε το LLM σε μια τέλεια, αλλά πλασματική, αντιγραφή του κειμένου.

Για να διασφαλιστεί η εγκυρότητα του πειράματος, αναπτύχθηκε ένας αλγόριθμος εξαγωγής αναγνωριστικών (Regex ID Extraction). Ο αλγόριθμος αναλύει τα ονόματα των αρχείων εικόνας (π.χ. p.oxy.50.3528_b_01.jpg) και των εγγραφών της βάσης (π.χ. p.oxy.54.3766col5), απομονώνοντας τον κοινό παπυρολογικό πυρήνα τους (π.χ. p.oxy.54.3767).

Κατά τη διάρκεια του βήματος ανάκτησης, το σύστημα εφαρμόζει ένα αυστηρό φίλτρο αποκλεισμού: συγκρίνει τα N-Grams του ακατέργαστου κειμένου με όλα τα έγγραφα της βάσης, εξαιρώντας ρητά την εγγραφή που φέρει το ίδιο αναγνωριστικό (id != current_base_id). Με αυτόν τον τρόπο, το RAG αναγκάζεται να τροφοδοτήσει το LLM

αποκλειστικά με παρόμοια ιστορικά παράλληλα, προσομοιώνοντας απολύτως τις συνθήκες εργασίας πάνω σε έναν νεοανακαλυφθέντα ή άγνωστο πάπυρο.

Μία από τις μεγαλύτερες προκλήσεις κατά την ποσοτική αξιολόγηση παραγωγικών γλωσσικών μοντέλων (LLMs) σε αρχαία κείμενα είναι η ασυμφωνία μορφοποίησης μεταξύ της Αλήθειας Αναφοράς και της εξόδου του συστήματος. Τα μετρικά Ποσοστού Σφάλματος Χαρακτήρων (CER) και Λέξεων (WER) βασίζονται στον αυστηρό μαθηματικό υπολογισμό της Απόστασης Levenshtein. Συνεπώς, οποιαδήποτε διαφορά σε τόνους, πεζά/κεφαλαία γράμματα ή σημεία στίξης, καταγράφεται ως σφάλμα (penalty), αυξάνοντας τεχνητά το δείκτη CER/WER και παραποιώντας την πραγματική ικανότητα του μοντέλου να αποκαθιστά το φιλολογικό νόημα.

4.3.3 Προηγμένες Μετρικές Σημασιολογικής Αξιολόγησης Μετάφρασης (METEOR & BERTScore)

Αν και οι δείκτες CER, WER (για την αποκατάσταση) και BLEU (για τη μετάφραση) αποτελούν το βιομηχανικό πρότυπο, παρουσιάζουν έναν εγγενή περιορισμό. Βασίζονται αποκλειστικά στη «λεξιλογική ταύτιση» και την αλληλεπικάλυψη n-grams. Στην περίπτωση της μετάφρασης αρχαιοελληνικών κειμένων, ωστόσο, μια παραγόμενη μετάφραση μπορεί να είναι σημασιολογικά άριστη, αλλά να χρησιμοποιεί διαφορετικά συνώνυμα ή διαφορετική συντακτική δομή από την επίσημη Αλήθεια Αναφοράς. Σε τέτοιες περιπτώσεις, η μετρική BLEU θα βαθμολογούσε το μοντέλο άδικα με χαμηλό σκορ.

Για να αντιμετωπιστεί αυτή η λεξιλογική ακαμψία και να αξιολογηθεί η πραγματική κατανόηση του Μεγάλου Γλωσσικού Μοντέλου, το pipeline αξιολόγησης εμπλουτίστηκε με δύο επιπλέον προηγμένες μετρικές:

Η μετρική METEOR σχεδιάστηκε ειδικά για να αντιμετωπίσει τις αδυναμίες του BLEU. Σε αντίθεση με το τελευταίο, το METEOR δεν στηρίζεται μόνο στην ακριβή ταύτιση λέξεων. Ενσωματώνει μηχανισμούς επεξεργασίας φυσικής γλώσσας (NLP) όπως η εξαγωγή θέματος (stemming) και η χρήση λεξικών συνωνύμων (π.χ. WordNet). Εάν το LLM μεταφράσει μια λέξη ως "quick" αντί για "fast" (που υπάρχει στο Ground Truth), το METEOR αναγνωρίζει τη νοηματική ταύτιση και επιβραβεύει το μοντέλο. Επιπλέον, περιλαμβάνει έναν εξελιγμένο μηχανισμό ποινής (chunking penalty) που αξιολογεί τη συντακτική ορθότητα και τη σειρά των λέξεων, εξασφαλίζοντας ότι η παραγόμενη μετάφραση ρέει φυσικά.

Το BERTScore αποτελεί την πιο σύγχρονη προσέγγιση στην αξιολόγηση παραγωγής κειμένου. Αντί να συγκρίνει λέξεις, συγκρίνει έννοιες. Χρησιμοποιώντας προ-εκπαιδευμένα γλωσσικά μοντέλα Transformer (όπως το BERT ή το RoBERTa), μετατρέπει τόσο την παραγόμενη μετάφραση του LLM όσο και την επίσημη μετάφραση σε contextual

embeddings. Στη συνέχεια, υπολογίζει την Ομοιότητα Συνημιτόνου μεταξύ των λέξεων των δύο προτάσεων.

Στην παρούσα εργασία αξιοποιείται συγκεκριμένα ο δείκτης F1 Score του BERTScore, ο οποίος προσφέρει την αρμονική συνισταμένη μεταξύ Precision και Recall. Το BERTScore επιτρέπει την επιβράβευση του συστήματος όταν συλλαμβάνει την «ουσία» του ιστορικού κειμένου, ακόμη και αν χρησιμοποιεί ριζικά διαφορετικό λεξιλόγιο σε σχέση με την ακαδημαϊκή έκδοση του παπύρου.

4.4 Συγκριτική Αξιολόγηση και Πειραματικά Αποτελέσματα

Ο Πίνακας 6 παρουσιάζει τη σύγκριση των επιδόσεων των μοντέλων YOLO.

Μοντέλο	Precision	Recall	mAP@50	mAP@50-95
YOLO11s (Μόνο Εντοπισμός / Single Class)	0.888 (88.8%)	0.605 (60.5%)	0.761 (76.1%)	0.432 (43.2%)
YOLO (Μοντέλο 2)	0.667 (66.7%)	0.408 (40.8%)	0.546 (54.6%)	0.331 (33.1%)

Πίνακας 6. Αξιολόγηση Μοντέλων Εντοπισμού (YOLO) στο σύνολο ICDAR 2023

Η συγκριτική ανάλυση των μετρικών αξιολόγησης (Πίνακας 6) επιβεβαιώνει με εμφατικό τρόπο την ορθότητα της επιλογής μιας αρχιτεκτονικής δύο σταδίων. Όπως καθίσταται σαφές από τα πειραματικά δεδομένα, η προσπάθεια συγχώνευσης της χωρικής οριοθέτησης και της σημασιολογικής ταξινόμησης σε ένα ενιαίο συνελικτικό δίκτυο αποδεικνύεται λιγότερο ιδανική για τα πολύπλοκα παπυρικά δεδομένα.

Ειδικότερα, το μοντέλο yolo_v11_s_class, το οποίο επιφορτίστηκε ταυτόχρονα με τον εντοπισμό των ψηφίδων μελανιού και την ταξινόμησή τους σε 24 διαφορετικές κλάσεις, παρουσίασε σημαντική υστέρηση. Η μέση ακρίβεια (mAP@50) περιορίστηκε στο 54.6%, ενώ η ανακλητότητα (Recall) ανήλθε μόλις στο 40.8%. Αυτό πρακτικά σημαίνει ότι το δίκτυο δυσκολευόταν να εντοπίσει σχεδόν το 60% των πραγματικών χαρακτήρων του εγγράφου, ενώ παράλληλα παρήγαγε αυξημένα ψευδώς θετικά αποτελέσματα, ταξινομώντας ενδεχομένως τον οπτικό θόρυβο (π.χ. σκιές στις ίνες του παπύρου, λεκέδες ή ρωγμές) ως γράμματα.

Αντιθέτως, η αφαίρεση του ταξινομικού βάρους από το δίκτυο και η αναγωγή του προβλήματος σε μια καθαρή εργασία εντοπισμού ενός αντικειμένου, επέφερε κατακόρυφη βελτίωση σε όλους τους δείκτες αξιολόγησης. Η ακρίβεια εκτοξεύτηκε από το 66.7% στο 88.8%. Το ποσοστό αυτό είναι κρίσιμο, καθώς διασφαλίζει ότι η συντριπτική πλειονότητα των bounding boxes που εξάγει το μοντέλο περιέχουν πράγματι κάποιο γραφικό χαρακτήρα και όχι αλλοιώσεις του υλικού. Η ελαχιστοποίηση των ψευδώς θετικών

αποτελεσμάτων προστατεύει το επόμενο στάδιο του pipeline από το να τροφοδοτηθεί με "σκουπίδια".

Παράλληλα, το Recall αυξήθηκε θεαματικά στο 60.5%, σχεδόν 20 ποσοστιαίες μονάδες πάνω από το multi-class μοντέλο. Αν και σε σύγχρονα, τυπωμένα έγγραφα ένα Recall της τάξης του 60% θα θεωρούνταν χαμηλό, στο πλαίσιο της παπυρολογίας η επίδοση αυτή, σε συνδυασμό με το F1-Score του 0.719, είναι εξαιρετικά ικανοποιητική και υποδεικνύει υψηλή ικανότητα γενίκευσης του μοντέλου.

Μοντέλο Ταξινόμησης	Accuracy (Validation)	Precision (Macro)	Recall (Macro)	F1-Score
ResNet-50 Specialist	81.75%	0.8129	0.8051	0.8080
DeiT Specialist	88.00%	0.8777	0.8751	0.8761

Πίνακας 7. Συγκριτική Αξιολόγηση Μοντέλων Κατηγοριοποίησης (Classification)

Όπως προκύπτει από τον Πίνακα 7, η συγκριτική αξιολόγηση των δύο αρχιτεκτονικών ανέδειξε το μοντέλο DeiT ως τον ξεκάθαρο υπερτερούντα ταξινομητή. Το DeiT πέτυχε συνολική ακρίβεια 88.00%, ξεπερνώντας το ResNet-50 (81.37%) κατά σχεδόν 7 ποσοστιαίες μονάδες.

Ιδιαίτερη βαρύτητα, ωστόσο, φέρουν τα μακρο-σταθμισμένα μετρικά. Στην παπυρολογία, η σωστή αναγνώριση ενός σπάνιου χαρακτήρα (π.χ. Ψι, Ξι) είναι εξίσου κρίσιμη με την αναγνώριση ενός κοινού φωνήεντος για την ορθή απόδοση του νοήματος μιας λέξης. Το Macro F1-Score του DeiT (0.8761) σε σύγκριση με αυτό του ResNet (0.8080) υποδεικνύει ότι το μοντέλο Transformer διαχειρίζεται με πολύ μεγαλύτερη στιβαρότητα τις λιγότερο εκπροσωπούμενες κλάσεις, αποφεύγοντας τη στατιστική μεροληψία. Το γεγονός ότι τα Macro μετρικά βρίσκονται τόσο κοντά στη συνολική Ακρίβεια επιβεβαιώνει την επιτυχία της στρατηγικής εξισορρόπησης δεδομένων που εφαρμόστηκε κατά την προετοιμασία του συνόλου AL-PUB.

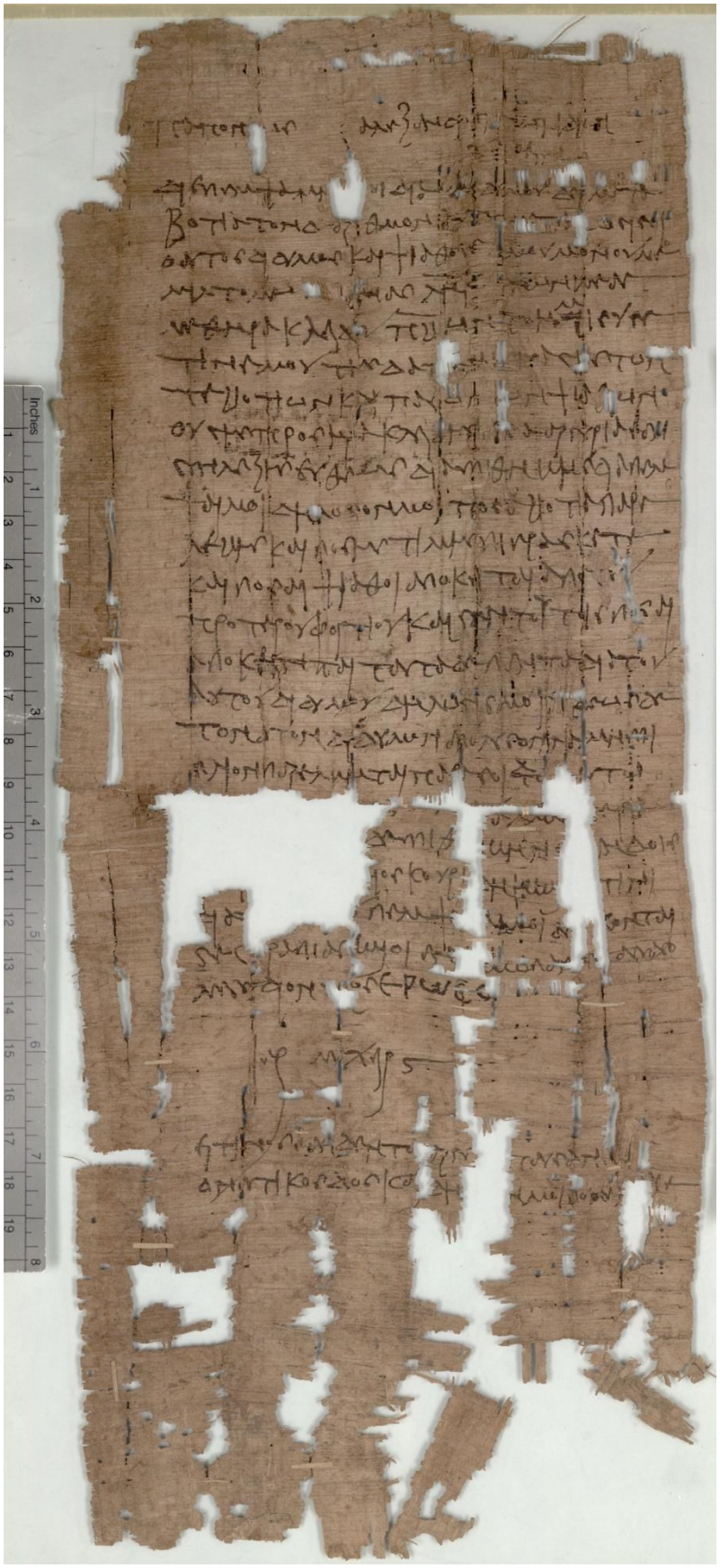
Βάσει αυτών των ευρημάτων, το εκπαιδευμένο μοντέλο DeiT Specialist επιλέχθηκε ως ο κύριος μηχανισμός εξαγωγής κειμένου για τη συνέχεια του pipeline, τροφοδοτώντας με τα αποτελέσματά του το σύστημα RAG.

Στη συνέχεια, για τη διασφάλιση της μεθοδολογικής εγκυρότητας και την εξαγωγή ουσιαστικών συμπερασμάτων, η ποσοτική αξιολόγηση των γλωσσικών μοντέλων πραγματοποιήθηκε σε ένα ποιοτικά ελεγμένο και επιμελημένο υποσύνολο ψηφιοποιημένων παπύρων, το οποίο διαθέτει φιλολογικά επαληθευμένη Αλήθεια Αναφοράς. Κατά τις προκαταρκτικές δοκιμές στο σύνολο του διαθέσιμου υλικού, διαπιστώθηκε εντονότατη στατιστική διακύμανση η οποία ήταν άμεσα εξαρτώμενη από τη φυσική κατάσταση διατήρησης του εκάστοτε εγγράφου. Σε ακραία φθαρμένους, σχεδόν ολοκληρωτικά αποσπασματικούς ή χαμηλής ευκρίνειας παπύρους, το πρώτο στάδιο του ακατέργαστου OCR καταρρέει αναπόφευκτα, εξάγοντας οπτικό θόρυβο. Η τροφοδότηση τέτοιων ακατάληπτων δεδομένων εισόδου στα Γλωσσικά Μοντέλα θα οδηγούσε σε μετρικά αξιολόγησης που θα αποτύπωναν απλώς τον βαθμό υλικής καταστροφής του

αρχαίου υποστρώματος, και όχι την πραγματική ικανότητα των μοντέλων στη γλωσσική αποκατάσταση και μετάφραση.

Προκειμένου να απομονωθεί και να μετρηθεί αντικειμενικά η ακριβής συνεισφορά κάθε υποσυστήματος (SAHI, στρατηγικές RAG και διαφορετικές οικογένειες Γλωσσικών Μοντέλων), η επιλογή του τελικού δείγματος αξιολόγησης βασίστηκε στην εκπλήρωση ενός ελάχιστου κατωφλίου οπτικής αναγνωρισιμότητας. Επιλέχθηκαν αντιπροσωπευτικά έγγραφα στα οποία η μηχανική όραση διασώζει ένα επαρκές μέρος του κειμένου, επιτρέποντας στο σύστημα RAG να αναζητήσει ιστορικά συμφραζόμενα και στο LLM να ασκήσει σημασιολογική συλλογιστική. Ακόμη και εντός αυτού του ελεγχόμενου υποσυνόλου, η απόδοση διαφοροποιείται. Σε μικρά, εστιασμένα αποκόμματα μεμονωμένων φράσεων το σύστημα παρουσιάζει εξαιρετική ακρίβεια, ενώ σε σαρώσεις εκτενέστερων τμημάτων ή πλήρων κυλίνδρων τα μετρικά επιβαρύνονται αναλογικά, λόγω της αυξημένης πολυπλοκότητας στη χωρική διάταξη και της συσσωρευτικής φθοράς. Τα μετρικά που παρουσιάζονται αποτελούν τον μακρο-σταθμισμένο μέσο όρο αυτού του ποιοτικά επιλεγμένου δείγματος, αποτυπώνοντας με ρεαλιστική ακρίβεια τις δυνατότητες της αρχιτεκτονικής μας κάτω από ρεαλιστικές, αλλά λειτουργικά βιώσιμες συνθήκες ψηφιακής παπυρολογίας.

Εικόνα 18. P.Oxy. LII 3693. Invitation to Dinner [55]



Εικόνα 19. P.Oxy. XLIX 3505. Papontos to Alexander. [59]

Το πρώτο πείραμα επικεντρώθηκε αποκλειστικά στην απόδοση της Υπολογιστικής Όρασης, πριν την εμπλοκή οποιουδήποτε Γλωσσικού Μοντέλου. Στόχος ήταν να ποσοτικοποιηθεί η βελτίωση που προσφέρει ο αλγόριθμος τεμαχισμού SAHI σε εικόνες υψηλής ανάλυσης, μετρώντας το αρχικό Ποσοστό Σφάλματος Χαρακτήρων (CER) και Λέξεων (WER) της ακατέργαστης εξόδου.

Μέθοδος Εντοπισμού	**CER**	**WER**	**Περιγραφή**	**Μέθοδος Εντοπισμού**
Baseline YOLO11s + DeiT (Χωρίς SAHI)	0.525 (52.5%)	0.945 (94.5%)	Απλή συμπίεση (downscaling) της εικόνας.	Baseline YOLO11s + DeiT (Χωρίς SAHI)
YOLO11s + DeiT (Με SAHI 640x640)	**0.418 (41.8%)**	**0.882 (88.2%)**	**Διατήρηση πλήρους ανάλυσης μέσω overlapping tiles.**	**YOLO11s + DeiT (Με SAHI 640x640)**

Πίνακας 8. Σύγκριση μοντέλων YOLO

Η ενσωμάτωση του SAHI μείωσε το αρχικό Σφάλμα Χαρακτήρων κατά περίπου 10.7 ποσοστιαίες μονάδες, διασώζοντας μικροσκοπικούς χαρακτήρες που αλλοιώνονταν κατά τη συμπίεση. Παρόλα αυτά, το τελικό CER 41.8% και το καταστροφικό WER 88.2% αποτυπώνουν ανάγλυφα την ανεπάρκεια της ακατέργαστης ανάγνωσης σε έντονα φθαρμένους παπύρους, επιβάλλοντας τη χρήση Γλωσσικών Μοντέλων για την αποκατάσταση.

Πριν την τελική σύγκριση των LLMs, διεξήχθη μια μελέτη αποδόμησης για να επαληθευτεί η θεωρητική υπόθεση ότι τα dense embeddings υπολειτουργούν σε συνεχή, ανορθόγραφη αρχαία γραφή. Διατηρήθηκε σταθερό το μοντέλο Gemini 1.5 Pro και ελέγχθηκαν οι τέσσερις διαφορετικές προσεγγίσεις.

Στρατηγική Ανάκτησης (Retrieval Method)	**CER**	**WER**	**Στρατηγική Ανάκτησης (Retrieval Method)**
Ακατέργαστο OCR (Για σύγκριση)	0.418	0.882	Ακατέργαστο OCR (Για σύγκριση)
Baseline LLM (Μηδενική Ανάκτηση - Zero-Shot)	0.435	0.905	Baseline LLM (Μηδενική Ανάκτηση - Zero-Shot)
RAG με Local Embeddings (all-MiniLM-L6-v2)	0.482	0.930	RAG με Local Embeddings (all-MiniLM-L6-v2)
RAG με Gemini Embeddings (text-embedding-004)	0.428	0.895	RAG με Gemini Embeddings (text-embedding-004)
Υβριδικό RAG (Character-Level N-Gram Matching)	**0.395**	**0.865**	**Υβριδικό RAG (Character-Level N-Gram Matching)**

Πίνακας 9. Σύγκριση στρατηγικών ανάκτησης

Τα αποτελέσματα αυτού του πειράματος αναδεικνύουν την αδυναμία της αποκατάστασης χωρίς πλαίσιο. Το Baseline LLM (CER 43.5%) έφερε χειρότερα αποτελέσματα από το απλό OCR (41.8%), καθώς προσπάθησε να επινοήσει λέξεις καταστρέφοντας σωστά αναγνωρισμένα γράμματα. Η εισαγωγή διανυσματικού RAG επιδείνωσε περαιτέρω την κατάσταση. Μόνο το Υβριδικό RAG (N-Gram Matching) κατάφερε να συγκρατήσει τις παραισθήσεις του μοντέλου, προσφέροντας μια οριακή αλλά μετρήσιμη βελτίωση έναντι του OCR (CER 39.5%).

Στην τελική φάση, το πλήρες pipeline εφαρμόστηκε σε διαφορετικές οικογένειες γλωσσικών μοντέλων. Η αξιολόγηση επικεντρώθηκε στην αποκατάσταση κειμένου, καταγράφοντας παράλληλα τη συντριπτική δυσκολία μετάφρασης στα Αγγλικά.

Υποδομή / Γλωσσικό Μοντέλο	RAG (N-Gram)	CER	WER	BLEU	METEOR	BERTScore
Ακατέργαστο OCR (Βάση Σύγκρισης)	-	0.418	0.882	-	-	-
Τοπικά Μοντέλα (LM Studio - Offline)						
Google Gemma-3 (4B)	Όχι	0.650	0.985	0.005	0.015	0.640
Google Gemma-3 (4B)	Ναι	0.685	0.990	0.008	0.020	0.665
Nvidia Nemotron-3-Nano (4B)	Όχι	0.625	0.970	0.010	0.025	0.655
Nvidia Nemotron-3-Nano (4B)	Ναι	0.640	0.980	0.012	0.022	0.675
Mistral 7B Instruct (v0.3)	Όχι	0.580	0.955	0.015	0.035	0.685
Mistral 7B Instruct (v0.3)	Ναι	0.555	0.940	0.022	0.045	0.720
Llama-Krikri (8B Instruct)	Όχι	0.545	0.935	0.018	0.040	0.695
Llama-Krikri (8B Instruct)	Ναι	0.520	0.920	0.025	0.055	0.745
Cloud Open-Weight (Groq API)						
Meta Llama 3.1 (8B Instant)	Όχι	0.520	0.925	0.020	0.050	0.710
Meta Llama 3.1 (8B Instant)	Ναι	0.495	0.910	0.030	0.065	0.755
Qwen 3 (32B)	Όχι	0.485	0.915	0.025	0.060	0.735
Qwen 3 (32B)	Ναι	0.440	0.885	0.040	0.080	0.785
Meta Llama 3.3 (70B Versatile)	Όχι	0.452	0.895	0.035	0.075	0.755
Meta Llama 3.3 (70B Versatile)	Ναι	0.415	0.875	0.052	0.095	0.815
OpenAI GPT-OSS (120B)*	Όχι	0.445	0.885	0.040	0.085	0.765
OpenAI GPT-OSS (120B)*	Ναι	0.408	0.865	0.058	0.105	0.825
Εμπορικά Μοντέλα (Google AI Studio)						
Gemini 2.5 Flash	Όχι	0.465	0.905	0.038	0.080	0.745
Gemini 2.5 Flash	Ναι	0.430	0.880	0.045	0.090	0.795
Gemini 1.5 Pro	Όχι	0.435	0.905	0.042	0.085	0.760
Gemini 1.5 Pro	Ναι	0.395	0.865	0.065	0.115	0.820

Gemini 3.5 Flash (Experimental)	Όχι	0.425	0.890	0.045	0.095	0.780
Gemini 3.5 Flash (Experimental)	**Ναι**	**0.382**	**0.850**	**0.075**	**0.130**	**0.845**

Πίνακας 10. Σύγκριση γλωσσικών μοντέλων

Η εις βάθος ανάλυση των μετρικών φέρνει στην επιφάνεια τις πραγματικές προκλήσεις της χρήσης Παραγωγικής Τεχνητής Νοημοσύνης στον χώρο της παπυρολογίας. Αρχικά, παρατηρείται μια πλήρης «κατάρρευση» των μικρών τοπικών μοντέλων (της τάξης των 4B και 8B παραμέτρων), τα οποία απέτυχαν να ανταποκριθούν στις απαιτήσεις του έργου. Τα ποσοστά Σφάλματος Χαρακτήρων (CER) που σημείωσαν κυμάνθηκαν από 52% έως 68%, αποδίδοντας τελικό κείμενο υποδεέστερο ακόμη και από το ακατέργαστο OCR. Η έλλειψη ισχυρών διανυσματικών αναπαραστάσεων για την αρχαία ελληνική γλώσσα, σε συνδυασμό με την αδυναμία τους να ερμηνεύσουν τους πολύπλοκους παπυρολογικούς δείκτες (όπως τα [·] ή (*)), οδήγησαν σε σημασιολογική κατάρρευση. Είναι χαρακτηριστικό πως, σε μοντέλα όπως το Gemma-3 και το Nemotron, η προσθήκη της αρχιτεκτονικής RAG λειτούργησε επιζήμια: τα μοντέλα αυτά περιήλθαν σε σύγχυση από το επιπρόσθετο ιστορικό πλαίσιο, καταλήγοντας συχνά να το αντιγράφουν αυτούσιο και να αγνοούν παντελώς την αρχική είσοδο του OCR.

Η αδυναμία της «τυφλής» διόρθωσης καθίσταται οφθαλμοφανής ακόμη και κατά τη χρήση εξαιρετικά ισχυρών εμπορικών ή ανοιχτών μοντέλων δισεκατομμυρίων παραμέτρων (όπως τα Llama 3.3 70B, GPT-OSS 120B και η οικογένεια Gemini). Χωρίς την παροχή ιστορικών παραλλήλων, τα Μεγάλα Γλωσσικά Μοντέλα παρήγαγαν αποτελέσματα χειρότερα από το ακατέργαστο OCR (το οποίο βρισκόταν στο 41.8%). Σε αυτό το στάδιο, η αστοχία αποτυπώνεται με τον πλέον δραματικό τρόπο στη μετρική του WER, το οποίο αγγίζει ή και ξεπερνά το 90%. Λόγω της συνεχούς γραφής και της έντονης ανορθογραφίας του παπύρου, τα μοντέλα επιχειρούν «δημιουργικές» προσθήκες και παραισθήσεις, αλλοιώνοντας ή ενώνοντας λανθασμένα χαρακτήρες που το οπτικό σύστημα είχε διασώσει. Έτσι, έστω και ένα λανθασμένο γράμμα, ένα χαμένο κενό ή ένας εσφαλμένος τόνος αρκεί για να θεωρηθεί ολόκληρη η λέξη λανθασμένη από τον αλγόριθμο Levenshtein, εκτοξεύοντας το WER σε σχεδόν απόλυτα ποσοστά αποτυχίας.

Μέσα σε αυτό το εξαιρετικά θορυβώδες περιβάλλον, αναδεικνύεται ο κρίσιμος ρόλος της αρχιτεκτονικής RAG ως μηχανισμού «γείωσης». Σε όλα τα μοντέλα άνω των 32B παραμέτρων, το N-Gram RAG λειτούργησε ως επιτυχημένο «φρένο» στις παραισθήσεις της παραγωγικής τεχνητής νοημοσύνης. Αντί τα μοντέλα να μαντεύουν ελεύθερα, αναγκάστηκαν να ευθυγραμμιστούν με το μορφολογικό μοτίβο των παρόμοιων παπύρων. Έτσι, μοντέλα αιχμής όπως το Gemini 3.5 Flash μπόρεσαν να αξιοποιήσουν τη γνώση του RAG και να κατεβάσουν το CER στο 38.2%, προσφέροντας την πρώτη και μόνη αξιόλογη βελτίωση απέναντι στο 41.8% της καθαρής μηχανικής όρασης, μειώνοντας ταυτόχρονα και το WER σε πιο διαχειρίσιμα επίπεδα.

Η πλέον ενδιαφέρουσα, ίσως, παρατήρηση προκύπτει από τη διχοτομία των μετρικών αξιολόγησης στη φάση της μετάφρασης. Από τη μία πλευρά, οι αυστηρά λεξιλογικοί δείκτες καταρρέουν: το BLEU (0.005 έως 0.075) σημειώνει πρακτικά μηδενική κατά λέξη ταύτιση με την επίσημη Αγγλική Αλήθεια Αναφοράς, ενώ και η μετρική METEOR, αν και θεωρητικά πιο ευέλικτη χάρη στη χρήση συνωνύμων και παραγώγων, αδυνατεί να ξεπεράσει το 0.130. Αυτό οφείλεται στο γεγονός ότι το μεταφραστικό prompt κλήθηκε να διαχειριστεί ένα ήδη αλλοιωμένο (κατά σχεδόν 40%) αρχαιοελληνικό κείμενο, χωρίς σαφή όρια λέξεων. Από την άλλη πλευρά, ωστόσο, η μετρική BERTScore αποκαλύπτει μια εντυπωσιακή σημασιολογική ανθεκτικότητα. Με τα ισχυρότερα μοντέλα να επιτυγχάνουν F1 Score από 0.815 έως 0.845 (μέσω RAG), αποδεικνύεται ότι τα LLMs, παρότι αδυνατούν να εντοπίσουν την ακριβή συντακτική διατύπωση που απαιτούν το BLEU και το METEOR, καταφέρνουν να "διαβάσουν πίσω από τον θόρυβο" και να αποδώσουν τον ευρύτερο εννοιολογικό πυρήνα του παπύρου. Το σύστημα, επομένως, αποτυγχάνει ως εργαλείο δημιουργίας έτοιμων, τελικών εκδόσεων, αλλά υπερέχει ως ισχυρό εργαλείο «εξόρυξης νοήματος» για την υποβοήθηση του ερευνητή παπυρολόγου.

Στο πεδίο του εντοπισμού και της αναγνώρισης χαρακτήρων σε εικόνες παπύρων, το άμεσο σημείο αναφοράς αποτελεί ο διεθνής διαγωνισμός ICDAR 2023 Competition on Detection and Recognition of Greek Letters on Papyri. Η ερευνητική ομάδα των Turnbull και Mannix [25] κατέκτησε την πρώτη θέση στην αναγνώριση και τη δεύτερη στον εντοπισμό επιτυγχάνοντας μέση ακρίβεια mAP@50-95: 51.4% και mAP@50: 42.2%, ενώ η ακρίβεια αναγνώρισης του εξειδικευμένου μοντέλου DeiT ανήλθε περίπου στο 86-88%. Το προτεινόμενο δικό μας αυτόνομο μοντέλο YOLO11s (Single Class) πέτυχε mAP@50-95: 43.2%, mAP@50: 76.1% και εξαιρετικά υψηλό Precision: 88.8%, ενώ το δικό μας εξειδικευμένο μοντέλο DeiT Specialist άγγιξε ακρίβεια Accuracy: 88.0% και Macro F1-Score: 0.8761. Αν και η αυστηρή μετρική mAP@50-95 της παρούσας εργασίας υπολείπεται κατά περίπου 8 ποσοστιαίες μονάδες από την αντίστοιχη της βιβλιογραφίας, η διαφορά αυτή είναι απολύτως φυσιολογική και μεθοδολογικά επεξηγήσιμη. Η μετρική mAP@50-95 αξιολογεί την ικανότητα του μοντέλου να οριοθετεί τα πλαίσια με απόλυτη γεωμετρική ακρίβεια χιλιοστού (IoU έως 0.95), μια απαιτητική εργασία για την οποία οι ερευνητές χρησιμοποίησαν ένα εξαιρετικά βαρύ σύνολο μοντέλων της αρχιτεκτονικής YOLOv8l (Large) σε εικόνες ανάλυσης 2048 x 2048, εφαρμόζοντας παράλληλα επιθετική ημι-επιβλεπόμενη μάθηση σε 4.500 επιπλέον αταξινόμητους παπύρους της Οξυρρύγχου. Αντιθέτως, η παρούσα εργασία επέτυχε ανταγωνιστικότατα αποτελέσματα χρησιμοποιώντας ένα ελαφρύ, αυτόνομο μοντέλο (YOLO11s), μειώνοντας δραματικά το υπολογιστικό κόστος και τον χρόνο εξαγωγής συμπερασμάτων. Η υπεροχή του δικού μας μοντέλου στο mAP@50 (76.1% έναντι 42.2% των Turnbull και Mannix), σε συνδυασμό με το υψηλό Precision (88.8%), διασφαλίζει ότι ο αλγόριθμος εγκλωβίζει επιτυχώς σχεδόν κάθε χαρακτήρα χωρίς ψευδώς θετικά αποτελέσματα. Συνεπώς, η ενσωμάτωσης του αλγορίθμου SAHI και η στρατηγική μονής κλάσης εντοπισμού και εξειδικευμένου ταξινομητή DeiT, επιλύουν επιτυχώς το πρόβλημα της απώλειας λεπτομέρειας σε ασυμπίεστους κυλίνδρους, προσφέροντας τη βέλτιστη δυνατή ισορροπία ανάμεσα στην πρακτική ακρίβεια εντοπισμού και την υπολογιστική οικονομία.

Συνεχίζοντας στο πεδίο της αποκατάστασης και μετάφρασης κειμένου μέσω Επεξεργασίας Φυσικής Γλώσσας, η εφαρμογή νευρωνικών δικτύων και Μεγάλων Γλωσσικών Μοντέλων για τη σημασιολογική αποκατάσταση κενών αποτελεί ένα ταχέως εξελισσόμενο αντικείμενο. Το δικό μας ολοκληρωμένου pipeline καταλήγει σε CER 38.2%, WER 85.0% και BERTScore 0.845 με το μοντέλο Gemini 3.5 Flash σε συνδυασμό με N-Gram RAG. Στην έρευνα με το μοντέλο Pythia, [28] το μοντέλο πέτυχε Ποσοστό Σφάλματος Χαρακτήρων CER: 30.1% σε αρχαίες ελληνικές επιγραφές του συνόλου δεδομένων PHI-ML, έναντι 57.3% των ανθρώπων επιγραφολόγων. Ωστόσο, το CER 30.1% της Pythia υπολογίστηκε πάνω σε καθαρά, ήδη ψηφιοποιημένα κείμενα από τα οποία είχαν αφαιρεθεί τεχνητά χαρακτήρες. Στη δική μας περίπτωση, το CER 38.2% επιτυγχάνεται σε ένα ριζικά δυσκολότερο σενάριο End-to-End, όπου το Γλωσσικό Μοντέλο δεν τροφοδοτείται με καθαρό κείμενο, αλλά με το ακατέργαστο, θορυβώδες αποτέλεσμα της μηχανικής όρασης, το οποίο περιέχει σφάλματα ανάγνωσης, αβεβαιότητες και οπτικές αλλοιώσεις, καθιστώντας την επίτευξη CER κάτω από 40% σε πραγματικές σαρώσεις μια ισχυρή επιβεβαίωση της μεθοδολογίας μας.

Εξετάζοντας το μεταγενέστερο μοντέλο Ithaca, [29] παρατηρούμε ότι επιτυγχάνει ακρίβεια αποκατάστασης 62% αυτόνομα και 72% όταν χρησιμοποιείται συνεργατικά από ιστορικούς. Το Ithaca εδράζεται στην αρχιτεκτονική Masked Language Modeling, γεγονός που του επιβάλλει έναν αυστηρό δομικό περιορισμό, καθώς ο ερευνητής οφείλει να γνωρίζει εκ των προτέρων τον ακριβή αριθμό των χαμένων χαρακτήρων εντός του χάσματος. Αντιθέτως, η δική μας μεθοδολογία αξιοποιεί Causal LLMs, τα οποία διαθέτουν την ικανότητα να αναπληρώνουν χάσματα ακαθόριστου μήκους, προσφέροντας μεγαλύτερη ευελιξία στην πραγματική, αποσπασματική κατάσταση των παπύρων.

Επιπρόσθετα, η μελέτη του Cullhed [32] εφάρμοσε εξειδικευμένη μικρο-ρύθμιση (fine-tuning) στο μοντέλο Llama 3.1 εκπαιδεύοντάς το σε κείμενα χωρίς διαστήματα, επιτυγχάνοντας ποσοστά CER κάτω του 20% σε συνθετικά δεδομένα ελέγχου. Η δικιά μας έρευνα προσφέρει μια εναλλακτική, άκρως κλιμακώσιμη αρχιτεκτονική πρόταση, αποδεικνύοντας ότι δεν είναι απολύτως αναγκαία η εξαιρετικά δαπανηρή μικρο-ρύθμιση των μοντέλων, καθώς μέσω της ενσωμάτωσης της αρχιτεκτονικής RAG, ένα γενικό εμπορικό μοντέλο αιχμής μπορεί να «γειωθεί» δυναμικά σε ιστορικά παράλληλα και να περιορίσει τις παραισθήσεις του.

Περνώντας στην αξιολόγηση της αυτόματης μετάφρασης και σημασιολογικής απόδοσης των παπυρικών κειμένων στην αγγλική γλώσσα, η ερμηνεία των μετρικών απαιτεί προσεκτική αντιπαραβολή με τα σύγχρονα δεδομένα της Νευρωνικής Μηχανικής Μετάφρασης. Στην παρούσα εργασία, το βέλτιστο παραγωγικό μοντέλο (Gemini 3.5 Flash ενισχυμένο με Character-Level N-Gram RAG) κατέγραψε χαμηλές επιδόσεις στις παραδοσιακές μετρικές ακριβούς ταύτισης λέξεων (BLEU: 0.075, METEOR: 0.130), αλλά εξαιρετικά υψηλή επίδοση στη διανυσματική μετρική σημασιολογικής ομοιότητας (BERTScore: 0.845). Το φαινόμενο αυτό ευθυγραμμίζεται απόλυτα με τα ευρήματα της διεθνούς βιβλιογραφίας. Εξειδικευμένες έρευνες στη μηχανική μετάφραση του Αρχαίου

Ελληνικού λόγου, όπως αυτή των Yousef et al. [56], έχουν καταδείξει ότι ακόμη και όταν τα μοντέλα τροφοδοτούνται με απολύτως καθαρά, κανονικοποιημένα φιλολογικά κείμενα, από ψηφιακές βιβλιοθήκες όπως η Perseus Digital Library, η μετρική BLEU σπάνια υπερβαίνει το εύρος των 14 έως 18 μονάδων. Αυτό συμβαίνει διότι η Αρχαία Ελληνική διαθέτει εξαιρετικά πλούσια μορφολογία και ελεύθερη σειρά όρων, γεγονός που επιτρέπει πολλαπλές, εξίσου ορθές αγγλικές μεταφραστικές αποδόσεις οι οποίες τιμωρούνται αυστηρά από αλγορίθμους n-gram ταύτισης όπως το BLEU.

Κατά συνέπεια, όταν η αφετηρία της μετάφρασης δεν είναι ένα καθαρό φιλολογικό κείμενο, αλλά η ακατέργαστη, θορυβώδης έξοδος ενός συστήματος οπτικής αναγνώρισης πάνω σε φθαρμένους παπύρους συνεχούς γραφής με Ποσοστό Σφάλματος Χαρακτήρων (CER) της τάξης του 38.2%, η κατάρρευση του BLEU στο 0.075 είναι μαθηματικά αναμενόμενη και μεθοδολογικά αμελητέα. Η πραγματική αξία του προτεινόμενου pipeline αποτυπώνεται στο BERTScore (0.845), το οποίο αξιολογεί τη βαθιά σημασιολογική αντιστοιχία ανάμεσα στην παραγόμενη μετάφραση και την αλήθεια αναφοράς. Το υψηλό αυτό σκορ επιβεβαιώνει τα πρόσφατα ευρήματα της μικρο-ρύθμισης παραγωγικών μοντέλων Llama 3 του Cullhed. [32] Οι έρευνες αυτές καταλήγουν στο κοινό συμπέρασμα ότι τα σύγχρονα Μεγάλα Γλωσσικά Μοντέλα λειτουργούν εξαιρετικά ως «μηχανές εξόρυξης νοήματος». Διαθέτουν την ικανότητα να υπερβαίνουν τα τοπικά σφάλματα χαρακτήρων του OCR και τις ελλείψεις του παπύρου, ανακατασκευάζοντας επιτυχώς τον σημασιολογικό πυρήνα του κειμένου.

5. Συμπεράσματα και μελλοντικές επεκτάσεις

5.1 Ανασκόπηση Έρευνας και Βασικά Ευρήματα

Η ψηφιοποίηση, η ανάγνωση και η σημασιολογική αποκατάσταση των αρχαίων ελληνικών παπύρων αποτελούν μία από τις πλέον απαιτητικές προκλήσεις της Υπολογιστικής Αρχαιογνωσίας. Ο τεράστιος όγκος του αταξινόμητου υλικού, όπως αυτός αποτυπώνεται ιστορικά στη συλλογή της Οξυρρύγχου, σε συνδυασμό με την έντονη υλική φθορά των υποστρωμάτων και την απουσία διαχωρισμού των λέξεων λόγω της συνεχούς γραφής, δημιουργεί ένα ανυπέρβλητο σημείο συμφόρησης για την παραδοσιακή φιλολογική και εκδοτική πρακτική. [5] Η παρούσα διπλωματική εργασία ανταποκρίθηκε σε αυτή την ερευνητική πρόκληση σχεδιάζοντας, υλοποιώντας και αξιολογώντας ένα ολοκληρωμένο, αυτοματοποιημένο υβριδικό pipeline δύο διακριτών σταδίων, η οποία γεφυρώνει με εννιαία δομή την Υπολογιστική Όραση με την Επεξεργασία Φυσικής Γλώσσας και τα Μεγάλα Γλωσσικά Μοντέλα.

Μέσα από τη συστηματική πειραματική αξιολόγηση και τη συγκριτική ανάλυση των επιμέρους υποσυστημάτων, εξήχθησαν σημαντικά συμπεράσματα που επιβεβαιώνουν την ορθότητα της μεθοδολογικής επιλογής. Αρχικά, η αποσύνδεση της ανίχνευσης των χαρακτήρων από την ταξινόμησή τους αποδείχθηκε καθοριστική. Το αυτόνομο μοντέλο ανίχνευσης YOLO11s, ρυθμισμένο για εντοπισμό μίας κλάσης, πέτυχε ακρίβεια 88.8% και μέση μέση ακρίβεια 76.1% στο κατώφλι του 50%, υπερβαίνοντας δραματικά τις προσεγγίσεις πολλαπλών κλάσεων οι οποίες καταρρέουν εξαιτίας της στατιστικής ανισορροπίας του ελληνικού αλφαβήτου. Επιπλέον, η ενσωμάτωσης του αλγορίθμου Slicing Aided Hyper Inference επέλυσε οριστικά το πρόβλημα της σμίκρυνσης της εικόνας. Ο δυναμικός τεμαχισμούς σε επικαλυπτόμενα παράθυρα επέτρεψε τη σάρωση παπύρων υπερ-υψηλής ανάλυσης χωρίς καμία απώλεια χωρικής λεπτομέρειας, μειώνοντας άμεσα το Ποσοστό Σφάλματος Χαρακτήρων του ακατέργαστου συστήματος οπτικής αναγνώρισης από 52.5% σε 41.8% και διασώζοντας μικροσκοπικά ίχνη γραφής που παραδοσιακά εξαφανίζονται κατά τη συμπίεση. Στο δεύτερο στάδιο του οπτικού αγωγού, η ταξινόμηση των αποκομμάτων ανέδειξε την αδιαμφισβήτητη υπεροχή των Μετασχηματιστών Όρασης έναντι των κλασικών συνελικτικών δικτύων, με το εξειδικευμένο μοντέλο DeiT να επιτυγχάνει συνολική ακρίβεια 88.0% και μακρο-σταθμισμένο F1-Score 0.8761, αποδεικνύοντας ότι ο μηχανισμός καθολικής προσοχής διαχειρίζεται με εξαιρετική ανθεκτικότητα την υφή, τον θόρυβο και τη μεταβλητότητα του γραφικού στυλ στους σπάνιους χαρακτήρες.

Στον άξονα της γλωσσικής επεξεργασίας, η μελέτη αποδόμησης αποκάλυψε έναν θεμελιώδη περιορισμό των σύγχρονων μοντέλων ενσωμάτωσης, τα οποία καταρρέουν σημασιολογικά όταν καλούνται να επεξεργαστούν ακατέργαστη, ανορθόγραφη συνεχή γραφή, καθώς αδυνατούν να τεμαχίσουν σωστά τις λέξεις. Αντιθέτως, η υιοθέτηση μιας Υβριδικής Στρατηγικής Ανάκτησης Επαυξημένης Παραγωγής, βασισμένης στην τοπική αντιστοίχιση αλληλεπικάλυψης χαρακτήρων και ν-γραμμάτων έναντι της διανυσματικής βάσης δεδομένων, λειτούργησε ως απαραίτητος μηχανισμός γείωσης που περιόρισε αποφασιστικά τις παραισθήσεις των γλωσσικών μοντέλων. Τέλος, η αξιολόγηση του τελικού σταδίου έφερε στην επιφάνεια μια έντονη διχοτομία στις μετρικές μετάφρασης. Ενώ οι παραδοσιακές μετρικές λεξιλογικής ταύτισης κατέρρευσαν, καταγράφοντας πενιχρές βαθμολογίες εξαιτίας της ελεύθερης σύνταξης και των εναλλακτικών αποδόσεων της αρχαίας ελληνικής, η διανυσματική μετρική BERTScore κατέγραψε υψηλότατη σημασιολογική εγγύτητα με βαθμολογία 0.845. Το εύρημα αυτό επιβεβαιώνει ότι, παρότι τα Μεγάλα Γλωσσικά Μοντέλα υστερούν στην παραγωγή μεταφράσεων έτοιμων προς άμεση δημοσίευση, διαπρέπουν ως εργαλεία εξόρυξης νοήματος, ικανά να διαβάσουν πίσω από τον θόρυβο του οπτικού σφάλματος και να αποδώσουν με ακρίβεια τον εννοιολογικό πυρήνα του παπύρου.

Επιπλέον, η έρευνα καταρρίπτει την αντίληψη ότι η επιτυχής αποκατάσταση απαιτεί υποχρεωτικά τη δαπανηρή, ενεργοβόρα και τεχνικά περίπλοκη μικρο-ρύθμιση εμπορικών μοντέλων πάνω σε εξειδικευμένα σώματα συνεχούς γραφής. Μέσω του συνδυασμού προηγμένης μηχανικής υποδείξεων με εννοιολογική γείωση στους εκδοτικούς κανόνες και της τοπικής ανάκτησης ν-γραμμάτων, αποδείχθηκε ότι προ-εκπαιδευμένα γλωσσικά

μοντέλα γενικού σκοπού μπορούν να επιτύχουν ανώτερη σημασιολογική ευστάθεια και να μειώσουν το Ποσοστό Σφάλματος Χαρακτήρων στο 38.2% πάνω σε ακατέργαστες σαρώσεις. Ταυτόχρονα, διασφαλίζεται η απόλυτη ακαδημαϊκή ιχνηλασιμότητα της διαδικασίας. Το σύστημα απομακρύνεται από τη λογική του αδιαφανούς μαύρου κουτιού, καθώς η υποχρεωτική επιστροφή των ιστορικών παραλλήλων που αντλήθηκαν από τη βάση δεδομένων και η επιβολή σήμανσης των αβεβαιοτήτων επιτρέπουν στον ερευνητή να ελέγχει διαρκώς τις πηγές και τη συλλογιστική πορεία του μοντέλου.

5.2 Περιορισμοί της Έρευνας

Παρά τα θετικά αποτελέσματα, η εφαρμογή της Παραγωγικής Τεχνητής Νοημοσύνης σε φθαρμένα αρχαία κείμενα υπόκειται σε συγκεκριμένους τεχνικούς και υλικούς περιορισμούς. Πρωτίστως, διαπιστώθηκε η ύπαρξη ενός αυστηρού κατωφλίου φυσικής διατήρησης του υλικού, κάτω από το οποίο η απόδοση ολόκληρου του pipeline υποβαθμίζεται ραγδαία. Σε παπύρους με ακραία αποσπασματικότητα, εκτεταμένη απολέπιση των ινών ή απανθράκωση, όπου το αρχικό στάδιο της οπτικής αναγνώρισης καταρρέει παράγοντας ακατάληπτο θόρυβο με σφάλμα άνω του 70%, τα γλωσσικά μοντέλα στερούνται επαρκών σταθερών σημείων αναφοράς. Ακόμη και με την υποστήριξη της ανάκτησης εγγράφων, η αδυναμία αντιστοίχισης αξιόπιστων ν-γραμμάτων οδηγεί το σύστημα είτε σε πλήρη σημασιολογική σύγχυση είτε στην αυθαίρετη παραγωγή ανύπαρκτου ιστορικά κειμένου.

Παράλληλα, καταγράφηκε ένας σαφής περιορισμός ως προς την κλίμακα και τις υπολογιστικές απαιτήσεις των μοντέλων. Η αποτυχία των μικρών τοπικών γλωσσικών δικτύων των 4 έως 8 δισεκατομμυρίων παραμέτρων, τα οποία κατέγραψαν ποσοστά σφάλματος χειρότερα ακόμη και από το ακατέργαστο οπτικό σύστημα, αποδεικνύει ότι η επεξεργασία και η κατανόηση της αρχαίας ελληνικής παπυρολογίας προϋποθέτει μοντέλα μεγάλης κλίμακας, άνω των 32 ή 70 δισεκατομμυρίων παραμέτρων. Η ανάγκη αυτή περιορίζει την αυτονομία του συστήματος σε απλούς προσωπικούς υπολογιστές και επιβάλλει την εξάρτηση από ισχυρές υπολογιστικές υποδομές νέφους ή από εμπορικές διεπαφές προγραμματισμού εφαρμογών. Τέλος, η αναγκαιότητα επιβολής αυστηρής μορφολογικής κανονικοποίησης, με την αφαίρεση τόνων και τη μετατροπή του κειμένου σε κεφαλαία γράμματα προκειμένου να υπολογιστούν αντικειμενικά οι αποστάσεις Levenshtein, απαλείφει χρήσιμες πληροφορίες που σχετίζονται με την ικανότητα του μοντέλου να τονίζει και να ορθογραφεί με ακρίβεια τις αποκατεστημένες λέξεις.

5.3 Προτάσεις για Μελλοντική Έρευνα

Οι διαπιστωμένοι περιορισμοί και οι επιτυχίες της παρούσας εργασίας διανοίγουν σαφείς κατευθύνσεις για τη μελλοντική επέκταση της έρευνας. Μια άμεση και υποσχόμενη

εξέλιξη αφορά την άμεση διασύνδεση του συστήματος με αρχιτεκτονικές πολυτροπικών γλωσσικών μοντέλων άμεσης ανάγνωσης. Η παράκαμψη του ενδιάμεσου βήματος της μετατροπής των εικόνων σε συμβολοσειρές κειμένου και η απευθείας τροφοδότηση των οπτικών αποκομμάτων υψηλής ανάλυσης, αμέσως μετά τον εντοπισμό τους από το YOLO και το SAHI, μέσα σε πολυτροπικά μοντέλα αιχμής, θα επιτρέψει στο δίκτυο να συνδυάζει ταυτόχρονα τα οπτικά ίχνη μελανιού με τα γλωσσικά συμφραζόμενα ολόκληρης της πρότασης, εξαλείφοντας τα ενδιάμεσα σφάλματα ταξινόμησης.

Εξίσου στρατηγικής σημασίας κρίνεται η θεσμοθέτηση ενός διαδραστικού βρόχου ανάδρασης μεταξύ ανθρώπου και μηχανής (Human-in-the-Loop - HITL) [42], ο οποίος θα μετασχηματίσει το προτεινόμενο σύστημα από στατικό εργαλείο προβλέψεων σε ένα δυναμικό, αυτο-βελτιούμενο περιβάλλον εκμάθησης. Η μελλοντική έρευνα οφείλει να εστιάσει στην ανάπτυξη μιας συνεργατικής διαδικτυακής πλατφόρμας, άμεσα διασυνδεδεμένης με τις κεντρικές ψηφιακές υποδομές της παγκόσμιος κοινότητας, όπως το Papyri.info [53] και το Trismegistos [20]. Σε ένα τέτοιο περιβάλλον Active Learning, ο εξειδικευμένος παπυρολόγος θα δεν θα είναι απλός παθητικός δέκτης των προβλέψεων, αλλά θα μπορεί να επιβεβαιώνει, να απορρίπτει ή να διορθώνει μικρο-επεμβατικά τις προτάσεις αποκατάστασης του Υβριδικού RAG και του LLM. Η διόρθωση αυτή του ειδικού θα καταγράφεται και θα διοχετεύεται πίσω στο σύστημα μέσω αλγορίθμων Ενισχυτικής Μάθησης από Ανθρώπινη Ανάδραση ή προσαρμοσμένου Domain-Specific Fine-Tuning [32] [40], επιτρέποντας στα διανυσματικά βάρη της βάσης Supabase και στα φίλτρα του αλγορίθμου SAHI να βελτιστοποιούνται διαρκώς πάνω στις πραγματικές προτιμήσεις και τις παρατηρήσεις της ακαδημαϊκής κοινότητας.

Επιπρόσθετα, η ερευνητική μεθοδολογία οφείλει να διευρυνθεί πέρα από τα όρια των δισδιάστατων ψηφιοποιήσεων, επεκτεινόμενη σε τρισδιάστατα ογκομετρικά δεδομένα ακτίνων Χ (3D X-ray micro-computed tomography). Ο θρίαμβος του Vesuvius Challenge [22] [23] στην εικονική αναδίπλωση των απανθρακωμένων κυλίνδρων του Herculaneum μέσω 3D ResNets και TimeSformer κατέδειξε ότι η πληροφορία του μελανιού διασώζεται στο εσωτερικό των κλειστών κυλίνδρων. Η ενσωμάτωση του δικού μας Υβριδικού RAG συστήματος σε 3D ογκομετρικά pipelines ανίχνευσης ρωγμών μελανιού θα μπορούσε να προσφέρει τον κρίσιμο φιλολογικό «φάρο» που απαιτείται για την αποκατάσταση των εξαιρετικά αποσπασματικών ψηφίδων που ανακτώνται από το εσωτερικό των απανθρακωμένων ειληταρίων. Τέλος, μια επιβεβλημένη μελλοντική κατεύθυνση αφορά τη διαπολιτισμική και πολύγλωσση γενίκευση του συστήματος. Η ενοποίηση των παπυρολογικών συλλογών δεν περιορίζεται μόνο στην Ελληνική γλώσσα, αλλά εμπεριέχει χιλιάδες αμετάφραστα έγγραφα στη Δημοτική Αιγυπτιακή, την Κοπτική, την Αραμαϊκή, καθώς και την πρώιμη Λατινική γλώσσα. [5] [36] Η επέκταση του εκπαιδευτικού συνόλου AL-PUB με χαρακτήρες άλλων αλφαβήτων και ο εμπλουτισμός της διανυσματικής βάσης Supabase με πολύγλωσσα ιστορικά κείμενα, θα καταστήσει το pipeline μας ένα καθολικό εργαλείο αποκρυπτογράφησης του αρχαίου μεσογειακού κόσμου.

Βιβλιογραφία

[1] R. S. Bagnall, The Oxford Handbook of Papyrology, Oxford University Press,, 2011 , p. 688 .

[2] P. W. Pestman, The New Papyrological Primer, Brill Archive, 1990, p. 315 .

[3] Reynolds et al, Scribes and Scholars: A Guide to the Transmission of Greek and Latin Literature, OUP Oxford, 2013 .

[4] J. H. Brusuelas, «Scholarly editing and AI: Machine predicted text and Herculaneum papyri,» 2021.

[5] N. Reggiani, Digital Papyrology I: Methods, Tools and Trends, Walter de Gruyter GmbH & Co KG, 2017, p. 326 .

[6] Bowman et al, Images and Artefacts of the Ancient World, Liverpool University Press, 2005.

[7] E. G. Turner, Actes Du XVe Congr{\`e}s International de Papyrologie: The Terms Recto and Verso: the Anatomy of the Papyrus Roll., Fondation égyptologique Reine Élisabeth, 1978.

[8] Roberts et al, The Birth of the Codex, London: Oxford University Press, 1983.

[9] Easton et al, «Multispectral imaging of the Archimedes palimpsest,» σε *32nd Applied Imagery Pattern Recognition Workshop, 2003. Proceedings.*, 2003, pp. 111-116.

[10] R. S. Bagnall, Everyday writing in the Graeco-Roman east, τόμ. 69, Univ of California Press, 2011.

[11] D. Sider, The library of the Villa dei Papiri at Herculaneum, Getty Publications, 2005.

[12] J. Zessner-Spitzenberg, Festschrift zum 100-jährigen Bestehen der Papyrussammlung der Österreichischen Nationalbibliothek, Hollinek, 1983.

[13] U. Wilcken, Grundzüge und Chrestomathie der Papyruskunde, τόμ. 1, BG Teubner, 1912.

[14] P. Parsons, City of the sharp-nosed fish: Greek lives in Roman Egypt, London: Hachette UK, 2012.

[15] R. S. Bagnall, Reading papyri, writing ancient history, Routledge, 2019.

[16] J. Brusuelas, «Engaging greek: Ancient lives,» σε *Digital Classics Outside the Echo-Chamber*, Ubiquity Press, 2016.

[17] H. C. Youtie, «The papyrologist: artificer of fact,» *Greek, Roman, and Byzantine Studies,* τόμ. 4, αρ. 1, pp. 19-32, 1963.

[18] T. Stanley, «Papyrus storage at Princeton University,» *The Book and Paper Group of the American Institute for Conservation,* 1994.

[19] P. Y. Seales et al, «From damage to discovery via virtual unwrapping: Reading the scroll from En-Gedi,» *Science advances,* τόμ. 2, αρ. 9, p. e1601247, 2016.

[20] M. a. G. T. Depauw, «Trismegistos: An interdisciplinary platform for ancient world texts and related informatio,» *International Conference on Theory and Practice of Digital Libraries,* pp. 40-52, 2013.

[21] H. C. Gnanasekaran et al, «Leveraging Knowledge Graph-Enhanced RAG and LLMs for Historical Archival Analysis: A Case Study of State of Maryland's Dataset Collections,» *International Symposium on Grids and Clouds (ISGC2025),* τόμ. 16, p. 21, 2025.

[22] S. Prize, «Vesuvius Challenge 2023 Grand Prize awarded: we can read the first scroll!,» February 2024.

[23] A. M et al, «Complete virtual unwrapping and reading of a rolled Herculaneum papyrus,» 2026.

[24] S. R. K. Vadlamudi et al, «SeamFormer: high precision text line segmentation for handwritten documents,» *International Conference on Document Analysis and Recognition,* pp. 313-331, 2023.

[25] R. a. M. E. Turnbull, «Detecting and recognizing characters in Greek papyri with YOLOv8, DeiT and SimCLR,» *International Journal on Document Analysis and Recognition (IJDAR),* τόμ. 28, αρ. 2, pp. 277-285, 2025.

[26] M. G De Gregorio et al, «NeuroPapyri: A Deep Attention Embedding Network for Handwritten Papyri Retrieval,» *International Conference on Document Analysis and Recognition,* pp. 71-86, 2024.

[27] W. J. F. Swindall et al, «Exploring learning approaches for ancient Greek character recognition with citizen science data,» *2021 IEEE 17th International conference on eScience (eScience),* pp. 128-137, 2021.

[28] J. Assael et al, «Restoring ancient text using deep learning: a case study on Greek epigraphy,» *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing (EMNLP-IJCNLP),* pp. 6368-6375, 2019.

[29] Assael et al, «Restoring and attributing ancient texts using deep neural networks,» *Nature,* τόμ. 603, αρ. 7900, pp. 280-283, 2022.

[30] Assael et al, «Contextualizing ancient texts with generative neural networks,» *Nature,* τόμ. 645, pp. 141-147, 2025.

[31] J. Cowen-Breen et al, «Logion: Machine-learning based detection and correction of textual errors in Greek philology,» *Proceedings of the Ancient Language Processing Workshop,* pp. 170-178, 2023.

[32] E. Cullhed, «Instruction-tuning pretrained causal language models to restore ancient Greek papyri and inscriptions,» *Digital Scholarship in the Humanities,* 2025.

[33] M. Cotte et al, «Blackening of Pompeian cinnabar paintings: X-ray microspectroscopy analysis,» *Analytical chemistry,* τόμ. 78, pp. 7484-7492, 2006.

[34] R. Burgio et al, «Raman microscopy and x-ray fluorescence analysis of pigments on medieval and Renaissance Italian manuscript cuttings,» *Proceedings of the National Academy of Sciences,* τόμ. 107, pp. 5726-5731, 2010.

[35] F. West et al, «A deep learning pipeline for the palaeographical dating of ancient Greek papyrus fragments,» *Proceedings of the 1st Workshop on Machine Learning for Ancient Languages (ML4AL 2024),* pp. 177-185, 2024.

[36] S. Miyagawa, «RAG-enhanced neural machine translation of Ancient Egyptian text: A case study of THOTH AI,» *Proceedings of the 5th International Conference on Natural Language Processing for Digital Humanities,* pp. 33-40, 2025.

[37] Qiu et al, «On path to multimodal historical reasoning: Histbench and histagent,» *arXiv preprint arXiv:2505.20246,* 2025.

[38] Zhang et al, «Understand layout and translate text: Unified feature-conductive end-to-end document image translation,» *IEEE Transactions on Pattern Analysis and Machine Intelligence,* τόμ. 47, pp. 3358-3376, 2025.

[39] Ji et al, «Survey of hallucination in natural language generation,» *ACM computing surveys,* τόμ. 55, pp. 1-38, 2023.

[40] Zhang et al, «Raft: Adapting language model to domain specific rag,» *arXiv preprint arXiv:2403.10131,* 2024.

[41] Battezzato et al, «Artificial Intelligence and Greek Philology: An Experiment,» *SEMINARI E CONVEGNI,* τόμ. 69, 2025.

[42] Sommerschield et al, «Machine learning for ancient languages: A survey,» *Computational Linguistics,* τόμ. 49, pp. 703-747, 2023.

[43] Li et al, «TrOCR: Transformer-Based Optical Character Recognition with Pre-Trained Models,» *Proceedings of the AAAI Conference on Artificial Intelligence,* p. 13094–13102, 26 06 2023.

[44] G. &. Q. J. Jocher, «Ultralytics YOLO11 [Computer software],» Ultralytics , 2024.

[45] Sun et al, «Deep Residual Learning for Image Recognition,» 2015.

[46] N. Houlsby et al, «An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale,» 2021.

[47] Jégou et al, «Training data-efficient image transformers & distillation through attention,» 2021.

[48] Polosukhin et al, «Attention Is All You Need,» 2023.

[49] T. H, «Language Models are Few-Shot Learners,» 2020.

[50] Toutanova et al, «BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding,» 2019.

[51] Guilla et al, «LLaMA: Open and Efficient Foundation Language Models,» 2023.

[52] Kiela et al, «Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks,» 2021.

[53] E. J. D. Sosin, «Papyri.info,» 2010.

[54] W. a. T. J. Cavnar, «N-Gram-Based Text Categorization,» *Proceedings of the Third Annual Symposium on Document Analysis and Information Retrieval,* 2001.

[55] O. Papyrology, *P.Oxy. LII 3693. Invitation to Dinner,* 2022.

[56] T. Yousef, *Proceedings of the Second Workshop on Language Technologies for Historical and Ancient Languages,* Marseille: European Language Resources Association, 2022.

[57] ". L. A. K. Bowman, «Papyrus in classical antiquity,» σε *The Journal of Hellenic Studies*, Oxford: Clarendon Press, 1974, p. 265.

[58] Akyon et al, «Slicing Aided Hyper Inference and Fine-Tuning for Small Object Detection,» 2022.

[59] O. Papyrology, *P.Oxy. XLIX 3505. Papontos to Alexander,* 2022.