ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΔΙΟΙΚΗΣΗ ΚΑΙ ΨΗΦΙΑΚΟΣ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ»

ΜΕΤΑ-ΑΞΙΟΛΟΓΗΣΗ ΜΕΓΑΛΩΝ ΓΛΩΣΣΙΚΩΝ ΜΟΝΤΕΛΩΝ-LLMS ΜΕ ΤΗΝ ΧΡΗΣΗ ΤΩΝ ΜΕΘΟΔΩΝ AHP ΚΑΙ FAHP

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Φοιτητής: ΚΥΡΙΑΚΟΣ ΠΙΠΕΡΑΚΗΣ, ΜΔΕ38
Επιβλέπων:Δρ Στέλιος Παπαδάκης , Καθηγητής, Μέλος ΔΕΠ

©
2025

ΕΛΛΗΝΙΚΟ ΜΕΣΟΓΕΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ
ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΔΙΟΙΚΗΣΗ ΚΑΙ ΨΗΦΙΑΚΟΣ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ»

ΜΕΤΑ-ΑΞΙΟΛΟΓΗΣΗ ΜΕΓΑΛΩΝ ΓΛΩΣΣΙΚΩΝ ΜΟΝΤΕΛΩΝ-LLMS ΜΕ ΤΗΝ ΧΡΗΣΗ ΤΩΝ ΜΕΘΟΔΩΝ AHP ΚΑΙ FAHP

Τριμελής Εξεταστική Επιτροπή

Στέλιος Παπαδάκης (Επιβλέπων/ουσα)
Καθηγητής, Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας, ΕΛ.ΜΕ.ΠΑ.

Μαρία Μαρκάκη
Επίκουρη Καθηγήτρια, Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας, ΕΛ.ΜΕ.ΠΑ.

Ιωάννης Δημοτίκαλης
Επίκουρος Καθηγητής , Τμήμα Διοικητικής Επιστήμης και Τεχνολογίας, ΕΛ.ΜΕ.ΠΑ.

HELLENIC MEDITERRANEAN UNIVERSITY
SCHOOL OF MANAGEMENT AND ECONOMICS SCIENCE
DEPARTMENT OF MANAGEMENT SCIENCE AND TECHNOLOGY

MSc in
MANAGEMENT AND DIGITAL TRANSFORMATION

META-EVALUATION OF LARGE LANGUAGE MODELS-LLMS USING THE AHP AND FAHP METHODS

MASTER THESIS

Student : KYRIAKOS PIPERAKIS, MDE 38
Supervisor : Dr. Stelios Papadakis, Faculty Member

©
2025

<u>Υπεύθυνη Δήλωση</u> *: Βεβαιώνω ότι είμαι συγγραφέας αυτής της διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στη διπλωματική εργασία. Επίσης έχω αναφέρει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επίσης βεβαιώνω ότι αυτή η διπλωματική εργασία προετοιμάστηκε από εμένα προσωπικά ειδικά για τις απαιτήσεις του Μεταπτυχιακού Προγράμματος Σπουδών Διοίκηση και Ψηφιακός Μετασχηματισμός του Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας του ΕΛΜΕΠΑ.*

ΠΕΡΙΛΗΨΗ

Η διάδοση και η χρήση των LLMs (Large Language Models-Μεγάλα Γλωσσικά Μοντέλα) σε πολλούς τομείς των επιστημών έχει κεντρίσει των ενδιαφέρον πολλών ερευνητών σχετικά με την απόδοση αυτών και την επιλογή των κατάλληλων μοντέλων ανάλογα με τις εφαρμογές που χρησιμοποιούνται. Στην παρούσα εργασία έγινε μέτρηση και εκτίμηση απόδοσης τριών γνωστών μεγάλων γλωσσικών μοντέλων (Phi, Mistral, Gemma) σύμφωνα με ερώτηση που έχει σχέση με την σύνταξη εντολής της γλώσσας προγραμματισμού Python ('What is the command to print output in Python?') και συγκεκριμένα με την εφαρμογή του Ollama framework (τοπική εκτέλεση LLMs) και την χρήση στατιστικών μεθόδων με την βοήθεια της γλώσσας προγραμματισμού Python . Οι μετρήσεις βασίζονται στον καθορισμό κριτηρίων απόδοσης των LLMs (π.χ. ακρίβεια, ταχύτητα εκτέλεσης κ.ά.) που στην συνέχεια με την χρήση των μεθόδων AHP (Analytical Hierarchy Process- Διαδικασία Αναλυτικής Ιεράρχησης) και FAHP (Fuzzy Analytical Hierarchy Process-Ασαφής Διαδικασία Αναλυτικής Ιεράρχησης) έγινε μετα- αξιολόγηση των μοντέλων με σκοπό την εύρεση του καταλληλότερου μοντέλου για το παράδειγμα μας. Οι μετρήσεις και με τις δύο μεθόδους έδειξαν παρόμοια αποτελέσματα με τα δύο μοντέλα (Phi, Mistral) να υπερτερούν σαφώς σε σχέση με το τρίτο Gemma). Η χρήση περισσότερων ερωτήσεων και περισσότερων κριτηρίων αλλά και ο έλεγχος παραπάνω LLMs μπορεί να αποτελέσει αντικείμενο συνέχειας της έρευνας μας.

Λέξεις Κλειδιά : LLMs, AHP, FAHP, MCDA, OLLAMA, PYTHON

ABSTRACT

The proliferation and use of LLMs (Large Language Models) in many fields of science has attracted the interest of many researchers on their performance and the choice of appropriate models depending on the applications used. In this paper, we measured and evaluated the performance of three well-known large language models (Phi, Mistral, Gemma) against a question related to Python programming language command syntax ('What is the command to print output in Python?') by applying the Ollama framework (local execution of LLMs) and using statistical methods with the help of Python programming language. The measurements are based on the definition of performance criteria of LLMs (e.g., accuracy, execution speed, etc.) which then using AHP (Analytical Hierarchy Process) and FAHP (Fuzzy Analytical Hierarchy Process) methods were used to meta-evaluate the models in order to find the most suitable model for our example. The measurements with both methods showed similar results with both models (Phi, Mistral) clearly outperforming the third one (Gemma). The use of more questions and more criteria and the testing of more LLMs can be a subject of our follow-up research.

Key Words : LLMs, AHP, FAHP, MCDA, OLLAMA, PYTHON

ΠΙΝΑΚΑΣ ΠΕΡΙΕΧΟΜΕΝΩΝ

Περιεχόμενα

ΛΙΣΤΑ ΠΙΝΑΚΩΝ vi
ΛΙΣΤΑ ΣΧΕΔΙΑΓΡΑΜΜΑΤΩΝ vii
ΛΙΣΤΑ ΕΙΚΟΝΩΝ viii
ΛΙΣΤΑ ΕΞΙΣΩΣΕΩΝ ix
ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ x
ΕΥΧΑΡΙΣΤΙΕΣ xi
ΚΕΦΑΛΑΙΟ 1- ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΠΡΟΒΛΗΜΑΤΩΝ ΜΕ ΠΟΛΛΑΠΛΑ ΚΡΙΤΗΡΙΑ (MCDA) 1
ΚΕΦΑΛΑΙΟ 2-AHP ΜΕΘΟΔΟΣ 3
2.1 ΕΙΣΑΓΩΓΗ 3
2.2 ΚΛΙΜΑΚΑ ΣΥΓΚΡΙΣΕΩΝ ΚΑΤΑ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ 4
2.2.1 ΒΑΘΜΟΝΟΜΗΣΗ ΤΗΣ ΣΗΜΑΣΙΑΣ 4
2.2.2 ΑΜΟΙΒΑΙΟΤΗΤΕΣ ΤΙΜΩΝ ΣΥΓΚΡΙΣΕΩΝ 5
2.2.3 ΧΡΗΣΗ ΤΩΝ ΣΥΓΚΡΙΣΕΩΝ ΑΝΑ ΖΕΥΓΗ 5
2.2.4 ΤΕΧΝΙΚΕΣ ΛΕΠΤΟΜΕΡΕΙΕΣ ΜΕΘΟΔΟΥ 6
2.3 ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΕΘΟΔΟΥ AHP 6
ΚΕΦΑΛΑΙΟ 3-FAHP ΜΕΘΟΔΟΣ 9
3.1 ΕΙΣΑΓΩΓΗ-ΕΦΑΡΜΟΓΗ ΤΗΣ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ ΣΤΗΝ AHP 9
3.2 ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΤΩΝ ΚΡΙΤΗΡΙΩΝ 10
3.2.1 ΧΡΗΣΗ ΑΣΑΦΩΝ ΑΡΙΘΜΩΝ: 11
3.2.2 ΣΥΓΚΡΙΣΕΙΣ ΚΑΤΑ ΖΕΥΓΗ 11
3.2.3 ΥΠΟΛΟΓΙΣΜΟΣ ΚΑΤΑΤΑΞΗΣ ΚΑΙ ΒΑΡΥΤΗΤΑΣ:\ 12
3.2.4 ΛΗΨΗ ΤΕΛΙΚΗΣ ΑΠΟΦΑΣΗΣ 12
3.3 ΑΣΑΦΗΣ ΛΟΓΙΚΗ 12
3.4 ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΕΘΟΔΟΥ FAHP 13
ΚΕΦΑΛΑΙΟ 4-LLMs ΜΟΝΤΕΛΑ 15
4.1 ΕΙΣΑΓΩΓΗ 15
4.2 ΕΦΑΡΜΟΓΕΣ LLMs 16
4.3 ΠΑΡΑΓΩΓΗ ΚΕΙΜΕΝΟΥ ΕΝΙΣΧΥΜΕΝΗ ΜΕ ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ (RAG) ΓΙΑ LLMS 18
4.3.1 ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ RAG 19
4.3.2. ΕΙΔΗ RAG – ΣΥΣΤΗΜΑΤΩΝ 19
ΚΕΦΑΛΑΙΟ 5-ΥΛΟΠΟΙΗΣΗ 22
5.1 LLM – OLLAMA PYTHON ΒΙΒΛΙΟΘΗΚΕΣ 22
5.2 ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ LLMS 23

5.3 ΕΠΙΠΛΕΟΝ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ LLMS 25
5.4 ΜΕΤΡΗΣΕΙΣ ΕΠΙΔΟΣΕΩΝ LLMs 26
5.5 ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΩΝ AHP/FAHP 29
5.5.1 ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ ΜΕΘΟΔΟΥ 29
5.5.2 ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ ΜΕΘΟΔΟΥ 35
ΚΕΦΑΛΑΙΟ 6- ΣΥΜΠΕΡΑΣΜΑΤΑ-ΠΡΟΤΑΣΕΙΣ 40
ΒΙΒΛΙΟΓΡΑΦΙΑ 42
Α. ΞΕΝΟΓΛΩΣΣΗ 42
ΠΑΡΑΡΤΗΜΑ Α 46
ΚΩΔΙΚΑΣ PYTHON ΜΕΘΟΔΟΣ AHP 46
ΚΩΔΙΚΑΣ PYTHON ΜΕΘΟΔΟΣ FAHP 47
ΠΑΡΑΡΤΗΜΑ Β 51
ΚΩΔΙΚΑΣ PYTHON ΚΡΙΤΗΡΙΟ - ROUGE 51
ΚΩΔΙΚΑΣ PYTHON ΚΡΙΤΗΡΙΑ- LOAD DURATION, PROMPT_RATE, EVALUATION_RATE 51

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

ΠΙΝΑΚΑΣ 5.1-ΠΙΝΑΚΑΣ ΚΡΙΤΗΡΙΩΝ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ 30
ΠΙΝΑΚΑΣ 5.2-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-LOAD DURATION.....31
ΠΙΝΑΚΑΣ 5.3-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-PROMPT RATE 32
ΠΙΝΑΚΑΣ 5.4-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-EVALUATION RATE .32
ΠΙΝΑΚΑΣ 5.5-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-1.....32
ΠΙΝΑΚΑΣ 5.6-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-2.....33
ΠΙΝΑΚΑΣ 5.7-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-L 33
ΠΙΝΑΚΑΣ 5.8-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-METEOR 34
ΠΙΝΑΚΑΣ 5.9-ΠΙΝΑΚΑΣ ΚΡΙΤΗΡΙΩΝ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ.....36
ΠΙΝΑΚΑΣ 5.10-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-LOAD DURATION36
ΠΙΝΑΚΑΣ 5.11-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-PROMPT RATE.....37
ΠΙΝΑΚΑΣ 5.12-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-EVALUATION RATE.....37
ΠΙΝΑΚΑΣ 5.13-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-1 37
ΠΙΝΑΚΑΣ 5.14-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-2 37
ΠΙΝΑΚΑΣ 5.15-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-L.....38
ΠΙΝΑΚΑΣ 5.16-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-METEOR.....38

ΛΙΣΤΑ ΣΧΕΔΙΑΓΡΑΜΜΑΤΩΝ

ΔΙΑΓΡΑΜΜΑ 5.1-ΑΠΟΔΟΣΗ LLMS ΩΣ ΠΡΟΣ CALC_DURATION, PROMPT_RATE, EVAL_RATE 27
ΔΙΑΓΡΑΜΜΑ 5.2-ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS AHP ΜΕΘΟΔΟΣ 34
ΔΙΑΓΡΑΜΜΑ 5.3-ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS FAHP ΜΕΘΟΔΟΣ 35
ΔΙΑΓΡΑΜΜΑ 5.4-ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS AHP ΜΕΘΟΔΟΣ 38
ΔΙΑΓΡΑΜΜΑ 5.5-ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS FAHP ΜΕΘΟΔΟΣ 39

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

ΕΙΚΟΝΑ 2.1 – ΑΗΡ ΔΙΑΔΙΚΑΣΙΑ ΛΥΣΗΣ....3
ΕΙΚΟΝΑ 2.2 – ΑΗΡ ΙΕΡΑΡΧΙΚΗ ΔΟΜΗ....4
ΕΙΚΟΝΑ 2.3 – ΠΙΝΑΚΑΣ ΑΜΟΙΒΑΙΟΤΗΤΑΣ ΚΡΙΤΗΡΙΩΝ ΜΕΘΟΔΟΥ....6
ΕΙΚΟΝΑ 3.1-ΑΝΑΠΑΡΑΣΤΑΣΗ ΜΕΘΟΔΟΥ FAHP....10
ΕΙΚΟΝΑ 3.2- Η ΣΥΝΑΡΤΗΣΗ ΣΥΜΜΕΤΟΧΗΣ ΤΟΥ ΤΡΙΓΩΝΙΚΟΥ ΑΣΑΦΟΥΣ ΑΡΙΘΜΟΥ....11
ΕΙΚΟΝΑ 4.1-ΑΝΑΠΑΡΑΣΤΑΣΗ LLMS....16
ΕΙΚΟΝΑ 4.2-ΣΧΗΜΑΤΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ RAG....18
ΕΙΚΟΝΑ 5.1-ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΩΤΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-AHP....34
ΕΙΚΟΝΑ 5.2-ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΩΤΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-FAHP....35
ΕΙΚΟΝΑ 5.3-ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΥΤΕΡΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-AHP....38
ΕΙΚΟΝΑ 5.4-ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΥΤΕΡΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-FAHP....39

ΛΙΣΤΑ ΕΞΙΣΩΣΕΩΝ

ΕΞΙΣΩΣΗ 5.1-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-N 23
ΕΞΙΣΩΣΗ 5.2-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-L 25
ΕΞΙΣΩΣΗ 5.3-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ METEOR 25
ΕΞΙΣΩΣΗ 5.4-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-W 26
ΕΞΙΣΩΣΗ 5.5-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-S 26

ΣΥΝΤΟΜΟΓΡΑΦΙΕΣ

LLMs: Large Language Models

AHP: Analytical Hierarchy Process

FAHP: Fuzzy Analytical Hierarchy Process

MCDA: Multiple-criteria decision analysis

RAG: Retrieval Augmented Generation

ΕΥΧΑΡΙΣΤΙΕΣ

Ευχαριστώ τον Επιβλέποντα Καθηγητή Κο Παπαδάκη για την άψογη συνεργασία που είχαμε και την πολύτιμη βοήθεια του καθώς και την οικογένεια μου και ιδιαίτερα την σύζυγο μου για την συμπαράσταση και την υπομονή που έδειξε κατά την διάρκεια της εργασίας.

ΚΕΦΑΛΑΙΟ 1- ΜΕΘΟΔΟΙ ΑΝΑΛΥΣΗΣ ΠΡΟΒΛΗΜΑΤΩΝ ΜΕ ΠΟΛΛΑΠΛΑ ΚΡΙΤΗΡΙΑ (MCDA)

Η χρήση των μεθόδων ανάλυσης προβλημάτων με πολλαπλά κριτήρια είναι ένας τρόπος να βοηθηθούν οι υπεύθυνοι λήψης αποφάσεων έτσι ώστε να επιλέξουν σωστά μεταξύ πολλαπλών επιλογών όταν υπάρχουν αντικρουόμενοι στόχοι. Η εφαρμογή τους είναι επιτακτική όταν υπάρχει ένας συνδυασμός κριτηρίων που δεν μπορούν να συγκριθούν και όταν υπάρχουν πολλαπλές προοπτικές που επηρεάζουν την λήψη αποφάσεων. Οι μέθοδοι συνδυάζουν ποιοτικά και ποσοτικά στοιχεία, όπου τα ποιοτικά αναφέρονται στην συνεργασία με τα ενδιαφερόμενα μέρη για την διερεύνηση των απόψεων τους ενώ τα ποσοτικά στην χρήση των μοντέλων για την αναπαράσταση των προτιμήσεων των ενδιαφερομένων και των επιδόσεων των επιλογών. Με αυτό τον τρόπο τα μοντέλα μπορούν να παράγουν χρήσιμες πληροφορίες για την επίλυση προβλημάτων.

Οι παραδοσιακές μέθοδοι λήψης αποφάσεων συχνά παραβλέπουν κρίσιμες πτυχές των προβλημάτων αφού οι αποφάσεις του πραγματικού κόσμου περιλαμβάνουν πολλές εξαρτήσεις και διαφορετικά κριτήρια. Οι MCDA στοχεύουν στον προσδιορισμό της καλύτερης επιλογής από ένα σύνολο εναλλακτικών λύσεων (Chaube et al., 2024). Ο προσδιορισμός και η οργάνωση είναι ζωτικής σημασίας στην φάση δόμησης του προβλήματος και η λάθος επιλογή κριτηρίων μπορεί να οδηγήσει σε παραπλανητικά συμπεράσματα. Ένας πίνακας αποφάσεων σε αυτά τα μοντέλα δείχνει τις εναλλακτικές λύσεις σε σχέση με τα κριτήρια ανάλυσης όπου κάθε στοιχείο του πίνακα καταδεικνύει την απόδοση μιας εναλλακτικής λύσης σε σχέση με ένα κριτήριο. Οι περισσότερες μέθοδοι επιλέγουν τιμές βαρών που δίνονται υποθετικά σύμφωνα με κάποια υποκειμενικά κριτήρια κάτι το οποίο έρχεται να διορθώσει η μέθοδος της ιεραρχικής ανάλυσης αποφάσεων (AHP). Στην συγκεκριμένη μέθοδο τα προβλήματα οργανώνονται σε ιεραρχία όπου διαμορφώνεται μια γραμμική ιεραρχική δομή για την εξαγωγή βαρών ενώ υπάρχει και έλεγχος συνέπειας για την διατήρηση της αξιοπιστίας στους πίνακες αποφάσεων. Με αυτό τον τρόπο αντιμετωπίζονται πιθανές ασυνέπειες δεδομένων που προκύπτουν κάτι που δεν είναι εφικτό στις περισσότερες μεθόδους ανάλυσης προβλημάτων με πολλαπλά κριτήρια (MCDA).

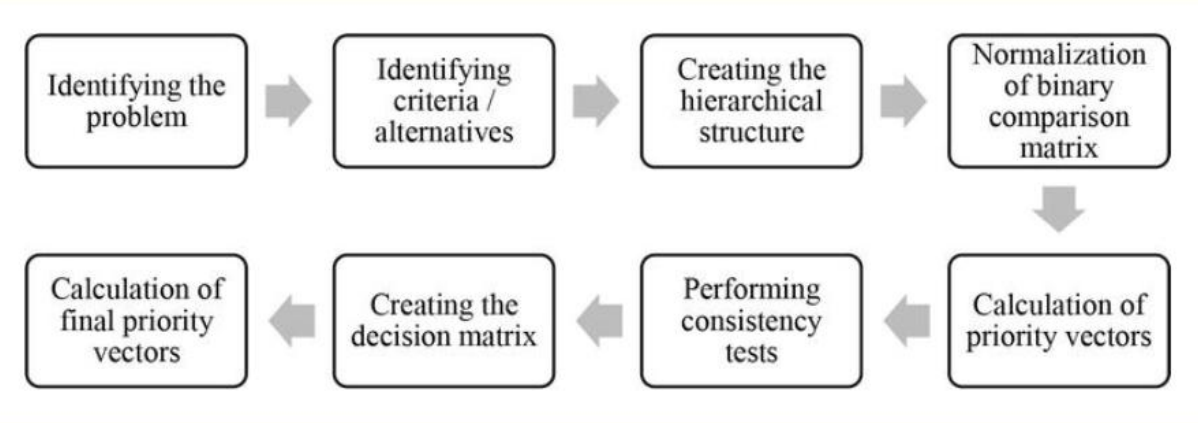
Η AHP μέθοδος χρησιμοποιείται για την επιλογή προμηθευτών και την βελτίωση των επιδόσεων στις αλυσίδες εφοδιασμού αλλά και στην λήψη κλινικών αποφάσεων, στην αξιολόγηση επιλογών ηλεκτρονικού εμπορίου και τη διαχείριση αποβλήτων. Η μέθοδος αυτή αποτελεί ένα παράδειγμα μεθόδου MCDA όμως σε σχέση με το σύνολο των μεθόδων της συγκεκριμένης κατηγορίας είναι πιο ευέλικτη και δομεί ιεραρχικά και όχι με βάση αυθαίρετες εκτιμήσεις ένα πρόβλημα απόφασης καθότι συγκρίνει τα κριτήρια ανά ζεύγη και υπολογίζει σταθμισμένες βαθμολογίες. Στην συνέχεια γίνεται περιγραφή των μεθόδων AHP και FAHP καθώς και εφαρμογών αυτών.

ΚΕΦΑΛΑΙΟ 2-AHP ΜΕΘΟΔΟΣ

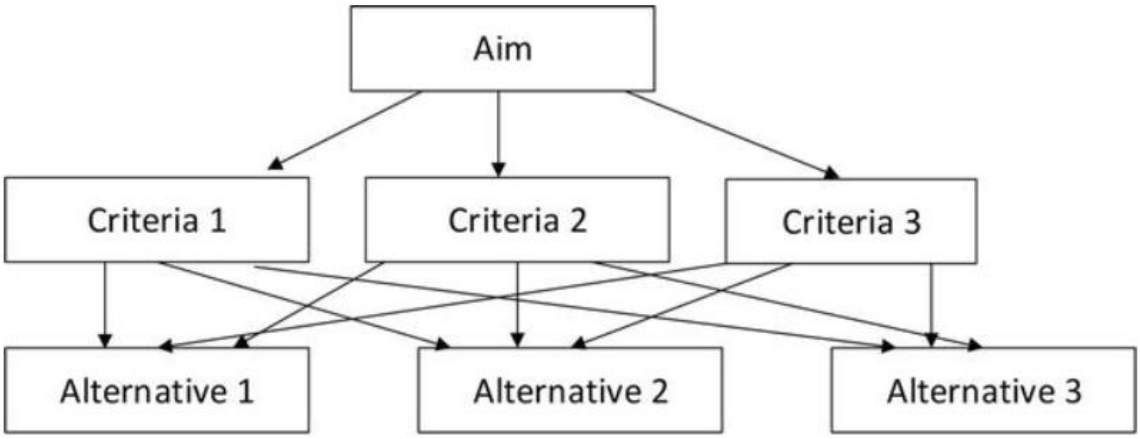
2.1 ΕΙΣΑΓΩΓΗ

Η μέθοδος της διαδικασίας αναλυτικής ιεράρχησης ενός προβλήματος απόφασης (AHP) περιλαμβάνει τα ακόλουθα βήματα (Saaty, 2008):

- Διαχωρισμός της διαδικασίας λήψης αποφάσεων σε συγκεκριμένα και σαφή βήματα. Ξεκινώντας με τον ορισμό του προβλήματος και τον εντοπισμό της απαιτούμενης γνώσης.
- Δημιουργία της ιεραρχίας αποφάσεων που περιγράφει τον κύριο στόχο (main goal), τους επιμέρους στόχους, τα κριτήρια και τις εναλλακτικές λύσεις με δομημένο τρόπο.
- Εφαρμογή πινάκων συγκρίσεων ανά ζεύγη προκειμένου να συγκριθούν τα στοιχεία σε διαφορετικά επίπεδα εστιάζοντας στον τρόπο με τον οποίο κάθε στοιχείο σχετίζεται με το στοιχείο που βρίσκεται πάνω από αυτό.
- Εφαρμογή αριθμητικής κλίμακας για την ανάδειξη της σημασία ενός στοιχείου έναντι ενός άλλου, φροντίζοντας να καταγραφούν οι αμοιβαίες τιμές για ακριβείς συγκρίσεις.



ΕΙΚΟΝΑ 2.1 – AHP ΔΙΑΔΙΚΑΣΙΑ ΛΥΣΗΣ
(Esen,2023)



ΕΙΚΟΝΑ 2.2 – AHP ΙΕΡΑΡΧΙΚΗ ΔΟΜΗ
(Esen,2023)

2.2 ΚΛΙΜΑΚΑ ΣΥΓΚΡΙΣΕΩΝ ΚΑΤΑ ΤΗΝ ΕΦΑΡΜΟΓΗ ΤΟΥ ΜΟΝΤΕΛΟΥ

Κατά τη λήψη αποφάσεων, ειδικά όταν χρησιμοποιούμε τη Διαδικασία Αναλυτικής Ιεράρχησης (AHP), χρειαζόμαστε έναν τρόπο σύγκρισης διαφορετικών δραστηριοτήτων ή επιλογών. Αυτή η σύγκριση μας βοηθά να καταλάβουμε ποια επιλογή είναι καλύτερη για την επίτευξη των στόχων μας. Ο πίνακας που εξετάζουμε παρέχει μια κλίμακα για να μετρήσουμε πόσο σημαντική είναι μια δραστηριότητα σε σύγκριση με μια άλλη. Ας το αναλύσουμε βήμα προς βήμα.

2.2.1 ΒΑΘΜΟΝΟΜΗΣΗ ΤΗΣ ΣΗΜΑΣΙΑΣ

Αυτός είναι ένας τρόπος για να περιγραφεί πόσο πιο σημαντική είναι μια δραστηριότητα σε σύγκριση με μια άλλη. Η κλίμακα κυμαίνεται από το 1 έως το 9, με κάθε αριθμό να αντιπροσωπεύει ένα διαφορετικό επίπεδο σημαντικότητας.

1: Ίδια σημαντικότητα

Αυτό σημαίνει ότι δύο δραστηριότητες είναι εξίσου σημαντικές για την επίτευξη του στόχου.

2: Ασθενής ή μικρή σημαντικότητα

Εδώ, η μία δραστηριότητα είναι απλώς λίγο πιο σημαντική από την άλλη.

3: Μέτρια σημαντικότητα

Στην περίπτωση αυτή, η εμπειρία ή η κρίση δείχνει ότι η μία δραστηριότητα είναι κάπως καλύτερη από την άλλη.

4: Μέτρια συν- σημαντικότητα

Αυτό δείχνει ότι η μία δραστηριότητα είναι πιο σημαντική από την άλλη, αλλά όχι με μεγάλη διαφορά.

5: Ισχυρή σημαντικότητα

Εδώ, η μία δραστηριότητα είναι σαφώς καλύτερη από την άλλη.

6: Ισχυρή συν-σημαντικότητα

Αυτό σημαίνει ότι η μία δραστηριότητα είναι σημαντικά καλύτερη από την άλλη.

7: Πολύ ισχυρή ή αποδεδειγμένη σημαντικότητα

Σε αυτό το επίπεδο, μια δραστηριότητα ευνοείται πολύ έντονα έναντι μιας άλλης.

8: Πολύ, πολύ ισχυρή σημαντικότητα

Αυτό δείχνει ότι μια δραστηριότητα είναι εξαιρετικά καλύτερη από την άλλη.

9: Ακραία σημαντικότητα

Αυτό είναι το υψηλότερο επίπεδο σπουδαιότητας.

2.2.2 ΑΜΟΙΒΑΙΟΤΗΤΕΣ ΤΙΜΩΝ ΣΥΓΚΡΙΣΕΩΝ

Όταν γίνεται σύγκριση δύο δραστηριοτήτων και αποδοθεί ένας αριθμός στη μία, τότε η άλλη θα έχει την αντίστροφη αξία. Επιπλέον, όταν οι δραστηριότητες είναι πολύ κοντά σε σημασία, μπορεί να είναι δύσκολο να επιλεγεί η καλύτερη τιμή. Παραδείγματος χάριν , μπορεί να χρησιμοποιηθεί αριθμός όπως 1,1 ή 1,2. Αυτοί οι μικροί αριθμοί δείχνουν ότι ενώ είναι κοντά, η μία εξακολουθεί να είναι λίγο πιο σημαντική από την άλλη. Συνοπτικά, η κλίμακα αυτή μας βοηθά να κάνουμε σαφείς συγκρίσεις μεταξύ διαφορετικών δραστηριοτήτων ή επιλογών. Χρησιμοποιώντας αυτούς τους αριθμούς, μπορούμε να κατανοήσουμε καλύτερα ποιες επιλογές είναι πιο σημαντικές για τους στόχους μας. Αυτή η δομημένη προσέγγιση καθιστά τη λήψη αποφάσεων ευκολότερη και πιο λογική.

2.2.3 ΧΡΗΣΗ ΤΩΝ ΣΥΓΚΡΙΣΕΩΝ ΑΝΑ ΖΕΥΓΗ

Η ακριβής εκτίμηση των δεδομένων είναι ζωτικής σημασίας στις μεθόδους λήψης αποφάσεων, συμπεριλαμβανομένης της AHP. Τα ποιοτικά δεδομένα είναι συχνά δύσκολο να

ποσοτικοποιηθούν σε απόλυτους όρους (Triantaphyllou and Mann, 1995). Οι κατά ζεύγη συγκρίσεις βοηθούν στον προσδιορισμό της σχετικής σημασίας των εναλλακτικών λύσεων. Οι λήπτες αποφάσεων εκφράζουν απόψεις για ζεύγη, επιλέγοντας από διακριτές γλωσσικές φράσεις. Η ποσοτικοποίηση των γλωσσικών επιλογών σε αριθμητικές τιμές αποτελεί βασική πρόκληση.

Υπάρχουν διάφορες κλίμακες για την ποσοτικοποίηση των συγκρίσεων ανά ζεύγη, με προεξέχουσα την κλίμακα του Saaty που αναφέραμε προηγουμένως. Οι ψυχολογικές θεωρίες παρέχουν πληροφορίες για την ανάπτυξη αυτών των κλιμάκων. Ενώ οι άνθρωποι δυσκολεύονται να συγκρίνουν πολλά αντικείμενα ταυτόχρονα, γεγονός που επηρεάζει το σχεδιασμό της κλίμακας.

	Criteria 1	**Criteria 2**	**Criteria...**	**criteria j**
Criteria 1	W_1/W_1	W_1/W_2	...	W_1/Wj
Criteria 2	W_2/W_1	W_1/W_1	...	W_2/Wj
Criteria...	...	...	...	...
Criteria j	Wj/W_1	Wj/W_2	...	Wj/Wj

ΕΙΚΟΝΑ 2.3 – ΠΙΝΑΚΑΣ ΑΜΟΙΒΑΙΟΤΗΤΑΣ ΚΡΙΤΗΡΙΩΝ ΜΕΘΟΔΟΥ
(Esen,2023)

2.2.4 ΤΕΧΝΙΚΕΣ ΛΕΠΤΟΜΕΡΕΙΕΣ ΜΕΘΟΔΟΥ

Οι υπεύθυνοι λήψης αποφάσεων αποδίδουν τιμές στις συγκρίσεις με βάση την αντιλαμβανόμενη σημασία. Το κύριο ιδιοδιάνυσμα του πίνακα κρίσης εκτιμά τις σχετικές σημαντικότητες. Στην συνέχεια ο γεωμετρικός μέσος κάθε γραμμής χρησιμοποιείται για την εξαγωγή διανυσμάτων προτεραιότητας. Η AHP επιτρέπει ελαφρώς μη συνεπείς συγκρίσεις ανά ζεύγη. Ο λόγος συνέπειας (CR) αξιολογεί την επάρκεια των συγκρίσεων, ιδανικά κάτω από 10%. Το βήμα σύνθεσης συνδυάζει τα επιμέρους διανύσματα προτεραιότητας σε έναν πίνακα αποφάσεων ενώ απαιτούνται πολλαπλοί πίνακες απόφασης για M εναλλακτικές λύσεις και N κριτήρια.

2.3 ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΕΘΟΔΟΥ AHP

Ένα παράδειγμα εφαρμογής της μεθόδου αφορά στη διαχείριση των αποβλήτων ακτινοβολημένου γραφίτη (Guidi, Goffo and Violante, 2024) όπου υπάρχουν ανεπαρκείς εγκαταστάσεις αποθήκευσης και στους κανονισμούς λειτουργίας αυτών. Η μέθοδος AHP αποδίδει σημαντική βαρύτητα στα περιβαλλοντικά κριτήρια και στα κριτήρια ασφάλειας των εργαζομένων. Η AHP παρέχει μια τεχνική βάση για την επιλογή βέλτιστων επιλογών διαχείρισης γραφίτη.

Επιπλέον, η διαχείριση των πολιτιστικών πόρων ως προς την κλιματική αλλαγή και τα συμφέροντα που υπάρχουν από τους ενδιαφερόμενους αποτελεί άλλη μια ενδιαφέρουσα εφαρμογή του μοντέλου AHP (Kibria et al., 2024) . Η συγκεκριμένη μελέτη παρέχει πληροφορίες σχετικά με τις προτεραιότητες των ενδιαφερομένων μερών και τη δημιουργία συναίνεσης. Οι στόχοι διαχείρισης επικεντρώνονται στη διατήρηση του ιστορικού χαρακτήρα, την ευαισθητοποίηση σε θέματα πολιτιστικής κληρονομιάς και τα οικονομικά οφέλη. Η μέθοδος AHP διευκόλυνε τη συναίνεση μεταξύ των συμμετεχόντων στην ιεράρχηση των στόχων.

Επιπροσθέτως, η χρήση βιομηχανικών ρομπότ για μηχανουργικές κατεργασίες και οι αποφάσεις επιλογής αυτών αποτελεί άλλο ένα πεδίο εφαρμογής της μεθόδου AHP (Breaz, Bologa and Racz, 2017). Η μέθοδος AHP εφαρμόζεται για την επιλογή της καλύτερης ρομποτικής δομής μεταξύ τριών επιλογών. Το κόστος αποκλείστηκε ως κριτήριο λόγω της παρόμοιας τιμολόγησης για συγκρίσιμα ρομπότ. Ενώ τα νέα κριτήρια που εισήχθησαν περιλαμβάνουν την κατανάλωση ενέργειας και την επιδεξιότητα, που αξιολογούνται ποσοτικά.

Η ενίσχυση της στρατηγικής για την υδροηλεκτρική ενέργεια του Νεπάλ (Singh and Nachtnebel, 2016) αποτελεί άλλο ένα παράδειγμα εφαρμογής της μεθόδου AHP. Μια αξιόπιστη μέθοδος λήψης αποφάσεων είναι ζωτικής σημασίας ελλείψει επαρκών δεδομένων. Η προτίμηση στρέφεται προς τα μεσαία και μεγάλα υδροηλεκτρικά έργα. Το πεδίο των γεωργικών συστημάτων ύδρευσης αποτελεί άλλο ένα παράδειγμα εφαρμογής. Συγκεκριμένα, η μελέτη υπογραμμίζει τη σημασία της εμπλοκής των ενδιαφερόμενων μερών στις διαδικασίες λήψης αποφάσεων που σχετίζονται με τα αρδευτικά συστήματα (Veisi et al., 2022). Υπογραμμίζει την ανάγκη για μια ολοκληρωμένη προσέγγιση που να λαμβάνει υπόψη τις κοινωνικοοικονομικές πτυχές για την επίτευξη της βιωσιμότητας στην αρδευόμενη

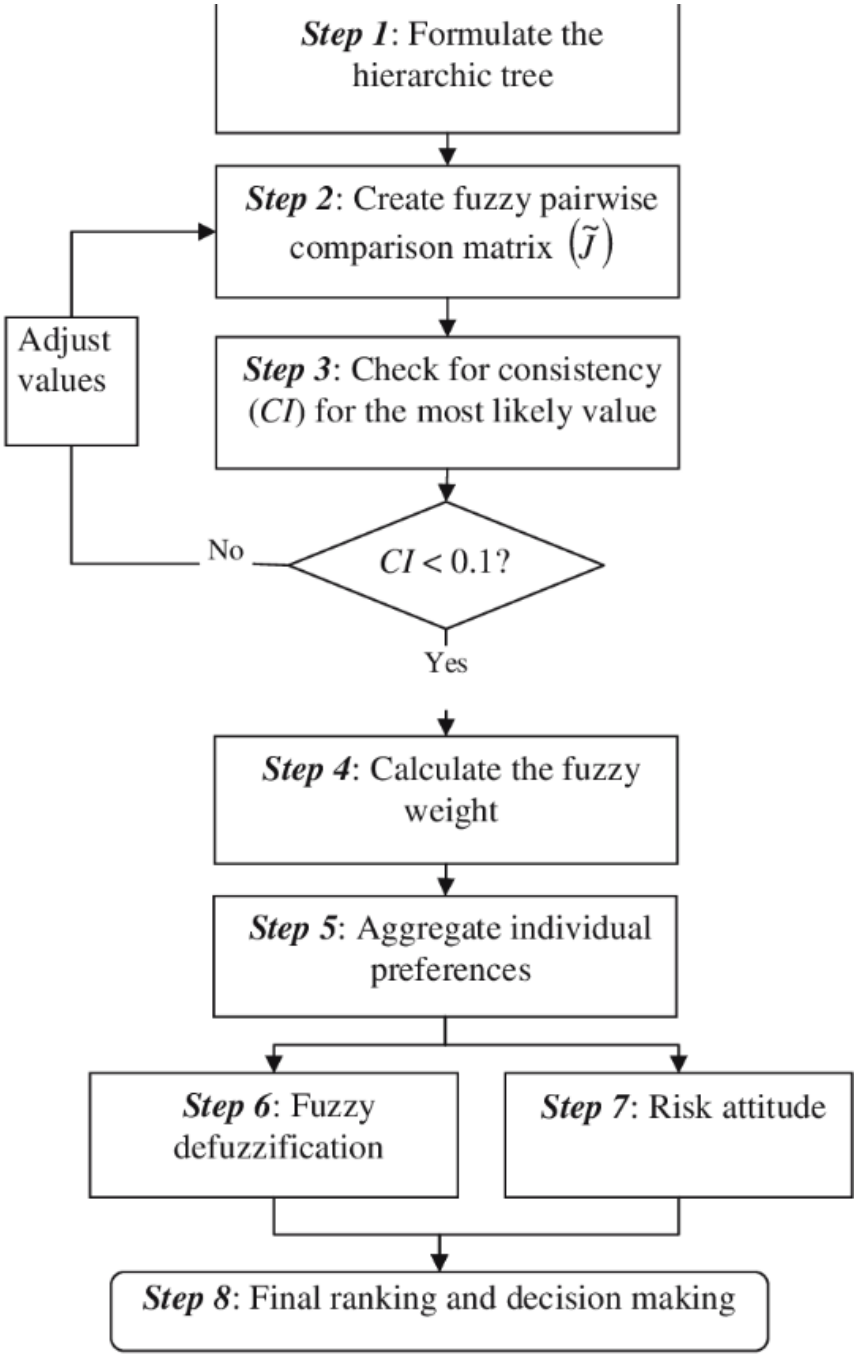
γεωργία. Το πλαίσιο AHP αποδείχθηκε αποτελεσματικό στη δόμηση πολύπλοκων αποφάσεων και θα μπορούσε να εφαρμοστεί και σε άλλα θέματα διαχείρισης των υδάτων.

Η χρήση της μεθόδου επεκτείνεται και στις εφαρμογές web (Khan, Al-Badi and Al-Kindi, 2019) όπου η αξιολόγηση AHP επιβεβαιώνει την υπεροχή του PWA (progressive web application) όσον αφορά το μέγεθος της εφαρμογής, την υποστήριξη πολλαπλών πλατφορμών και τη διαθεσιμότητα εκτός σύνδεσης. Η λειτουργία εκτός σύνδεσης, το μικρό μέγεθος και η ευκολία πρόσβασης της PWA την καθιστούν προτιμότερη. Οι χρήστες μπορούν να περιηγηθούν στην PWA χωρίς εγκατάσταση, σε αντίθεση με τις εγγενείς και τις υβριδικές εφαρμογές.

ΚΕΦΑΛΑΙΟ 3-FAHP ΜΕΘΟΔΟΣ

3.1 ΕΙΣΑΓΩΓΗ-ΕΦΑΡΜΟΓΗ ΤΗΣ ΑΣΑΦΟΥΣ ΛΟΓΙΚΗΣ ΣΤΗΝ AHP

Η ασαφής λογική ενσωματώνεται στη Διαδικασία Αναλυτικής Ιεράρχησης (AHP) για την αντιμετώπιση της αβεβαιότητας και της ασάφειας στη λήψη αποφάσεων, ιδίως στην επιλογή προμηθευτών. Η παραδοσιακή AHP βασίζεται σε ακριβείς αριθμητικές τιμές, οι οποίες ενδέχεται να μην αποτυπώνουν την πολυπλοκότητα των σεναρίων του πραγματικού κόσμου. Με την ενσωμάτωση της ασαφούς λογικής, η διαδικασία λήψης αποφάσεων μπορεί να αντικατοπτρίζει καλύτερα τις υποκειμενικές κρίσεις των υπευθύνων λήψης αποφάσεων (Digalwar, Borade and Metri, 2014).



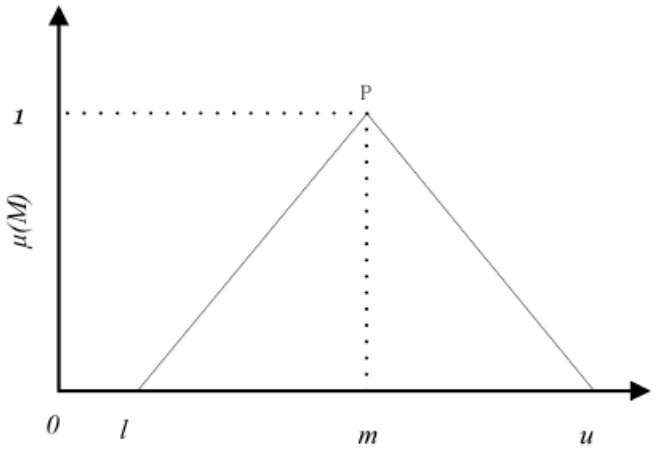
ΕΙΚΟΝΑ 3.1-ΑΝΑΠΑΡΑΣΤΑΣΗ ΜΕΘΟΔΟΥ FAHP
(Tesfamariam et al, 2006)

3.2 ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΣ ΤΩΝ ΚΡΙΤΗΡΙΩΝ

Τα πραγματικά στοιχεία του πίνακα σύγκρισης μετατρέπονται σε ασαφείς αριθμούς. Αυτός ο μετασχηματισμός επιτρέπει την αναπαράσταση ποιοτικών αξιολογήσεων με τη χρήση ασαφών γλωσσικών μεταβλητών, όπως «χαμηλό», «μεσαίο» και «υψηλό». Το βήμα αυτό είναι ζωτικής σημασίας, καθώς επιτρέπει τη μοντελοποίηση των ανακριβών δεδομένων που συχνά αντιμετωπίζουν οι υπεύθυνοι λήψης αποφάσεων.

3.2.1 ΧΡΗΣΗ ΑΣΑΦΩΝ ΑΡΙΘΜΩΝ:

Η μέθοδος χρησιμοποιεί τριγωνικούς ασαφείς αριθμούς (TFN) για την αναπαράσταση των ασαφών γλωσσικών μεταβλητών. Ένας TFN ορίζεται από τρία σημεία, γεγονός που επιτρέπει μια πιο διαφοροποιημένη αναπαράσταση της σημασίας των κριτηρίων. Για παράδειγμα, η ασαφής μεταβλητή «πολύ χαμηλό» μπορεί να αναπαρασταθεί από έναν TFN με συγκεκριμένες ελάχιστες, μέσες και μέγιστες τιμές, διευκολύνοντας την σαφέστερη κατανόηση της σχετικής σημασίας του κριτηρίου.



ΕΙΚΟΝΑ 3.2- Η ΣΥΝΑΡΤΗΣΗ ΣΥΜΜΕΤΟΧΗΣ ΤΟΥ ΤΡΙΓΩΝΙΚΟΥ ΑΣΑΦΟΥΣ ΑΡΙΘΜΟΥ
(Peng *et al.*, 2021)

3.2.2 ΣΥΓΚΡΙΣΕΙΣ ΚΑΤΑ ΖΕΥΓΗ

Στο πλαίσιο της AHP, οι υπεύθυνοι λήψης αποφάσεων πραγματοποιούν συγκρίσεις ανά ζεύγη κριτηρίων και υποκριτηρίων. Με την ασαφή λογική, αυτές οι συγκρίσεις εκφράζονται με ασαφείς αριθμούς, γεγονός που βοηθά στην αποτύπωση της αβεβαιότητας στις κρίσεις

των φορέων λήψης αποφάσεων. Αυτή η διαδικασία ενισχύει την αξιοπιστία των συγκρίσεων, επιτρέποντας ένα εύρος τιμών αντί για έναν ακριβή αριθμό.

3.2.3 ΥΠΟΛΟΓΙΣΜΟΣ ΚΑΤΑΤΑΞΗΣ ΚΑΙ ΒΑΡΥΤΗΤΑΣ:\

Μετά τον καθορισμό των ασαφών πινάκων σύγκρισης, τα επόμενα βήματα περιλαμβάνουν τον υπολογισμό του μέσου όρου των ασαφών αριθμών για την εξαγωγή των βαρών των κριτηρίων. Αυτή η διαδικασία κατάταξης είναι απαραίτητη για τον προσδιορισμό των καταλληλότερων προμηθευτών με βάση τις ασαφείς αξιολογήσεις των κριτηρίων και των υποκριτηρίων.

3.2.4 ΛΗΨΗ ΤΕΛΙΚΗΣ ΑΠΟΦΑΣΗΣ

Το τελικό βήμα στην προσέγγιση ασαφούς AHP περιλαμβάνει την κατάταξη των κριτηρίων με βάση τις ασαφείς συγκρίσεις. Αυτή η κατάταξη βοηθά τους υπεύθυνους λήψης αποφάσεων να κάνουν τεκμηριωμένες επιλογές για την επιλογή προμηθευτών, λαμβάνοντας υπόψη τις εγγενείς αβεβαιότητες στις αξιολογήσεις τους.
Συνοπτικά, η ασαφής λογική ενισχύει τη μεθοδολογία AHP επιτρέποντας μια πιο ευέλικτη και ρεαλιστική αναπαράσταση των κριτηρίων λήψης αποφάσεων, οδηγώντας τελικά σε καλύτερα αποτελέσματα.

3.3 ΑΣΑΦΗΣ ΛΟΓΙΚΗ

Η ασαφής λογική, που εισήχθη από τον Lotfi Zadeh, αντιμετωπίζει την αβεβαιότητα στη λήψη αποφάσεων. Μοντελοποιεί γλωσσικές κατασκευές και παρέχει ένα πλαίσιο για την ανθρώπινη συλλογιστική. Σε αντίθεση με την παραδοσιακή θεωρία δυαδικών συνόλων, η ασαφής λογική δέχεται την ανακρίβεια και την ασάφεια. Τα ασαφή σύνολα είναι χρήσιμα για τη μοντελοποίηση αβέβαιων δεδομένων, ιδίως σε πολύπλοκες εργασίες. Τα προβλήματα του πραγματικού κόσμου συχνά υπερβαίνουν την ακρίβεια των παραδοσιακών μαθηματικών μοντέλων.

Οι ειδικοί χρησιμοποιούν γλωσσικές μεταβλητές για να αξιολογήσουν κριτήρια και εναλλακτικές λύσεις στη λήψη αποφάσεων. Η μελέτη μετατρέπει ακριβείς τιμές σε ασαφείς γλωσσικές μεταβλητές για ανάλυση. Οι τριγωνικοί και τραπεζοειδείς ασαφείς αριθμοί προτιμώνται για την απλότητά τους. Ένας τριγωνικός ασαφής αριθμός (TFN) χρησιμοποιείται για την αναπαράσταση των ασαφών γλωσσικών μεταβλητών στην παρούσα μελέτη. Για τη διευκόλυνση της ανάλυσης ορίζονται συναρτήσεις συμμετοχής για διάφορες ασαφείς μεταβλητές.

3.4 ΕΦΑΡΜΟΓΕΣ ΤΗΣ ΜΕΘΟΔΟΥ FAHP

Μια εφαρμογή της μεθόδου αφορά την επιλογή προμηθευτών κινητήρων γραναζιών σε μια εταιρεία (Ayhan, 2013). Το πρόβλημα αφορά την επιλογή του καλύτερου προμηθευτή για το υλικό «δακτύλιος έδρασης». Τα κριτήρια και τα εναλλακτικά βάρη υπολογίζονται μέσω συγκρίσεων ανά ζεύγη. Η επιλογή προμηθευτή είναι κρίσιμη και περιλαμβάνει αντικρουόμενα κριτήρια. Η ασαφής AHP αντιμετωπίζει αποτελεσματικά τη λήψη αποφάσεων υπό συνθήκες αβεβαιότητας.

Επιπλέον, ένα ασαφές μοντέλο AHP αναπτύχθηκε για την αξιολόγηση των προμηθευτών σε μια εταιρεία χάλυβα (Tahriri, 2014) . Το μοντέλο αντιμετωπίζει αποτελεσματικά τις αβεβαιότητες στη λήψη αποφάσεων. Προσδιορίστηκαν τα βασικά κριτήρια για την επιλογή προμηθευτών, ενισχύοντας την ακρίβεια της λήψης αποφάσεων. Το μοντέλο είναι εφαρμόσιμο σε διάφορα σενάρια επιλογής προμηθευτών στη βιομηχανία χάλυβα.

Εκτός από τα παραπάνω, η μέθοδος FAHP βρίσκει εφαρμογή και στον προσδιορισμό κατάλληλων ζωνών υπόγειου υδάτινου δυναμικού με τη χρήση υδρογεωφυσικών μετρήσεων πεδίου και υδραυλικών παραγόντων του εδάφους σε ένα σύνθετο υδρογεωλογικό έδαφος (Halder et al., 2024). Η μελέτη συγκρίνει τεχνικές μηχανικής μάθησης και ασαφούς AHP για τη χαρτογράφηση του δυναμικού των υπόγειων υδάτων. Η έρευνα τονίζει τη σημασία της διαπερατότητας και της υδραυλικής αγωγιμότητας στο δυναμικό των υπόγειων υδάτων.

Επιπροσθέτως, η μέθοδος χρησιμοποιείται στην επιλογή σύνθετων υλικών που χρησιμοποιούνται σε δομικές εφαρμογές (Singh et al., 2020). Η επιλογή των κατάλληλων υλικών είναι κρίσιμη για τις μηχανολογικές εφαρμογές λόγω της πολυπλοκότητάς τους. Η

εργασία βοηθά τους σχεδιαστές στην επιλογή υλικών με βάση τα απαιτούμενα χαρακτηριστικά.

Η μέθοδος μπορεί να χρησιμοποιηθεί και στην λήψη αποφάσεων σχετικά με την επιλογή διευθυντή ανθρωπίνου δυναμικού (Kusumawardani and Agintiara, 2015). Οι προτεραιότητες επιλογής των υποψηφίων είναι συνεπείς μεταξύ των αξιολογητών, αντανακλώντας τις αξίες της εταιρείας. Ορισμένες περιοχές παρουσιάζουν διαφορετικές προτεραιότητες, οδηγώντας σε αναντιστοιχίες επιλογής. Η ασαφής AHP είναι ένα ικανοποιητικό εργαλείο για τις διαδικασίες επιλογής υποψηφίων.

Η χρήση της μεθόδου υπάρχει και στην ενεργειακή απόδοση της λειτουργίας των πλοίων (Beşikçi et al., 2016). Η εφαρμογή μέτρων ενεργειακής απόδοσης είναι απαραίτητη για την αντιμετώπιση των οικονομικών και περιβαλλοντικών προκλήσεων. Η ασαφής AHP ιεραρχεί αποτελεσματικά τα επιχειρησιακά μέτρα, βοηθώντας τους υπεύθυνους λήψης αποφάσεων στη ναυτιλιακή βιομηχανία.

ΚΕΦΑΛΑΙΟ 4-LLMs ΜΟΝΤΕΛΑ

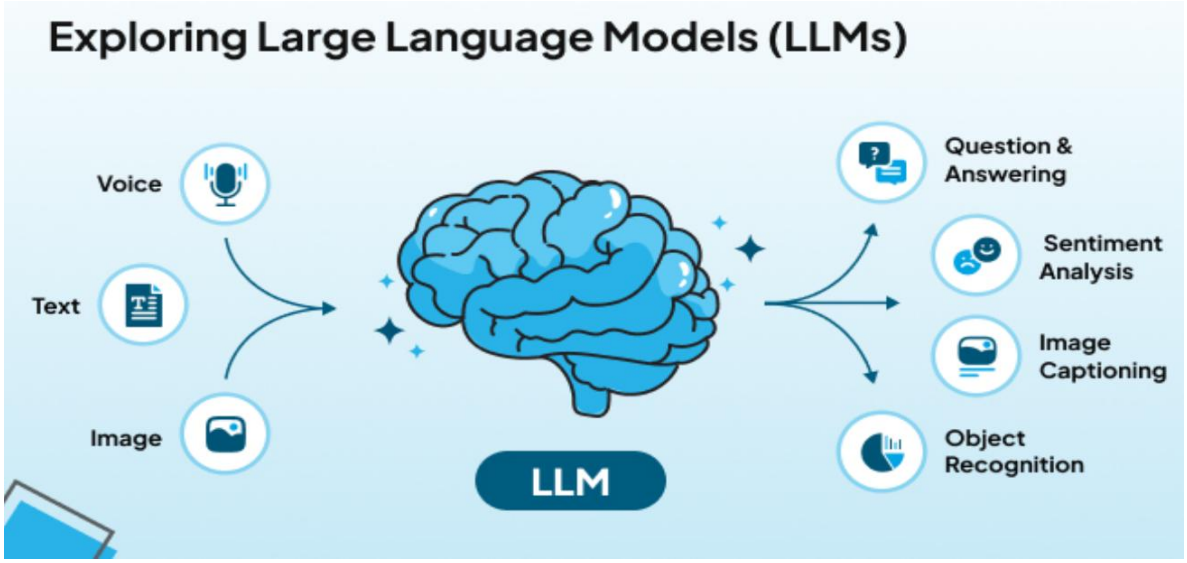
4.1 ΕΙΣΑΓΩΓΗ

Είναι υπολογιστικά μοντέλα που έχουν την ικανότητα να καταλαβαίνουν και να δημιουργούν την ανθρώπινη γλώσσα. Έχουν την δυνατότητα μετασχηματισμού να προβλέπουν την πιθανότητα εμφάνισης ακολουθιών λέξεων ή και απλών λέξεων όπως επίσης και να δημιουργούν καινούργιο κείμενο βασισμένα σε μια δοσμένη είσοδο. Τα N-GRAM μοντέλα εστιάζουν στις πιθανότητες λέξεων βασισμένα σε προηγούμενο περιεχόμενο. Εντούτοις, τα μοντέλα αντιμετωπίζουν κάποιες προκλήσεις όπως π.χ. το γεγονός σπάνιων ή και μη υπαρχόντων λέξεων, το πρόβλημα της υπερ-προσαρμογής (over-fitting) και την δυσκολία αποτύπωσης πολύπλοκων γλωσσικών φαινομένων (Chang et al., 2024).

Τα LLMs είναι προχωρημένα γλωσσικά μοντέλα με τεράστια μεγέθη παραμέτρων και εξαιρετικές μαθησιακές ικανότητες. Σε πολλά μοντέλα υπάρχει η μονάδα αυτό-προσοχής (self-attention) μέσα στον μετασχηματιστή η οποία χρησιμεύει ως το θεμελιώδες δομικό στοιχείο (Module) για τις εργασίες μοντελοποίησης γλώσσας. Οι μετασχηματιστές που υπάρχουν στα LLMs έφεραν επανάσταση στον τομέα της επεξεργασίας φυσικών γλωσσών επιτρέποντας τον χειρισμό διαδοχικών δεδομένων και τον παραλληλισμό καθώς και την αποτύπωση εξαρτήσεων μεγάλης έκτασης μέσα στο κείμενο. Υπάρχει μάθηση του πλαισίου όπου το μοντέλο εκπαιδεύεται να δημιουργεί κείμενο με βάση ένα δεδομένο πλαίσιο. Με αυτό τον τρόπο τα LLMs δημιουργούν πιο συνεκτικές και συναφείς με τα συμφραζόμενα απαντήσεις, καθιστώντας τα κατάλληλα για διαδραστικές και συνομιλητικές εφαρμογές.

Τα LLMs χρησιμοποιούνται κυρίως σε εφαρμογές NLP, με εξαιρετικές επιδόσεις στη δημιουργία κειμένων. Εξειδικευμένα σχέδια, όπως οι μετασχηματιστές, ενισχύουν τις δυνατότητες του μοντέλου- το GPT-3 έχει 175 δισεκατομμύρια παραμέτρους. Το σύνολο δεδομένων εκπαίδευσης του GPT-3 είναι 45 terabytes, αποδίδοντας 570 GB μετά το φιλτράρισμα. Υπάρχουν πολυάριθμα LLM, συμπεριλαμβανομένων των PaLM2, Gemini και Llama από διάφορα ιδρύματα. Οι ερευνητές ΤΝ επικεντρώνονται στη μάθηση μεταφοράς και στην προ-εκπαίδευση μοντέλων από το 2018. Τα μοντέλα BERT και GPT έχουν προχωρήσει σημαντικά τις εργασίες NLP, όπως οι ερωτήσεις και οι απαντήσεις και η ταξινόμηση. Οι βελτιωμένες παραλλαγές BERT όπως οι RoBERTa και DeBERTa χρησιμοποιούνται σε διάφορες εφαρμογές. Οι ανησυχίες σχετικά με την ασφάλεια και τα ηθικά ζητήματα που

αφορούν τα LLM αυξάνονται. Τα LLMs έχουν μετατοπιστεί από την ποσότητα στην ποιότητα, με τάση προς περισσότερες παραμέτρους (Chen et al., 2024).



ΕΙΚΟΝΑ 4.1-ΑΝΑΠΑΡΑΣΤΑΣΗ LLMS
(Team Innovatics,2024)

Οι πρόσφατες εξελίξεις στα LLMs επιτρέπουν την επικοινωνία και τις εκπαιδευτικές εφαρμογές που μοιάζουν με την ανθρώπινη ανταπόκριση. Οι βασικές αρχιτεκτονικές περιλαμβάνουν τις BERT και GPT, με σημαντικές εξελίξεις από την Google και το OpenAI. Τα μοντέλα GPT-3 και GPT-4 παρουσιάζουν εξαιρετικές επιδόσεις σε διάφορες εργασίες φυσικής γλώσσας. Αυτές οι εξελίξεις αναδεικνύουν τον δυνητικό αντίκτυπο των LLMs σε πολλούς τομείς (Dao, 2023).

4.2 ΕΦΑΡΜΟΓΕΣ LLMs

Τα LLM έχουν φέρει επανάσταση στα εκπαιδευτικά καθήκοντα, ενισχύοντας τις δεξιότητες υποβολής ερωτήσεων και την παραγωγή περιεχομένου. Οι τεράστιες δυνατότητές τους υποστηρίζουν διάφορες εκπαιδευτικές εφαρμογές, συμπεριλαμβανομένης της ανατροφοδότησης και των επεξηγήσεων. Τα LLMs παρουσιάζουν δυνατότητες στη δημιουργία εκπαιδευτικών ερωτήσεων, αλλά απαιτούν σε κάποιες περιπτώσεις περαιτέρω βελτίωση (Al Faraby, Romadhony, and Adiwijaya, 2024) . Ένα επιπλέον παράδειγμα εφαρμογής των μοντέλων έχει σχέση με την εκμάθηση ξένης γλώσσας από Βιετναμέζους

μαθητές (Dao, 2023). Η μελέτη συνέκρινε το ChatGPT, το BingChat και το Bard στο σύνολο αγγλικών δεδομένων VNHSGE. Τα αποτελέσματα υποδεικνύουν υψηλή ακρίβεια και δυνατότητες για LLMs στην εκπαίδευση. Η αξιοποίηση των δυνατών σημείων των LLM μπορεί να ενισχύσει τις εξατομικευμένες μαθησιακές εμπειρίες για τους μαθητές.

Εκτός των παραπάνω, τα LLMs χρησιμοποιούνται στον τομέα της υγείας και συγκεκριμένα στο πεδίο του καρκίνου στις γυναίκες (Gumilar et al., 2024). Η μελέτη αναδεικνύει τις δυνατότητες των LLM στην παροχή ακριβών ιατρικών απαντήσεων για τον γυναικολογικό καρκίνο. Το GemAdv αναδείχθηκε ως η κορυφαία επίδοση, με το GPT-4 να παρουσιάζει κάποια πλεονεκτήματα. Το CG-4 όσο και το Copilot παρουσίασαν σφάλματα σε δύσκολες ερωτήσεις, υποδεικνύοντας τομείς προς βελτίωση. Τα LLMs χρησιμεύουν και στην ανάπτυξη κώδικα γλωσσών προγραμματισμού (Husein, Aburajouh and Catal, 2025). Τα LLMs εκπαιδεύονται σε διάφορες πηγές δεδομένων, όπως αποθήκες κώδικα και διαδικτυακό περιεχόμενο. Η προ-εκπαίδευση χρησιμοποιεί μαζικά σύνολα δεδομένων, ακολουθούμενη από λεπτομερή ρύθμιση για συγκεκριμένες εργασίες. Η προετοιμασία των δεδομένων περιλαμβάνει την αφαίρεση του θορύβου, το φιλτράρισμα της ποιότητας και την αποδελτίωση. Οι μέθοδοι Tokenization επηρεάζουν το μέγεθος του λεξιλογίου και την απόδοση του μοντέλου.

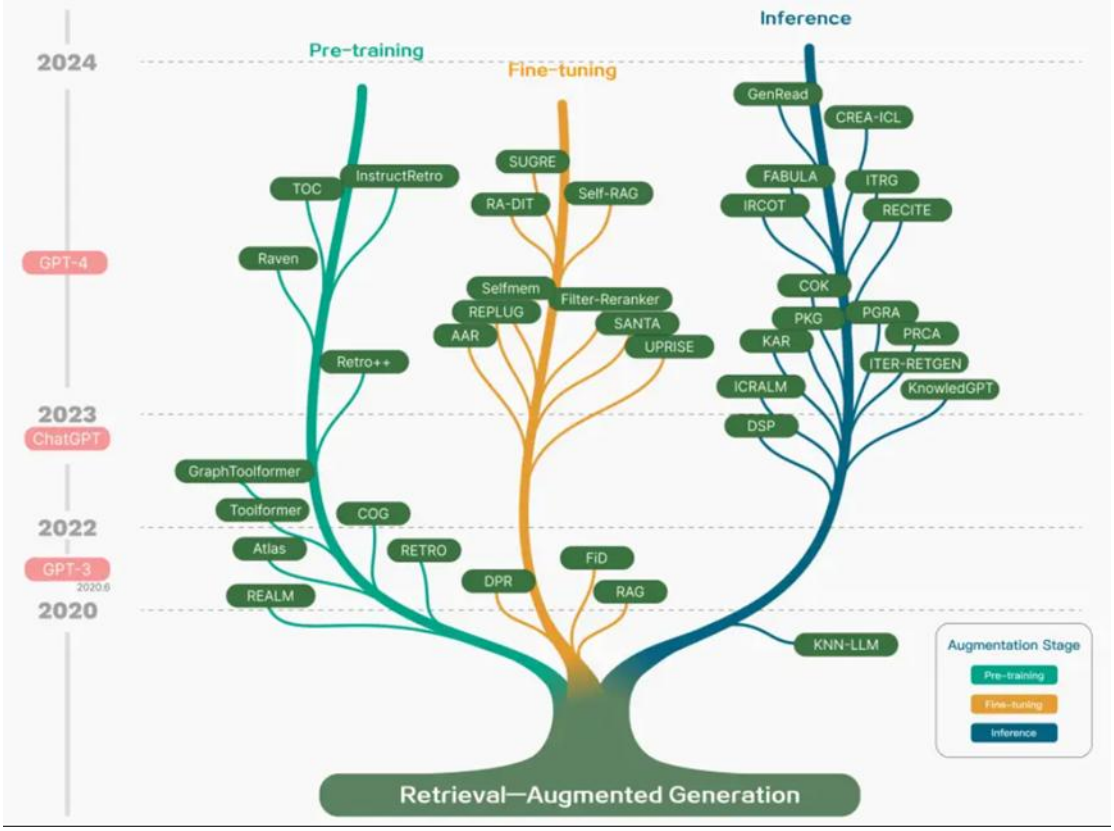
Η χρήση των LLMs στηρίζεται στην εφαρμογή ενός μετασχηματιστή όπου μια νέα αρχιτεκτονική νευρωνικού δικτύου που βασίζεται αποκλειστικά σε μηχανισμούς προσοχής, εξαλείφοντας την ανάγκη για επαναλήψεις και συνελίξεις συνδέει τον κωδικοποιητή και τον αποκωδικοποιητή μέσω της προσοχής. Ο μετασχηματιστής αποδεικνύεται πολλές πιο παραλληλοποιήσιμος, ταχύτερος στην εκπαίδευση και γενικεύεται καλά σε διάφορες εργασίες πέραν της μετάφρασης, αναδεικνύοντας τον σημαντικό αντίκτυπό του στην αποδοτικότητα και την απόδοση των μοντέλων στην επεξεργασία φυσικής γλώσσας (Vaswani et al., 2023).

Γενικά, τα LLMs χειρίζονται διάφορες εργασίες κωδικοποίησης, συμπεριλαμβανομένης της συμπλήρωσης, της κλωνοποίησης και της σύνοψης. Η αξιολόγηση των επιδόσεων λαμβάνει υπόψη την αρχιτεκτονική του μοντέλου και τις ειδικές απαιτήσεις των εργασιών. Οι αρχιτεκτονικές μετασχηματιστών κυριαρχούν στα LLM για εργασίες κωδικοποίησης, ενισχύοντας τη σημασιολογική κατανόηση. Τα διάφορα μοντέλα μετασχηματιστών περιλαμβάνουν αρχιτεκτονικές μόνο κωδικοποιητή, μόνο αποκωδικοποιητή και κωδικοποιητή-αποκωδικοποιητή. Τα LLMs αντιμετωπίζουν προκλήσεις όπως η έλλειψη

γνώσης του τομέα και οι αδιαφανείς παράμετροι. Η λεπτομερής ρύθμιση απαιτεί μεγάλα σύνολα δεδομένων και υπολογιστικούς πόρους, περιπλέκοντας τη διαδικασία.

4.3 ΠΑΡΑΓΩΓΗ ΚΕΙΜΕΝΟΥ ΕΝΙΣΧΥΜΕΝΗ ΜΕ ΑΝΑΚΤΗΣΗ ΠΛΗΡΟΦΟΡΙΑΣ (RAG) ΓΙΑ LLMS

Πολύ ενδιαφέρον παρουσιάζουν τα LLMs με εξωτερική γνώση όπως βάσεις δεδομένων. Η τεχνική ονομάζεται Ανάκτηση Επαυξημένης Γενιάς (RAG) όπου το μοντέλο παίρνει είσοδο και κάνει ανάκτηση από ένα σύνολο υποστηρικτικών εγγράφων από μια πηγή (π.χ. Wikipedia). Τα έγγραφα ενώνονται σας context με την αρχική είσοδο prompt και πηγαίνουν στο text generator όπου παράγει την τελική έξοδο. Βοηθάει όταν υπάρχουν γεγονότα που μετεξελίσσονται. Έτσι δεν χρειάζεται ξανά να επανεκπαιδευτούν. Υπάρχει πρόσβαση στην τελευταία πληροφορία για την αναπαραγωγή αξιόπιστων εξόδων μέσω της retriavel-based generation. Αυτό έχει ως αποτέλεσμα την ακρίβεια, διαχειρισιμότητα , σχετικότητα των απαντήσεων των LLMs.



ΕΙΚΟΝΑ 4.2-ΣΧΗΜΑΤΙΚΗ ΑΝΑΠΑΡΑΣΤΑΣΗ RAG
(Prompting Guide,2024)

4.3.1 ΒΑΣΙΚΑ ΣΤΟΙΧΕΙΑ RAG

Α)Input: Η ερώτηση που το LLM σύστημα απαντάει. Εάν δεν λειτουργεί το RAG το LLM απαντάει άμεσα στην ερώτηση.
Β)Indexing: Ένα το RAG χρησιμοποιείται , βάζουμε τα έγγραφα σε ένα vector store. Το ερώτημα ενσωματώνεται με παρόμοιο τρόπο.
Γ)Retrieval: Τα σχετικά έγγραφα παρατηρούνται μέσω συγκρίσεων με τα διανύσματα. Relevant Documents
Δ)Generation:Τα σχετικά έγγραφα συνδυάζονται με την original prompt σαν επιπλέον περιεχόμενο. Το συνδυασμένο κείμενο και το prompt περνάνε στο μοντέλο για απάντηση – generation που προετοιμάζεται σαν έξοδο για τον χρήστη.

Σε πολλά παραδείγματα όταν χρησιμοποιούμε ένα μοντέλο απευθείας δεν μπορεί να απαντήσει άμεσα λόγω έλλειψης γνώσεων . Με το RAG βάζουμε την σχετική πληροφορία για να απαντήσουμε στην ερώτηση κατάλληλα.

4.3.2. ΕΙΔΗ RAG – ΣΥΣΤΗΜΑΤΩΝ

1)Native RAG
2)Advanced RAG
3)Modular RAG

<u>Native RAG</u>
Παραδοσιακή διαδικασία. Είσοδος χρήστη- query relevant documents, συνδυάζονται με το prompt πηγαίνουν στο μοντέλο και παράγουν στην έξοδο. Χαμηλή ακρίβεια και Low recall.
Όταν εφαρμόζεται η αύξηση, μπορεί επίσης να υπάρχουν προβλήματα με πλεονασμό και επανάληψη. Όταν χρησιμοποιούνται πολλά ανακτημένα αποσπάσματα, η κατάταξη και η συμφωνία του στυλ/τόνου είναι επίσης βασικές. Μια άλλη πρόκληση είναι να διασφαλιστεί ότι η εργασία δημιουργίας δεν εξαρτάται υπερβολικά από τις επαυξημένες πληροφορίες που μπορεί να οδηγήσουν στο μοντέλο να επαναλαμβάνει απλώς το ανακτηθέν περιεχόμενο.

Advanced RAG

Βελτιώνει την ποιότητα ανάκτησης πληροφοριών , βελτιστοποιεί τις pre-retrieval, retrieval και post-retrieval διαδικασίες.

Η διαδικασία προ-ανάκτησης περιλαμβάνει τη βελτιστοποίηση της ευρετηρίασης δεδομένων, η οποία στοχεύει στη βελτίωση της ποιότητας των δεδομένων που ευρετηριάζονται μέσω πέντε σταδίων: ενίσχυση της ευρεσιτεχνίας των δεδομένων, βελτιστοποίηση δομών ευρετηρίου, προσθήκη μεταδεδομένων, βελτιστοποίηση στοίχισης και μικτή ανάκτηση.

Το στάδιο ανάκτησης μπορεί να βελτιωθεί περαιτέρω βελτιστοποιώντας το ίδιο το μοντέλο ενσωμάτωσης, το οποίο επηρεάζει άμεσα την ποιότητα των τμημάτων που συνθέτουν το πλαίσιο.

Αυτό μπορεί να γίνει με βελτιστοποίηση της ενσωμάτωσης για βελτιστοποίηση της συνάφειας ανάκτησης ή χρησιμοποιώντας δυναμικές ενσωματώσεις που αποτυπώνουν καλύτερα την κατανόηση των συμφραζομένων (π.χ. το μοντέλο OpenAI's embeddings-ada-02).

Η βελτιστοποίηση της μετά την ανάκτηση εστιάζει στην αποφυγή των ορίων παραθύρων περιβάλλοντος και στην αντιμετώπιση θορυβωδών ή δυνητικά αποσπών πληροφοριών.

Μια κοινή προσέγγιση για την αντιμετώπιση αυτών των ζητημάτων είναι η ανακατάταξη, η οποία θα μπορούσε να περιλαμβάνει προσεγγίσεις όπως η μετεγκατάσταση του σχετικού περιβάλλοντος στις άκρες της προτροπής ή ο επανυπολογισμός της σημασιολογικής ομοιότητας μεταξύ του ερωτήματος και των σχετικών τμημάτων κειμένου. Η άμεση συμπίεση μπορεί επίσης να βοηθήσει στην αντιμετώπιση αυτών των ζητημάτων.

Modular RAG

Functional Modules , search module για ανάκτηση ομοιοτήτων και εφαρμογή fine-tuning στο retriever. Search, memory, fusion, routing , predict και task adapter που λύνουν διαφορετικά προβλήματα. Επομένως, το Modular RAG επωφελείται από μεγαλύτερη ποικιλομορφία και ευελιξία, καθώς μπορείτε να προσθέσετε ή να αντικαταστήσετε μονάδες ή να προσαρμόσετε τη ροή μεταξύ των μονάδων με βάση τις απαιτήσεις εργασιών.

Δεδομένης της αυξημένης ευελιξίας στην κατασκευή συστημάτων RAG, άλλες σημαντικές τεχνικές βελτιστοποίησης έχουν προταθεί για τη βελτιστοποίηση των αγωγών RAG, όπως:

1)Εξερεύνηση υβριδικής αναζήτησης: Αυτή η προσέγγιση αξιοποιεί έναν συνδυασμό τεχνικών αναζήτησης, όπως αναζήτηση βάσει λέξης-κλειδιού και σημασιολογική αναζήτηση, για την ανάκτηση σχετικών πληροφοριών και πληροφοριών πλούσιες σε περιεχόμενο. Αυτό είναι χρήσιμο όταν αντιμετωπίζουμε διαφορετικούς τύπους ερωτημάτων και ανάγκες πληροφοριών.

2)Αναδρομική ανάκτηση και μηχανή ερωτημάτων: Περιλαμβάνει μια αναδρομική διαδικασία ανάκτησης που μπορεί να ξεκινήσει με μικρά σημασιολογικά κομμάτια και στη συνέχεια να ανακτήσει μεγαλύτερα κομμάτια που εμπλουτίζουν το περιβάλλον. Αυτό είναι χρήσιμο για την εξισορρόπηση της αποτελεσματικότητας και των πληροφοριών πλούσιες σε περιβάλλον.

3)StepBack-prompt: Μια τεχνική προτροπής που επιτρέπει στα LLM να εκτελούν αφαίρεση που παράγει έννοιες και αρχές που καθοδηγούν τη συλλογιστική. Αυτό οδηγεί σε καλύτερα θεμελιωμένες απαντήσεις όταν υιοθετείται σε ένα πλαίσιο RAG, επειδή το LLM απομακρύνεται από συγκεκριμένες περιπτώσεις και επιτρέπεται να αιτιολογεί ευρύτερα εάν χρειάζεται.

ΚΕΦΑΛΑΙΟ 5-ΥΛΟΠΟΙΗΣΗ

5.1 LLM – OLLAMA PYTHON ΒΙΒΛΙΟΘΗΚΕΣ

Στην παρούσα έρευνα χρησιμοποιούνται οι βιβλιοθήκες ollama και διάφοροι μετασχηματιστές με σκοπό την ενσωμάτωση Μεγάλων Γλωσσικών Μοντέλων (LLMs) με την Python για να παρέχουν δυνατότητες δημιουργίας chatbot και κειμένου. Αρχικά έγινε εγκατάσταση της βιβλιοθήκης Ollama και του διακομιστή Ollama. Με εγκατεστημένο τον διακομιστή ollama και το αντίστοιχο πακέτο της python, έγινε ανάκτηση του μοντέλου LLM mistral της βιβλιοθήκης ollama. Το μοντέλο mistral είναι ένα σχετικά μικρό (7B παράμετροι) LLM που μπορεί να εκτελεστεί στις περισσότερες CPUs. Εκτός από το παραπάνω μοντέλο χρησιμοποιήθηκαν δύο ακόμα το phi και το gemma.

Η βιβλιοθήκη ollama python διευκολύνει τις LLMs σε εφαρμογές όπως chatbots, πράκτορες υποστήριξης πελατών και εργαλεία παραγωγής περιεχομένου. Η δημιουργία κώδικα, η αποσφαλμάτωση και η υποστήριξη προγραμματισμού σε διαφορετικές γλώσσες μπορούν να επιταχυνθούν με LLMs, αν χρησιμοποιηθούν αποτελεσματικά. Η βιβλιοθήκη ollama απλοποιεί την αλληλεπίδραση με προηγμένα μοντέλα LLM που επιτρέπουν πιο εξελιγμένες απαντήσεις και δυνατότητες.

Μετά την εγκατάσταση των μοντέλων δόθηκε σε κάθε μοντέλο η ερώτηση: «'What is the command to print output in Python?». Το κάθε μοντέλο σαν έξοδο έδωσε τις δικές του απαντήσεις . Ενδεικτικά έχουμε τις παρακάτω απαντήσεις:

- **Mistral**:In Python, you can use the `print()` function to display output.
- **Gemma**:The `print()` function is used to display data or text on the console or other output devices.
- **Phi**:To print output in Python, you can use the `print()` function

Το κείμενο αναφοράς με βάση το οποίο έγινε η σύγκριση ακρίβειας των μοντέλων είναι το ακόλουθο:

Reference: The print() function.

5.2 ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ LLMS

Επειδή η αξιολόγηση της LLM απαιτεί πολλές εκτιμήσεις, καμία μετρική δεν μπορεί να αποτυπώσει κάθε πτυχή της απόδοσης. Ακόμη και οι στατιστικές μετρικές που εξετάζονται παρακάτω αφορούν μόνο ορισμένες πτυχές της συμπεριφοράς της LLM. Ειδικότερα, ενώ αυτές οι μέθοδοι είναι χρήσιμες για τη μέτρηση πτυχών όπως η επιφανειακή ομοιότητα, μπορεί να μην αποτυπώνουν πλήρως τη βαθύτερη συλλογιστική ή τη σημασιολογική κατανόηση. Πρόσθετες ή συμπληρωματικές μέθοδοι αξιολόγησης (όπως νεότερες μετρικές π.χ. το BERTScore) μπορεί να είναι απαραίτητες για μια ολοκληρωμένη αξιολόγηση. Στην συνέχεια αναφέρουμε τις μετρικές που χρησιμοποιήθηκαν στην εφαρμογή μας.

Το ROUGE, ή Recall-Oriented Understudy for Gisting Evaluation, είναι μια συλλογή από μετρικές που χρησιμοποιούνται για την αξιολόγηση της απόδοσης εξόδου LLM από την άποψη της ανάκλησης. Αρχικά δημοσιεύθηκε από τον Lin, το ROUGE σχεδιάστηκε για την αξιολόγηση της αυτόματης περίληψης, αλλά έκτοτε έχει εφαρμοστεί σε διάφορες εργασίες γλωσσικών μοντέλων, συμπεριλαμβανομένης της μετάφρασης. Το ROUGE μετρά την επικάλυψη μεταξύ της παραγόμενης εξόδου και των κειμένων αναφοράς. Ωστόσο, το ROUGE δίνει μεγαλύτερη έμφαση στην ανάκληση, γεγονός που το καθιστά ιδιαίτερα χρήσιμο όταν ο στόχος είναι η σύλληψη όλων των κρίσιμων πληροφοριών από την αναφορά. Υπάρχουν διάφορες παραλλαγές του ROUGE.

ROUGE-N

Το ROUGE-N υπολογίζεται ως η επικάλυψη των n-γραμμάτων μεταξύ του κειμένου εξόδου και του κειμένου αναφοράς. Η ακόλουθη εικόνα δείχνει την εξίσωσή του:

$$ROUGE - N = \frac{\sum_{n-gram \in reference} min(count_{candidate}, count_{reference})}{\sum_{n-gram \in reference} count_{reference}}$$

ΕΞΙΣΩΣΗ 5.1-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-N

Αυτή η μετρική υπολογίζει την αναλογία των επικαλυπτόμενων n-grams, περικόπτοντας τις μετρήσεις για να αποφευχθεί η υπεραντιπροσώπευση, και κανονικοποιεί με τον συνολικό αριθμό των n-grams στην αναφορά.

Για τις ανάγκες της εργασίας χρησιμοποιούνται δύο παραλλαγές του συγκεκριμένου κριτηρίου που ονομάζονται ROUGE-1 και ROUGE-2.

ROUGE-1

Μετρά την επικάλυψη μονόγραμμα (μία λέξη) μεταξύ του παραγόμενου κειμένου και του κειμένου αναφοράς. Επιπλέον, μετράει πόσες λέξεις στο παραγόμενο κείμενο υπάρχουν και στο κείμενο αναφοράς. Επικεντρώνεται κυρίως στην επικάλυψη περιεχομένου σε επίπεδο λέξης. Μια υψηλή βαθμολογία ROUGE-1 υποδηλώνει ότι το παραγόμενο κείμενο χρησιμοποιεί πολλές από τις ίδιες λέξεις με το κείμενο αναφοράς. Συνήθως αναφέρει ακρίβεια, ανάκληση και βαθμολογία F1.

ROUGE-2

Μετρά την επικάλυψη bigram (δύο διαδοχικές λέξεις) μεταξύ του παραγόμενου κειμένου και του κειμένου αναφοράς. Συγκεκριμένα, μετράει πόσα ζεύγη διαδοχικών λέξεων στο παραγόμενο κείμενο υπάρχουν και στο κείμενο αναφοράς, διατηρώντας τη σειρά τους. Πηγαίνει πέρα από τις μεμονωμένες λέξεις και εξετάζει την ομοιότητα σε επίπεδο ροής και φράσεων. Μια υψηλή βαθμολογία ROUGE-2 υποδηλώνει ότι το παραγόμενο κείμενο χρησιμοποιεί πολλά από τα ίδια ζεύγη λέξεων με το κείμενο αναφοράς, γεγονός που υποδηλώνει καλύτερη τοπική συνοχή και προσήλωση στη διατύπωση του κειμένου αναφοράς. Ο υπολογισμός γίνεται με παρόμοιο τρόπο με το ROUGE-1, αναφέρει ακρίβεια, ανάκληση και ένα F1-score με βάση την επικάλυψη bigram.

ROUGE-L

Σε αντίθεση με το ROUGE-N, το ROUGE-L χρησιμοποιεί τη μεγαλύτερη κοινή υποδιαστολή (LCS) για τη μέτρηση της ομοιότητας των προτάσεων. Βρίσκει τη μεγαλύτερη ακολουθία λέξεων που εμφανίζεται τόσο στην έξοδο όσο και στην αναφορά, ακόμη και αν οι λέξεις δεν είναι διαδοχικές, εφόσον διατηρούν την ίδια σειρά.

$$ROYGE - L = \frac{LCS\ length}{reference\ length}$$

ΕΞΙΣΩΣΗ 5.2-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-L

Αυτή η μετρική είναι ιδιαίτερα καλή για την αξιολόγηση της ευχέρειας και της γραμματικής συνοχής σε εργασίες περίληψης και παραγωγής κειμένων.

METEOR (Μετρική για την αξιολόγηση της μετάφρασης με ρητή εντολή)

Το METEOR, ή Metric for Evaluation of Translation with Explicit ORdering, είναι μια μετρική που εισήχθη από τους Banerjee και Lavie για την αξιολόγηση των αποτελεσμάτων LLM με σύγκριση με κείμενα αναφοράς. Αν και παρόμοια με τις BLEU και ROUGE, η METEOR τις βελτιώνει ενσωματώνοντας εκτιμήσεις για τα συνώνυμα, το stemming και τη σειρά των λέξεων. Το METEOR βασίζεται στο σκορ F1 - τον αρμονικό μέσο όρο της ακρίβειας και της ανάκλησης - δίνοντας πρόσθετη βαρύτητα στην ανάκληση. Αυτή η έμφαση διασφαλίζει ότι η μετρική επιβραβεύει τις εξόδους που καταγράφουν περισσότερο από το περιεχόμενο αναφοράς.

Ο τύπος METEOR έχει ως εξής:

$$METEOR = (1 - P) \times F1 + P \times Fragmentation\ Penalty$$

ΕΞΙΣΩΣΗ 5.3-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ METEOR

Στην εξίσωση αυτή, το P αντιπροσωπεύει το βάρος που αποδίδεται στην ποινή και το F1 είναι ο αρμονικός μέσος όρος της ακρίβειας και της ανάκλησης. Η βαθμολογία METEOR, η οποία συνήθως κυμαίνεται από 0 έως 1, με βαθμολογίες άνω του 0,4 να θεωρούνται καλές.

5.3 ΕΠΙΠΛΕΟΝ ΣΤΑΤΙΣΤΙΚΕΣ ΜΕΘΟΔΟΙ ΑΞΙΟΛΟΓΗΣΗΣ LLMS

Εκτός των παραπάνω κριτηρίων γίνεται παράθεση επιπλέον κριτηρίων τα οποία μπορούν να χρησιμοποιηθούν σε μελλοντικές έρευνες.

ROUGE-W

Το ROUGE-W είναι μια σταθμισμένη έκδοση του ROUGE-L, δίνοντας πρόσθετη σημασία. Οι μακρύτερες διαδοχικές ακολουθίες αποδίδουν υψηλότερη βαθμολογία λόγω της τετραγωνικής στάθμισης.

$$ROUGE - W = \frac{L_W}{reference\ length}$$

ΕΞΙΣΩΣΗ 5.4-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-W

ROUGE-S

Το ROUGE-S επιτρέπει την αντιστοίχιση με παραλείψεις, δηλαδή εξετάζει ζεύγη λέξεων που εμφανίζονται με τη σωστή σειρά αλλά δεν είναι απαραίτητα διαδοχικά. Αυτό παρέχει ένα ευέλικτο μέτρο σημασιολογικής ομοιότητας.

$$ROUGE - S = \frac{Matched\ skip - bigrams}{Total\ skip - bigrams\ in\ reference}$$

ΕΞΙΣΩΣΗ 5.5-ΤΥΠΟΣ ΥΠΟΛΟΓΙΣΜΟΥ ROUGE-S

Η ευελιξία του ROUGE-S το καθιστά κατάλληλο για την αξιολόγηση αποτελεσμάτων όπου η ακριβής αντιστοίχιση φράσεων είναι λιγότερο κρίσιμη από την αποτύπωση του συνολικού νοήματος.

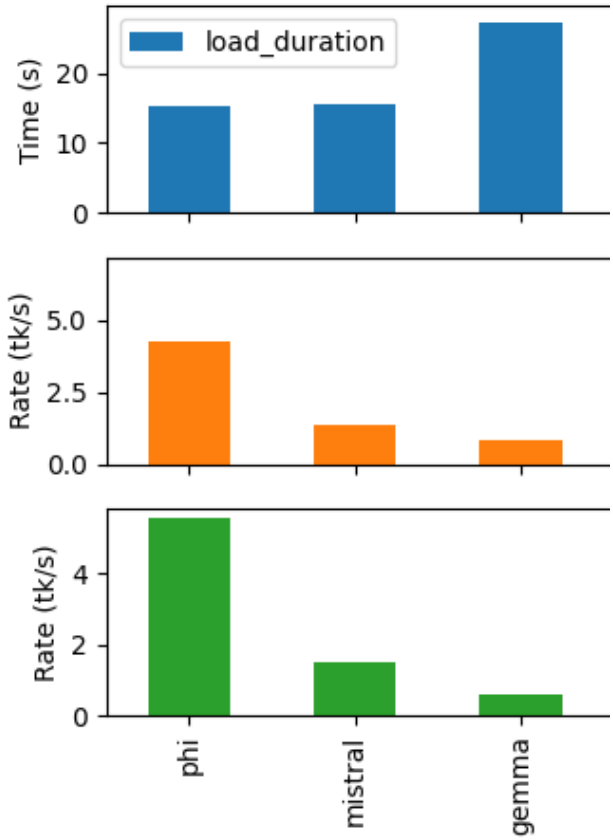
Οι βαθμολογίες ROUGE κυμαίνονται συνήθως από 0 έως 1. Σε πολλές εφαρμογές, μια βαθμολογία άνω του 0,4 θεωρείται καλή. Το παραπάνω παράδειγμα δείχνει ότι η έξοδος LLM έχει καλή απόδοση σύμφωνα με αυτές τις μετρικές. Αυτή η ενότητα καταδεικνύει ότι ενώ η ROUGE προσφέρει πολύτιμες πληροφορίες σχετικά με την ανάκληση και την ευχέρεια, θα πρέπει ιδανικά να χρησιμοποιείται μαζί με άλλες μετρικές για μια πλήρη αξιολόγηση.

5.4 ΜΕΤΡΗΣΕΙΣ ΕΠΙΔΟΣΕΩΝ LLMs

Στην συνέχεια έγινε σύγκριση των επιδόσεων των μοντέλων τοπικά στον υπολογιστή μας. Μία από τις μετρήσεις αφορά την ταχύτητα απόκρισης σε σχέση με το μέγεθος του κάθε μοντέλου. Έγινε συλλογή μετρήσεις επιδόσεων, όπως ο χρόνος που απαιτείται για τη δημιουργία απαντήσεων, καθώς και τα ποσοστά αξιολόγησης και προτροπής. Η αποθήκευση των αποτελεσμάτων υλοποιήθηκε σε ένα πλαίσιο δεδομένων pandas DataFrame όπου έγινε και η εκτέλεση ορισμένων πρόσθετων υπολογισμών (π.χ. calc_duration, prompt_rate, eval_rate).

Παρακάτω παραθέτουμε σε γραφικές παραστάσεις τα αποτελέσματα που προέκυψαν:

ΔΙΑΓΡΑΜΜΑ 5.1-ΑΠΟΔΟΣΗ LLMS ΩΣ ΠΡΟΣ CALC_DURATION, PROMPT_RATE, EVAL_RATE



Τα αποτελέσματα των παραπάνω κριτηρίων στο περιβάλλον Python με ακριβείς αριθμητικές τιμές είναι τα ακόλουθα:

1)**phi**:LOAD_DURATION=15.303293

,PROMPT_RATE=4.265219,EVALUATION_RATE=5.540166

2)**mistral**: LOAD_DURATION

=15.693258, PROMPT_RATE=1.367439, EVALUATION_RATE=1.511097

3)**gemma**:LOAD_DURATION=27.167443,PROMPT_RATE=0.844511,EVALUATION_RATE=0.624581

Ο κώδικας επιλέγει τρεις στήλες (load_duration, prompt_rate, eval_rate) και δημιουργεί ένα ραβδόγραμμα για κάθε μία, το οποίο δείχνει την απόδοση κάθε μοντέλου. Στην πραγματικότητα ο κώδικας ρωτά τα μοντέλα τεχνητής νοημοσύνης ('phi', 'mistral', 'gemma') για τη δημιουργία απαντήσεων με βάση την ίδια ερώτηση. Συλλέγει μετρήσεις απόδοσης, όπως ο χρόνος που απαιτείται για τη δημιουργία απαντήσεων, καθώς και τα ποσοστά αξιολόγησης και προτροπής. Αποθηκεύει τα αποτελέσματα σε ένα πλαίσιο δεδομένων pandas DataFrame και εκτελεί ορισμένους πρόσθετους υπολογισμούς (π.χ. calc_duration, prompt_rate, eval_rate). Τέλος, ο κώδικας παράγει ραβδογράμματα για την οπτικοποίηση της απόδοσης των μοντέλων και αποθηκεύει τα γραφήματα ως εικόνα όπως παρουσιάστηκε παραπάνω.

Το calc_duration και load_duration αναφέρονται στον χρόνο υπολογισμού και «φόρτωσης» της κάθε απάντησης από το κάθε μοντέλο. Η μονάδα μέτρησης είναι σε δευτερόλεπτα. Ο όρος "prompt rate" χρησιμοποιείται σε διάφορα πλαίσια, τα οποία σχετίζονται με την ταχύτητα και την αποτελεσματικότητα των prompts-απαντήσεων των μοντέλων. Ο όρος "eval rate" αναφέρεται στον ρυθμό αξιολόγησης ή τον ρυθμό με τον οποίο κάτι αξιολογείται.

Στην συνέχεια παραθέτουμε ενδεικτικές τιμές των LLMs σύμφωνα με τα κριτήρια που προαναφέρθηκαν.

Rouge-N

Mistral:

ROUGE1:Score(precision=0.2727272727272727,recall=1.0,fmeasure=0.42857142857142855)

ROUGE-2: Score(precision=0.2, recall=1.0, fmeasure=0.33333333333333337)

ROUGE-L:Score(precision=0.2727272727272727,recall=1.0,
fmeasure=0.42857142857142855)

Gemma:

ROUGE-1: Score(precision=0.17647058823529413, recall=1.0, fmeasure=0.3)

ROUGE-2: Score(precision=0.125, recall=1.0, fmeasure=0.2222222222222222)

ROUGE-L: Score(precision=0.17647058823529413, recall=1.0, fmeasure=0.3)

Phi:

ROUGE-1:Score(precision=0.2727272727272727,recall=1.0,
fmeasure=0.42857142857142855)

ROUGE-2: Score(precision=0.2, recall=1.0, fmeasure=0.33333333333333337)

ROUGE-L:Score(precision=0.2727272727272727,recall=1.0,
fmeasure=0.42857142857142855)

METEOR

Mistral:

METEOR Score: 0.7311475409836066

Gemma:

METEOR Score: 0.6757575757575758

Phi:

METEOR Score: 0.7311475409836066

5.5 ΕΦΑΡΜΟΓΗ ΜΕΘΟΔΩΝ AHP/FAHP

5.5.1 ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ ΜΕΘΟΔΟΥ

Η πρώτη υλοποίηση των μεθόδων ιεραρχικής προτεραιότητας εστιάζεται περισσότερο στην ομοιότητα παρά στην χρονική απόδοση και τους ρυθμούς προτροπής και αξιολόγησης. Παρακάτω παραθέτουμε τις βαρύτητες των κριτηρίων της πρώτης υλοποίησης.

1. **LOAD DURATION (Διάρκεια Φόρτωσης):**
    - Βαθμολογία: 6/9
2. **PROMPT RATE (Ρυθμός Προτροπής):**
    - Βαθμολογία: 7/9

3. **EVALUATION RATE (Ρυθμός Αξιολόγησης):**
    - Βαθμολογία 6/9
4. **ROUGE-1, ROUGE-2, ROUGE-L:**
    - Βαθμολογία: 8/9
5. **METEOR:**
    - Βαθμολογία: 9/9 (Κρίσιμο για την αξιολόγηση της ποιότητας της παραγωγής κειμένου)

Δημιουργήθηκε ο παρακάτω πίνακας κριτηρίων με βάση την συγκεκριμένη σπουδαιότητα.

ΠΙΝΑΚΑΣ 5.1-ΠΙΝΑΚΑΣ ΚΡΙΤΗΡΙΩΝ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ

Κριτήριο	LOAD DURATION	PROMPT RATE	EVALUATION RATE	ROUGE-1	ROUGE-2	ROUGE-L	METEOR
LOAD DURATION	1	1/2	1	1/3	1/4	1/4	1/5
PROMPT RATE	2	1	2	1/2	1/3	1/3	1/4
EVALUATION RATE	1	1/2	1	1/3	1/4	1/4	1/5
ROUGE-1	3	2	3	1	1/2	1/2	1/3
ROUGE-2	4	3	4	2	1	1	1/2
ROUGE-L	4	3	4	2	1	1	1/2
METEOR	5	4	5	3	2	2	1

Οι παραπάνω τιμές έχουν την κάτωθι ερμηνεία:

- 1: Ίση σημασία
- 3: Μέτρια σημαντικότερο
- 5: Αρκετά σημαντικότερο
- 7: Πολύ σημαντικότερο
- 9: Απόλυτα σημαντικότερο.
- Οι ενδιάμεσες τιμές (2, 4, 6, 8) χρησιμοποιούνται για ενδιάμεσες τιμές.

- Τα κλάσματα (1/2, 1/3, κ.λπ.) αντιπροσωπεύουν την αντίστροφη σύγκριση.

Κάθε κελί στον πίνακα αντιπροσωπεύει τη σύγκριση δύο κριτηρίων. Για παράδειγμα, το κελί στη γραμμή "PROMPT RATE" και τη στήλη "LOAD DURATION" έχει την τιμή 2, που σημαίνει ότι το PROMPT RATE είναι ελαφρώς πιο σημαντικό από το LOAD DURATION.Τα κελιά στην διαγώνιο του πίνακα έχουν την τιμή 1, διότι ένα κριτήριο έχει την ίδια σημασία με τον εαυτό του. Τα κελιά κάτω από την διαγώνιο είναι οι αντίστροφες τιμές των κελίων πάνω από την διαγώνιο.

Στην συνέχεια δημιουργήθηκαν πίνακες σύγκρισης για κάθε κριτήριο ξεχωριστά, χρησιμοποιώντας τις τιμές των LLMs. Χρησιμοποιείται η κλίμακα της AHP (1-9) για την σύγκριση κάθε ζεύγους LLM για κάθε κριτήριο.

1. Κριτήριο: LOAD DURATION (Διάρκεια Φόρτωσης)

- **Τιμές:**
  - phi: 15.303293
  - mistral: 15.693258
  - gemma: 27.167443
- **Πίνακας Σύγκρισης:**

ΠΙΝΑΚΑΣ 5.2-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-LOAD DURATION

LOAD DURATION	phi	mistral	gemma
phi	1	2	5
mistral	1/2	1	4
gemma	1/5	1/4	1

2. Κριτήριο: PROMPT RATE (Ρυθμός Προτροπής)

- **Τιμές:**
  - phi: 4.265219
  - mistral: 1.367439
  - gemma: 0.844511
- **Πίνακας Σύγκρισης:**

ΠΙΝΑΚΑΣ 5.3-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-PROMPT RATE

PROMPT RATE	phi	mistral	gemma
phi	1	4	7
mistral	1/4	1	3
gemma	1/7	1/3	1

3. Κριτήριο: EVALUATION RATE (Ρυθμός Αξιολόγησης)

- **Τιμές:**
  - phi: 5.540166
  - mistral: 1.511097
  - gemma: 0.624581
- **Πίνακας Σύγκρισης:**

ΠΙΝΑΚΑΣ 5.4-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-EVALUATION RATE

EVALUATION RATE	phi	mistral	gemma
phi	1	4	7
mistral	1/4	1	3
gemma	1/7	1/3	1

4. Κριτήριο: ROUGE-1 (F-measure)

- **Τιμές:**
  - phi: 0.4286
  - mistral: 0.4286
  - gemma: 0.3
- **Πίνακας Σύγκρισης:**

ΠΙΝΑΚΑΣ 5.5-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-1

ROUGE-1	phi	mistral	gemma
phi	1	1	5
mistral	1	1	5

gemma	1/5	1/5	1

5. Κριτήριο: ROUGE-2 (F-measure)

- **Τιμές:**
  - phi: 0.3333
  - mistral: 0.3333
  - gemma: 0.2222
- **Πίνακας Σύγκρισης:**

ΠΙΝΑΚΑΣ 5.6-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-2

ROUGE-2	phi	mistral	gemma
phi	1	1	5
mistral	1	1	5
gemma	1/5	1/5	1

6. Κριτήριο: ROUGE-L (F-measure)

- **Τιμές:**
  - phi: 0.4286
  - mistral: 0.4286
  - gemma: 0.3
- **Πίνακας Σύγκρισης:**

ΠΙΝΑΚΑΣ 5.7-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-L

ROUGE-L	phi	mistral	gemma
phi	1	1	5
mistral	1	1	5
gemma	1/5	1/5	1

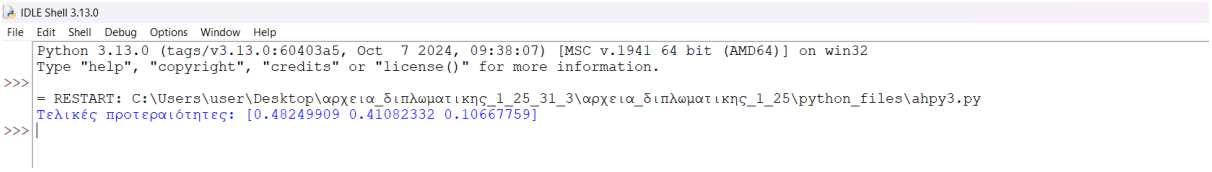
7. Κριτήριο: METEOR

- **Τιμές:**
  - phi: 0.7311
  - mistral: 0.7311
  - gemma: 0.6758
- **Πίνακας Σύγκρισης:**

ΠΙΝΑΚΑΣ 5.8-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ-METEOR

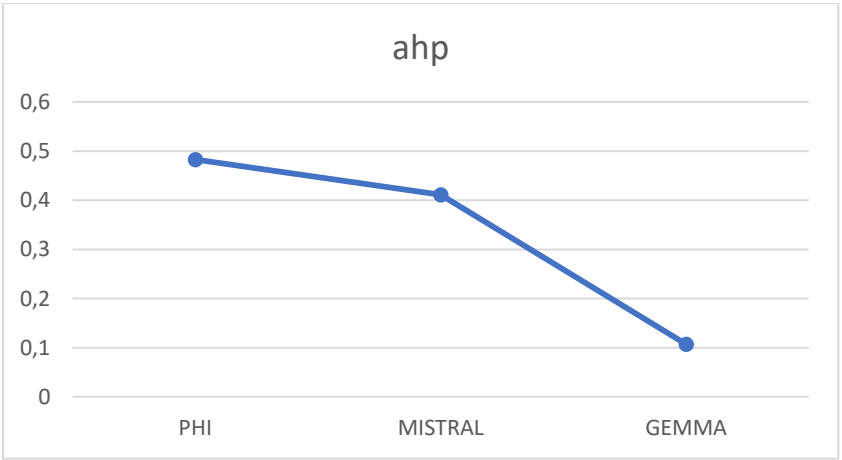
METEOR	phi	mistral	gemma
phi	1	1	3
mistral	1	1	3
gemma	1/3	1/3	1

Οι τελικές προτεραιότητες μετά την εκτέλεση της μεθόδου AHP σε περιβάλλον Python είναι οι παρακάτω:



ΕΙΚΟΝΑ 5.1-ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΩΤΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-AHP

ΔΙΑΓΡΑΜΜΑ 5.2-ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS AHP ΜΕΘΟΔΟΣ

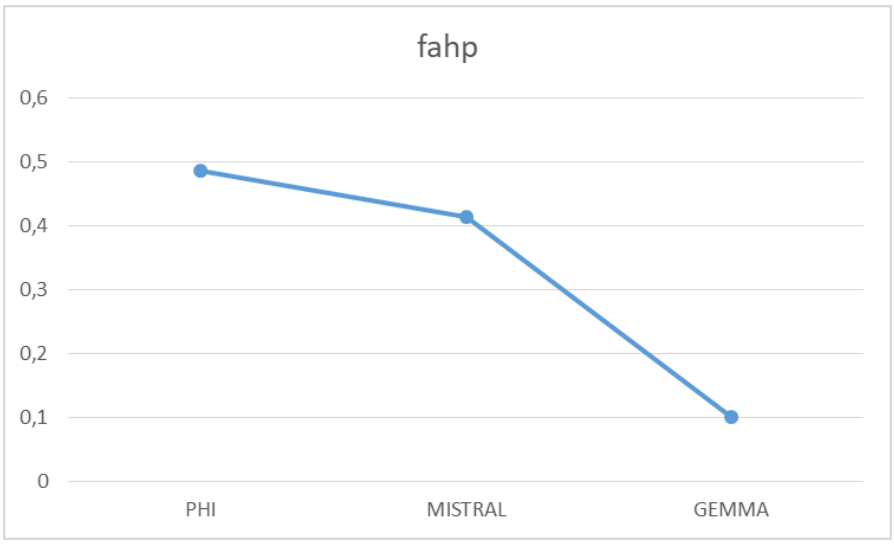


Ενώ μετά την εκτέλεση της μεθόδου FAHP προέκυψαν τα παρακάτω αποτελέσματα:

{'criteria': {'LOAD DURATION': np.float64(0.047953546544956066), 'PROMPT RATE': np.float64(0.07475431214532051), 'EVALUATION RATE': np.float64(0.04795354
6544956066), 'ROUGE-1': np.float64(0.12013538823394469), 'ROUGE-2': np.float64(0.2013731243399024), 'ROUGE-L': np.float64(0.2013731243399024), 'METEOR':
np.float64(0.30974987336908993)}, 'alternatives': {'Phi': np.float64(0.4849618966364544), 'Mistral': np.float64(0.41356488828458476), 'Gemma': np.float64
(0.10147321507896094)}}

ΕΙΚΟΝΑ 5.2-ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΩΤΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-FAHP

ΔΙΑΓΡΑΜΜΑ 5.3-ΠΡΩΤΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS FAHP ΜΕΘΟΔΟΣ



Βάσει των αποτελεσμάτων τα μοντέλα PHI και MISTRAL με βάση την σημαντικότητα των κριτηρίων που παραθέσαμε παραπάνω εμφανίζουν παρόμοια αποτελέσματα με σαφή διαφορά σε σχέση με το τελευταίο σε απόδοση μοντέλο που είναι το GEMMA.

5.5.2 ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ ΜΕΘΟΔΟΥ

Η δεύτερη υλοποίηση των μεθόδων ιεραρχικής προτεραιότητας βασίζεται περισσότερο στα κριτήρια του LOAD DURATION σε σχέση με τα υπόλοιπα που είχαν αυξημένη προτεραιότητα στην προηγούμενη εφαρμογή των μεθόδων. Παρακάτω παραθέτουμε τις βαρύτητες των κριτηρίων της δεύτερης υλοποίησης.

- LOAD DURATION: 9
- PROMPT RATE: 6
- EVALUATION RATE: 6
- ROUGE-1, ROUGE-2, ROUGE-L: 7

- METEOR: 7

Δημιουργήθηκε ο παρακάτω πίνακας κριτηρίων .

ΠΙΝΑΚΑΣ 5.9-ΠΙΝΑΚΑΣ ΚΡΙΤΗΡΙΩΝ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ

Κριτήριο	LOAD DURATION	PROMPT RATE	EVALUATION RATE	ROUGE-1	ROUGE-2	ROUGE-L	METEOR
LOAD DURATION	1	4	3	2	2	2	2
PROMPT RATE	1/4	1	1/2	1/3	1/3	1/3	1/3
EVALUATION RATE	1/3	2	1	1/2	1/2	1/2	1/2
ROUGE-1	1/2	3	2	1	1	1	1
ROUGE-2	1/2	3	2	1	1	1	1
ROUGE-L	1/2	3	2	1	1	1	1
METEOR	1/2	3	2	1	1	1	1

Ακολουθούν οι πίνακες σύγκρισης για κάθε κριτήριο ξεχωριστά όπως και στην προηγούμενη υλοποίηση:

1. Πίνακας Σύγκρισης για LOAD DURATION:

ΠΙΝΑΚΑΣ 5.10-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-LOAD DURATION

Μοντέλο	Phi	Mistral	Gemma
Phi	1	1/1.025488	1/1.775178
Mistral	1.025488	1	1/1.725178
Gemma	1.775178	1.725178	1

2. Πίνακας Σύγκρισης για PROMPT RATE:

ΠΙΝΑΚΑΣ 5.11-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-PROMPT RATE

Μοντέλο	Phi	Mistral	Gemma
Phi	1	3.119.864	5.050.519
Mistral	1/3.119864	1	1.618.641
Gemma	1/5.050519	1/1.618641	1

3.Πίνακας Σύγκρισης για EVALUATION RATE:

ΠΙΝΑΚΑΣ 5.12-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-EVALUATION RATE

Μοντέλο	Phi	Mistral	Gemma
Phi	1	3.666271	8.870634
Mistral	1/3.666271	1	2.413327
Gemma	1/8.870634	1/2.413327	1

4. Πίνακας Σύγκρισης για ROUGE-1:

ΠΙΝΑΚΑΣ 5.13-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-1

Μοντέλο	Phi	Mistral	Gemma
Phi	1	1	1.428571
Mistral	1	1	1.428571
Gemma	1/1.428571	1/1.428571	1

5.Πίνακας Σύγκρισης για ROUGE-2:

ΠΙΝΑΚΑΣ 5.14-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-2

Μοντέλο	Phi	Mistral	Gemma
Phi	1	1	1.5
Mistral	1	1	1.5
Gemma	1/1.5	1/1.5	1

6.Πίνακας Σύγκρισης για ROUGE-L:

ΠΙΝΑΚΑΣ 5.15-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-ROUGE-L

Μοντέλο	Phi	Mistral	Gemma
Phi	1	1	1.428571
Mistral	1	1	1.428571
Gemma	1/1.428571	1/1.428571	1

7.Πίνακας Κριτηρίων για METEOR:

ΠΙΝΑΚΑΣ 5.16-ΠΙΝΑΚΑΣ ΣΥΓΚΡΙΣΗΣ ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ-METEOR

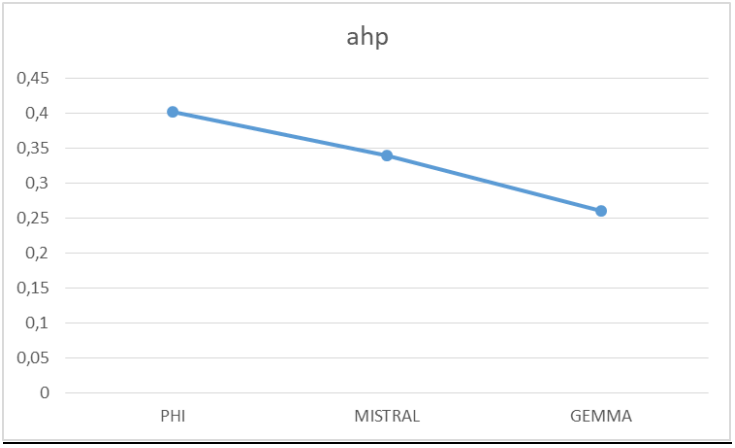
Μοντέλο	Phi	Mistral	Gemma
Phi	1	1	1.081967
Mistral	1	1	1.081967
Gemma	1/1.081967	1/1.081967	1

Οι τελικές προτεραιότητες μετά την εκτέλεση της μεθόδου AHP σε περιβάλλον Python είναι οι παρακάτω:

```
============== RESTART: C:\Users\user\Desktop\αρχεια_διπλωματικης_1_25_31_3\αρχ
Τελικές προτεραιότητες: [0.40117472 0.33836268 0.2604626 ]
```

ΕΙΚΟΝΑ 5.3-ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΥΤΕΡΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-AHP

ΔΙΑΓΡΑΜΜΑ 5.4-ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS AHP ΜΕΘΟΔΟΣ

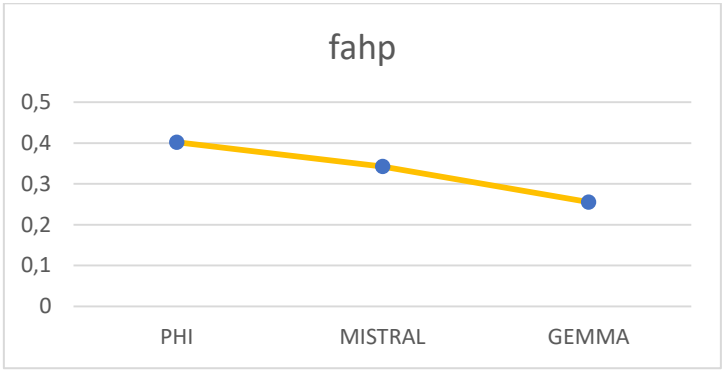


Ενώ μετά την εκτέλεση της μεθόδου FAHP προέκυψαν τα παρακάτω αποτελέσματα:

{'criteria': {'LOAD DURATION': np.float64(0.2580198115806321), 'PROMPT RATE': np.float64(0.048805946208791175), 'EVALUATION RATE': np.float64(0.07741771931669296), 'ROUGE-1': np.float64(0.15542461654826284), 'ROUGE-2': np.float64(0.15542461654826284), 'ROUGE-L': np.float64(0.15542461654826284), 'METEOR': np.float64(0.15542461654826284)}, 'alternatives': {'Phi': np.float64(0.40196613198068093), 'Mistral': np.float64(0.34270903892803256), 'Gemma': np.float64(0.2553248290912865 6)}}

ΕΙΚΟΝΑ 5.4-ΑΠΟΤΕΛΕΣΜΑΤΑ ΔΕΥΤΕΡΗΣ ΥΛΟΠΟΙΗΣΗΣ ΠΕΡΙΒΑΛΛΟΝ PYTHON-FAHP

ΔΙΑΓΡΑΜΜΑ 5.5-ΔΕΥΤΕΡΗ ΥΛΟΠΟΙΗΣΗ, ΑΠΟΔΟΣΗ LLMS FAHP ΜΕΘΟΔΟΣ



ΚΕΦΑΛΑΙΟ 6- ΣΥΜΠΕΡΑΣΜΑΤΑ-ΠΡΟΤΑΣΕΙΣ

Βάσει των αποτελεσμάτων το LLM μοντέλο PHI παραμένει σταθερά πρώτο σε απόδοση με το μοντέλο MISTRAL να το ακολουθεί σε μικρή απόσταση ελαφρά μειωμένο στις δύο υλοποιήσεις των μεθόδων ενώ είναι αξιοσημείωτη η απόδοση του μοντέλου GEMMA στην δεύτερη υλοποίηση που σχεδόν τριπλασιάστηκε η σπουδαιότητα του σχέση με την πρώτη εφαρμογή των μεθόδων.
Ανάμεσα στα μοντέλα AHP και FAHP δεν προέκυψε κάποια διαφορά αποτελεσμάτων σημειώνοντας ελάχιστες διαφορές ως προς την εμφάνιση των αποτελεσμάτων των τιμών των σπουδαιοτήτων των τριών μοντέλων.

Μελλοντικά, είναι εφικτό να χρησιμοποιηθούν περισσότερες ερωτήσεις και περισσότερα κριτήρια αξιολόγησης όπως και ο έλεγχος παραπάνω LLMs. Επιπλέον, είναι χρήσιμο να μελετηθεί η απόδοση των μοντέλων με εξωτερική γνώση όπως βάσεις δεδομένων όπου τα έγγραφα ενώνονται σας context με την αρχική είσοδο (prompt) και πηγαίνουν στην γεννήτρια κειμένου όπου παράγει την τελική έξοδο. Αυτό αυξάνει την ακρίβεια, διαχειρισιμότητα και την σχετικότητα των απαντήσεων των LLMs.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Α. ΞΕΝΟΓΛΩΣΣΗ

1. Al Faraby, S., Romadhony, A., and Adiwijaya (2024) 'Analysis of LLMs for educational question classification and generation', Computers and Education: Artificial Intelligence, 7, p. 100298. Available at: https://doi.org/10.1016/j.caeai.2024.100298.
2. Ayhan, M.B. (2013) 'A Fuzzy Ahp Approach For Supplier Selection Problem: A Case Study In A Gearmotor Company', International Journal of Managing Value and Supply Chains, 4(3), pp. 11–23. Available at: https://doi.org/10.5121/ijmvsc.2013.4302.
3. Beşikçi, E.B. et al. (2016) 'An application of fuzzy-AHP to ship operational energy efficiency measures', Ocean Engineering, 121, pp. 392–402. Available at: https://doi.org/10.1016/j.oceaneng.2016.05.031.
4. Breaz, R.E., Bologa, O. and Racz, S.G. (2017) 'Selecting industrial robots for milling applications using AHP', Procedia Computer Science, 122, pp. 346–353. Available at: https://doi.org/10.1016/j.procs.2017.11.379.
5. Chang, Yupeng et al. (2024) 'A Survey on Evaluation of Large Language Models', ACM Transactions on Intelligent Systems and Technology, 15(3), pp. 1–45. Available at: https://doi.org/10.1145/3641289.
6. Chaube, S. et al. (2024) 'An Overview of Multi-Criteria Decision Analysis and the Applications of AHP and TOPSIS Methods', International Journal of Mathematical, Engineering and Management Sciences, 9(3), pp. 581–615. Available at: https://doi.org/10.33889/IJMEMS.2024.9.3.030.
7. Chen, Y. et al. (2024) 'A survey of large language models for cyber threat detection', Computers & Security, 145, p. 104016. Available at: https://doi.org/10.1016/j.cose.2024.104016.
8. Dao, X.-Q. (2023) 'Performance Comparison of Large Language Models on VNHSGE English Dataset: OpenAI ChatGPT, Microsoft Bing Chat, and Google Bard'. arXiv. Available at: https://doi.org/10.48550/arXiv.2307.02288.

9. Digalwar, A.K., Borade, A. and Metri, B. (2014) ‘A Fuzzy AHP Approach for Supplier Selection’, Operations and Supply Chain Management: An International Journal, pp. 46–53. Available at: https://doi.org/10.31387/oscm0170107.
10. Esen H (2023) ANALYTICAL HIERARCHY PROCESS PROBLEM SOLUTION. Analytic Hierarchy Process - Models, Methods, Concepts, and Applications [Working Title]. IntechOpen. Available at: http://dx.doi.org/10.5772/intechopen.1001072.
11. Guidi, G., Goffo, G. and Violante, A.C. (2024) ‘Application of the Analytic Hierarchy Process (AHP) method to identify the most suitable approach for managing irradiated graphite’, Nuclear Engineering and Technology, 56(11), pp. 4820–4825. Available at: https://doi.org/10.1016/j.net.2024.06.045.
12. Gumilar, K.E. et al. (2024) ‘Assessment of Large Language Models (LLMs) in decision-making support for gynecologic oncology’, Computational and Structural Biotechnology Journal, 23, pp. 4019–4026. Available at: https://doi.org/10.1016/j.csbj.2024.10.050.
13. Halder, S. et al. (2024) ‘Application of machine learning and fuzzy AHP for identification of suitable groundwater potential zones using field based hydrogeophysical and soil hydraulic factors in a complex hydrogeological terrain’, Groundwater for Sustainable Development, 27, p. 101329. Available at: https://doi.org/10.1016/j.gsd.2024.101329.
14. Husein, R.A., Aburajouh, H. and Catal, C. (2025) ‘Large language models for code completion: A systematic literature review’, Computer Standards & Interfaces, 92, p. 103917. Available at: https://doi.org/10.1016/j.csi.2024.103917.
15. Khan, A.I., Al-Badi, A. and Al-Kindi, M. (2019) ‘Progressive Web Application Assessment Using AHP’, Procedia Computer Science, 155, pp. 289–294. Available at: https://doi.org/10.1016/j.procs.2019.08.041.
16. Kibria, A.S. et al. (2024) ‘Multi-criteria decision approach for climate adaptation of cultural resources along the Atlantic coast of the southeastern United States: Application of AHP method’, Climate Risk Management, 43, p. 100587. Available at: https://doi.org/10.1016/j.crm.2024.100587.
17. Kusumawardani, R.P. and Agintiara, M. (2015) ‘Application of Fuzzy AHP-TOPSIS Method for Decision Making in Human Resource Manager Selection Process’, Procedia Computer Science, 72, pp. 638–646. Available at: https://doi.org/10.1016/j.procs.2015.12.173.

18. Peng, G. et al. (2021) ‘An Application of Fuzzy Analytic Hierarchy Process in Risk Evaluation Model’, Frontiers in Psychology, 12, p. 715003. Available at: https://doi.org/10.3389/fpsyg.2021.715003.
19. Prompting Guide. (2024). Retrieval-Augmented Generation (RAG). [Διαδικτυακή σελίδα]. Διαθέσιμο από: https://www.promptingguide.ai/research/rag (Προσπελάστηκε: 08 Μαΐου 2025)
20. Saaty, T.L. (2008) ‘Decision making with the analytic hierarchy process’, International Journal of Services Sciences, 1(1), p. 83. Available at: https://doi.org/10.1504/IJSSCI.2008.017590.
21. Singh, A.K. et al. (2020) ‘A fuzzy-AHP and M − TOPSIS based approach for selection of composite materials used in structural applications’, Materials Today: Proceedings, 26, pp. 3119–3123. Available at: https://doi.org/10.1016/j.matpr.2020.02.644.
22. Singh, R.P. and Nachtnebel, H.P. (2016) ‘Analytical hierarchy process (AHP) application for reinforcement of hydropower strategy in Nepal’, Renewable and Sustainable Energy Reviews, 55, pp. 43–58. Available at: https://doi.org/10.1016/j.rser.2015.10.138.
23. Tahriri, F. (2014) ‘Supplier Assessment and Selection Using Fuzzy Analytic Hierarchy Process in a Steel Manufacturing Company’, Journal of Scientific Research and Reports, 3(10), pp. 1319–1338. Available at: https://doi.org/10.9734/JSRR/2014/8627.
24. Team Innovatics. (2024). Large Language Models (LLMs): An Overview. [Διαδικτυακό άρθρο ιστολογίου]. Διαθέσιμο από: https://teaminnovatics.com/blogs/large-language-models-llms-overview/ (Προσπελάστηκε: 08 Μαΐου 2025)
25. Tesfamariam, Solomon & Sadiq, Rehan. (2006). Risk-based environmental decision-making using fuzzy analytic hierarchy process (F-AHP). Stochastic Environmental Research and Risk Assessment. 21. 35-50. 10.1007/s00477-006-0042-9.
26. Triantaphyllou, E. and Mann, S.H. (1995) ‘USING THE ANALYTIC HIERARCHY PROCESS FOR DECISION MAKING IN ENGINEERING APPLICATIONS: SOME CHALLENGES’.
27. Vaswani, A. et al. (2023) ‘Attention Is All You Need’. arXiv. Available at: https://doi.org/10.48550/arXiv.1706.03762.
28. Veisi, H. et al. (2022) ‘Application of the analytic hierarchy process (AHP) in a multi-criteria selection of agricultural irrigation systems’, Agricultural Water Management, 267, p. 107619. Available at: https://doi.org/10.1016/j.agwat.2022.107619.

ΠΑΡΑΡΤΗΜΑ Α

ΚΩΔΙΚΑΣ PYTHON ΜΕΘΟΔΟΣ AHP

```
import numpy as np

def ahp(criteria_matrix, alternative_matrices):

    # 1. Υπολογισμός των προτεραιοτήτων των κριτηρίων
    criteria_priorities = calculate_priorities(criteria_matrix)

    # 2. Υπολογισμός των προτεραιοτήτων των εναλλακτικών για κάθε κριτήριο
    alternative_priorities = []
    for matrix in alternative_matrices:
        alternative_priorities.append(calculate_priorities(matrix))

    # 3. Υπολογισμός των συνολικών προτεραιοτήτων των εναλλακτικών
    num_alternatives = len(alternative_priorities[0])
    final_priorities = np.zeros(num_alternatives)

    for i in range(num_alternatives):
        for j in range(len(criteria_priorities)):
            final_priorities[i] += criteria_priorities[j] * alternative_priorities[j][i]

    return final_priorities

def calculate_priorities(matrix):
    """
    Υπολογίζει τις προτεραιότητες από έναν πίνακα σύγκρισης.

    Args:
        matrix: Πίνακας σύγκρισης (NxN).

    Returns:
        Ένας πίνακας με τις προτεραιότητες.
    """

    # 1. Υπολογισμός του μέσου γεωμετρικού κάθε γραμμής
    geometric_means = np.power(np.prod(matrix, axis=1), 1/matrix.shape[0])

    # 2. Κανονικοποίηση των μέσων γεωμετρικών
    priorities = geometric_means / np.sum(geometric_means)

    return priorities

# Παράδειγμα χρήσης
```

```
criteria_matrix = np.array([[1, 1/2, 1,1/3,1/4,1/4,1/5],      #LOAD_DURATION=1,PROMPT
RATE=2, EVAL_RATE=3
                    [2, 1, 2,1/2,1/3,1/3,1/4],
                    [1, 1/2, 1,1/3,1/4,1/4,1/5],
                    [3,2,3,1,1/2,1/2,1/3],
                    [4,3,4,2,1,1,1/2],
                    [4,3,4,2,1,1,1/2],
                    [5,4,5,3,2,2,1]])

alternative_matrices = [
   np.array([[1, 2, 5],
          [1/2, 1, 4],
          [1/5, 1/4, 1]]),
   np.array([[1, 4, 7],
          [1/4, 1, 3],
          [1/7, 1/3, 1]]),

   np.array([[1, 4, 7],
          [1/4, 1, 3],
          [1/7, 1/3, 1]]),
   np.array([[1, 1, 5],
          [1, 1, 5],
          [1/5, 1/5, 1]]),
   np.array([[1, 1, 5],
          [1, 1, 5],
          [1/5, 1/5, 1]]),
   np.array([[1, 1, 5],
          [1, 1, 5],
          [1/5, 1/5, 1]]),
   np.array([[1, 1, 3],                  #EVAL_RATE
          [1, 1, 3],
          [1/3, 1/3, 1]])

]
final_priorities = ahp(criteria_matrix, alternative_matrices)
print("Τελικές προτεραιότητες:", final_priorities)
```

ΚΩΔΙΚΑΣ PYTHON ΜΕΘΟΔΟΣ FAHP

```
import numpy as np
from fuzzywuzzy import fuzz

def fuzzy_ahp_fuzzywuzzy_with_alternatives(criteria_matrix, alternatives_matrices,
criteria_names=None, alternative_names=None):
   num_criteria = len(criteria_matrix)
   num_alternatives = len(alternatives_matrices[0])
```

```
    # Υπολογισμός βαρών κριτηρίων
    criteria_weights = _calculate_fuzzy_weights(criteria_matrix)

    # Υπολογισμός βαρών εναλλακτικών για κάθε κριτήριο
    alternatives_weights = []
    for matrix in alternatives_matrices:
        alternatives_weights.append(_calculate_fuzzy_weights(matrix))

    # Συνδυασμός βαρών κριτηρίων και εναλλακτικών
    final_weights = np.zeros(num_alternatives)
    for i in range(num_alternatives):
        for j in range(num_criteria):
            final_weights[i] += criteria_weights[j] * alternatives_weights[j][i]

    # Κανονικοποίηση τελικών βαρών
    final_weights /= np.sum(final_weights)

    # Δημιουργία αποτελέσματος
    result = {}
    if criteria_names:
        result["criteria"] = {criteria_names[i]: criteria_weights[i] for i in range(num_criteria)}
    else:
        result["criteria"] = {f"Criterion {i+1}": criteria_weights[i] for i in range(num_criteria)}

    if alternative_names:
        result["alternatives"] = {alternative_names[i]: final_weights[i] for i in
range(num_alternatives)}
    else:
        result["alternatives"] = {f"Alternative {i+1}": final_weights[i] for i in
range(num_alternatives)}

    return result

def _calculate_fuzzy_weights(matrix):
    n = len(matrix)
    fuzzified_matrix = np.zeros((n, n, 3))

    for i in range(n):
        for j in range(n):
            if matrix[i][j] >= 1:
                fuzzified_matrix[i, j] = [matrix[i][j] - 0.1, matrix[i][j], matrix[i][j] + 0.1]
            else:
                reciprocal = 1 / matrix[i][j]
                fuzzified_matrix[i, j] = [1 / (reciprocal + 0.1), 1 / reciprocal, 1 / (reciprocal - 0.1)]

            ratio = fuzz.ratio(str(matrix[i][j]), str(matrix[j][i])) / 100.0
            for k in range(3):
```

```
                fuzzified_matrix[i, j, k] = fuzzified_matrix[i, j, k] * (1 + (ratio - 0.5) / 5.0)

    fuzzy_geometric_mean = np.ones((n, 3))
    for i in range(n):
        for j in range(3):
            for k in range(n):
                fuzzy_geometric_mean[i, j] *= fuzzified_matrix[i, k, j]
            fuzzy_geometric_mean[i, j] = fuzzy_geometric_mean[i, j] ** (1 / n)

    sum_fuzzy_gm = np.zeros(3)
    for j in range(3):
        for i in range(n):
            sum_fuzzy_gm[j] += fuzzy_geometric_mean[i, j]

    inv_sum_fuzzy_gm = 1 / sum_fuzzy_gm[::-1]

    fuzzy_weights = np.zeros((n, 3))
    for i in range(n):
        for j in range(3):
            fuzzy_weights[i, j] = fuzzy_geometric_mean[i, j] * inv_sum_fuzzy_gm[j]

    return np.mean(fuzzy_weights, axis=1)

criteria_matrix = [
        [1, 4, 3,2,2,2,2],
        [1/4, 1, 1/2,1/3,1/3,1/3,1/3],
                        [1/3, 2, 1,1/2,1/2,1/2,1/2],
                        [1/2,3,2,1,1,1,1],
                        [1/2,3,2,1,1,1,1],
                        [1/2,3,2,1,1,1,1],
                        [1/2,3,2,1,1,1,1]

]

alternatives_matrices = [
    [[1, 1/1.025488, 1/1.775178],
            [1.025488, 1, 1/1.725178],
            [1.775178, 1/4, 1]],  # Για το κριτήριο 1
    [[1, 3.119864, 5.050519],
            [1/3.119864, 1, 1.618641],
            [1/5.050519, 1/1.618641, 1]],  # Για το κριτήριο 2
    [[1, 3.666271, 8.870634],
            [1/3.666271, 1, 2.413327],
            [1/8.870634, 1/2.413327, 1]],
    [[1, 1, 1.428571],
            [1, 1, 1.428571],
            [1/1.428571, 1/1.428571, 1]],
```

```
    [[1, 1, 1.5],
        [1, 1, 1.5],
        [1/1.5, 1/1.5, 1]],
    [[1, 1, 1.428571],
        [1, 1, 1.428571],
        [1/1.428571, 1/1.428571, 1]],
    [[1, 1, 1.081967],
        [1, 1, 1.081967],
        [1/1.081967, 1/1.081967, 1]]



# Για το κριτήριο 3
]

criteria_names = ["LOAD DURATION", "PROMPT RATE", "EVALUATION RATE","ROUGE-1","ROUGE-2","ROUGE-L","METEOR"]
alternative_names = ["Phi", "Mistral", "Gemma"]

weights = fuzzy_ahp_fuzzywuzzy_with_alternatives(criteria_matrix, alternatives_matrices, criteria_names, alternative_names)
print(weights)
```

ΠΑΡΑΡΤΗΜΑ Β

ΚΩΔΙΚΑΣ PYTHON ΚΡΙΤΗΡΙΟ - ROUGE

```
from rouge_score import rouge_scorer

reference = "The print() function"
candidate = "To print output in Python, you can use the `print()` function"

scorer = rouge_scorer.RougeScorer(['rouge1', 'rouge2', 'rougeL'], use_stemmer=True)
scores = scorer.score(reference, candidate)

print("ROUGE-1:", scores['rouge1'])
print("ROUGE-2:", scores['rouge2'])
print("ROUGE-L:", scores['rougeL'])
```

ΚΩΔΙΚΑΣ PYTHON ΚΡΙΤΗΡΙΟ – METEOR

```
from nltk.translate.meteor_score import meteor_score
from nltk.tokenize import word_tokenize

reference = "The print() function"
candidate = "To print output in Python, you can use the `print()` function"

reference_tokens = word_tokenize(reference)
candidate_tokens = word_tokenize(candidate)

score = meteor_score([reference_tokens], candidate_tokens)
print(f"METEOR Score: {score}")
```

ΚΩΔΙΚΑΣ PYTHON ΚΡΙΤΗΡΙΑ- LOAD DURATION, PROMPT_RATE, EVALUATION_RATE

```
import ollama
import pandas as pd
import matplotlib.pyplot as plt

# question
q = '"What is the command to print output in Python?"'

# context
cntx = '"The print() function"'

# models
```

```
models = ['phi','mistral','gemma'] #,'mixtral']

# store results
r = []
idx = []
# prompt without context
pmpt = f'Question: {q}'
for i,mx in enumerate(models):
    r.append(ollama.generate(model=mx, prompt=pmpt))
    idx.append(mx)
    print(f"Model: {idx[-1]}, Time: {r[-1]['total_duration']/1e9}s")

# prompt with context
pmpt = f'Context: {cntx} Question: {q}'
for i,mx in enumerate(models):
    r.append(ollama.generate(model=mx, prompt=pmpt))
    idx.append(mx+'+ctx')
    print(f"Model: {idx[-1]}, Time: {r[-1]['total_duration']/1e9}s")

# put results in DataFrame
rcols = ['total_duration','load_duration',
         'prompt_eval_count','eval_count',
         'prompt_eval_duration','eval_duration']
data = {}
for i,x in enumerate(rcols):
    if (i==2) or (i==3):
        data[x] = [ri[x] for ri in r]
    else:
        data[x] = [ri[x]/1e9 for ri in r]
data = pd.DataFrame(data, index=idx)
data['calc_duration'] = data['total_duration']-data['load_duration']
data['prompt_rate'] = data['prompt_eval_count']/data['prompt_eval_duration']
data['eval_rate'] = data['eval_count']/data['eval_duration']
print(data)

pcols = ['load_duration','prompt_rate','eval_rate']
axs = data[pcols].plot(figsize=(6,5),kind='bar',subplots=True)
ylb = ['Time (s)','Rate (tk/s)','Rate (tk/s)']
for i,ax in enumerate(axs):
    ax.set_title('')
    ax.set_ylabel(ylb[i])
plt.tight_layout(); plt.savefig('results.png',dpi=300); plt.show()
```