



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

Ανάπτυξη ψηφιακού βοηθού ενεργειακής μετάβασης

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Γ. Κουνάς

Επιβλέπων : Αναστάσιος Δ. Δουλάμης
Καθηγητής Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2025



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ – ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»

Ανάπτυξη ψηφιακού βοηθού ενεργειακής μετάβασης

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Δημήτριος Γ. Κουνάς

Επιβλέπων : Αναστάσιος Δ. Δουλάμης

Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 19^η Φεβρουαρίου 2025.

.....
Αναστάσιος Δουλάμης
Καθηγητής Ε.Μ.Π.

.....
Νικόλαος Δουλάμης
Καθηγητής Ε.Μ.Π.

.....
Θεοδώρα Βαρβαρίγου
Καθηγήτρια Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2025

.....
Δημήτριος Γ. Κουνάς

Διπλωματούχος / Κάτοχος Μεταπτυχιακού Προγράμματος «Δ.Π.Μ.Σ Τεχνο-Οικονομικά
Συστήματα» της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών Ε.Μ.Π

Copyright © Δημήτριος Γ. Κουνάς 2025
Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Περίληψη

Τα τελευταία χρόνια ο πλανήτης βρίσκεται αντιμέτωπος με την κλιματική αλλαγή και την ενεργειακή κρίση. Η κλιματική αλλαγή, από τη μία, οφείλεται στα αέρια του θερμοκηπίου που παράγονται από την καύση ορυκτών καυσίμων και εμφανίζεται με πληθώρα καταστροφών και ακραίων καιρικών συνθηκών. Η μέση θερμοκρασία της Γης ολοένα και αυξάνεται, τα καλοκαίρια γίνονται θερμότερα, αυξάνονται οι δασικές πυρκαγιές και οι πλημμύρες, ενώ άλλοτε παρουσιάζεται ξηρασία και λειψυδρία. Από την άλλη, η συνεχής πληθυσμιακή αύξηση, οι γεωπολιτικές εντάσεις, οι πόλεμοι και άλλες κρίσεις (π.χ. πανδημίες) δημιουργούν το φαινόμενο της ενεργειακής κρίσης, η οποία εμφανίζεται με αυξημένες τιμές ενέργειας ή ακόμα και αδυναμία πρόσβασης στην ενέργεια για κάποιο μερίδιο του πληθυσμού.

Για την αντιμετώπιση των παραπάνω δύο φαινομένων γίνεται λόγος για ενεργειακή μετάβαση από την καύση ορυκτών καυσίμων στη χρήση ανανεώσιμων πηγών ενέργειας (ΑΠΕ), με ταυτόχρονη μείωση της κατανάλωσης ενέργειας. Η μετάβαση αυτή από τη μία θα συντελέσει στη μείωση των εκπομπών αερίων του θερμοκηπίου, περιορίζοντας την επίδραση της κλιματικής αλλαγής, ενώ από την άλλη θα ενισχύσει την ενεργειακή ανεξαρτησία των χωρών, οι οποίες δεν θα εξαρτώνται από επιρρεπείς σε κρίσεις εισαγωγές καυσίμων. Ταυτόχρονα, θα αυξήσει την ενεργειακή ασφάλεια, την πρόσβαση δηλαδή των καταναλωτών σε ενέργεια σε προσιτές τιμές. Στην κατεύθυνση της ενεργειακής μετάβασης, οι κυβερνήσεις ανά τον κόσμο σχεδιάζουν προγράμματα με σκοπό να δώσουν κίνητρα στους καταναλωτές, ώστε να αρχίσουν να χρησιμοποιούν νέες, αποδοτικότερες τεχνολογίες, να αποσύρουν τις παλιές τους συσκευές και να αναβαθμίσουν ενεργειακά τα κτήριά τους.

Παράλληλα, τα τελευταία χρόνια ο πλανήτης παρακολουθεί ένα ακόμα φαινόμενο που δεν είναι άλλο από την εξάπλωση των εφαρμογών τεχνητής νοημοσύνης και ιδιαίτερα των εφαρμογών ερωταπαντήσεων (AI chatbots). Οι εφαρμογές αυτές έχουν διεισδύσει στην καθημερινότητά μας και η απόδοσή τους στην κατανόηση και παραγωγή κειμένου τις καθιστά ιδιαίτερα χρήσιμες, με αποτέλεσμα να επιλέγονται συχνά ως αντικαταστάτες των παραδοσιακών ψηφιακών βοηθών.

Στην εργασία αυτή αναπτύσσεται ένας ψηφιακός βοηθός ενεργειακής μετάβασης βασισμένος σε τεχνητή νοημοσύνη. Συλλέγονται δεδομένα προγραμμάτων ενεργειακής μετάβασης και αφού δομηθούν κατάλληλα, τροφοδοτούνται σε ένα μεγάλο γλωσσικό μοντέλο (LLM), δημιουργώντας έτσι ένα σύστημα που είναι σε θέση να δίνει πληροφορίες και να επιλύει πιθανές απορίες του χρήστη σχετικά με αυτά. Σκοπός ήταν να επιδειχθεί το πώς η σύγχρονη τεχνολογία τεχνητής νοημοσύνης αποτελεί ισχυρό εργαλείο για την ενίσχυση της συμμετοχής των καταναλωτών σε προγράμματα ενεργειακής μετάβασης, με απώτερο σκοπό την καταπολέμηση της κλιματικής αλλαγής και της ενεργειακής κρίσης.

Λέξεις Κλειδιά: κλιματική αλλαγή, ενεργειακή μετάβαση, τεχνητή νοημοσύνη, μεγάλα γλωσσικά μοντέλα (LLMs), ψηφιακός βοηθός

Abstract

In recent years, the planet has been facing climate change and energy crisis. Climate change, on the one hand, is caused by greenhouse gases produced by the burning of fossil fuels and is manifested in a multitude of disasters and extreme weather conditions. The average temperature of the Earth is constantly increasing, summers are getting hotter, forest fires and floods are increasing, while droughts and water shortages are sometimes occurring. On the other hand, continuous population growth, geopolitical tensions, wars and other crises (e.g. pandemics) create the phenomenon of the energy crisis, which is manifested by increased energy prices or even the inability to access energy for a certain share of the population.

To address the above two phenomena, there is talk of an energy transition from the combustion of fossil fuels to the use of renewable energy, while simultaneously reducing energy consumption. This transition will, on the one hand, contribute to the reduction of greenhouse gas emissions, limiting the impact of climate change, while on the other hand, it will strengthen the energy independence of countries, which will not depend on crisis-prone fuel imports. At the same time, it will increase energy security, that is, consumers' access to energy at affordable prices. In the direction of the energy transition, governments around the world are designing programs to provide incentives to consumers to start using new, more efficient technologies, to retire their old appliances and to upgrade their buildings in terms of energy efficiency.

At the same time, in recent years the planet has been witnessing another phenomenon that is none other than the spread of artificial intelligence applications and especially question-and-answer applications (AI chatbots). These applications have penetrated our daily lives and their performance in understanding and producing text makes them particularly useful, as a result of which, they are often chosen as replacements for traditional digital assistants.

In this work, a digital energy transition assistant based on artificial intelligence is developed. Data from energy transition programs are collected and, after being appropriately structured, are fed into a large language model (LLM), thus creating a system that is able to provide information and resolve possible user questions about them. The aim was to demonstrate how modern artificial intelligence technology is a powerful tool for enhancing consumer participation in energy transition programs, with the ultimate goal of combating climate change and energy crisis.

Keywords: climate change, energy transition, artificial intelligence, large language models (LLMs), digital assistant

Ευχαριστώ τον επιβλέποντα καθηγητή κύριο Δουλάμη Αναστάσιο για την ευκαιρία που μου έδωσε να ασχοληθώ με τα επίκαιρα και καίρια ζητήματα της κλιματικής αλλαγής, της ενεργειακής κρίσης και της τεχνητής νοημοσύνης.

Επιπλέον, ευχαριστώ όσους ήταν δίπλα μου κατά το χρόνο φοίτησής μου και εκπόνησης της εργασίας αυτής.

Πίνακας Περιεχομένων

1	Εισαγωγή.....	13
1.1	Επιστημονικό αντικείμενο	13
1.2	Σκοπός της εργασίας.....	14
1.3	Δομή περιεχομένου	15
2	Κλιματική αλλαγή.....	16
2.1	Περιγραφή του φαινομένου	16
2.2	Αίτια.....	17
2.3	Συνέπειες.....	17
2.3.1	Συνέπειες στη φύση	17
2.3.2	Συνέπειες στην ανθρώπινη ζωή και δραστηριότητα	18
2.4	Αντιμετώπιση	19
2.5	Προβλέψεις για το μέλλον	20
3	Ενεργειακή κρίση.....	21
3.1	Αίτια.....	21
3.2	Συνέπειες.....	22
3.3	Αντιμετώπιση	23
3.4	Η ενεργειακή κρίση ως ευκαιρία	23
4	Ενεργειακή μετάβαση	24
4.1	Η ανάγκη για ενεργειακή μετάβαση	24
4.2	Πλεονεκτήματα και μειονεκτήματα της ενεργειακής μετάβασης.....	25
4.3	Προσπερνώντας εμπόδια	26
5	Δράσεις και κίνητρα ενεργειακής μετάβασης.....	27
5.1	Δράσεις ενεργειακής μετάβασης στην Ελλάδα	27
5.2	Χαρακτηριστικά των δράσεων ενεργειακής μετάβασης.....	28
5.3	Κατηγορίες δράσεων ενεργειακής μετάβασης.....	29

5.4	Ανάγκη για αξιολόγηση των δράσεων ενεργειακής μετάβασης	29
6	Τεχνητή Νοημοσύνη και LLMs	30
6.1	Βασικοί ορισμοί	30
6.1.1	Τεχνητή Νοημοσύνη	30
6.1.2	Large Language Models – LLMs.....	31
6.1.3	Εκπαίδευση των LLMs	31
6.1.4	LLMs και κατανάλωση ενέργειας.....	32
6.1.5	Αξιολόγηση των LLM	33
6.2	Παραισθήσεις (hallucinations).....	35
6.3	Αιτίες παραισθήσεων	36
6.4	Αξιολόγηση παραισθήσεων	38
6.5	Αντιμετώπιση των παραισθήσεων	40
6.5.1	Τεχνικές αντιμετώπισης παραισθήσεων	40
6.5.2	Retrieval-Augmented Generation (RAG)	41
6.5.3	RAG και μηχανική προτροπών	45
6.6	Επισκόπηση των LLMs.....	48
7	LLMs και προγράμματα ενεργειακής μετάβασης	49
7.1	Υπάρχουσες λύσεις	49
7.2	Ένας ψηφιακός βοηθός για προγράμματα ενεργειακής μετάβασης	51
8	Υλοποίηση του βοηθού ενεργειακής μετάβασης	52
8.1	Χρήσιμα εργαλεία – Προγραμματιστικό υπόβαθρο	52
8.1.1	LangChain – Ένα πλαίσιο ανάπτυξης LLM εφαρμογών.....	52
8.1.2	HuggingFace – Μία βιβλιοθήκη μοντέλων τεχνητής νοημοσύνης	52
8.2	Συλλογή των δεδομένων	53
8.3	Μετατροπή των δεδομένων σε CSV	57
8.4	Φόρτωση και κατακερματισμός των δεδομένων	57
8.5	Ενσωματώσεις.....	58

8.6	Εισαγωγή των ενσωματώσεων στη βάση δεδομένων.....	61
8.7	Meltemi - Το πρώτο ανοιχτό Μεγάλο Γλωσσικό Μοντέλο για τα Ελληνικά ..	62
8.7.1	Χρήση του μοντέλου.....	63
8.7.2	Διαμορφώνοντας το επαυξημένο ερώτημα.....	65
8.8	Κατασκευάζοντας την αλυσίδα RAG	66
8.9	Δίνοντας στην εφαρμογή μας ένα γραφικό περιβάλλον	68
9	Αποτελέσματα δοκιμών	70
9.1	Παρουσίαση και σχολιασμός αποτελεσμάτων.....	70
9.2	Προκλήσεις που αντιμετωπίστηκαν.....	81
10	Συμπεράσματα και προτάσεις	83
10.1	Αντιμετώπιση προβλημάτων.....	84
10.2	Βελτιώσεις και προεκτάσεις	84
11	Παράρτημα: Κώδικας της υλοποίησης.....	86
11.1	Μετατροπή του Excel σε CSV - excel2csv.py	86
11.2	Κώδικας της εφαρμογής RAG – main.py	86
11.3	Κώδικας της ενσωμάτωσης του Streamlit	89

1 Εισαγωγή

1.1 Επιστημονικό αντικείμενο

Η δραστηριότητα του ανθρώπου πάνω στη Γη είναι συνυφασμένη με την κατανάλωση ενέργειας. Από τα προϊστορικά χρόνια όπου ο άνθρωπος άναψε φωτιά προκειμένου να ζεσταθεί μέχρι και τη σύγχρονη θέρμανση με αντλίες θερμότητας, η ύπαρξη διαθέσιμης ενέργειας διαδραματίζει σημαντικό ρόλο τόσο για την επιβίωση όσο και για την ευζωία του.

Η κατανάλωση ενέργειας όμως, ιδιαίτερα τα τελευταία χρόνια, έχει επιβαρύνει σημαντικά τον πλανήτη και τα αποτελέσματα είναι πλέον έκδηλα. Η θερμοκρασία της Γης αυξάνεται, καταστρέφοντας οικοσυστήματα, οδηγώντας σε ακόμα πιο έντονες φυσικές καταστροφές όπως πυρκαγιές και πλημμύρες. Ο καλοκαιρινός καύσωνας οδηγεί σε αυξανόμενους θανάτους, ενώ η λειψυδρία είναι πλέον πραγματικότητα. Ως αιτία για το φαινόμενο αυτό θεωρείται η καύση ορυκτών καυσίμων και τα αέρια του θερμοκηπίου που αυτή παράγει, εμποδίζοντας τη θερμότητα να φύγει από τη Γη.

Από την άλλη, αν και υπήρχαν και στο παρελθόν ενεργειακές κρίσεις, τα τελευταία χρόνια βιώσαμε εκείνη με τον πλέον παγκόσμιο χαρακτήρα μέχρι στιγμής. Οι γεωπολιτικές ταραχές, η πανδημία COVID-19 και ιδιαίτερα ο πόλεμος Ρωσίας-Ουκρανίας δημιούργησαν τις κατάλληλες προϋποθέσεις ώστε να προκληθεί αύξηση των τιμών της ενέργειας. Η παράλληλη αύξηση του πληθυσμού εντείνει ακόμα περισσότερο το φαινόμενο. Το αποτέλεσμα ήταν τα έξοδα ενέργειας να γίνουν για πολλούς δυσβάσταχτα, ενώ σε αναπτυσσόμενες χώρες πολλοί δεν μπορούσαν να έχουν καν πρόσβαση σε ενέργεια.

Η αντιμετώπιση της κλιματικής αλλαγής και της ενεργειακής κρίσης απαιτούν δράση, που δεν είναι άλλη από την ενεργειακή μετάβαση, τη στροφή δηλαδή από τα ορυκτά καύσιμα προς τις ανανεώσιμες πηγές ενέργειας. Η εκτεταμένη χρήση ΑΠΕ μπορεί από τη μία να μειώσει σημαντικά τις εκπομπές αερίων του θερμοκηπίου οδηγώντας σε αναχαίτιση της κλιματικής αλλαγής και από την άλλη να δώσει ενεργειακή ανεξαρτησία στα κράτη ώστε να μην χρειάζεται να πραγματοποιούν εισαγωγές, δημιουργώντας έτσι ασπίδα απέναντι σε ενεργειακές κρίσεις. Προς την ίδια κατεύθυνση οδηγεί και η εξοικονόμηση ενέργειας αλλά και ο ενεργειακά αποδοτικός σχεδιασμός συσκευών και κτηρίων. Για τους παραπάνω λόγους, τα κράτη συχνά σχεδιάζουν και υλοποιούν προγράμματα ενεργειακής αναβάθμισης και εξοικονόμησης ενέργειας, δίνοντας κίνητρα (π.χ. επιδοτήσεις) στους καταναλωτές.

Ενώ συμβαίνουν τα παραπάνω, ο τομέας της τεχνητής νοημοσύνης καλπάζει με ταχύτατους ρυθμούς, με τα γλωσσικά μοντέλα να πρωτοστατούν. Αυτά παρουσιάζουν πρωτοφανείς δυνατότητες κατανόησης και παραγωγής κειμένου και για το λόγο αυτό χρησιμοποιούνται σε πληθώρα ακαδημαϊκών και βιομηχανικών εφαρμογών. Για να εκπαιδευτούν όμως απαιτούν μεγάλα ποσά ενέργειας, διογκώνοντας περαιτέρω τα παραπάνω προβλήματα, αν η ενέργεια που καταναλώνουν δεν προέρχεται από ανανεώσιμες πηγές.

Τα μοντέλα αυτά μπορούν ωστόσο να γίνουν σύμμαχοι στην προσπάθεια ενεργειακής μετάβασης, βοηθώντας τους καταναλωτές να ανακαλύψουν προγράμματα ενεργειακής μετάβασης στα οποία μπορούν να ενταχθούν και να ερμηνεύσουν τις πολλές φορές δυσνόητες τεχνικές λεπτομέρειες από τις οποίες αυτά αποτελούνται. Στην εργασία αυτή, σχεδιάζεται και υλοποιείται ένα σύστημα τεχνητής νοημοσύνης το οποίο τροφοδοτείται με πληροφορίες προγραμμάτων ενεργειακής μετάβασης και μπορεί να απαντά σε ερωτήσεις σχετικά με αυτά, αποδεικνύοντας πως η σύγχρονη τεχνολογία τεχνητής νοημοσύνης μπορεί να αποτελέσει ισχυρό εργαλείο για τη βιωσιμότητα του μέλλοντος.

1.2 Σκοπός της εργασίας

Η εργασία αυτή έχει ως κύριο σκοπό να παρουσιάσει πώς με τη χρήση των σύγχρονων εργαλείων τεχνητής νοημοσύνης είναι δυνατό να υλοποιηθεί ένας ψηφιακός βοηθός ο οποίος θα ενημερώνει, θα επιλύει απορίες και θα καθοδηγεί τον χρήστη στην προσπάθειά του να ανακαλύψει, να κατανοήσει και να ενταχθεί σε προγράμματα ενεργειακής μετάβασης.

Επιπλέον, έχει ως σκοπό να επισημάνει τη σπουδαιότητα των φαινομένων της κλιματικής αλλαγής και της ενεργειακής κρίσης, τα οποία επηρεάζουν τόσο το περιβάλλον όσο και τον άνθρωπο. Παράλληλα, αποσκοπεί στο να παρουσιάσει την αναγκαιότητα για ενεργειακή μετάβαση και για παροχή κινήτρων στους καταναλωτές. Μέσα από τη συλλογή πραγματικών δεδομένων από προγράμματα ενεργειακής μετάβασης στην Ελλάδα, δημιουργείται μια μεθοδολογία επεξεργασίας των δεδομένων και τροφοδότησής τους σε μοντέλο τεχνητής νοημοσύνης. Το αποτέλεσμα είναι η υλοποίηση του ψηφιακού βοηθού και η παράθεση διαπιστώσεων και συμπερασμάτων σχετικά με τη λειτουργία και την αποτελεσματικότητά του.

1.3 Δομή περιεχομένου

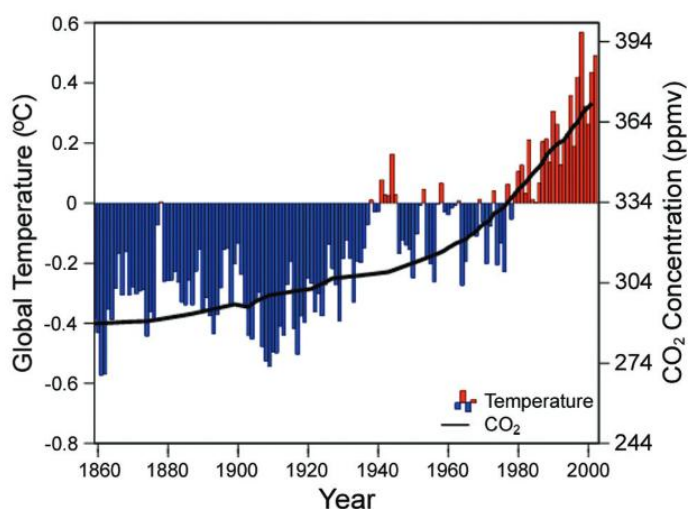
Η εργασία αποτελείται από 11 κεφάλαια. Στο παρόν και πρώτο κεφάλαιο γίνεται εισαγωγή στο αντικείμενο και τον σκοπό της εργασίας. Στο δεύτερο κεφάλαιο γίνεται αναφορά στο φαινόμενο της κλιματικής αλλαγής και τη σημασία της. Στο τρίτο κεφάλαιο παρουσιάζεται η έννοια της ενεργειακής κρίσης. Στο τέταρτο κεφάλαιο διαπιστώνεται πώς τα δύο παραπάνω φαινόμενα καθιστούν την ενεργειακή μετάβαση απαραίτητη για ένα βιώσιμο μέλλον. Στο πέμπτο κεφάλαιο αναλύονται οι δράσεις ενεργειακής μετάβασης που σχεδιάζονται από τα κράτη προκειμένου να δοθούν κίνητρα στους καταναλωτές να προβούν σε ενέργειες ενεργειακής αναβάθμισης. Στο έκτο κεφάλαιο δίνονται βασικοί ορισμοί σχετικά με την τεχνητή νοημοσύνη και παρουσιάζονται τα μεγάλα γλωσσικά μοντέλα (LLMs). Δίνεται έμφαση στο φαινόμενο των παραισθήσεων και παρουσιάζεται η τεχνική Retrieval-Augmented Generation (RAG) η οποία χρησιμοποιείται από τον ψηφιακό βοηθό που υλοποιείται στα πλαίσια της εργασίας. Στο έβδομο κεφάλαιο γίνεται η σύνδεση τεχνητής νοημοσύνης και ενεργειακής μετάβασης, όπου αναφέρονται αντίστοιχα υπάρχοντα εγχειρήματα και παρουσιάζεται η πρόταση για τον ψηφιακό βοηθό που υλοποιήθηκε. Στο όγδοο κεφάλαιο περιγράφεται αναλυτικά η διαδικασία υλοποίησης του ψηφιακού βοηθού. Στο ένατο κεφάλαιο παρουσιάζονται τα αποτελέσματα των δοκιμών που πραγματοποιήθηκαν. Στο δέκατο κεφάλαιο γίνονται διαπιστώσεις και προτείνονται κατευθύνσεις για μελλοντική εργασία. Η εργασία ολοκληρώνεται με το ενδέκατο κεφάλαιο, το οποίο αποτελεί το παράρτημα στο οποίο παρουσιάζεται ο πλήρης κώδικας της εφαρμογής που υλοποιήθηκε.

2 Κλιματική αλλαγή

2.1 Περιγραφή του φαινομένου

Η κλιματική αλλαγή, ένας όρος που ακούγεται όλο και περισσότερο τα τελευταία χρόνια, αφορά στη θέρμανση του πλανήτη και τις γενικότερες επιπτώσεις αυτής στο κλίμα της Γης. Τα Ηνωμένα Έθνη, στη «Σύμβαση-Πλαίσιο των Ηνωμένων Εθνών για τις Κλιματικές Μεταβολές» (United Nations Framework Convention on Climate Change) [1], που συνάφθηκε το 1992, ορίζουν ως κλιματική αλλαγή την **μεταβολή στο κλίμα που οφείλεται άμεσα ή έμμεσα σε ανθρώπινες δραστηριότητες**. Με τον ορισμό αυτό ξεχωρίζει από την κλιματική μεταβλητότητα, η οποία αναφέρεται στις γενικότερες μεταβολές του κλίματος της Γης ανά τα χρόνια.

Ο κύριος τρόπος με τον οποίο η ανθρώπινη δραστηριότητα μεταβάλλει το κλίμα είναι παρεμβαίνοντας στις φυσικές ροές της ενέργειας μέσω αλλαγών στην ατμοσφαιρική σύνθεση και όχι από την θερμότητα που παράγεται κατά τη χρήση της ενέργειας. Οι αλλαγές αυτές στη σύνθεση της ατμόσφαιρας συμβαίνουν λόγω εκπομπών αερίων του θερμοκηπίου, όπως το διοξείδιο του άνθρακα, το οποίο είναι αποτέλεσμα της καύσης ορυκτών καυσίμων, μεθανίου κ.α. Τα αέρια αυτά, βρισκόμενα στην ατμόσφαιρα εμποδίζουν την έξοδο της ακτινοβολίας και συνεπώς της θερμότητας από τη Γη, οδηγώντας στη σταδιακή της θέρμανση [2].



Εικόνα 1: Αύξηση της θερμοκρασίας της επιφάνειας της Γης σε σχέση με την συγκέντρωση CO₂ στην ατμόσφαιρα Πηγή: [2]

2.2 Αίτια

Όπως αναφέρθηκε, ο κύριος μηχανισμός της κλιματικής αλλαγής είναι το **φαινόμενο του θερμοκηπίου**, όπου η παγίδευση αερίων στην ατμόσφαιρα αποτρέπει την διάχυση της θερμότητας της Γης προς το διάστημα, οδηγώντας σε αύξηση τη θερμοκρασίας της.

Οι κύριες ανθρώπινες δραστηριότητες που συμβάλλουν στη συγκέντρωση αερίων του θερμοκηπίου στην ατμόσφαιρα είναι οι ακόλουθες [3]:

- Η **καύση ορυκτών καυσίμων** (π.χ. βενζίνη, πετρέλαιο, φυσικό αέριο) παράγει διοξείδιο του άνθρακα και υποξείδιο του αζώτου.
- Η **αποψίλωση των δασών** στερεί τη θετική συμβολή των δέντρων, τα οποία απορροφούν το διοξείδιο του άνθρακα από την ατμόσφαιρα.
- Η **αύξηση της κτηνοτροφίας** οδηγεί σε αυξημένη διοχέτευση μεθανίου, το οποίο παράγεται από τα ζώα κατά την πέψη της τροφής τους.
- Η **χρήση αζωτούχων λιπασμάτων** αυξάνει τις εκπομπές υποξειδίου του αζώτου.
- Τα **φθοριούχα αέρια** που χρησιμοποιεί ο άνθρωπος προκαλούν έως και 23.000 φορές μεγαλύτερη αύξηση της θερμοκρασίας σε σχέση με το διοξείδιο του άνθρακα.

2.3 Συνέπειες

Η κλιματική αλλαγή επηρεάζει τόσο τη φύση όσο και τον άνθρωπο και τη δραστηριότητά του [4].

2.3.1 Συνέπειες στη φύση

Η κλιματική αλλαγή έχει ήδη επηρεάσει και θα συνεχίσει να επηρεάζει τα ζώα, τα φυτά και το περιβάλλον γενικότερα. Οι αλλαγές στο κλίμα μεταβάλλουν την κατανομή και την ποικιλία φυτικών και ζωικών ειδών, τα οποία δυσκολεύονται να επιβιώσουν λόγω της απώλειας οικοτόπων και της ρύπανσης. Παράλληλα, επηρεάζεται η συμπεριφορά και ο κύκλος ζωής τους, οδηγώντας σε αύξηση του αριθμού των επιβλαβών οργανισμών.

Οι περίοδοι ξηρασίας έχουν αυξηθεί ήδη τόσο σε συχνότητα, όσο και σε ένταση αλλά και διάρκεια. Το γεγονός αυτό οφείλεται αφενός στην αύξηση της εξάτμισης των υδάτων λόγω των υψηλών θερμοκρασιών και αφετέρου στην έλλειψη βροχοπτώσεων. Η ξηρασία επηρεάζει φυσικά και τη διαθεσιμότητα του γλυκού νερού αλλά και την ποιότητά του. Τα αποτελέσματα στη χώρα μας έχουν αρχίσει ήδη να γίνονται ορατά [5].

Μαζί με την ξηρασία, αυξάνεται η διάρκεια και η ένταση των δασικών πυρκαγιών, ιδιαίτερα στην περιοχή της Μεσογείου. Μία έρευνα αποκάλυψε πως η συχνότητα των ακραίων γεγονότων δασικών πυρκαγιών αυξήθηκε κατά 2,2 φορές από το 2003 ως το 2023 [6].

Από την άλλη, η κλιματική αλλαγή αναμένεται να αυξήσει τα φαινόμενα έντονων πλημμύρων. Οι έντονες καταιγίδες αναμένεται να γίνουν συχνότερες και ισχυρότερες και σε συνδυασμό με την αποψίλωση των δασών λόγω των αυξημένων πυρκαγιών, το φαινόμενο θα ενισχυθεί περαιτέρω καθώς θα μειωθεί σημαντικό τμήμα του όγκου συγκράτησης των ορμητικών νερών στα βουνά.

Επιπλέον, η στάθμη της θάλασσας έχει ήδη αρχίσει να ανεβαίνει από τον 20^ο αιώνα και επιταχύνεται τις τελευταίες δεκαετίες. Το φαινόμενο αυτό δημιουργείται καθώς η αύξηση της θερμοκρασίας προκαλεί την θερμική έκταση των ωκεανών αλλά και το λιώσιμο των πάγων. Η άνοδος της στάθμης των υδάτων θα θέσει σε κίνδυνο τους ανθρώπους και το οικοσύστημα των παράκτιων περιοχών.

2.3.2 Συνέπειες στην ανθρώπινη ζωή και δραστηριότητα

Η ζωή του ανθρώπου επηρεάζεται ποικιλοτρόπως από την κλιματική αλλαγή. Οι αυξημένες και εντονότερες περίοδοι καύσωνα αναμένεται να αυξήσουν τους θανάτους και τις ασθένειες που προκαλούνται λόγω υψηλών θερμοκρασιών. Επιπλέον, τα ακραία καιρικά φαινόμενα όπως πλημμύρες, πυρκαγιές και καταιγίδες θα θέσουν σε κίνδυνο τη ζωή πολλών συνανθρώπων μας. Παράλληλα, οι αλλαγές στο οικοσύστημα θα επιφέρουν αλλαγές στα νοσήματα που μεταφέρονται μέσω ζώων, υδάτων και τροφών. Αναμένεται ακόμα να επανεμφανιστούν ή να αναδυθούν οργανισμοί επιβλαβείς για τα φυτά και κατά συνέπεια για τις καλλιέργειες. Η υγεία των ανθρώπων θα επηρεαστεί επιπλέον από την κακή ποιότητα του αέρα.

Όσον αφορά την ανθρώπινη δραστηριότητα, θα επηρεαστεί η παραγωγικότητα και η βιωσιμότητα όλων των οικονομικών τομέων. Οι συνθήκες εργασίας ενδέχεται να υποβαθμιστούν, ενώ ορισμένοι οικονομικοί τομείς θα επηρεαστούν περισσότερο λόγω της άμεσης σχέσης τους με τις κλιματικές συνθήκες (π.χ. γεωργία, τουρισμός). Από την άλλη, οι επενδύσεις στην προσαρμογή στη νέα πραγματικότητα θα μπορούσαν να προσφέρουν ευκαιρίες απασχόλησης σε περιοχές που έχουν ανάγκη. Επιπλέον, οι συχνότερες και εντονότερες φυσικές καταστροφές θα προκαλέσουν σημαντικές φθορές σε κτίρια και υποδομές. Σχετικά με την ενέργεια, η ζήτηση για θέρμανση θα μειωθεί στις ψυχρότερες περιοχές, αλλά στις θερμότερες αναμένεται σημαντική αύξηση της ζήτησης για ψύξη κατά

τους θερμούς μήνες. Οι αλλαγές στο κλίμα δημιουργούν επιπλέον αβεβαιότητα σχετικά με τις ανανεώσιμες πηγές ενέργειας, καθώς αλλάζουν οι συνθήκες ηλιοφάνειας και ανέμου. Στον τομέα της γεωργίας, αναμένονται μικρότερες σοδειές και μείωση των εκτάσεων που είναι κατάλληλες για καλλιέργεια. Στον ασφαλιστικό τομέα αναμένονται αυξήσεις των ασφαλιστρών καθώς τα ακραία καιρικά φαινόμενα και οι καταστροφές γίνονται συχνότερα. Το γεγονός αυτό μπορεί να δημιουργήσει κοινωνικές ανισότητες καθώς οι οικονομικά ασθενέστεροι θα αδυνατούν να ασφαλιστούν. Σε διάφορες χώρες έχει ήδη αρχίσει να γίνεται λόγος για υποχρεωτική ασφάλιση απέναντι σε φυσικές καταστροφές [7], [8].

2.4 Αντιμετώπιση

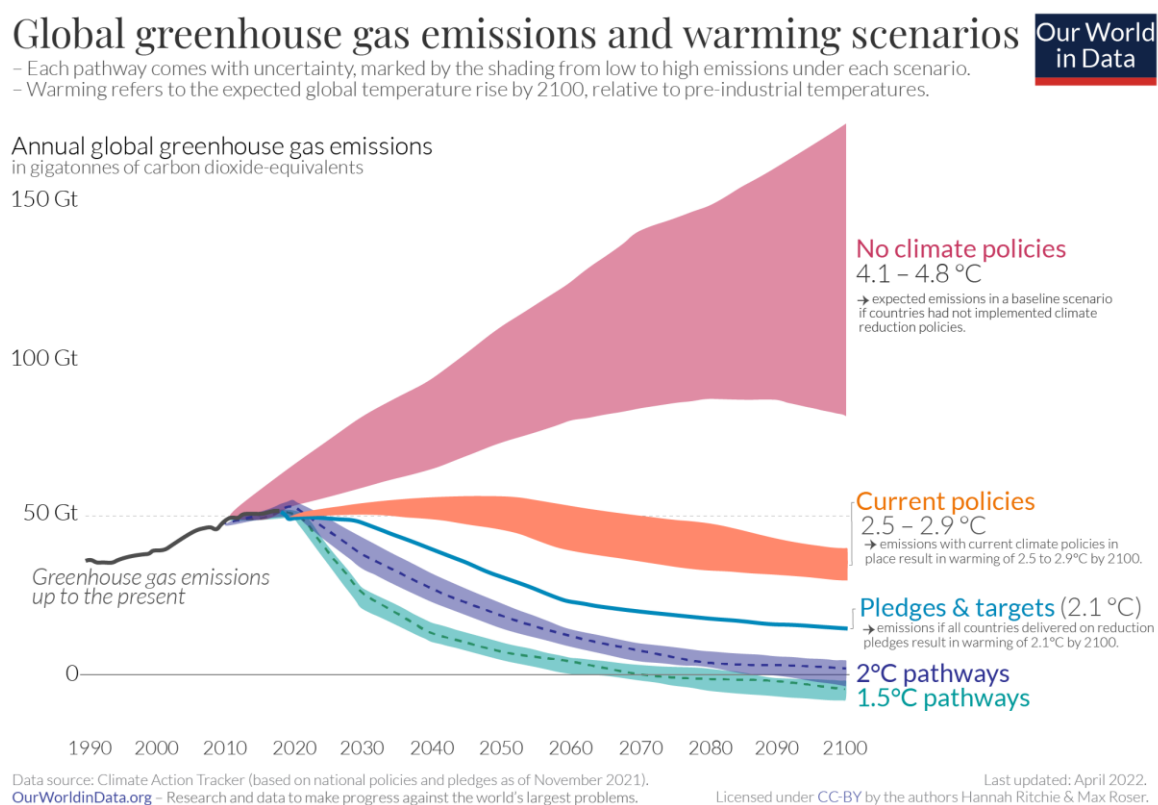
Καθώς η κλιματική αλλαγή οφείλεται πρωτίστως στα αέρια του θερμοκηπίου, για την αντιμετώπισή της γίνονται σε παγκόσμιο επίπεδο προσπάθειες περιορισμού των εκπομπών των αερίων αυτών. Οι δράσεις για την αντιμετώπιση της κλιματικής αλλαγής αναφέρονται συχνά στη βιβλιογραφία ως decarbonization.

Οι προσπάθειες αντιμετώπισης της κλιματικής αλλαγής επικεντρώνονται σε τεχνολογίες που αφορούν τέσσερις τομείς: από την πλευρά της παραγωγής στον τομέα της ενέργειας και από την πλευρά της ζήτησης στη βιομηχανία, τις μεταφορές και τα κτίρια. Στον τομέα της ενέργειας, η κλιματική αλλαγή μπορεί να αντιμετωπιστεί με τη χρήση ανανεώσιμων πηγών ενέργειας, πυρηνικής ενέργειας, αποθήκευσης ενέργειας και χρήση καυσίμων χαμηλών εκπομπών άνθρακα. Στους τομείς της ζήτησης οι προσπάθειες περιλαμβάνουν ενεργειακά αποδοτικές διαδικασίες και τεχνολογίες που μειώνουν την κατανάλωση ενέργειας, μετάβαση από ορυκτά καύσιμα σε ανανεώσιμα καύσιμα (π.χ. βιοντίζελ, βιοαιθανολη) καθώς και ενσωμάτωση τεχνολογιών ανανεώσιμης ενέργειας. Εκτός από τους παραπάνω τρόπους αντιμετώπισης της κλιματικής αλλαγής (συμβατικές μέθοδοι), έχουν προταθεί επίσης τεχνολογίες αρνητικών εκπομπών (negative emissions technologies), οι οποίες ως στόχο έχουν να αφαιρέσουν το διοξείδιο του άνθρακα από την ατμόσφαιρα [9].

Στη Συμφωνία του Παρισιού [10], που υπογράφηκε το 2016, τέθηκε ως στόχος ο περιορισμός της αύξησης της θερμοκρασίας του πλανήτη σε κάτω από 2 °C πάνω από τα προβιομηχανικά επίπεδα μέχρι το 2100 καθώς και ένας περισσότερος αισιόδοξος στόχος περιορισμού στον 1,5 °C. Ωστόσο, με την εφαρμογή των υφιστάμενων πολιτικών προβλέπεται αύξηση της θερμοκρασίας στους 2,7 °C μέχρι το 2100 [11].

2.5 Προβλέψεις για το μέλλον

Στην παρακάτω εικόνα παρουσιάζονται προβλέψεις για την εκπομπή αερίων του θερμοκηπίου και την αύξηση της θερμοκρασίας, ανάλογα με τις κλιματικές πολιτικές που θα εφαρμοστούν. Γίνεται ολοφάνερο πως αν δεν υπάρξει συντονισμένη δράση σε διακρατικό επίπεδο αλλά και αίσθημα ατομικής ευθύνης από το σύνολο του πληθυσμού τότε ίσως ο κόσμος του μέλλοντος να μην είναι πια όπως τον γνωρίζουμε. Είναι απαραίτητο να εντυπωθεί στη συνείδηση του πληθυσμού πως η απειλή βρίσκεται ήδη στο κατώφλι μας, αφήνοντας το στίγμα της στα καλοκαίρια και τους χειμώνες που βιώνουμε. Αν δεν αναλάβουμε δράση, δεν κινδυνεύει μόνο η φύση και το περιβάλλον αλλά και η ανθρωπότητα του μέλλοντος.



Εικόνα 2: Πρόβλεψη εκπομπής αερίων του θερμοκηπίου και μεταβολής της θερμοκρασίας της Γης

ανάλογα με τις εφαρμοζόμενες πολιτικές. Πηγή: [11]

3 Ενεργειακή κρίση

Ως ενεργειακή κρίση ορίζεται μία κατάσταση όπου **η συνολική ζήτηση για ενέργεια υπερβαίνει τη συνολική προσφορά σε μία οικονομία**, οδηγώντας σε αύξηση των τιμών ενέργειας. Η ενεργειακή κρίση μπορεί να έχει τοπικό ή παγκόσμιο χαρακτήρα. Κατά τη διάρκεια μιας ενεργειακής κρίσης, διαταράσσεται η προμήθεια, η διανομή και το κόστος των ενεργειακών πόρων, προκαλώντας ποικίλα προβλήματα στην οικονομία και την κοινωνία.

Η ενεργειακή κρίση δεν είναι ένα φαινόμενο που πρωτοεμφανίστηκε τα τελευταία χρόνια. Ήδη λίγο μετά τον 1^ο Παγκόσμιο Πόλεμο η Αμερική αντιμετωπίζει προβλήματα έλλειψης πετρελαίου, κάτι που συνέβη και μετά τη συμμετοχή της στον 2^ο Παγκόσμιο Πόλεμο [12]. Σημαντική είναι επιπλέον η ενεργειακή κρίση της δεκαετίας του 1970, η οποία επηρέασε πολλές χώρες του δυτικού κόσμου.

Η πιο πρόσφατη ενεργειακή κρίση, η οποία έχει αποκτήσει παγκόσμιο χαρακτήρα, ξέσπασε το 2021 και τα αποτελέσματά της είναι αισθητά ακόμα και σήμερα. Καθώς ο πλανήτης έκανε προσπάθειες ανάκαμψης από την πανδημία του COVID-19, όπου τα επίπεδα ζήτησης και παραγωγής ενέργειας είχαν πέσει, συνέβη η εισβολή της Ρωσίας στην Ουκρανία, επιφέροντας διαταραχές στην προμήθεια ενέργειας από την Ρωσία προς την Ευρώπη [13]. Η τιμή του φυσικού αερίου έφτασε σε ιστορικά υψηλά, συμπαρασύροντας και το κόστος του ηλεκτρισμού. Ταυτόχρονα, οι τιμές πετρελαίου έφτασαν στο υψηλότερο επίπεδο από το 2008 [14].

3.1 Αίτια

Κάθε ενεργειακή κρίση έχει διαφορετικά αίτια. Στο σημείο αυτό θα αναφέρουμε εκείνα τα οποία είναι κοινά στις περισσότερες από αυτές, καθώς και εκείνα τα οποία αφορούν τη σύγχρονη πραγματικότητα.

- Οι **πόλεμοι και οι γεωπολιτικές εντάσεις** ανέκαθεν διατάρασσαν την παγκόσμια παραγωγή και διανομή ενέργειας [12], [14].
- Η **συνεχής αύξηση του πληθυσμού** οδηγεί σε ολοένα και μεγαλύτερες ενεργειακές ανάγκες [15].
- Παγκόσμιες **κρίσεις**, όπως η COVID-19, οι οποίες αποσταθεροποιούν την οικονομία [16].

- Η ενεργειακή μετάβαση σε ανανεώσιμες πηγές ενέργειας, καθώς κλείνουν εργοστάσια παραγωγής ενέργειας τα οποία βασίζονται σε συμβατικά καύσιμα, μειώνοντας έτσι την συνολική παραγωγική ικανότητα.

3.2 Συνέπειες

Η ενεργειακή κρίση επηρεάζει τη ζωή και τη δραστηριότητα του ανθρώπου με τους παρακάτω τρόπους:

- **Υψηλό κόστος ενέργειας**, το οποίο με τη σειρά του προκαλεί αύξηση του κόστους των μεταφορών και τελικά των προϊόντων και υπηρεσιών [15].
- **Πληθωρισμός** και μείωση της αγοραστικής δύναμης των καταναλωτών [15].
- Εργοστάσια αναγκάζονται να **περιορίσουν την παραγωγή** [14].
- **Μείωση του περιθωρίου κέρδους** των επιχειρήσεων, το οποίο οδηγεί σε μειωμένες επενδύσεις και ανάπτυξη και αυξημένη ανεργία [15].
- **Φτώχεια** λόγω της υψηλής τιμής της ενέργειας, ιδιαίτερα στις αναπτυσσόμενες χώρες [14].
- **Αδυναμία πρόσβασης στην ενέργεια** και ιδιαίτερα στον ηλεκτρισμό, η οποία επηρεάζει την ποιότητα ζωής και προκαλεί κοινωνικές ανισότητες. Είναι χαρακτηριστικό πως στην Αφρική οι άνθρωποι που ζουν χωρίς ηλεκτρισμό αυξήθηκαν κατά παραπάνω από 15 εκατομμύρια μεταξύ του 2019 και του 2021 [15].
- Η **ανάγκη χρήσης ορυκτών καυσίμων** για την κάλυψη των αναγκών επιβαρύνει το περιβάλλον και δυσκολεύει τις προσπάθειες αντιμετώπισης της κλιματικής αλλαγής [15].
- **Αυξημένη κερδοφορία** για τους κατόχους μονοπωλίων ενέργειας ή αποθεμάτων, όσους ενισχύουν την ενεργειακή τους αυτονομία και όσους επενδύουν σε διακυμάνσεις των τιμών ενέργειας [16].

3.3 Αντιμετώπιση

Ορισμένες ενέργειες που είναι δυνατόν να πραγματοποιηθούν για την αντιμετώπιση της ενεργειακής κρίσης είναι οι παρακάτω:

- Χρήση **ενεργειακά αποδοτικών συσκευών** προκειμένου να επιτευχθεί εξοικονόμηση ενέργειας [15].
- Χρήση **ανανεώσιμων πηγών ενέργειας** ώστε να μην υπάρχει εξάρτηση από εισαγόμενα ορυκτά καύσιμα [15].
- Εύρεση **εναλλακτικών πηγών εφοδιασμού** και αποθήκευση ενέργειας [17].
- **Εξοικονόμηση ενέργειας** μέσω της αλλαγής των συνηθειών των καταναλωτών: Ρύθμιση της θέρμανσης σε χαμηλότερη θερμοκρασία και της ψύξης σε υψηλότερη, ρύθμιση χαμηλότερης θερμοκρασίας στον θερμοσίφωνα, κλείσιμο παραθύρων και βελτιώσεις στη μόνωση, χρήση ποδηλάτου και μέσων μαζικής μεταφοράς, οικονομική οδήγηση, κ.α. [14].
- Επιβολή **ελεγκτικών μέτρων σχετικά με το κέρδος** των επιχειρήσεων παραγωγής ενέργειας και παροχή επιδοτήσεων προς τους καταναλωτές [17].
- **Πολεοδομικοί κανονισμοί** ώστε να προωθηθεί ο ενεργειακά αποδοτικός σχεδιασμός και κατασκευαστική διαδικασία [15].

3.4 Η ενεργειακή κρίση ως ευκαιρία

Μέσα από την ενεργειακή κρίση μπορούν να γεννηθούν ευκαιρίες. Όπως τη δεκαετία του 1970, όπου η τότε κρίση προκάλεσε μεγάλη πρόοδο στην ενεργειακή αποδοτικότητα, έτσι και σήμερα, με αφορμή την πρόσφατη ενεργειακή κρίση δίνεται η ευκαιρία για επιτάχυνση της εγκατάστασης καθαρότερων προς το περιβάλλον πηγών ενέργειας [14].

Η υιοθέτηση ανανεώσιμης ενέργειας σε μεγάλη κλίμακα σε συνδυασμό με τη χρήση ενεργειακά αποδοτικών συσκευών αλλά και την καλλιέργεια της νοοτροπίας εξοικονόμησης ενέργειας όχι μόνο θα βοηθήσουν τον πλανήτη να ανακάμψει από τη βλάβη που έχει υποστεί αλλά και τους ανθρώπους να έχουν πρόσβαση σε οικονομικότερη ενέργεια απαλλαγμένη από γεωπολιτικές εξαρτήσεις.

4 Ενεργειακή μετάβαση

4.1 Η ανάγκη για ενεργειακή μετάβαση

Βρισκόμαστε σε ένα σημείο της ιστορίας όπου βιώνουμε ταυτόχρονα την κλιματική αλλαγή και την ενεργειακή κρίση, άλλοτε με εντονότερο και άλλοτε με ηπιότερο τρόπο. Έχοντας πλέον κατανοήσει τα δύο αυτά πολυδιάστατα φαινόμενα, μπορούμε να επικεντρωθούμε στην αντιμετώπισή τους. Από τα προηγούμενα κεφάλαια, γίνεται αντιληπτό πως υπάρχουν κοινές κατευθύνσεις που θα συμβάλλουν ταυτόχρονα στην επίλυση και των δύο προβλημάτων.

Προκειμένου λοιπόν να μειώσουμε το περιβαλλοντικό μας αποτύπωμα χρειάζεται να πραγματοποιήσουμε στροφή προς βιώσιμες μορφές ενέργειας, που δεν είναι άλλες από τις ανανεώσιμες πηγές ενέργειας. Παράλληλα, μπορούμε να μειώνουμε όσο είναι δυνατόν την κατανάλωση ενέργειας είτε μέσω ατομικών επιλογών είτε χρησιμοποιώντας ενεργειακά αποδοτικότερες συσκευές. Ακόμα, μπορούμε να αντικαταστήσουμε συστήματα που χρησιμοποιούν ορυκτά καύσιμα με νέα, τα οποία χρησιμοποιούν καθαρή ηλεκτρική ενέργεια, η οποία το βέλτιστο είναι να προέρχεται από ανανεώσιμες πηγές. Για παράδειγμα, η αντικατάσταση καυστήρων πετρελαίου με αντλίες θερμότητας ή η χρήση ηλεκτροκίνησης είναι βήματα που μας οδηγούν προς την απεξάρτησή μας από τα ορυκτά καύσιμα και τις αρνητικές συνέπειες που επιφέρουν.

Έτσι, γίνεται λόγος για ενεργειακή μετάβαση, μία μεταβολή στις τεχνολογίες και τις συνήθειες εκείνες ώστε μία πηγή ενέργειας να αντικατασταθεί από μία άλλη. Η ενεργειακή μετάβαση έχει συμβεί ξανά στο παρελθόν, όπως για παράδειγμα με την Βιομηχανική Επανάσταση, όπου σταδιακά έγινε μετάβαση από το ξύλο στον άνθρακα και στη συνέχεια στο πετρέλαιο και το φυσικό αέριο [18]. Στις ημέρες μας, όταν γίνεται λόγος για ενεργειακή μετάβαση, αυτή αφορά στην αντικατάσταση των ορυκτών καυσίμων από τις ανανεώσιμες πηγές ενέργειας [19].

4.2 Πλεονεκτήματα και μειονεκτήματα της ενεργειακής μετάβασης

Η ενεργειακή μετάβαση είναι μία αλλαγή που παρουσιάζει αδιαμφισβήτητα θετικά στοιχεία, ωστόσο δεν θα μπορούσαμε να παραβλέψουμε και τις αρνητικές επιπτώσεις της [20].

Η ενεργειακή μετάβαση, απαρτιζόμενη από τη στροφή προς τις ανανεώσιμες πηγές ενέργειας και την ταυτόχρονη μείωση της ζήτησης ενέργειας έχει τα εξής πλεονεκτήματα τόσο για το περιβάλλον όσο και για τον άνθρωπο:

- **Μετριαάζεται η κλιματική κρίση** και τα αποτελέσματά της, αφού περιορίζονται οι εκπομπές αερίων του θερμοκηπίου.
- Ενισχύεται η **κατανόηση των ανθρώπων** πως οι ενεργειακοί πόροι δεν είναι ανεξάντλητοι και θα πρέπει να χρησιμοποιούνται με σύνεση.
- **Μειώνεται η εξάρτηση από εισαγωγές ορυκτών καυσίμων**, οι οποίες επηρεάζονται από κοινωνικοπολιτικές αστάθειες και γεωπολιτικές εντάσεις.
- **Αυξάνεται η ενεργειακή ασφάλεια**, δηλαδή η απρόσκοπτη διαθεσιμότητα ενέργειας σε τιμές που μπορούν να αντέξουν οι καταναλωτές.
- Δημιουργούνται **νέες θέσεις εργασίας** σε διάφορους τομείς, όπως κατασκευή, συντήρηση, έρευνα, κ.τ.λ.
- Πραγματοποιούνται **υποδομικά έργα** σε περιοχές που βρίσκονται κοντά σε ΑΠΕ. Για παράδειγμα, ανοίγονται δρόμοι για την εγκατάσταση ανεμογεννητριών, οι οποίοι μπορούν να χρησιμοποιηθούν και από τους κατοίκους της περιοχής.

Από την άλλη, η ενεργειακή μετάβαση εμφανίζει σημαντικά μειονεκτήματα:

- Το **κόστος επένδυσης** που απαιτείται για την τοποθέτηση μονάδων ΑΠΕ είναι αρκετά υψηλό. Το γεγονός αυτό μπορεί να αποτελέσει εμπόδιο για την εγκατάστασή τους αλλά και να προκαλέσει αύξηση της τιμής ενέργειας.
- Οι μονάδες που εγκαθίστανται απαιτούν ορισμένες φορές την **παρέμβαση στο περιβάλλον** (χρήση γης, κοπή δέντρων), οδηγώντας στην απώλεια βιοποικιλότητας, καταστρέφοντας το τοπίο και δημιουργώντας οπτική και ηχητική ενόχληση στις τοπικές κοινότητες.

- Οι ανανεώσιμες πηγές ενέργειας επηρεάζονται από τα καιρικά φαινόμενα και η παραγωγή τους δεν μπορεί να ελεγχθεί. Έτσι, άλλοτε απαιτείται η χρήση συσσωρευτών για αποθήκευση ενέργειας και άλλοτε ενδέχεται η παραγωγή τους να μην είναι επαρκής για την κάλυψη των αναγκών.
- Άνθρωποι που εργάζονται σε μονάδες παραγωγής ενέργειας που βασίζονται σε παραδοσιακούς τρόπους ενδέχεται να χάσουν τη δουλειά τους.
- Ο εξοπλισμός που χρησιμοποιούν οι ΑΠΕ έχει περιορισμένη διάρκεια ζωής και η ανακύκλωσή του δεν είναι πάντα αποτελεσματική.
- Ενδέχεται να δημιουργηθούν νέες εξαρτήσεις από εισαγωγές υλικών όπως το λίθιο, το οποίο χρησιμοποιείται για την παραγωγή συσσωρευτών.

4.3 Προσπερνώντας εμπόδια

Οι αρνητικές επιπτώσεις της ενεργειακής μετάβασης είναι αδύνατο να παραβλεφθούν. Χαρακτηριστικά παραδείγματα αποτελούν η αύξηση των τιμών ενέργειας που παρατηρείται τα τελευταία χρόνια στη χώρα μας, παρόλο που ολοένα και αυξάνεται η χρήση ανανεώσιμων πηγών για την παραγωγή της. Για το γεγονός αυτό όμως υπεύθυνη είναι και η πρόσφατη ενεργειακή κρίση. Επιπλέον, ουκ ολίγες είναι οι περιπτώσεις όπου κάτοικοι, ενδεχομένως και δικαιολογημένα, διαμαρτύρονται για την εγκατάσταση ανεμογεννητριών στην περιοχή τους.

Οι αρνητικές επιπτώσεις είναι ωστόσο δυνατό και απαραίτητο να αντιμετωπιστούν. Με τη συνειδητοποίηση του προβλήματος της κλιματικής αλλαγής και ενεργειακής κρίσης και της αναγκαιότητας της ενεργειακής μετάβασης, με σωστό προγραμματισμό, συνεργασία και κατανόηση μεταξύ πολιτείας και κοινοτήτων είναι δυνατό να ζήσουν οι επόμενες γενιές σε έναν περισσότερο βιώσιμο κόσμο.

5 Δράσεις και κίνητρα ενεργειακής μετάβασης

Η ενεργειακή μετάβαση απαιτεί την υιοθέτηση νέων τεχνολογιών, κάτι που αρκετές φορές προϋποθέτει μία σημαντική επένδυση. Ακόμα, η περιορισμένη πληροφόρηση, η σύγκρουση συμφερόντων, η έλλειψη γνώσεων αλλά και η ελλιπής εκπροσώπηση των ενεργειακά αποδοτικών προϊόντων στην αγορά αποτελούν επιπλέον εμπόδια για τους παραγωγούς και τους καταναλωτές.

Στο πλαίσιο αυτό, τα τελευταία χρόνια οι κυβερνήσεις ανά τον κόσμο δημιουργούν προγράμματα και νομοθετικά πλαίσια ώστε να επιτευχθεί εξοικονόμηση ενέργειας και μείωση εκπομπών. Τα προγράμματα (δράσεις) συνήθως αφορούν σε κίνητρα που δίνονται στους κατασκευαστές, τους εμπόρους και τους καταναλωτές προκειμένου να στραφούν προς ενεργειακά αποδοτικότερα προϊόντα και τεχνολογίες. Τα νομοθετικά πλαίσια υποχρεώνουν τους παρόχους ενέργειας στο να πετύχουν μειωμένη κατανάλωση ενέργειας ή να επενδύσουν σε αύξηση της αποδοτικότητας.

Στην παρούσα εργασία επικεντρωνόμαστε σε δράσεις που σχεδιάζονται ώστε να ωθήσουν τους καταναλωτές στην αγορά και χρήση ενεργειακά αποδοτικότερων συσκευών, αλλά και σε ενεργειακές παρεμβάσεις ώστε να αυξηθεί η αποδοτικότητα των κτηρίων. Γενικά, τα κίνητρα που δίνονται στους καταναλωτές είτε παρέχονται με οικονομικές ελαφρύνσεις, όπως συμβαίνει σε πολλές χώρες της Ευρωπαϊκής Ένωσης, είτε με πόντους ανταμοιβής που μπορούν να εξαργυρωθούν σε διάφορες υπηρεσίες, όπως συνέβη στη Νότια Κορέα και την Ιαπωνία, είτε με επιδότηση αντικατάστασης παλαιών συσκευών [21].

5.1 Δράσεις ενεργειακής μετάβασης στην Ελλάδα

Στην Ελλάδα έχουν υλοποιηθεί διάφορες δράσεις ενεργειακής μετάβασης που απευθύνονται στους καταναλωτές. Ο τρόπος με τον οποίο παρέχεται κίνητρο είναι η επιδότηση για αγορά/εγκατάσταση αποδοτικών συσκευών με ή χωρίς προϋπόθεση αντικατάστασης καθώς και η επιδότηση για διαφόρων ειδών ενεργειακές παρεμβάσεις (π.χ. αντικατάσταση κουφωμάτων). Για παράδειγμα, η δράση «Ανακυκλώνω-Αλλάζω Συσκευή» επιδοτεί την αγορά ενεργειακά αποδοτικότερων συσκευών (ψυγεία, καταψύκτες, κλιματιστικά) με προϋπόθεση ανακύκλωσης των παλιών [22], ενώ η δράση «Ανακυκλώνω-Αλλάζω Θερμοσίφωνα» επιδοτεί την αντικατάσταση ηλεκτρικών θερμοσίφωνων από ηλιακούς [23]. Αντίστοιχη είναι η δράση «Αλλάζω συσκευή για τις επιχειρήσεις», η οποία απευθύνεται σε επιχειρηματίες και αφορά

στην αγορά συστήματος θέρμανσης με αντλία θερμότητας ή ηλιακή θέρμανση [24]. Στον τομέα της ηλεκτροκίνησης, δίνονται κίνητρα αγοράς ή χρονομίσθωσης ηλεκτρικών οχημάτων με το πρόγραμμα «Κινούμαι Ηλεκτρικά» [25], ενώ το πρόγραμμα «Πράσινα Ταξί» επιδοτεί και τα ηλεκτρικά ταξί [26]. Άλλα προγράμματα όπως το «Εξοικονομώ 2023» επιδοτούν δαπάνες για την ενεργειακή αναβάθμιση κτηρίων (π.χ. αντικατάσταση κουφωμάτων, τοποθέτηση θερμομόνωσης) [27], ενώ άλλα όπως το «Φωτοβολταϊκά στη Στέγη» επιδοτούν την εγκατάσταση συστήματος παραγωγής ηλεκτρικής ενέργειας με φωτοβολταϊκά [28].

5.2 Χαρακτηριστικά των δράσεων ενεργειακής μετάβασης

Στο [21] οι συγγραφείς παρουσιάζουν τα στοιχεία που απαρτίζουν τον σχεδιασμό ενός προγράμματος ενεργειακής αναβάθμισης. Από αυτά, θα κρατήσουμε εκείνα που σχετίζονται με δράσεις που απευθύνονται σε καταναλωτές, αφού εκείνοι θα είναι οι χρήστες του ψηφιακού βοηθού που θα κατασκευάσουμε. Παράλληλα, βάσει των δράσεων στη χώρα μας θα διακρίνουμε επιπλέον χαρακτηριστικά που θα ήταν χρήσιμο να γνωρίζει ο εν δυνάμει ωφελούμενος.

Στον παρακάτω πίνακα παρουσιάζονται τα χαρακτηριστικά των δράσεων ενεργειακής μετάβασης που απευθύνονται στους καταναλωτές:

Χαρακτηριστικό	Ερμηνεία
Όνομα δράσης	Η επίσημη ονομασία του προγράμματος
Κατηγορία δράσης	Η κατηγορία στην οποία ανήκει η δράση (βλέπε παρακάτω)
Κριτήρια απόδοσης	Τα κριτήρια που πρέπει να πληροί το αντικείμενο (π.χ. ενεργειακή κλάση μεγαλύτερη από A+)
Ποσό κινήτρου	Το ποσό ωφέλειας (π.χ. επιδότηση 30%)
Είδος κινήτρου	Ο τρόπος με τον οποίο δίνεται κίνητρο (π.χ. επιδότηση αγοράς, χαμηλότοκο δάνειο)
Προϋποθέσεις συμμετοχής	Οποιοσδήποτε προϋποθέσεις (π.χ. εισοδηματικές)
Απαίτηση ανακύκλωσης	Αν απαιτείται η παράδοση παλαιάς συσκευής ώστε να ανακυκλωθεί
Έναρξη αιτήσεων	Ημερομηνία έναρξης υποβολής αιτήσεων
Προθεσμία υποβολής αιτήσεων	Ημερομηνία λήξης υποβολής αιτήσεων

Πίνακας 1: Χαρακτηριστικά δράσεων ενεργειακής μετάβασης

5.3 Κατηγορίες δράσεων ενεργειακής μετάβασης

Βάσει των όσων αναπτύχθηκαν παραπάνω, διακρίνουμε τις εξής κατηγορίες δράσεων ενεργειακής μετάβασης:

Κατηγορία δράσης	Περιγραφή
Αντικατάσταση συσκευών	Επιδoteίται η αγορά συσκευών με προϋπόθεση ανακύκλωσης των παλιών.
ΑΠΕ (Ανανεώσιμες Πηγές Ενέργειας)	Επιδoteίται η αγορά/εγκατάσταση συστημάτων παραγωγής ηλεκτρικής ενέργειας από ανανεώσιμες πηγές.
Ενεργειακή αναβάθμιση	Επιδoteούνται ενεργειακές παρεμβάσεις (αγορά, εγκατάσταση, εργασία) χωρίς κάποια προϋπόθεση παράδοσης παλαιών συσκευών.
Ηλεκτροκίνηση	Επιδoteίται η αγορά ή χρονομίσθωση ηλεκτρικών οχημάτων.

Πίνακας 2: Κατηγορίες δράσεων ενεργειακής μετάβασης

5.4 Ανάγκη για αξιολόγηση των δράσεων ενεργειακής μετάβασης

Με το σχεδιασμό και την εφαρμογή των δράσεων ενεργειακής μετάβασης, τα κράτη επιδιώκουν να αυξήσουν το μερίδιο αγοράς των ενεργειακά αποδοτικών προϊόντων, σε συνδυασμό με τα υπάρχοντα κανονιστικά πλαίσια περί ενεργειακής αποδοτικότητας. Η αποτελεσματικότητα των εφαρμοζόμενων δράσεων και η μελλοντική τους βελτίωση εξαρτάται από την αξιολόγηση και παρακολούθησή τους, ώστε να εξάγονται αποτελέσματα σχετικά με την εξοικονόμηση ενέργειας και μείωση εκπομπών που επιτεύχθηκε, αλλά και αναφορικά με πιθανά εμπόδια που παρουσιάστηκαν κατά την εφαρμογή τους. Δυστυχώς, επειδή η διαδικασία αυτή απαιτεί χρόνο και χρήμα, οι δράσεις δεν αξιολογούνται σε συστηματική βάση από τα κράτη, στερώντας δεδομένα που θα μπορούσαν να βελτιώσουν την εφαρμογή των δράσεων μέσα στο ίδιο το κράτος αλλά και να φανούν χρήσιμα σε άλλα κράτη, μέσω της κοινοποίησής τους προς αυτά.

6 Τεχνητή Νοημοσύνη και LLMs

Η τεχνητή νοημοσύνη (Artificial Intelligence - AI) βιώνει μεγάλη άνθιση και εξάπλωση τα τελευταία χρόνια κυρίως λόγω της ευρείας διάδοσης των AI chatbots (π.χ. ChatGPT [29]) με τα οποία ο χρήστης αλληλεπιδρά ρωτώντας και παίρνοντας απαντήσεις για σχεδόν οτιδήποτε μπορεί να φανταστεί. Στο κεφάλαιο αυτό περιγράφονται ορισμένες έννοιες που αφορούν την τεχνητή νοημοσύνη, με έμφαση στα Μεγάλα Γλωσσικά Μοντέλα (Large Language Models – LLMs), στα οποία βασίζονται τόσο τα δημοφιλή chatbots όσο και ο ψηφιακός βοηθός τον οποίο θα κατασκευάσουμε στα πλαίσια της εργασίας αυτής.

6.1 Βασικοί ορισμοί

6.1.1 Τεχνητή Νοημοσύνη

Η τεχνητή νοημοσύνη στη γενική της έννοια αναφέρεται στην **νοημοσύνη που μπορεί να επιδεικνύουν τα υπολογιστικά συστήματα**. Ένας σημαντικός τομέας της επιστήμης της τεχνητής νοημοσύνης είναι η μηχανική μάθηση, η οποία αφορά στην ανάπτυξη και μελέτη αλγορίθμων οι οποίοι μπορούν με είσοδο δεδομένα να μαθαίνουν και να εκτελούν εργασίες χωρίς να έχουν προγραμματιστεί ρητά για αυτές.

Η μηχανική μάθηση βασίζεται σε **μοντέλα τεχνητής νοημοσύνης**, τα οποία είναι **προγράμματα που έχουν εκπαιδευτεί σε ένα σύνολο δεδομένων για να αναγνωρίζουν μοτίβα και να λαμβάνουν αποφάσεις χωρίς ανθρώπινη παρέμβαση**. Τα μοντέλα μπορούν σε γενικές γραμμές να χωριστούν σε παραγωγικά μοντέλα (generative models) και σε διακριτικά μοντέλα (discriminative models). Τα παραγωγικά μοντέλα μπορούν να παράγουν ως έξοδο νέο περιεχόμενο όπως κείμενο, εικόνες, βίντεο κ.τ.λ., ενώ τα διακριτικά μοντέλα εκπαιδεύονται στο να κατανοούν τις κατηγορίες στις οποίες χωρίζονται τα δεδομένα και μπορούν να χρησιμοποιηθούν για την ταξινόμηση αντικειμένων (π.χ. δεδομένης μιας ακτινογραφίας να αποφαινούνται για την πάθηση από την οποία πάσχει ο ασθενής).

Ένας ακόμα τρόπος για την κατηγοριοποίηση των μοντέλων τεχνητής νοημοσύνης είναι βάσει της χρήσης τους. Από τη μία έχουμε τα μοντέλα παλινδρόμησης (regression), τα οποία προβλέπουν συνεχείς τιμές, όπως η τιμή ενός προϊόντος και χρησιμοποιούνται για προβλέψεις, αξιολόγηση ρίσκου κ.α. Από την άλλη έχουμε τα μοντέλα ταξινόμησης (classification), τα οποία προβλέπουν διακριτές τιμές, όπως το αν μία εικόνα περιέχει ένα αυτοκίνητο και χρησιμοποιούνται για την κατηγοριοποίηση δεδομένων.

Ένα σημαντικό χαρακτηριστικό των μοντέλων τεχνητής νοημοσύνης είναι ο αριθμός παραμέτρων. Οι παράμετροι είναι οι μεταβλητές του μοντέλου που λαμβάνουν τιμές κατά τη διάρκεια της εκπαίδευσής του. Όσο περισσότερες είναι οι παράμετροι, τόσο περισσότερα δεδομένα χρειάζονται για την εκπαίδευσή του και τόσο πιο σύνθετο είναι το μοντέλο [30], [31].

6.1.2 Large Language Models – LLMs

Ως **LLM** ορίζεται ένα γλωσσικό μοντέλο γενικής χρήσης που βασίζεται σε **αρχιτεκτονική με μετατροπείς (transformers)** και έχει εκπαιδευτεί πάνω σε μεγάλο όγκο **κειμένου**. Τα LLM ανήκουν στην κατηγορία των παραγωγικών μοντέλων τεχνητής νοημοσύνης.

Οι μετατροπείς προτάθηκαν πρώτη φορά το 2017 στο [32]. Βάσει της αρχιτεκτονικής αυτής, **το κείμενο μετατρέπεται σε αριθμητικές αναπαραστάσεις που ονομάζονται tokens** και κάθε token μετατρέπεται σε ένα διάνυσμα. Ο τρόπος με τον οποίο το κείμενο χωρίζεται σε tokens εξαρτάται από το μοντέλο. Συνήθως, δημιουργείται ένα token ανα λέξη ή ανά χαρακτήρα ή ανά υπολέξη (αλληλουχία χαρακτήρων εντός μίας λέξης) [33]. Η OpenAI παρέχει ένα εργαλείο που οπτικοποιεί τον τρόπο με τον οποίο το κείμενο χωρίζεται σε tokens από τα μοντέλα της [34]. Με τη διαδικασία μετατροπής του κειμένου σε διανύσματα, λέξεις που βρίσκονται κοντά μεταξύ τους στον διανυσματικό χώρο έχουν παρόμοια σημασία. Με το παραπάνω ως θεμέλιο είναι δυνατή η κατανόηση και η δημιουργία όχι μόνο κειμένου, αλλά και άλλων μέσων όπως φωτογραφίες και βίντεο.

Ένα ακόμα χαρακτηριστικό των LLM στο οποίο αξίζει να αναφερθούμε είναι το **μήκος περιεχομένου (context length)**, το οποίο αφορά στο **μέγιστο αριθμό tokens που ένα μοντέλο μπορεί να δεχθεί ως είσοδο**.

6.1.3 Εκπαίδευση των LLMs

Στο σημείο αυτό θα αναφερθούμε επιγραμματικά στη διαδικασία εκπαίδευσης των LLM, προκειμένου να είμαστε σε θέση στη συνέχεια να κατανοήσουμε πιθανές πηγές μη ικανοποιητικών αποτελεσμάτων.

Η εκπαίδευση των LLMs χωρίζεται σε τρία στάδια [35]:

- 1. Προ-εκπαίδευση (pre-training):** Στο βασικό αυτό στάδιο το LLM αποκτά γνώση της γλώσσας, παγκόσμια γνώση και ικανότητες λογικής μέσω αυτό-επιτηρούμενης (self-supervised) εκπαίδευσης σε πολύ μεγάλο όγκο κειμένου. Στην ουσία, το LLM μπορεί να προβλέπει την κατανομή πιθανότητας της επόμενης λέξης μέσα σε μία πρόταση.
- 2. Επιτηρούμενη τελειοποίηση (supervised fine-tuning):** Μέχρι το στάδιο αυτό το LLM μπορεί να προβλέπει ποια θα είναι η επόμενη λέξη στην πρόταση που κατασκευάζει βάσει του κειμένου στο οποίο έχει εκπαιδευτεί, αλλά ενδεχομένως δεν μπορεί να ευθυγραμμιστεί με τον στόχο που έχει θέσει ο χρήστης στην είσοδο. Στο στάδιο αυτό χρησιμοποιούνται ζεύγη εντολής-απάντησης ώστε να βελτιωθούν περαιτέρω οι δυνατότητες του LLM να ακολουθεί τις εντολές που του δίνονται.
- 3. Ενισχυμένη εκπαίδευση μέσω ανθρώπινης ανατροφοδότησης (Reinforcement Learning from Human Feedback):** Στο στάδιο αυτό γίνεται εναρμονισμός του μοντέλου (μέχρι αυτό το σημείο ονομάζεται βασικό μοντέλο – base model) με τις ανθρώπινες προτιμήσεις. Για να πραγματοποιηθεί αυτό, χρησιμοποιείται ένα μοντέλο προτίμησης (preference model) το οποίο είναι εκπαιδευμένο να προβλέπει ποια είναι η προτιμώμενη απάντηση δεδομένης της εισόδου. Η έξοδος του μοντέλου αυτού ανατροφοδοτείται στο βασικό μοντέλο, δίνοντάς του πλέον τη δυνατότητα να παράγει ποιοτικότερα αποτελέσματα τα οποία είναι πιο κοντά στις ανθρώπινες αξίες και προτιμήσεις [36].

6.1.4 LLMs και κατανάλωση ενέργειας

Μία ανησυχία που έχει προκύψει με την εξάπλωση της τεχνητής νοημοσύνης και συγκεκριμένα των LLMs έχει να κάνει με την κατανάλωση ενέργειας που απαιτείται για την εκπαίδευση και τη χρήση τους. Τα LLMs τα τελευταία χρόνια έχουν μεγαλώσει από άποψη παραμέτρων, με το GPT-3 να διαθέτει 175 δισεκατομμύρια παραμέτρους. Όσο τα LLM μεγαλώνουν, τόσο αυξάνεται η ενέργεια που απαιτούν και συνεπώς το περιβαλλοντικό τους αποτύπωμα. Τα data centers στα οποία εκπαιδεύονται τα μοντέλα ορισμένες φορές εξακολουθούν να χρησιμοποιούν ενέργεια που παράγεται (και) με παραδοσιακούς τρόπους, συνεισφέροντας στο φαινόμενο του θερμοκηπίου και την κλιματική αλλαγή. Εκτός αυτού, η ενέργεια που απαιτείται για την εκπαίδευση ενός μοντέλου επιφέρει μεγάλο κόστος, κάτι που

μπορεί αν δυσκολεύει μικρότερες επιχειρήσεις οι οποίες θέλουν να παραμείνουν ανταγωνιστικές. Προκειμένου να είναι βιώσιμη η χρήση και η ανάπτυξη των LLMs αλλά και της τεχνητής νοημοσύνης γενικότερα, είναι σημαντικό να πραγματοποιηθούν δράσεις ώστε η ενέργεια που χρησιμοποιείται για την εκπαίδευση των μοντέλων να προέρχεται από ανανεώσιμες πηγές.

Η κατανάλωση ενέργειας των LLMs επηρεάζεται από πολλούς παράγοντες, με τους σημαντικότερους να είναι οι εξής:

- Μέγεθος του μοντέλου (αριθμός παραμέτρων)
- Αποδοτικότητα του υλικού (hardware) που χρησιμοποιείται
- Ώρες που αφιερώνονται στην εκπαίδευση
- Αποδοτικότητα υποδομής data center
- Διαδικασία τελειοποίησης (fine-tuning)
- Αποδοτικότητα αλγορίθμων
- Προετοιμασία δεδομένων εκπαίδευσης

Υπολογίζεται πως η εκπαίδευση ενός μικρού μοντέλου 7 δισεκατομμυρίων παραμέτρων και η χρήση του από ένα εκατομμύριο ανθρώπους απαιτεί 55,1 MWh, ενώ μόνο η εκπαίδευση του GPT-3, το οποίο έχει 175 δισεκατομμύρια παραμέτρους, απαίτησε 1.287 MWh, ενέργεια που καταναλώνει το μέσο αμερικανικό νοικοκυριό σε 120 χρόνια [37].

6.1.5 Αξιολόγηση των LLM

Η αξιολόγηση των LLMs παρέχει πληροφορία σχετικά με την απόδοσή τους πάνω σε διαφορετικές εργασίες. Στον πίνακα «Open LLM Leaderboard» [38] που παρουσιάζει η δημοφιλής πλατφόρμα HuggingFace μπορεί κανείς να δει την κατάταξη των ανοιχτών μοντέλων βάσει της επίδοσής τους σε μια σειρά από αξιολογήσεις. Παράλληλα, φαίνονται οι εκπομπές διοξειδίου του άνθρακα που έγιναν προκειμένου να αξιολογηθεί το μοντέλο, σε μια προσπάθεια ευαισθητοποίησης του χρήστη.

Οι συγγραφείς του [39] μοντελοποιούν τη διαδικασία αξιολόγησης των LLM ως την αλυσίδα

Μοντέλο → Τι (Εργασία) → Που (Δεδομένα) → Πώς (Διαδικασία)

Το **πρώτο βήμα** είναι φυσικά η **επιλογή του μοντέλου** που επιθυμούμε να αξιολογήσουμε.

Το **δεύτερο βήμα** αφορά στην **επιλογή των εργασιών** εκείνων πάνω στις οποίες θέλουμε να αξιολογήσουμε το μοντέλο.

Οι συγγραφείς εδώ διακρίνουν τις εξής κατηγορίες:

- **Επεξεργασία φυσικής γλώσσας (natural language processing – NLP):** Κατανόηση φυσικής γλώσσας και συναισθήματος, ταξινόμηση κειμένου, παραγωγή περίληψης, απάντηση ερωτήσεων κ.α.
- **Ανθεκτικότητα, ηθική, προκαταλήψεις και αξιοπιστία:** Η δυνατότητα του συστήματος να είναι ανθεκτικό σε απροσδόκητες εισόδους που προσπαθούν να παρακάμψουν δικλείδες ασφαλείας, το κατά πόσον αναπαράγει στερεότυπα και ρητορική μίσους κ.α.
- **Κοινωνική επιστήμη:** Κατανόηση οικονομικών, κοινωνιολογίας, πολιτικής επιστήμης, νομικής επιστήμης, ψυχολογίας κ.α. Η βελτίωση της απόδοσης των LLMs στους τομείς αυτούς μπορεί να συνεισφέρει σε ακαδημαϊκή έρευνα, νομικές διαδικασίες και επίλυση κοινωνικών προβλημάτων.
- **Φυσικές επιστήμες και μηχανική:** Η κατανόηση επιστημών όπως μαθηματικά, φυσική, χημεία και εφαρμογές τους στη μηχανική μπορεί να βοηθήσει σε επιστημονική έρευνα, ανάπτυξη της τεχνολογίας και διεξαγωγή τεχνικών μελετών.
- **Ιατρικές εφαρμογές:** Απαντήσεις σε ιατρικά ερωτήματα, συμβολή σε ιατρική εξέταση κ.α.
- **Εφαρμογές πρακτόρων (agents):** Χρήση επιπλέον εργαλείων για την περαιτέρω διεύρυνση των δυνατοτήτων των LLMs. Για παράδειγμα, δυνατότητα σύνδεσης του LLM με διαδικτυακές υπηρεσίες (web services) για την εμφάνιση δεδομένων καιρού πραγματικού χρόνου.
- **Άλλες εφαρμογές:** Χρήση των LLMs σε άλλους τομείς όπως εκπαίδευση, συστήματα προτάσεων (recommendation systems), ανάλυση προσωπικότητας ή εξειδικευμένες εφαρμογές όπως δημιουργία παιχνιδιών.

Το **τρίτο βήμα** είναι η **επιλογή του σετ δεδομένων και του συστήματος αξιολόγησης (benchmark)** που θα χρησιμοποιηθεί για την αξιολόγηση του μοντέλου. Τα σετ δεδομένων

προσομοιάζουν πραγματικά σενάρια επεξεργασίας της γλώσσας και καλύπτουν ευρύ φάσμα εργασιών, όπως αυτές περιγράφηκαν παραπάνω. Τα συστήματα αξιολόγησης εστιάζουν σε διαφορετικά σημεία και κριτήρια αξιολόγησης. Οι συγγραφείς τα διακρίνουν σε γενικών γλωσσικών εργασιών, ειδικών εργασιών (π.χ. ιατρικές ερωτήσεις) και πολυτροπικών εργασιών (multimodal tasks), όπως για παράδειγμα μοντέλα που δέχονται ή παράγουν εικόνες.

Το **τέταρτο και τελευταίο βήμα** έχει να κάνει με τη **διαδικασία αξιολόγησης**. Η διαδικασία αυτή μπορεί να είναι αυτόματη ή χειροκίνητη, δηλαδή με τη συμμετοχή ανθρώπου. Η αυτόματη διαδικασία αξιολόγησης είναι σύμφωνα με τους συγγραφείς πιθανώς η πιο διαδεδομένη και χρησιμοποιεί συγκεκριμένες μετρικές και εργαλεία για να αξιολογήσει τα μοντέλα. Επειδή δεν απαιτεί την ενεργή συμμετοχή του ανθρώπου, εκτός του ότι απαιτεί λιγότερο χρόνο, μειώνει την επίδραση του υποκειμενικού παράγοντα. Από την άλλη, η χειροκίνητη αξιολόγηση είναι μία επιλογή που προτιμάται σε πιο εξειδικευμένες εφαρμογές. Για παράδειγμα, εφαρμογές που έχουν να κάνουν με δημιουργία περιεχομένου, όπου δεν αρκεί μια σύγκριση ομοιότητας, η αξιολόγηση από άνθρωπο είναι προτιμότερη. Σε σύγκριση με την αυτόματη αξιολόγηση, η χειροκίνητη είναι πιο κοντά στην πραγματική εφαρμογή των LLMs, δηλαδή τη χρήση τους από ανθρώπους, έχοντας τη δυνατότητα να παράγει πιο περιεκτικά και ακριβή αποτελέσματα. Ωστόσο, οι ανθρώπινοι αξιολογητές μπορεί να δώσουν αποτελέσματα που χαρακτηρίζονται από αποκλίσεις και αστάθεια, λόγω των πολιτιστικών, γνωστικών και προσωπικών τους διαφορών. Για το λόγο αυτό έχει σημασία η συνετή επιλογή των αξιολογητών. Στην πράξη, η αυτόματη και η χειροκίνητη αξιολόγηση συνδυάζονται προκειμένου να παραχθεί όσο το δυνατόν πιο αξιόπιστο αποτέλεσμα.

6.2 Παραισθήσεις (hallucinations)

Όσον αφορά τα LLMs, **οι παραισθήσεις αναφέρονται στην δημιουργία περιεχομένου το οποίο είναι παράλογο ή δεν μπορεί να υποστηριχθεί από τις δεδομένες πληροφορίες**. Σε κρίσιμες εφαρμογές το φαινόμενο αυτό θα μπορούσε να έχει καταστροφικά αποτελέσματα, όπως για παράδειγμα η παραγωγή λανθασμένων οδηγιών λήψης φαρμάκου. Στην βιβλιογραφία, οι παραισθήσεις χωρίζονται βάσει της σχέσης της εξόδου με την είσοδο ως εξής [40], [41]:

- **Εσωτερικές (intrinsic)**, όπου η έξοδος του μοντέλου δεν συμφωνεί με το περιεχόμενο που δίνεται ως είσοδος. Για παράδειγμα, στην είσοδο αναφέρεται πως η ενεργειακή

δράση X επιδοτεί με 3.000€, ενώ σε ερώτημα το LLM θα ανταποκριθεί πως η επιδότηση είναι 4.000€.

- **Εξωτερικές (extrinsic)**, όπου η έξοδος του μοντέλου δεν μπορεί να υποστηριχθεί από τα δεδομένα που δίνονται ως είσοδος, δηλαδή δεν βρίσκεται καθόλου σε αυτά. Για παράδειγμα, ενώ στην είσοδο δεν δίνονται καθόλου πληροφορίες για επιδότηση αντικατάστασης οθονών ηλεκτρονικού υπολογιστή, το LLM σε ερώτημα για το ποια είναι τα διαθέσιμα προγράμματα επιδότησης θα συμπεριλάβει ένα πρόγραμμα αντικατάστασης οθονών στην απάντησή του.

Η δεύτερη περίπτωση δεν παράγει πάντα λανθασμένα αποτελέσματα, καθώς το μοντέλο μπορεί να χρησιμοποιεί το σύνολο δεδομένων πάνω στο οποίο εκπαιδεύτηκε για να διανθίσει την απάντησή του. Ωστόσο, το γεγονός πως οι πληροφορίες αυτές δεν είναι δυνατόν να επιβεβαιωθούν μας καθιστά επιφυλακτικούς στο να τις χρησιμοποιήσουμε σε κρίσιμες εφαρμογές ή σε εφαρμογές που η έξοδος πρέπει να στηρίζεται ως επί το πλείστον στην είσοδο (π.χ. παραγωγή περίληψης κειμένου, ερωτήματα πάνω σε συγκεκριμένα δεδομένα/βιβλιογραφία).

6.3 Αιτίες παραισθήσεων

Ο ψηφιακός οδηγός που θα υλοποιηθεί θα πρέπει να είναι όσο το δυνατόν αξιόπιστος σχετικά με τις απαντήσεις που θα δίνει. Συνεπώς, επιδιώκεται ελαχιστοποίηση των παραισθήσεων και για το λόγο αυτό είναι ωφέλιμο να εντοπίσουμε τις κύριες αιτίες εμφάνισής τους.

Οι παραισθήσεις οφείλονται κατά κανόνα στους εξής παράγοντες [40]:

1. Δεδομένα εκπαίδευσης

Είναι σημαντικό τα δεδομένα με τα οποία έχει εκπαιδευτεί το μοντέλο (training data) να είναι ποιοτικά και να παρουσιάζουν μεγάλη διασπορά. Από τη μία, αν δεν είναι ποιοτικά - περιέχουν δηλαδή λανθασμένα ή ανακριβή στοιχεία - το μοντέλο ενδέχεται να δώσει ως έξοδο στοιχεία που έρχονται σε σύγκρουση με την πραγματικότητα (factuality). Από την άλλη, αν τα δεδομένα δεν παρουσιάζουν ικανοποιητική διασπορά (γνώση για συγκεκριμένα γνωστικά αντικείμενα και χρονικά επίκαιρη), τότε το

μοντέλο δεν θα μπορεί να παράγει επιθυμητά αποτελέσματα σε περισσότερο συγκεκριμένες εργασίες ενώ παράλληλα θα δώσει αυξημένες παραισθήσεις.

2. Μέθοδοι εκπαίδευσης

Έχει παρατηρηθεί πώς κατά την προ-εκπαίδευση (1^ο στάδιο) η απόδοση του μοντέλου φθίνει καθώς αυξάνεται το μήκος της παραγόμενης αλληλουχίας λέξεων. Παράλληλα, σε μεγάλα μοντέλα πολλές φορές η έξοδος του προηγούμενου βήματος χρησιμοποιείται ως είσοδος για το επόμενο, ενισχύοντας ακόμα περισσότερο το φαινόμενο. Ακόμα, πολλές φορές κατά τη διάρκεια τελειοποίησης του μοντέλου (2^ο στάδιο), χρησιμοποιείται η έξοδος ενός ισχυρότερου μοντέλου. Αν η έξοδος του ισχυρότερου μοντέλου έχει παραισθήσεις, τότε θα δημιουργηθούν ακόμα περισσότερες παραισθήσεις στην έξοδο του μοντέλου που εκπαιδεύεται. Επιπλέον, έχει παρατηρηθεί πως κατά το 3^ο στάδιο εκπαίδευσης μπορεί να δημιουργηθούν επιπλέον παραισθήσεις, καθώς τα μοντέλα παράγουν απαντήσεις που είναι περισσότερο αρεστές στους ανθρώπους αλλά δεν συμβαδίζουν με την πραγματικότητα.

3. Δημιουργία της απάντησης

Το στάδιο κατά το οποίο το LLM παράγει την απάντησή του δεδομένης της εισόδου ονομάζεται αποκωδικοποίηση. Με τη διαδικασία αυτή το LLM μετατρέπει τις κωδικοποιημένες αναπαραστάσεις σε κείμενο κατανοητό στον άνθρωπο. Υπάρχουν πολλές στρατηγικές αποκωδικοποίησης, κάθε μία με τα πλεονεκτήματα και τα μειονεκτήματά της [42]. Η στρατηγική αποκωδικοποίησης είναι ουσιαστικά η μέθοδος με την οποία επιλέγεται κάθε φορά η επόμενη λέξη στην απάντηση που κατασκευάζει το μοντέλο. Συνεπώς, ο τρόπος με τον οποίο επιλέγεται η επόμενη λέξη (π.χ. το κατά πόσον το μοντέλο επιλέγει την επόμενη λέξη βάσει της πιθανότητας εμφάνισής της στα δεδομένα στα οποία έχει εκπαιδευτεί) μπορεί να συμβάλει σε λιγότερο ή περισσότερο απρόσμενη έξοδο και συνεπώς σε παραισθήσεις. Παράλληλα, μεγάλη συμβολή στην ποιότητα της απάντησης έχει και η είσοδος που δίνεται από τον χρήστη. Η διαδικασία της κατασκευής των οδηγιών που δίνονται στο μοντέλο ώστε να το καθοδηγήσουν σχετικά με την λειτουργία που έχει να επιτελέσει είναι γνωστή ως μηχανική προτροπών (prompt engineering). Μέσω προσεκτικά σχεδιασμένων οδηγιών είναι εφικτή η βελτίωση της απόδοσης των μοντέλων σε διάφορες λειτουργίες και εφαρμογές, χωρίς να χρειάζεται εκπαίδευση πάνω σε διαφορετικά δεδομένα. Ο αναγνώστης παραπέμπεται στο [43], όπου παρουσιάζεται πληθώρα τεχνικών.

6.4 Αξιολόγηση παραισθήσεων

Οι συγγραφείς του [44] χρησιμοποιούν δύο λέξεις για να περιγράψουν την ικανότητα του LLM να παράγει αξιόπιστο και σύμφωνο με τα δεδομένα περιεχόμενο. Συγκεκριμένα, ως **«πιστότητα» (faithfulness)** ορίζουν την ικανότητα να παράγεται έξοδος που βασίζεται και συμφωνεί με τη δεδομένη είσοδο, ενώ ως **«πραγματικότητα» (factuality)** ορίζουν την ικανότητα να παράγεται έξοδος που είναι σύμφωνη με γεγονότα και δεδομένα της παγκόσμιας γνώσης.

Στη βιβλιογραφία έχουν προταθεί αρκετές μεθοδολογίες σχετικά με την αξιολόγηση των παραισθήσεων (hallucination benchmarks) [35], [41]. Αυτές χωρίζονται σε δύο κατηγορίες:

- Οι μεθοδολογίες αξιολόγησης παραισθήσεων αξιολογούν το κατά πόσον τα LLMs παράγουν παραισθήσεις.
- Οι μεθοδολογίες αξιολόγησης εντοπισμού παραισθήσεων αξιολογούν τις υπάρχουσες τεχνικές εντοπισμού παραισθήσεων.

Οι **μεθοδολογίες αξιολόγησης παραισθήσεων** ποσοτικοποιούν την τάση των LLMs να παράγουν παραισθήσεις, δίνοντας έμφαση στην πιστότητα και την πραγματικότητα, όπως διατυπώθηκαν παραπάνω. Εστιάζουν στο να ελέγχουν τις απαντήσεις των LLMs πάνω σε σπάνια γνώση και δύσκολα ερωτήματα ή ερωτήσεις που προσπαθούν να παραπλανήσουν το LLM ώστε να παράγει λανθασμένες απαντήσεις. Για τη μέτρηση της επίδοσης χρησιμοποιούν ένα ερωτηματολόγιο πολλαπλής επιλογής που συμπληρώνεται είτε από τον άνθρωπο είτε από άλλα μοντέλα (proxy models).

Από την άλλη, οι **μεθοδολογίες αξιολόγησης εντοπισμού παραισθήσεων** εστιάζουν περισσότερο σε συγκεκριμένες εφαρμογές των LLMs. Σε γενικές γραμμές, αποτελούνται από κατασκευασμένες και μη πληροφορίες καθώς και από ένδειξη για το αν τα στοιχεία που περιέχουν είναι αληθινά ή όχι. Με τα σετ δεδομένων αυτά μπορεί να ελεγχθεί το κατά πόσον ένα LLM είναι σε θέση να αναγνωρίσει την ύπαρξη ψευδών πληροφοριών.

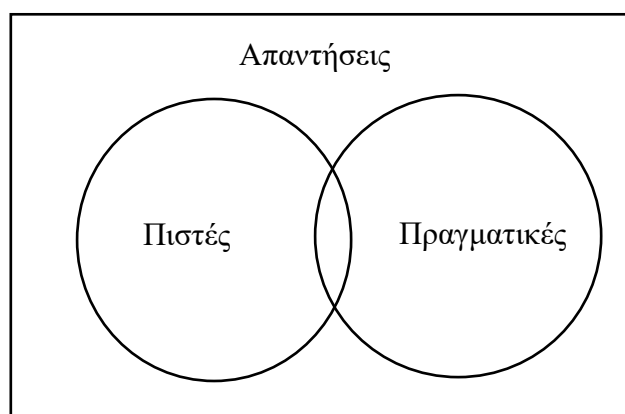
Όπως και με την αξιολόγηση των LLMs που παρουσιάστηκε παραπάνω, έτσι και η αξιολόγηση των παραισθήσεων μπορεί να γίνει αυτόματα ή χειροκίνητα, δηλαδή με τη συμμετοχή ανθρώπου. Η διαδικασία αυτόματου εντοπισμού παραισθήσεων είναι ακόμα σε πρωταρχικό στάδιο. Συνεπώς, η αξιολόγηση των παραισθήσεων γίνεται ως επί το πλείστον από τον άνθρωπο.

Υπάρχουν δύο κύριες κατηγορίες αξιολόγησης παραισθήσεων από τον άνθρωπο: Η πρώτη δίνει μία βαθμολογία του επιπέδου παραισθήσεων σε μία κλίμακα και η δεύτερη συγκρίνει την έξοδο ενός μοντέλου με κείμενα αναφοράς και παραπομπές.

Μία μεθοδολογία αξιολόγησης παραισθήσεων περιγράφεται στο [44] και έχει ως εξής:

- Η έξοδος του μοντέλου παρουσιάζεται στον αξιολογητή, μαζί με την αντίστοιχη είσοδο.
- Ο αξιολογητής καλείται να εντοπίσει αν στην έξοδο υπάρχουν παραισθήσεις. Σημειώνει τα σημεία του κειμένου στα οποία βρίσκονται παραισθήσεις καθώς και αν η παραίσθηση είναι εσωτερική ή εξωτερική.
- Υπολογίζονται τα ποσοστά των εξόδων που εμφάνισαν έστω μία εσωτερική, εξωτερική, ή τουλάχιστον ένα από τα δύο είδη παραισθήσεων.
- Όταν μία έξοδος δεν περιέχει ούτε εσωτερική ούτε εξωτερική παραίσθηση τότε χαρακτηρίζεται ως «πιστή» (faithful). Το ποσοστό των πιστών απαντήσεων υπολογίζεται ως $100 - \text{ποσοστό εξόδων που εμφάνισαν τουλάχιστον μία εσωτερική ή εξωτερική παραίσθηση}$.
- Όταν μία έξοδος είναι σύμφωνη με την πραγματικότητα τότε χαρακτηρίζεται ως «πραγματική» (factual). Υπολογίζεται το ποσοστό faithful+factual, δηλαδή το ποσοστό των πιστών ή/και πραγματικών απαντήσεων.

Παρακάτω παρουσιάζουμε ένα διάγραμμα Venn που οπτικοποιεί τα παραπάνω:



Εικόνα 3: Διάγραμμα Venn πιστών και πραγματικών απαντήσεων

Γίνεται κατανοητό πως όταν θέτουμε στο LLM ένα ερώτημα πάνω στην είσοδο που του δίνουμε, **προκειμένου η απάντηση να γίνεται αποδεκτή θα πρέπει να είναι τόσο πιστή όσο και πραγματική**. Επιπλέον, για την περίπτωση του ψηφιακού βοηθού που θα κατασκευάσουμε, θα απαιτήσουμε να είναι και **πλήρης**, δηλαδή **να περιέχει όλες εκείνες τις πληροφορίες που θα περιμέναμε να λάβουμε**. Για παράδειγμα, αν ρωτήσουμε το μοντέλο

ποιες είναι οι προϋποθέσεις ένταξης σε ένα πρόγραμμα ενεργειακής μετάβασης, παρέχοντάς του παράλληλα τις πληροφορίες του προγράμματος και αυτό μας απαντήσει μόνο με ορισμένες από τις προϋποθέσεις που βρίσκονται στις πληροφορίες που του δίνουμε, τότε μπορεί η απάντηση να είναι πιστή και πραγματική (ακολουθεί τις πληροφορίες που του δόθηκαν στην είσοδο και συμφωνεί με την παγκόσμια γνώση), αλλά δεν είναι πλήρης, καθώς έχουν παραληφθεί απαραίτητες για τον χρήστη πληροφορίες.

6.5 Αντιμετώπιση των παραισθήσεων

6.5.1 Τεχνικές αντιμετώπισης παραισθήσεων

Στη βιβλιογραφία αναφέρονται αρκετές τεχνικές αντιμετώπισης των παραισθήσεων. Αναφέρουμε επιγραμματικά ορισμένες από αυτές, δίνοντας έμφαση σε εκείνη την οποία τελικά θα χρησιμοποιήσουμε.

Ορισμένες μέθοδοι αναφέρονται στα **δεδομένα πάνω στα οποία εκπαιδεύεται το μοντέλο**. Στο [41] οι συγγραφείς παραπέμπουν σε μεθοδολογίες κατασκευής ενός έμπιστου συνόλου δεδομένων (dataset). Οι συγγραφείς του [35] κάνουν μία εκτενή αναφορά στη διαδικασία φιλτραρίσματος των δεδομένων εκπαίδευσης, επισημαίνοντας τη δυσκολία που υπάρχει καθώς τα δεδομένα εκπαίδευσης όλο και αυξάνονται, ενώ παράλληλα στα διαθέσιμα δεδομένα προστίθεται συνεχώς περιεχόμενο κατασκευασμένο από άλλα LLMs. Στο [44], οι συγγραφείς συμπεραίνουν πως σε όσο περισσότερο κείμενο εκτίθεται ένα μοντέλο κατά την προεκπαίδευση, τότε τόσο πιο έγκυρες απαντήσεις θα δώσει.

Όπως ήδη αναφέραμε, κατά το 3^ο στάδιο εκπαίδευσης είναι πιθανό να δημιουργηθούν περαιτέρω παραισθήσεις, λόγω του γεγονότος πως τα μοντέλα παράγουν απαντήσεις που είναι περισσότερο αρεστές παρά αληθείς. Για να αντιμετωπιστεί το φαινόμενο αυτό, προτείνεται η **χρήση πολλαπλών ανθρωπίνων προτιμήσεων** και η τελειοποίηση της εκπαίδευσης με εισόδους όπου η αλήθεια είναι ανεξάρτητη της άποψης του χρήστη [40].

Επιπλέον, γίνεται αναφορά σε **βελτιώσεις στις στρατηγικές αποκωδικοποίησης** που χρησιμοποιούνται για τη δημιουργία απάντησης [35], [40], [41], αλλά και σε τεχνικές επεξεργασίας «προβληματικών» περιοχών του μοντέλου [35].

Για τη δημιουργία ενός chatbot τεχνητής νοημοσύνης το οποίο θα βασίζεται σε συγκεκριμένες πληροφορίες και θα απαντά μόνο βάσει αυτών, η κατάλληλη μέθοδος για να ελέγξουμε τη συμπεριφορά του και να αποφύγουμε τις παραισθήσεις είναι να εφαρμόσουμε την τεχνική RAG, η οποία περιγράφεται αναλυτικά στη συνέχεια.

6.5.2 Retrieval-Augmented Generation (RAG)

Μία ευρέως χρησιμοποιούμενη τεχνική τόσο μείωσης των παραισθήσεων όσο και ενίσχυσης των γνώσεων και των δυνατοτήτων των LLMs πάνω σε συγκεκριμένους τομείς και εργασίες είναι η Retrieval-Augmented Generation (RAG).

Η RAG ενισχύει τα LLMs με πληροφορία που παρέχεται από εξωτερικές πηγές με χρήση υπολογισμού σημασιολογικής ομοιότητας. Συμβάλει στη σημαντική μείωση της πιθανότητας δημιουργίας λανθασμένου αποτελέσματος, κάτι που την έχει καταστήσει βασική τεχνική για την κατασκευή chatbots βασισμένα σε LLM [45].

Η διαδικασία στην οποία στηρίζεται η RAG αποτελείται από τα παρακάτω τέσσερα βασικά βήματα [45], [46]:

1. Ευρετηριοποίηση (Indexing)

Τα δεδομένα (π.χ. έγγραφα) που επιθυμούμε να εισάγουμε στο μοντέλο χωρίζονται σε κομμάτια (chunks) και μετατρέπονται σε αριθμητικές αναπαραστάσεις με τη μορφή διανυσμάτων, οι οποίες ονομάζονται ενσωματώσεις (embeddings). Η διαδικασία αυτή ονομάζεται κωδικοποίηση (encoding). Τα διανύσματα αποθηκεύονται σε μία διανυσματική βάση δεδομένων (vector database). Οι διανυσματικές βάσεις δεδομένων αποτελούν μια ειδική κατηγορία βάσεων δεδομένων, σχεδιασμένες για την αποθήκευση διανυσμάτων. Διαθέτουν λειτουργικότητα αναζήτησης διανυσμάτων (vector search), επιτρέποντας έτσι την αναζήτηση παρόμοιων αντικειμένων βάσει των χαρακτηριστικών τους [47].

2. Ανάκτηση (Retrieval)

Το ερώτημα του χρήστη μετατρέπεται σε διανύσματα με την ίδια στρατηγική κωδικοποίησης που χρησιμοποιήθηκε στο πρώτο βήμα. Έπειτα, υπολογίζεται η ομοιότητα μεταξύ του ερωτήματος και των αποθηκευμένων στην διανυσματική βάση διανυσμάτων των κομματιών πληροφορίας. Ανακτώνται τα κομμάτια εκείνα των οποίων οι διανυσματικές αναπαραστάσεις παρουσιάζουν τη μεγαλύτερη ομοιότητα με τις αντίστοιχες του ερωτήματος του χρήστη και χρησιμοποιούνται σαν επιπλέον πληροφορία στο ερώτημα που θα παραδοθεί στο LLM.

3. Επαύξηση (Augmentation)

Η πληροφορία που έχει ανακτηθεί καθώς και το ερώτημα που έχει δώσει ο χρήστης συντίθεται χρησιμοποιώντας τεχνικές μηχανικής προτροπών (prompt engineering), ώστε να διαμορφωθεί το τελικό επαυξημένο ερώτημα που θα υποβληθεί στο LLM. Στο στάδιο αυτό δίνονται συγκεκριμένες οδηγίες προς το LLM σχετικά με το έργο το οποίο έχει να επιτελέσει καθώς και με τον τρόπο με τον οποίο θα πρέπει να χειριστεί το ερώτημα του χρήστη και τις ανακτημένες πληροφορίες.

4. Δημιουργία (Generation)

Το LLM έχοντας στη διάθεσή του το τελικό ερώτημα, το οποίο περιέχει το ερώτημα του χρήστη, τις ανακτημένες πληροφορίες καθώς και τις πληροφορίες επαύξησης είναι σε θέση να δημιουργήσει την απάντησή του.

Παράδειγμα RAG

Επιθυμούμε τη δημιουργία ενός chatbot το οποίο εξειδικεύεται στην παροχή ιατρικών συμβουλών σε εγκύους. Τα στάδια που θα πρέπει να ακολουθήσουμε για την κατασκευή και χρήση ενός συστήματος RAG για το συγκεκριμένο σκοπό είναι τα ακόλουθα:

A. Κατασκευή

1. Συλλέγουμε έναν ικανοποιητικό όγκο αξιόπιστων πληροφοριών σχετικά με ιατρικά ζητήματα των εγκύων. Πάνω σε αυτή την πληροφορία θα έχει τη δυνατότητα να απαντά το LLM.
2. Τροφοδοτούμε την πληροφορία αυτή σε μία διανυσματική βάση δεδομένων.
3. Γράφουμε το πρότυπο του επαυξημένου ερωτήματος, το οποίο δηλώνει το σκοπό και τα δεδομένα που καλείται να επεξεργαστεί το LLM. Ένα τέτοιο ερώτημα θα μπορούσε να έχει την εξής μορφή:
«Είσαι ένας ψηφιακός βοηθός που καλείται να απαντάει σε ερωτήματα του χρήστη σχετικά με ιατρικά ζητήματα εγκύων. Θα πρέπει πάντα να απαντάς βάσει του δεδομένου περιεχομένου. Αν κάτι δεν βρίσκεται στο περιεχόμενο που σου δίνεται, θα πρέπει να απαντάς με: “Δεν γνωρίζω την απάντηση για το συγκεκριμένο ερώτημα”.

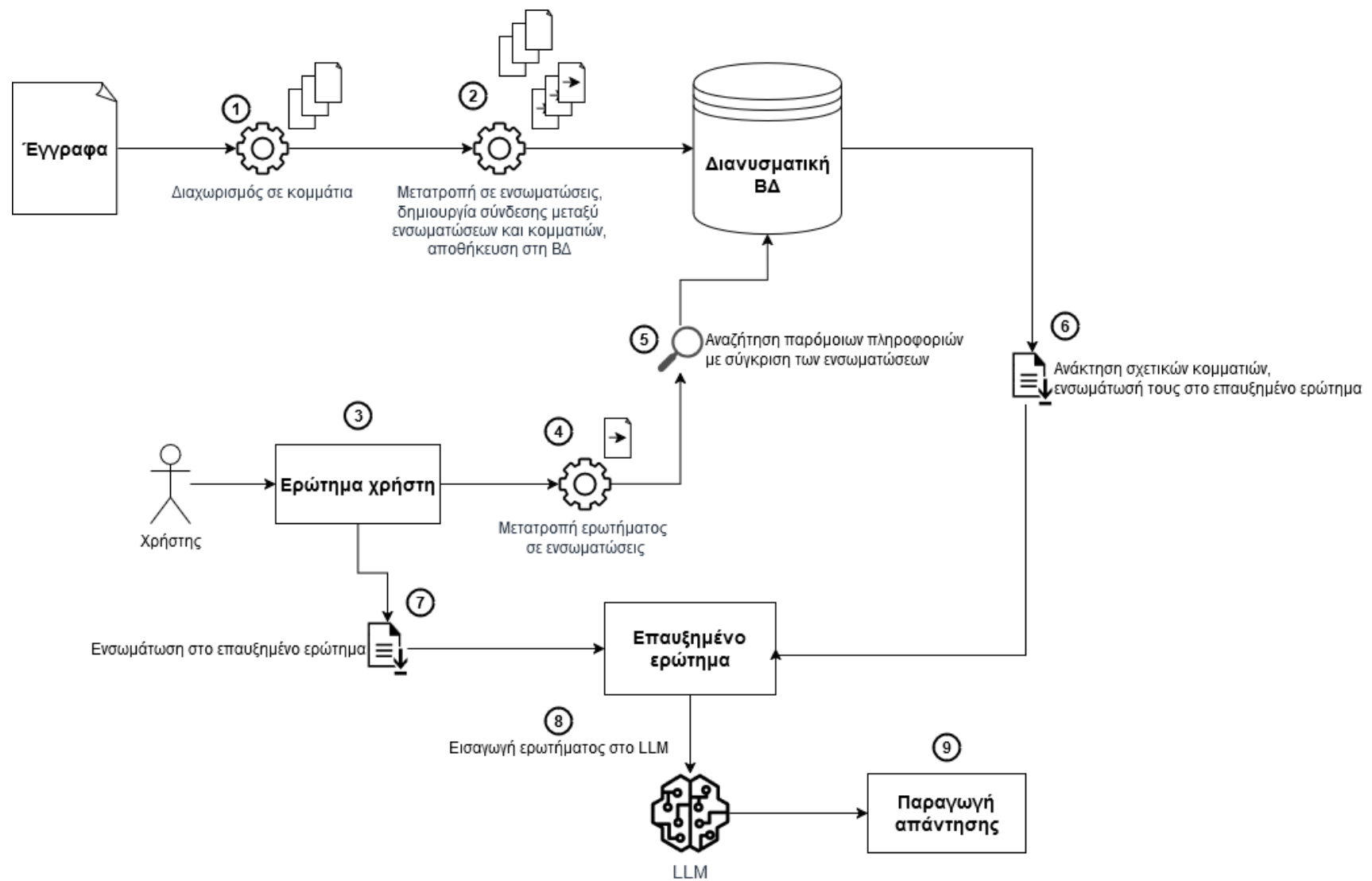
Περιεχόμενο: {Context}

Ερώτημα: {Question}»

B. Χρήση

1. Ο χρήστης θέτει το ερώτημά του στο σύστημα.
2. Το RAG σύστημα εντάσσει το ερώτημα του χρήστη καθώς και τις πλέον όμοιες με αυτό πληροφορίες από τη διανυσματική βάση στο παραπάνω επανυζημένο ερώτημα (μεταβλητές Question και Context) και το τροφοδοτεί στο LLM.
3. Το LLM παράγει την απάντησή του.

Η αρχιτεκτονική ενός συστήματος RAG φαίνεται στο παρακάτω σχήμα.



Εικόνα 4: Αρχιτεκτονική ενός συστήματος RAG

6.5.3 RAG και μηχανική προτροπών

Όπως αναφέρθηκε, η μηχανική προτροπών επικεντρώνεται στο βέλτιστο σχεδιασμό του ερωτήματος που τίθεται προς το μοντέλο, ώστε να βελτιωθεί η απόδοσή του πάνω σε μία συγκεκριμένη εργασία. Δεν υπάρχει σχεδιασμός που να είναι κατάλληλος για όλες τις περιπτώσεις και κάθε εργασία απαιτεί τη δική της προσέγγιση. Στο σημείο αυτό, διερευνούμε τη δομή και το περιεχόμενο που πρέπει να περιέχει το (επαυξημένο) ερώτημα προς το LLM σε ένα σύστημα RAG, καθώς και τα γενικότερα χαρακτηριστικά που πρέπει να το διέπουν.

Όσον αφορά τη **δομή και το περιεχόμενο** του επαυξημένου ερωτήματος, αυτή είναι σε γενικές γραμμές η εξής [48], [49], [50]:

- **Οδηγίες προς το μοντέλο.** Αυτές μπορεί να περιέχουν τα παρακάτω στοιχεία:
 - Χαρακτήρας του μοντέλου - Persona: Ορίζεται ο χαρακτήρας (ρόλος) τον οποίο καλείται να υποδυθεί το μοντέλο, ανάλογα με την εργασία την οποία καλείται να εκτελέσει. Για παράδειγμα «Είσαι ένας βοηθός σχετικά με την ερμηνεία νομικών κειμένων».
 - Αναφορά στο περιεχόμενο (context): Αναφέρεται πως θα πρέπει να απαντά βάσει του περιεχομένου που δίνεται στη συνέχεια.
 - Περιορισμοί: Γίνεται αναφορά σε τυχόν περιορισμούς στους οποίους θα πρέπει να υπόκειται η απάντηση του μοντέλου. Για παράδειγμα «Θα πρέπει η απάντησή σου να αποτελείται από 3 προτάσεις κατά μέγιστο» ή «Αν δεν ξέρεις την απάντηση θα πρέπει να απαντάς με “Δεν γνωρίζω την απάντηση για αυτό”».
- **Περιεχόμενο:** Δίνεται το περιεχόμενο που ανακτά το σύστημα RAG από τα δεδομένα έγγραφα, ανάλογα με το ερώτημα του χρήστη.
- **Ερώτημα χρήστη:** Δίνεται το ερώτημα του χρήστη.
- **Προτροπή για απάντηση:** Δίνεται το έναυσμα στο μοντέλο για να συμπληρώσει την απάντησή του.

Είναι σημαντικό να αναφερθεί πως ορισμένα μοντέλα ορίζουν συγκεκριμένη δομή (template) βάσει της οποίας θα πρέπει να γράφεται το ερώτημα, χρησιμοποιώντας λέξεις-κλειδιά. Για παράδειγμα, το μοντέλο Llama ορίζει το παρακάτω template [51]:

[INST]<<SYS>>{Οδηγίες προς το μοντέλο}<</SYS>>

Question: {question}

Context: {context}

Answer: [/INST]

Εκτός αυτού, δεν υπάρχουν όρια στο σχεδιασμό του ερωτήματος. Ο σχεδιαστής του συστήματος έχει την ελευθερία να δίνει οδηγίες στο μοντέλο σε φυσική γλώσσα, κάνοντάς το να λειτουργεί όπως αυτός επιθυμεί. Μπορεί έτσι να ρυθμίζει ακόμη και τον τόνο και το ύφος με τον οποίο θα απαντά το μοντέλο, ή να αποτρέπει τη χρήση συγκεκριμένων λέξεων/φράσεων από τον χρήστη.

Όσον αφορά στα **χαρακτηριστικά** που πρέπει να διέπουν το επαυξημένο ερώτημα, αυτά ακολουθούν τις καλές πρακτικές που πρέπει να διαθέτουν γενικά τα ερωτήματα προς τα LLMs και είναι τα ακόλουθα [52], [53]:

- Οι οδηγίες πρέπει να είναι **ξεκάθαρες και περιγραφικές** ως προς το ζητούμενο, δίνοντας την απαραίτητη λεπτομέρεια, ώστε να μην μένει χώρος στο μοντέλο να χρησιμοποιεί τη «φαντασία» του.
- Πρέπει να τηρείται **ισορροπία στο πλήθος πληροφοριών** που δίνονται. Υπερβολικά λεπτομερείς και φλύαρες οδηγίες θα φέρουν αρνητικό αποτέλεσμα.
- Τα **διαφορετικά στοιχεία της εισόδου είναι καλό να διαχωρίζονται**, όπως για παράδειγμα οι οδηγίες προς το μοντέλο με το περιεχόμενο πάνω στο οποίο θα πρέπει να βασίζει την απάντησή του. Για το σκοπό αυτό μπορεί π.χ. να χρησιμοποιείται μία κενή γραμμή.
- Είναι σημαντικό να δίνεται **ρόλος (persona)** στο μοντέλο, ώστε να κατανοεί την εργασία που καλείται να επιτελέσει και να προσαρμόζει την απάντησή του.
- Συμβάλει θετικά η **περιγραφή των βημάτων** που απαιτούνται ώστε να επιτευχθεί το ζητούμενο και αν χρειάζεται να δίνονται παραδείγματα.
- Μία **σύνθετη οδηγία πρέπει να χωρίζεται** σε απλές, σύντομες οδηγίες.

- Αντί να περιγράφεται τι δεν πρέπει να κάνει το μοντέλο, είναι προτιμότερο να περιγράφεται το **τι πρέπει να κάνει**.
- Αν χρειάζεται **ιδιαίτερη μορφή** στην έξοδο (π.χ. πίνακες) αυτό πρέπει να περιγράφεται στις οδηγίες.

Έχοντας τα παραπάνω κατά νου, παρουσιάζεται ένα παράδειγμα ενός επαυξημένου ερωτήματος που ακολουθεί τις βέλτιστες πρακτικές.

Παράδειγμα επαυξημένου ερωτήματος

Είσαι ένας νομικός σύμβουλος. Το ύφος σου είναι σοβαρό και εξυπηρετικό. Ο χρήστης θα σου θέτει ερωτήσεις σχετικά με νομικά ζητήματα.

Θα πρέπει να χρησιμοποιείς μόνο το ακόλουθο περιεχόμενο για να απαντήσεις την ερώτηση.

Αν δεν γνωρίζεις την απάντηση, θα απαντάς με «Δεν γνωρίζω την απάντηση».

Η απάντησή σου θα πρέπει να είναι το πολύ 5 προτάσεις και να είναι περιεκτική.

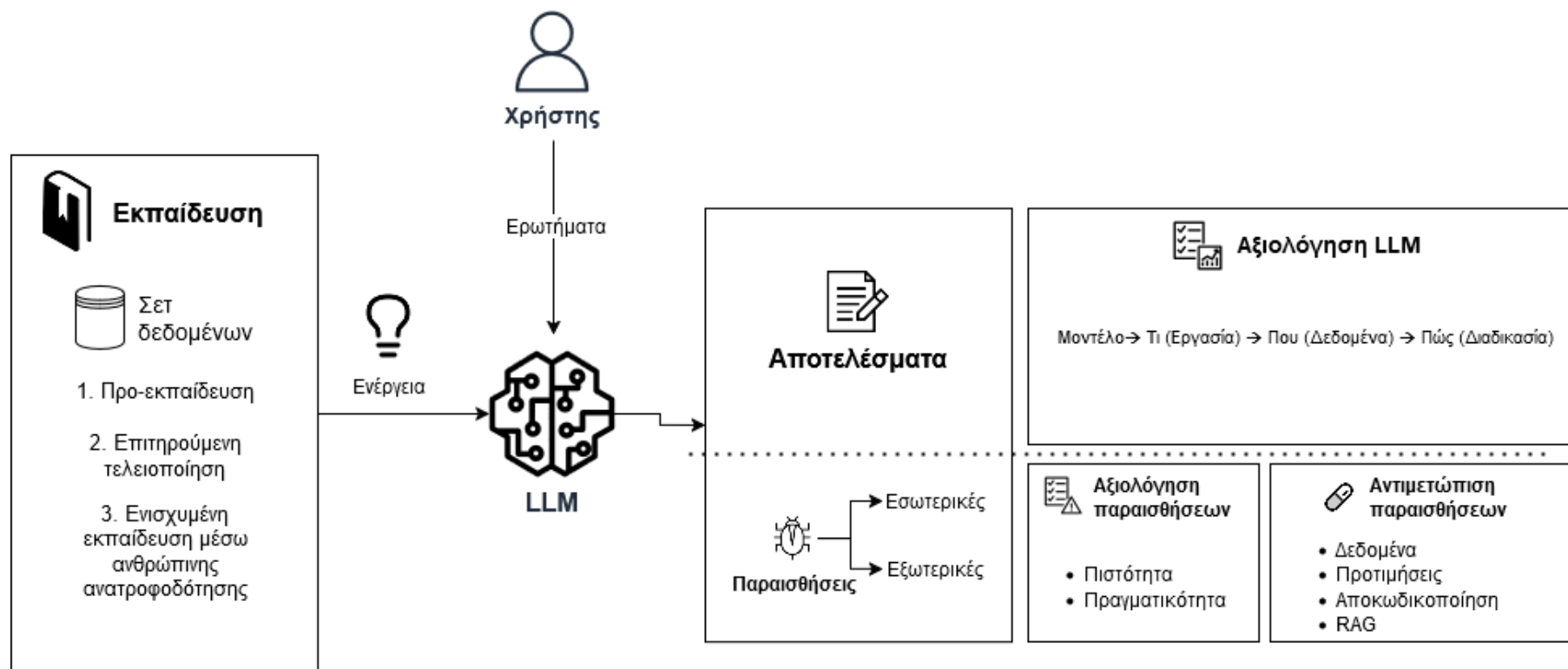
Περιεχόμενο: {περιεχόμενο}

Ερώτηση: {ερώτηση}

Απάντηση:

6.6 Επισκόπηση των LLMs

Στο παρόν κεφάλαιο πραγματοποιήθηκε μία παρουσίαση των βασικών στοιχείων, χαρακτηριστικών και διαδικασιών γύρω από τα LLMs και το περιβάλλον τους και τα οποία συνοψίζονται στο παρακάτω σχήμα.



Εικόνα 5: Στοιχεία και χαρακτηριστικά του LLM και του περιβάλλοντός του

7 LLMs και προγράμματα ενεργειακής μετάβασης

7.1 Υπάρχουσες λύσεις

Στο γενικό πλαίσιο της ενεργειακής μετάβασης και αναβάθμισης, οποιοδήποτε AI chatbot γενικής χρήσης θα μπορούσε να χρησιμοποιηθεί για την παροχή σχετικών συμβουλών. Επιπλέον, πολλά chatbots γενικής χρήσης έχουν πλέον τη δυνατότητα να εκτελούν αναζήτηση στο διαδίκτυο, κάτι που τα καθιστά ικανά στο να ανακτήσουν πληροφορίες για ισχύοντα προγράμματα ενεργειακής μετάβασης. Ωστόσο, τίθεται θέμα αμφισβήτησης της πληρότητας και της αξιοπιστίας των δεδομένων που ανακτούν. Στη λύση που προτείνεται στην εργασία αυτή, οι πληροφορίες έχουν ανακτηθεί ήδη από έγκυρες πηγές και δίνονται στο μοντέλο, το οποίο καλείται να απαντά μόνο με βάση αυτές, ενισχύοντας την αξιοπιστία του.

Εκτός από τα chatbots γενικής χρήσης, έχουν αναπτυχθεί διάφορα εργαλεία συγκεκριμένα για τον ενεργειακό τομέα. Το eco-bot [54] είναι μία λύση που βασίζεται σε διεπαφή τύπου chatbot και δίνει τη δυνατότητα στους χρήστες να παρακολουθούν την ενεργειακή τους κατανάλωση, το κόστος, τις εκπομπές CO², κ.α., δίνοντάς τους παράλληλα συμβουλές για μεγαλύτερη ενεργειακή αποδοτικότητα. Μέσω του εργαλείου οι χρήστες μπορούν επιπλέον να θέτουν ερωτήματα σχετικά με πληροφορίες και διαδικασίες που αφορούν τον πάροχο ενέργειάς τους. Το εργαλείο παρέχεται σε εταιρείες ενέργειας, σε εταιρείες BEMS (Building and Energy Management System), σε ESCOs (Energy Saving Companies), καθώς και σε εταιρείες λογισμικού. Η πρωτοβουλία αυτή έχει χρηματοδοτηθεί από το Ευρωπαϊκό πρόγραμμα Horizon 2020. Σε γενικές γραμμές στην Ελλάδα οι εταιρείες ενέργειας έχουν αναπτύξει ψηφιακούς βοηθούς [55], οι οποίοι είναι περισσότερο προσανατολισμένοι προς την παροχή πληροφοριών σχετικά με τα προϊόντα τους και την εξυπηρέτηση πελατών. Στο ίδιο πλαίσιο κινείται και ο ΔΕΔΔΗΕ [56]. Ένα ακόμα εγχείρημα που σχετίζεται με την ενέργεια στη χώρα μας είναι το EnergyBot [57], το οποίο υποστηρίζει πως βοηθά τον καταναλωτή να επιλέξει το κατάλληλο τιμολόγιο ενέργειας, βάσει ορισμένων στοιχείων όπως μηνιαία κατανάλωση ενέργειας, τύπος κατοικίας κτλ. Σε δοκιμές που πραγματοποιήθηκαν φαίνεται να μην τεκμηριώνει επαρκώς την απάντησή του, ενώ ωθεί τον χρήστη στην αγορά πακέτου ώστε να παρέχει εξατομικευμένες προτάσεις.

Στο πλαίσιο της ανακάλυψης και κατανόησης προγραμμάτων ενεργειακής μετάβασης, το πλέον κοντινό εργαλείο που υπάρχει στη χώρα μας είναι ο ψηφιακός βοηθός mAigon [58] που αναπτύχθηκε από το Υπουργείο Ψηφιακής Διακυβέρνησης [59]. Το mAigon είναι στην ουσία

ένα RAG σύστημα το οποίο πραγματοποιεί αναζήτηση στις υπηρεσίες του gov.gr και στις διαδικασίες του Εθνικού Μητρώου Διοικητικών Διαδικασιών «MITOS». Σε δοκιμές που πραγματοποιήθηκαν παρατηρήθηκε πως μπορεί να δώσει ικανοποιητικές πληροφορίες σχετικά με τα διαθέσιμα κρατικά προγράμματα ενεργειακής μετάβασης, όπως π.χ. ηλεκτροκίνησης, αντικατάστασης συσκευών κ.α. Ωστόσο, ορισμένες φορές παρουσίαζε δυσκολία στο να συνθέσει απάντηση από διαφορετικές σελίδες, δίνοντας έτσι ελλείψεις απαντήσεις, ενώ άλλοτε εμφάνιζε πλήρη αδυναμία παραγωγής απάντησης. Σημειώνεται πως κατά τη συγγραφή της εργασίας αυτής βρίσκεται ακόμα σε στάδιο beta και συνεπώς ενδέχεται να βελτιωθεί στο μέλλον.

Ο κρατικός ψηφιακός βοηθός mAigon διαφέρει από την πρόταση που παρουσιάζεται στα πλαίσια της εργασίας αυτής στα εξής σημεία:

mAigon	Πρόταση παρούσης εργασίας
Βασίζεται στο Azure Open AI [59], το οποίο δίνει πρόσβαση στα κλειστά μοντέλα της Open AI.	Βασίζεται σε ανοιχτό και ελεύθερα διαθέσιμο μοντέλο.
Διαθέτει πρόσβαση σε διάφορες κυβερνητικές πληροφορίες.	Αφορά αποκλειστικά στην παροχή πληροφοριών και την καθοδήγηση σχετικά με προγράμματα ενεργειακής μετάβασης.
Διαθέτει πρόσβαση μόνο σε κρατικά προγράμματα.	Ο τρόπος σχεδιασμού του επιτρέπει την τροφοδότησή του με ιδιωτικές πρωτοβουλίες που δίνουν κίνητρα ενεργειακής αναβάθμισης στους καταναλωτές, μέσω απλής εισαγωγής δεδομένων σε ένα αρχείο Excel.
Διαθέτει περιορισμό στο χρόνο συνεδρίας, αποσυνδέοντας τον χρήστη μετά από κάποια λεπτά.	Δεν υπάρχει περιορισμός στο χρόνο συνεδρίας.
Διατηρεί ιστορικότητα.	Δεν διατηρεί ιστορικότητα, κάθε ερώτημα αντιμετωπίζεται ως νέο.
Λαμβάνει ως είσοδο ιστοσελίδες.	Λαμβάνει ως είσοδο αρχείο δομημένο αρχείο Excel μετά από μετατροπή σε CSV.

Πίνακας 3: Σύγκριση του mAigon με τον ψηφιακό οδηγό που παρουσιάζεται στην παρούσα εργασία

7.2 Ένας ψηφιακός βοηθός για προγράμματα ενεργειακής μετάβασης

Εξ όσων γνωρίζουμε και όπως παρουσιάστηκαν παραπάνω, δεν υπάρχει ένας ψηφιακός βοηθός ο οποίος να βασίζεται σε LLM και να αφορά αποκλειστικά στην παροχή πληροφοριών και την απάντηση σε ερωτήματα σχετικά με τα προγράμματα ενεργειακής μετάβασης και τις ευκαιρίες που δίνονται στους καταναλωτές, ιδιαίτερα όσον αφορά στην Ελλάδα.

Θεωρούμε πως παρόλο που διατίθενται κατά διαστήματα διάφορα προγράμματα ενεργειακής μετάβασης, αυτά δεν βρίσκονται συγκεντρωμένα σε ένα σημείο και συχνά ο καταναλωτής δυσκολεύεται να τα κατανοήσει, να υπολογίσει το όφελος που θα έχει καθώς και ποια θα πρέπει να είναι η διαδικασία που θα πρέπει να ακολουθήσει, αναγκαζόμενος έτσι να καταφύγει σε ειδικούς. Η κατάσταση αυτή δημιουργεί αφενός καθυστέρηση στην ενεργειακή μετάβαση και την επίτευξη των στόχων που αυτή έχει θέσει και αφετέρου δρα κατά των ίδιων των καταναλωτών, οι οποίοι θα μπορούσαν να επωφεληθούν από την εκάστοτε επιδότηση και τα μακροπρόθεσμα οφέλη από την χρήση των νέων τεχνολογιών.

Στα πλαίσια της εργασίας αυτής πραγματοποιείται η δημιουργία ενός ψηφιακού βοηθού βασιζόμενου σε LLM, ο οποίος θα συγκεντρώνει πληροφορίες για κρατικά και μη κίνητρα για ενεργειακή αναβάθμιση και θα μπορεί με φιλικό προς το χρήστη τρόπο να αποτελέσει ένα ισχυρό εργαλείο, βοηθώντας τον να ανακαλύψει ποια είναι τα διαθέσιμα προγράμματα, να κατανοήσει τη σημασία τους και να καθοδηγηθεί σχετικά με τη διαδικασία που θα πρέπει να ακολουθήσει προκειμένου να ωφεληθεί από αυτά.

Τα βασικά χαρακτηριστικά του ψηφιακού βοηθού που προτείνεται και υλοποιείται είναι τα εξής:

- Μπορεί να δεχτεί πληροφορίες για προγράμματα ενεργειακής μετάβασης σε δομημένη μορφή και εύκολα διαχειρίσιμη από άτομα που διαθέτουν βασικές γνώσεις χειρισμού ηλεκτρονικού υπολογιστή (αρχείο excel).
- Βασίζεται σε ανοιχτό και ελεύθερα διαθέσιμο LLM, ώστε να μπορεί να χρησιμοποιηθεί ελεύθερα και με διάφορες μελλοντικές προεκτάσεις.
- Αφορά στα προγράμματα ενεργειακής μετάβασης που είναι διαθέσιμα στην Ελλάδα.
- Χρησιμοποιεί ως γλώσσα τα Ελληνικά.

8 Υλοποίηση του βοηθού ενεργειακής μετάβασης

Στο κεφάλαιο αυτό παρουσιάζονται τα εργαλεία που χρησιμοποιήθηκαν και περιγράφεται η διαδικασία που ακολουθήθηκε για την υλοποίηση του ψηφιακού βοηθού. Στην ουσία κατασκευάστηκε ένα σύστημα RAG, όπως αυτό περιγράφηκε στο κεφάλαιο 6. Το σύστημα αυτό τροφοδοτείται με πληροφορίες προγραμμάτων ενεργειακής μετάβασης στην Ελλάδα και ως σκοπό έχει να παρέχει ακριβείς απαντήσεις σε ερωτήσεις σχετικά με αυτά, παράγοντας τις λιγότερες δυνατές παραισθήσεις.

Ως γλώσσα προγραμματισμού επιλέχθηκε η Python, η οποία είναι η πλέον διαδεδομένη γλώσσα για εφαρμογές τεχνητής νοημοσύνης, ενώ σημαντικό ρόλο διαδραμάτισαν τα σύγχρονα εργαλεία ανάπτυξης λογισμικού τεχνητής νοημοσύνης που είναι ελεύθερα διαθέσιμα προς χρήση.

8.1 Χρήσιμα εργαλεία – Προγραμματιστικό υπόβαθρο

Για την ανάπτυξη του ψηφιακού βοηθού ενεργειακής μετάβασης χρησιμοποιήθηκαν δύο εργαλεία τα οποία απλοποίησαν σημαντικά το προγραμματιστικό κομμάτι και μας έδωσαν τη δυνατότητα να μπορούμε να εκτελούμε μοντέλα σε υποδομή νέφους (cloud infrastructure). Έτσι ήταν εφικτή η δοκιμή διαφόρων μοντέλων ενσωματώσεων αλλά και LLM και η επιλογή εκείνων που παρουσίαζαν την καλύτερη απόδοση.

8.1.1 LangChain – Ένα πλαίσιο ανάπτυξης LLM εφαρμογών

Η ανάπτυξη του ψηφιακού βοηθού διευκολύνθηκε ιδιαίτερα με τη χρήση του **LangChain** [60], ενός δημοφιλούς πλαισίου λογισμικού (software framework) για την ανάπτυξη εφαρμογών που βασίζονται σε LLMs. Το LangChain απλοποιεί τις διαδικασίες που απαιτούνται από ένα σύστημα RAG, όπως είναι η φόρτωση των δεδομένων και η μετατροπή τους σε ενσωματώσεις, αλλά και η αλληλεπίδραση με μοντέλα που είτε τρέχουν τοπικά είτε απομακρυσμένα.

8.1.2 HuggingFace – Μία βιβλιοθήκη μοντέλων τεχνητής νοημοσύνης

Αξίζει να αναφερθεί η συμβολή του **HuggingFace** [61] στη διαδικασία ανάπτυξης του ψηφιακού βοηθού αλλά και στην ευελιξία που μας προσέφερε. Η πλατφόρμα αυτή παρέχει

πρόσβαση σε εκατοντάδες μοντέλα για ποικίλες εφαρμογές όπως δημιουργία κειμένου, παραγωγή και σύγκριση ενσωματώσεων, αναγνώριση αντικειμένων, ταξινόμηση κ.α. Ο χρήστης μπορεί να κατεβάσει ένα μοντέλο και να το εκτελέσει στον υπολογιστή του ή αν επιθυμεί να το εκτελέσει απομακρυσμένα στις υποδομές του HuggingFace. Για την κατασκευή του ψηφιακού βοηθού έγινε χρήση της πλατφόρμας αυτής προκειμένου να ανακαλύψουμε τα μοντέλα που τελικά χρησιμοποιήσαμε, να δοκιμάσουμε την απόδοσή τους αλλά και να τα τρέξουμε απομακρυσμένα, όπως θα δούμε στη συνέχεια.

8.2 Συλλογή των δεδομένων

Το πρώτο βήμα της διαδικασίας υλοποίησης του βοηθού ενεργειακής μετάβασης ήταν η συλλογή πληροφοριών σχετικά με τα προγράμματα ενεργειακής μετάβασης στην Ελλάδα. Μετά από αναζήτηση στο διαδίκτυο, οι δράσεις ενεργειακής μετάβασης που συλλέχθηκαν τοποθετήθηκαν σε πίνακα, ο οποίος φαίνεται παρακάτω.

Οι δράσεις με αριθμό από 1 έως 10 αφορούν σε πραγματικά προγράμματα που ίσχυσαν ή ισχύουν ακόμα κατά την περίοδο συγγραφής της παρούσης εργασίας. Στη στήλη «Περιγραφή» συμπληρώνουμε τις σημαντικότερες πληροφορίες από την ιστοσελίδα της αντίστοιχης στήλης «Περισσότερες πληροφορίες». Το περιεχόμενο της στήλης παραλείπεται, ώστε να παραμένει ευανάγνωστος ο πίνακας. Εξαίρεση αποτελούν τα προγράμματα «Φωτοβολταϊκά στη Στέγη» και «Φωτοβολταϊκά στο χωράφι» όπου η επίσημη ιστοσελίδα δεν παρέχει επαρκείς πληροφορίες, καθώς και οι δράσεις «Εξοικονομώ 2023» και «Εξοικονομώ – Ανακαινίζω για νέους», όπου οι επίσημες ιστοσελίδες δεν περιέχουν συγκεντρωμένες τις απαραίτητες πληροφορίες, οπότε και χρησιμοποιήθηκαν δεδομένα από τις ιστοσελίδες

- <https://www.gov.gr/ipiresies/polites-kai-kathemerinoteta/periballon-kai-poiotetazoes/photovoltaika-ste-stege>
- <https://greece20.gov.gr/?calls=fotovoltaiika-sto-chorafi>
- <https://www.gov.gr/ipiresies/periouisia-kai-phorologia/epidoteseis-politon/exoikonomo-2023>
- <https://www.gov.gr/ipiresies/periouisia-kai-phorologia/epidoteseis-politon/exoikonomo-anakainizo-to-spiti-mou-gia-neous>

Οι δράσεις με αριθμούς 11 και 12 αφορούν σε φανταστικά προγράμματα ιδιωτικού παρόχου (η εταιρεία Acme Corporation), τα οποία ωστόσο λίγο διαφέρουν από αυτά που συναντούμε στην πραγματικότητα.

Οι πληροφορίες του πίνακα τοποθετούνται σε ένα υπολογιστικό φύλλο σε αρχείο τύπου Excel. Επιλέχθηκε η μορφή αυτή καθώς είναι δομημένη, επιτρέπει γρήγορη επισκόπηση της πληροφορίας και είναι εύκολα συντηρήσιμη από άτομα που δεν έχουν μεγάλη εξοικείωση με την τεχνολογία.

Οι στήλες του πίνακα βασίζονται στα χαρακτηριστικά των δράσεων ενεργειακής μετάβασης που αναπτύχθηκαν στην ενότητα 5.2, τροποποιημένα ώστε να δομηθεί κατάλληλα η πληροφορία για να βοηθηθεί το LLM στην απάντηση ερωτήσεων υψηλού ενδιαφέροντος (π.χ. ερώτηση για το αν ισχύει μία δράση ή για το που μπορεί ο χρήστης να βρει περισσότερες πληροφορίες).

A/A	Όνομα δράσης	Πάροχος	Κατηγορία	Περιγραφή	Σε ισχύ (ανοιχτές αιτήσεις)	Περισσότερες πληροφορίες
1	Ανακυκλώνω - Αλλάζω Συσκευή	Ελληνικό δημόσιο	Αντικατάσταση συσκευών	<παραλείπεται>	Όχι	https://allazosyskevi.gov.gr/
2	Αλλάζω συσκευή για τις επιχειρήσεις	Ελληνικό δημόσιο	Ενεργειακή αναβάθμιση	<παραλείπεται>	Ναι	https://exoikonomo-epixeiro2023.gov.gr/
3	Ανακυκλώνω - Αλλάζω Θερμοσίφωνα	Ελληνικό δημόσιο	Αντικατάσταση συσκευών	<παραλείπεται>	Όχι	https://allazothermosifona.gov.gr/
4	Κινούμαι Ηλεκτρικά Γ' Κύκλος	Ελληνικό δημόσιο	Ηλεκτροκίνηση	<παραλείπεται>	Ναι	https://kinoumeilektrika3.gov.gr/
5	Πράσινα Ταξί	Ελληνικό δημόσιο	Ηλεκτροκίνηση	<παραλείπεται>	Ναι	https://prasinataxi.gov.gr/
6	Φωτοβολταϊκά στη στέγη	Ελληνικό δημόσιο	ΑΠΕ (Ανανεώσιμες Πηγές Ενέργειας)	<παραλείπεται>	Όχι	https://ape4all.deddie.gr/
7	Φωτοβολταϊκά στο χωράφι	Ελληνικό δημόσιο	ΑΠΕ (Ανανεώσιμες Πηγές Ενέργειας)	<παραλείπεται>	Όχι	https://greece20.gov.gr/?calls=fotovoltaika-sto-chorafi
8	Επιδότηση εγκατάστασης	ΕΔΑ Αττικής	Ενεργειακή αναβάθμιση	<παραλείπεται>	Όχι	https://edaattikis.gr/gia-to-spiti/neo-programma-epidotisis-egkatastasis-

	θέρμανσης φυσικού αερίου					thermansis-fysikou-aeriou-se-katoikies-apo-tin-eda-attikis-2022-2/
9	Εξοικονομώ 2023	Ελληνικό δημόσιο	Ενεργειακή αναβάθμιση	<παραλείπεται>	Όχι	https://exoikonomo2023.gov.gr/
10	Εξοικονομώ – Ανακαινίζω για νέους	Ελληνικό δημόσιο	Ενεργειακή αναβάθμιση	<παραλείπεται>	Όχι	https://exoikonomoneon.gov.gr/
11	Δωροεπιταγή ανακύκλωσης για αντικατάσταση κλιματιστικού	Acme Corporation	Αντικατάσταση συσκευών	Με την αγορά νέου κλιματιστικού και την προσκόμιση του παλιού σας επωφελείστε από δωροεπιταγή 100 Ευρώ για τις επόμενες αγορές σας στα καταστήματά μας.	Ναι	https://www.example.com/air-condition
12	Δωρεάν εγκατάσταση αντλίας θερμότητας	Acme Corporation	Ενεργειακή αναβάθμιση	Με την αγορά οποιασδήποτε αντλίας θερμότητας σας κάνουμε δώρο την εγκατάστασή της.	Ναι	https://www.example.com/heat-pump

Πίνακας 4: Δεδομένα δράσεων ενεργειακής μετάβασης με τα οποία τροφοδοτούμε το μοντέλο

8.3 Μετατροπή των δεδομένων σε CSV

Το πρώτο βήμα της επεξεργασίας των δεδομένων είναι η μετατροπή τους σε μορφή CSV. Η μορφή αυτή διευκολύνει τη μετέπειτα φόρτωση, κατακερματισμό και αποθήκευσή τους στη βάση δεδομένων, καθώς το CSV είναι μία ανοιχτή μορφή αρχείου που υποστηρίζεται από ποικίλες βιβλιοθήκες λογισμικού. Η μετατροπή του αρχείου μορφής Excel σε CSV γίνεται με χρήση της βιβλιοθήκης pandas. Ο κώδικας είναι διαθέσιμος στο παράρτημα.

Αξίζει να σημειωθεί πως επιλέγουμε να αντικαταστήσουμε τον προεπιλεγμένο οριοθέτη (delimiter) του CSV που είναι το κόμμα (,) με το ελληνικό ερωτηματικό (semi-colon - ;), καθώς τα δεδομένα που εισάγουμε ενδέχεται να περιέχουν κόμμα και δεν θέλουμε αυτό να θεωρηθεί ως οριοθέτης κατά την ανάγνωση του αρχείου CSV.

8.4 Φόρτωση και κατακερματισμός των δεδομένων

Προκειμένου να αποθηκεύσουμε στη διανυσματική βάση δεδομένων τα δεδομένα και τις αντίστοιχες ενσωματώσεις τους, πρέπει πρώτα να τα φορτώσουμε και να τα κατακερματίσουμε, όπως είδαμε στην ενότητα 6.5.

Η φόρτωση των δεδομένων γίνεται με τη βοήθεια της κλάσης CSVLoader, ως εξής:

```
from langchain.document_loaders import CSVLoader
loader = CSVLoader('./data.csv', encoding='UTF-8', csv_args={"delimiter": ";"})
documents = loader.load()
```

Έτσι, τα δεδομένα του αρχείου data.csv φορτώνονται στο περιβάλλον της εφαρμογής. Παρατηρούμε πως εφόσον έχουμε επιλέξει διαφορετικό οριοθέτη για το αρχείο μας, πρέπει να τον δηλώσουμε και στο σημείο αυτό.

Στη συνέχεια, κατακερματίζουμε τα δεδομένα σε κομμάτια, κάνοντας χρήση της κλάσης CharacterTextSplitter, ως εξής:

```
from langchain.text_splitter import CharacterTextSplitter
text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=150)
docs = text_splitter.split_documents(documents)
```

Στο σημείο αυτό αξίζει να αναφερθούμε στις δύο σημαντικές παραμέτρους που ορίζουμε για τον κατακερματισμό των δεδομένων, το `chunk_size` και το `chunk_overlap`.

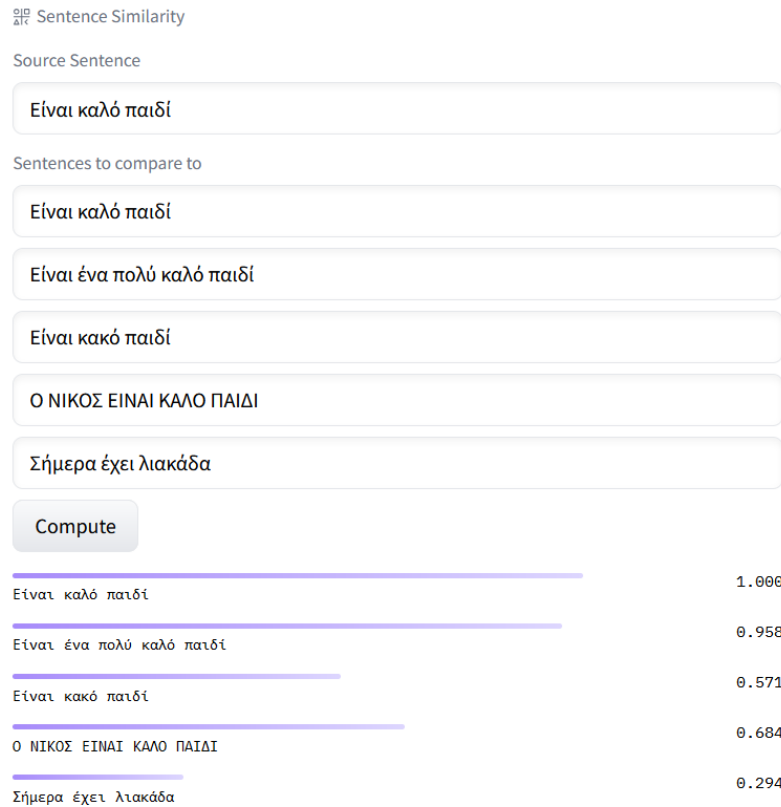
Το **chunk_size** ορίζει τον μέγιστο αριθμό χαρακτήρων που θα περιέχεται σε κάθε κομμάτι. Αν τα κομμάτια είναι πολύ μικρά, τότε είναι πιθανό το μοντέλο να μην μπορεί να κατανοήσει το περιεχόμενο τους, ενώ αν είναι πολύ μεγάλα, τότε δημιουργούν υπολογιστικό φόρτο.

Το **chunk_overlap** ορίζει τον αριθμό χαρακτήρων που επικαλύπτονται μεταξύ διαδοχικών κομματιών. Η επικάλυψη βοηθά το μοντέλο να κατανοεί τη σημασιολογική συνέχεια μεταξύ των διαδοχικών κομματιών. Είναι ιδιαίτερα χρήσιμη όταν οι προτάσεις είναι μεγάλες και έχουν χωριστεί σε κομμάτια.

8.5 Ενσωματώσεις

Η επιλογή του μοντέλου ενσωματώσεων που θα χρησιμοποιήσουμε παίζει πολύ σημαντικό ρόλο στο να ανακτώνται από τη βάση δεδομένων τα κομμάτια εκείνα που έχουν τη μεγαλύτερη σχέση με το ερώτημα του χρήστη. Για την παραγωγή των ενσωματώσεων επιλέχθηκε το μοντέλο **sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2** [62]. Το μοντέλο αυτό έχει εκπαιδευτεί σε 50 γλώσσες, συμπεριλαμβανομένων των ελληνικών. Μετατρέπει προτάσεις και παραγράφους σε διανύσματα 384 διαστάσεων και μπορεί να χρησιμοποιηθεί για ταξινόμηση και σημασιολογική αναζήτηση.

Μέσω του περιβάλλοντος του HuggingFace μας δίνεται η δυνατότητα να δοκιμάσουμε την επίδοση του μοντέλου, ζητώντας του να μας δώσει έναν βαθμό συσχέτισης μίας πρότασης με ορισμένες άλλες. Ο μέγιστος βαθμός συσχέτισης είναι 1 και δίνεται όταν οι δύο προτάσεις είναι όμοιες. Στην ακόλουθη εικόνα συγκρίνεται η φράση «Είναι καλό παιδί» με άλλες 5 φράσεις και εμφανίζεται ο αντίστοιχος βαθμός συσχέτισης.



Εικόνα 6: Σύγκριση προτάσεων με το μοντέλο *sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2*

Παρατηρούμε πως το μοντέλο αναγνωρίζει την σημασιολογική ομοιότητα ανεξάρτητα από το αν χρησιμοποιούνται μικρά ή κεφαλαία.

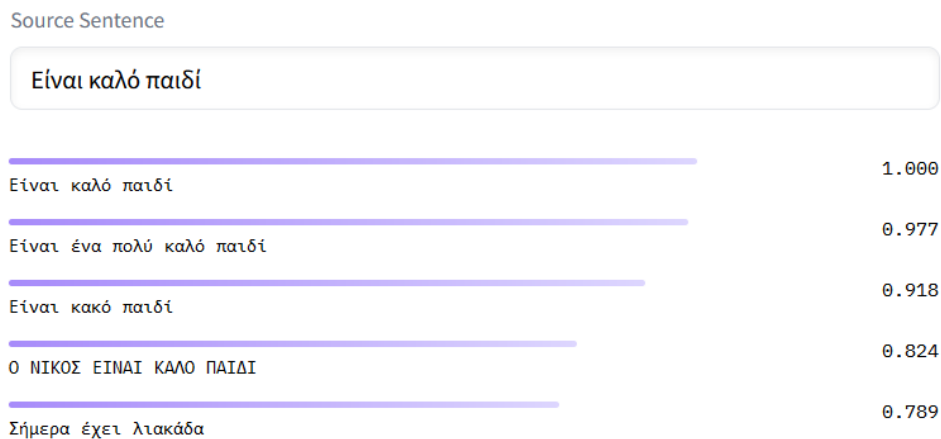
Για λόγους σύγκρισης, παραθέτουμε παρακάτω την αντίστοιχη εκτέλεση για το μοντέλο *sentence-transformers/distiluse-base-multilingual-cased-v2*, το οποίο είναι μεν πολύγλωσσο αλλά ξεχωρίζει μικρά και κεφαλαία.



Εικόνα 7: Σύγκριση προτάσεων με το μοντέλο *sentence-transformers/distiluse-base-multilingual-cased-v2*

Στο παραπάνω παρατηρούμε πως η πρόταση «Ο ΝΙΚΟΣ ΕΙΝΑΙ ΚΑΛΟ ΠΑΙΔΙ» εμφανίζεται να έχει μικρότερη ομοιότητα με την πρόταση «Είναι καλό παιδί» σε σχέση με την πρόταση «Σήμερα έχει λιακάδα», κάτι που είναι σημασιολογικά άτοπο. Αυτό συμβαίνει γιατί το μοντέλο αυτό θεωρεί διαφορετικά τα μικρά και κεφαλαία γράμματα.

Ακόμα, παραθέτουμε την ίδια εκτέλεση για το μοντέλο sentence-transformers/all-MiniLM-L6-v2, το οποίο είναι την τρέχουσα χρονική στιγμή το δημοφιλέστερο μοντέλο ομοιότητας προτάσεων στην πλατφόρμα HuggingFace και έχει εκπαιδευτεί μόνο στα αγγλικά.



Εικόνα 8: Σύγκριση προτάσεων με το μοντέλο sentence-transformers/all-MiniLM-L6-v2

Στο παραπάνω παρατηρούμε πως η πρόταση «Ο ΝΙΚΟΣ ΕΙΝΑΙ ΚΑΛΟ ΠΑΙΔΙ» εμφανίζεται να έχει μικρότερη ομοιότητα με την πρόταση «Είναι καλό παιδί» σε σχέση με την πρόταση «Είναι κακό παιδί», κάτι που είναι σημασιολογικά άτοπο. Ενδιαφέρον παρουσιάζει και το γεγονός πως αν γράψουμε αυτή την πρόταση με μικρά, θα εμφανίζεται πως παρουσιάζει ακόμα μικρότερη ομοιότητα, κάτι που φυσικά δεν ισχύει. Γίνεται αισθητό, συνεπώς, πως έχει πολύ μεγάλη σημασία η γλώσσα πάνω στην οποία έχει εκπαιδευτεί το μοντέλο που θα χρησιμοποιήσουμε για τις ενσωματώσεις.

Για περισσότερες πληροφορίες σχετικά με τον υπολογισμό της ομοιότητας μεταξύ προτάσεων ο αναγνώστης παραπέμπεται στο [63].

Η αρχικοποίηση του μοντέλου ενσωματώσεων γίνεται προγραμματιστικά με χρήση της κλάσης HuggingFaceEmbeddings, ως εξής:

```
from langchain.embeddings import HuggingFaceEmbeddings
modelPath = "sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2"
model_kwargs = {'device':'cpu'}
```

```
embeddings = HuggingFaceEmbeddings(
    model_name=modelPath,
    model_kwargs=model_kwargs
)
```

Αξίζει να σημειωθεί πως με τη ρύθμιση `model_kwargs = {'device': 'cpu'}` δίνουμε οδηγία στο μοντέλο να εκτελεστεί στον επεξεργαστή της συσκευής μας.

8.6 Εισαγωγή των ενσωματώσεων στη βάση δεδομένων

Τα δεδομένα μας, έχοντας πλέον κατακερματιστεί εισάγονται σε μία διανυσματική βάση δεδομένων με τη μορφή ενσωματώσεων (διανυσμάτων). Επιλέχθηκε η **Pinecone** [64], μία δημοφιλής διανυσματική βάση δεδομένων η οποία παρέχεται ως υπηρεσία στο cloud, υποστηρίζεται άμεσα από το LangChain και μπορούμε να τη χρησιμοποιήσουμε χωρίς χρέωση για εφαρμογές χαμηλών απαιτήσεων.

Για να χρησιμοποιήσουμε την Pinecone θα χρειαστεί να δημιουργήσουμε λογαριασμό και να εκδώσουμε κλειδί API (API key).

Αρχικοποιούμε τον πελάτη (client) που θα χρησιμοποιείται για την επικοινωνία της εφαρμογής μας με την Pinecone, δίνοντας το API key μας ως εξής:

```
from pinecone import Pinecone
pc = Pinecone(
    api_key= os.getenv('PINECONE_API_KEY')
)
```

Στη συνέχεια δημιουργούμε το ευρετήριο (index) στο οποίο θα αποθηκευτούν τα δεδομένα μας.

```
index_name = "energy-rag"
pc.create_index(name=index_name, metric="cosine", dimension=384,
spec=ServerlessSpec(
    cloud="aws",
    region="us-east-1")
```

Είναι σημαντικό να επισημάνουμε πως κατά τη δημιουργία του ευρετηρίου, εκτός από το όνομά του, καθορίζουμε και τον τρόπο με τον οποίο θα συγκρίνονται τα διανύσματα καθώς

και τη διάστασή τους. Για τον τρόπο σύγκρισης των διανυσμάτων έχουμε επιλέξει τον cosine, ο οποίος είναι και ο προεπιλεγμένος. Αυτό σημαίνει πως για την σύγκριση των διανυσμάτων υπολογίζεται το συνημίτονό τους. Έτσι, όσο πιο μικρή είναι η γωνία μεταξύ των διανυσμάτων, δηλαδή τα διανύσματα παρουσιάζουν μεγαλύτερη ομοιότητα, τόσο πιο πολύ το συνημίτονό τους πλησιάζει το ένα. Αυτό εξηγεί και τα αποτελέσματα που παρουσιάστηκαν στην ενότητα 8.6. Σχετικά με τη διάσταση των διανυσμάτων, αυτή ορίζεται σε 384, καθώς τέτοιας διάστασης διανύσματα παράγει το μοντέλο ενσωματώσεων που επιλέξαμε παραπάνω.

Αφού δημιουργηθεί το ευρετήριο εισάγονται σε αυτό τα δεδομένα μας υπό μορφή διανυσμάτων κάνοντας χρήση της κλάσης Pinecone του Langchain, ως εξής:

```
from langchain.vectorstores import Pinecone as PineconeVectorStore
docsearch = PineconeVectorStore.from_documents(docs, embeddings,
index_name=index_name)
```

Πλέον, οποτεδήποτε χρειαστούμε τα δεδομένα μας και τα αντίστοιχα διανύσματά τους μπορούμε να τα φορτώσουμε απευθείας από την Pinecone ως εξής:

```
docsearch = PineconeVectorStore.from_existing_index(index_name, embeddings)
```

8.7 Meltemi - Το πρώτο ανοιχτό Μεγάλο Γλωσσικό Μοντέλο για τα Ελληνικά

Σε πρώτο στάδιο πραγματοποιήθηκαν δοκιμές με διάφορα δημοφιλή μοντέλα που είναι διαθέσιμα στο HuggingFace, όπως είναι το Llama της Meta, το Gemma της Google, το Phi της Microsoft και το Mistral της Mistral AI. Κανένα από αυτά δεν παρουσίασε ικανοποιητική απόδοση στη χρήση των ελληνικών, καθώς κανένα δεν έχει εκπαιδευτεί επαρκώς πάνω σε ελληνικά δεδομένα.

Μετά από αναζήτηση στο περιβάλλον του HuggingFace για μοντέλα που υποστηρίζουν ελληνικά βρέθηκε το **Meltemi** [65], [66], το οποίο και τελικά επιλέχθηκε για την ανάπτυξη του ψηφιακού βοηθού. Το Meltemi αναπτύχθηκε από το Ινστιτούτο Επεξεργασίας του Λόγου του Ερευνητικού Κέντρου Αθηνά και αποτελεί **το πρώτο ανοιχτό Μεγάλο Γλωσσικό Μοντέλο για την ελληνική γλώσσα**.

Το Meltemi βασίζεται στο μοντέλο Mistral-7B και έχει εκπαιδευτεί σε ένα μεγάλο σώμα ελληνικών κειμένων υψηλής ποιότητας. Είναι ένα LLM 7 δισεκατομμυρίων παραμέτρων με

8192 μέγιστο μήκος μονάδων κειμένου (tokens) που μπορεί να επεξεργαστεί (context length). Διατίθεται σε δύο εκδοχές, το θεμελιώδες μοντέλο Meltemi-7B καθώς και η παραλλαγή του, Meltemi-Instruct-7B, η οποία έχει εκπαιδευτεί περαιτέρω σε ερωτήσεις και απαντήσεις και είναι περισσότερο κατάλληλη για χρήση σε chatbots. Και οι δύο παραλλαγές βρίσκονται ήδη στην έκδοση 1.5.

Για την εφαρμογή μας θα χρησιμοποιήσουμε την παραλλαγή **ilsp/Meltemi-7B-Instruct-v1.5** [67]. Το μοντέλο αυτό, βάσει των δοκιμών που πραγματοποίησε το Ερευνητικό Κέντρο Αθηνά, παρουσιάζει μέση βελτίωση στα ελληνικά test sets κατά 7,8% σε σχέση με το ilsp/Meltemi-7B-Instruct-v1 και κατά 18,1% σε σχέση με το Mistral 7B.

8.7.1 Χρήση του μοντέλου

Το μοντέλο μπορεί κανείς να το κατεβάσει και να το εκτελέσει τοπικά στον υπολογιστή του από τη σελίδα του στο HuggingFace [67]. Ωστόσο, στις προσπάθειες που έγιναν η απόδοση ήταν μη ικανοποιητική και δημιουργούσε εμπόδιο στην πραγματοποίηση αρχικών δοκιμών. Οι απαιτήσεις συστήματος για την τοπική εκτέλεση του μοντέλου ήταν πολύ μεγαλύτερες από αυτές που είχαμε στη διάθεσή μας.

Το HuggingFace μας παρέχει τη δυνατότητα να εκτελέσουμε μοντέλα απομακρυσμένα σε δικές του υποδομές με χαμηλό έως και μηδενικό κόστος και να αλληλεπιδράσουμε με αυτά με έναν πολύ εύκολο τρόπο, χάρη στα Inference Endpoints [68]. Μέσω του περιβάλλοντος του HuggingFace, μπορούμε να επιλέξουμε το μοντέλο και τις προδιαγραφές συστήματος στο οποίο αυτό θα εκτελείται και στη συνέχεια να του στέλνουμε ερωτήματα και να λαμβάνουμε απαντήσεις μέσω ενός REST API.

Το Meltemi-7B-Instruct-v1.5, όντας σχεδιασμένο να δέχεται οδηγίες και να απαντά σε ερωτήσεις, έχει συγκεκριμένη μορφή (template) στην οποία αναμένει το ερώτημα, η οποία είναι η παρακάτω:

```
<|system|>
{Οδηγίες προς το μοντέλο}</s>
<|user|>
{Ερώτημα χρήστη}</s>
<|assistant|>
{Απάντηση μοντέλου}
```

Παρακάτω φαίνεται ένα παράδειγμα ερωτήματος προς το Meltemi-7B-Instruct-v1.5, μέσω του γραφικού περιβάλλοντος του HuggingFace.

Text Generation

```
<|system|>
Είσαι ο Μελτέμης, ένας σύγχρονος ποιητής.</s>
<|user|>
Γράψε μου ένα μικρό ποίημα για το καλοκαιρινό μελτέμι</s>
<|assistant|>
Στο νησί, ένα καλοκαιρινό μελτέμι φυσά,
Κουνώντας τα δέντρα, ψιθυρίζοντας στα φύλλα,
Μια απαλή υπενθύμιση της θάλασσας κοντά,
Όπου τα κύματα χορεύουν, σε έναν αιώνιο χορό.
```

Εικόνα 9: Ερώτηση και απάντηση χρησιμοποιώντας το μοντέλο *ilsp/Meltemi-7B-Instruct-v1.5*

Για να καλέσουμε απομακρυσμένα το μοντέλο χρησιμοποιούμε το παρακάτω κομμάτι κώδικα (αποκρύπτεται το πλήρες URL):

```
import requests

API_URL = "https://*****.us-east-1.aws.endpoints.huggingface.cloud"
headers = {
    "Accept" : "application/json",
    "Authorization": "Bearer hf_XXXXX",
    "Content-Type": "application/json"
}

def query(payload):
    response = requests.post(API_URL, headers=headers, json=payload)
    return response.json()

prompt = """
<|system|>
{Οδηγίες προς το μοντέλο}</s>
<|user|>
{Ερώτημα χρήστη}</s>
<|assistant|>
"""
```



```
input = {
    "inputs": prompt,
    "parameters": {
        "max_new_tokens": 150
    }
}

output = query(input)
```

Με την παράμετρο `max_new_tokens` μπορούμε να ορίσουμε το μέγιστο αριθμό λέξεων που θα παράγει το μοντέλο στην απάντησή του.

8.7.2 Διαμορφώνοντας το επαυξημένο ερώτημα

Ακολουθώντας τις καλές πρακτικές για την δημιουργία ενός ερωτήματος που αναπτύχθηκαν στην ενότητα 6.5.3, σχεδιάστηκε το παρακάτω επαυξημένο ερώτημα το οποίο και εν τέλει χρησιμοποιήθηκε στην εφαρμογή.

```
prompt = """
<|system|>
Είσαι ένας ευγενικός και εξυπηρετικός βοηθός που σκοπός σου είναι να απαντάς στις
ερωτήσεις του χρήστη σχετικά με τα προγράμματα ενεργειακής αναβάθμισης που
δίνονται στο παρακάτω περιεχόμενο.
Θα πρέπει να χρησιμοποιείς μόνο το περιεχόμενο που σου δίνεται παρακάτω για να
απαντήσεις στην ερώτηση.
Αν δεν γνωρίζεις την απάντηση βάσει του παρακάτω περιεχομένου θα πρέπει να απαντάς
με "Δεν γνωρίζω την απάντηση για αυτό".
Η απάντησή σου πρέπει πάντα να είναι σύντομη και περιεκτική.
```

```
Περιεχόμενο: {context}
</s>
<|user|>
{question}</s>
<|assistant|>
"""
```

Όπως θα δούμε στη συνέχεια, σε κάθε ερώτηση που θέτει ο χρήστης προς το σύστημα, στο παραπάνω ερώτημα στο πεδίο {question} τοποθετείται η ερώτηση του χρήστη ενώ στο {context} τα σχετικότερα κομμάτια πληροφορίας που ανακτήθηκαν από τη βάση δεδομένων, βάσει της σχετικότητάς τους με το ερώτημα του χρήστη.

Το επαυξημένο ερώτημα στο οποίο καταλήξαμε είναι πιστό στις καλές πρακτικές, καθώς:

- Δίνει ρόλο (persona) στο μοντέλο («Είσαι ένας ευγενικός και εξυπηρετικός βοηθός...»)
- Περιέχει ξεκάθαρες και περιγραφικές οδηγίες, χωρίς να δίνει υπερβολική λεπτομέρεια.
- Διαχωρίζονται οι προτάσεις σε νέα γραμμή, ενώ υπάρχει κενή γραμμή πριν το περιεχόμενο.
- Ορίζεται η μορφή της εξόδου («... να είναι σύντομη και περιεκτική»).
- Περιγράφεται το τι πρέπει να κάνει το μοντέλο σε περίπτωση που δεν βρει απάντηση στο περιεχόμενο και όχι το τι δεν πρέπει.

8.8 Κατασκευάζοντας την αλυσίδα RAG

Η αλυσίδα RAG αποτελεί το τελευταίο στάδιο, όπου το ερώτημα του χρήστη μαζί με τα κομμάτια πληροφορίας που ανακτώνται από τη βάση συντίθεται στο επαυξημένο ερώτημα.

Χάρη στο LangChain αυτό μπορεί να γίνει πολύ απλά ως εξής:

Αρχικά ορίζεται το πρότυπο (template) του ερωτήματος προς το LLM, ενσωματώνοντας τις μεταβλητές context (τα κομμάτια που θα ανακτηθούν) και question (το ερώτημα του χρήστη):

```
prompt = PromptTemplate(  
    template=template,  
    input_variables=["context", "question"]  
)
```

Πλέον είμαστε σε θέση να κατασκευάσουμε την αλυσίδα RAG ως εξής:

```
rag_chain = (  
    {"context": docsearch.as_retriever(), "question": RunnablePassthrough()  
    | prompt  
    | query  
)
```

Η αλυσίδα αυτή περνάει στο πρότυπο του ερωτήματος τον ανακτητή εγγράφων (retriever) από την Pinecone καθώς και το ερώτημα του χρήστη. Έχοντας πλέον διαμορφωθεί, το τελικό επαυξημένο ερώτημα περνάει στη συνάρτηση query, η οποία το αποστέλλει στο μοντέλο και μας επιστρέφει το αποτέλεσμα.

Στο σημείο αυτό είναι χρήσιμο να αναφερθεί πως μπορούμε να ελέγξουμε τον αριθμό των κομματιών πληροφορίας που θα μας επιστρέψει ο ανακτητής ρυθμίζοντας την παράμετρο k ως εξής:

```
docsearch.as_retriever(search_kwargs={"k": X}),
```

όπου X ο αριθμός των σχετικότερων με το ερώτημα του χρήστη κομματιών πληροφορίας (chunks) που θα ανακτηθούν.

Για παράδειγμα, ορίζοντας το k ίσο με 8 ζητάμε από τον ανακτητή να μας φέρει τα οκτώ πιο σχετικά με το ερώτημα του χρήστη κομμάτια πληροφορίας, τα οποία και θα ενσωματωθούν στο επαυξημένο ερώτημα. Αν δεν ρυθμίσουμε τη συγκεκριμένη παράμετρο, τότε από προεπιλογή ανακτώνται τα 4 σχετικότερα κομμάτια.

Μπορούμε πλέον να χρησιμοποιήσουμε το RAG σύστημα που κατασκευάσαμε ως εξής:

```
input = input("Είμαι ο ψηφιακός βοηθός ενεργειακής μετάβασης. Τι θα ήθελες να μάθεις σήμερα; ")
result = rag_chain.invoke(input)
print(result)
```

Επισημαίνεται πως το μοντέλο θα γυρίσει όλη την πληροφορία (είσοδος και έξοδος) βάσει του προτύπου που χρησιμοποιεί (βλέπε ενότητα 8.8.1) και μάλιστα στη μορφή

```
[{'generated_text': <Απόκριση μοντέλου>}]
```

Μπορούμε να απομονώσουμε την απάντηση ως εξής:

```
answer = result[0]['generated_text'].split('<|assistant|>')[-1].strip()
```

Παρακάτω φαίνεται η εκτέλεση της εφαρμογής με ερώτημα: «Μπορώ να αγοράσω τηλεόραση με το ανακυκλώνω αλλάζω συσκευή;». Διακρίνονται τα ανακτημένα κομμάτια

πληροφορίας με σειρά σχετικότητας ως προς το ερώτημά μας. Για τη συγκεκριμένη εκτέλεση έχουμε θέσει `search_kwargs={'k': 2}` στον ανακτητή, ώστε να ανακτηθούν μόνο τα δύο σχετικότερα. Παρατηρούμε πως το πρώτο κομμάτι πληροφορίας σε σειρά αφορά το πρόγραμμα Ανακυκλώνω-Αλλάζω Συσκευή, καθώς είναι το σχετικότερο με το ερώτημα που τέθηκε.

```
Είμαι ο ψηφιακός βοηθός ενεργειακής μετάβασης. Τι θα ήθελες να μάθεις σήμερα;
> Μπορώ να αγοράσω τηλεόραση με το ανακυκλώνω αλλάζω συσκευή;
[{'generated_text': '<|system|>\nΕίσαι ένας ευγενικός και εξυπηρετικός βοηθός που σκοπός σου είναι να απαντάς στις ερωτήσεις του χρήστη σχετικά με τα προγράμματα ενεργειακής αναβάθμισης που δίνονται στο παρακάτω περιεχόμενο.\nΘα πρέπει να χρησιμοποιείς μόνο το περιεχόμενο που σου δίνεται παρακάτω για να απαντήσεις στην ερώτηση.\nΑν δεν γνωρίζεις την απάντηση βάσει του παρακάτω περιεχομένου θα πρέπει να απαντάς με "Δεν γνωρίζω την απάντηση για αυτό".\nΗ απάντησή σου πρέπει πάντα να είναι σύντομη και περιεκτική.\n\nΠεριεχόμενο: [Document(metadata={'row': 0.0, 'source': './energy_programs.csv'}, page_content='Όνομα Δράσης: Ανακυκλώνω - Αλλάζω Συσκευή\nΠάροχος: Ελληνικό δημόσιο\nΚατηγορία: Αντικατάσταση συσκευών\nΠεριγραφή: Το Πρόγραμμα «ΑΝΑΚΥΚΛΩΝΩ-ΑΛΛΑΖΩ ΣΥΣΚΕΥΗ» επιχορηγεί νοικοκυριά για την αντικατάσταση παλαιών ηλεκτρικών συσκευών με νέες, φιλικές προς το περιβάλλον και ενεργειακά πιο αποδοτικές.\n\nΣτο Πρόγραμμα συμμετέχουν μόνο οι ακόλουθες κατηγορίες συσκευών:\n\nΚλιματιστικά\nΨυγεία\nΚαταψύκτες\n\nΗ επιχορήγηση παρέχεται μέσω επιταγών (vouchers) του Προγράμματος, οι οποίες μπορούν να χρησιμοποιηθούν κατά την αγορά νέας συσκευής σε καταστήματα λιανικής πώλησης ηλεκτρικών συσκευών και να καλύψουν ένα μέρος της συνολικής δαπάνης.\n\nΓια κάθε αγορά νέας συσκευής με επιδότηση, δημιουργείται παράλληλα η υποχρέωση παράδοσης μίας παλιάς προς ανακύκλωση που ανήκει στην ίδια κατηγορία.\n\nΓια περισσότερες πληροφορίες σχετικά με το Πρόγραμμα, επιλέξτε την κατάλληλη ενότητα που αντιστοιχεί στο ρόλο σας.\n\nΣε ισχύ: Όχι\nΑπαιτείται ανακύκλωση: Ναι\nΠερισσότερες πληροφορίες: https://allazosyskevi.gov.gr/'), Document(metadata={'row': 2.0, 'source': './energy_programs.csv'}, page_content='Προσοχή:\n\nΓίνονται δεκτοί προς αντικατάσταση παλαιοί Ηλεκτρικοί Θερμοσίφωνες χωρητικότητας τουλάχιστον 40 λίτρων.\n\nΔεν είναι δυνατή η επιδότηση νέου Ηλιακού Θερμοσίφωνα χωρίς την παράλληλη παράδοση του παλαιού Ηλεκτρικού Θερμοσίφωνα.\n\nΗ επιχορήγηση παρέχεται μέσω επιταγών (vouchers) του Προγράμματος, οι οποίες μπορούν να χρησιμοποιηθούν σε εγκεκριμένους Προμηθευτές του Προγράμματος και να καλύψουν μέρος της συνολικής δαπάνης των σχετικών προϊόντων/υπηρεσιών.\n\nΓια περισσότερες πληροφορίες, επιλέξτε την κατάλληλη ενότητα που αντιστοιχεί στο ρόλο σας.\n\nΣε ισχύ: Όχι\nΑπαιτείται ανακύκλωση: Ναι\nΠερισσότερες πληροφορίες: https://allazothermosifona.gov.gr/')]\n</s>\n<|user|>\nΜπορώ να αγοράσω τηλεόραση με το ανακυκλώνω αλλάζω συσκευή;\n</s>\n<|assistant|>\n\nΌχι, δεν μπορείτε να αγοράσετε τηλεόραση με το ανακυκλώνω αλλάζω συσκευή. Το πρόγραμμα επιχορηγεί νοικοκυριά για την αντικατάσταση παλαιών ηλεκτρικών συσκευών με νέες, φιλικές προς το περιβάλλον και ενεργειακά πιο αποδοτικές. Οι επιλέξιμες συσκευές είναι κλιματιστικά, ψυγεία και καταψύκτες.'}]
```

Όχι, δεν μπορείτε να αγοράσετε τηλεόραση με το ανακυκλώνω αλλάζω συσκευή. Το πρόγραμμα επιχορηγεί νοικοκυριά για την αντικατάσταση παλαιών ηλεκτρικών συσκευών με νέες, φιλικές προς το περιβάλλον και ενεργειακά πιο αποδοτικές. Οι επιλέξιμες συσκευές είναι κλιματιστικά, ψυγεία και καταψύκτες.

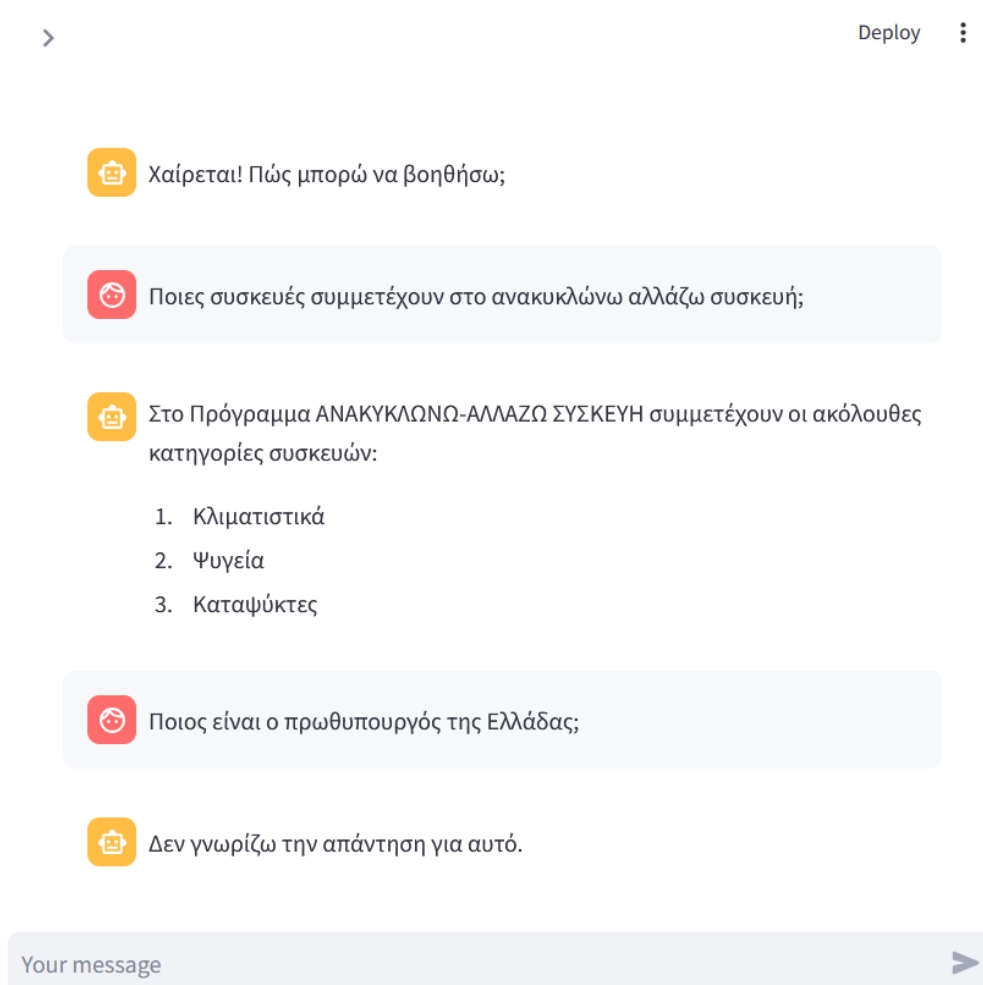
Εικόνα 10: Παράδειγμα εκτέλεσης της εφαρμογής

Ο πλήρης κώδικας της εφαρμογής είναι διαθέσιμος στο παράρτημα.

8.9 Δίνοντας στην εφαρμογή μας ένα γραφικό περιβάλλον

Η εφαρμογή λειτουργεί σωστά, ωστόσο δεν είναι ιδιαίτερα βολική η χρήση της μέσω του τερματικού. Στην ενότητα αυτή, θα χρησιμοποιήσουμε το **Streamlit** [69], ένα εργαλείο που βοηθά στη γρήγορη δημιουργία ιστοσελίδων για την παρουσίαση δεδομένων και την αλληλεπίδραση με αυτά. Το Streamlit είναι γραμμένο σε Python και ενσωματώνεται άμεσα στην εφαρμογή μας, χωρίς να χρειαστεί να ασχοληθούμε με σχεδιασμό ιστοσελίδων.

Παρακάτω φαίνεται ο ψηφιακός μας βοηθός, έχοντας πλέον αποκτήσει γραφικό περιβάλλον:



Εικόνα 11: Το γραφικό περιβάλλον του ψηφιακού βοηθού

Ο κώδικας για την ενσωμάτωση του Streamlit είναι διαθέσιμος στο παράρτημα.

9 Αποτελέσματα δοκιμών

9.1 Παρουσίαση και σχολιασμός αποτελεσμάτων

Προκειμένου να αξιολογήσουμε τον ψηφιακό οδηγό που κατασκευάσαμε, θέσαμε σε αυτόν ορισμένα ερωτήματα και λάβαμε τις αντίστοιχες απαντήσεις. Χρησιμοποιήθηκε η μεθοδολογία που παρουσιάστηκε στην ενότητα 6.4. Συγκεκριμένα, για κάθε απάντηση σημειώσαμε αν υπάρχει εσωτερική ή εξωτερική παραίτηση και κατόπιν αν η απάντηση είναι «πιστή» και «πραγματική». Επειδή ορισμένες φορές παρατηρήθηκε πως ναι μεν μπορεί μία απάντηση να ικανοποιούσε τα παραπάνω αλλά να παρέλειπε πληροφορίες, προστέθηκε η ιδιότητα «πλήρης», όπως περιγράφηκε στην παραπάνω ενότητα. Για να χαρακτηριστεί μία απάντηση ως πλήρης, θα πρέπει το μοντέλο βάσει των δεδομένων πληροφοριών να απαντά με όλη την πληροφορία που θα περίμενε κανείς.

Τα αποτελέσματα της αξιολόγησης παρουσιάζονται στον παρακάτω πίνακα. Σημειώνεται ότι με αστεράκι (*) στην πρώτη στήλη σημειώνονται οι απαντήσεις που κρίθηκαν μη ικανοποιητικές.

A/A	Ερώτηση	Απάντηση	Παραισθήσεις Εσωτερικές Εξωτερικές	Ιδιότητες Πιστή Πραγματική Πλήρης	Σχόλια
1	Ποιες συσκευές συμμετέχουν στο ανακυκλώνω αλλάζω συσκευή;	Το ανακυκλώνω αλλάζω συσκευή περιλαμβάνει τις ακόλουθες κατηγορίες συσκευών: 1. Κλιματιστικά 2. Ψυγεία 3. Καταψύκτες	Όχι Όχι	Ναι Ναι Ναι	Δίνονται ακριβώς οι πληροφορίες που ζητούνται.
2 *	ποιες είναι οι διαθέσιμες δράσεις ηλεκτροκίνησης;	Το Πρόγραμμα Κινούμαι Ηλεκτρικά Γ' Κύκλος είναι ένα πρόγραμμα επιδότησης αγοράς ηλεκτρικών οχημάτων και φορτιστών.	Όχι Όχι	Ναι Ναι Όχι	Το περιεχόμενο της απάντησης είναι σωστό βάσει του δεδομένου κειμένου, όμως έχει παραληφθεί αναφορά στο πρόγραμμα «Πράσινα Ταξί».

3	υπάρχει πρόγραμμα ηλεκτροκίνησης για ταξί;	Ναι, υπάρχει πρόγραμμα ηλεκτροκίνησης για ταξί. Το πρόγραμμα ονομάζεται "Πράσινα Ταξί" και είναι ένα πρόγραμμα επιδότησης επιβατηγών οχημάτων δημόσιας χρήσεως (Ε.Δ.Χ.). Το ύψος των ενισχύσεων καθορίζεται ως εξής: - Για αμιγώς ηλεκτρικά Ε.Δ.Χ. οχήματα ΤΑΞΙ (BEV), το οικολογικό bonus ανέρχεται σε ποσοστό 40% επί της Λ.Τ.Π.Φ. με μέγιστο ποσό τα 17.500 €. - Η υποχρεωτική απόσυρση Ε.Δ.Χ. οχήματος ΤΑΞΙ επιβραβεύεται με ποσό ίσο με 5.000 €. - Η αγορά "έξυπνου" φορτιστή επιδοτείται με ποσό 500 €. Για περισσότερες πληροφορίες, μπορείτε να επισκεφθείτε την ιστοσελίδα https://prasinataxi.gov.gr/.	Όχι Όχι	Ναι Ναι Ναι	Αν και η απάντηση θα αρκούσε να είναι «Ναι υπάρχει», εδώ το μοντέλο έδωσε τις βασικές πληροφορίες του προγράμματος και παρέπεμψε στην επίσημη ιστοσελίδα.
4	Ποιος είναι ο πρωθυπουργός της Ελλάδας;	Δεν γνωρίζω την απάντηση για αυτό.	Όχι Όχι	Ναι Ναι Ναι	Το μοντέλο απάντησε ορθά βάσει των οδηγιών και πληροφοριών που του έχουν δοθεί.
5	Θα βρέξει αύριο στην Αττική	Δεν γνωρίζω την απάντηση για αυτό.	Όχι Όχι	Ναι Ναι Ναι	Το μοντέλο απάντησε ορθά βάσει των οδηγιών και πληροφοριών που του έχουν δοθεί.

6	Ο θερμοσίφωνάς μου είναι 80 λίτρα, μπορώ να μπω στο πρόγραμμα αλλαγής θερμοσίφωνα;	Ναι, μπορείτε να μπείτε στο πρόγραμμα αλλαγής θερμοσίφωνα. Το πρόγραμμα επιδοτεί θερμοσίφωνες χωρητικότητας τουλάχιστον 40 λίτρων.	Όχι Όχι	Ναι Ναι Ναι	Στην απάντηση αυτή φαίνονται οι ικανότητες σύγκρισης αριθμών του μοντέλου. Το κείμενο αναφοράς είναι: «Γίνονται δεκτοί προς αντικατάσταση παλιοί Ηλεκτρικοί Θερμοσίφωνες χωρητικότητας τουλάχιστον 40 λίτρων.»
7	Υπάρχει πρόγραμμα για αντλίες θερμότητας; Ποιος το παρέχει;	Ναι, υπάρχει πρόγραμμα για αντλίες θερμότητας. Το πρόγραμμα παρέχεται από την Acme Corporation.	Όχι Όχι	Ναι Ναι Ναι	Εδώ ελέγχουμε τη δυνατότητα του μοντέλου να επεξεργαστεί τη στήλη «Πάροχος» των δεδομένων.
8	Τι περιλαμβάνει το πρόγραμμα για αντλίες θερμότητας της Acme;	Το πρόγραμμα για αντλίες θερμότητας της Acme περιλαμβάνει δωρεάν εγκατάσταση αντλίας θερμότητας με την αγορά οποιασδήποτε αντλίας θερμότητας.	Όχι Όχι	Ναι Ναι Ναι	Δίνεται πλήρης απάντηση σε σχέση με το ερώτημα που τίθεται.
9	Υπάρχει κάποιο πρόγραμμα για αγορά κλιματιστικού και αν ναι, που	Ναι, υπάρχει ένα πρόγραμμα για αγορά κλιματιστικού. Το πρόγραμμα ονομάζεται Δωροεπιταγή ανακύκλωσης για αντικατάσταση κλιματιστικού και προσφέρεται από την Acme Corporation. Για να μάθετε περισσότερα,	Όχι Όχι	Ναι Ναι Ναι	Εδώ ελέγχουμε τη δυνατότητα του μοντέλου να επεξεργαστεί τη στήλη «Περισσότερες

	μπορώ να μάθω περισσότερα;	μπορείτε να επισκεφθείτε την ιστοσελίδα https://www.example.com/air-condition .			πληροφορίες» των δεδομένων.
10	Θέλω να εγκαταστήσω θέρμανση φυσικού αερίου. Ποιος πάροχος είναι υπεύθυνος για το πρόγραμμα; Πρέπει η περιοχή μου να έχει δίκτυο;	Το πρόγραμμα για την εγκατάσταση θέρμανσης φυσικού αερίου διαχειρίζεται η ΕΔΑ Αττικής. Το πρόγραμμα ισχύει μόνο για συμβάσεις σύνδεσης φυσικού αερίου για κατοικίες που βρίσκονται επί ενεργού δικτύου φυσικού αερίου χαμηλής πίεσης (4bar).	Όχι Όχι	Ναι Ναι Ναι	Ελέγχεται η απάντηση του μοντέλου σε πολλαπλά ερωτήματα σχετικά με μία δράση αναβάθμισης. Το μοντέλο εντοπίζει σωστά τα κομμάτια κειμένου στα οποία βρίσκεται η ζητούμενη πληροφορία.
11 *	Πώς ωφελούμαι από το πρόγραμμα για φωτοβολταϊκα στη στέγη και υπάρχει προϋπόθεση;	Το πρόγραμμα για φωτοβολταϊκα στη στέγη επιχορηγεί νοικοκυριά για την εγκατάσταση φωτοβολταϊκών συστημάτων με σύστημα αποθήκευσης για αυτοκατανάλωση με εφαρμογή ενεργειακού συμψηφισμού. Η προϋπόθεση είναι ότι ο ωφελοούμενος δεν πρέπει να έχει ξεκινήσει καμία εργασία ή παραγγελία για τον φωτοβολταϊκό σταθμό.	Ναι Όχι	Όχι Ναι Όχι	Στην πρώτη πρόταση το μοντέλο έχει παραγάγει εσωτερική παραίσθηση, παραποιώντας το κείμενο του προγράμματος «Φωτοβολταϊκά στη Στέγη». Στη δεύτερη πρόταση έχει χρησιμοποιήσει κείμενο που αφορά το πρόγραμμα «Φωτοβολταϊκά στο χωράφι».

12	Τι ποσοστό επιδότησης μπορώ να πάρω με το Κινούμαι ηλεκτρικά για δίκυκλο L1e;	Το Κινούμαι Ηλεκτρικά προσφέρει επιδότηση 30% της τιμής προ ΦΠΑ για την αγορά ηλεκτρικών δικύκλων κατηγορίας L1e έως L4e, με μέγιστο ποσό επιδότησης 1.300€.	Όχι Όχι	Ναι Ναι Ναι	Ορθή και πλήρης απάντηση.
13	ποιες δαπάνες επιδοτεί το πρόγραμμα εξοικονομώ 2023;	Οι επιλέξιμες δαπάνες που μπορούν να επιδοτηθούν από το πρόγραμμα Εξοικονομώ 2023 είναι: 1. Αντικατάσταση κουφωμάτων 2. Τοποθέτηση ή αναβάθμιση θερμομόνωσης 3. Αναβάθμιση συστήματος θέρμανσης/ψύξης 4. Σύστημα ζεστού νερού με χρήση Ανανεώσιμων Πηγών Ενέργειας 5. Άλλες παρεμβάσεις (σύστημα διαχείρισης smart home)	Όχι Όχι	Ναι Ναι Ναι	Μία σχετικά εύκολη ερώτηση στην οποία το μοντέλο απαντά με επιτυχία.
14 *	τι ποσό δίνεται για αυτόνομη θέρμανση φυσικού αερίου;	Το επιδοτούμενο ποσό για αυτόνομη θέρμανση φυσικού αερίου είναι κλιμακούμενο και ορίζεται από την κατηγορία σύνδεσης και την ισχύ του λέβητα εγκατάστασης σύμφωνα με τον παρακάτω πίνακα: Κατηγορία Σύνδεσης Ισχύς Λέβητα (kW) Επιδοτούμενο Ποσό --- --- --- Κεντρική Θέρμανση ≤ 35 3.000€ Κεντρική Θέρμανση > 35 4.000€ Αυτόνομη Θέρμανση ≤ 35 2.000€	Ναι Ναι	Όχι Ναι Όχι	Εδώ το μοντέλο ενσωματώσεων δεν έφερε το κατάλληλο κομμάτι πληροφορίας που περιέχει τον πίνακα (τα κομμάτια έχουν χωριστεί ακριβώς πριν τον πίνακα επιδοτήσεων) με αποτέλεσμα το LLM να παραγάγει από μόνο του έναν

		Αυτόνομη Θέρμανση > 35 3.000€ Αυτά τα ποσά είναι τα μέγιστα που μπορούν να επιδοτηθούν και μπορεί να διαφέρουν ανάλογα με τις συγκεκριμένες λεπτομέρειες της αίτησης.			πίνακα με αυθαίρετες τιμές. Η τελευταία πρόταση μετά τον πίνακα δεν υπάρχει στα δεδομένα κείμενα.
15 *	ποιες προϋποθέσεις έχει το πρόγραμμα ανακαινίζω για νέους;	Οι προϋποθέσεις για το πρόγραμμα "Εξοικονομώ - Ανακαινίζω για νέους" είναι: 1. Να είστε φυσικό πρόσωπο. 2. Να έχετε γεννηθεί από την 1η Ιανουαρίου 1984 έως και την 31η Δεκεμβρίου 2005. 3. Να έχετε εμπράγματο δικαίωμα στην κατοικία (πλήρη κυριότητα / επικαρπία / ψιλή κυριότητα). 4. Η κατοικία να είναι κύρια. 5. Η κατοικία να πληροί τα εισοδηματικά κριτήρια. Προϋποθέσεις επιλεξιμότητας κατοικίας: 1. Να είναι νόμιμη και να μην έχει χαρακτηριστεί κατεδαφιστέα. 2. Να είναι κύρια κατοικία.	Ναι Ναι	Όχι Όχι Όχι	Στο πρώτο σκέλος θα έπρεπε να αναφέρονται οι προϋποθέσεις που πρέπει να πληροί ο ενδιαφερόμενος. Εδώ το μοντέλο έχει προσθέσει και κάποιες προϋποθέσεις της κατοικίας, έχοντας μάλιστα «αναμείξει» την κατοικία με τα εισοδηματικά κριτήρια, παράγοντας μια πρόταση χωρίς λογική. Έτσι, έχουμε εσωτερική και

		3. Να ανήκει σε κατηγορία χαμηλότερη ή ίση της Γ βάσει του πρώτου πιστοποιητικού Ενεργειακής Απόδοσης.			εξωτερική παραίσθηση. Το δεύτερο σκέλος σχετικά με τις προϋποθέσεις της κατοικίας είναι ορθό.
16	το πρόγραμμα ανακυκλώνω αλλάζω συσκευή ισχύει ακόμα	Όχι, το πρόγραμμα "Ανακυκλώνω - Αλλάζω Συσκευή" δεν ισχύει πλέον.	Όχι Όχι	Ναι Ναι Ναι	Εδώ ελέγχουμε τη δυνατότητα του μοντέλου να επεξεργαστεί τη στήλη «Σε ισχύ» των δεδομένων.
18	Μπορώ ως ιδιώτης να επωφεληθώ από το φωτοβολταϊκά στο χωράφι;	Όχι, δεν μπορείτε να επωφεληθείτε από το φωτοβολταϊκά στο χωράφι ως ιδιώτης. Το πρόγραμμα απευθύνεται ειδικά σε επαγγελματίες αγρότες, αγρότες ειδικού καθεστώτος και αγροτικές επιχειρήσεις.	Όχι Όχι	Ναι Ναι Ναι	Το σύστημα εντοπίζει το σωστό κομμάτι πληροφορίας και απαντά ορθά,

19	Ενδιαφέρομαι για ανανεώσιμες πηγές ενέργειας. Ποια προγράμματα υπάρχουν διαθέσιμα;	Υπάρχουν αρκετά προγράμματα διαθέσιμα για ανανεώσιμες πηγές ενέργειας. Μερικά από αυτά είναι: 1. Φωτοβολταϊκά στη Στέγη: Αυτό το πρόγραμμα παρέχει ενίσχυση για την εγκατάσταση φωτοβολταϊκών συστημάτων με ή χωρίς σύστημα αποθήκευσης για αυτοκατανάλωση με εφαρμογή ενεργειακού συμψηφισμού. 2. Φωτοβολταϊκά στο χωράφι: Αυτό το πρόγραμμα επιχορηγεί τους αγρότες για την εγκατάσταση Φωτοβολταϊκών συστημάτων για αυτοκατανάλωση ενέργειας. 3. Αλλάζω συσκευή για τις επιχειρήσεις: Αυτό το πρόγραμμα επιχορηγεί τις επιχειρήσεις για την εγκατάσταση ενεργειακά αποδοτικού εξοπλισμού. 4. Ανακυκλώνω - Αλλάζω Θερμοσίφωνα: Αυτό το πρόγραμμα επιχορηγεί τα νοικοκυριά για την αντικατάσταση ενεργοβόρων Ηλεκτρικών Θερμοσιφώνων με νέους Ηλιακούς Θερμοσίφωνες. 5. Εξοικονομώ 2023: Αυτό το πρόγραμμα παρέχει επιδοτήσεις για παρεμβάσεις που εξοικονομούν	Όχι Όχι	Ναι Ναι Όχι	Εκ πρώτης όψεως θα περίμενε κανείς το μοντέλο να απαντήσει μόνο για τα προγράμματα στην κατηγορία των ΑΠΕ. Ωστόσο, ορθά το μοντέλο δίνει πληροφορίες και για άλλα προγράμματα που αφορούν ΑΠΕ, βάσει του κειμένου που του έχει δοθεί. Εδώ φαίνονται και οι ικανότητες του μοντέλου να συνοψίζει πληροφορία σχετικά με το κάθε πρόγραμμα. Η απάντηση μπορεί να μην είναι πλήρης (δίνονται μερικά προγράμματα), είναι όμως ικανοποιητική.
----	--	--	------------	-------------------	---

		ενέργεια και ενισχύουν την ενεργειακή αυτονομία της κατοικίας σας.			
20	Είμαι φυσικό πρόσωπο και θέλω να αγοράσω ηλεκτρικό αυτοκίνητο με επιδότηση. Αν το αυτοκίνητο κοστίζει 15.000 ευρώ, τι επιδότηση θα πάρω	Αν είστε φυσικό πρόσωπο και αγοράσετε ηλεκτρικό αυτοκίνητο με κόστος 15.000 ευρώ, θα λάβετε επιδότηση 30% της τιμής προ ΦΠΑ, με μέγιστο ποσό τις 9.000 ευρώ. Για να υπολογίσετε την επιδότηση που θα λάβετε, μπορείτε να χρησιμοποιήσετε τον ακόλουθο τύπο: Επιδοτούμενη Ποσό = (Τιμή Προ ΦΠΑ * Ποσοστό Επιδότησης) / 100 Επιδοτούμενη Ποσό = (15.000 * 30) / 100 Επιδοτούμενη Ποσό = 4.500 ευρώ Επομένως, θα λάβετε επιδότηση 4.500 ευρώ για	Όχι Όχι	Ναι Ναι Ναι	Εδώ το μοντέλο έχει χρησιμοποιήσει το σωστό κομμάτι πληροφορίας και μάλιστα έχει κάνει μαθηματικούς υπολογισμούς.

		την αγορά ηλεκτρικού αυτοκινήτου με κόστος 15.000 ευρώ.			
21	Για να πάρω δώρο την εγκατάσταση της αντλίας θερμότητας πρέπει να δώσω κάποια συσκευή για ανακύκλωση;	Όχι, δεν χρειάζεται να δώσετε κάποια συσκευή για ανακύκλωση για να πάρετε δώρο την εγκατάσταση της αντλίας θερμότητας.	Όχι Όχι	Ναι Ναι Ναι	Εδώ ελέγχουμε τη δυνατότητα του μοντέλου να επεξεργαστεί τη στήλη «Απαιτείται ανακύκλωση» των δεδομένων.
22	Είμαι οδηγός ταξί αμέα, υπάρχει κάποια επιπλέον παροχή για εμένα;	Ναι, υπάρχει επιπλέον παροχή για οδηγούς ταξί αμέα. Σύμφωνα με το πρόγραμμα "Πράσινα Ταξί", προβλέπεται προσαύξηση έως και 1.000€ στο ποσό του οικολογικού bonus για οδηγούς ταξί αμέα.	Όχι Όχι	Ναι Ναι Ναι	Το μοντέλο απαντά συνοπτικά και περιεκτικά, δίνοντας την απαραίτητη πληροφορία.

Πίνακας 5: Αποτελέσματα αξιολόγησης του ψηφιακού βοηθού ενεργειακής μετάβασης

Παρατηρούμε ότι ως επί το πλείστον το σύστημα δίνει ικανοποιητικές απαντήσεις, σπάνια παράγοντας παραισθήσεις. Επιπλέον, μόλις μία φορά εμφανίστηκε εξωτερική παραίσθηση, δηλαδή το μοντέλο έδωσε απάντηση που δεν συμβαδίζει με την πραγματικότητα. Σε ορισμένα ερωτήματα φάνηκαν ακόμα οι ικανότητες του μοντέλου να παράγει σύνοψη και να εκτελεί μαθηματικούς υπολογισμούς, βασιζόμενο στα δεδομένα στοιχεία.

9.2 Προκλήσεις που αντιμετωπίστηκαν

Τα σημαντικότερα ζητήματα που εντοπίστηκαν κατά την εκτέλεση των παραπάνω δοκιμών είναι τα εξής:

- Καθοριστικός παράγοντας ήταν το **μέγεθος των κομματιών** στα οποία κατακερματίζεται η δεδομένη πληροφορία καθώς και η **επικάλυψη χαρακτήρων** μεταξύ διαδοχικών κομματιών. Στις δοκιμές μας **καταλήξαμε σε μέγεθος κομματιών 1000 χαρακτήρες με επικάλυψη 150 χαρακτήρων**. Παρατηρήθηκε πως όσο αυξανόταν το μέγεθος των κομματιών, σε συνδυασμό με τον αριθμό κομματιών που ανακτώνται, ορισμένες φορές η είσοδος του μοντέλου ξεπερνούσε το μήκος περιεχομένου (context length) και λαμβάναμε σφάλμα. Από την άλλη, όσο πιο μικρό ήταν το μήκος του κομματιού, τόσο πιο επιρρεπές ήταν το μοντέλο σε παραισθήσεις, καθώς αδυνατούσε να βρει όλη τη σχετική πληροφορία. Χαρακτηριστικό παράδειγμα αποτελεί η ερώτηση 14 του παραπάνω πίνακα, όπου τα δύο συνεχόμενα κομμάτια χωρίζονταν ακριβώς στον πίνακα. Έτσι, το μοντέλο ενσωματώσεων έφερε μόνο το πρώτο κομμάτι (πριν τον πίνακα) καθώς μόνο με αυτό έβλεπε ομοιότητα. Κατά συνέπεια, το μοντέλο παρήγαγε μόνο του έναν αυθαίρετο πίνακα. Σχετικά με το μήκος επικάλυψης, παρατηρήθηκε πως το μοντέλο έδινε πιο πλήρεις και συνεκτικές απαντήσεις όταν αυτό αυξήθηκε από τους 50 στους 100 και στη συνέχεια στους 150 χαρακτήρες.
- Μεγάλη σημασία διαδραματίζει και ο **αριθμός των σχετικότερων κομματιών που ανακτώνται**. Επειδή η συγκεκριμένη εφαρμογή είναι τέτοια που μπορεί να χρειαστεί να ανακτηθούν πολλά κομμάτια πληροφορίας (π.χ. στην ερώτηση 19 του πίνακα), επιλέχθηκε να **ανακτώνται τα οκτώ σχετικότερα** με την ερώτηση του χρήστη έγγραφα, έναντι της προεπιλογής των τεσσάρων σχετικότερων. Διατηρώντας μία ισορροπία σχετικά τόσο με τον αριθμό εγγράφων όσο και με το μέγεθος των κομματιών

καταφέραμε να έχουμε ικανοποιητικά αποτελέσματα χωρίς να ξεπερνάμε το μήκος περιεχομένου του μοντέλου.

- Όπως αναμέναμε, παρατηρήθηκε **καθυστέρηση στις απαντήσεις** που λαμβάναμε από το μοντέλο. Για να πάρουμε κάθε απάντηση έπρεπε να περιμένουμε συνήθως λίγο λιγότερο από μισό λεπτό. Αυτό οφείλεται στο γεγονός πως χρησιμοποιήσαμε τη δωρεάν παροχή του HuggingFace για την εκτέλεση του μοντέλου.
- Δοκιμάστηκε η **ρύθμιση της θερμοκρασίας του μοντέλου** αλλά δεν διαπιστώθηκε κάποια διαφορά. Η θερμοκρασία (temperature) είναι μία παράμετρος που ορίζει πόσο «δημιουργική» θα είναι η απάντηση του μοντέλου. Όσο μεγαλύτερη είναι η θερμοκρασία, τόσο μεγαλύτερη η τυχαιότητα στην επιλογή των λέξεων. Όσο μικρότερη είναι η θερμοκρασία, τόσο πιο ακριβής είναι η απάντηση του μοντέλου, χρησιμοποιώντας τις πιο πιθανές λέξεις βάσεις της εκπαίδευσής του. Το συγκεκριμένο μοντέλο δεν φαίνεται να επηρεάζεται από την ρύθμιση της θερμοκρασίας όπως ορίζεται από το Hugging Face API.

10 Συμπεράσματα και προτάσεις

Στην εργασία αυτή σχεδιάστηκε και υλοποιήθηκε ένας ψηφιακός βοηθός ενεργειακής μετάβασης βασισμένος σε τεχνητή νοημοσύνη. Ως στόχο έχει να βοηθήσει τους καταναλωτές στην ανακάλυψη και κατανόηση δράσεων - κρατικών και μη - που προωθούν την ενεργειακή αναβάθμιση και εξοικονόμηση αλλά και στην παροχή οδηγιών σχετικά με το πώς μπορούν να ενταχθούν σε αυτές.

Στην αρχή πραγματοποιήθηκε μία επισκόπηση δύο πολύ σημαντικών και επίκαιρων ζητημάτων, αυτών της κλιματικής αλλαγής και της ενεργειακής κρίσης. Στη συνέχεια, επισημάνθηκε το πώς τα δύο παραπάνω φαινόμενα δημιουργούν μία νέα πραγματικότητα και φέρνουν την ανάγκη για ενεργειακή μετάβαση από τα ορυκτά καύσιμα στις ανανεώσιμες πηγές ενέργειας, με ταυτόχρονη μείωση της κατανάλωσης. Έπειτα, έγινε αναφορά στις δράσεις και τα κίνητρα που δίνονται για την επίτευξη της ενεργειακής μετάβασης και αναλύθηκαν τα βασικά χαρακτηριστικά και οι κατηγορίες τους. Τέθηκαν ακόμα ορισμένες γνωστικές βάσεις σχετικά με την τεχνητή νοημοσύνη και τα LLMs, με έμφαση στο φαινόμενο των παραισθήσεων και την αρχιτεκτονική ενός συστήματος RAG. Στη συνέχεια έγινε η σύνδεση της τεχνητής νοημοσύνης με την ενεργειακή μετάβαση, παρουσιάζοντας υπάρχουσες λύσεις ενεργειακών βοηθών και παρουσιάζοντας την πρόταση της εργασίας αυτής. Έπειτα, δόθηκαν όλες οι απαραίτητες λεπτομέρειες υλοποίησης και εν τέλει παρουσιάστηκαν τα αποτελέσματα και οι προκλήσεις που αντιμετωπίστηκαν.

Τα συμπεράσματα που προέκυψαν από την υλοποίηση του ψηφιακού βοηθού και τις δοκιμές του είναι τα εξής:

- Σε περίπου 100 γραμμές κώδικα είναι δυνατόν να υλοποιηθεί ένας βοηθός τεχνητής νοημοσύνης. Αυτό δείχνει πόσο έχει αναπτυχθεί ο τομέας των LLMs και τα συναφή frameworks που διατίθενται και μάλιστα χωρίς κόστος.
- Πραγματοποιείται πρόοδος στα ελληνικά LLMs, με το Meltemi να αποτελεί ένα αποτελεσματικό μοντέλο που μπορεί να χρησιμοποιηθεί για ποικίλες εφαρμογές.
- Το Meltemi διαθέτει σχετικά μικρό μήκος περιεχομένου (8192 tokens), κάτι που δημιουργεί προβλήματα αν υπάρχει ανάγκη τροφοδότησης με μεγάλο όγκο δεδομένων. Τα σύγχρονα μοντέλα όπως το Google Gemini 1.5 Pro φτάνουν τα δύο εκατομμύρια tokens μήκος περιεχομένου [70].

- Η επιλογή να δομηθούν τα δεδομένα σε αρχείο Excel βοήθησε στην οπτική επισκόπησή τους και την εύκολη εισαγωγή τους. Ωστόσο, η ύπαρξη μεγάλων κειμένων μέσα στα κελιά το καθιστούσε δυσλειτουργικό.
- Τα αποτελέσματα των δοκιμών απέδειξαν πως με ανοιχτά και δωρεάν εργαλεία είναι δυνατόν να κατασκευαστεί ένας ψηφιακός βοηθός ικανοποιητικής απόδοσης χωρίς να χρειάζεται αυτός που το υλοποιεί να ασχοληθεί με αλγορίθμους και κανόνες. Εκεί ακριβώς είναι που αποκαλύπτεται η «μαγεία» της τεχνητής νοημοσύνης.

Στη συνέχεια γίνεται αναφορά στο πώς θα μπορούσαν να επιλυθούν τα προβλήματα που αντιμετωπίστηκαν και παρουσιάζονται σημεία στα οποία υπάρχει περιθώριο μελλοντικών βελτιώσεων και προεκτάσεων.

10.1 Αντιμετώπιση προβλημάτων

- Για την αντιμετώπιση ορισμένων παραισθήσεων που παρατηρήθηκαν μπορεί να πραγματοποιηθεί δοκιμή χρησιμοποιώντας κάποιο μοντέλο με μεγαλύτερο μήκος περιεχομένου.
- Για την ταχύτερη απόκριση του μοντέλου θα μπορούσε αυτό να «τρέξει» είτε τοπικά σε ισχυρή κάρτα γραφικών είτε απομακρυσμένα σε κάποια υποδομή κατάλληλη για εφαρμογές τεχνητής νοημοσύνης.
- Η παροχή αναλυτικών πληροφοριών για παρεμφερή προγράμματα ίσως «μπέρδευε» το μοντέλο και γι' αυτό παράγονταν ορισμένες από τις παραισθήσεις. Θα μπορούσε να δοκιμαστεί η παροχή λιγότερων πληροφοριών μέσα στο αρχείο Excel και η προτροπή του μοντέλου να ενημερώνει σε κάθε απάντησή του για τον ιστότοπο στον οποίο ο χρήστης μπορεί να βρει περισσότερες πληροφορίες για το συγκεκριμένο πρόγραμμα.

10.2 Βελτιώσεις και προεκτάσεις

- Είναι δυνατό να υλοποιηθεί λειτουργία μνήμης στον ψηφιακό βοηθό, δηλαδή να θυμάται τη μέχρι τώρα συνομιλία με το χρήστη. Αυτό είναι τεχνικά εφικτό κάνοντας χρήση του template που διαθέτει το εκάστοτε μοντέλο και περνώντας του με κάθε ερώτημα όλο το ιστορικό της συνομιλίας μέχρι εκείνη τη στιγμή.

- Για να μην απαιτείται η συντήρηση ενός αρχείου Excel με όλη την πληροφορία σχετικά με τα προγράμματα ενεργειακής αναβάθμισης, θα μπορούσε να συνδυαστεί ανάκτηση της λίστας των προγραμμάτων από ένα αρχείο Excel και έπειτα ανάκτηση των αναλυτικών πληροφοριών για κάθε ένα από αυτά από τον αντίστοιχο ιστότοπο με προγραμματιστικό τρόπο.
- Ο ψηφιακός οδηγός θα μπορούσε να επεκταθεί περαιτέρω και να γίνει κάτι παραπάνω από ένα σύστημα RAG, παίρνοντας τη μορφή πράκτορα (agent). Οι πράκτορες έχουν στη διάθεσή τους επιπλέον εργαλεία που αυξάνουν τις δυνατότητές τους. Για παράδειγμα, είναι δυνατόν να πραγματοποιείται κλήση συναρτήσεων (function calling), ανάλογα με το ερώτημα του χρήστη, ώστε να συγκρίνεται το ενεργειακό όφελος μεταξύ της εγκατάστασης φωτοβολταϊκών και ανεμογεννητριών και μάλιστα χρησιμοποιώντας επίκαιρα δεδομένα τιμών ενέργειας.

11 Παράρτημα: Κώδικας της υλοποίησης

11.1 Μετατροπή του Excel σε CSV - excel2csv.py

```
import pandas as pd

# Καθορισμός ονομασίας και φύλλου του αρχείου Excel
excel_file = 'energy_programs.xlsx'
sheet_name = 'Sheet1'

# Φορτωση αρχείου
df = pd.read_excel(excel_file, sheet_name=sheet_name)

# Αποθήκευση αρχείου σε μορφή CSV, με καθορισμό μη προεπιλεγμένου οριοθέτη
csv_file = 'energy_programs.csv'
separator = ';' # Customize your separator here
df.to_csv(csv_file, sep=separator, index=False)
```

11.2 Κώδικας της εφαρμογής RAG – main.py

```
from langchain.text_splitter import CharacterTextSplitter
from langchain.document_loaders import CSVLoader
from langchain.embeddings import HuggingFaceEmbeddings
from langchain.vectorstores import Pinecone as PineconeVectorStore
from pinecone import Pinecone, ServerlessSpec
from langchain import PromptTemplate
from langchain.schema.runnable import RunnablePassthrough
import os
import requests

# Επικεφαλίδες HTTP για την επικοινωνία με το μοντέλο
headers = {
    "Accept": "application/json",
    "Authorization": "Bearer " + os.getenv('HUGGINGFACE_API_KEY'),
    "Content-Type": "application/json"
}
```

```

# Συνάρτηση κλήσης του μοντέλου
def query(input):

    payload = {
        "inputs": input.to_string(),
        "parameters": {
            "max_new_tokens": 250, # Μέγιστος αριθμός tokens στην έξοδο
        }
    }

    response = requests.post("https://*****.us-east-1.aws.endpoints.huggingface.cloud", headers=headers, json=payload)
    # Αποκρύπτεται το ακριβές URL
    return response.json()

class ChatBot():

    # Φόρτωση και καρακερματισμός δεδομένων
    loader = CSVLoader('./energy_programs.csv', encoding='UTF-8',
csv_args={"delimiter": ";"})
    documents = loader.load()
    docs = documents
    text_splitter = CharacterTextSplitter(chunk_size=1000, chunk_overlap=150)
    docs = text_splitter.split_documents(documents)

    # Αρχικοποίηση του μοντέλου ενσωματώσεων
    modelPath = "sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2"
    model_kwargs = {'device': 'cpu'}
    embeddings = HuggingFaceEmbeddings(
        model_name=modelPath,
        model_kwargs=model_kwargs
    )

    # Αρχικοποίηση του πελάτη Pinecone
    pc = Pinecone(
        api_key= os.getenv('PINECONE_API_KEY')
    )

    # Δημιουργία του ευρετηρίου αν δεν υπάρχει
    index_name = "energy-programs"

```

```

indexes = pc.list_indexes()
indexnames = [index.name for index in indexes]
if index_name not in indexnames:
    print('Creating new index with name: ' + index_name)
    pc.create_index(name=index_name, metric="cosine", dimension=384,
spec=ServerlessSpec(
    cloud="aws",
    region="us-east-1")
    )
    docsearch = PineconeVectorStore.from_documents(docs, embeddings,
index_name=index_name)
else:
    docsearch = PineconeVectorStore.from_existing_index(index_name,
embeddings)

```

Πρότυπο του επαυξημένου ερωτήματος

```
template = """<|system|>
```

Είσαι ένας ευγενικός και εξυπηρετικός βοηθός που σκοπός σου είναι να απαντάς στις ερωτήσεις του χρήστη σχετικά με τα προγράμματα ενεργειακής αναβάθμισης που δίνονται στο παρακάτω περιεχόμενο.

Θα πρέπει να χρησιμοποιείς μόνο το περιεχόμενο που σου δίνεται παρακάτω για να απαντήσεις στην ερώτηση.

Αν δεν γνωρίζεις την απάντηση βάσει του παρακάτω περιεχομένου θα πρέπει να απαντάς με "Δεν γνωρίζω την απάντηση για αυτό".

Η απάντησή σου πρέπει πάντα να είναι σύντομη και περιεκτική.

Περιεχόμενο: {context}

```
</s>
```

```
<|user|>
```

```
{question}</s>
```

```
<|assistant|>"""
```

Ορισμός του ερωτήματος βάσει των εισόδων (context, question)

```
prompt = PromptTemplate(
    template=template,
    input_variables=["context", "question"]
)
```

Καθορισμός της αλυσίδας RAG

```
rag_chain = (
```



```

        {"context": docsearch.as_retriever(search_kwargs={"k": 8}), "question":
RunnablePassthrough()
    | prompt
    | query
)

```

```

# Ο παρακάτω κώδικας θα πρέπει να σχολιαστεί για εκτέλεση με το Streamlit
bot = ChatBot()

```

```

print("Είμαι ο ψηφιακός βοηθός ενεργειακής μετάβασης. Τι θα ήθελες να μάθεις
σήμερα; ")

```

```

# Ανάγνωση της εισόδου μέχρι να εισαχθεί η λέξη "exit" που τερματίζει
while True:
    user_input = input("> ")
    if user_input.lower() == "exit":
        print("Τερματισμός.")
        break
    result = bot.rag_chain.invoke(user_input)
    # Εκτύπωση του αποτελέσματος
    print(result[0]['generated_text'].split('<|assistant|>')[-1].strip())

```

11.3 Κώδικας της ενσωμάτωσης του Streamlit

```

from main import ChatBot
import streamlit as st

bot = ChatBot()

# Καθορισμός λεκτικών
st.set_page_config(page_title="Ψηφιακός βοηθός ενεργειακής μετάβασης")
with st.sidebar:
    st.title('Ψηφιακός βοηθός ενεργειακής μετάβασης')

# Συνάρτηση κλήσης του συστήματος RAG
def generate_response(input):
    result = bot.rag_chain.invoke(input)
    return result[0]['generated_text'].split('<|assistant|>')[-1].strip()

```

```

# Εμφάνιση του μηνύματος καλωσορίσματος
if "messages" not in st.session_state.keys():
    st.session_state.messages = [{"role": "assistant", "content": "Χαίρεται! Πώς  
μπορώ να βοηθήσω;"}]

# Εμφάνιση των αποθηκευμένων μηνυμάτων
for message in st.session_state.messages:
    with st.chat_message(message["role"]):
        st.write(message["content"])

# Αποθήκευση των μηνυμάτων του χρήστη
if input := st.chat_input():
    st.session_state.messages.append({"role": "user", "content": input})
    with st.chat_message("user"):
        st.write(input)

# Αν το τελευταίο μήνυμα δεν προέρχεται από το chatbot (είναι δηλαδή μήνυμα  
χρήστη, τότε κάλεσε το σύστημα RAG και αποθήκευσε την απάντηση
if st.session_state.messages[-1]["role"] != "assistant":
    with st.chat_message("assistant"):
        with st.spinner("Getting your answer.."):
            response = generate_response(input)
            st.write(response)
    message = {"role": "assistant", "content": response}
    st.session_state.messages.append(message)

```

Βιβλιογραφία

- [1] United Nations, « United Nations Framework Convention on Climate Change,» [Ηλεκτρονικό]. Available: <http://unfccc.int/resource/docs/convkp/conveng.pdf>. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [2] T. R. Karl και K. E. Trenberth, «Modern global climate change.,» *science*, τόμ. 302, αρ. 5651, pp. 1719-1723, 2003.
- [3] Ευρωπαϊκή Επιτροπή - Γενική Διεύθυνση Δράσης για το Κλίμα, «Αίτια της κλιματικής αλλαγής,» [Ηλεκτρονικό]. Available: https://climate.ec.europa.eu/climate-change/causes-climate-change_el. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [4] Ευρωπαϊκή Επιτροπή - Γενική Διεύθυνση Δράσης για το Κλίμα, «Συνέπειες της κλιματικής αλλαγής,» [Ηλεκτρονικό]. Available: https://climate.ec.europa.eu/climate-change/consequences-climate-change_el. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [5] in.gr, «Λειψυδρία: Η Αττική και τα νησιά κινδυνεύουν να μείνουν χωρίς νερό,» 3 Ιουλίου 2024. [Ηλεκτρονικό]. Available: <https://www.in.gr/2024/07/03/greece/leipsydria-attiki-kai-ta-nisia-kindyneyou-n-na-meinoun-xoris-nero/>. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [6] C. Cunningham, G. Williamson και D. Bowman, «Increasing frequency and intensity of the most extreme wildfires on Earth.,» *Nat Ecol Evol*, τόμ. 8, αρ. 8, p. 1420–1425, 2024.
- [7] Σ. Κράλογλου, «Υποχρεωτική ασφάλιση σε ακίνητα και Ι.Χ. από φυσικές καταστροφές,» *newmoney*, 31 Οκτωβρίου 2024. [Ηλεκτρονικό]. Available: <https://www.newmoney.gr/roh/palmos-oikonomias/oikonomia/ipochreotiki-asfalisi-se-akinita-ke-i-ch-apo-fisikes-katastrofes/>. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [8] KfW, «Majority of Germans in favour of compulsory insurance for natural hazards,» 6 Ιουλίου 2022. [Ηλεκτρονικό]. Available: https://www.kfw.de/About-KfW/Newsroom/Latest-News/Pressemitteilungen-Details_717824.html. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [9] S. Fawzy, A. I. Osman, J. Doran και D. W. Rooney, «Strategies for mitigation of climate change: a review,» *Environmental Chemistry Letters*, τόμ. 18, αρ. 6, pp. 2069 - 2094, 2020.

- [10] United Nations, «Paris Agreement,» 2015. [Ηλεκτρονικό]. Available: https://unfccc.int/sites/default/files/english_paris_agreement.pdf. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [11] H. Ritchie, M. Roser και P. Rosado, «CO₂ and Greenhouse Gas Emissions,» OurWorldinData.org, 2023. [Ηλεκτρονικό]. Available: <https://ourworldindata.org/co2-and-greenhouse-gas-emissions>. [Πρόσβαση 14 Δεκεμβρίου 2024].
- [12] G. D. Nash, «Energy Crises in Historical Perspective,» *Nat. Resources J.*, τόμ. 21, p. 341, 1981.
- [13] The World Bank, «ENERGY CRISIS,» Οκτώβριος 22. [Ηλεκτρονικό]. Available: <https://openknowledge.worldbank.org/server/api/core/bitstreams/b39442da-9f3e-57e1-9b69-0e893fad4fcc/content>. [Πρόσβαση 15 Δεκεμβρίου 2024].
- [14] International Energy Agency, «Global Energy Crisis,» [Ηλεκτρονικό]. Available: <https://www.iea.org/topics/global-energy-crisis>. [Πρόσβαση 15 Δεκεμβρίου 2024].
- [15] M. Farghali et al., «Strategies to save energy in the context of the energy crisis: a review,» *Environmental Chemistry Letters*, τόμ. 21, αρ. 4, pp. 2003-2039, 2023.
- [16] Ελεύθερος Τύπος, «Ενεργειακή Κρίση: Αιτίες, Συνέπειες και Λύσεις,» 16 Νοεμβρίου 2022. [Ηλεκτρονικό]. Available: <https://elftherostypos.gr/business/energeiaki-krisi-aities-synepeies-kai-lyseis>. [Πρόσβαση 18 Δεκεμβρίου 2024].
- [17] Ευρωπαϊκή Επιτροπή, «Δράσεις της ΕΕ για την αντιμετώπιση της ενεργειακής κρίσης,» [Ηλεκτρονικό]. Available: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/european-green-deal/eu-action-address-energy-crisis_el. [Πρόσβαση 15 Δεκεμβρίου 2024].
- [18] S. Davidsson, «Global energy transitions,» 2015. [Ηλεκτρονικό]. Available: <https://www.diva-portal.org/smash/get/diva2:791155/FULLTEXT01.pdf>. [Πρόσβαση 18 Δεκεμβρίου 2024].
- [19] B. D. Solomon και K. Krishna, «The coming sustainable energy transition: History, strategies, and outlook,» *Energy Policy*, τόμ. 39, αρ. 11, pp. 7422-7431, 2011.
- [20] Α. Λαμπρίδη, Ε. Τσαμαντάνη και Α. Σαλονίκα, «Πλεονεκτήματα και Μειονεκτήματα Ενεργειακής Μετάβασης,» [Ηλεκτρονικό]. Available: https://www.dept.aueb.gr/sites/default/files/msclawecon/Λαμπρίδη_Τσαμαντάνη_Σαλονίκα.pdf. [Πρόσβαση 23 Δεκεμβρίου 2024].

- [21] S. de la Rue du Can, G. Leventis, A. Phadke και A. Gopal, «Design of incentive programs for accelerating penetration of energy-efficient appliances,» *Energy Policy*, τόμ. 72, pp. 56-66, 2014.
- [22] Υπουργείο Περιβάλλοντος και Ενέργειας, «ΑΝΑΚΥΚΛΩΝΩ-ΑΛΛΑΖΩ ΣΥΣΚΕΥΗ,» [Ηλεκτρονικό]. Available: <https://allazosyskevi.gov.gr/>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [23] Υπουργείο Περιβάλλοντος και Ενέργειας, «ΑΝΑΚΥΚΛΩΝΩ-ΑΛΛΑΖΩ ΘΕΡΜΟΣΙΦΩΝΑ,» [Ηλεκτρονικό]. Available: <https://allazothermosifona.gov.gr/>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [24] Υπουργείο Περιβάλλοντος και Ενέργειας, «Εξοικονομώ Επιχειρώντας «Αλλάζω συσκευή για τις επιχειρήσεις»,» [Ηλεκτρονικό]. Available: <https://exoikonomo-epixeiro2023.gov.gr/>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [25] Υπουργείο Υποδομών και Μεταφορών, «Κινούμαι Ηλεκτρικά Γ' κύκλος,» [Ηλεκτρονικό]. Available: <https://kinoumeilektrika3.gov.gr/>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [26] Υπουργείο Υποδομών και Μεταφορών, «Πράσινα Ταξί,» [Ηλεκτρονικό]. Available: <https://prasinataxi.gov.gr/>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [27] Υπουργείο Περιβάλλοντος και Ενέργειας, «ΕΞΟΙΚΟΝΟΜΩ 2023,» [Ηλεκτρονικό]. Available: <https://exoikonomo2023.gov.gr/>. [Πρόσβαση 27 Δεκεμβρίου 2024].
- [28] ΔΕΔΔΗΕ, «ρνstegi,» [Ηλεκτρονικό]. Available: <https://ape4all.deddie.gr/>. [Πρόσβαση 27 Δεκεμβρίου 2024].
- [29] OpenAI, «<https://chatgpt.com/>,» [Ηλεκτρονικό]. Available: <https://chatgpt.com/>.
- [30] IBM, «What is an AI model?,» [Ηλεκτρονικό]. Available: <https://www.ibm.com/think/topics/ai-model>. [Πρόσβαση 23 Δεκεμβρίου 2024].
- [31] SAP, «What is generative AI?,» [Ηλεκτρονικό]. Available: <https://www.sap.com/products/artificial-intelligence/what-is-generative-ai.html>. [Πρόσβαση 24 December 2024].
- [32] A. Vaswani et. al, «Attention is all you need.,» *Advances in Neural Information Processing Systems*, 2017.

- [33] Microsoft, «Understanding tokens,» 21 Δεκεμβρίου 2024. [Ηλεκτρονικό]. Available: <https://learn.microsoft.com/en-us/dotnet/ai/conceptual/understanding-tokens>. [Πρόσβαση 23 Δεκεμβρίου 2024].
- [34] OpenAI, «Tokenizer - OpenAI API,» [Ηλεκτρονικό]. Available: <https://platform.openai.com/tokenizer>. [Πρόσβαση 23 Δεκεμβρίου 2024].
- [35] L. Huang et al., «A Survey on Hallucination in Large Language Models: Principles, Taxonomy, Challenges, and Open Questions,» 2024.
- [36] M. Ramponi, «How RLHF Preference Model Tuning Works (And How Things May Go Wrong),» 3 Αυγούστου 2023. [Ηλεκτρονικό]. Available: <https://www.assemblyai.com/blog/how-rlhf-preference-model-tuning-works-and-how-things-may-go-wrong/>. [Πρόσβαση 8 Δεκεμβρίου 2024].
- [37] S. Mehta, «<https://adasci.org/how-much-energy-do-llms-consume-unveiling-the-power-behind-ai/>,» <https://adasci.org/how-much-energy-do-llms-consume-unveiling-the-power-behind-ai/>, 3 Ιουλίου 2024. [Ηλεκτρονικό]. Available: <https://adasci.org/how-much-energy-do-llms-consume-unveiling-the-power-behind-ai/>. [Πρόσβαση 23 Δεκεμβρίου 2024].
- [38] HuggingFace, «Open LLM Leaderboard,» [Ηλεκτρονικό]. Available: https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard#/. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [39] Y. Chang et al., «A survey on evaluation of large language models,» *ACM Transactions on Intelligent Systems and Technology*, τόμ. 15, αρ. 3, pp. 1-45, 2024.
- [40] Zhao, Wayne Xin, et al., «A Survey of Large Language Models,» *arXiv preprint arXiv:2303.18223*, 2023.
- [41] Ziwei Ji et al., «Survey of hallucination in natural language generation,» *ACM Computing Surveys*, τόμ. 55, αρ. 12, pp. 1-38, 2023.
- [42] M. Faheem, «Decoding Demystified : How LLMs Generate Text - III,» [Ηλεκτρονικό]. Available: <https://dev.to/mahakfaheem/decoding-demystified-how-llms-generate-text-iii-3a0d>. [Πρόσβαση 9 Δεκεμβρίου 2024].
- [43] P. Sahoo et al., «A Systematic Survey of Prompt Engineering in Large Language Models:,» *arXiv preprint arXiv:2402.07927*, 2024.

- [44] J. Maynez, S. Narayan, B. Bohnet και R. McDonald, «On Faithfulness and Factuality in Abstractive Summarization,» *arXiv preprint arXiv:2005.00661*, 2020.
- [45] Y. Gao et al., «Retrieval-augmented generation for large language models: A survey,» *arXiv preprint arXiv:2312.10997*, 2023.
- [46] Amazon Web Services, Inc., «What is RAG (Retrieval-Augmented Generation)?,» [Ηλεκτρονικό]. Available: <https://aws.amazon.com/what-is/retrieval-augmented-generation/>. [Πρόσβαση 12 Δεκεμβρίου 2024].
- [47] Microsoft, «Vector database,» [Ηλεκτρονικό]. Available: <https://learn.microsoft.com/en-us/azure/cosmos-db/vector-database>. [Πρόσβαση 12 Δεκεμβρίου 2024].
- [48] DAIR.AI, «Elements of a Prompt | Prompt Engineering Guide,» [Ηλεκτρονικό]. Available: <https://www.promptingguide.ai/introduction/elements>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [49] LangChain, «rlm/rag-prompt,» [Ηλεκτρονικό]. Available: <https://smith.langchain.com/hub/rlm/rag-prompt>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [50] LlamaIndex, «Prompt Engineering for RAG,» [Ηλεκτρονικό]. Available: https://docs.llamaindex.ai/en/stable/examples/prompts/prompts_rag/. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [51] LangChain, «rlm/rag-prompt-llama,» [Ηλεκτρονικό]. Available: <https://smith.langchain.com/hub/rlm/rag-prompt-llama>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [52] DigitalOcean, LLC., «Prompt Engineering Best Practices: Tips, Tricks, and Tools,» [Ηλεκτρονικό]. Available: <https://www.digitalocean.com/resources/articles/prompt-engineering-best-practices>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [53] OpenAI, «Prompt engineering - OpenAI API,» [Ηλεκτρονικό]. Available: <https://platform.openai.com/docs/guides/prompt-engineering>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [54] eco-bot, «eco-bot,» [Ηλεκτρονικό]. Available: <http://eco-bot.eu/>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [55] ΔΕΗ Α.Ε., «Αυτόματη Ψηφιακή Βοηθός,» [Ηλεκτρονικό]. Available: <https://chatbot.dei.gr/>. [Πρόσβαση 26 Δεκεμβρίου 2024].

- [56] ΔΕΔΔΗΕ, «Αυτόματος Ψηφιακός Βοηθός του ΔΕΔΔΗΕ,» [Ηλεκτρονικό]. Available: <https://chatbot.deddie.gr/>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [57] EnergyBot, «Chatbot,» [Ηλεκτρονικό]. Available: <https://chat.energybot.gr/>. [Πρόσβαση 26 Δεκεμβρίου 2024].
- [58] Υπουργείο Ψηφιακής Διακυβέρνησης, «mAIBot,» [Ηλεκτρονικό]. Available: <https://aibot.gov.gr/>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [59] Υπουργείο Ψηφιακής Διακυβέρνησης, «mAigov – Ο «Ψηφιακός Βοηθός» του gov.gr στην υπηρεσία των πολιτών – Το Δημόσιο στην εποχή της Τεχνητής Νοημοσύνης,» 12 Δεκεμβρίου 2023. [Ηλεκτρονικό]. Available: <https://mindigital.gr/archives/5671>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [60] «LangChain,» [Ηλεκτρονικό]. Available: <https://www.langchain.com/>. [Πρόσβαση 19 Δεκεμβρίου 2024].
- [61] «HuggingFace,» [Ηλεκτρονικό]. Available: <https://huggingface.co/>. [Πρόσβαση 19 Δεκεμβρίου 2024].
- [62] Sentence Transformers - SBERT, «sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2 - HuggingFace,» [Ηλεκτρονικό]. Available: <https://huggingface.co/sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2>. [Πρόσβαση 21 Δεκεμβρίου 2024].
- [63] HuggingFace, «The power of Sentence Similarity,» [Ηλεκτρονικό]. Available: <https://huggingface.co/learn/ml-games-course/unit1/sentence-similarity-explained>. [Πρόσβαση 21 Δεκεμβρίου 2024].
- [64] Pinecone Systems, Inc., «The vector database to build knowledgeable AI,» [Ηλεκτρονικό]. Available: <https://www.pinecone.io/>. [Πρόσβαση 21 Δεκεμβρίου 2024].
- [65] Institute for Language and Speech Processing / Athena RC, «Meltemi: A Large Language Model for Greek,» 26 Μαρτίου 2024. [Ηλεκτρονικό]. Available: <https://medium.com/institute-for-language-and-speech-processing/meltemi-a-large-language-model-for-greek-9f5ef1d4a10f>. [Πρόσβαση 21 Δεκεμβρίου 2024].
- [66] Ινστιτούτο Επεξεργασίας του Λόγου - Ερευνητικό Κέντρο Αθηνά, «Meltemi, το πρώτο ανοιχτό Μεγάλο Γλωσσικό Μοντέλο για τα Ελληνικά,» 28 Μαρτίου 2024. [Ηλεκτρονικό]. Available: <https://www.ilsp.gr/news/meltemi/>. [Πρόσβαση 21 Δεκεμβρίου 2024].

- [67] Ινστιτούτο Επεξεργασίας του Λόγου, «ilsp/Meltemi-7B-Instruct-v1.5 - HuggingFace,» [Ηλεκτρονικό]. Available: <https://huggingface.co/ilsp/Meltemi-7B-Instruct-v1.5>. [Πρόσβαση 22 Δεκεμβρίου 2024].
- [68] HuggingFace, «Inference Endpoints,» [Ηλεκτρονικό]. Available: <https://huggingface.co/inference-endpoints/dedicated>. [Πρόσβαση 22 Δεκεμβρίου 2024].
- [69] Snowflake Inc., «Streamlit - A faster way to build and share data apps,» [Ηλεκτρονικό]. Available: <https://streamlit.io/>. [Πρόσβαση 24 Δεκεμβρίου 2024].
- [70] Google Inc., «Long context | Gemini API | Google AI for Developers,» [Ηλεκτρονικό]. Available: <https://ai.google.dev/gemini-api/docs/long-context>. [Πρόσβαση 30 Δεκεμβρίου 2024].