



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

**Αυτοματοποιημένο Σύστημα Επικοινωνίας Φωνής με
Μεγάλα Γλωσσικά Μοντέλα: Υλοποίηση και Αξιολόγηση**

Μεταπτυχιακή Διπλωματική Εργασία

Βαλιάκος Απόστολος

Επιβλέπων: Βασιλακόπουλος Μιχαήλ

Ιούνιος 2024



ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΙΑΣ

ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ

ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

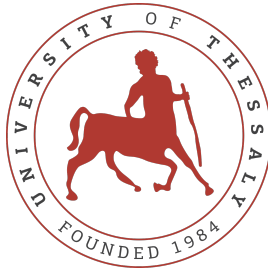
**Αυτοματοποιημένο Σύστημα Επικοινωνίας Φωνής με
Μεγάλα Γλωσσικά Μοντέλα: Υλοποίηση και Αξιολόγηση**

Μεταπτυχιακή Διπλωματική Εργασία

Βαλιάκος Απόστολος

Επιβλέπων: Βασιλακόπουλος Μιχαήλ

Ιούνιος 2024



UNIVERSITY OF THESSALY

SCHOOL OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

Automated Voice Communication System with Large Language Models: Implementation and Evaluation

MSc Thesis

Valiakos Apostolos

Supervisor: Teacher Vasilakopoulos Michail

June 2024

Εγκρίνεται από την Επιτροπή Εξέτασης:

Επιβλέπων **Βασιλακόπουλος Μιχαήλ**

Καθηγητής, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών, Πανεπιστήμιο Θεσσαλίας

Μέλος **Δασκαλοπούλου Ασπασία**

Αναπληρώτρια Καθηγήτρια, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

Μέλος **Τσομπανοπούλου Παναγιώτα**

Αναπληρώτρια Καθηγήτρια, Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών

ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ ΠΕΡΙ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ ΚΑΙ ΠΝΕΥΜΑΤΙΚΩΝ ΔΙΚΑΙΩΜΑΤΩΝ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ρητά ότι η παρούσα μεταπτυχιακή διπλωματική εργασία, καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας αυτής, αποτελούν αποκλειστικά προϊόν προσωπικής μου εργασίας, δεν προσβάλλουν οποιασδήποτε μορφής δικαιώματα διανοητικής ιδιοκτησίας, προσωπικότητας και προσωπικών δεδομένων τρίτων, δεν περιέχουν έργα/εισφορές τρίτων για τα οποία απαιτείται άδεια των δημιουργών/δικαιούχων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον και πληρούν τους κανόνες της επιστημονικής παράθεσης. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή/και πηγές άλλων συγγραφέων αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Δηλώνω επίσης ότι τα αποτελέσματα της εργασίας δεν έχουν χρησιμοποιηθεί για την απόκτηση άλλου πτυχίου. Αναλαμβάνω πλήρως, ατομικά και προσωπικά, όλες τις νομικές και διοικητικές συνέπειες που δύναται να προκύψουν στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής.

Ο/Η Δηλών/ούσα

Βαλιάκος Απόστολος

Μεταπτυχιακή Διπλωματική Εργασία
Αυτοματοποιημένο Σύστημα Επικοινωνίας Φωνής με Μεγάλα
Γλωσσικά Μοντέλα: Υλοποίηση και Αξιολόγηση
Βαλιάκος Απόστολος

Περίληψη

Στην παρούσα εργασία παρουσιάζονται τα βήματα που ακολουθήθηκαν για την ανάπτυξη εφαρμογής αυτοματοποιημένου συστήματος επικοινωνίας με μεγάλα γλωσσικά μοντέλα. Ο χρήστης έχει τη δυνατότητα να στείλει είσοδο σε ένα μεγάλο γλωσσικό μοντέλο χρησιμοποιώντας το μικρόφωνό του και η απάντηση που θα λάβει θα παραχθεί σε φωνή με τη χρήση άλλου μοντέλου τεχνητής νοημοσύνης. Στη συνέχεια, αναπτύχθηκε και η λειτουργία για παραγωγή εικόνων με βάση την περιγραφή που δίνει ο χρήστης από το μικρόφωνό του ενώ η είσοδος βελτιστοποιείται για να υπάρχει η καλύτερη απόδοση της περιγραφής του χρήστη.

Για την επίτευξη του συγκεκριμένου στόχου, ακολουθήθηκε μια σειρά από στάδια ανάπτυξης και βελτιστοποίησης. Αρχικά, έγινε η επιλογή των κατάλληλων γλωσσικών μοντέλων, τα οποία μπορούν να κατανοήσουν και να επεξεργαστούν φυσική γλώσσα με υψηλή ακρίβεια.

Στη συνέχεια, προσαρμόστηκε το λογισμικό ώστε να μπορεί να δέχεται είσοδο από το μικρόφωνο του χρήστη. Η είσοδος αυτή μετατρέπεται σε κείμενο μέσω μιας διαδικασίας αναγνώρισης φωνής, η οποία απαιτεί υψηλή ακρίβεια και ταχύτητα για να είναι αποτελεσματική. Έπειτα, το κείμενο που παράγεται από τη φωνητική είσοδο του χρήστη υποβάλλεται στο μεγάλο γλωσσικό μοντέλο για επεξεργασία. Το μοντέλο αναλύει το κείμενο και παράγει την κατάλληλη απάντηση.

Μόλις παραχθεί η απάντηση από το γλωσσικό μοντέλο, περνάει από το λογισμικό το οποίο αναπαράγει την απάντηση σε φυσική, κατανοητή φωνή, ώστε ο χρήστης να μπορεί να την ακούσει. Η διαδικασία αυτή εξασφαλίζει ότι η επικοινωνία είναι ομαλή και αμφίδρομη, παρέχοντας μια φυσική εμπειρία στον χρήστη.

Παράλληλα με την επεξεργασία κειμένου και φωνής, η εφαρμογή επεκτάθηκε για να περιλαμβάνει και τη δυνατότητα παραγωγής εικόνων βάσει περιγραφών. Ο χρήστης μπορεί να περιγράψει μια σκηνή ή ένα αντικείμενο μέσω του μικροφώνου του, και η περιγραφή αυτή

μετατρέπεται σε εικόνα με τη χρήση προηγμένων αλγορίθμων τεχνητής νοημοσύνης. Αυτοί οι αλγόριθμοι αναλύουν την περιγραφή και δημιουργούν μια εικόνα που ανταποκρίνεται όσο το δυνατόν πιστότερα στις απαιτήσεις του χρήστη.

Η βελτιστοποίηση της εισόδου για την παραγωγή εικόνων είναι επίσης κρίσιμης σημασίας. Η περιγραφή του χρήστη υποβάλλεται σε μια διαδικασία ανάλυσης και προσαρμογής, έτσι ώστε οι αλγόριθμοι να κατανοήσουν με ακρίβεια τις λεπτομέρειες και τις ιδιαιτερότητες της περιγραφής. Αυτό περιλαμβάνει τη διόρθωση πιθανών ασαφειών και την προσθήκη απαραίτητων πληροφοριών που μπορεί να λείπουν, προκειμένου να επιτευχθεί η καλύτερη δυνατή απόδοση της περιγραφής.

Το συνολικό σύστημα που αναπτύχθηκε προσφέρει μια ολοκληρωμένη εμπειρία επικοινωνίας και δημιουργίας, συνδυάζοντας την επεξεργασία φυσικής γλώσσας, τη σύνθεση φωνής και την παραγωγή εικόνων. Οι χρήστες μπορούν να αλληλεπιδρούν με το σύστημα με φυσικό και αβίαστο τρόπο, ενώ το σύστημα ανταποκρίνεται με ακρίβεια και αποτελεσματικότητα στις ανάγκες τους. Η εφαρμογή αυτή αποτελεί ένα σημαντικό βήμα προς την κατεύθυνση της ανάπτυξης προηγμένων τεχνολογιών τεχνητής νοημοσύνης που διευκολύνουν την ανθρώπινη επικοινωνία και δημιουργικότητα.

Λέξεις Κλειδιά

Large Language Model (LLM), Natural Language Processing (NLP), Speech Recognition, Text-to-Speech Synthesis, Dialogue Management

MSc Thesis

Automated Voice Communication System with Large Language

Models: Implementation and Evaluation

Valiakos Apostolos

Abstract

This thesis presents the steps followed for the development of an automated communication system application using large language models. The user has the ability to send input to a large language model using their microphone, and the response they receive will be produced in voice by using another artificial intelligence model. Additionally, a feature was developed for generating images based on the description given by the user through their microphone, while the input is optimized to provide the best possible rendering of the user's description.

To achieve this goal, a series of development and optimization stages were followed. Initially, the appropriate language models were selected, which can understand and process natural language with high accuracy.

Next, the software was adapted to accept input from the user's microphone. This input is converted into text through a speech recognition process, which requires high accuracy and speed to be effective. Then, the text generated from the user's vocal input is submitted to the large language model for processing. The model analyzes the text and produces the appropriate response.

Once the response is generated by the language model, it is passed through the software that reproduces the response in a natural, comprehensible voice so that the user can hear it. This process ensures smooth and bidirectional communication, providing a natural experience for the user.

In parallel with text and voice processing, the application was expanded to include the ability to generate images based on descriptions. The user can describe a scene or an object through their microphone, and this description is converted into an image using advanced artificial intelligence algorithms. These algorithms analyze the description and create an image that as closely as possible matches the user's requirements.

Optimizing the input for image generation is also critically important. The user's description undergoes a process of analysis and adjustment so that the algorithms can accurately

understand the details and specifics of the description. This includes correcting potential ambiguities and adding necessary information that may be missing to achieve the best possible rendering of the description.

The developed system offers a comprehensive experience of communication and creation, combining natural language processing, voice synthesis, and image generation. Users can interact with the system in a natural and effortless manner, while the system responds accurately and effectively to their needs. This application represents a significant step towards the development of advanced artificial intelligence technologies that facilitate human communication and creativity.

Λέξεις Κλειδιά

Large Language Model (LLM), Natural Language Processing (NLP), Speech Recognition, Text-to-Speech Synthesis, Dialogue Management

Πίνακας περιεχομένων

Περίληψη	xi
Abstract	xiii
Πίνακας περιεχομένων	xv
Κατάλογος σχημάτων	xix
Κατάλογος πινάκων	xxi
Συντομογραφίες	xxiii
1 Εισαγωγή	1
1.1 Αντικείμενο της διπλωματικής	3
1.1.1 Συνεισφορά	4
1.2 Οργάνωση της Εργασίας	4
2 Βασικές Αρχές και Τεχνολογίες	7
2.1 Ιστορική αναδρομή	7
2.1.1 Μπορούν οι Μηχανές να Σκεφτούν;	7
2.1.2 Η Διάσκεψη που τα Ξεκίνησε Όλα	8
2.1.3 Επεξεργασία Φυσικής Γλώσσας	9
2.1.4 Σύνθεση Φωνής	10
2.1.5 Ιστορία της Σύνθεσης Φωνής από Υπολογιστές	10
2.1.6 Ιστορία της Ανάλυσης Φυσικής Γλώσσας από Υπολογιστές	11
2.1.7 Τελευταίες Έρευνες στη Σύνθεση Φωνής	11
2.1.8 Τελευταίες Έρευνες στην Ανάλυση Φυσικής Γλώσσας	11

2.2	Εισαγωγή στις Βασικές Αρχές και Έννοιες	12
2.2.1	Ορισμοί και Βασικές Αρχές	12
2.2.2	Πώς Λειτουργεί η Σύνθεση Φωνής (TTS)	12
2.3	Βασικοί Μηχανισμοί Πίσω από τη Σύνθεση Φωνής	13
2.3.1	Φωνολογική και Φωνητική Ανάλυση	14
2.3.2	Γλωσσολογική και Σημασιολογική Ανάλυση	14
2.3.3	Τεχνητά Νευρωνικά Δίκτυα	14
2.3.4	Ανάλυση Φυσικής Γλώσσας	14
2.3.5	Ορισμός της NLP	15
2.4	Κύριες Προκλήσεις της NLP	15
2.4.1	Ασάφεια και Πολυπλοκότητα της Φυσικής Γλώσσας	15
2.4.2	Συντακτική και Σημασιολογική Ανάλυση	15
2.4.3	Πολυγλωσσικότητα και Πολιτισμικές Διαφορές	15
2.5	Βασικές Τεχνικές της NLP	16
2.5.1	Στατιστικά Μοντέλα και Μοντέλα Μηχανικής Μάθησης	16
2.5.2	Νευρωνικά Δίκτυα και Μοντέλα Βαθιάς Μάθησης	16
2.6	Ιστορική Ανασκόπηση	17
2.6.1	Σύνθεση Φωνής (TTS)	17
2.6.2	Ανάλυση Φυσικής Γλώσσας (NLP)	18
2.7	Θεωρητικό Υπόβαθρο και Τεχνολογίες	19
2.7.1	Φωνολογική και Φωνητική Ανάλυση	19
2.7.2	Φωνολογία	19
2.7.3	Αναγνώριση Φωνητικών Μοτίβων	19
2.7.4	Διαχείριση Τονισμού και Προσοδίας	20
2.7.5	Φωνητική	20
2.8	Ενσωμάτωση της Φωνολογίας και της Φωνητικής στα Σύγχρονα TTS Συστήματα	20
2.8.1	Στάδια Ενσωμάτωσης	20
2.8.2	Παράδειγμα: WaveNet	21
2.9	Συντακτική και Σημασιολογική Ανάλυση	21
2.9.1	Τεχνικές NLP για Συντακτική και Σημασιολογική Ανάλυση Κειμένου	22
2.9.2	Συντακτική Ανάλυση:	22

2.10 Στατιστικά Μοντέλα και Μοντέλα Μηχανικής Μάθησης	29
2.10.1 Μοντέλα Μίγματος Gaussian (Gaussian Mixture Models - GMMs)	33
2.11 Συνοπτικά	34
2.12 Μετάβαση από Στατιστικά Μοντέλα σε Νευρωνικά Δίκτυα και Βαθιά Μάθηση	34
2.13 Νευρωνικά Δίκτυα: Κινητήρια δύναμη για Σύνθεση Φωνής και Ανάλυση Φυ- σικής Γλώσσας	36
2.13.1 Ευρέως Διαδεδομένα Νευρωνικά Δίκτυα	36
2.13.2 Εφαρμογές στη Σύνθεση Φωνής	37
2.13.3 Εφαρμογές στην Ανάλυση Φυσικής Γλώσσας	38
2.14 Τεχνολογίες και Εφαρμογές	38
2.15 Εφαρμογές της Ανάλυσης Φυσικής Γλώσσας	40
3 Τελευταίες Έρευνες και Τάσεις	43
3.1 Προηγμένες Τεχνολογίες και Καινοτομίες	43
3.1.1 Τελευταίες Έρευνες και Καινοτομίες στη Σύνθεση Φωνής και Ανά- λυση Φυσικής Γλώσσας	43
3.1.2 Ηθικές και Κοινωνικές Επιπτώσεις	44
4 Μεθοδολογία Ανάπτυξης της εφαρμογής	47
4.1 Βιβλιοθήκες που χρησιμοποιήθηκαν	47
4.1.1 Διαδικασία Ανάπτυξης Εφαρμογής	52
4.1.2 Χρήση της Εφαρμογής	54
5 Περαιτέρω Έρευνα	59
5.1 Μελλοντικές Προοπτικές να μεταφερθεί σε διαφορετικό κεφάλαιο στο τέλος με τίτλο περαιτέρω έρευνα	59
Βιβλιογραφία	63
ΠΑΡΑΡΤΗΜΑΤΑ	67
A Παράθεση κώδικα	69
A.0.1 Αποθετήριο κώδικα και εγκατάσταση εφαρμογής	78
A.1 Οδηγίες Εγκατάστασης	79

A.1.1	Εγκατάσταση Γλώσσας προγραμματισμού Python σε περιβάλλον Win-	
	dows	79
A.1.2	Εγκατάσταση Βιβλιοθηκών	80
A.1.3	Χρήση του OpenAI API	80

Κατάλογος σχημάτων

2.1	Alan Touring	8
2.2	Αριστερά Marvin Minsky, Claude Shannon, Ray Solomonoff και άλλοι επι- στήμονες στο Dartmouth Summer Research Project on Artificial Intelligence (Photo: Margaret Minsky)	9
2.3	Ένα δευτερόλεπτο φωνής που παράχθηκε μηχανικά	10
4.1	The First version of the app that runs in terminal	50
4.2	The Second version of the app that runs with User Interface	51
4.3	The Third version of the app	52
4.4	The Fourth version of the app	53
4.5	The Fourth version of the app	54
4.6	The Fourth version of the app	55
4.7	The Last version of the app that runs in terminal	56
4.8	Ενδεικτική Παραγόμενη εικόνα	57
A.1	Python Web Page	79
A.2	Last Version Of Python	79
A.3	Step 1 of python installation	80
A.4	Step 2 of python installation	81
A.5	Step 3 of python installation	82
A.6	Step 4 of python installation	82
A.7	Increase Limit	83
A.8	Create API key	83
A.9	API key options	83

Κατάλογος πινάκων

Συντομογραφίες

LLM	Large Language Model
βλπ	βλέπε
κ.λπ.	και λοιπά
κ.ο.κ	και ούτω καθεξής
TEI	Τεχνολογικό Εκπαιδευτικό Ίδρυμα
BPF	Band Pass Filter

Κεφάλαιο 1

Εισαγωγή

Η επικοινωνία ανθρώπων με μηχανές αποτελεί θεμελιώδες στοιχείο της αλληλεπίδρασης με την τεχνολογία από τις απαρχές της ψηφιακής εποχής. Από τις πρώτες γραμμές κώδικα που πληκτρολογήθηκαν σε τερματικά μέχρι τις φωνητικές εντολές που δίνονται σε έξυπνα ηχεία σήμερα, η αναζήτηση ολοένα και πιο φυσικών και διαισθητικών τρόπων επικοινωνίας με υπολογιστές παραμένει διαρκής επιδίωξη. Η εξέλιξη της τεχνολογίας μας έχει φέρει από τα πρώτα τερματικά πληκτρολόγησης σε μια εποχή όπου οι φωνητικές εντολές και οι φυσικές γλωσσικές αλληλεπιδράσεις είναι καθημερινές πραγματικότητες, επιτρέποντας στους χρήστες να χειρίζονται τις συσκευές τους με τρόπο που είναι πιο κοντά στην ανθρώπινη επικοινωνία.

Στο πλαίσιο αυτής της αναζήτησης, η ανάπτυξη και η εξέλιξη των Large Language Models (LLMs) σηματοδοτεί μια σημαντική αλλαγή. Τα LLMs, τροφοδοτούνται με τεράστιες ποσότητες δεδομένων κειμένου και κώδικα, κατέχουν πλέον την ικανότητα να παράγουν κείμενο ανθρώπινης ποιότητας, να μεταφράζουν γλώσσες, να γράφουν διαφορετικά είδη δημιουργικού περιεχομένου και να απαντούν στις ερωτήσεις σας με ενημερωτικό τρόπο. Η ικανότητά τους να κατανοούν και να παράγουν φυσική γλώσσα με υψηλή ακρίβεια τα καθιστά πολύτιμα εργαλεία για ποικίλες εφαρμογές, από την εξυπηρέτηση πελατών μέχρι τη δημιουργία περιεχομένου και την παροχή υποστήριξης στην έρευνα.

Τα LLMs είναι ικανά να επεξεργάζονται και να κατανοούν περίπλοκες ερωτήσεις, παρέχοντας απαντήσεις που όχι μόνο είναι σωστές αλλά και εξηγούνται με τρόπο που έχει νόημα για τον χρήστη. Μπορούν να διαχειρίζονται μεγάλες ποσότητες δεδομένων και να εξάγουν χρήσιμες πληροφορίες, κάνοντας τη διαδικασία εύρεσης πληροφοριών πιο αποτελεσματική και γρήγορη. Επιπλέον, οι δυνατότητες τους στη δημιουργική γραφή επιτρέπουν τη σύνθεση

κειμένων που μπορούν να χρησιμοποιηθούν σε ποικίλα πλαίσια, από τη λογοτεχνία και τη δημοσιογραφία μέχρι τη διαφήμιση και το μάρκετινγκ.

Η ανάπτυξη αυτών των μοντέλων έχει ανοίξει νέους δρόμους για την επικοινωνία και την αλληλεπίδραση με την τεχνολογία. Οι δυνατότητες των LLMs στην αναγνώριση φωνής και τη φυσική γλωσσική επεξεργασία έχουν διευρύνει το φάσμα των εφαρμογών που μπορούν να υποστηρίξουν, προσφέροντας μια πιο ανθρωποκεντρική προσέγγιση στην τεχνολογία. Είτε πρόκειται για την παροχή βοήθειας μέσω φωνητικών εντολών είτε για τη δημιουργία περιεχομένου βάσει σύνθετων περιγραφών, τα LLMs προσφέρουν μια γέφυρα μεταξύ της ανθρώπινης δημιουργικότητας και της τεχνολογικής καινοτομίας.

Σε αυτό το συνεχώς εξελισσόμενο τοπίο, η ενσωμάτωση των LLMs σε καθημερινές εφαρμογές και διαδικασίες δείχνει πως η τεχνολογία μπορεί να γίνει ακόμα πιο προσβάσιμη και εύχρηστη, διευκολύνοντας την ανθρώπινη αλληλεπίδραση και ενισχύοντας την παραγωγικότητα. Οι προοπτικές που ανοίγονται με τη χρήση των LLMs είναι τεράστιες, και το μέλλον της επικοινωνίας ανθρώπων και μηχανών φαίνεται να είναι πιο φωτεινό και γεμάτο δυνατότητες από ποτέ.

Η σημασία της επικοινωνίας χρηστών με LLMs γίνεται ολοένα και πιο φανερή σε διάφορους τομείς:

Βελτιωμένη Προσβασιμότητα: Τα LLMs μπορούν να διασφαλίσουν την απρόσκοπτη πρόσβαση στην πληροφορία και την επικοινωνία για άτομα με αναπηρίες, όπως προβλήματα όρασης ή ακοής.

Ενισχυμένη Εκπαίδευση: Η δυνατότητα των LLMs να εξατομικεύουν την εκπαίδευση και να παρέχουν προσαρμοσμένη υποστήριξη στους μαθητές και δύναται να φέρει επανάσταση στον τρόπο μάθησης και διδασκαλίας.

Ευφύεστερη Εξυπηρέτηση Πελατών: Τα LLMs μπορούν να αυτοματοποιήσουν εργασίες εξυπηρέτησης πελατών, να παρέχουν 24/7 υποστήριξη και να προσφέρουν εξατομικευμένες προτάσεις, βελτιώνοντας την εμπειρία των πελατών.

Ενισχυμένη Δημιουργικότητα: Τα LLMs μπορούν να λειτουργήσουν ως συνεργάτες στη δημιουργική διαδικασία, βοηθώντας συγγραφείς, καλλιτέχνες και μουσικούς να ξεπεράσουν τα δημιουργικά τους όρια.

Πρόσβαση σε Πληροφορίες: Η ικανότητα των LLMs να αντλούν πληροφορίες από διάφορες πηγές και να τις παρουσιάζουν με εύληπτο τρόπο μπορεί να ενδυναμώσει την λήψη τεκμηριωμένων αποφάσεων και την επίλυση προβλημάτων.

Καθώς η τεχνολογία των LLMs συνεχίζει να εξελίσσεται, η επικοινωνία με αυτούς τα μοντέλα θα γίνεται ολοένα και πιο φυσική και ουσιαστική. Η ικανότητά τους να κατανοούν τις αποχρώσεις της ανθρώπινης γλώσσας και να ανταποκρίνονται με ευφυή τρόπο θα ανοίξει νέες δυνατότητες σε διάφορους τομείς, διαμορφώνοντας το μέλλον της αλληλεπίδρασης ανθρώπου-υπολογιστή.

Εκτός από τα προαναφερθέντα, η επικοινωνία με LLMs φέρει στο προσκήνιο και ορισμένες ηθικές ανησυχίες που χρήζουν προσοχής. Η πιθανότητα παραπληροφόρησης, προκατάληψης και χειραγώγησης οφείλει να αντιμετωπιστεί με υπευθυνότητα και διαφάνεια. Η ανάπτυξη ηθικών πλαισίων και η υιοθέτηση υπεύθυνων πρακτικών αποτελούν απαραίτητες προϋποθέσεις για την χρήση τους.

1.1 Αντικείμενο της διπλωματικής

Στην παρούσα διπλωματική εργασία αναπτύχθηκε μια καινοτόμα εφαρμογή που επιτρέπει στους χρήστες να αλληλεπιδρούν με μεγάλα γλωσσικά μοντέλα (LLM) μέσω φωνητικής εισόδου. Συγκεκριμένα, ο χρήστης χρησιμοποιεί το μικρόφωνό του για να δώσει φωνητική εντολή ή ερώτηση, η οποία στη συνέχεια υποβάλλεται στο LLM. Το μοντέλο επεξεργάζεται την ερώτηση και παράγει μια κατάλληλη απάντηση. Η απάντηση αυτή δεν παραμένει μόνο σε κείμενο, αλλά μετατρέπεται σε φωνή μέσω ενός συστήματος σύνθεσης φωνής και απαγγέλεται στον χρήστη, προσφέροντας μια άμεση και φυσική μορφή αλληλεπίδρασης.

Ένα από τα σημαντικότερα και πιο εντυπωσιακά χαρακτηριστικά των μεγάλων γλωσσικών μοντέλων είναι η ικανότητά τους να παράγουν εικόνες. Η συγκεκριμένη εφαρμογή εκμεταλλεύεται αυτήν την ικανότητα, επιτρέποντας στους χρήστες να δημιουργούν εικόνες βάσει των περιγραφών που δίνουν φωνητικά μέσω του μικροφώνου τους. Η διαδικασία αυτή περιλαμβάνει την ανάλυση της περιγραφής από το LLM και την παραγωγή μιας εικόνας που ανταποκρίνεται όσο το δυνατόν πιστότερα στην περιγραφή του χρήστη.

Τελικά, η παραγόμενη εικόνα εμφανίζεται στην οθόνη του χρήστη. Είναι σημαντικό να σημειωθεί ότι η εικόνα αυτή δημιουργείται από το μηδέν, χωρίς να υπάρχει ήδη προϋπάρχουσα εικόνα, και το περιεχόμενό της διαμορφώνεται βάσει της κατανόησης του LLM για τα στοιχεία που επιθυμεί ο χρήστης να περιλαμβάνει. Αυτό σημαίνει ότι το μοντέλο πρέπει να αναγνωρίζει και να κατανοεί τα αντικείμενα, τις σκηνές και τις λεπτομέρειες που περιγράφει ο χρήστης για να παράξει μια εικόνα που βγάζει λογικό νόημα και ανταποκρίνεται στις

προσδοκίες του.

Με αυτήν την εφαρμογή, επιτυγχάνεται ένας συνδυασμός προηγμένων τεχνολογιών επεξεργασίας φυσικής γλώσσας, σύνθεσης φωνής και δημιουργίας εικόνων, προσφέροντας μια ολοκληρωμένη και διαδραστική εμπειρία στους χρήστες. Η εφαρμογή αυτή αποτελεί σημαντικό βήμα προόδου στην ανάπτυξη εργαλείων τεχνητής νοημοσύνης που ενισχύουν την επικοινωνία και τη δημιουργικότητα του ανθρώπου.

1.1.1 Συνεισφορά

Εδώ παραθέτουμε τις ενέργειες και μεθοδολογίες που χρησιμοποιήθηκαν κατά την υλοποίηση της διπλωματικής.

1. Μελετήθηκαν Large Language Models και ο τρόπος λειτουργίας τους
2. Αναπτύχθηκε η εφαρμογή με απλό περιβάλλον χρήστη για να την κάνει όσο πιο απλή στην χρήση για οποιονδήποτε χρήστη
3. Αξιολογήθηκε η επίδοση των αλγορίθμων και βρέθηκε ότι πιο σύνθετοι αλγόριθμοι speech to text μπορεί να είναι πιο ακριβείς αλλά πιο χρονοβόροι στην αναγνώριση του κειμένου και η μετατροπή του σε κείμενο.

1.2 Οργάνωση της Εργασίας

Βασικές Αρχές και Τεχνολογίες Η παρούσα διπλωματική εργασία είναι δομημένη σε διάφορα κεφάλαια, καθένα από τα οποία εξυπηρετεί ένα συγκεκριμένο σκοπό και καλύπτει διαφορετικές πτυχές του θέματος που εξετάζεται. Στόχος είναι η αναλυτική παρουσίαση και ανάλυση του αντικειμένου της έρευνας, η εξήγηση του σκοπού της εργασίας, η περιγραφή της μεθοδολογίας ανάπτυξης της εφαρμογής, καθώς και η παρουσίαση των αποτελεσμάτων και των συμπερασμάτων που προέκυψαν. Παρακάτω παρουσιάζεται η δομή της εργασίας με περιγραφή των θεμάτων που εξετάζονται σε κάθε κεφάλαιο.

Κεφάλαιο 1: Εισαγωγή και επεξήγηση Αντικειμένου Σε αυτήν την ενότητα γίνεται η εισαγωγή στο θέμα της διπλωματικής εργασίας και εξηγείται αναλυτικά το αντικείμενο της μελέτης. Παρουσιάζονται οι βασικές έννοιες και οι τεχνολογίες των μεγάλων γλωσσικών μοντέλων (LLM) στα σύγχρονα συστήματα μετατροπής κειμένου σε ομιλία (Text-to-Speech, TTS).

Κεφάλαιο 2: Βασικές Αρχές και Τεχνολογίες Σε αυτήν την ενότητα παρουσιάζονται οι πιο πρόσφατες έρευνες και τάσεις στον τομέα των συστημάτων μετατροπής κειμένου σε ομιλία και των μεγάλων γλωσσικών μοντέλων. Αναλύονται οι τεχνολογίες που χρησιμοποιούνται και οι πιθανές εφαρμογές τους στον τομέα της διπλωματικής εργασίας.

Κεφάλαιο 3: Τελευταίες Έρευνες και Τάσεις Σε αυτήν την ενότητα παρουσιάζεται η μεθοδολογία που ακολουθήθηκε για την ανάπτυξη της εφαρμογής. Αναλύονται οι διάφορες φάσεις της ανάπτυξης και οι τεχνικές που χρησιμοποιήθηκαν για την υλοποίηση του συστήματος Text-to-Speech.

Κεφάλαιο 4: Μεθοδολογία Ανάπτυξης της εφαρμογής Σε αυτήν την ενότητα περιλαμβάνονται τα αποτελέσματα της έρευνας και η ανάλυση των αποτελεσμάτων από τις δοκιμές της εφαρμογής. Επίσης, γίνεται αναφορά στις συνολικές επιδόσεις του συστήματος και στις πιθανές βελτιώσεις για μελλοντικές αναβαθμίσεις.

Κεφάλαιο 2

Βασικές Αρχές και Τεχνολογίες

2.1 Ιστορική αναδρομή

Κατά το πρώτο μισό του 20ού αιώνα, η επιστημονική κοινότητα επιδίωξε ενεργά την ανάπτυξη τεχνολογιών που θα μπορούσαν να μετατρέπουν κείμενο σε φωνητική απεικόνιση. Αυτή η προσπάθεια συμπεριλάμβανε την ανάπτυξη αυτόματων συστημάτων παραγωγής ομιλίας και την εφαρμογή νέων τεχνολογιών TTS (Text-to-Speech). Αυτές οι πρώτες προσπάθειες αποτέλεσαν την αφετηρία για την εξέλιξη των τεχνολογιών TTS που γνωρίζουμε σήμερα, προσφέροντας νέες δυνατότητες στην ανθρώπινη επικοινωνία και διασφαλίζοντας την πρόσβαση σε πληροφορίες μέσω προφορικής μορφής.

2.1.1 Μπορούν οι Μηχανές να Σκεφτούν;

Κατά το πρώτο μισό του 20ού αιώνα, η επιστημονική φαντασία εξοικείωσε τον κόσμο με την έννοια της τεχνητής νοημοσύνης. Μέχρι τη δεκαετία του 1950, είχαμε μια γενιά επιστημόνων, μαθηματικών και φιλοσόφων με την έννοια της τεχνητής νοημοσύνης (ή TN) ενσωματωμένη στο μυαλό τους. Ένας τέτοιος άνθρωπος ήταν ο Άλαν Τούρινγκ, ένας νεαρός Βρετανός πολυμαθής που εξέτασε τη μαθηματική πιθανότητα της τεχνητής νοημοσύνης. Ο Τούρινγκ υπέδειξε ότι οι άνθρωποι χρησιμοποιούν διαθέσιμες πληροφορίες καθώς και λογική για να λύσουν προβλήματα και να προβαίνουν σε αποφάσεις, οπότε γιατί δεν μπορούν οι μηχανές να κάνουν το ίδιο; Αυτό ήταν το λογικό πλαίσιο της εργασίας του το 1950, "Υπολογιστικά Μηχανήματα και Νοημοσύνη", στην οποία συζήτησε πώς να χτίσει ευφυείς μηχανές και πώς να δοκιμάσει τη νοημοσύνη τους. Δυστυχώς, τα λόγια δεν αρκούν. Τι εμπόδισε τον Τούρινγκ να αρχίσει να εργάζεται αμέσως; Πρώτον, οι υπολογιστές χρειαζόταν να αλλάξουν



Σχήμα 2.1: Alan Turing

θεμελιωδώς. Πριν από το 1949, οι υπολογιστές έλειπαν από ένα βασικό προαπαιτούμενο για τη νοημοσύνη: δεν μπορούσαν να αποθηκεύσουν εντολές, αλλά μόνο να τις εκτελούν. Με άλλα λόγια, οι υπολογιστές μπορούσαν να κάνουν ό,τι τους ζητούσαν, αλλά δεν μπορούσαν να θυμούνται τι έκαναν. Δεύτερον, οι υπολογιστές ήταν εξαιρετικά ακριβοί. Στις αρχές της δεκαετίας του 1950, το κόστος ενοικίασης ενός υπολογιστή ανερχόταν σε 200.000 δολάρια τον μήνα. Μόνο κορυφαία πανεπιστήμια και μεγάλες τεχνολογικές εταιρείες μπορούσαν να αντιμετωπίσουν αυτά τα άγνωστα ύδατα. Χρειάζοταν ένα προσχέδιο και υποστήριξη από προφίλ υψηλής προβολής για να πείσουν οι πηγές χρηματοδότησης ότι η μηχανική νοημοσύνη άξιζε τον κόπο. [1]

2.1.2 Η Διάσκεψη που τα Ξεκίνησε Όλα

Πέντε χρόνια αργότερα, το προσχέδιο ξεκίνησε μέσω του Logic Theorist 2 των Allen Newell, Cliff Shaw και Herbert Simon. Το Logic Theorist ήταν ένα πρόγραμμα σχεδιασμένο να μιμηθεί τις δεξιότητες επίλυσης προβλημάτων ενός ανθρώπου και χρηματοδοτήθηκε από την Research and Development (RAND) Corporation. Θεωρείται από πολλούς ως το πρώτο πρόγραμμα τεχνητής νοημοσύνης και παρουσιάστηκε στο Καλοκαιρινό Ερευνητικό Έργο για την Τεχνητή Νοημοσύνη του Dartmouth (DSRPAI) που διοργάνωσαν οι John McCarthy και Marvin Minsky το 1956. Σε αυτήν την ιστορική διάσκεψη, ο McCarthy, φανταζόμενος μια μεγάλη συνεργατική προσπάθεια, συγκέντρωσε κορυφαίους ερευνητές από διάφορα πεδία για μια ανοικτή συζήτηση για την τεχνητή νοημοσύνη, τον όρο που ίδιος δημιούργησε κατά τη διάρκεια του συνεδρίου. Δυστυχώς, η διάσκεψη δεν ήταν αυτό που περίμενε ο McCarthy. Οι άνθρωποι έφευγαν και έρχονταν όπως ήθελαν, και απέτυχαν να συμφωνήσουν σε κοινές



Σχήμα 2.2: Αριστερά Marvin Minsky, Claude Shannon, Ray Solomonoff και άλλοι επιστήμονες στο Dartmouth Summer Research Project on Artificial Intelligence (Photo: Margaret Minsky)

μεθόδους για το πεδίο. Παρά ταύτα, όλοι συμφώνησαν θερμά με το συναίσθημα ότι η TN ήταν εφικτή. Η σημασία αυτού του γεγονότος δεν μπορεί να υποτιμηθεί, καθώς έδωσε ώθηση στα επόμενα είκοσι χρόνια της έρευνας για την TN. 2.2

2.1.3 Επεξεργασία Φυσικής Γλώσσας

Είναι αρκετά δύσκολο να φανταστούμε ένα έξυπνο σύστημα όπως ένας φωνητικός βοηθός (alexa, siri) ή ένα chat-bot ή ένα σύστημα συστάσεων χωρίς τη φυσική επεξεργασία γλώσσας (Natural Language Processing - NLP). Όλα ξεκινούν με μια αρχική μονάδα που πρώτα ερμηνεύει τα δεδομένα (ήχος ή κείμενο) που παρέχονται και στη συνέχεια αρχίζει να δίνει νόημα στα δεδομένα, και μετά τη σωστή επεξεργασία των δεδομένων, ακολουθούνται βήματα (αλγοριθμικές πράξεις) από τη μηχανή για να εκτελέσει κάποιες απαντήσεις ή να ολοκληρώσει την διεργασία. Η NLP δεν εμπίπτει σε μία ειδική επιστήμη· μάλλον απο-



Σχήμα 2.3: Ένα δευτερόλεπτο φωνής που παράχθηκε μηχανικά

τελεί ξεχωριστός κλάδος διάφορων επιστημών, δηλαδή, της επιστήμης υπολογιστών, της μηχανικής πληροφορικής, της τεχνητής νοημοσύνης (TN) και της γλωσσολογίας. Το αντικείμενο της NLP είναι η αλληλεπίδραση μεταξύ ενός υπολογιστή και ανθρώπινων γλωσσών. Οι περιοχές της NLP περιλαμβάνουν αναγνώριση ομιλίας, αυτόματη μετάφραση κειμένου, αυτόματη περίληψη κειμένου, επισήμανση μέρους του λόγου, κ.λπ. Γενικά, η NLP χρησιμοποιείται σε πολλές εφαρμογές πραγματικού χρόνου όπως έξυπνα σπίτια, έξυπνα γραφεία όπως οι Alexa, Cortana, Siri και Google Assistant. Η ιστορία της NLP ξεκίνησε γενικά στη δεκαετία του 1950 και έχει διανύσει μακρύ δρόμο από τότε και έχει βελτιωθεί πολύ. Αυτή η εργασία συζητά την ιστορία της NLP, την εξέλιξή της, τα εργαλεία και τις τεχνικές της, καθώς και τις εφαρμογές της σε διάφορους τομείς. Η εργασία αναφέρεται επίσης στον ρόλο της μηχανικής μάθησης και των τεχνητών νευρωνικών δικτύων (ANNs) για τη βελτίωση της NLP.[2]

2.1.4 Σύνθεση Φωνής

Η σύνθεση φωνής από υπολογιστές και η ανάλυση φυσικής γλώσσας από υπολογιστές αποτελούν δύο σημαντικούς τομείς της επιστήμης των υπολογιστών και της τεχνητής νοημοσύνης. Στην μεταπτυχιακή σας εργασία, μπορείτε να αναφερθείτε στην ιστορία και τις τελευταίες έρευνες στους τομείς αυτούς με τις ακόλουθες πληροφορίες και πηγές.

2.1.5 Ιστορία της Σύνθεσης Φωνής από Υπολογιστές

Η σύνθεση φωνής από υπολογιστές, γνωστή και ως TTS (Text-to-Speech), έχει τις ρίζες της στη δεκαετία του 1930 με την πρώτη ηλεκτρομηχανική συσκευή, το Voder, που παρουσιάστηκε από την Bell Labs στην Παγκόσμια Έκθεση της Νέας Υόρκης το 1939. Οι πρώτες προσπάθειες για τη δημιουργία συνθετικής φωνής βασίζονταν σε απλές ακολουθίες προηχογραφημένων λέξεων ή φθόγγων. Στη δεκαετία του 1950, η εφεύρεση των πρώτων προγραμματισμένων συστημάτων υπολογιστών έδωσε ώθηση στη σύνθεση φωνής. Το 1968, η Bell Labs ανέπτυξε το σύστημα "Daisy Bell," το οποίο ήταν το πρώτο σύστημα που παρήγαγε κα-

τανοητή συνθετική φωνή μέσω ψηφιακής τεχνολογίας.[3] Μια σημαντική εξέλιξη ήρθε τη δεκαετία του 1980 με την εισαγωγή του συστήματος DECtalk, που χρησιμοποιήθηκε ευρέως σε συσκευές για άτομα με αναπηρίες. Από τότε, οι τεχνολογίες σύνθεσης φωνής έχουν εξελιχθεί σημαντικά με την ενσωμάτωση των μεθόδων μηχανικής μάθησης και των νευρωνικών δικτύων.

2.1.6 Ιστορία της Ανάλυσης Φυσικής Γλώσσας από Υπολογιστές

Η ανάλυση φυσικής γλώσσας από υπολογιστές (Natural Language Processing - NLP) ξεκίνησε ουσιαστικά με την ανάπτυξη του πρώτου προγράμματος μηχανικής μετάφρασης στη δεκαετία του 1950. Οι πρώτες προσπάθειες ήταν απλοϊκές, βασισμένες σε λεξικά και συντακτικούς κανόνες, αλλά παρουσίαζαν περιορισμένη επιτυχία λόγω της πολυπλοκότητας και της ασάφειας της φυσικής γλώσσας. Η δεκαετία του 1970 και του 1980 είδε την ανάπτυξη των πρώτων συστημάτων κατανόησης φυσικής γλώσσας, όπως το SHRDLU, που μπορούσε να αλληλεπιδρά με χρήστες σε περιορισμένα περιβάλλοντα. Η πρόοδος στη θεωρητική γλωσσολογία και η εισαγωγή στατιστικών μεθόδων στα τέλη του 1980 και 1990 βελτίωσε σημαντικά τις επιδόσεις των συστημάτων NLP.

2.1.7 Τελευταίες Έρευνες στη Σύνθεση Φωνής

Η σύγχρονη έρευνα στη σύνθεση φωνής επικεντρώνεται κυρίως στις τεχνικές που βασίζονται σε νευρωνικά δίκτυα. Μία από τις πιο σημαντικές εξελίξεις είναι το WaveNet, που αναπτύχθηκε από την DeepMind το 2016. Το WaveNet χρησιμοποιεί μοντέλα βαθιάς μάθησης για την παραγωγή φωνητικών κυμάτων από το μηδέν, προσφέροντας εξαιρετικά φυσικό ήχο.[4] Πιο πρόσφατα, η χρήση των μετασχηματιστικών μοντέλων (transformers), όπως το Tacotron 2 της Google, έχει επιτρέψει την ακόμη πιο ακριβή και φυσική σύνθεση φωνής. Αυτά τα μοντέλα εκπαιδεύονται σε μεγάλες βάσεις δεδομένων και μπορούν να μιμηθούν την ανθρώπινη φωνή με εξαιρετική ακρίβεια και ευχρηστία.

2.1.8 Τελευταίες Έρευνες στην Ανάλυση Φυσικής Γλώσσας

Στην ανάλυση φυσικής γλώσσας, τα μοντέλα βάθους μάθησης και ιδιαίτερα τα μοντέλα μετασχηματιστών έχουν φέρει επανάσταση στον τομέα. Το BERT (Bidirectional Encoder Representations from Transformers), που αναπτύχθηκε από την Google το 2018, αποτελεί

μια από τις σημαντικότερες εξελίξεις. Το BERT και τα επόμενα μοντέλα του, όπως το GPT-3 της OpenAI[5], χρησιμοποιούν μεγάλης κλίμακας αδράνεια μάθηση για να κατανοήσουν και να παράγουν κείμενα με τρόπους που ήταν αδιανόητοι πριν από λίγα χρόνια.[6] Οι πρόσφατες έρευνες επικεντρώνονται επίσης στη βελτίωση της πολυγλωσσικής κατανόησης και της μετάφρασης, με έργα όπως το mBERT και το XLM-R της Facebook AI, που μπορούν να χειριστούν δεκάδες γλώσσες ταυτόχρονα.[7] Η σύνθεση φωνής και η ανάλυση φυσικής γλώσσας από υπολογιστές συνεχίζουν να εξελίσσονται, ενσωματώνοντας νέες τεχνολογίες και μεθόδους που υπόσχονται να βελτιώσουν περαιτέρω την αλληλεπίδραση ανθρώπου-υπολογιστή. Η κατανόηση της ιστορίας και των τελευταίων ερευνών στους τομείς αυτούς είναι κρίσιμη για την ανάπτυξη και την εφαρμογή αυτών των τεχνολογιών στο μέλλον. Για να εμπλουτίσετε το θεωρητικό κομμάτι της διπλωματικής σας εργασίας σχετικά με τη σύνθεση φωνής από υπολογιστές και την ανάλυση φυσικής γλώσσας από υπολογιστές.

2.2 Εισαγωγή στις Βασικές Αρχές και Έννοιες

Στην ενότητα αυτή γίνεται εισαγωγή στις βασικές αρχές και έννοιες που σχετίζονται με τη μετατροπή κειμένου σε ομιλία (TTS). Τα συστήματα TTS αποτελούν σύνθετα συστήματα λογισμικού που επεξεργάζονται γραπτό κείμενο και το μετατρέπουν αυτόματα σε ανθρώπινη φωνητική παραγωγή. Βασικές αρχές των συστημάτων αυτών περιλαμβάνουν την αναγνώριση και την ερμηνεία της γραπτής γλώσσας, τη σύνθεση φωνητικών μονάδων και την προσαρμογή της παραγωγής ομιλίας σύμφωνα με τις γλωσσικές και ποιοτικές απαιτήσεις.

2.2.1 Ορισμοί και Βασικές Αρχές

Η Σύνθεση Φωνής (Text-to-Speech, TTS) είναι μια τεχνολογία που μετατρέπει γραπτό κείμενο σε φυσικά ηχητική μορφή. Αυτή η τεχνολογία χρησιμοποιείται ευρέως σε εφαρμογές όπως οι ψηφιακοί βοηθοί, οι συσκευές πλοήγησης, και οι βοηθητικές τεχνολογίες για άτομα με προβλήματα όρασης ή αναπηρίες. Η λειτουργία της σύνθεσης φωνής περιλαμβάνει διάφορα στάδια επεξεργασίας και χρήσης τεχνολογιών μηχανικής μάθησης.

2.2.2 Πώς Λειτουργεί η Σύνθεση Φωνής (TTS)

Η διαδικασία της σύνθεσης φωνής μπορεί να αναλυθεί σε τρία κύρια στάδια:

1. Προεπεξεργασία Κειμένου Σε αυτό το στάδιο, το γραπτό κείμενο μετατρέπεται σε μια μορφή που μπορεί να επεξεργαστεί περαιτέρω. Περιλαμβάνει την επεξεργασία συντομογραφιών, αριθμών και συμβόλων, ώστε να μετατραπούν σε πλήρεις λέξεις ή φράσεις. Για παράδειγμα, το "Dr." μετατρέπεται σε "Doctor".[3]
2. Γλωσσολογική Ανάλυση Το κείμενο υποβάλλεται σε γλωσσολογική ανάλυση για να κατανοηθούν τα συντακτικά και σημασιολογικά χαρακτηριστικά του. Αυτό περιλαμβάνει τη μορφολογική ανάλυση (ανάλυση της μορφής των λέξεων), τη συντακτική ανάλυση (πώς οι λέξεις συνδέονται μεταξύ τους) και την προσοδική ανάλυση (καθορισμός τονισμού και προσωδίας για φυσικό ήχο). Το σύστημα καθορίζει επίσης την προφορά των λέξεων, χρησιμοποιώντας λεξικά προφοράς ή κανόνες γραμματικής και φωνολογίας.
3. Δημιουργία Φωνής Σε αυτό το στάδιο, το αναλυθέν κείμενο μετατρέπεται σε ηχητικό σήμα. Υπάρχουν διάφορες τεχνικές για την παραγωγή της φωνής: Συγκολλητική Σύνθεση (Concatenative Synthesis): Χρησιμοποιεί ηχογραφημένα τμήματα φωνής από ανθρώπινους ομιλητές. Αυτά τα τμήματα (π.χ., φθόγγοι, συλλαβές, λέξεις) συνδυάζονται για να παραχθεί ο ήχος της επιθυμητής πρότασης. Παραμετρική Σύνθεση (Parametric Synthesis): Χρησιμοποιεί μοντέλα που παράγουν ήχο με βάση παραμέτρους όπως η συχνότητα και η ένταση. Ένα δημοφιλές μοντέλο σε αυτή την κατηγορία είναι το Hidden Markov Model (HMM). Σύνθεση με Νευρωνικά Δίκτυα (Neural TTS): Χρησιμοποιεί προηγμένα νευρωνικά δίκτυα για τη δημιουργία φυσικού ήχου. Μια σημαντική εξέλιξη σε αυτή την κατηγορία είναι το WaveNet, που αναπτύχθηκε από την DeepMind το 2016. Το WaveNet χρησιμοποιεί μοντέλα βαθιάς μάθησης για την παραγωγή φωνητικών κυμάτων από το μηδέν, προσφέροντας εξαιρετικά φυσικό ήχο.[4]

2.3 Βασικοί Μηχανισμοί Πίσω από τη Σύνθεση Φωνής

Στο παρόν κεφάλαιο γίνεται εμβάθυνση στους κύριους τεχνικούς μηχανισμούς που απαιτούνται για τη σύνθεση φωνής στα συστήματα Text-to-Speech (TTS). Η φωνολογική ανάλυση εστιάζει στην αναγνώριση και την αποσαφήνιση των φωνητικών μονάδων μιας γλώσσας, όπως φωνήεντα και σύμφωνα, καθώς και την προσαρμογή τους στις γλωσσικές κανονιστικές. Από την άλλη, η φωνητική ανάλυση εξετάζει τους φυσικούς μηχανισμούς παραγωγής ήχου και πώς αυτοί μπορούν να προσομοιωθούν μέσω υπολογιστικών μεθόδων. Η κατα-

νόηση και η εφαρμογή αυτών των μηχανισμών είναι θεμελιώδης για την επίτευξη φυσικής και εύληπτης αναπαραγωγής φωνής μέσω συστημάτων TTS.

2.3.1 Φωνολογική και Φωνητική Ανάλυση

Η φωνολογική ανάλυση ασχολείται με τον τρόπο που οι ήχοι της γλώσσας οργανώνονται και χρησιμοποιούνται για να παράγουν λέξεις. Η φωνητική ανάλυση ασχολείται με τις φυσικές ιδιότητες αυτών των ήχων και πώς παράγονται από τα ανθρώπινα όργανα ομιλίας.

2.3.2 Γλωσσολογική και Σημασιολογική Ανάλυση

Η γλωσσολογική ανάλυση εξασφαλίζει ότι η συντακτική δομή του κειμένου διατηρείται κατά τη μετατροπή του σε ομιλία. Η σημασιολογική ανάλυση βοηθάει στην κατανόηση του νοήματος του κειμένου, προσδιορίζοντας τον κατάλληλο τονισμό και την προσωδία.

2.3.3 Τεχνητά Νευρωνικά Δίκτυα

Τα σύγχρονα συστήματα TTS χρησιμοποιούν νευρωνικά δίκτυα για την προσαρμογή και την παραγωγή πιο φυσικού και ρεαλιστικού ήχου. Τα μοντέλα όπως το Tacotron 2 της Google χρησιμοποιούν ακολουθιακά νευρωνικά δίκτυα (Sequence-to-Sequence Models) για την παραγωγή φωνής με ανθρώπινη προσωδία και φυσικότητα.[8] Η σύνθεση φωνής αποτελεί μια σύνθετη και πολυδιάστατη τεχνολογία, που βασίζεται σε διάφορους επιστημονικούς και τεχνολογικούς τομείς, συμπεριλαμβανομένων της γλωσσολογίας, της φωνητικής, της μηχανικής μάθησης και της τεχνητής νοημοσύνης. Οι συνεχείς έρευνες και βελτιώσεις σε αυτούς τους τομείς συμβάλλουν στην ανάπτυξη πιο φυσικών και αποτελεσματικών συστημάτων σύνθεσης φωνής.

2.3.4 Ανάλυση Φυσικής Γλώσσας

Η Ανάλυση Φυσικής Γλώσσας (Natural Language Processing - NLP) είναι ένας τομέας της τεχνητής νοημοσύνης που επικεντρώνεται στην αλληλεπίδραση μεταξύ των υπολογιστών και της ανθρώπινης γλώσσας. Συγκεκριμένα, η NLP επιδιώκει να επιτρέψει στους υπολογιστές να κατανοήσουν, να ερμηνεύσουν και να παράγουν ανθρώπινη γλώσσα με τρόπο που είναι τόσο φυσικός και αποτελεσματικός όσο η ανθρώπινη επικοινωνία.

2.3.5 Ορισμός της NLP

Η NLP αφορά την ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να επεξεργάζονται και να κατανοούν μεγάλα ποσά φυσικής γλώσσας. Αυτό περιλαμβάνει την ανάλυση κειμένων, τη δημιουργία νέων κειμένων, τη μετάφραση μεταξύ γλωσσών, την ανάλυση συναισθημάτων, την απάντηση σε ερωτήσεις και πολλές άλλες εφαρμογές.

2.4 Κύριες Προκλήσεις της NLP

Η επεξεργασία φυσικής γλώσσας (Natural Language Processing - NLP) αντιμετωπίζει πολλές σημαντικές προκλήσεις λόγω της πολυπλοκότητας και της αόριστης φύσης της ανθρώπινης γλώσσας. Αυτές οι προκλήσεις περιλαμβάνουν την αναγνώριση και την αποσαφήνιση της σημασίας των λέξεων στο πλαίσιο του περιβάλλοντος τους, την κατανόηση της γραμματικής δομής και της σημασιολογικής συνεκτίμησης προτάσεων, καθώς και την αντιμετώπιση της ποικιλίας στυλ και εκφραστικών μορφών που μπορεί να παρουσιάζει η γλώσσα. Επιπλέον, η πρόοδος στην επίλυση αυτών των προκλήσεων συχνά απαιτεί την εφαρμογή σύγχρονων τεχνικών μηχανικής μάθησης και ανάλυσης δεδομένων, προκειμένου να επιτευχθούν υψηλά επίπεδα ακρίβειας και αξιοπιστίας στις εφαρμογές NLP.

2.4.1 Ασάφεια και Πολυπλοκότητα της Φυσικής Γλώσσας

Η ανθρώπινη γλώσσα είναι πλούσια σε ασάφειες και σημασιολογικές πολυπλοκότητες. Λέξεις μπορεί να έχουν πολλαπλές σημασίες, και οι προτάσεις μπορούν να ερμηνευτούν με διαφορετικούς τρόπους ανάλογα με τα συμφραζόμενα.[9]

2.4.2 Συντακτική και Σημασιολογική Ανάλυση

Η ανάλυση της συντακτικής δομής και της σημασιολογικής σημασίας μιας πρότασης είναι μια περίπλοκη διαδικασία που απαιτεί την κατανόηση των κανόνων γραμματικής και της σημασιολογικής σχέσης μεταξύ των λέξεων.[10]

2.4.3 Πολυγλωσσικότητα και Πολιτισμικές Διαφορές

Η NLP πρέπει να αντιμετωπίσει τη δυσκολία της πολυγλωσσικότητας και των πολιτισμικών διαφορών, κάτι που περιπλέκει τη δημιουργία μοντέλων που λειτουργούν εξίσου καλά

σε πολλές γλώσσες και πολιτισμούς.[11]

2.5 Βασικές Τεχνικές της NLP

1. Προεπεξεργασία Κειμένου Περιλαμβάνει διαδικασίες όπως η αφαίρεση των σταματημένων λέξεων (stop words), η αποστέλεση (stemming), και η λεματοποίηση (lemmatization). Αυτές οι τεχνικές βοηθούν στη μείωση της πολυπλοκότητας και στην προετοιμασία του κειμένου για περαιτέρω ανάλυση.[9]
2. Συντακτική Ανάλυση (Parsing) Η συντακτική ανάλυση περιλαμβάνει την αναγνώριση της δομής μιας πρότασης, προσδιορίζοντας τις σχέσεις μεταξύ των λέξεων σύμφωνα με τους κανόνες της γραμματικής. Χρησιμοποιούνται τεχνικές όπως τα Dependency Parsing και Constituency Parsing.[10]
3. Σημασιολογική Ανάλυση Η σημασιολογική ανάλυση επικεντρώνεται στην κατανόηση του νοήματος των λέξεων και των προτάσεων. Αυτό περιλαμβάνει τεχνικές όπως η αναγνώριση οντοτήτων (Named Entity Recognition - NER) και η ανάλυση συναισθημάτων (Sentiment Analysis).[9]

2.5.1 Στατιστικά Μοντέλα και Μοντέλα Μηχανικής Μάθησης

Τα μοντέλα αυτά χρησιμοποιούν στατιστικές μεθόδους για την πρόβλεψη και την ανάλυση κειμένων. Περιλαμβάνουν τα Hidden Markov Models (HMMs), τα Naive Bayes Classifiers, και τα Support Vector Machines (SVMs).[10]

2.5.2 Νευρωνικά Δίκτυα και Μοντέλα Βαθιάς Μάθησης

Τα σύγχρονα μοντέλα NLP χρησιμοποιούν νευρωνικά δίκτυα και τεχνικές βαθιάς μάθησης για την κατανόηση και παραγωγή κειμένων. Το BERT (Bidirectional Encoder Representations from Transformers) της Google είναι ένα από τα πιο προηγμένα μοντέλα που χρησιμοποιούνται σήμερα. Το BERT εκπαιδεύεται σε μεγάλης κλίμακας δεδομένα και μπορεί να κατανοήσει τα συμφραζόμενα καλύτερα από τα παραδοσιακά μοντέλα.[6] Η NLP είναι ένας συνεχώς εξελισσόμενος τομέας, με πολλές προκλήσεις και τεχνικές που στοχεύουν στη βελτίωση της κατανόησης και της επεξεργασίας της ανθρώπινης γλώσσας από τους υπολογιστές. Οι πρόσφατες εξελίξεις, όπως τα νευρωνικά δίκτυα και τα μοντέλα βαθιάς μάθησης,

έχουν φέρει επανάσταση στον τομέα, επιτρέποντας πιο φυσική και ακριβή αλληλεπίδραση μεταξύ ανθρώπων και μηχανών.

2.6 Ιστορική Ανασκόπηση

Η ιστορική ανασκόπηση των φυσικών γλωσσών και των συστημάτων μετατροπής κειμένου σε ομιλία (Text-to-Speech, TTS) αποτελεί ένα σημαντικό κομμάτι στην εξέλιξη της τεχνολογίας. Η έναρξη των φυσικών γλωσσών χρονολογείται από τη δεκαετία του 1950, όπου πρωτοεμφανίστηκαν μελέτες για την ανάλυση και την κατασκευή συστημάτων που μπορούσαν να κατανοήσουν και να παράγουν φυσική γλώσσα. Από τότε, η επιστημονική κοινότητα έχει εργαστεί επί μακρόν για την ανάπτυξη αλγορίθμων και τεχνικών που να επιτρέπουν την ακριβή κατανόηση και παραγωγή γλωσσικού περιεχομένου από υπολογιστές.

Η εξέλιξη των συστημάτων TTS ξεκίνησε από απλές μηχανικές μεθόδους για τη σύνθεση ομιλίας μέχρι την εποχή της μηχανικής μάθησης και της νευρωνικής δικτύωσης. Αρχικά, οι συστηματικές προσπάθειες εστιάστηκαν στη βελτίωση της φυσικότητας της παραγόμενης ομιλίας και στην προσαρμογή των συστημάτων σε διάφορες γλωσσικές παραλλαγές και ποιοτικά επίπεδα. Με την πρόοδο της τεχνητής νοημοσύνης και της μηχανικής μάθησης, οι προσεγγίσεις βασίζονται σήμερα σε σύγχρονες μεθόδους όπως η βαθιά μάθηση, η οποία έχει επιτρέψει τη δημιουργία εξαιρετικά ρεαλιστικών και φυσικών φωνητικών εξόδων.

Η σύγκλιση της έρευνας στα πεδία της NLP και των συστημάτων TTS έχει οδηγήσει σε καινοτόμες εφαρμογές σε τομείς όπως οι προσωπικοί βοηθοί, η εκπαίδευση, η ψυχαγωγία και η πρόσβαση στην πληροφορία για άτομα με αναπηρία. Η συνεχής ανάπτυξη και η εξέλιξη αυτών των τεχνολογιών υπόσχονται να αλλάξουν ριζικά τον τρόπο που αλληλεπιδρούμε με τους υπολογιστές και την τεχνολογία γενικότερα στο μέλλον.

2.6.1 Σύνθεση Φωνής (TTS)

- 1939 - Voder Στη δεκαετία του 1930, η Bell Labs παρουσίασε το Voder, την πρώτη ηλεκτρομηχανική συσκευή που μπορούσε να παράγει ανθρώπινη φωνή. Παρουσιάστηκε στην Παγκόσμια Έκθεση της Νέας Υόρκης το 1939 και θεωρείται η πρώτη σημαντική καινοτομία στον τομέα της σύνθεσης φωνής.[3]
- 1968 - Daisy Bell Η Bell Labs ανέπτυξε το σύστημα Daisy Bell, το οποίο ήταν το πρώτο σύστημα που παρήγαγε κατανοητή συνθετική φωνή μέσω ψηφιακής τεχνολο-

γίας. Αυτό το σύστημα χρησιμοποίησε βασικές τεχνικές σύνθεσης φωνής για να παράγει τη γνωστή μελωδία "Daisy Bell".[3]

- 1990s - Festival και Flite Στη δεκαετία του 1990, το Πανεπιστήμιο του Εδιμβούργου ανέπτυξε το Festival Speech Synthesis System, ένα σύστημα σύνθεσης φωνής ανοιχτού κώδικα. Αργότερα, αναπτύχθηκε το Flite, μια ελαφριά έκδοση του Festival, η οποία χρησιμοποιείται σε εφαρμογές με περιορισμένους πόρους.
- 2016 - WaveNet To 2016, η DeepMind παρουσίασε το WaveNet, ένα σύστημα σύνθεσης φωνής που χρησιμοποιεί βαθιά νευρωνικά δίκτυα για την παραγωγή φυσικού ήχου. Το WaveNet δημιουργεί φωνητικά κύματα από το μηδέν, προσφέροντας εξαιρετικά φυσικό ήχο.[4]

2.6.2 Ανάλυση Φυσικής Γλώσσας (NLP)

- 1950 - Test Turing O Alan Turing εισήγαγε το "Turing Test", ένα κριτήριο για να εκτιμήσει αν μια μηχανή μπορεί να επιδείξει νοημοσύνη ισοδύναμη ή αδιάκριτη από αυτήν ενός ανθρώπου. Αυτό το κριτήριο αποτέλεσε θεμέλιο για την ανάπτυξη της NLP.
- 1960s - ELIZA Στη δεκαετία του 1960, ο Joseph Weizenbaum ανέπτυξε το ELIZA, ένα από τα πρώτα προγράμματα επεξεργασίας φυσικής γλώσσας. Το ELIZA μπορούσε να μιμηθεί μια συζήτηση με έναν ψυχολόγο, απαντώντας σε ερωτήσεις με προκαθορισμένες φράσεις.
- 1990s - Στατιστικά Μοντέλα Η δεκαετία του 1990 είδε τη μετάβαση από κανόνες βάσει γραμματικής σε στατιστικά μοντέλα για την επεξεργασία της φυσικής γλώσσας. Η χρήση των Hidden Markov Models (HMMs) και των Naive Bayes έγινε δημοφιλής για εργασίες όπως η αναγνώριση ομιλίας και η ανάλυση κειμένου.
- 2013 - Word2Vec Η Google παρουσίασε το Word2Vec, ένα μοντέλο που δημιουργεί διανυσματικές αναπαραστάσεις λέξεων. Αυτό το μοντέλο έκανε επανάσταση στην κατανόηση της σημασιολογικής σχέσης μεταξύ λέξεων, επιτρέποντας στα συστήματα να κατανοούν τις λέξεις βάσει των συμφραζόμενων τους.[12]
- 2018 - BERT To 2018, η Google ανέπτυξε το BERT (Bidirectional Encoder Representations from Transformers), ένα από τα πιο προηγμένα μοντέλα NLP. Το BERT μπορεί

να κατανοήσει το πλαίσιο των λέξεων και των φράσεων σε μια πρόταση, επιτυγχάνοντας κορυφαία αποτελέσματα σε πολλές εργασίες NLP.[6]

- 2019 - GPT-3 Το 2019, η OpenAI παρουσίασε το GPT-3, ένα από τα μεγαλύτερα και πιο ισχυρά γλωσσικά μοντέλα. Το GPT-3 μπορεί να δημιουργεί κείμενο που είναι σχεδόν αδιακρίτως από κείμενο που γράφεται από ανθρώπους, έχοντας εφαρμογές σε διάφορους τομείς από την αυτόματη συγγραφή έως τη δημιουργία διαλόγων.[5] Οι τομείς της σύνθεσης φωνής και της ανάλυσης φυσικής γλώσσας έχουν δει σημαντικές εξελίξεις από την πρώτη εμφάνισή τους. Από τις πρώτες μηχανικές συσκευές και τα απλά προγράμματα διαλόγου, έχουμε φτάσει σε προηγμένα νευρωνικά δίκτυα και πολυγλωσσικά μοντέλα που επιτρέπουν στους υπολογιστές να κατανοούν και να παράγουν ανθρώπινη γλώσσα με υψηλή ακρίβεια και φυσικότητα. Οι συνεχείς έρευνες και καινοτομίες υπόσχονται ακόμα πιο συναρπαστικές εφαρμογές και βελτιώσεις στο μέλλον.

2.7 Θεωρητικό Υπόβαθρο και Τεχνολογίες

2.7.1 Φωνολογική και Φωνητική Ανάλυση

Η φωνολογία και η φωνητική παίζουν κρίσιμο ρόλο στην ανάπτυξη συστημάτων σύνθεσης φωνής (TTS). Αυτοί οι τομείς της γλωσσολογίας είναι απαραίτητοι για την κατανόηση και τη δημιουργία φυσικού και κατανοητού ήχου.

2.7.2 Φωνολογία

Η φωνολογία ασχολείται με τη μελέτη των φθόγγων (ήχων) και των κανόνων που διέπουν τη χρήση τους σε μια συγκεκριμένη γλώσσα.

2.7.3 Αναγνώριση Φωνητικών Μοτίβων

Η φωνολογία αναλύει πώς οι ήχοι αλλάζουν ανάλογα με το περιβάλλον τους μέσα στη λέξη και τη φράση. Για παράδειγμα, οι κανόνες για τη χρήση των διαφορετικών φωνητικών παραλλαγών είναι απαραίτητοι για την ορθή προφορά των λέξεων.

2.7.4 Διαχείριση Τονισμού και Προσοδίας

Οι φωνολογικοί κανόνες βοηθούν στον καθορισμό του κατάλληλου τονισμού και της προσωδίας (ρυθμός και μελωδία της ομιλίας). Αυτό περιλαμβάνει την τοποθέτηση του τόνου σε λέξεις και τη διαμόρφωση της φράσης, προσφέροντας μια φυσική αίσθηση στην ομιλία.

2.7.5 Φωνητική

Η φωνητική ασχολείται με τη φυσική παραγωγή και τις ακουστικές ιδιότητες των φθόγων. Στα TTS συστήματα, η φωνητική είναι ζωτικής σημασίας για:

1. Αναπαράσταση Ήχων Η φωνητική παρέχει τις βασικές πληροφορίες για το πώς να παράγονται και να συνδυάζονται οι φθόγοι, χρησιμοποιώντας το Διεθνές Φωνητικό Αλφάβητο (IPA) για την ακριβή αναπαράσταση των ήχων.
2. Ανάλυση και Σύνθεση Ήχων Η φωνητική ανάλυση βοηθά στην κατανόηση των φυσικών ιδιοτήτων των φθόγων, όπως η συχνότητα, η διάρκεια και η ένταση. Αυτό επιτρέπει την ακριβή αναπαραγωγή των ήχων σε ένα TTS σύστημα.

2.8 Ενσωμάτωση της Φωνολογίας και της Φωνητικής στα Σύγχρονα TTS Συστήματα

Τα σύγχρονα TTS συστήματα χρησιμοποιούν προχωρημένες τεχνικές για να ενσωματώσουν τις αρχές της φωνολογίας και της φωνητικής, επιτυγχάνοντας έτσι πιο φυσικό και κατανοητό ήχο.

2.8.1 Στάδια Ενσωμάτωσης

1. Προεπεξεργασία Κειμένου Τα συστήματα TTS ξεκινούν με την προεπεξεργασία του κειμένου για να αντιμετωπιστούν ζητήματα όπως οι συντομογραφίες, τα αριθμητικά και τα ειδικά σύμβολα, μετατρέποντας τα σε πλήρεις λέξεις ή φράσεις. Αυτό το στάδιο εξασφαλίζει ότι το κείμενο είναι έτοιμο για περαιτέρω γλωσσολογική επεξεργασία.
2. Γλωσσολογική Ανάλυση Περιλαμβάνει τη μορφολογική και συντακτική ανάλυση του κειμένου. Τα φωνολογικά χαρακτηριστικά των λέξεων αναγνωρίζονται και χρησιμο-

ποιούνται για την καθοδήγηση της προφοράς και της προσωδίας. Η φωνολογική ανάλυση εξασφαλίζει ότι οι λέξεις τονίζονται σωστά και οι φθόγγοι προφέρονται με ακρίβεια.

3. Παραγωγή Ήχου Τα συστήματα χρησιμοποιούν φωνητικά δεδομένα για την παραγωγή των ήχων. Υπάρχουν διάφορες τεχνικές για την παραγωγή του ήχου:

- Συγκολλητική Σύνθεση (Concatenative Synthesis): Χρησιμοποιεί αποθηκευμένα τμήματα ηχογραφημένης φωνής (όπως φθόγγοι, συλλαβές ή λέξεις) που συνδέονται μεταξύ τους για να δημιουργηθεί η επιθυμητή πρόταση.[13]
- Παραμετρική Σύνθεση (Parametric Synthesis): Χρησιμοποιεί μαθηματικά μοντέλα για να δημιουργήσει ήχους από παραμέτρους, βασισμένα σε χαρακτηριστικά όπως η συχνότητα και η ένταση.[14]
- Σύνθεση με Νευρωνικά Δίκτυα (Neural TTS): Χρησιμοποιεί προηγμένα νευρωνικά δίκτυα, όπως το WaveNet, για την παραγωγή εξαιρετικά φυσικού ήχου από το μηδέν.[15]

2.8.2 Παράδειγμα: WaveNet

Το WaveNet, ένα μοντέλο που αναπτύχθηκε από την DeepMind, χρησιμοποιεί ένα νευρωνικό δίκτυο για να μάθει τα φωνολογικά και φωνητικά πρότυπα από μεγάλα σύνολα δεδομένων φωνής. Το WaveNet δημιουργεί φωνητικά κύματα που προσομοιώνουν τη φυσική παραγωγή της ομιλίας, ενσωματώνοντας φωνολογικές και φωνητικές πληροφορίες για την παραγωγή φυσικού ήχου.[4] Με την ενσωμάτωση φωνολογικών και φωνητικών γνώσεων, τα σύγχρονα TTS συστήματα μπορούν να παράγουν φωνή που είναι όχι μόνο κατανοητή αλλά και φυσική, προσεγγίζοντας την ανθρώπινη ομιλία σε μεγάλο βαθμό. Αυτές οι τεχνικές συνεχώς βελτιώνονται, καθιστώντας τα TTS συστήματα πιο αποτελεσματικά και χρήσιμα σε πολλές εφαρμογές.

2.9 Συντακτική και Σημασιολογική Ανάλυση

Η συντακτική και σημασιολογική ανάλυση αποτελούν κρίσιμα στάδια στη φυσική γλωσσική επεξεργασία (Natural Language Processing - NLP), που αφορούν την κατανόηση και την ανάλυση της γλώσσας από τους υπολογιστές.

Η συντακτική ανάλυση ασχολείται με τη γραμματική δομή της φράσης ή του κειμένου, με στόχο την αναγνώριση των συντακτικών στοιχείων όπως τα ρήματα, τα ουσιαστικά, οι επιθέτα και οι προτάσεις. Αυτή η ανάλυση βασίζεται σε γραμματικούς κανόνες και αλγορίθμους, οι οποίοι επιτρέπουν τη διάκριση της δομής της φράσης και την κατανόηση των συνδυασμών λέξεων που σχηματίζουν νόημα.

Από την άλλη πλευρά, η σημασιολογική ανάλυση επικεντρώνεται στην κατανόηση του νοήματος πίσω από τις φράσεις και τα κείμενα. Αυτή η διαδικασία χρησιμοποιεί διάφορες τεχνικές όπως οντολογίες, θεωρίες σημασιολογικής διακριτής κατηγορίας και αναπαραστάσεις κειμένου σε διανυσματικό χώρο, προκειμένου να κατανοήσει το πραγματικό νόημα και την περιεχόμενη πληροφορία που μεταφέρεται μέσω της γλώσσας.

Συνολικά, η συνδυασμένη χρήση συντακτικής και σημασιολογικής ανάλυσης είναι κρίσιμη για την ανάπτυξη προηγμένων εφαρμογών NLP, όπως η αυτόματη μετάφραση, η ανάλυση συναισθημάτων σε κείμενα, η αυτόματη παραγωγή κειμένου και οι εξατομικευμένοι προσωπικοί βοηθοί.

2.9.1 Τεχνικές NLP για Συντακτική και Σημασιολογική Ανάλυση Κειμένου

Η Φυσική Επεξεργασία Γλώσσας (NLP) αξιοποιεί διάφορες τεχνικές για την ανάλυση κειμένου σε συντακτικό και σημασιολογικό επίπεδο. Ακολουθούν μερικές από τις πιο συνηθισμένες:

2.9.2 Συντακτική Ανάλυση:

- **Μορφολογική Ανάλυση:** Διασπά το κείμενο σε βασικές μονάδες, όπως λέξεις και φράσεις, και τα ταξινομεί σε γραμματικές κατηγορίες (π.χ., ουσιαστικό, ρήμα, επίθετο).
- **Μορφολογική Ανάλυση** Η μορφολογική ανάλυση είναι ένα κρίσιμο στάδιο στην επεξεργασία φυσικής γλώσσας (NLP) που ασχολείται με τη διάσπαση του κειμένου σε βασικές μονάδες, όπως λέξεις και φράσεις, και την ταξινόμησή τους σε γραμματικές κατηγορίες. Αυτό το βήμα είναι απαραίτητο για την κατανόηση της δομής και της σημασίας του κειμένου και αποτελεί τη βάση για πολλές άλλες διαδικασίες επεξεργασίας κειμένου. Αρχικά, η μορφολογική ανάλυση περιλαμβάνει τη διάσπαση του συνεχούς

κειμένου σε ξεχωριστές λέξεις. Αυτό γίνεται μέσω της ανάλυσης των διαστημάτων και των σημείων στίξης που διαχωρίζουν τις λέξεις. Ακολουθεί η αναγνώριση των ριζών των λέξεων και των μορφολογικών παραλλαγών τους. Για παράδειγμα, η λέξη "τρέχω" μπορεί να έχει παράγωγες μορφές όπως "έτρεξα" και "τρέχεις". Αυτές οι παραλλαγές αναλύονται και αναγνωρίζονται για να κατανοηθεί η βασική ρίζα της λέξης. Στη συνέχεια, οι λέξεις ταξινομούνται σε γραμματικές κατηγορίες όπως ουσιαστικά, ρήματα, επίθετα και επιρρήματα. Αυτή η ταξινόμηση είναι απαραίτητη για την περαιτέρω ανάλυση της σύνταξης και της σημασιολογίας του κειμένου [10]. Ένα σημαντικό μέρος αυτής της διαδικασίας είναι η σημειολογική ετικετοθέτηση (POS tagging), όπου κάθε λέξη ετικετοποιείται με το μέρος του λόγου της, όπως ουσιαστικό, ρήμα, επίθετο. Αυτή η διαδικασία βασίζεται σε λεξικά και στατιστικά μοντέλα που εκπαιδεύονται σε μεγάλα σύνολα δεδομένων [6]. Η μορφολογική ανάλυση είναι εξαιρετικά σημαντική γιατί επιτρέπει στα συστήματα να κατανοούν καλύτερα το περιεχόμενο των κειμένων αναγνωρίζοντας τις βασικές μονάδες και τις γραμματικές τους σχέσεις. Επιπλέον, για τις εφαρμογές μετάφρασης και σύνθεσης φωνής, η μορφολογική ανάλυση είναι απαραίτητη για την ορθή απόδοση των λέξεων και των προτάσεων, διασφαλίζοντας την ακρίβεια και τη φυσικότητα. Τέλος, η ανάλυση των μορφολογικών ιδιοτήτων των λέξεων υποστηρίζει τη συντακτική ανάλυση, δηλαδή την κατανόηση της σχέσης μεταξύ των λέξεων σε μια πρόταση, και τη σημασιολογική ανάλυση, δηλαδή την κατανόηση του νοήματος. Η μορφολογική ανάλυση αποτελεί θεμελιώδη διαδικασία στην επεξεργασία φυσικής γλώσσας, επιτρέποντας την κατανόηση και την ανάλυση της γραμματικής δομής των κειμένων. Μέσω της διάσπασης του κειμένου σε βασικές μονάδες και της ταξινόμησής τους σε γραμματικές κατηγορίες, τα συστήματα NLP μπορούν να επιτύχουν μεγαλύτερη ακρίβεια και αποτελεσματικότητα σε διάφορες εφαρμογές, από τη μετάφραση μέχρι τη σύνθεση φωνής και την κατανόηση κειμένου.

Παράδειγμα Μορφολογικής Ανάλυσης

Κείμενο: "Ο μαθητής διαβάζει το βιβλίο."

Διάσπαση σε Λέξεις:

Ο, μαθητής, διαβάζει, το, βιβλίο.

Αναγνώριση Ριζών και Μορφών:

μαθητής (ρίζα: μαθητ-)

διαβάζει (ρίζα: διαβάζ-)

βιβλίο (ρίζα: βιβλί-)

Ταξινόμηση σε Γραμματικές Κατηγορίες:

Ο (άρθρο)

μαθητής (ουσιαστικό)

διαβάζει (ρήμα)

το (άρθρο)

βιβλίο (ουσιαστικό)

Σημειολογική Ετικετοθέτηση:

O/DT (Determiner)

μαθητής/NN (Noun)

διαβάζει/VB (Verb)

το/DT (Determiner)

βιβλίο/NN (Noun)

Συντακτική Ανάλυση: Καθορίζει τη δομή του κειμένου, αναγνωρίζοντας τους ρόλους των λέξεων μέσα στις φράσεις και τις προτάσεις. Αυτό επιτυγχάνεται με τεχνικές όπως η ανάλυση εξάρτησης και η ανάλυση φράσεων. **Συντακτική Ανάλυση** Η συντακτική ανάλυση είναι μια διαδικασία στην επεξεργασία φυσικής γλώσσας (NLP) που καθορίζει τη δομή του κειμένου, αναγνωρίζοντας τους ρόλους των λέξεων μέσα στις φράσεις και τις προτάσεις. Αυτή η ανάλυση είναι σημαντική για την κατανόηση της σύνθεσης των προτάσεων και της σχέσης μεταξύ των λέξεων. Για να επιτευχθεί αυτό, χρησιμοποιούνται τεχνικές όπως η ανάλυση εξάρτησης και η ανάλυση φράσεων. Η ανάλυση εξάρτησης εστιάζει στις σχέσεις μεταξύ των λέξεων σε μια πρόταση, δημιουργώντας ένα γράφημα εξαρτήσεων όπου κάθε λέξη συνδέεται με άλλες λέξεις με συγκεκριμένες σχέσεις. Για παράδειγμα, σε μια πρόταση όπως "Ο μαθητής διαβάζει το βιβλίο", η λέξη "διαβάζει" θα ήταν το κεντρικό ρήμα, με τις λέξεις "μαθητής" και "βιβλίο" να συνδέονται με αυτό ως υποκείμενο και αντικείμενο αντίστοιχα. Η ανάλυση φράσεων, από την άλλη πλευρά, δημιουργεί μια ιεραρχική δομή των φράσεων μέσα σε μια πρόταση, αποκαλύπτοντας πώς οι λέξεις συνδυάζονται για να σχηματίσουν μεγαλύτερες φράσεις και προτάσεις. Για παράδειγμα, η φράση "ο μαθητής" αναγνωρίζεται ως ονομαστική φράση, και "το βιβλίο" ως αντικείμενο. Αυτές οι φράσεις εντάσσονται στη συνολική δομή της πρότασης, βοηθώντας στην κατανόηση του τρόπου με τον οποίο οι λέξεις συνεργάζονται για να εκφράσουν πλήρη νοήματα. **Παράδειγμα Συντακτικής Ανάλυσης** Ας εξετάσουμε την πρόταση: "Ο μαθητής διαβάζει το βιβλίο."

Ανάλυση Εξάρτησης:

- "διαβάζει" (κεντρικό ρήμα)
- "μαθητής" (υποκείμενο, εξαρτάται από το "διαβάζει")
- "βιβλίο" (αντικείμενο, εξαρτάται από το "διαβάζει")
- "Ο" (καθοριστικό άρθρο για το "μαθητής")
- "το" (καθοριστικό άρθρο για το "βιβλίο")

Ανάλυση Φράσεων:

- Ονοματική φράση (NP): "ο μαθητής"
- Άρθρο (DT): "ο"
- Ουσιαστικό (NN): "μαθητής"
- Ρηματική φράση (VP): "διαβάζει το βιβλίο"
- Ρήμα (VB): "διαβάζει"
- Ονοματική φράση (NP): "το βιβλίο"
- Άρθρο (DT): "το"
- Ουσιαστικό (NN): "βιβλίο"

Η συντακτική ανάλυση είναι θεμελιώδης για την πλήρη κατανόηση και επεξεργασία του κειμένου σε επίπεδο φυσικής γλώσσας. Μέσω της αναγνώρισης των σχέσεων και των δομών των λέξεων, επιτρέπει την αποτελεσματική ανάλυση της σημασίας και της λειτουργίας του κειμένου, ενισχύοντας την ακρίβεια και την αποδοτικότητα σε εφαρμογές όπως η μηχανική μετάφραση, η σύνθεση φωνής, και οι ερωταποκρίσεις.

Ανάλυση Συντακτικής Σχέσης: Εντοπίζει σχέσεις εξάρτησης μεταξύ λέξεων και φράσεων, όπως υποκειμενικό-ρήμα, αντικείμενο-ρήμα, κ.λπ.

Ανάλυση Συντακτικής Σχέσης Η ανάλυση συντακτικής σχέσης αποτελεί ένα σημαντικό μέρος της συντακτικής ανάλυσης στην επεξεργασία φυσικής γλώσσας (NLP). Εστιάζει στην αναγνώριση και τον εντοπισμό των σχέσεων εξάρτησης μεταξύ λέξεων και φράσεων σε μια πρόταση. Αυτό περιλαμβάνει την ταυτοποίηση των σχέσεων όπως υποκείμενο-ρήμα, αντικείμενο-ρήμα, και άλλες σημαντικές συντακτικές δομές. Η ανάλυση συντακτικής σχέσης βοηθά στην κατανόηση του τρόπου με τον οποίο οι λέξεις συνδέονται μεταξύ τους για να σχηματίσουν προτάσεις με νόημα. Για παράδειγμα, σε μια πρόταση, το υποκείμενο είναι συνήθως η λέξη ή η φράση που εκτελεί την ενέργεια που περιγράφεται από το ρήμα, ενώ το αντικείμενο είναι αυτό που δέχεται την ενέργεια.

Διαδικασία Ανάλυσης Συντακτικής Σχέσης Η διαδικασία ξεκινά με την ανάλυση των λέ-

ξεων σε μια πρόταση και την αναγνώριση των γραμματικών τους ρόλων. Ακολουθεί η χαρτογράφηση των σχέσεων εξάρτησης μεταξύ αυτών των λέξεων. Αυτή η διαδικασία συχνά χρησιμοποιεί τεχνικές όπως η ανάλυση εξάρτησης και η ανάλυση φράσεων. Στην ανάλυση εξάρτησης, κάθε λέξη συνδέεται με άλλες λέξεις μέσω γραμμών που αντιπροσωπεύουν συντακτικές σχέσεις. Το αποτέλεσμα είναι ένα γράφημα εξαρτήσεων που αποτυπώνει τη δομή της πρότασης. Για παράδειγμα, το ρήμα είναι συνήθως το κεντρικό στοιχείο, με τις άλλες λέξεις να εξαρτώνται από αυτό. Η ανάλυση φράσεων, από την άλλη πλευρά, οργανώνει τις λέξεις σε ιεραρχικές δομές που αποκαλύπτουν πώς οι λέξεις συνδυάζονται για να σχηματίσουν φράσεις και προτάσεις. Αυτή η ιεραρχική δομή βοηθά στην κατανόηση των σχέσεων εξάρτησης σε πιο σύνθετο επίπεδο.

Παράδειγμα Ανάλυσης Συντακτικής Σχέσης

Ας εξετάσουμε την πρόταση: "Ο μαθητής διαβάζει το βιβλίο."

Ανάλυση Συντακτικής Σχέσης: - "διαβάζει" (κεντρικό ρήμα)

- "μαθητής" (υποκείμενο, εξαρτάται από το "διαβάζει")
- "βιβλίο" (αντικείμενο, εξαρτάται από το "διαβάζει")
- "ο" (καθοριστικό άρθρο για το "μαθητής")
- "το" (καθοριστικό άρθρο για το "βιβλίο")

Στο παράδειγμα αυτό, το "μαθητής" είναι το υποκείμενο που εκτελεί την ενέργεια του ρήματος "διαβάζει", ενώ το "βιβλίο" είναι το αντικείμενο που δέχεται την ενέργεια. Τα άρθρα "ο" και "το" καθορίζουν τα ουσιαστικά "μαθητής" και "βιβλίο" αντίστοιχα, και συνδέονται συντακτικά μαζί τους. Η ανάλυση συντακτικής σχέσης είναι θεμελιώδης για την πλήρη κατανόηση της σύνταξης μιας πρότασης και παρέχει τη βάση για πιο προηγμένες επεξεργασίες κειμένου, όπως η σημασιολογική ανάλυση και η ανάλυση προτάσεων. Μέσω της ταυτοποίησης των σχέσεων εξάρτησης, τα συστήματα NLP μπορούν να κατανοούν καλύτερα το περιεχόμενο και τη δομή των κειμένων, βελτιώνοντας την ακρίβεια και την αποτελεσματικότητα σε εφαρμογές όπως η αυτόματη μετάφραση και η ανάλυση κειμένων.

Σημασιολογική Ανάλυση:

- **Ανάλυση Λεξιλογίου:** Καθορίζει το νόημα των λέξεων με βάση ορισμούς, λεξικά και στατιστικά δεδομένα.
- **Ανάλυση Σημασιολογικής Σχέσης:** Εντοπίζει σημασιολογικές σχέσεις μεταξύ λέξεων και φράσεων, όπως συνωνυμία, αντίθεση, υπερωνυμία, υποωνυμία, κ.λπ.

- Ανάλυση Συναισθήματος: Αναγνωρίζει το συναίσθημα που εκφράζεται στο κείμενο, όπως χαρά, θλίψη, θυμός, κ.λπ.
- Ανάλυση Σκοπού: Καθορίζει τον σκοπό του κειμένου, όπως ενημέρωση, πειθώ, ψυχαγωγία, κ.λπ.

Εκτός από τις προαναφερθείσες τεχνικές, η NLP αξιοποιεί και άλλες μεθόδους, όπως:

- Μηχανική Μάθηση: Χρησιμοποιεί αλγόριθμους για να εκπαιδευτεί σε μεγάλα σύνολα δεδομένων κειμένου και να μάθει να εκτελεί αυτόματα εργασίες NLP, όπως ταξινόμηση κειμένου, εξαγωγή πληροφοριών και απάντηση σε ερωτήσεις.
- Στατιστική Μοντελοποίηση: Χρησιμοποιεί στατιστικά μοντέλα για να αναπαραστήσει τη γλωσσική δομή και να προβλέψει την επόμενη λέξη σε μια πρόταση ή την πιθανότητα εμφάνισης μιας λέξης σε ένα δεδομένο πλαίσιο.

Η επιλογή των κατάλληλων τεχνικών NLP εξαρτάται από το σκοπό της ανάλυσης και το είδος του κειμένου που εξετάζεται. Η NLP βρίσκει εφαρμογή σε πλήθος τομέων, όπως η μηχανική μετάφραση, η αναζήτηση πληροφοριών, η εξαγωγή κειμένου, η ανάλυση συναισθημάτων, η δημιουργία chatbot και η δημιουργία κειμένου.

Πώς τα σύγχρονα μοντέλα NLP αντιμετωπίζουν τις πολυπλοκότητες της ανθρώπινης γλώσσας

Η ανθρώπινη γλώσσα είναι γεμάτη πολυπλοκότητες που θέτουν προκλήσεις στα σύγχρονα μοντέλα NLP. Ακολουθούν μερικές από τις πιο σημαντικές:

- Ασάφεια: Η γλώσσα συχνά δεν είναι σαφής και μπορεί να έχει πολλαπλές ερμηνείες. Για παράδειγμα, η φράση "Η Μαρία είδε τον άνδρα με το τηλεσκόπιο" μπορεί να ερμηνευτεί με δύο τρόπους: η Μαρία χρησιμοποίησε ένα τηλεσκόπιο για να δει τον άνδρα ή η Μαρία είδε έναν άνδρα που κρατούσε ένα τηλεσκόπιο. Τα μοντέλα NLP πρέπει να είναι σε θέση να λαμβάνουν υπόψη όλα τα πιθανά νοήματα μιας πρότασης για να την ερμηνεύσουν σωστά.
- Συμφραζόμενο: Η σημασία μιας λέξης ή φράσης μπορεί να εξαρτηθεί από το πλαίσιο στο οποίο χρησιμοποιείται. Για παράδειγμα, η λέξη "μήλο" μπορεί να αναφέρεται σε φρούτο, εταιρεία υπολογιστών ή πόλη, ανάλογα με το πλαίσιο. Τα μοντέλα NLP

πρέπει να λαμβάνουν υπόψη το πλαίσιο μιας πρότασης για να κατανοήσουν σωστά τη σημασία των λέξεων που χρησιμοποιούνται.

- **Πολυσημία:** Πολλές λέξεις στην ανθρώπινη γλώσσα έχουν πολλαπλές σημασίες. Για παράδειγμα, η λέξη "κλειδί" μπορεί να αναφέρεται σε ένα αντικείμενο που χρησιμοποιείται για να ανοίξει μια πόρτα, ένα βασικό στοιχείο κατανόησης ενός προβλήματος ή μια μουσική νότα. Τα μοντέλα NLP πρέπει να είναι σε θέση να προσδιορίζουν τη σωστή σημασία μιας λέξης με βάση το πλαίσιο στο οποίο χρησιμοποιείται.
- **Σαρκασμός και χιούμορ:** Η ανθρώπινη γλώσσα συχνά χρησιμοποιεί σαρκασμό και χιούμορ, τα οποία μπορεί να είναι δύσκολο να κατανοηθούν από τα μοντέλα NLP. Για παράδειγμα, η φράση "Αυτό ήταν υπέροχο" μπορεί να ειπωθεί με σαρκασμό για να εκφράσει το αντίθετο. Τα μοντέλα NLP πρέπει να είναι σε θέση να αναγνωρίζουν και να κατανοούν τέτοιου είδους γλωσσικά φαινόμενα.
- **Γλωσσικές αλλαγές:** Η γλώσσα εξελίσσεται συνεχώς, με νέες λέξεις και φράσεις να δημιουργούνται και παλαιές να εξαφανίζονται. Τα μοντέλα NLP πρέπει να είναι σε θέση να προσαρμόζονται σε αυτές τις αλλαγές για να παραμένουν ακριβή.

Για να αντιμετωπίσουν αυτές τις προκλήσεις, τα σύγχρονα μοντέλα NLP χρησιμοποιούν διάφορες τεχνικές, όπως:

- **Μηχανική Μάθηση:** Τα μοντέλα NLP εκπαιδεύονται σε μεγάλα σύνολα δεδομένων κειμένου και κώδικα για να μάθουν τα στατιστικά μοτίβα της ανθρώπινης γλώσσας.
- **Σημασιολογική Εμφάνιση:** Τα μοντέλα NLP εκπροσωπούν τη σημασία των λέξεων και φράσεων με τρόπο που μπορεί να κατανοηθεί από υπολογιστές.
- **Προσοχή:** Τα μοντέλα NLP εστιάζουν σε συγκεκριμένα μέρη μιας πρότασης ή ενός εγγράφου για να κατανοήσουν καλύτερα το πλαίσιο.
- **Γνώσεις Κόσμου:** Τα μοντέλα NLP ενσωματώνουν γνώσεις για τον κόσμο για να βοηθήσουν στην ερμηνεία του κειμένου.

Παρά τις προκλήσεις, τα σύγχρονα μοντέλα NLP έχουν κάνει σημαντική πρόοδο στην κατανόηση και την επεξεργασία της ανθρώπινης γλώσσας. Πλέον, μπορούν να

2.10 Στατιστικά Μοντέλα και Μοντέλα Μηχανικής Μάθησης

Στατιστικά Μοντέλα στην Σύνθεση Φωνής και Αναγνώριση Ομιλίας Τα Στατιστικά Μοντέλα παίζουν καθοριστικό ρόλο στην Σύνθεση Φωνής και Αναγνώριση Ομιλίας, προσφέροντας ισχυρά εργαλεία για την επεξεργασία και την ανάλυση των φωνητικών σημάτων. Ας δούμε δύο βασικά μοντέλα που χρησιμοποιούνται ευρέως σε αυτούς τους τομείς:

1. Κρυφά Μοντέλα Markov (Hidden Markov Models - HMMs): Περιγραφή των Κρυφών Μαρκοβιανών Μοντέλων (HMMs)

Τα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models, HMMs) είναι μια βασική τεχνική που χρησιμοποιείται ευρέως στην επεξεργασία σήματος ομιλίας, καθώς και σε πολλές άλλες εφαρμογές στην επεξεργασία φυσικής γλώσσας. Η βασική ιδέα πίσω από τα HMMs είναι να θεωρήσουν το σήμα ομιλίας ως μια σειρά από κρυφές καταστάσεις, όπου κάθε κατάσταση αντιπροσωπεύει ένα διαφορετικό φωνητικό φαινόμενο, όπως ένα φωνήεν ή ένα σύμφωνο. Κάθε κατάσταση στο HMM αντιπροσωπεύει ένα συγκεκριμένο φωνητικό γεγονός, και η μετάβαση από μια κατάσταση σε μια άλλη διέπεται από πιθανότητες μετάβασης. Αυτές οι πιθανότητες καθορίζουν την πιθανότητα μετάβασης από την τρέχουσα κατάσταση σε μια άλλη συγκεκριμένη κατάσταση. Επιπλέον, η παραγωγή ενός παρατηρημένου σήματος σε κάθε κατάσταση ορίζεται από πιθανότητες εκπομπής, οι οποίες καθορίζουν την πιθανότητα να παρατηρηθεί ένα συγκεκριμένο σύμβολο ή χαρακτηριστικό δεδομένης της κατάστασης. Για παράδειγμα, ένα HMM για την αναγνώριση φωνής θα έχει καταστάσεις που αντιπροσωπεύουν διάφορους ήχους της ομιλίας, όπως φωνήεντα και σύμφωνα. Καθώς το μοντέλο εξελίσσεται, θα μεταβαίνει από τη μία κατάσταση στην άλλη με βάση τις πιθανότητες μετάβασης. Κάθε φορά που το μοντέλο βρίσκεται σε μια συγκεκριμένη κατάσταση, θα παράγει έναν παρατηρούμενο ήχο με βάση τις πιθανότητες εκπομπής. Η ισχύς των HMMs έγκειται στην ικανότητά τους να χειρίζονται αλληλουχίες δεδομένων όπου η σειρά των παρατηρήσεων έχει σημασία. Η διαδικασία εκπαίδευσης ενός HMM περιλαμβάνει την εκτίμηση των πιθανοτήτων μετάβασης και εκπομπής από τα δεδομένα εκπαίδευσης, συνήθως χρησιμοποιώντας τον αλγόριθμο Baum-Welch. Μόλις το μοντέλο εκπαιδευτεί, μπορεί να χρησιμοποιηθεί για την ανάλυση νέων ακολουθιών, προσδιορίζοντας τις πιο πιθανές καταστάσεις και τις πιθανότητες των παρατηρήσεων. Για να κατανο-

ήσουμε καλύτερα πώς λειτουργούν τα HMMs, ας εξετάσουμε ένα παράδειγμα. Ας υποθέσουμε ότι έχουμε ένα HMM που αναγνωρίζει φωνήεντα και σύμφωνα. Οι καταστάσεις του μοντέλου μπορεί να είναι:

- - K1: φωνήεν
- - K2: σύμφωνο

Οι πιθανότητες μετάβασης θα μπορούσαν να είναι:

- $P(K1 \rightarrow K1)$: Πιθανότητα να παραμείνει σε φωνήεν
- $P(K1 \rightarrow K2)$: Πιθανότητα να μεταβεί από φωνήεν σε σύμφωνο
- $P(K2 \rightarrow K1)$: Πιθανότητα να μεταβεί από σύμφωνο σε φωνήεν
- $P(K2 \rightarrow K2)$: Πιθανότητα να παραμείνει σε σύμφωνο

Οι πιθανότητες εκπομπής καθορίζουν τις πιθανότητες παρατήρησης συγκεκριμένων φωνητικών χαρακτηριστικών σε κάθε κατάσταση. Συνοψίζοντας, τα Κρυφά Μαρκοβιανά Μοντέλα αποτελούν μια ισχυρή μέθοδο για την ανάλυση και την κατανόηση σήματος ομιλίας. Μέσω της μοντελοποίησης των φωνητικών φαινομένων ως κρυφών καταστάσεων και της χρήσης πιθανοτήτων μετάβασης και εκπομπής, τα HMMs επιτρέπουν την αποτελεσματική αναγνώριση και ανάλυση των αλληλουχιών ομιλίας.

2. Εφαρμογές:

Σύνθεση Φωνής με Κρυφά Μαρκοβιανά Μοντέλα (HMMs)

Στη σύνθεση φωνής, τα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models, HMMs) χρησιμοποιούνται για να μοντελοποιήσουν τη στατιστική σχέση μεταξύ του κειμένου και του αντίστοιχου φωνητικού σήματος. Η τεχνική αυτή επιτρέπει τη σύνθεση φυσικότερης και πιο κατανοητής ομιλίας από μηχανές, κάνοντας χρήση των πιθανοτήτων μετάβασης και εκπομπής για την αναπαραγωγή του ήχου.

Πώς Λειτουργούν τα HMMs στη Σύνθεση Φωνής

Στη διαδικασία σύνθεσης φωνής με HMMs, το κείμενο που πρόκειται να μετατραπεί σε ομιλία πρώτα αναλύεται ώστε να εξάγει τις απαραίτητες φωνητικές πληροφορίες, όπως τα φωνήεντα, τα σύμφωνα, και τη διάρκεια των ήχων. Κάθε λέξη ή φράση στο κείμενο αντιστοιχεί σε μια σειρά από κρυφές καταστάσεις, καθεμία από τις οποίες αντιπροσωπεύει ένα συγκεκριμένο φωνητικό φαινόμενο.

Οι πιθανότητες μετάβασης και εκπομπής παίζουν κεντρικό ρόλο σε αυτό το μοντέλο:

- Πιθανότητες Μετάβασης: Καθορίζουν την πιθανότητα μετάβασης από μια κατάσταση σε μια άλλη. Για παράδειγμα, η μετάβαση από ένα φωνήεν σε ένα σύμφωνο.
- - Πιθανότητες Εκπομπής: Καθορίζουν την πιθανότητα παραγωγής ενός συγκεκριμένου ήχου ή φωνητικού σήματος δεδομένης της κατάστασης.

Διαδικασία Σύνθεσης

- (α') Ανάλυση Κειμένου: Το κείμενο διασπάται σε προτάσεις, λέξεις και φωνήματα (οι βασικές μονάδες ήχου).
- (β') Μοντελοποίηση Καταστάσεων: Κάθε φώνημα αντιστοιχεί σε μια κρυφή κατάσταση του HMM. Για παράδειγμα, η λέξη "σπίτι" μπορεί να αναλυθεί σε φωνήματα που αντιστοιχούν στις καταστάσεις /s/, /p/, /i/, /t/, /i/.
- (γ') Υπολογισμός Πιθανοτήτων: Για κάθε φώνημα, υπολογίζονται οι πιθανότητες μετάβασης στις επόμενες καταστάσεις και οι πιθανότητες εκπομπής των αντίστοιχων φωνητικών σημάτων.
- (δ') . Σύνθεση Ομιλίας: Χρησιμοποιώντας τις πιθανότητες αυτές, το μοντέλο παράγει το φωνητικό σήμα που αντιστοιχεί στο κείμενο. Οι ήχοι συντίθενται διαδοχικά με βάση τις πιθανότητες μετάβασης μεταξύ των φωνημάτων και τις πιθανότητες εκπομπής για την παραγωγή των σωστών ηχητικών κυμάτων.

Παράδειγμα Σύνθεσης Φωνής

Ας εξετάσουμε τη σύνθεση της λέξης "σπίτι":

- Το κείμενο "σπίτι" διασπάται στα φωνήματα /s/, /p/, /i/, /t/, /i/.
- Κάθε φώνημα αντιστοιχεί σε μια κρυφή κατάσταση του HMM. - Οι πιθανότητες μετάβασης μπορεί να είναι:
- $P(/s/ \rightarrow /p/)$: Πιθανότητα μετάβασης από /s/ σε /p/
- $P(/p/ \rightarrow /i/)$: Πιθανότητα μετάβασης από /p/ σε /i/
- $P(/i/ \rightarrow /t/)$: Πιθανότητα μετάβασης από /i/ σε /t/
- $P(/t/ \rightarrow /i/)$: Πιθανότητα μετάβασης από /t/ σε /i/
- Οι πιθανότητες εκπομπής καθορίζουν τον ακριβή ήχο που παράγεται σε κάθε φώνημα.

Η σύνθεση φωνής μέσω HMMs επιτρέπει τη δημιουργία φωνητικών σημάτων που ακούγονται φυσικά και κατανοητά. Τα μοντέλα αυτά επιτρέπουν τη λεπτομερή και ακριβή αναπαραγωγή των ήχων, λαμβάνοντας υπόψη τις φυσικές μεταβολές και τις εξαρτήσεις μεταξύ των φωνημάτων.

Χρησιμοποιώντας τις τεχνικές αυτές, τα HMMs συμβάλλουν στην ανάπτυξη προηγμένων συστημάτων σύνθεσης φωνής, βελτιώνοντας τη φυσικότητα και την ευελιξία της παραγόμενης ομιλίας και καθιστώντας την τεχνολογία αυτή ιδιαίτερα χρήσιμη σε διάφορες εφαρμογές, από την υποβοήθηση ανθρώπων με αναπηρίες μέχρι τις ψηφιακές βοηθούς και τα συστήματα πλοήγησης. **Αναγνώριση Ομιλίας με Κρυφά Μαρκοβιανά Μοντέλα (HMMs)** Τα Κρυφά Μαρκοβιανά Μοντέλα (Hidden Markov Models, HMMs) αποτελούν μια από τις κύριες μεθόδους που χρησιμοποιούνται στην αναγνώριση ομιλίας. Τα HMMs επιτρέπουν την αποδοτική και ακριβή αναγνώριση λέξεων ή φράσεων από ένα ηχητικό σήμα, συγκρίνοντας το παρατηρούμενο σήμα με προετοιμασμένα μοντέλα που έχουν εκπαιδευτεί σε ένα σύνολο δεδομένων φωνής. **Πώς Λειτουργούν τα HMMs στην Αναγνώριση Ομιλίας** Η διαδικασία αναγνώρισης ομιλίας με HMMs βασίζεται στην πιθανότητα ταύτισης του παρατηρούμενου ηχητικού σήματος με τα εκπαιδευμένα μοντέλα. Η αναγνώριση ομιλίας περιλαμβάνει τα ακόλουθα βήματα:

1. Εκπαίδευση των Μοντέλων: Τα HMMs εκπαιδεύονται σε ένα μεγάλο σύνολο δεδομένων φωνής. Κατά την εκπαίδευση, τα μοντέλα μαθαίνουν τις πιθανότητες μετάβασης μεταξύ καταστάσεων και τις πιθανότητες εκπομπής για κάθε φώνημα ή λέξη.
2. Προεπεξεργασία Σήματος: Το ηχητικό σήμα που πρόκειται να αναγνωριστεί υποβάλλεται σε προεπεξεργασία για να εξαχθούν τα φωνητικά χαρακτηριστικά του. Αυτό περιλαμβάνει τη μετατροπή του σήματος σε μια σειρά από φωνητικές παρατηρήσεις, όπως τα φασματογραφικά χαρακτηριστικά.
3. Ανάλυση και Ταύτιση: Το παρατηρούμενο σήμα συγκρίνεται με τα εκπαιδευμένα HMMs για να βρεθεί η ακολουθία καταστάσεων που ταιριάζει καλύτερα με το σήμα. Χρησιμοποιείται ο αλγόριθμος Viterbi για να βρεθεί η πιο πιθανή ακολουθία καταστάσεων.
4. Εκτίμηση Πιθανοτήτων: Ο αλγόριθμος υπολογίζει τις πιθανότητες μετάβασης και εκπομπής για να καθορίσει ποιο HMM ταιριάζει καλύτερα με το παρατηρούμενο σήμα. Το μοντέλο με την υψηλότερη πιθανότητα είναι αυτό που αναγνωρίζεται ως το πιο πιθανό αποτέλεσμα.

Παράδειγμα Αναγνώρισης Ομιλίας

Ας εξετάσουμε την αναγνώριση της φράσης "καλημέρα":

1. Εκπαίδευση των Μοντέλων: Εκπαιδεύουμε τα HMMs χρησιμοποιώντας δεδομένα φωνής από πολλούς ομιλητές που λένε τη φράση "καλημέρα". Το μοντέλο μαθαίνει τις πιθανότητες μετάβασης μεταξύ φωνημάτων (/k/, /a/, /l/, /i/, /m/, /e/, /r/, /a/) και τις πιθανότητες εκπομπής των αντίστοιχων ηχητικών χαρακτηριστικών.
2. Προεπεξεργασία Σήματος: Το παρατηρούμενο ηχητικό σήμα "καλημέρα" μετατρέπεται σε μια σειρά φωνητικών χαρακτηριστικών.
3. Ανάλυση και Ταύτιση: Χρησιμοποιώντας τον αλγόριθμο Viterbi, το σύστημα βρίσκει την πιο πιθανή ακολουθία καταστάσεων του HMM που ταιριάζει με το παρατηρούμενο σήμα.
4. Εκτίμηση Πιθανοτήτων: Το σύστημα υπολογίζει τις πιθανότητες μετάβασης και εκπομπής και συγκρίνει αυτές τις πιθανότητες με τα εκπαιδευμένα HMMs για τη φράση "καλημέρα". Το μοντέλο με την υψηλότερη συνολική πιθανότητα αναγνωρίζεται ως το αποτέλεσμα.

Πλεονεκτήματα και Χρήσεις

Η χρήση των HMMs στην αναγνώριση ομιλίας προσφέρει σημαντικά πλεονεκτήματα, όπως η ικανότητα χειρισμού των μεταβαλλόμενων χαρακτηριστικών της ανθρώπινης ομιλίας και η δυνατότητα αναγνώρισης σε θορυβώδη περιβάλλοντα. Οι τεχνικές αυτές χρησιμοποιούνται ευρέως σε εφαρμογές όπως οι ψηφιακοί βοηθοί, τα συστήματα πλοήγησης, και τα προγράμματα εκπαίδευσης γλώσσας. Τα HMMs επιτρέπουν στα συστήματα αναγνώρισης ομιλίας να κατανοούν και να ερμηνεύουν την ανθρώπινη φωνή με μεγάλη ακρίβεια, καθιστώντας τα εργαλεία αυτά απαραίτητα για την αλληλεπίδραση ανθρώπου-μηχανής.

2.10.1 Μοντέλα Μίγματος Gaussian (Gaussian Mixture Models - GMMs)

Περιγραφή: Τα GMMs αποτελούν μια γενίκευση των HMMs, όπου η πιθανότητα εκπομπής σε κάθε κατάσταση δεν είναι μονοδιάστατη, αλλά μοντελοποιείται από ένα μίγμα Gaussian. Αυτό επιτρέπει στα GMMs να απεικονίζουν πιο περίπλοκες κατανομές χαρακτηριστικών φωνής, βελτιώνοντας την ακρίβεια.

Εφαρμογές:

Σύνθεση Φωνής: Τα GMMs χρησιμοποιούνται για να μοντελοποιήσουν με μεγαλύτερη ακρίβεια την κατανομή χαρακτηριστικών φωνής σε κάθε κατάσταση HMM, οδηγώντας σε πιο φυσική και ρεαλιστική φωνή. **Αναγνώριση Ομιλίας:** Τα GMMs μπορούν να χειριστούν φωνητική διακύμανση και θόρυβο, βελτιώνοντας την αναγνώριση λέξεων, ειδικά σε περιβάλλοντα με θόρυβο.

2.11 Συνοπτικά

- Τα **HMMs** και **GMMs** αποτελούν θεμελιώδη εργαλεία για την επεξεργασία φωνητικών σημάτων.
- Τα **HMMs** μοντελοποιούν τις καταστάσεις φωνητικών φαινομένων και τις πιθανότητες μετάβασης μεταξύ αυτών.
- Τα **GMMs** αποτελούν μια γενίκευση των HMMs, επιτρέποντας την απεικόνιση πιο περίπλοκων κατανομών χαρακτηριστικών φωνής.
- Η χρήση HMMs και GMMs συμβάλλει στην **ακρίβεια και φυσικότητα της σύνθεσης φωνής**, καθώς και στην **αναγνώριση ομιλίας σε θορυβώδη περιβάλλοντα**.

Εκτός από τα HMMs και GMMs, αξίζει να αναφερθούν και άλλα στατιστικά μοντέλα που χρησιμοποιούνται στην σύνθεση φωνής και αναγνώριση ομιλίας, όπως τα **Dynamical Neural Networks (DNNs)** και τα **Recurrent Neural Networks (RNNs)**, τα οποία προσφέρουν ακόμα πιο βελτιωμένες αποδόσεις.

2.12 Μετάβαση από Στατιστικά Μοντέλα σε Νευρωνικά Δίκτυα και Βαθιά Μάθηση

Η επεξεργασία φωνητικών σημάτων, τόσο στη σύνθεση φωνής όσο και στην αναγνώριση ομιλίας, βασίστηκε για πολλά χρόνια σε στατιστικά μοντέλα, όπως **Hidden Markov Models (HMMs)** και **Gaussian Mixture Models (GMMs)**. Τα μοντέλα αυτά προσέφεραν ισχυρά εργαλεία για την ανάλυση και επεξεργασία φωνητικών δεδομένων, αλλά παρουσίαζαν κάποιους περιορισμούς:

- **Περιορισμένη εκφραστικότητα:** Τα στατιστικά μοντέλα δυσκολεύονταν να μοντελοποιήσουν με ακρίβεια τις περίπλοκες μη γραμμικές σχέσεις που χαρακτηρίζουν τη φωνή.
- **Χειροκίνητη μηχανική:** Η δημιουργία και η βελτιστοποίηση των στατιστικών μοντέλων απαιτούσε συχνά χειροκίνητη μηχανική, η οποία ήταν χρονοβόρα και δεν πάντα οδηγούσε σε βέλτιστες λύσεις.

Η άνοδος της **βαθιάς μάθησης** έφερε επανάσταση στον τομέα της επεξεργασίας φωνητικών σημάτων. Τα **νευρωνικά δίκτυα**, ιδιαίτερα τα **αναδρομικά νευρωνικά δίκτυα (RNNs)** και **Convolutional Neural Networks (CNNs)**, προσφέρουν σημαντικά πλεονεκτήματα σε σχέση με τα στατιστικά μοντέλα:

- **Αυτόματη μάθηση:** Τα νευρωνικά δίκτυα μπορούν να μάθουν από δεδομένα χωρίς την ανάγκη χειροκίνητης μηχανικής.
- **Βελτιωμένη εκφραστικότητα:** Τα νευρωνικά δίκτυα, με την πολυπλοκότερη δομή τους, μπορούν να μοντελοποιήσουν με μεγαλύτερη ακρίβεια τις μη γραμμικές σχέσεις στη φωνή.
- **Καλύτερη γενίκευση:** Τα νευρωνικά δίκτυα τείνουν να γενικεύονται καλύτερα σε νέα δεδομένα, άρα μπορούν να λειτουργήσουν αποτελεσματικά σε ποικίλες συνθήκες.

Ως αποτέλεσμα, η βαθιά μάθηση έχει οδηγήσει σε σημαντική πρόοδο στην σύνθεση φωνής και αναγνώριση ομιλίας:

- **Πιο φυσική και ρεαλιστική φωνή:** Τα νευρωνικά δίκτυα παράγουν φωνή που είναι πιο κοντά στην ανθρώπινη φωνή, με λιγότερους τεχνητούς ή ρομποτικούς ήχους.
- **Βελτιωμένη ακρίβεια αναγνώρισης:** Τα νευρωνικά δίκτυα μπορούν να αναγνωρίσουν ομιλία με μεγαλύτερη ακρίβεια, ακόμη και σε θορυβώδη περιβάλλοντα ή με ασαφή προφορά.
- **Νέα δυνατότητες:** Η βαθιά μάθηση ανοίγει τον δρόμο για νέες δυνατότητες, όπως η προσαρμογή της φωνής σε συγκεκριμένα χαρακτηριστικά ή η δημιουργία συνθετικής φωνής σε πραγματικό χρόνο.

Η μετάβαση από στατιστικά μοντέλα σε νευρωνικά δίκτυα και βαθιά μάθηση σηματοδοτεί μια **νέα εποχή στην επεξεργασία φωνητικών σημάτων**. Τα νευρωνικά δίκτυα προσφέρουν ατελείωτες δυνατότητες για την βελτίωση της σύνθεσης φωνής και αναγνώρισης ομιλίας, με στόχο την ολοένα και πιο φυσική και ανθρώπινη αλληλεπίδραση με υπολογιστές και συσκευές.

2.13 Νευρωνικά Δίκτυα: Κινητήρια δύναμη για Σύνθεση Φωνής και Ανάλυση Φυσικής Γλώσσας

Τα **Νευρωνικά Δίκτυα** (Neural Networks), εμπνευσμένα από τη λειτουργία του ανθρώπινου εγκεφάλου, έχουν εξελιχθεί σε ισχυρά εργαλεία για την επεξεργασία δεδομένων σε διάφορους τομείς, συμπεριλαμβανομένης της **Σύνθεσης Φωνής** και της **Ανάλυσης Φυσικής Γλώσσας** (NLP). Η ικανότητά τους να μαθαίνουν από δεδομένα και να ανακαλύπτουν περίπλοκες σχέσεις τα καθιστά ιδανικά για την αντιμετώπιση των προκλήσεων που ελλοχεύουν σε αυτούς τους τομείς.

2.13.1 Ευρέως Διαδεδομένα Νευρωνικά Δίκτυα

- **Convolutional Neural Networks (CNNs):**

- **Εξαιρετικά στην αναγνώριση μοτίβων** σε δεδομένα εικόνων και ακολουθιών.
- Χρησιμοποιούνται στην NLP για εργασίες όπως η ταξινόμηση κειμένου, η εξαγωγή χαρακτηριστικών και η ανάλυση συναισθημάτων.

- **Recurrent Neural Networks (RNNs):**

- **Ιδανικά για την επεξεργασία ακολουθιών δεδομένων**, όπως το κείμενο, λαμβάνοντας υπόψη το πλαίσιο.
- Χρησιμοποιούνται στην NLP για εργασίες όπως η μετάφραση μηχανών, η δημιουργία κειμένου και η ανάλυση συναισθημάτων.

- **Long Short-Term Memory (LSTM):**

- Μία υποκατηγορία RNNs, σχεδιασμένη για να χειρίζεται μακροπρόθεσμες εξαρτήσεις σε δεδομένα ακολουθιών.

- Χρησιμοποιούνται στην NLP για εργασίες όπως η αναγνώριση ομιλίας, η ανάλυση συναισθημάτων και η απάντηση σε ερωτήσεις.
- **Transformer μοντέλα:**
 - Αρχιτεκτονική νευρωνικών δικτύων που βασίζεται σε μηχανισμούς προσοχής.
 - **Προσφέρουν παράλληλη επεξεργασία** δεδομένων ακολουθιών, επιτυγχάνοντας υψηλή ταχύτητα και ακρίβεια.
 - Χρησιμοποιούνται στην NLP για εργασίες όπως η μετάφραση μηχανών, η περίληψη κειμένου και η απάντηση σε ερωτήσεις.

2.13.2 Εφαρμογές στη Σύνθεση Φωνής

- **Νευροφωνή:** Χρησιμοποιεί RNNs, ειδικά LSTM, για να μοντελοποιήσει τις χρονικές εξαρτήσεις στη φωνή και να παράγει ρεαλιστική και φυσική φωνή. Επίσης εκμεταλλεύεται τα **Αναδρομικά Νευρωνικά Δίκτυα (RNNs)**, ειδικότερα τα **Μνημονικά Συνελικτικά Δίκτυα (LSTM)**, για να μοντελοποιήσει τις χρονικές εξαρτήσεις στην παραγωγή φυσικής φωνής [16].
- **WaveNet:** Βασίζεται σε CNNs για να απεικονίσει τη σχέση μεταξύ φωνητικών χαρακτηριστικών και ηχητικών κυμάτων, δημιουργώντας **υψηλής ποιότητας φωνή**. Βασίζεται σε **Συνελικτικά Νευρωνικά Δίκτυα (CNNs)** και δημιουργεί υψηλής ποιότητας φωνή προσομοιώνοντας τη σχέση μεταξύ φωνητικών χαρακτηριστικών και ηχητικών κυμάτων [4].
- **Tacotron 2:** Συνδυάζει RNNs και CNNs για να δημιουργήσει φωνή από κείμενο, λαμβάνοντας υπόψη την προφορά, τον τονισμό και τη ρυθμική δομή. Συνδυάζει RNNs και CNNs για να δημιουργήσει φωνή από κείμενο, λαμβάνοντας υπόψη την προφορά, τον τονισμό και τη ρυθμική δομή [8].

Αυτές οι εφαρμογές αντιπροσωπεύουν την κορυφαία τεχνολογία στον τομέα της σύνθεσης φωνής, χρησιμοποιώντας προηγμένες τεχνικές μηχανικής μάθησης για τη δημιουργία φυσικής και ρεαλιστικής φωνής από κείμενο.

2.13.3 Εφαρμογές στην Ανάλυση Φυσικής Γλώσσας

1. **Machine Translation:** Χρησιμοποιεί Transformer μοντέλα για να μεταφράσει κείμενο από τη μία γλώσσα στην άλλη, λαμβάνοντας υπόψη το γραμματικό και σημασιολογικό πλαίσιο.
2. **Text Summarization:** Χρησιμοποιεί RNNs ή Transformer μοντέλα για να συνοψίσει κείμενο, διατηρώντας τα βασικά σημεία και την ουσία του.
3. **Question Answering:** Χρησιμοποιεί RNNs ή Transformer μοντέλα για να απαντήσει σε ερωτήσεις σχετικά με κείμενο, λαμβάνοντας υπόψη το πλαίσιο και το νόημα της ερώτησης.

Η ανάπτυξη και η εφαρμογή νευρωνικών δικτύων στους τομείς της Σύνθεσης

2.14 Τεχνολογίες και Εφαρμογές

1. **Εφαρμογές της Σύνθεσης Φωνής** Η Σύνθεση Φωνής, η οποία μετατρέπει κείμενο σε ομιλία, έχει πλέον αμέτρητες πρακτικές εφαρμογές που αγγίζουν διάφορους τομείς της καθημερινότητας. Ας δούμε μερικά ενδεικτικά παραδείγματα:

- **Συστήματα Πλοήγησης:**

GPS: Η φωνητική καθοδήγηση από συστήματα GPS έχει καταστεί απαραίτητη για ασφαλή και άνετη οδήγηση, παρέχοντας οδηγίες με σαφή και κατανοητό τρόπο.

Χάρτες: Η φωνητική περιγραφή διαδρομών και σημείων ενδιαφέροντος σε εφαρμογές χαρτών προσφέρει εύχρηστη πρόσβαση σε πληροφορίες, ιδανική για άτομα που οδηγούν ή αδυνατούν να κοιτάξουν την οθόνη.

- **Βοηθητικές Τεχνολογίες για Άτομα με Αναπηρίες:**

Αναγνώστες Οθόνης: Η μετατροπή κειμένου σε ομιλία επιτρέπει σε άτομα με προβλήματα όρασης να έχουν πρόσβαση σε πληροφορίες από υπολογιστές, tablets και smartphones, βελτιώνοντας την αυτονομία και την ποιότητα ζωής τους. Εργαλεία Επικοινωνίας: Η σύνθεση φωνής μπορεί να βοηθήσει άτομα με δυσκολίες στην ομιλία να επικοινωνούν αποτελεσματικά, προσφέροντας εναλλακτικές λύσεις έκφρασης και αλληλεπίδρασης.

- **Ψηφιακοί Βοηθοί:**

Εικονικοί Βοηθοί: Η φωνητική επικοινωνία με ψηφιακούς βοηθούς, όπως Amazon Alexa, Google Assistant και Siri, έχει γίνει καθημερινή συνήθεια για πολλούς, επιτρέποντας τον έλεγχο έξυπνων συσκευών, την αναζήτηση πληροφοριών, τη ρύθμιση υπενθυμίσεων και πολλά άλλα.

Chatbots: Η σύνθεση φωνής δίνει "φωνή" στα chatbots, επιτρέποντας τους να αλληλεπιδρούν με τους χρήστες με πιο φυσικό και ελκυστικό τρόπο, βελτιώνοντας την εξυπηρέτηση πελατών και την εμπειρία χρήστη.

- **Ψυχαγωγικές Εφαρμογές:**

Αφήγηση Ιστοριών: Η φωνητική αφήγηση παραμυθιών, βιβλίων και ιστοριών προσφέρει μια καθηλωτική εμπειρία ψυχαγωγίας, ιδανική για παιδιά και ενήλικες. Ακουστικά Βιβλία: Η μετατροπή βιβλίων σε ακουστική μορφή μέσω σύνθεσης φωνής επιτρέπει την κατανάλωση λογοτεχνίας εν κινήσει ή σε στιγμές που η ανάγνωση δεν είναι εφικτή. Διαδραστικά Παιχνίδια: Η ενσωμάτωση φωνητικών στοιχείων σε παιχνίδια προσφέρει μια πιο ρεαλιστική και συναρπαστική εμπειρία παιχνιδιού, ενισχύοντας την αλληλεπίδραση του χρήστη.

- **Εκπαιδευτικές Εφαρμογές:**

Γλωσσική Εκμάθηση: Η σύνθεση φωνής μπορεί να χρησιμοποιηθεί για την εκμάθηση προφοράς, την εξάσκηση στην κατανόηση προφορικής ομιλίας και την ανάπτυξη δεξιοτήτων ακρόασης. Εκπαίδευση Ατόμων με Δυσκολίες Μάθησης: Η φωνητική περιγραφή εκπαιδευτικού υλικού και η παροχή φωνητικής υποστήριξης μπορεί να βοηθήσει άτομα με δυσκολίες μάθησης να αφομοιώσουν πληροφορίες και να βελτιώσουν την ακαδημαϊκή τους επίδοση.

2. Επιχειρηματικές Εφαρμογές:

- **Εξυπηρέτηση Πελατών:** Η φωνητική αυτοματοποίηση τηλεφωνικών κέντρων και η υλοποίηση chatbot με φωνή μπορούν να βελτιώσουν την εξυπηρέτηση πελατών, προσφέροντας ταχύτερη και πιο αποτελεσματική υποστήριξη.
- **Διαφήμιση και Marketing:** Η χρήση φωνητικών μηνυμάτων σε διαφημίσεις και marketing campaigns μπορεί να αυξήσει την ελκυστικότητα του μηνύματος και να ενισχύσει την αναγνωρισιμότητα της μάρκας.
- **Εκπαίδευση και Ενημέρωση Εργαζομένων:** Η φωνητική παρουσίαση εκπαι-

δευτικού υλικού ή η παροχή ενημερωτικών μηνυμάτων μέσω σύνθεσης φωνής μπορεί να βελτιώσει την παραγωγικότητα και την ενημέρωση των εργαζομένων.

3. Ιατρικές Εφαρμογές:

- **Διαγνωστικές Εφαρμογές:** Η φωνητική ανάλυση μπορεί να χρησιμοποιηθεί για την ανίχνευση σημείων ασθενειών, όπως η κατάθλιψη ή η νόσος του Parkinson, με βάση τα φωνητικά χαρακτηριστικά.
- **ο Πληροφόρηση Ασθενών:** Η φωνητική παρουσίαση ιατρικών πληροφοριών ή η παροχή οδηγιών σε ασθενείς μπορεί να βελτιώσει την κατανόηση και την τήρηση της θεραπείας.

4. Ερευνητικές Εφαρμογές:

- **Ανάλυση Συναισθημάτων:** Η φωνητική ανάλυση μπορεί να χρησιμοποιηθεί για την ανίχνευση συναισθημάτων, όπως χαρά, θλίψη ή θυμός, με βάση φωνητικά χαρακτηριστικά.
- **Ανάπτυξη Τεχνητής Νοημοσύνης:** Η σύνθεση φωνής αποτελεί σημαντικό εργαλείο στην ανάπτυξη τεχνητής νοημοσύνης, επιτρέποντας την υλοποίηση συστημάτων που μπορούν να αλληλεπιδρούν με τους ανθρώπους με φυσικό και ανθρώπινο τρόπο.
- **Διατήρηση Γλωσσών:** Η σύνθεση φωνής μπορεί να χρησιμοποιηθεί για την αναβίωση και διατήρηση γλωσσών που κινδυνεύουν με εξαφάνιση, δημιουργώντας φωνητικά ηχητικά ντοκουμέντα και εκπαιδευτικά υλικά.

2.15 Εφαρμογές της Ανάλυσης Φυσικής Γλώσσας

Η φυσική γλώσσα (NLP) έχει πλέον διεισδύσει σε πλήθος τομέων, προσφέροντας λύσεις που αλλάζουν τον τρόπο με τον οποίο αλληλεπιδρούμε με πληροφορίες και τεχνολογία.

Μία από τις πιο συνηθισμένες εφαρμογές της NLP είναι η αυτόματη μετάφραση. Η τεχνολογία αυτή τροφοδοτεί υπηρεσίες όπως το Google Translate, το DeepL και το Microsoft Translator, επιτρέποντας την άμεση μετάφραση κειμένων και ιστοσελίδων σε διάφορες γλώσσες. Επιπλέον, η δυνατότητα μετάφρασης ομιλίας σε πραγματικό χρόνο ανοίγει νέες δυνατότητες για επικοινωνία και συνεργασία σε πολυγλωσσικά περιβάλλοντα. Αυτές

οι εφαρμογές επιτρέπουν την άνετη επικοινωνία και την ανταλλαγή πληροφοριών ανεξάρτητα από τη γλώσσα που μιλά ο χρήστης, κάνοντας την πληροφορία πιο προσβάσιμη και κατανοητή σε διαφορετικά περιβάλλοντα. Η ανάλυση συναισθημάτων αποτελεί μία ακόμη σημαντική εφαρμογή της Φυσικής Γλώσσας (NLP). Μέσω της κατανόησης απόψεων, η NLP επεξεργάζεται κείμενα όπως κριτικές προϊόντων, σχόλια στα social media ή απαντήσεις σε έρευνες, προσφέροντας πολύτιμες πληροφορίες για την κατανόηση της κοινής γνώμης. Αυτό μπορεί να βοηθήσει επιχειρήσεις και οργανισμούς να αντλήσουν εποικοδομητικά συμπεράσματα σχετικά με την αντίληψη του κοινού για προϊόντα, υπηρεσίες ή γεγονότα. Επιπλέον, η ανάλυση συναισθήματος μπορεί να βοηθήσει στην εκτίμηση της δυσαρέσκειας πελατών και να προσφέρει έγκαιρες παρεμβάσεις για την επίλυση προβλημάτων. Αυτό είναι ιδιαίτερα χρήσιμο στην παροχή υπηρεσιών πελατών, καθώς επιτρέπει στις επιχειρήσεις να ανιχνεύουν και να αντιμετωπίζουν προβλήματα ή αντιδράσεις πελατών πριν αυτά εξελιχθούν σε μεγαλύτερα ζητήματα. Μία ακόμη εφαρμογή της Φυσικής Γλώσσας (NLP) είναι η αυτόματη περίληψη κειμένου, η οποία προσφέρει διάφορα οφέλη: Αρχικά, μέσω της συνοπτικής παρουσίασης πληροφοριών, η NLP μπορεί να δημιουργήσει περιλήψεις μεγάλων κειμένων, διατηρώντας τα βασικά σημεία και την ουσία του περιεχομένου. Αυτό επιτρέπει στους χρήστες να αποκτήσουν γρήγορα επισκόπηση του κειμένου χωρίς να χρειάζεται να διαβάσουν ολόκληρο το περιεχόμενο. Επιπλέον, η αυτόματη περίληψη συμβάλλει στην εξοικονόμηση χρόνου κατά την ανάγνωση μεγάλων άρθρων, εγγράφων ή επιστημονικών εργασιών. Αυτό επιτρέπει στους χρήστες να επικεντρωθούν μόνο στα κυρίαρχα σημεία του κειμένου χωρίς να χάνουν χρόνο στην ανάγνωση περιττών λεπτομερειών. Συνολικά, αυτή η δυνατότητα βελτιώνει την παραγωγικότητα και την αποτελεσματικότητα της εργασίας. Τα chatbots αποτελούν έναν ακόμη τομέα εφαρμογής της Φυσικής Γλώσσας (NLP) που προσφέρει πολλαπλά οφέλη: Αρχικά, οι εικονικοί βοηθοί που λειτουργούν με τη χρήση chatbots και NLP, όπως η Siri, η Alexa και η Cortana, επιτρέπουν τη φυσική γλωσσική αλληλεπίδραση με συσκευές και υπηρεσίες. Αυτό σημαίνει ότι οι χρήστες μπορούν να επικοινωνούν με τις συσκευές τους χρησιμοποιώντας τη φυσική τους φωνή για να πραγματοποιούν ερωτήσεις, να εκτελούν εντολές και να λαμβάνουν απαντήσεις ή να εκτελούν διάφορες λειτουργίες. Επιπλέον, οι chatbots χρησιμοποιούνται ευρέως για την εξυπηρέτηση πελατών. Αυτοί οι εικονικοί βοηθοί μπορούν να παρέχουν υποστήριξη στους πελάτες 24/7, απαντώντας σε ερωτήσεις, επιλύοντας προβλήματα και διεκπεραιώνοντας απλές συναλλαγές. Αυτό μπορεί να βελτιώσει σημαντικά την εμπειρία των πελατών και να μειώσει το φόρτο εργασίας του προσωπικού εξυπηρέτησης πε-

λατών, επιτρέποντάς τους να επικεντρωθούν σε πιο πολύπλοκες και εξειδικευμένες ανάγκες των πελατών. Τα συστήματα ερωτήσεων και απαντήσεων (Q&A) αποτελούν μια σημαντική εφαρμογή της Φυσικής Γλώσσας (NLP) με διάφορες χρήσεις: Πρώτον, στην αναζήτηση πληροφοριών, συστήματα Q&A όπως η Google Search και η Siri χρησιμοποιούν την τεχνολογία NLP για να κατανοήσουν τις ερωτήσεις των χρηστών και να παρέχουν σχετικές απαντήσεις από διάφορες πηγές. Αυτό επιτρέπει στους χρήστες να αναζητούν πληροφορίες με βάση τις ανάγκες τους, λαμβάνοντας γρήγορα και ακριβείς απαντήσεις. Επιπλέον, η τεχνολογία Q&A μπορεί να ενσωματωθεί σε εκπαιδευτικά εργαλεία, παρέχοντας στους μαθητές τη δυνατότητα να θέτουν ερωτήσεις και να λαμβάνουν άμεσες και εμπεριστατωμένες απαντήσεις. Αυτό ενισχύει την εκπαιδευτική διαδικασία, επιτρέποντας στους μαθητές να αναζητούν και να αποκτούν γνώσεις με τρόπο πιο αποτελεσματικό και διαδραστικό. Εκτός από αυτές τις εφαρμογές, η τεχνολογία NLP αξιοποιείται σε πολλούς άλλους τομείς. Για παράδειγμα, εργαλεία γραμματικής διόρθωσης και spell checkers χρησιμοποιούν NLP για τον εντοπισμό και τη διόρθωση σφαλμάτων γραφής και στίξης. Επιπλέον, η τεχνολογία NLP χρησιμοποιείται για το φιλτράρισμα ανεπιθύμητων emails και μηνυμάτων, βοηθώντας να διατηρηθεί η επικοινωνία αποτελεσματική και ασφαλής.

Κεφάλαιο 3

Τελευταίες Έρευνες και Τάσεις

3.1 Προηγμένες Τεχνολογίες και Καινοτομίες

Σε αυτό το κεφάλαιο της εργασίας εξετάζονται οι προηγμένες τεχνολογίες και οι καινοτομίες που συνδέονται με τις πιο πρόσφατες εξελίξεις στον τομέα της επεξεργασίας φυσικής γλώσσας (NLP) και των συστημάτων μετατροπής κειμένου σε ομιλία (TTS). Αναλύονται οι συνέπειες και οι επιδράσεις της εφαρμογής προηγμένων μοντέλων μηχανικής μάθησης, η ολοκλήρωση της τεχνητής νοημοσύνης στις καθημερινές εφαρμογές, καθώς και οι σημαντικές προκλήσεις και οι δυνατότητες που προκύπτουν από αυτήν την τεχνολογία για το μέλλον. Το κεφάλαιο αναδεικνύει πώς αυτές οι καινοτομίες αλλάζουν την αντίληψη και την εφαρμογή των συστημάτων NLP και TTS σε διάφορους τομείς, όπως η ρομποτική, η ψυχαγωγία και η υγεία, παρέχοντας έτσι μια ολοκληρωμένη εικόνα της σύγχρονης κατάστασης και των προοπτικών των τεχνολογιών αυτών.

3.1.1 Τελευταίες Έρευνες και Καινοτομίες στη Σύνθεση Φωνής και Ανάλυση Φυσικής Γλώσσας

Οι τομείς της Σύνθεσης Φωνής (TTS) και Ανάλυσης Φυσικής Γλώσσας (NLP) εξελίσσονται ραγδαία, με νέες έρευνες και τεχνολογικές καινοτομίες να φέρνουν επανάσταση στις δυνατότητες επεξεργασίας φωνής και κειμένου. Ας δούμε μερικά από τα πιο συναρπαστικά τρέχοντα ερευνητικά πεδία:

Σύνθεση Φωνής:

- **Neuronal Vocoders:** Η νευρωνική κωδικοποίηση φωνής βελτιώνει την ποιότητα της

παραγόμενης φωνής, μειώνοντας τα τεχνητά ή ρομποτικά χαρακτηριστικά και προσφέροντας πιο ρεαλιστικά και φυσικά ηχητικά αποτελέσματα.[17]

- **Multispeaker TTS:** Η δυνατότητα σύνθεσης φωνής από πολλούς ομιλητές ανοίγει νέες δυνατότητες για την δημιουργία φωνητικών περιεχομένων, όπως audiobooks, podcast και εικονικούς βοηθούς με διαφορετικές φωνές.[18]
- **Emotional TTS:** Η έρευνα εστιάζει στην δημιουργία φωνής με συναισθήματα, επιτρέποντας την έκφραση συναισθημάτων όπως χαρά, θλίψη ή θυμός μέσω φωνητικών χαρακτηριστικών.[19]

Ανάλυση Φυσικής Γλώσσας:

- **Transformers:** Η αρχιτεκτονική Transformers έχει φέρει επανάσταση στην NLP, προσφέροντας παράλληλη επεξεργασία δεδομένων ακολουθιών και βελτιώνοντας σημαντικά την ακρίβεια σε εργασίες όπως η μετάφραση μηχανών, η περίληψη κειμένου και η απάντηση σε ερωτήσεις.
- **Explainable AI:** Η έρευνα εστιάζει στην ανάπτυξη συστημάτων NLP που είναι πιο "διαφανή" και εξηγήσιμα, επιτρέποντας την κατανόηση του τρόπου με τον οποίο λαμβάνουν αποφάσεις.[20]
- **Generative NLP:** Η δημιουργική NLP εστιάζει στην ανάπτυξη μοντέλων που μπορούν να δημιουργήσουν κείμενο, κώδικα, μουσική και άλλο περιεχόμενο, ανοίγοντας νέες δυνατότητες για δημιουργικές εφαρμογές.[21]

3.1.2 Ηθικές και Κοινωνικές Επιπτώσεις

Ηθικά Ζητήματα και Ρύθμιση Σύγχρονων Τεχνολογιών NLP και TTS

Η ραγδαία ανάπτυξη και αξιοποίηση σύγχρονων τεχνολογιών, όπως τα Transformers, GANs και πολυτροπικά μοντέλα, στους τομείς της Σύνθεσης Φωνής (TTS) και Ανάλυσης Φυσικής Γλώσσας (NLP), φέρνει μαζί της και σημαντικά ηθικά ζητήματα που χρήζουν προσοχής και διεξοδικής συζήτησης. **Παραβίαση Ιδιωτικότητας:**

- **Συλλογή και Αποθήκευση Δεδομένων:** Η ανάπτυξη και εκπαίδευση μοντέλων NLP και TTS απαιτεί μεγάλο όγκο δεδομένων, συμπεριλαμβανομένων φωνητικών εγγράφων και κειμένου. Η συλλογή και αποθήκευση αυτών των δεδομένων εγείρει ανησυ-

χίες σχετικά με την προστασία της ιδιωτικότητας και την πιθανότητα μη εξουσιοδοτημένης πρόσβασης ή χρήσης.[22]

- **Βιομετρική Αναγνώριση:** Η ικανότητα αναγνώρισης ατόμων με βάση τη φωνή τους, που τροφοδοτείται από μοντέλα TTS, δημιουργεί κινδύνους για παράνομη παρακολούθηση και καταγραφή φωνητικών δεδομένων.

Παραπληροφόρηση και Deepfakes:

- **Δημιουργία Ψεύτικων Φωνών:** Η τεχνολογία TTS μπορεί να χρησιμοποιηθεί για τη δημιουργία ρεαλιστικών φωνών που μιλούν ψευδή ή παραπλανητικά μηνύματα, προκαλώντας σύγχυση και διασπώντας την κοινωνική συνοχή. [23]
- **Δυσκολία Διακρίσεων:** Η αυξημένη ρεαλιστικότητα των φωνών που δημιουργούνται με TTS καθιστά δυσκολότερη την διάκριση αληθινών από ψεύτικες. [24]

Κοινωνικές Επιπτώσεις της Αυτοματοποίησης:

- **Απώλεια Εργασίας:** Η αυτοματοποίηση εργασιών που βασίζονται σε γλώσσα, όπως η μετάφραση, η συγγραφή και η εξυπηρέτηση πελατών, μπορεί να οδηγήσει σε απώλεια θέσεων εργασίας και κοινωνικές αναταραχές. [25]
- **Αλγόριθμοι και Προκαταλήψεις:** Η ανάπτυξη μοντέλων NLP και TTS βασίζεται σε δεδομένα που μπορεί να φέρουν προκαταλήψεις, οι οποίες ενδέχεται να αναπαραχθούν και να ενισχυθούν από τα ίδια τα μοντέλα, οδηγώντας σε άνιση μεταχείριση και διακρίσεις. [26]

Προτάσεις για Υπεύθυνη Χρήση και Ρύθμιση:

- **Διαφάνεια και Εξηγήσιμη Τεχνητή Νοημοσύνη:** Η ανάπτυξη διαφανών και εξηγήσιμων μοντέλων NLP και TTS είναι απαραίτητη για την κατανόηση του τρόπου λειτουργίας τους και την ενίσχυση της εμπιστοσύνης. [27]
- **Εθνικοί και Διεθνείς Κανονισμοί:** Απαιτείται η θέσπιση νομικών πλαισίων και κανονισμών που να διασφαλίζουν την ηθική χρήση και την προστασία δεδομένων στους τομείς NLP και TTS. [28]
- **Εκπαίδευση και Ευαισθητοποίηση:** Η εκπαίδευση του κοινού σχετικά με τις δυνατότητες και τους κινδύνους των σύγχρονων τεχνολογιών NLP και TTS είναι απαραίτητη για την προώθηση της υπεύθυνης χρήσης και την αποφυγή παραπληροφόρησης. [29]

- **Συνεργασία Πολλαπλών Φορέων:** Η αντιμετώπιση των ηθικών ζητημάτων απαιτεί συνεργασία μεταξύ κυβερνήσεων, ερευνητικών φορέων, οργανισμών της κοινωνίας των πολιτών και της βιομηχανίας.

Κεφάλαιο 4

Μεθοδολογία Ανάπτυξης της εφαρμογής

Η ανάπτυξη της εφαρμογής έγινε σταδιακά και στα διάφορα στάδια προστίθενταν λειτουργίας μου φαινόοντουσαν χρήσιμες για έναν χρήστη και την άμεση επαφή με ένα large language model και τις δυνατότητές του. Για τη για την ανάπτυξη της συγκεκριμένης εφαρμογής χρησιμοποιήθηκαν βιβλιοθήκες η γλώσσα προγραμματισμού Python3.7.9 και η πλατφόρμα του ChatGPT. Η όλη αρχιτεκτονική της εφαρμογής αποτελείται από ένα απλό περιβάλλον χρήστη μέσα από το οποίο χρήστης μπορεί να επιλέξει αν θέλει να κάνει μια συζήτηση με αναγνώριση φωνής και παραγωγή φωνής από την απάντηση την οποία λαμβάνει ή αν θέλει με βάση την φωνή του να δημιουργήσει μια εικόνα.

4.1 Βιβλιοθήκες που χρησιμοποιήθηκαν

Οι βιβλιοθήκες που χρησιμοποιήθηκαν για την ανάπτυξη της εφαρμογής και λεπτομέρειες σχετικά με αυτές είναι οι ακόλουθες

- **tkinter**

- Το Tkinter είναι μια Python που συνδέεται με την εργαλειοθήκη Tk GUI. Είναι η τυπική διεπαφή Python για την εργαλειοθήκη Tk GUI και είναι το de facto τυπικό GUI της Python. Το Tkinter περιλαμβάνεται με τυπικές εγκαταστάσεις Linux, Microsoft Windows και macOS της Python. Το όνομα Tkinter προέρχεται από τη διεπαφή Tk.

- **speech_recognition**

- Η βιβλιοθήκη SpeechRecognition, γνωστή και ως SpeechRecognition ή speak_recognition, είναι μια βιβλιοθήκη Python που παρέχει μια απλή και φιλική προς το χρήστη διεπαφή για εργασία με διάφορες μηχανές ASR, όπως το Google Web Speech API, το CMU Sphinx και άλλα.

- **pyttsx3**

- Το pyttsx3 είναι μια βιβλιοθήκη Python που μας επιτρέπει να μετατρέπουμε κείμενο σε ομιλία. Έτσι, θα του παρέχουμε το κείμενό μας και θα μετατρέψει αυτό το κείμενο σε ήχο. Είναι ένα περιτύλιγμα γύρω από πολλές μηχανές μετατροπής κειμένου σε ομιλία, συμπεριλαμβανομένου του μηχανισμού μετατροπής κειμένου σε ομιλία (TTS) της Microsoft.

- **openai**

- Η βιβλιοθήκη OpenAI Python παρέχει εύκολη πρόσβαση στο OpenAI REST API από οποιαδήποτε εφαρμογή Python 3.7+. Η βιβλιοθήκη περιλαμβάνει ορισμούς τύπων για όλες τις παραμέτρους αιτημάτων και τα πεδία απόκρισης και προσφέρει τόσο σύγχρονους όσο και ασύγχρονους πελάτες που υποστηρίζονται από httpx.

- **pathlib**

- Αυτή η ενότητα προσφέρει κλάσεις που αντιπροσωπεύουν διαδρομές συστήματος αρχείων με σημασιολογία κατάλληλη για διαφορετικά λειτουργικά συστήματα. Οι κατηγορίες μονοπατιών χωρίζονται σε καθαρά μονοπάτια, τα οποία παρέχουν καθαρά υπολογιστικές λειτουργίες χωρίς I/O, και σε συγκεκριμένα μονοπάτια, τα οποία κληρονομούν από καθαρά μονοπάτια αλλά παρέχουν επίσης λειτουργίες I/O.

- **PIL**

- Η Python Imaging Library προσθέτει δυνατότητες επεξεργασίας εικόνας στον διερμηνέα Python. Αυτή η βιβλιοθήκη παρέχει εκτεταμένη υποστήριξη μορφής αρχείου, αποτελεσματική εσωτερική αναπαράσταση και αρκετά ισχυρές δυνατότητες επεξεργασίας εικόνας. Η βασική βιβλιοθήκη εικόνων έχει σχεδιαστεί για γρήγορη πρόσβαση σε δεδομένα που είναι αποθηκευμένα σε μερικές βασικές

μορφές pixel. Θα πρέπει να παρέχει μια σταθερή βάση για ένα γενικό εργαλείο επεξεργασίας εικόνας.

- **requests**

- Requests allows you to send HTTP/1.1 requests extremely easily. There's no need to manually add query strings to your URLs, or to form-encode your PUT & POST data — but nowadays, just use the json method!

- **BytesIO**

- Η ενότητα Python io μας επιτρέπει να διαχειριζόμαστε τις λειτουργίες εισόδου και εξόδου που σχετίζονται με το αρχείο. Το πλεονέκτημα της χρήσης της μονάδας IO είναι ότι οι διαθέσιμες κλάσεις και συναρτήσεις μας επιτρέπουν να επεκτείνουμε τη λειτουργικότητα για να ενεργοποιήσουμε την εγγραφή στα δεδομένα Unicode.

Στην πρώτη περίπτωση ο χρήστης πατάει το κουμπί για να ηχογραφήσει η εφαρμογή στη συνέχεια το μετατρέπει σε κείμενο ρωτάει το ChatGPT και η απάντηση η οποία λαμβάνει σε κείμενο μετατρέπεται σε φωνή. Με τον ίδιο τρόπο και στην δεύτερη λειτουργία της εφαρμογής ο χρήστης ηχογραφεί την φωνή του το κείμενο μετατρέπεται σε ερώτηση προς το ChatGPT και στη συνέχεια η απάντηση περιλαμβάνει ένα σύνδεσμο διαδικτύου ο οποίος περιέχει την εικόνα την οποία το ChatGPT παράγαγε. Υπήρχαν διάφορες εκδόσεις της εφαρμογής πριν αυτή φτάσει στην τελική της μορφή και θα υπάρξουν διάφορα στιγμιότυπα οθόνης τα οποία θα δείχνουν τα βήματα τα οποία ακολουθήθηκαν για την ολοκλήρωσή της. Η πρώτη έκδοση της εφαρμογής αποτελούνταν από μία εφαρμογή η οποία έτρεχε σε τερματικό στον υπολογιστή.

Στην δεύτερη κιόλας έκδοση της εφαρμογής δημιουργήθηκε και ένα απλό περιβάλλον χρήστη το οποίο έκανε την όλη διαδικασία χρήσης της εφαρμογής πολύ πιο εύκολη πολύ πιο κατανοητή για χρήστη ο οποίος πρώτη φορά έρχεται σε επαφή μαζί της.

Με τη χρήση της συγκεκριμένης εφαρμογής ο χρήστης μπορούσε πιο εύκολα να διαλέξει την γλώσσα στην οποία θα μιλήσει το μοντέλο το οποίο θα παράγει την απάντησή του καθώς επίσης και το κουμπί της έναρξης της ηχογράφησης και την λήξη της ηχογράφησης επίσης το κείμενο το οποίο έλεγε ο χρήστης και λάμβανε σαν απάντηση από το μοντέλο εμφανίζονταν στο ακόλουθο πλαίσιο. Στην έκδοση η οποία ακολούθησε βάλαμε περισσότερο

```

Please choose a language: English or Greek
English
result2:
{ 'alternative': [ { 'confidence': 0.97457761,
                    'transcript': 'tell me one sentence about computer '
                    'science'},
                  { 'confidence': 0.97457761,
                    'transcript': 'tell me one sentence about komputer '
                    'science'},
                  { 'confidence': 0.97457761,
                    'transcript': 'tell me one sentence about computer '
                    'psyence'},
                  { 'confidence': 0.97457761,
                    'transcript': 'tell me one sentence about computer '
                    'psience'},
                  { 'confidence': 0.97457761,
                    'transcript': 'tell me one sentence about komputer '
                    'psyence'}],
  'final': True}
You said: tell me one sentence about computer science
GPT's response: Computer science is the study of the theory, design, development, and application of computer systems and software.

```

Σχήμα 4.1: The First version of the app that runs in terminal

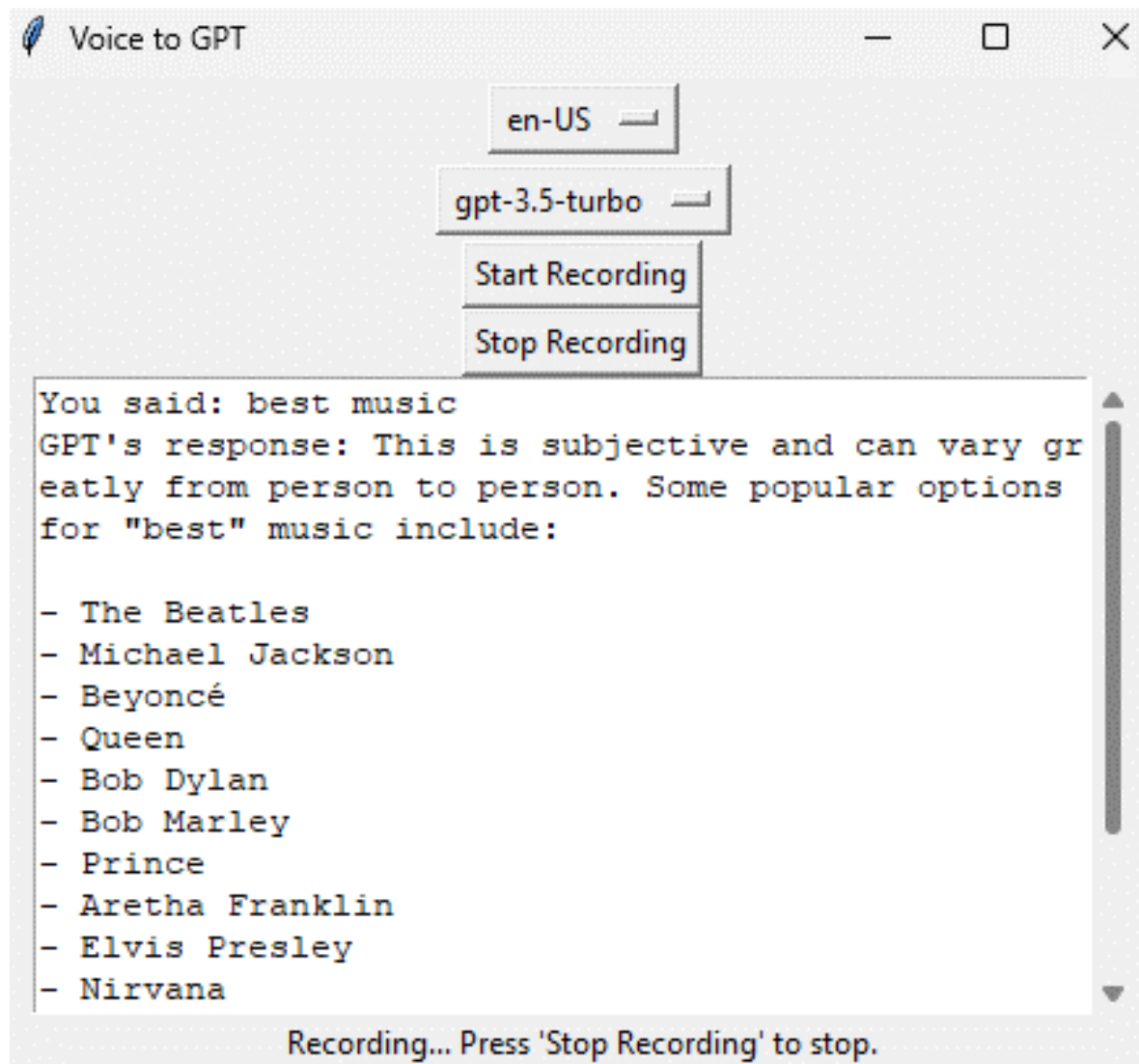
χρώμα στο σχεδιασμό της εφαρμογής f όπου μπορούμε εύκολα να ξεχωρίσουμε το κείμενο εισόδου που έδωσε ο χρήστης κατά την ηχογράφηση και την απάντηση την οποία λάβαμε από το εκπαιδευμένο μοντέλο f όπως φαίνεται και στην εικόνα που ακολουθεί

Στην έκδοση της εφαρμογής η οποία ακολούθησε αντιμετωπίστηκαν προβλήματα τα οποία εμπόδιζαν την ομαλή λειτουργία της προηγούμενης και προστέθηκαν μερικά χαρακτηριστικά όπως ένα παράθυρο το οποίο θα μπορεί ο χρήστης να κάνει scroll προς τα πάνω και να βλέπει να βλέπει τα προηγούμενα μηνύματα και απαντήσεις κατά την χρήση της εφαρμογής αλλά επίσης προστέθηκε και η δυνατότητα μετάφρασης τις απαντήσεις τις οποίες είχαμε λάβει από το μοντέλο.

Επίσης το μεταφρασμένο κείμενο το οποίο εμφανίζουμε με διαφορετικό χρώμα στην εφαρμογή παρέχεται η δυνατότητα ανάγνωσης του με τη βοήθεια του ίδιου μοντέλου επίσης το αντιπροσωπεύουμε με διαφορετικό χρώμα για να είναι πιο εύκολος εντοπισμός του στο παράθυρο με τα προηγούμενα μηνύματα.

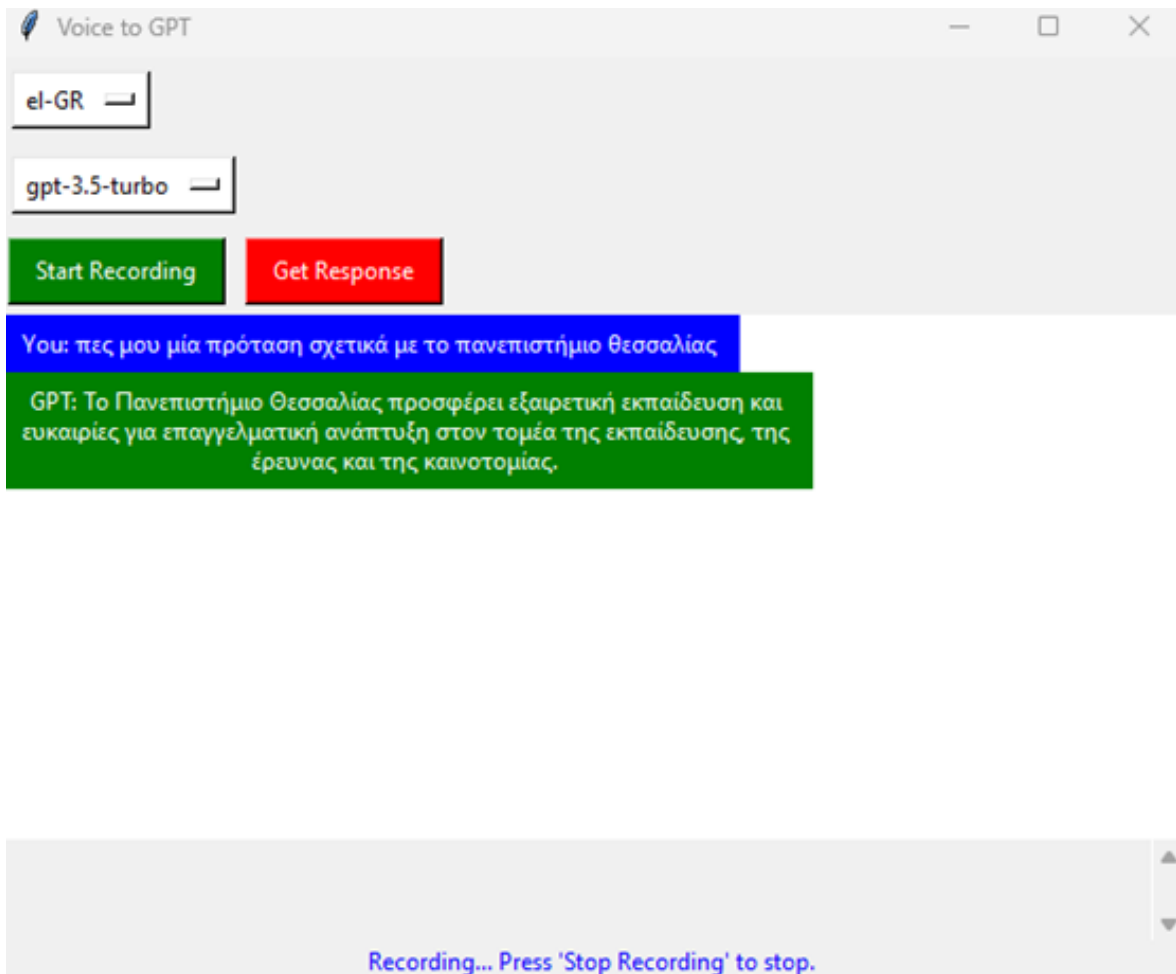
Στην πέμπτη και τελική έκδοση της εφαρμογής έχει προστεθεί η δυνατότητα παραγωγής εικόνας μέσα από την είσοδο την οποία δίνει ο χρήστης στην εφαρμογή ηχογραφώντας πάλι τη φωνή του και στην απάντηση την οποία λαμβάνουμε από το μοντέλο έχουμε την εικόνα η οποία μπορεί να είναι οτιδήποτε ζητήσουμε όπως και στο τραβηγμένο παράδειγμα που βλέπουμε να ακολουθεί

Ενδεικτικά το transcript το οποίο χρησιμοποιήθηκε για να παραχθεί η ανωτέρα φωτογραφία 'transcript': 'a cat riding a unicorn on the moon and in the background I want Earth a nuclear explosion seen from space' Η παρούσα φωτογραφία είναι παράδειγμα στο οποίο μπορούμε να καταλάβουμε ότι η φωτογραφία την οποία παράγει το μοντέλο δεν είναι αναγ-



Σχήμα 4.2: The Second version of the app that runs with User Interface

καίο να έχει ήδη υπάρξει νωρίτερα ούτε να βγάλει κάποιο νόημα το περιεχόμενό της όσο εμείς τις περιγράφουμε και τις ζητάμε να συμπεριλάβει αντικείμενα τα οποία ήδη γνωρίζει όπως στην προκειμένη περίπτωση είναι η γάτα ο μονόκερος είχε και μια πυρηνική έκρηξη. Στο Παράρτημα Α φαίνονται οι αναλυτικές οδηγίες χρήσης της εφαρμογής παρά τον απλό τρόπο με τον οποίο σχεδιάστηκε έτσι ώστε να είναι εύκολα χρησιμοποιούμενη από οποιονδήποτε χρήστη οποιοδήποτε τεχνολογικού μπαγκράουντ. Επίσης θα ακολουθήσουν οδηγίες για την εγκατάσταση της εφαρμογής σε έναν νέο υπολογιστή και ότι χρειάζεται ο χρήστης να γνωρίζει προκειμένου να την χρησιμοποιήσει.

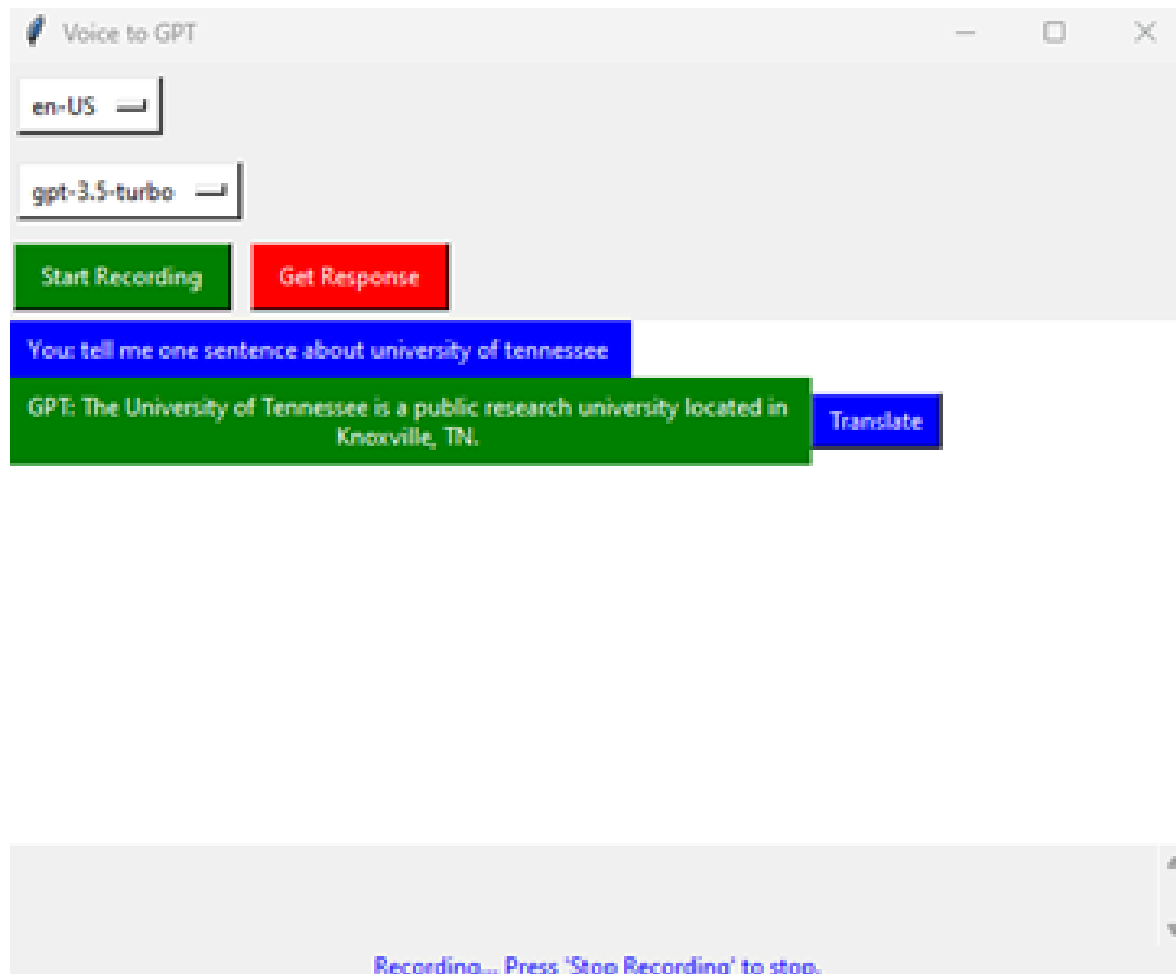


Σχήμα 4.3: The Third version of the app

4.1.1 Διαδικασία Ανάπτυξης Εφαρμογής

Η ανάπτυξη της εφαρμογής ξεκίνησε με την εγκατάσταση της γλώσσας προγραμματισμού Python, όπως αναφέρεται στο Παράρτημα Α, καθώς και των απαραίτητων βιβλιοθηκών. Αρχικά, η πιο απλή μορφή της εφαρμογής υλοποιήθηκε με περιβάλλον χρήσης στο τερματικό των Windows. Σε αυτή την αρχική φάση, ο χρήστης είχε τη δυνατότητα να επιλέξει τη γλώσσα στην οποία θα μιλήσει, και στη συνέχεια η φωνητική είσοδος μετατρεπόταν σε κείμενο.⁵⁰

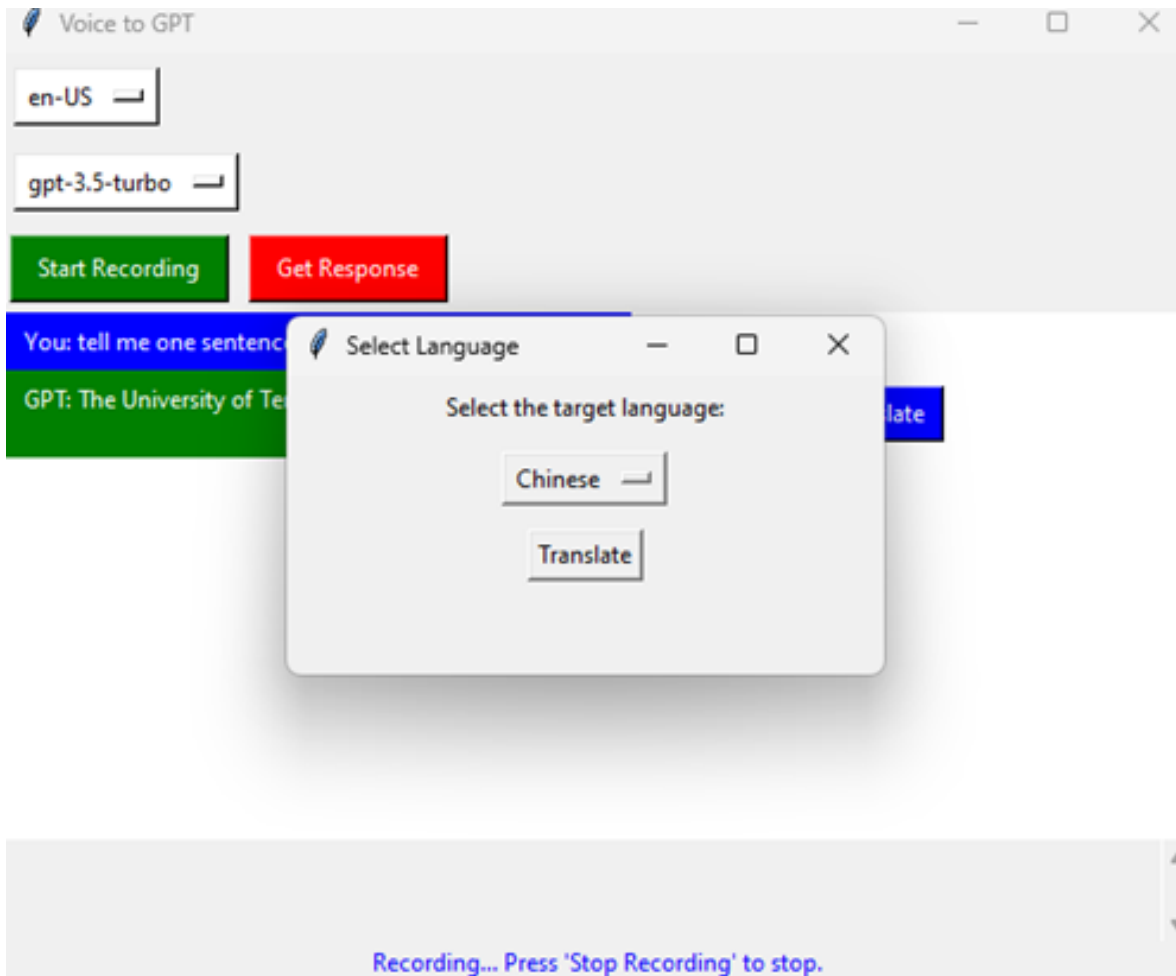
Μετά από αυτή τη βασική υλοποίηση, πραγματοποιήθηκε έρευνα σχετικά με τον τρόπο αποστολής του κειμένου σε ένα Μεγάλο Γλωσσικό Μοντέλο (Large Language Model, LLM) και πώς θα μπορούσαμε να εμφανίσουμε την απάντηση στον χρήστη. Αναδείχθηκε η ανάγκη για ένα πιο φιλικό και εύχρηστο περιβάλλον χρήστη σε σύγκριση με την απλή εκκίνηση της εφαρμογής μέσω του τερματικού. Ως εκ τούτου, εγκαταστάθηκε η βιβλιοθήκη Tkinter, η οποία χρησιμοποιήθηκε για τη δημιουργία ενός γραφικού περιβάλλοντος χρήστη (GUI).⁵¹



Σχήμα 4.4: The Fourth version of the app

Μέσα σε αυτό το περιβάλλον προστέθηκαν διάφορες λειτουργίες, όπως η δυνατότητα μετάφρασης και η δημιουργία φωνής για την ανάγνωση της απάντησης. Η σημαντικότερη προσθήκη όμως ήταν η ενσωμάτωση της λειτουργίας δημιουργίας εικόνας, αναδεικνύοντας τις ικανότητες των Μεγάλων Γλωσσικών Μοντέλων. Αυτή η λειτουργία επιτρέπει στο σύστημα να δημιουργεί οπτικό περιεχόμενο βάσει της εισόδου του χρήστη, επιδεικνύοντας την πολυπλοκότητα και τη δύναμη της τεχνολογίας των LLMs.⁵³

Συνολικά, η ανάπτυξη της εφαρμογής ακολούθησε μια συστηματική και σταδιακή προσέγγιση, από τις βασικές λειτουργίες έως τις πιο προχωρημένες, με στόχο τη δημιουργία ενός ολοκληρωμένου και εύχρηστου εργαλείου. Αυτή η διαδικασία περιλάμβανε την επιλογή κατάλληλων τεχνολογιών, τη συνεχή βελτίωση του περιβάλλοντος χρήστη και την προσθήκη νέων και καινοτόμων λειτουργιών που αξιοποιούν τις δυνατότητες των σύγχρονων γλωσσικών μοντέλων.⁵⁶



Σχήμα 4.5: The Fourth version of the app

4.1.2 Χρήση της Εφαρμογής

Η χρήση της εφαρμογής που αναπτύχθηκε σύμφωνα με την παραπάνω περιγραφή είναι απλή και φιλική προς τον χρήστη, ενσωματώνοντας προηγμένες τεχνολογίες NLP και TTS για πολυλειτουργικές δυνατότητες. Η διαδικασία χρήσης της εφαρμογής περιλαμβάνει τα εξής βήματα:

1. Εκκίνηση της Εφαρμογής:

- Εμφανίζεται το γραφικό περιβάλλον χρήστη (GUI) που έχει δημιουργηθεί με τη βιβλιοθήκη Tkinter.

2. Επιλογή Γλώσσας:

- Στην κύρια οθόνη, ο χρήστης επιλέγει τη γλώσσα που προτιμά να χρησιμοποιήσει από ένα αναπτυσσόμενο μενού, καθώς και το μεγάλο γλωσσικό μοντέλο που θέλει



Σχήμα 4.6: The Fourth version of the app

να χρησιμοποιήσει.

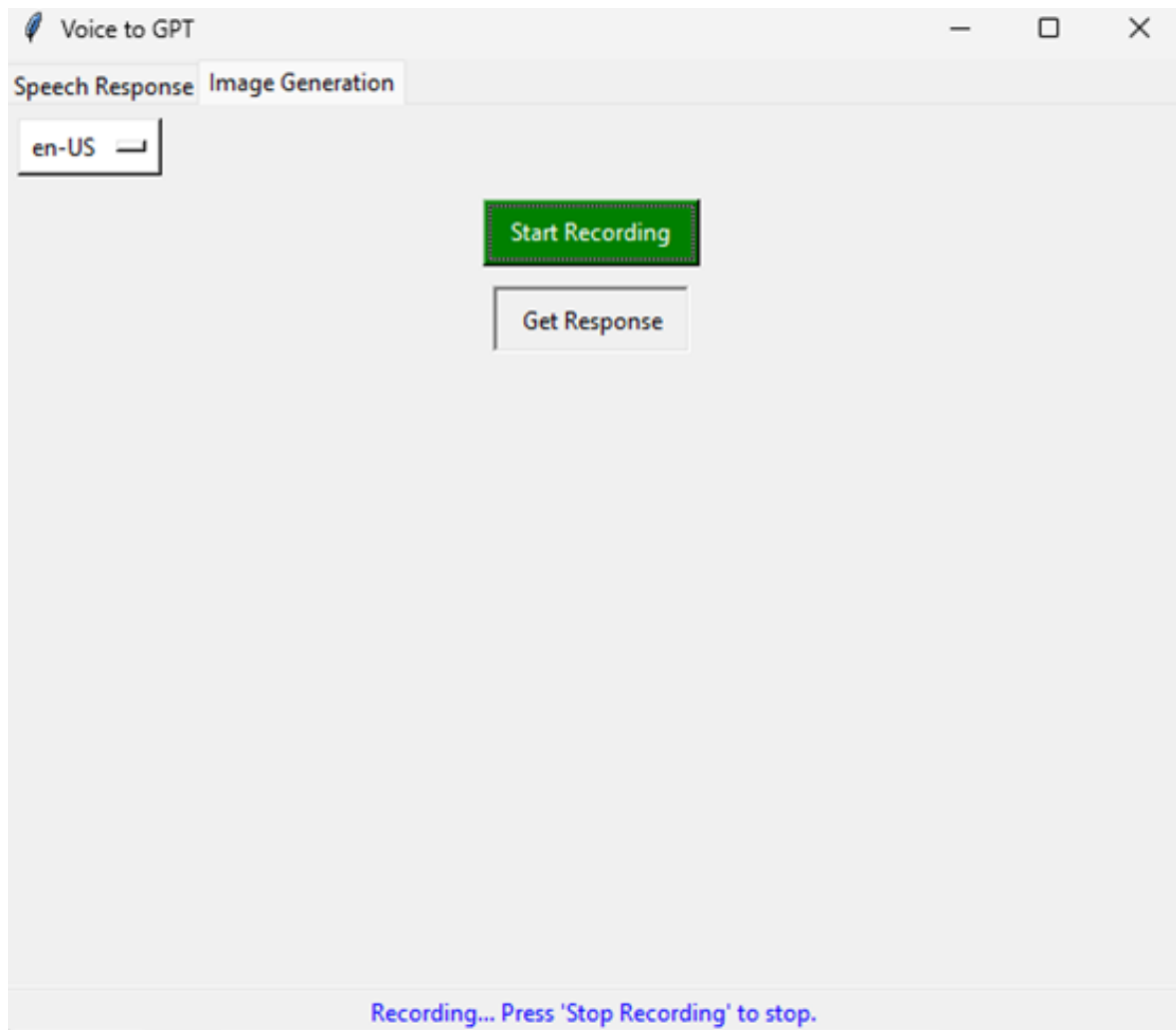
- Αυτή η επιλογή καθορίζει τη γλώσσα που θα χρησιμοποιηθεί για την ανάλυση της φωνητικής εισόδου και την παραγωγή της φωνητικής εξόδου.

3. Είσοδος Φωνής:

- Ο χρήστης μπορεί να πατήσει ένα κουμπί για να ξεκινήσει η καταγραφή της φωνητικής του εισόδου.
- Η εφαρμογή καταγράφει τη φωνή του χρήστη και τη μετατρέπει σε κείμενο χρησιμοποιώντας προηγμένες τεχνολογίες φωνητικής αναγνώρισης.

4. Αποστολή Κειμένου σε LLM:

- Το κείμενο που προέκυψε από τη φωνητική αναγνώριση αποστέλλεται σε ένα Μεγάλο Γλωσσικό Μοντέλο (LLM) για επεξεργασία.



Σχήμα 4.7: The Last version of the app that runs in terminal

- Το LLM αναλύει το κείμενο, κατανοεί το περιεχόμενο και παράγει μια απάντηση.

5. Απεικόνιση και Ανάγνωση Απάντησης:

- Η απάντηση από το LLM εμφανίζεται στην οθόνη του χρήστη.
- Παράλληλα, η εφαρμογή χρησιμοποιεί τεχνολογίες TTS για να μετατρέψει την κειμενική απάντηση σε φωνή, την οποία αναπαράγει για τον χρήστη.

6. Δημιουργία Εικόνας:

- Σε περίπτωση που το επιλέξει ο χρήστης, η εφαρμογή μπορεί να δημιουργήσει και μια σχετική εικόνα βάσει της απάντησης του LLM.
- Αυτή η εικόνα εμφανίζεται δίπλα στο κείμενο της απάντησης, ενισχύοντας την κατανόηση και την οπτικοποίηση της πληροφορίας.



Σχήμα 4.8: Ενδεικτική Παραγώμενη εικόνα

7. Περαιτέρω Δράσεις:

- Ο χρήστης έχει τη δυνατότητα να ζητήσει περισσότερες πληροφορίες, μετάφραση της απάντησης, να κάνει νέες ερωτήσεις ή να χρησιμοποιήσει άλλες λειτουργίες της εφαρμογής μέσω του φιλικού GUI.
- Όλες οι ενέργειες γίνονται μέσω εύχρηστων κουμπιών και μενού που προσφέρουν ξεκάθαρες επιλογές και οδηγίες.

Η εφαρμογή αυτή προσφέρει μια ολοκληρωμένη εμπειρία χρήστη, συνδυάζοντας τη φωνητική αλληλεπίδραση με τις δυνατότητες των σύγχρονων γλωσσικών μοντέλων και την

παραγωγή φωνητικού και οπτικού περιεχομένου. Είναι ένα ισχυρό εργαλείο που μπορεί να χρησιμοποιηθεί σε ποικίλα πεδία, από την εκπαίδευση μέχρι την ψυχαγωγία και την υποστήριξη ατόμων με αναπηρίες.

Κεφάλαιο 5

Περεταίρω Έρευνα

5.1 Μελλοντικές Προοπτικές να μεταφερθεί σε διαφορετικό κεφάλαιο στο τέλος με τίτλο περεταίρω ερευνα

Η συνεχής βελτίωση των Τεχνολογιών Φυσικής Γλώσσας (NLP) αποτελεί έναν ακατάπαυστο πορευτικό κλάδο με αμέτρητες προκλήσεις και ευκαιρίες. Μια από τις κύριες προκλήσεις είναι η επίτευξη μιας πιο βαθιάς κατανόησης και επεξεργασίας της ανθρώπινης γλώσσας από μηχανές, έτσι ώστε να μπορούν να αλληλεπιδρούν με τους ανθρώπους με πιο φυσικό τρόπο. Αυτό περιλαμβάνει την ανάπτυξη μοντέλων που μπορούν να αντιληφθούν τον συμβολισμό και τη σημασία πίσω από τις λέξεις, καθώς και την ικανότητα να εξάγουν πληροφορίες από το κείμενο και τον ομιλητή.

Με την υπέρβαση αυτών των προκλήσεων, υπάρχουν αμέτρητες ευκαιρίες για την εφαρμογή των NLP σε ποικίλους τομείς. Μια από αυτές είναι η δημιουργία ολοένα πιο έξυπνων εικονικών βοηθών και συστημάτων συνομιλίας που μπορούν να παρέχουν εξελιγμένη υποστήριξη στους χρήστες. Αυτό μπορεί να περιλαμβάνει την παροχή εξατομικευμένων πληροφοριών, τη διευκόλυνση της εκπαίδευσης και της εκμάθησης, καθώς και τη βελτίωση της ποιότητας των υπηρεσιών υποστήριξης πελατών.

Ωστόσο, με τις νέες ευκαιρίες που προκύπτουν, εμφανίζονται και νέες προκλήσεις. Μία από αυτές είναι η ανάγκη για την προστασία της ιδιωτικότητας καθώς και τη διαχείριση των δεδομένων που χρησιμοποιούνται για την εκπαίδευση των μοντέλων NLP. Επιπλέον, η ανάπτυξη αυτοματοποιημένων συστημάτων NLP θέτει προκλήσεις όσον αφορά την εξισορρόπηση μεταξύ της αυξημένης παραγωγικότητας και της απώλειας θέσεων εργασίας.

Επομένως, η πρόκληση που αντιμετωπίζουμε είναι να εκμεταλλευτούμε τις ευκαιρίες που προσφέρουν οι NLP με υπεύθυνο τρόπο, λαμβάνοντας υπόψη τις προκλήσεις που προκύπτουν και εργαζόμενοι προς την επίλυσή τους.

Ιδέες για Περαιτέρω Έρευνα για LLMs και AI Chatbots

1. Βελτίωση Κατανόησης και Παραγωγής Κειμένου

Η αντιμετώπιση των προκαταλήψεων στα γλωσσικά μοντέλα είναι κρίσιμη, καθώς τα μοντέλα μπορεί να αναπαράγουν ή και να ενισχύσουν τις υπάρχουσες κοινωνικές προκαταλήψεις που βρίσκονται στα δεδομένα εκπαίδευσης. Έρευνα μπορεί να εστιάσει στην ανάπτυξη αλγορίθμων που εντοπίζουν και διορθώνουν αυτές τις προκαταλήψεις. Η πολυγλωσσική κατανόηση είναι επίσης σημαντική, καθώς τα περισσότερα μοντέλα εκπαιδεύονται κυρίως σε αγγλικά δεδομένα. Η ενίσχυση της ικανότητας των μοντέλων να κατανοούν και να παράγουν κείμενο σε λιγότερο διαδεδομένες γλώσσες μπορεί να τα καταστήσει πιο προσιτά και χρήσιμα σε παγκόσμιο επίπεδο. Επιπλέον, η εξήγηση των αποφάσεων των μοντέλων μέσω διαφανών διαδικασιών είναι απαραίτητη για την οικοδόμηση εμπιστοσύνης μεταξύ των χρηστών και των AI συστημάτων.

2. Διασύνδεση με Εξωτερικά Συστήματα

Η ενσωμάτωση API μπορεί να επεκτείνει τις δυνατότητες των LLMs, επιτρέποντας την πρόσβαση σε πραγματικά δεδομένα και υπηρεσίες. Ένα AI chatbot που μπορεί να τραβήξει δεδομένα από εξωτερικά συστήματα, όπως χρηματιστηριακές τιμές ή ιατρικές βάσεις δεδομένων, μπορεί να παρέχει πιο ενημερωμένες και χρήσιμες απαντήσεις. Η έρευνα μπορεί να εστιάσει στην ανάπτυξη ασφαλών και αποδοτικών μεθόδων για την ενσωμάτωση αυτών των API. Η βελτίωση της διαδραστικότητας και της ανταπόκρισης σε πραγματικό χρόνο είναι επίσης κρίσιμη, ιδιαίτερα για εφαρμογές που απαιτούν άμεσες απαντήσεις, όπως η εξυπηρέτηση πελατών ή τα παιχνίδια.

3. Ανθρώπινη-Αυτόματη Συνεργασία

Τα εργαλεία υποστήριξης απόφασης που βασίζονται σε AI chatbots μπορούν να βελτιώσουν την αποτελεσματικότητα και την ακρίβεια των ανθρώπινων αποφάσεων. Για παράδειγμα, ένα chatbot που παρέχει πληροφορίες ή προτάσεις σε γιατρούς μπορεί να βελτιώσει τις διαγνώσεις και τις θεραπείες. Η προσωποποίηση και προσαρμογή της εμπειρίας χρήστη, λαμβάνοντας υπόψη τις ατομικές προτιμήσεις και το προφίλ του

χρήστη, μπορεί να καταστήσει τα AI chatbots πιο χρήσιμα και αποδοτικά. Έρευνα μπορεί να επικεντρωθεί στην ανάπτυξη αλγορίθμων που μαθαίνουν και προσαρμόζονται στις προτιμήσεις των χρηστών με την πάροδο του χρόνου.

4. Ασφάλεια και Ηθικά Ζητήματα

Η ασφαλής και υπεύθυνη χρήση των LLMs είναι υψίστης σημασίας, καθώς τα μοντέλα μπορεί να χρησιμοποιηθούν κακόβουλα για τη δημιουργία παραπληροφόρησης ή για άλλες επιβλαβείς ενέργειες. Έρευνα μπορεί να επικεντρωθεί στην ανάπτυξη αλγορίθμων και πρακτικών που ανιχνεύουν και προλαμβάνουν τέτοιες χρήσεις. Η προστασία της ιδιωτικότητας των δεδομένων των χρηστών είναι επίσης κρίσιμη, ειδικά όταν τα chatbots επεξεργάζονται ευαίσθητες πληροφορίες. Η ανάπτυξη τεχνικών που διασφαλίζουν την ανωνυμία και την ασφάλεια των δεδομένων μπορεί να αυξήσει την εμπιστοσύνη των χρηστών στα AI συστήματα.

5. Εφαρμογές σε Εξειδικευμένα Πεδία

Στην υγειονομική περίθαλψη, τα AI chatbots μπορούν να χρησιμοποιηθούν για τη διάγνωση και τη διαχείριση ασθενειών, καθώς και για την παροχή ψυχολογικής υποστήριξης. Έρευνα μπορεί να επικεντρωθεί στην ανάπτυξη ειδικών μοντέλων που κατανοούν ιατρική ορολογία και παρέχουν ακριβείς πληροφορίες και συμβουλές. Στον τομέα της εκπαίδευσης, τα LLMs μπορούν να προσαρμόσουν την εμπειρία μάθησης στις ανάγκες κάθε μαθητή, παρέχοντας εξατομικευμένη υποστήριξη και πόρους. Η έρευνα μπορεί να εξετάσει πώς αυτά τα μοντέλα μπορούν να ενσωματωθούν στις εκπαιδευτικές πλατφόρμες για να βελτιώσουν την ποιότητα της μάθησης. Στην εξυπηρέτηση πελατών, τα AI chatbots μπορούν να αυτοματοποιήσουν και να βελτιώσουν την αλληλεπίδραση με τους πελάτες, παρέχοντας γρήγορες και ακριβείς απαντήσεις.

6. Τεχνολογικές Βελτιώσεις

Η έρευνα στις αρχιτεκτονικές νευρωνικών δικτύων μπορεί να οδηγήσει σε νέα μοντέλα που προσφέρουν καλύτερη απόδοση και αποδοτικότητα. Εξετάζοντας και αναπτύσσοντας νέες αρχιτεκτονικές, οι ερευνητές μπορούν να βελτιώσουν την κατανόηση και την παραγωγή κειμένου από τα μοντέλα. Η βελτιστοποίηση των υπολογιστικών πόρων που απαιτούνται για την εκπαίδευση και τη λειτουργία των μοντέλων είναι επίσης σημαντική. Με την ανάπτυξη τεχνικών που μειώνουν τις απαιτήσεις σε υπολογιστικούς πόρους, τα LLMs μπορούν να γίνουν πιο προσιτά και βιώσιμα, επιτρέποντας την ευ-

ρύτερη χρήση τους σε διάφορες εφαρμογές και περιβάλλοντα.

Βιβλιογραφία

- [1] Science in the News. History of artificial intelligence. *Harvard University*, 2017.
- [2] P. Johri, S. K. Khatri, A. T. Al-Taani, M. Sabharwal, S. Suvanov, and A. Kumar. Natural language processing: History, evolution, application, and future work. In *Proceedings of 3rd International Conference on Computing Informatics and Networks: ICCIN 2020*, pages 365–375. Springer Singapore, 2021.
- [3] Bell Labs. Ιστορία της σύνθεσης φωνής, n.d.
- [4] A. V. D. Oord, S. Dieleman, H. Zen, K. Simonyan, O. Vinyals, A. Graves, and K. Kavukcuoglu. Wavenet: A generative model for raw audio. *arXiv preprint arXiv:1609.03499*, 2016.
- [5] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, and D. Amodei. Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33:1877–1901, 2020.
- [6] J. Devlin, M. W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018.
- [7] A. Conneau, K. Khandelwal, N. Goyal, V. Chaudhary, G. Wenzek, F. Guzmán, and V. Stoyanov. Unsupervised cross-lingual representation learning at scale. *arXiv preprint arXiv:1911.02116*, 2019.
- [8] J. Shen and R. Pang. Tacotron 2: Generating human-like speech from text. *Google*, 2017.
- [9] D. Jurafsky and J. H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. n.d.

- [10] C. D. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.
- [11] N. Goyal, J. Du, M. Ott, G. Anantharaman, and A. Conneau. Larger-scale transformers for multilingual masked language modeling. *arXiv preprint arXiv:2105.00572*, 2021.
- [12] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*, 2013.
- [13] R. A. Khan and J. S. Chitode. Concatenative speech synthesis: A review. *International Journal of Computer Applications*, 136(3):1–6, 2016.
- [14] Socrates Yiannoudes and Alexandros Peteinarelis. Parametric design: From code to fabrication. In *Proceedings of the 21st CAADRIA Conference*, 2016.
- [15] X. Tan, T. Qin, F. Soong, and T. Y. Liu. A survey on neural speech synthesis. *arXiv preprint arXiv:2106.15561*, 2021.
- [16] T. Sercu, C. Puhersch, B. Kingsbury, and J. Le Roux. End-to-end attention-based large vocabulary speech recognition. *arXiv preprint arXiv:1508.04395*, 2016.
- [17] Kundan Kumar, Rithesh Kumar, Thibault De Boissiere, Louis Gestin, Wai Zhong Teoh, Jose Sotelo, and Aaron C Courville. Melgan: Generative adversarial networks for conditional waveform synthesis. In *Advances in Neural Information Processing Systems*, volume 32. 2019.
- [18] Ye Jia, Yu Zhang, Ron Weiss, Quan Wang, Jonathan Shen, Fei Ren, and Yonghui Wu. Transfer learning from speaker verification to multispeaker text-to-speech synthesis. In *Advances in Neural Information Processing Systems*, volume 31. 2018.
- [19] C. Dai, Y. Cao, and X. Hu. Estimation of finger joint angle based on neural drive extracted from high-density electromyography. In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4820–4823, Honolulu, HI, USA, 2018.
- [20] R. Adhikary, K. Bose, S. Mukherjee, and B. Roy. Complexity of maximum cut on interval graphs. *Discrete & Computational Geometry*, 70(2):307–322, 2023.

- [21] Nicolas Gontier, Koustuv Sinha, Siva Reddy, and Christopher Pal. Measuring systematic generalization in neural proof generation with transformers. In *Advances in Neural Information Processing Systems*, volume 33, pages 22231–22242. 2020.
- [22] Ananda Bhowmick, John Duchi, Julien Freudiger, Gautam Kapoor, and Robert Rogers. Protection against reconstruction and its applications in private federated learning. *arXiv preprint arXiv:1812.00984*, 2018.
- [23] Xiang Yu, Yue Wang, Yujie Chen, Zhen Tao, Dong Xi, Shuyang Song, and Shuzhi Niu. Fake artificial intelligence generated contents (faigc): A survey of theories, detection methods, and opportunities. *arXiv preprint arXiv:2405.00711*, 2024.
- [24] Farhana Khanam, Fatema Akter Munmun, Nazia Afrose Ritu, Arka Khatua Saha, and Md Firoz. Text to speech synthesis: A systematic review, deep learning based architecture and future research direction. *Journal of Advances in Information Technology*, 13(5), 2022.
- [25] Carl Benedikt Frey and Michael A. Osborne. The future of employment: How susceptible are jobs to computerisation? *Technological forecasting and social change*, 114:254–280, 2017.
- [26] Tolga Bolukbasi, Kai-Wei Chang, James Y. Zou, Venkatesh Saligrama, and Adam T. Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems*, volume 29. 2016.
- [27] Hajo Felzmann, Eduard Fosch-Villaronga, Christoph Lutz, and Aurelia Tamò-Larrieux. Towards transparency by design for artificial intelligence. *Science and engineering ethics*, 26(6):3333–3361, 2020.
- [28] Max U. Scherer. Regulating artificial intelligence systems: Risks, challenges, competencies, and strategies. *Harv. JL & Tech.*, 29:353, 2015.
- [29] Hui Luan, Peter Geczy, Hoa Lai, Janice Gobert, Sheng-Jen Yang, Hiroaki Ogata, and Chin-Chung Tsai. Challenges and future directions of big data and artificial intelligence in education. *Frontiers in psychology*, 11:580820, 2020.

ΠΑΡΑΡΤΗΜΑΤΑ

Παράρτημα Α

Παράθεση κώδικα

```
import tkinter as tk
from tkinter import ttk, simpledialog
import speech_recognition as sr
import pyttsx3
import openai
from pathlib import Path
from PIL import Image, ImageTk
import requests
from io import BytesIO

# Initialize the OpenAI API client with your API key
openai.api_key = 'YOUR_API_KEY_HERE'

# Initialize the recognizer
r = sr.Recognizer()

audio_data = None

# Function to handle the recording process
def start_recording():
    global audio_data
```

```
try:
    with sr.Microphone() as source:
        # Adjust for ambient noise
        r.adjust_for_ambient_noise(source, duration=0.2)
        # Update the status label
        status_label.config(text="Recording...
        Press 'Stop Recording' to stop.")
        # Start recording
        audio_data = r.listen(source)
except Exception as e:
    status_label.config(text=f"Recording error: {e}")

# Function to handle stopping the recording and processing
#the input based on the active tab
def stop_recording(tab_name):
    global audio_data
    try:
        # Recognize audio using Google in the selected language
        MyText = r.recognize_google(audio_data, language=language_var.get())
        MyText = MyText.lower()

        if tab_name == "Speech Response":
            add_message("You", MyText, 'blue', 'w')
            response = ask_gpt(MyText, selected_model.get())
            add_message("GPT", response, 'green', 'e', True)
            SpeakText(response, language_var.get())
        elif tab_name == "Image Generation":
            generate_image(MyText)

    except Exception as e:
        status_label.config(text=f"Error: {e}")
```

```

# Function to convert text to speech in Greek or English
def SpeakText(command, language, speed=180):
    engine = pyttsx3.init()
    engine.setProperty('rate', speed)
    voice_mapping = {
        'el-GR': 'Microsoft Stefanos - Greek (Greece)',
        'en-US': 'Microsoft Zira Desktop - English (United States)'
    }
    try:
        voice_name = voice_mapping.get(language)
        voices = engine.getProperty('voices')
        for voice in voices:
            if voice.name.lower() == voice_name.lower():
                engine.setProperty('voice', voice.id)
                break
        engine.say(command)
        engine.runAndWait()
    except Exception as e:
        print(f"Error in SpeakText function: {e}")

# Function to interact with GPT API and keep track of
#previous messages
conversation_history = []

def ask_gpt(question, model):
    conversation_history.append({"role": "user", "content": question})
    response = openai.chat.completions.create(
        model=model,
        messages=conversation_history
    )
    gpt_response_text = response.choices[0].message.content.strip()
    conversation_history.append({"role": "assistant",

```

```
        "content": gpt_response_text}))
    return gpt_response_text

# Function to generate an image using
# DALL-E and display it in an iframe
def generate_image(prompt):
    try:
        response = openai.images.generate(
            model="dall-e-3",
            prompt=prompt,
            n=1,
            size="1024x1024"
        )
        image_url = response.data[0].url

        # Create a new window to display the image in an iframe
        image_window = tk.Toplevel(root)
        image_window.title("Generated Image")

        # Load the image
        image_response = requests.get(image_url)
        image_data = Image.open(BytesIO(image_response.content))
        photo = ImageTk.PhotoImage(image_data)

        # Create a label to display the image
        image_label = tk.Label(image_window, image=photo)
        image_label.image = photo
        image_label.pack()

    except Exception as e:
        status_label.config(text=f"Image generation error: {e}")

def update_language_code(*args):
```

```

global language_code
language_code = language_var.get()

# Function to add a message bubble to the conversation canvas
def add_message(sender, message, color,
                alignment, add_translate_button=False):
    global conversation_height
    bubble = tk.Label(conversation_frame_in_canvas,
                      text=f"{sender}: {message}", bg=color, fg="white", padx=10, pady=5,
                      bubble.grid(row=conversation_height, column=0, sticky=tk.W)
    if add_translate_button:
        translate_button =
            tk.Button(conversation_frame_in_canvas, text="Translate",
                      command=lambda m=message: open_translate_dialog(m),
                      bg="blue", fg="white", padx=5, pady=2)
        translate_button.grid(row=conversation_height,
                              column=1, sticky=tk.E)
    conversation_height += 1

def open_translate_dialog(message):
    translate_dialog = TranslateDialog(root, message)

# Class for the translation dialog
class TranslateDialog(tk.Toplevel):
    def __init__(self, parent, message):
        super().__init__(parent)
        self.message = message
        self.title("Select Language")
        self.geometry("300x150")

        tk.Label(self, text="Select the target language:").pack(pady=5)

```

```

self.target_language_var = tk.StringVar(value=language_options[0])
self.language_menu = tk.OptionMenu(self, self.target_language_var,
                                    *language_options)
self.language_menu.pack(pady=5)

self.translate_button = tk.Button(self, text="Translate",
                                   command=self.translate_message)
self.translate_button.pack(pady=5)

def translate_message(self):
    target_language = self.target_language_var.get()
    translation_request = f"Translate the following
                           text to {target_language}: {self.message}"
    try:
        translated_text = ask_gpt(translation_request, selected_model)
        add_message("Translated", translated_text, 'purple', 'e')
        SpeakText(translated_text, language_var.get())
    except Exception as e:
        status_label.config(text=f"Translation error: {e}")
    finally:
        self.destroy()

# Create the main application window
root = tk.Tk()
root.title("Voice to GPT")
root.minsize(600, 400)

# Define the available GPT models
gpt_models = [
    "gpt-3.5-turbo",
    "gpt-3.5-turbo-0125",
    "gpt-3.5-turbo-0301",

```

```
    "gpt-3.5-turbo-0613",  
    "gpt-3.5-turbo-1106",  
    "gpt-3.5-turbo-16k",  
    "gpt-3.5-turbo-16k-0613",  
    "gpt-3.5-turbo-instruct",  
    "gpt-3.5-turbo-instruct-0914",  
    "gpt-4",  
    "gpt-4-0613",  
]
```

```
languages = [  
    ("English", "en-US"),  
    ("Greek", "el-GR"),  
]
```

```
language_options = [  
    "Chinese",  
    "Spanish",  
    "English",  
    "Hindi",  
    "Arabic",  
    "Portuguese",  
    "Bengali",  
    "Russian",  
    "Japanese",  
    "Punjabi",  
    "German",  
    "Javanese",  
    "Wu (Shanghainese)",  
    "Korean",  
    "French"  
]
```

```
# Define a variable to hold the language code
language_var = tk.StringVar(value="en-US")
language_var.trace_add('write', update_language_code)

# Create a notebook for tabs
notebook = ttk.Notebook(root)
notebook.pack(expand=1, fill="both")

# Function to add common widgets to a tab
def add_common_widgets(tab_frame, tab_name):
    # Add the language dropdown
    language_frame = tk.Frame(tab_frame)
    language_frame.pack(side="top", fill="x")

    language_menu = tk.OptionMenu(language_frame, language_var, *[code
    language_menu.config(bg="white", fg="black", padx=5, pady=5)
    language_menu.pack(side="left", padx=5, pady=5)

    # Add the start recording button
    start_button = tk.Button(tab_frame, text="Start Recording",
                             command=start_recording, bg="green", fg="white", pa
    start_button.pack(side="top", padx=5, pady=5)

    # Add the stop recording button
    stop_button = tk.Button(tab_frame, text="Get Response",
                             command=lambda: stop_recording(tab_name), bg="red",
                             fg="white", padx=10, pady=5)
    stop_button.pack(side="top", padx=5, pady=5)

# Speech Response Tab
speech_response_tab = ttk.Frame(notebook)
```

```

notebook.add(speech_response_tab, text="Speech Response")
add_common_widgets(speech_response_tab, "Speech Response")

# Frame to contain the conversation canvas
conversation_frame = tk.Frame(speech_response_tab)
conversation_frame.pack(fill=tk.BOTH, expand=True)

# Canvas for displaying the conversation bubbles
conversation_canvas = tk.Canvas(conversation_frame, bg="white",
                                highlightthickness=0)
conversation_canvas.pack(side="top", fill="both", expand=True)

# Dropdown menu for selecting a GPT model
selected_model = tk.StringVar(value=gpt_models[0])
model_menu = tk.OptionMenu(conversation_frame, selected_model, *gpt_mod
model_menu.config(bg="white", fg="black", padx=5, pady=5)
model_menu.pack(side="left", padx=5, pady=5)

# Create a frame inside the canvas to hold the messages
conversation_frame_in_canvas = tk.Frame(conversation_canvas, bg="white"
conversation_canvas.create_window((0, 0), window=conversation_frame_in_

# Make the conversation area scrollable
scrollbar = tk.Scrollbar(conversation_frame, orient="vertical", command
scrollbar.pack(side="right", fill="y")
conversation_canvas.configure(yscrollcommand=scrollbar.set)

# Bind scrollbar to canvas
conversation_canvas.bind("<Configure>", lambda e:
    conversation_canvas.configure(
        scrollregion=conversation_canvas.bbox("all"))
conversation_canvas.bind_all("<MouseWheel>", lambda event:

```

```
        conversation_canvas.yview_scroll(int(-1*(event.delta/120)),
        "units"))

# Initial height of the conversation canvas
conversation_height = 0

# Image Generation Tab
image_generation_tab = ttk.Frame(notebook)
notebook.add(image_generation_tab, text="Image Generation")
add_common_widgets(image_generation_tab, "Image Generation")

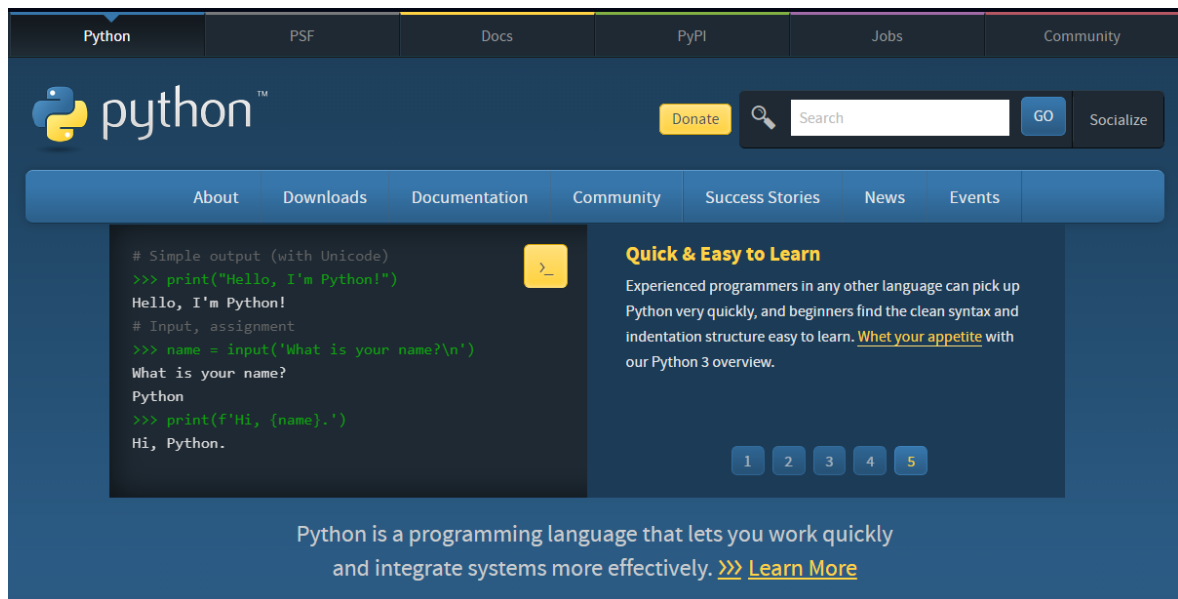
# Translation Tab
translation_tab = ttk.Frame(notebook)
# notebook.add(translation_tab, text="Translation")
# add_common_widgets(translation_tab, "Translation")

# Label for displaying status messages
status_label = tk.Label(root, text="", fg="blue")
status_label.pack()

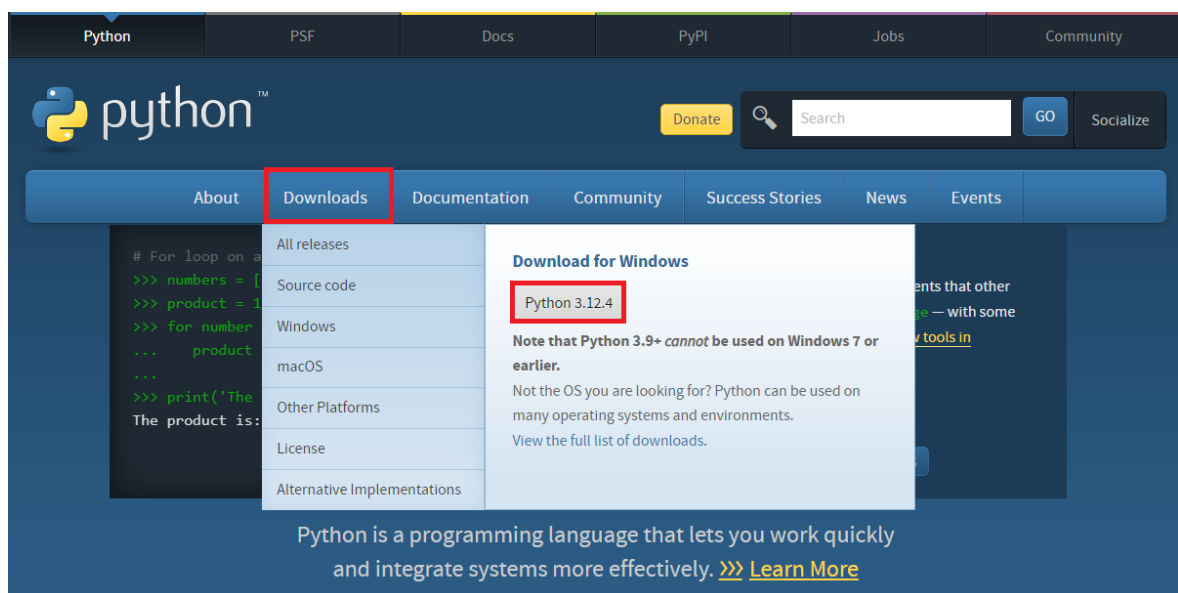
# Run the main event loop
root.mainloop()
```

A.0.1 Αποθετήριο κώδικα και εγκατάσταση εφαρμογής

Το σύνολο του κώδικα και οι διάφορες εκδόσεις της εφαρμογής βρίσκονται στο αποθε-
τήριο <https://github.com/Apostolos-Valiakos/Speakify>



Σχήμα A.1: Python Web Page

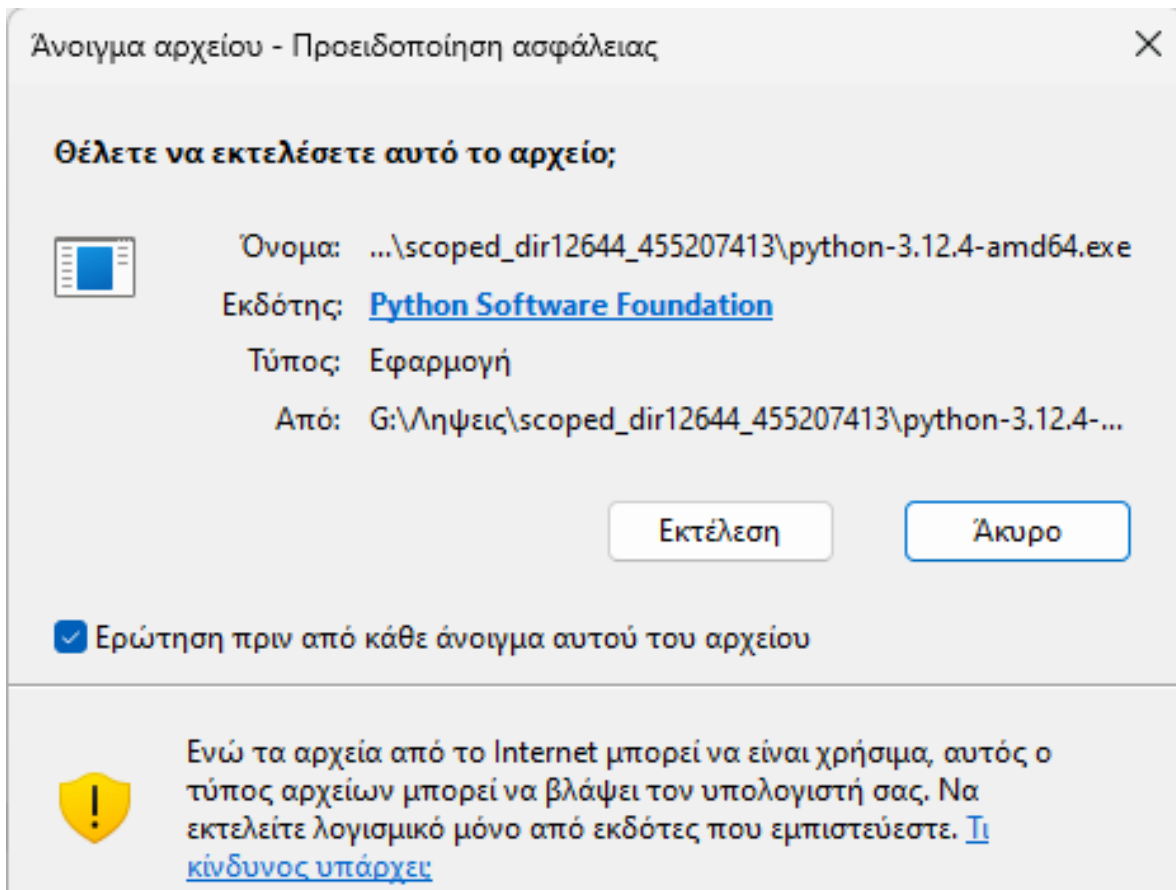


Σχήμα A.2: Last Version Of Python

A.1 Οδηγίες Εγκατάστασης

A.1.1 Εγκατάσταση Γλώσσας προγραμματισμού Python σε περιβάλλον Windows

Για να εγκαταστήσουμε την γλώσσα προγραμματισμού python επισκεπτόμαστε την επίσημη ιστοσελίδα από το σύνδεσμο <https://www.python.org> Επιλέγουμε το μενού downloads και εμφανίζεται η τελευταία έκδοση της Python Κάνοντας κλικ στο αρχείο που



Σχήμα Α.3: Step 1 of python installation

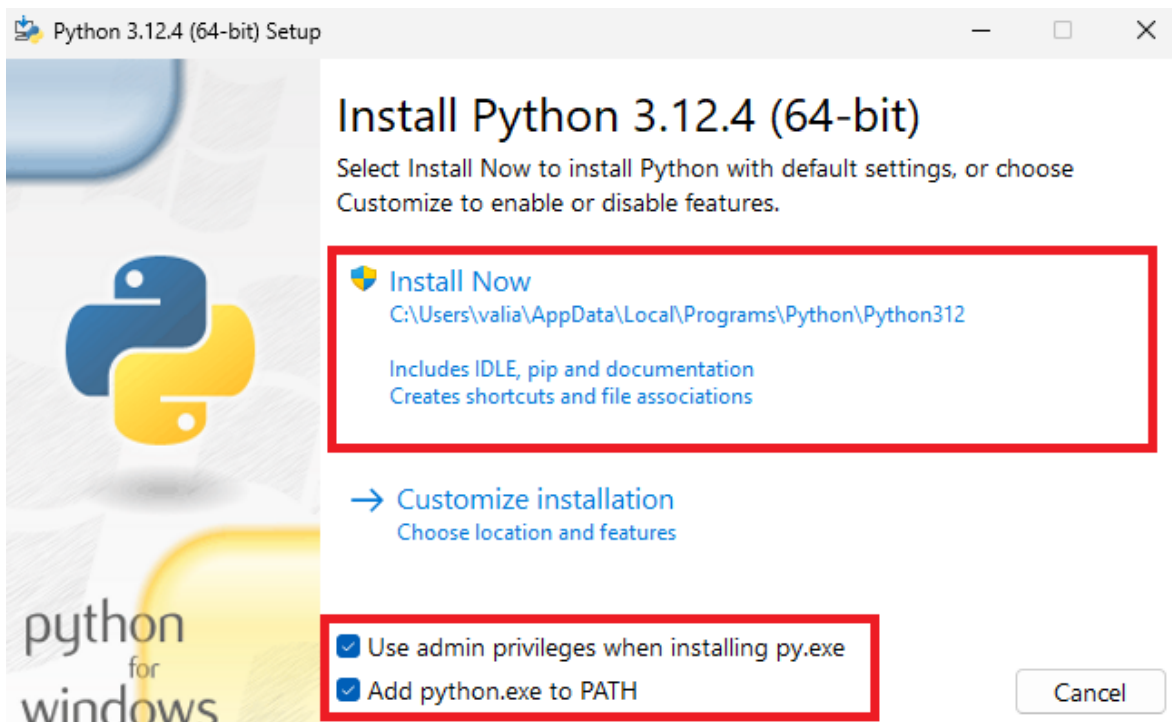
κατέβηκε, εμφανίζεται το παράθυρο το οποίο μας ζητάει τι θέλουμε να κάνουμε με το συγκεκριμένο αρχείο και επιλέγουμε "Εκτέλεση". Στη συνέχεια επιλέγουμε και τα 2 checkboxes που εμφανίζονται στο κάτω μέρος (σχήμα Α3) και πατάμε Install Now. Η διαδικασία της εγκατάστασης ξεκινά (σχήμα Α4). Τελικά εμφανίζεται το μήνυμα επιτυχούς εγκατάστασης.

Α.1.2 Εγκατάσταση Βιβλιοθηκών

Στη συνέχεια χρειάζεται να κάνουμε εγκατάσταση τα απαραίτητα πακέτα (βιβλιοθήκες) που είναι απαραίτητα για την λειτουργία της εφαρμογής. Για την εγκατάσταση χρειάζεται να τρέξουμε ως διαχειριστής το libraries.bat αρχείο που βρίσκεται στο project.

Α.1.3 Χρήση του OpenAI API

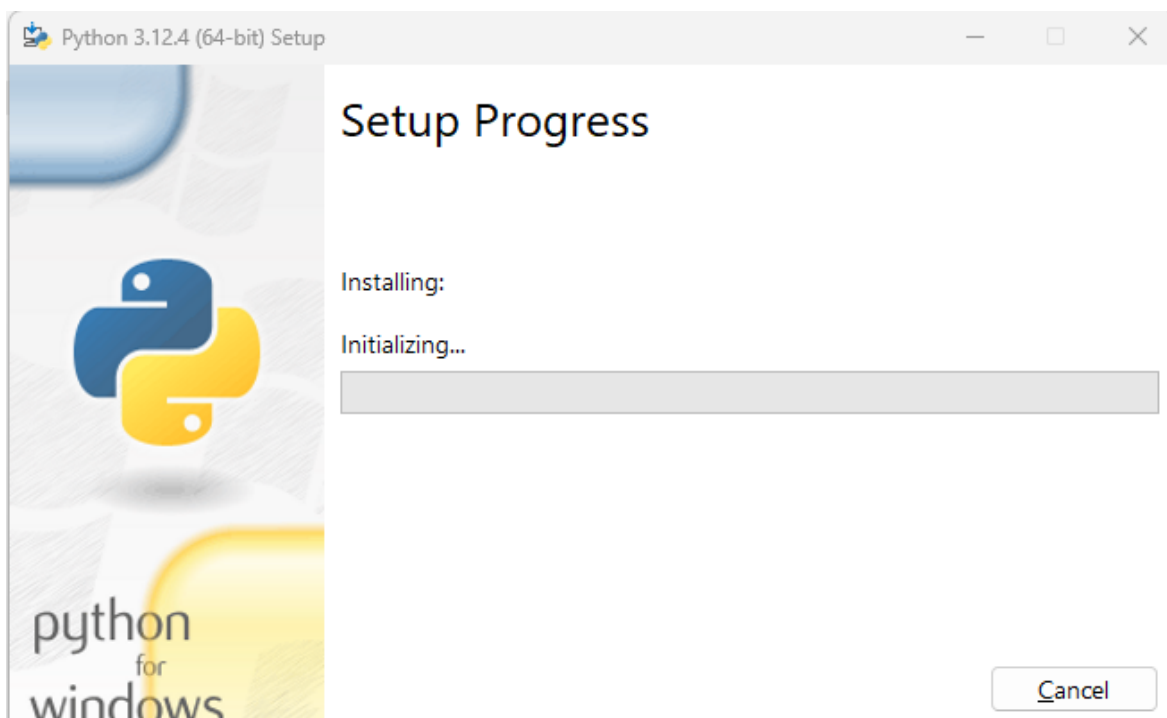
Στη συνέχεια, από την σελίδα που OpenAI <https://platform.openai.com/organization/usage>, κάνουμε εγγραφή πατώντας στο κουμπί register. Στη συνέχεια επιλέγουμε το πράσινο κουμπί increase limit στο δεξί μέρος της σελίδας όπως φαίνεται στο



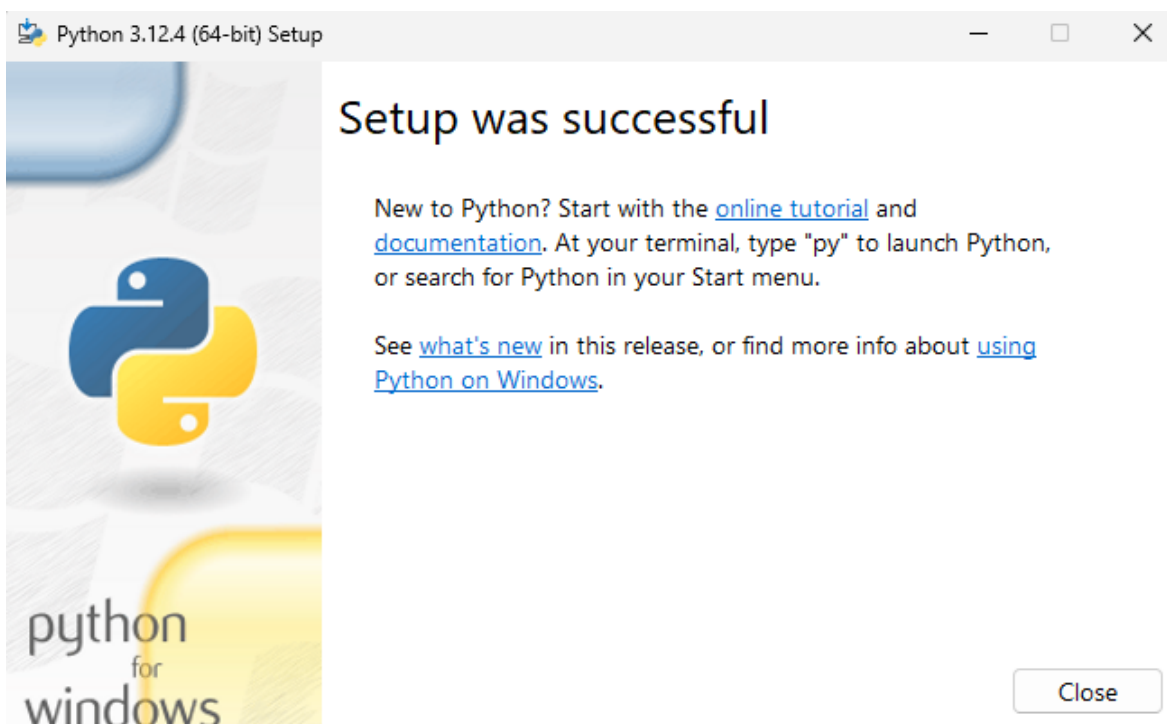
Σχήμα A.4: Step 2 of python installation

σχήμα A7. Από αυτό το μενού μπορούμε να δούμε την χρέωση των ερωτημάτων που πραγματοποιήσαμε με τη χρήση της εφαρμογής μας. Στη συνέχεια, αριστερά επιλέγουμε API KEYS, και create new Key όπως φαίνεται στο σχήμα A8. Αφού επιλέξουμε όνομα και πατήσουμε αποθήκευση Αφού εμφανιστεί το κλειδί το αντιγράφουμε και το κάνουμε επικόλληση στη γραμμή 12 του αρχείου thesis_V5.py

```
openai.api_key = 'ENTER-YOUR-API-KEY-HERE'
```



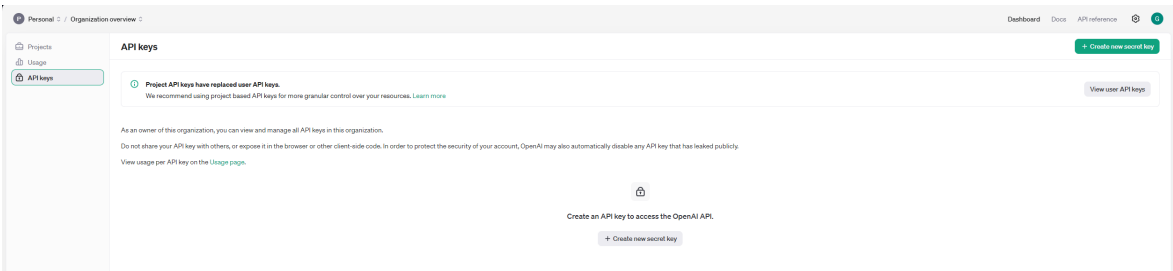
Σχήμα Α.5: Step 3 of python installation



Σχήμα Α.6: Step 4 of python installation



Σχήμα A.7: Increase Limit



Σχήμα A.8: Create API key

The screenshot shows a 'Create new secret key' dialog box. It has a title 'Create new secret key' and a section 'Owned by' with two radio buttons: 'You' (selected) and 'Service account'. Below this is a warning: 'This API key is tied to your user and can make requests against the selected project. If you are removed from the organization or project, this key will be disabled.' The 'Name' field is optional and contains 'My Test Key'. The 'Project' field is a dropdown menu labeled 'Select project'. The 'Permissions' section has three buttons: 'All' (selected), 'Restricted', and 'Read Only'. At the bottom, there are 'Cancel' and 'Create secret key' buttons.

Σχήμα A.9: API key options