

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

Speech to text Technologies for People with Deafness and hearing Loss

Ησαΐας Κωνσταντίνος

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

Κωνσταντίνος Κολομβάτσος

Λαμία Φεβρουάριος έτος 2025



SPEECH TO TEXT TECHNOLOGIES FOR PEOPLE WITH DEAFNESS AND HEARING LOSS

ΗΣΑΙΑΣ ΚΩΝΣΤΑΝΤΊΝΟΣ

ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΥΠΕΥΘΥΝΟΣ

Κωνσταντίνος Κολομβάτσος

Λαμία Φεβρουάριος έτος 2025



SCHOOL OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE & TELECOMMUNICATIONS

Speech to text Technologies for People with Deafness and hearing Loss

Isaias Konstantinos

FINAL THESIS

ADVISOR

Konstantinos Kolomvatsos

Lamia February year 2025

|  |
| --- |
| *«Με ατομική μου ευθύνη και γνωρίζοντας τις κυρώσεις (1), που προβλέπονται από της διατάξεις της παρ. 6 του άρθρου 22 του Ν. 1599/1986, δηλώνω ότι:* |
| *1.    Δεν παραθέτω κομμάτια βιβλίων ή άρθρων ή εργασιών άλλων αυτολεξεί****χωρίς να τα περικλείω σε εισαγωγικά****και χωρίς να αναφέρω το συγγραφέα, τη χρονολογία, τη σελίδα. Η αυτολεξεί παράθεση χωρίς εισαγωγικά χωρίς αναφορά στην πηγή, είναι λογοκλοπή. Πέραν της αυτολεξεί παράθεσης, λογοκλοπή θεωρείται και η παράφραση εδαφίων από έργα άλλων, συμπεριλαμβανομένων και έργων συμφοιτητών μου, καθώς και η παράθεση στοιχείων που άλλοι συνέλεξαν ή επεξεργάσθηκαν, χωρίς αναφορά στην πηγή. Αναφέρω πάντοτε με πληρότητα την πηγή κάτω από τον πίνακα ή σχέδιο, όπως στα παραθέματα.*  *2.    Δέχομαι ότι η αυτολεξεί****παράθεση χωρίς εισαγωγικά****, ακόμα κι αν συνοδεύεται από αναφορά στην πηγή σε κάποιο άλλο σημείο του κειμένου ή στο τέλος του, είναι αντιγραφή. Η αναφορά στην πηγή στο τέλος π.χ. μιας παραγράφου ή μιας σελίδας, δεν δικαιολογεί συρραφή εδαφίων έργου άλλου συγγραφέα, έστω και παραφρασμένων, και παρουσίασή τους ως δική μου εργασία.*  *3.    Δέχομαι ότι υπάρχει επίσης περιορισμός στο μέγεθος και στη συχνότητα των παραθεμάτων που μπορώ να εντάξω στην εργασία μου εντός εισαγωγικών. Κάθε μεγάλο παράθεμα (π.χ. σε πίνακα ή πλαίσιο, κλπ), προϋποθέτει ειδικές ρυθμίσεις, και όταν δημοσιεύεται προϋποθέτει την άδεια του συγγραφέα ή του εκδότη. Το ίδιο και οι πίνακες και τα σχέδια*  *4. Δέχομαι όλες τις συνέπειες σε περίπτωση λογοκλοπής ή αντιγραφής.* |

Ημερομηνία: ……/..…/20……

Ο – Η Δηλ.

*(1)*   *«Όποιος εν γνώσει του δηλώνει ψευδή γεγονότα ή αρνείται ή αποκρύπτει τα αληθινά με έγγραφη υπεύθυνη δήλωση*

*του άρθρου 8 παρ. 4 Ν. 1599/1986 τιμωρείται με φυλάκιση τουλάχιστον τριών μηνών. Εάν ο υπαίτιος αυτών των πράξεων*

*σκόπευε να προσπορίσει στον εαυτόν του ή σε άλλον περιουσιακό όφελος βλάπτοντας τρίτον ή σκόπευε να βλάψει άλλον, τιμωρείται με κάθειρξη μέχρι 10 ετών.»*

### ΠΕΡΙΛΗΨΗ

Η συγκεκριμένη πτυχιακή εργασία αφορά την ανάπτυξη ενός προγράμματος για την “μετάφραση” κειμένου από τον χρήστη είτε από την αναγνώριση της φωνής του είτε ( για μεγαλύτερη ευκολία και ακρίβεια) από εισαγωγή μέσο πληκτρολογίου. Ο συγκεκριμένος μεταφραστής θα έχει την δυνατότητα να αντιστοιχεί κάθε γράμμα του χρήστη με ένα συγκεκριμένο σύμβολο στην νοηματική γλώσσα. Ο μεταφραστής έχει την δυνατότητα αναγνώρισης φωνής στα Αγγλικά και Ελληνικά. Η συγκεκριμένη εφαρμογή εκτός από την αναγνώριση και αντιστοίχιση γραμμάτων με σύμβολα νοηματικής, έχει και την δυνατότητα(Μόνο στην αγγλική γλώσσα) να απλοποιεί ως ένα βαθμό και το κείμενο που δίνει ο χρήστης σαν εισαγωγή όταν μιλάει. Το συγκεκριμένο χαρακτηριστικό της εφαρμογής υλοποιείται μέσο μοντέλου μηχανικής μάθησης το οποίο έχει “προπονηθεί” σε ‘ένα μεγάλο σύνολο δεδομένων με απλές προτάσεις και το απλοποιημένο περιεχόμενό τους. Συνοψίζοντας όλα τα παραπάνω υλοποιούνται από την χρήση τεχνολογιών όπως αναγνώριση φωνής, chatbot(μόνο για εκφώνηση έτοιμων διαλόγων) , βιβλιοθήκη επεξεργασίας και προβολής εικόνων, βιβλιοθήκη για την καταγραφή ημερομηνιών ( χρησιμοποιείται στο csv για την σωστή και οργανωμένη καταγραφή δεδομένων), βιβλιοθήκη για την συμπερίληψη ANN(Advanced Neural Networks) όπως το t5-small και BART για NLP ( Natural Language Processing ) που ευθύνονται για την απλοποίηση κειμένου στην εφαρμογή μας.

### ABSTRACT

This thesis project involves the development of a program for “translating” text from the user either through voice recognition or (for greater ease and accuracy) through keyboard input. This translator will have the capability to map each letter entered by the user to a specific symbol in sign language. The translator can recognize speech in both English and Greek. In addition to recognizing and mapping letters to sign language symbols, the application also has the capability (only in English) to simplify, to some extent, the text provided by the user during speech input. This specific feature of the application is implemented through a machine learning model that has been “trained” on a large dataset of simple sentences and their simplified content. In summary, all of the above functionalities are achieved using technologies such as voice recognition, a chatbot (only for rendering pre-defined dialogues), a library for image processing and displaying it, a library for date recording (which is used in the CSV for proper and organized data logging), and a library for the inclusion of advanced neural networks (ANN) such as T5-small and BART for natural language processing (NLP) which are responsible for the text simplification.

Table of Contents

[ΠΕΡΙΛΗΨΗ i](#_Toc24624054)

[ABSTRACT iii](#_Toc24624055)

[ΚΕΦΑΛΑΙΟ 1 ΕΙΣΑΓΩΓΉ ΚΑΙ ΚΊΝΗΤΡΟ 1](#_Toc24624056)

**1.1** Λιγα λόγια για την νοηματική γλώσσα [1](#_Toc24624057)

**1.2** σκοπος και κινητρο της πτυχιακης εργασιας [1](#_Toc24624058)

**1.3** Εκμαθηση γραμμάτων στη νοηματική

**1.4** Εισαγωγή ατόμων στη Νοηματική γλωσσα(Με ή χωρις ακουστικά προβληματα)

[ΚΕΦΑΛΑΙΟ 2 Μοντελα μηχανικης μαθησης 2](#_Toc24624059)

**2.1** bart-base [2](#_Toc24624060)

**2.2** t5-small [2](#_Toc24624061)

**2.3** 3ο μοντέλο μηχανικής μαθησης

**2.4** Συγκριτική αξιολόγηση μεταξύ των μοντέλων

[ΚΕΦΑΛΑΙΟ 3 SPEECH RECOGNITION……… 3](#_Toc24624062)

3.1 speech recognition βιβλιοθηκη [3](#_Toc24624063)

3.2 Αναγνωριση ομιλιας με recognize\_google

3.3 λειτουργια recognizer  [3](#_Toc24624064)

[ΚΕΦΑΛΑΙΟ 4 προτεινομενο συστημα……… 4](#_Toc24624065)

4.1 γλωσσα προγραμματισμου και πακετα [4](#_Toc24624066)

4.2 αναλυση λειτουργιας του προγραμματος

4.3 διαρθωση κωδικα

4.4 εκπαιδευση μοντελων bart και t5-small [4](#_Toc24624067)

[ΚΕΦΑΛΑΙΟ 5 πειραματικη αποτιμηση 5](#_Toc24624068)

5.1 ελεγχος αποθηκευση στο csv

5.2 δοκιμες speech\_recognition σε δυο γλωσσες

5.3 Ελεγχος λειτουργιας switch/stop

5.4

5.5 Δοκιμή αντιστοιχησης εικόνων με γραμματα

5.6 Ελεγχος αναγνωρισης χαρακτηρων, αριθμων και συμβολων

5.7 Δοκιμη προβολης εικονων (ταχυτητα και ομαλοτητα)

5.8 Προτασεις απο χρηστη ASL για βελτιωση εφαρμογης

[κεφαλαιο 6 συμπερασματα και μελλοντικες προεκτασεις……………….](#_Toc24624069)

6.1 δυσκολιες στην υλοποιηση του προγραμματος

6.2 προβληματα ευρεσης λεξιλογίου νοηματικης γλωσσας

6.3 αναπτυξη της νοηματικης γλωσσας

6.4 Μελλοντικες προεκτασεις: χρηση νεων τεχνολογιων(animation bot)

6.5 προοπτικοη για προγραμμα αναγνωρισης νοηματων απο βιντεο

6.6 Δημιουργια μοντελου που συγκρινει βιντεο με κειμενα

[ΒΙΒΛΙΟΓΡΑΦΙΑ 6](#_Toc24624069)

# ΚΕΦΑΛΑΙΟ 1

# Εισαγωγή και κίνητρο

## **1.1 Λίγα λόγια για τη νοηματική γλώσσα**

Η νοηματική γλώσσα είναι μια προσπάθεια του ανθρώπου να εκφραστεί μέσο νοημάτων με τας χέρια του, με την στάση του σώματος του αλλά και τις εκφράσεις του προσώπου του για να μπορέσει να εκφράσει κάποια συναισθήματα ή καταστάσεις χωρίς να μιλήσει. Η νοηματική γλώσσα διαθέτει λεκτικές και συντακτικές δομές για να εκφράσει οποιοδήποτε αφηρημένη έννοια, και καθημερινές ανάγκες μας ( καλημέρα, χαίρω πολύ, πεινάω κ.α. ). Οι γλωσσολόγοι διακρίνουν τις φυσικές νοηματικές γλώσσες από άλλα συστήματα που προέρχονται από αυτές , όπως είναι οι τεχνητοί ανθρώπινοι κώδικες (αλγόριθμοι, γλώσσες προγραμματισμού κ.α. ), η ομιλούμενες γλώσσες, η οικιακή, η “βρεφική” και νοήματα που αποκτούνται από την βρεφική ηλικία.

Έτσι όπως υπάρχουν πολλές ομιλούμενες γλώσσες υπάρχουν και πολλές νοηματικές γλώσσες . Για παράδειγμα, η ASL ( American sign language ) είναι διαφορετική από την BSL ( British sign language) και τίς άλλες υπάρχουσες γλώσσες. [1]

Σύμφωνα με τον ethnologue υπάρχουν περισσότερες από 130 νοηματικές γλώσσες παγκοσμίως. Αυτές οι γλώσσες χρησιμοποιούνται από διάφορες κοινότητες κωφών και ημί-κωφων και ποικίλλουν από χώρα σε χώρα καθώς και σε διαφορετικές περιοχές εντός των ίδιων χωρών.[2]

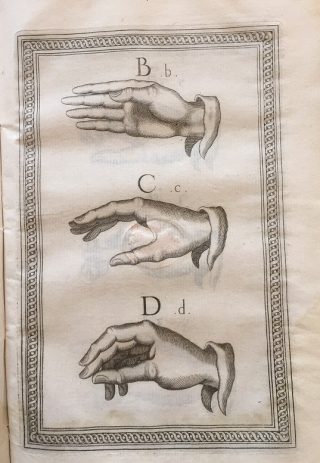
Σύμφωνα με τον Π.Ο.Υ (παγκόσμιος οργανισμός υγείας) περισσότερο από το 5% του παγκόσμιου πληθυσμού ή περίπου 430 εκατομμύρια άνθρωποι έχουν μόνιμη βλάβη στο ακουστικό τους σύστημα, είτε υπάρχει εκ γενετής ή μετά από ατύχημα. Μέχρι το 2050 εκτιμάται ότι πάνω από 700 εκατομμύρια άνθρωποι θα έχουν προβλήματα ακοής.[3]

Να σημειωθεί ότι η νοηματική γλώσσα δεν θα πρέπει να συσχετιστεί με την γλώσσα του σώματος , που θεωρείται ένα είδος μη λεκτικής επικοινωνίας.[4]

Πρώτες καταγεγραμμένες αναφορές για την νοηματική γλώσσα μπορούμε να βρούμε στο βιβλίο του Πλάτωνα “Κρατύλο” από συνομιλίες του με τον Σωκράτη.[5] Σε αυτό το σημείο να αναφέρω ότι τα άτομα με προβλήματα ακοής βρίσκονταν στο περιθώριο, διότι εκείνη την εποχή επικρατούσε η άποψη ότι ή γλώσσα μπορεί να μαθευτεί μόνο ακούγοντας τον προφορικό λόγο. Για παράδειγμα στην ρώμη σύμφωνα με το ρωμαϊκό δίκαιο επικρατούσε η άποψη ότι οι άνθρωποι που γεννιούνται κωφοί δεν είχαν ίσα δικαιώματα με τους υπόλοιπους πολίτες καθώς θεωρούταν ότι “δεν καταλάβαιναν τίποτα”.

Αυτή η προκατάληψη ενάντια στα κωφά άτομα άρχισε να εξαλείφεται την περίοδο της αναγέννησης.[6]

Μάλιστα το πρώτο άτομο που δημιούργησε μια “επίσημη” νοηματική γλώσσα είναι ο Ισπανός μοναχός Pedro Ponce de Leon τον 16ο αιώνα μΧ. Στη συνέχεια ένας άλλος Ισπανός κληρικός ο Juan Pablo Bonet κυκλοφόρησε το 1620 το πρώτο έργο για την εκμάθηση νοηματικής γλώσσας σε ανθρώπους με προβλήματα ακοής (Είναι από τα πρώτα έργα που έχουμε καταγεγραμμένα). Παρακάτω είναι μερικές από τις σελίδες του βιβλίου του Pedro.[7]



**Εικόνα 1** [**https://shorturl.at/qtZB5**](https://shorturl.at/qtZB5) **Εικόνα 2** [**https://shorturl.at/zxkDn**](https://shorturl.at/zxkDn)

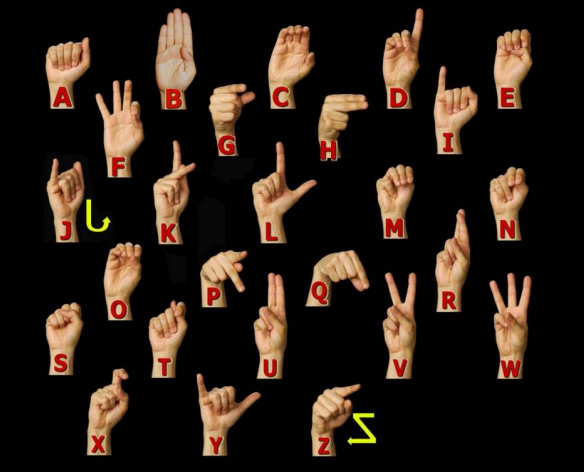
Η ASL( American sign language ) έχει ρίζες από την Αμερικανική σχολή για κωφούς που ιδρύθηκε στο Hartford, Connecticut το 1817[8]. Η συγκεκριμένη σχολή ιδρύθηκε από τον απόφοιτο του Yale Thomas Hopkins Gallaudet[9] και εκείνη την εποχή θεωρούταν σχολή για “χαζούς”.

Στην νοηματική γλώσσα ASL για να γίνει σωστά η εκτέλεση των λέξεων είναι απαραίτητες η παρακάτω 5 παράμετροι.[10]

1. Σχηματισμός χεριού
2. Προσανατολισμός παλάμης
3. κίνηση του χεριού
4. Τοποθεσία χεριού
5. Εκφράσεις σώματος (NMS)

**Σχηματισμός χεριού**

Σχηματισμοί χεριών αποτελούνται από το αλφάβητο που υπάρχει με της κινήσεις των χεριών σε διάφορες παραλλαγές σχημάτων όπως στην εικόνα παραπάνω.

****

**Εικόνα 3.** [**https://shorturl.at/iMmTu**](https://shorturl.at/iMmTu)

**Προσανατολισμός Παλάμης**

Ο σχηματισμός αναφέρεται σε ποια κατεύθυνση βλέπει η παλάμη για ένα συγκεκριμένο νοηματικό σήμα.

* Παλάμη προς τα έξω
* Παλάμη προς τα μέσα
* Παλάμη σε οριζόντια θέση
* Παλάμη βλέπεις προς τα Δεξιά/Αριστερά
* Παλάμη βλέπεις προς παλάμη
* Παλάμη βλέπει προς τα Πάνω/Κάτω

**Κίνηση του χεριού**

Ένα νόημα μπορεί να διακριθεί με κάποιες από τις παραπάνω αναριθμηζόμενες

κινήσεις.

* Κυκλική κίνηση
* Πάνω και κάτω
* Προς τα μπροστά
* Προς τα πίσω
* χτύπημα
* Μπροστά και πίσω
* Διάφορα κουνήματα

**Τοποθεσία χεριού**

Η τοποθεσία είναι ο χώρος στον οποίο λαμβάνει χώρα το νόημα που κάνει ο άνθρωπος σε σχέση με την τοποθεσία του σώματος του . Παραπάνω είναι όλες οι τοποθεσίες που μπορούν να γίνουν τα νοήματα, να σημειωθεί ότι ένα νόημα μπορεί να αρχίσει για παράδειγμα από το κεφάλι και να ολοκληρωθεί στο στήθος.

* Πιγούνι
* Ώμος
* Μπροστά από το σώμα
* Αριστερά και δεξιά του σώματος
* Μέτωπο

**Εκφράσεις σώματος**

Η παραπάνω εκφράσεις σώματος αναφέρονται ως συμπληρωματικές για να υπάρχει μεγαλύτερη κατανόηση σε νοήματα.

* Κούνημα κεφαλιού ( Καταφατικά/αρνητικά)
* Κούνημα φρυδιών
* Κούνημα μύτης
* Νοήματα με τα μάτια
* Χρήση των χειλιών

## **1.2 Σκοπός και κίνητρο της πτυχιακής εργασίας**

Ο σκοπός της συγκεκριμένης πτυχιακής είναι η εξοικείωση ανθρώπων χωρίς νοηματικά προβλήματα με το αλφάβητο της ASL και GSL. Είναι μια απλή προσπάθεια της τριβής ανθρώπων που δεν έχουν ξαναδεί νοήματα μέσο της ομιλίας τούς ή από την είσοδο στο πληκτρολόγιο στο πρόγραμμα που έχω υλοποιήσει. Περαιτέρω θα ήθελα μελλοντικά να ασχοληθώ με την αλληλεπίδραση προγραμμάτων χρήστη-υπολογιστή αλλά κυρίως με την κατανόηση της Νοηματικής και δημιουργία μιας πιο εξελιγμένης εφαρμογής.

Αυτό που με ώθησε και ήθελα να ασχοληθώ με το συγκεκριμένο θέμα είναι το συναισθηματικό κομμάτι, καθώς με μια απλή εφαρμογή μπορείς να κάνεις πιο εύκολη την ζωή κάποιου με κάποιες γραμμές κώδικα. Πέρα από την ευκολία της μετάφρασης των νοημάτων μπορείς να διευκολύνεις και κάποιον που θέλει να κάνει μια αρχική εισαγωγή στην νοηματική γλώσσα.

Η συγκεκριμένη εφαρμογή για εμένα δεν απευθύνεται σε άτομα που γνωρίζουν ήδη την Νοηματική γλώσσα αλλά σε άτομα που σε μια μεγαλύτερη ή και μικρότερη ηλικία θέλουν να βοηθηθούν στην εκμάθηση της . Κανείς δεν ξέρει πότε θα γνωρίσει κάποιον σημαντικό άνθρωπο που θα είναι κωφός και θα χρειαστεί να επικοινωνήσει μαζί του μέσο νοημάτων , οι περισσότεροι από εμάς θα θέλαμε να γνωρίζουμε την γλώσσα του για να μπορούμε να εμβαθύνουμε την σχέση μας με το συγκεκριμένο άτομο.

Στην τελική αυτός είναι ο σκοπός της εφαρμογής, μια μικρή εισαγωγή στην νοηματική γλώσσα και το να φέρουμε τους ανθρώπους πιο κοντά.

## **1.3 Εκμάθηση γραμμάτων στην νοηματική**

Η εκμάθηση του αλφαβήτου της Νοηματικής γλώσσας είναι σημαντική για την ανάπτυξη επικοινωνίας με βαρήκοα και κωφά άτομα. Η εκμάθηση είναι πολύ πιο αποδοτική με άτομα που την γνωρίζουν και μιλάνε, αλλά δεν έχουν όλοι την ευκαιρία να έχουν στον κύκλο τους ένα άτομο που να μιλάει και να χρησιμοποιεί την νοηματική γλώσσα. Για αυτό τον λόγο με το συγκεκριμένο πρόγραμμα ο μέσος χρήστης θα έχει την δυνατότητα να μιλάει με το πρόγραμμα ή να γράφει τι θέλει να πει και με τον καιρό να αρχίσει να μαθαίνει το κάθε νόημα σε ποιο γράμμα αντιστοιχεί.

Παρακάτω είναι το αλφάβητο της ASL.

Εικόνα που περιέχει πεταλούδα

Περιγραφή που δημιουργήθηκε αυτόματα

**Εικόνα 4** [**https://shorturl.at/pyVqo**](https://shorturl.at/pyVqo)

## **1.4 Εισαγωγή ατόμων στη Νοηματική γλώσσα( με ή χωρίς ακουστικά προβλήματα)**

Πέρα από την βοήθεια στην εκμάθηση της νοηματικής το πρόγραμμα βοηθάει και στην εισαγωγή ατόμων που δεν έχουν καμιά επαφή με την Νοηματική γλώσσα στα σύμβολα που χρησιμοποιούνται. Μπορεί να υπάρξουν περιπτώσεις οι οποίες δεν έχουν ξαναδεί κανένα σύμβολο.

Ακόμα η νοηματική προσφέρει μια πολυδιάστατη επικοινωνία στον άνθρωπο.

Ενισχύει την ικανότητα επικοινωνίας του είτε έχει κάποιο ακουστικό πρόβλημα ή

απλά την χρησιμοποιεί ως εναλλακτικό μέσο επικοινωνίας, επιπλέον συμβάλει και στην κοινωνική του συνοχή με άτομα της κωφής κοινότητας προάγοντας την συμπερίληψη και την κατανόηση διαφορετικών πολιτισμών.

Πέρα από όλα αυτά ανοίγονται και επαγγελματικής φύσεως ευκαιρίες διότι η εκμάθηση της νοηματικής γλώσσας μπορεί να προσφέρει εργασία σε όλους τους επαγγελματικούς χώρους όπου υπάρχει η κοινότητα των κωφών.[11]

Τέλος έχουν δείξει έρευνες ότι με την εκμάθηση της νοηματικής δίνεται η δυνατότητα ανάπτυξης γλωσσικών αλλά και γνωστικών δεξιοτήτων σε άτομα όλων των ηλικιών και πως αυτά τα άτομα έχουν βελτιωμένη νοητική ευελιξία, επίλυση καθημερινών προβλημάτων και αποκτούν δημιουργική σκέψη. [12]

# ΚΕΦΑΛΑΙΟ 2 Μοντέλα μηχανικής μάθησης

## 2.1 BART (Bidirectional and Auto-Regressive Transformers)

# ΚΕΦΑΛΑΙΟ 3 ………

## (Υποκεφάλαιο 3.1)

### (Ενότητα 3.1.α)

# ΚΕΦΑΛΑΙΟ 4 ………

## (Υποκεφάλαιο 4.1)

### (Ενότητα 4.1.α)

# ΚΕΦΑΛΑΙΟ 5 Συμπεράσματα

# ΒΙΒΛΙΟΓΡΑΦΙΑ

Εδώ θα προστεθεί όλη η βιβλιογραφία