



Πολυτεχνική Σχολή
Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Διπλωματική Εργασία

Υλοποίηση υποδομής chatbot για την αξιοποίηση βιοϊατρικής πληροφορίας

Ευδοκία Ξύγκη
Α.Μ. 1047172

Επιβλέπων
Δημήτριος Κουτσομητρόπουλος

Συνεπιβλέπων
Ανδρέας Ανδριόπουλος

Πάτρα, 2022

Θα ήθελα να ευχαριστήσω ιδιαίτερα τους Κουτσομητρόπουλο Δημήτριο, Ανδριόπουλο Ανδρέα, Σιούτα Σπύρο και Μακρή Χρήστο.

Τριμελής Επιτροπή

Ονοματεπώνυμο	Ιδιότητα	Τμήμα/Ιδρυμα	Γνωστικό Αντικείμενο
Δημήτρης Κουτσομητόπουλος	ΕΔΙΠ	Μηχανικών Η/Υ και Πληροφορικής	Διαχείριση Δεδομένων και Αναπαράσταση Γνώσης
Σιούτας Σπύρος	Καθηγητής	Μηχανικών Η/Υ και Πληροφορικής	Δομές Δεδομένων
Μακρής Χρήστος	Αναπληρωτής Καθηγητής	Μηχανικών Η/Υ και Πληροφορικής	Δόμηση Δεδομένων και Ανάκτηση Πληροφορίας

Περίληψη

Η συγκεκριμένη διπλωματική εργασία δημιουργήθηκε στα πλαίσια της έρευνας όσον αφορά την σχέση που θα μπορούσε να αναπτυχθεί ανάμεσα σε chatbots και βιοϊατρικά δεδομένα. Πιο συγκεκριμένα, η ιστορία των chatbots μας έχει αποδείξει πως η εξέλιξη τους είναι συνεχής και οι δυνατότητές τους αέναες. Ωστόσο, οι δυνατότητές τους σε βιοϊατρικά δεδομένα βρίσκονται ακόμη υπό εξερεύνηση και αυτός είναι και ο λόγος ύπαρξης της συγκεκριμένης διπλωματικής. Επιλέγοντας την δημιουργία ενός qa chatbot και την εισαγωγή βιοϊατρικών δεδομένων - και μη - σε αυτό, στοχεύουμε μέσω διαφόρων δοκιμών να αξιολογήσουμε πιθανές επιδόσεις και να εξάγουμε παρατηρήσεις σχετικά με την αντίδραση του chatbot σε κάθε περίπτωση.

Πιο συγκεκριμένα, όσον αφορά τα δεδομένα που χρησιμοποιήθηκαν, χρησιμοποιήσαμε 1100 Abstracts από ιατρικά papers της βιβλιοθήκης PubMed, καθώς και 250 SQuAD v2.0 δεδομένα, λόγω περιορισμένης υπολογιστικής ισχύος. Τα παραπάνω δείγματα, διαχωρίστηκαν σε προτάσεις, μετά την είσοδό τους στο πρόγραμμα, προσφέροντας 8737 προτάσεις για σύγκριση. Κατά τις δοκιμές χρησιμοποιήθηκε ο αλγόριθμος Bert σε τέσσερις διαφορετικές μορφές, το Roberta, Bert Large, Xlm-r και Biobert. Ωστόσο, θέλαμε να παρατηρήσουμε και πως θα αντιδρούσε το chatbot και χωρίς αλγόριθμο και έτσι αξιοποιήσαμε τον CountVectorizer, ο οποίος αποδείχτηκε και ο πιο αποτελεσματικός εκ των πέντε.

Όσον αφορά την αξιολόγηση, πραγματοποιήθηκαν 100 πειράματα σε κάθε περίπτωση, άρα συνολικά 500 πειραματικές ερωτήσεις και απαντήσεις, οι οποίες αξιολογήθηκαν με βάση τις απαντήσεις που έδινε το εκάστοτε chatbot και τις απαντήσεις που κρίναμε εμείς ως ορθές. Οι συγκεκριμένες απαντήσεις συγκρίθηκαν και αξιολογήθηκαν όσον αφορά τις ομοιότητές τους, μέσω των μετρικών f1 score, precision και recall, μετρικές οι οποίες υπολογίστηκαν συνολικά και έδωσαν συγκρίσιμα αποτελέσματα, φέρνοντας τον CounterVetconizer στην πρώτη θέση με τα Roberta, Bert Large, Xlm-r και Biobert να ακολουθούν.

Abstract

This thesis was created in the context of research on the relationship that could be developed between chatbots and biomedical data. More specifically, the history of chatbots has shown us that their evolution is continuous and their potential is endless. However, their potential in biomedical data is still being explored and that is the reason for the existence of this thesis. By choosing to create a qa chatbot and inserting biomedical data - and not - into it, we aim through various tests to evaluate possible performance and to make observations about the reaction of the chatbot in each case.

More specifically, in terms of the data used, we used 1100 abstracts from medical papers of the PubMed library, as well as 250 SQuAD v2.0 data, due to limited computing power. The above samples were divided into sentences, after entering the program, offering 8,737 sentences for comparison. The Bert algorithm was used in the tests in four different models, Roberta, Bert Large, Xlm-r and Biobert. However, we also wanted to observe how the chatbot would react without an algorithm, so we used CountVectorizer, which proved to be the most effective of the five.

In terms of evaluation, 100 experiments were performed in each case, so a total of 500 experimental questions and answers, which were evaluated based on the answers given by each chatbot and the answers that we judged to be correct. The specific answers were compared and evaluated in terms of their similarities, through the metrics f1 score, precision and recall, metrics which were calculated overall and gave comparable results, bringing CounterVetionizer in the first place with Roberta, Bert Large, Xlm-r and Biobert to follow.

Λίστα Εικόνων

Εικόνα 1. Αποτελέσματα αναζήτησης στο Scopus ανά έτος για "chatbot" ή "παράγοντα συνομιλίας" ή "διεπαφή συνομιλίας" ως λέξεις -κλειδιά από το 2000 έως το 2019.[2]

Εικόνα 2. Κατηγοριοποίηση chatbot συνδυαστικά με τα κριτήρια της έδρας ελέγχου (οριζόντιος άξονας) και του είδους της σχέσης (κάθετος άξονας) [40]

Εικόνα 3. Κατηγορίες και είδη Chatbot. [4]

Εικόνα 4. Προτίμηση Chatbots [4]

Εικόνα 5. Ταξινόμηση Chatbot, με βάση την φιλοσοφία σχεδίασης (οριζόντια) και τον τομέα γνώσης (κάθετα)[44]

Εικόνα 6. Σημείωση: Το "S" χρησιμοποιείται για κοινή χρήση και "N" για μη κοινόχρηστο. Τα στοιχεία που επισημαίνονται με «*» είναι υποχρεωτικά και «**» είναι στοιχεία που μοιράζονται χωρίς κανένα ΠΙ. Η σημείωση "με αστερί" σημαίνει ότι μπορούν να κοινοποιηθούν δημοφιλή μηνύματα κειμένου. Τα κενά κελιά δεν καθορίζονται στην πολιτική απορρήτου.

Εικόνα 7. Διαδικασία Υλοποίησης με χρήση NLP.

Εικόνα 8. Pre-training και Fine-tuning μέθοδοι

Εικόνα 9. Ανιχνευτής χωρίς παραμέτρους για συντακτική γνώση: λέξεις που μοιράζονται συντακτικά υπο δέντρα έχουν μεγαλύτερο αντίκτυπο το ένα στο άλλο στην πρόβλεψη MLM (Wu et al., 2020).

Εικόνα 10. BERT Παγκόσμια Γνώση

Εικόνα 11. Ονόματα όπως στην βιβλιοθήκη μετασχηματιστών: <https://github.com/huggingface/transformers> [29].

Εικόνα 12. Η αναλογία κατακερματισμού λέξεων των μοντέλων που βασίζονται στο BERT υπολογίζεται με βάση τα δεδομένα ανάπτυξης όλων των συνόλων δεδομένων. Θεωρούμε πως οι χαμηλότερες αναλογίες είναι καλύτερες με τα καλύτερα αποτελέσματα να εμφανίζονται με έντονους χαρακτήρες.

Εικόνα 13. Η συνολική διαδικασία προεκπαίδευσης και λεπτομέρειας του BioBERT.[60]

Εικόνα 14. Διάγραμμα υλοποίησης του Chatbot μας.

Εικόνα 15. Διάγραμμα ροής του κώδικα.

Εικόνα 16. Ένα παράδειγμα του cosine similarity, όπου το x_2 είναι παρόμοιο του x_1 , και διάφορο του x_3 . Το h_1 είναι η περιλαμβανόμενη γωνία μεταξύ των x_1 και x_2 , και το h_2 είναι η περιλαμβανόμενη γωνία μεταξύ των x_2 και x_3 . Η διακεκομμένη γραμμή είναι το όριο απόφασης.[75]

Εικόνα 17. Διάγραμμα ροής αξιολόγησης.

Εικόνα 18. Τύποι μετρικών αξιολόγησης.

Εικόνα 19. Περιεχόμενα μετρικών αξιολόγησης[95].

Εικόνα 20. Μέθοδος καταγραφής αξιολόγησης.

Εικόνα 21. Αναπαράσταση εκτέλεσης της αξιολόγησης του Chatbot μας.

Περιεχόμενα

1	1
Εισαγωγή	1
1.1 Η Ιστορία των Chatbots	1
1.2 Τυπολογία των Chatbots	6
1.3 Συνεισφορά της Διπλωματικής	15
2	17
Chatbot στην Βιοϊατρική	17
2.1 Η Ιστορία των Chatbots στην Βιοϊατρική	17
2.2 Σχετικό Έργο των Chatbots στην Βιοϊατρική	20
2.3 Ηθική και Ασφάλεια στον Τομέα των Chatbots	23
2.3.1 Ηθικές Επιπτώσεις των Chatbots προς τους Χρήστες	23
2.3.2 Ασφάλεια Προσωπικών Δεδομένων των Χρηστών	26
3	29
Αλγόριθμος Επεξεργασίας Φυσικής Γλώσσας BERT	29
3.1 Μοντέλα Επεξεργασίας Φυσικής Γλώσσας Κειμένου	29
3.2 Αλγόριθμος BERT	33
3.2.1 Συντακτική γνώση Μοντέλου	35
3.2.2 Σημασιολογική Γνώση	37
3.2.3 Παγκόσμια γνώση	37
3.2.4 BERT Embeddings	38
3.3 Αλγοριθμικά Μοντέλα BERT	39
3.3.1 Roberta Model	39
3.3.2 Bert-Large Model	40
3.3.3 Xlm-Large Model	41
3.3.4 BioBert Model	42
4	44
Υλοποίηση Υποδομής Chatbot με Χρήση Βιοϊατρικής Πληροφορίας	44
4.1 Δομή Συστήματος Chatbot	44
4.2 Μέθοδος Σύγκρισης Δεδομένων Εισόδου	47
4.3 Ενδεικτικός Κώδικας Υλοποίησης	49

5	56
Αξιολόγηση Υποδομής Chatbot με Χρήση Βιοϊατρικής Πληροφορίας	56
5.1 Μετρικές Σύγκρισης Συστημάτων	56
5.2 Πειραματική Αξιολόγηση Συστημάτων	60
5.3 Συγκρίσεις Αποτελεσμάτων Αξιολόγησης	64
6	68
Συμπεράσματα και Προεκτάσεις	68
6.1 Συμπεράσματα	68
6.2 Μελλοντική Έρευνα	69
Βιβλιογραφία - Αναφορές	70

1

Εισαγωγή

1.1 Η Ιστορία των Chatbots

Η τεχνητή νοημοσύνη AI (Artificial Intelligence) αποτελεί έναν πολύ σημαντικό κλάδο της επιστήμης των υπολογιστών, ο οποίος αφορά την προσομοίωση της ευφυούς συμπεριφοράς των υπολογιστών και ορίζεται ως η επιστήμη μέσω της οποίας εκτελούνται ανθρώπινες εργασίες από υπολογιστές. Πιο συγκεκριμένα, πρόκειται για υπολογιστές οι οποίοι επιλύουν πολύπλοκα προβλήματα, με τρόπο που προσομοιάζει την ανθρώπινη νοημοσύνη(Ralyte) [84].

Η Τεχνητή νοημοσύνη χρησιμοποιείται πλέον σε μεγάλη κλίμακα από οργανισμούς και επιχειρήσεις για την αυτοματοποίηση των επιχειρηματικών τους διαδικασιών και στόχων. Ο σχεδιασμός σχετικών εφαρμογών βρίσκεται μονίμως σε εξέλιξη, όσον αφορά το πεδίο της λειτουργίας του ανθρώπινου εγκεφάλου και της ευρύτερης κατανόησής του σε διαδικασίες όπως η βαθιά εκμάθηση, η λήψη αποφάσεων και οι ενέργειες κατά την επίλυση προβλημάτων, θέτοντας έτσι τα θεμέλια για την δημιουργία ευφύων συστημάτων, γνωστών και ως chatbots. Τα chatbots έχουν φέρει επανάσταση τόσο στην διαδικασία επικοινωνίας των επιχειρήσεων, όσο και στην κάλυψη αναγκών των χρηστών σε επίπεδο χρήστη.

Ο όρος chatbot αποτελείται από τους όρους chat και robot. Πιο συγκεκριμένα, αναφέρεται σε ένα υπολογιστικό πρόγραμμα που έχει την δυνατότητα προσομοίωσης της ανθρώπινης επικοινωνίας και γλώσσας με την χρήση ενός συστήματος διαλόγου με γραπτή ή και φωνητική επικοινωνία, ενώ παράλληλα προσφέρει ένα περιβάλλον συνομιλίας, δίνοντας στον χρήστη την αίσθηση επικοινωνίας με ένα αληθινό άτομο. Τέλος, ο όρος chatbot αναφέρεται σε κάθε πρόγραμμα λογισμικού, το οποίο εμπλέκεται σε έναν διάλογο με τον χρήστη και χρησιμοποιεί σαν μέσο την φυσική γλώσσα.

Ο Alan Turing το 1950 αναρωτήθηκε εάν ένα πρόγραμμα υπολογιστή θα μπορούσε να μιλήσει σε έναν άνθρωπο, χωρίς ο ίδιος να συνειδητοποιήσει ότι ο συνομιλητής τους είναι τεχνητός. Αυτή η ερώτηση ("Μπορούν οι μηχανές να

σκεφτούν;"), που ονομάζεται "Turing Test", θεωρείται από πολλούς ως η γενεσιουργός ιδέα των chatbots (Turing, 1950).[100]

Το πρώτο γνωστό chatbot ονομαζόταν Eliza και αναπτύχθηκε το 1966. Ο σκοπός του ήταν να λειτουργήσει ως ψυχοθεραπευτής, επιστρέφοντας τις προτάσεις του χρήστη με την ερωτηματική μορφή Weizenbaum (1966) χρησιμοποιώντας απλή αντιστοίχιση μοτίβου και ένα μηχανισμό απόκρισης βάσει προτύπων (Brandtzaeg & Følstad, 2017). Ένα μειονέκτημα του ELIZA ήταν ότι είχε περιορισμένες γνώσεις και ως εκ τούτου, η συζήτηση με αυτό περιορίζεται σε έναν συγκεκριμένο τομέα θεμάτων. Επίσης, δε μπορεί να κρατήσει μακρές συνομιλίες, να μάθει ή να ανακαλύψει το πλαίσιο της συζήτησης μέσα από αυτήν. Ωστόσο, αποτέλεσε πηγή έμπνευσης για την επακόλουθη ανάπτυξη των επόμενων chatbots (Klopfenstein, Delpriori, Malatini, & Bogliolo, 2017).[109]

Η επόμενη, βελτιωμένη έκδοση του ELIZA ήταν ένα chatbot με προσωπικότητα που ονομάστηκε PARRY, το οποίο αναπτύχθηκε το 1972 και λειτούργησε ως ασθενής με σχιζοφρένεια. Το PARRY θεωρείται πιο προηγμένο από το ELIZA, καθώς υποτίθεται ότι έχει "Προσωπικότητα" και καλύτερη δομή ελέγχου. Όριζε τις απαντήσεις του με βάση ένα σύστημα υποθέσεων και "συναισθηματικών απαντήσεων" που ενεργοποιείται από την αλλαγή των βαρών στις εκφράσεις του χρήστη. Επιπλέον, χρησιμοποιήθηκε σε ένα πείραμα το 1979, όταν πέντε ψυχιατρικοί δικαστές πήραν συνέντευξη από έναν ασθενή για να αποφασίσουν εάν πρόκειται για πρόγραμμα υπολογιστή ή για πραγματικό σχιζοφρενή ασθενή και έδωσαν δέκα διαγνώσεις. Ο πρώτος ψυχίατρος έδωσε δύο σωστές διαγνώσεις, ενώ ένας άλλος έδωσε δύο λάθος. Ο τρίτος θεώρησε ότι και τα δύο άτομα ήταν πραγματικοί ασθενείς και οι άλλοι δύο διέγινωσαν ότι και τα δύο άτομα ήταν chatbots. Ωστόσο, το δείγμα των πέντε ψυχιάτρων είναι πολύ μικρό και η έννοια των ευρημάτων δε μπορεί να θεωρεί σαφής ή επαρκής, καθώς τα άτομα με σχιζοφρένεια έχουν ένα βαθμό ασυνέπειας στην ομιλία τους. Σε γενικές γραμμές, το PARRY θεωρείται ένα chatbot με χαμηλές δυνατότητες όσον αφορά την κατανόηση της γλώσσας και την ικανότητα έκφρασης συναισθημάτων και έχει επίσης χαμηλή ταχύτητα απόκρισης, ενώ δε μπορεί να εκπαιδευτεί από την συνομιλία.[27]

Στη συνέχεια ήρθε η χρήση της τεχνητής νοημοσύνης, η οποία χρησιμοποιήθηκε αρχικά στον τομέα των chatbots με την κατασκευή του Jabberwacky (1988). Το Jabberwacky γράφτηκε με CleverScript, μια γλώσσα βασισμένη σε υπολογιστικά φύλλα που διευκόλυνε την ανάπτυξη των chatbots και χρησιμοποίησε αντιστοίχιση προτύπων με βάση τα συμφραζόμενα, ώστε να σχηματίζει τις απαντήσεις του με βάση τις προηγούμενες συζητήσεις. Ωστόσο, το Jabberwacky δε μπορούσε να δώσει απαντήσεις με υψηλή ταχύτητα απόκρισης ή να συνεργαστεί με μεγάλο αριθμό χρηστών.[51]

Αργότερα, αναφέρθηκε για πρώτη φορά ο όρος Chatterbot, το 1991. Πιο συγκεκριμένα, εκπροσωπούσε έναν τεχνητό παίκτη - TINYMUD (εικονικός κόσμος σε πραγματικό χρόνο για πολλούς παίκτες) - του οποίου η κύρια λειτουργία ήταν η συνομιλία. Αρκετοί πραγματικοί παίκτες έδειχναν να προτιμούν την συνομιλία με το Chatterbot, παρά με έναν πραγματικό παίκτη. Το Chatterbot είχε επιτυχία στον κόσμο

του TINYMUD, καθώς οι παίκτες είχαν την λανθασμένη εντύπωση πως όλοι οι παίκτες ήταν άνθρωποι και υπήρχαν αμφιβολίες μόνο εάν κάποιος έκανε κάποιο σημαντικό λάθος (Mauldin, 1994). [65] Στη συνέχεια, εμφανίστηκε το Dr. Sbitso (Sound Blaster Artificial Intelligent Text to Speech Operator), ένα chatbot που δημιουργήθηκε το 1992, σχεδιάστηκε για να εμφανίζει τις ψηφιοποιημένες φωνές που μπορούσαν να παράγουν οι κάρτες ήχου. Προσπάθησε να πάρει το ρόλο του ψυχολόγου χωρίς κάποια περίπλοκη αλληλεπίδραση (Zemčik, 2019).[36]

Στην πορεία της ιστορίας των chatbots, ήρθε το 1995 το ALICE (Artificial Linguistic Internet Computer Entity). Υπήρξε το πρώτο διαδικτυακό chatbot εμπνευσμένο από τον ELIZA (Wallace, 2009) και βασίστηκε σε έναν απλό αλγόριθμο αντιστοίχισης προτύπων, χρησιμοποιώντας υποκείμενη νοημοσύνη, βασισμένη στην γλώσσα σήμανσης της τεχνητής νοημοσύνης (AIML), η οποία επιτρέπει στους προγραμματιστές να ορίσουν τα δομικά στοιχεία της γνώσης του. Δεν είχε την δυνατότητα για πραγματική αντίληψη της συνομιλίας (Marietto et al., 2013), αλλά μπορούσε να συζητήσει για μία αρκετά μεγάλη εμβέλεια θεμάτων. Ωστόσο, χρειάστηκε να περάσει κάποιο χρονικό διάστημα πριν την βελτίωσή του και την κατάκτηση του βραβείου Loebner τα έτη 2000, 2001 και 2004, όντας έτσι ο πρώτος υπολογιστής που κέρδισε τον βαθμό του “πιο ανθρώπινου υπολογιστή”. Η γλώσσα στην οποία αναπτύχθηκε δημιουργήθηκε αποκλειστικά για αυτόν τον σκοπό - γλώσσα σήμανσης τεχνητής νοημοσύνης (AIML) -, η οποία αποτελεί την πιο κρίσιμη διαφορά μεταξύ του ALICE και του ELIZA. Όσον αφορά την γνωσιακή βάση του ALICE, αποτελείται από περίπου 41.000 πρότυπα και μοτίβα, γεγονός που αποτελεί μεγάλη διαφορά ανάμεσα σε αυτό και το ELIZA, το οποίο είχε μόνο 200 λέξεις-κλειδιά και κανόνες (Heller, Procter, Mah, Jewell, & Cheung, 2005). Ωστόσο, το ALICE δεν είχε έξυπνα χαρακτηριστικά και δε μπορούσε να παράγει απαντήσεις φαινομενικά “ανθρώπινες” που να εκφράζουν συναισθήματα ή στάσεις. [106]

Το 2001, υπήρξε μια πραγματικά μεγάλη εξέλιξη στον τομέα της τεχνολογίας των chatbot με την έλευση του SmarterChild, το οποίο ήταν διαθέσιμο σε μορφή Messenger, όπως το America Online (AOL) και το Microsoft(MSN). Ήταν η πρώτη φορά που ένα chatbot θα μπορούσε να βοηθήσει άτομα σε πρακτικές καθημερινές εργασίες, καθώς αξιοποιούσε πληροφορίες από βάσεις δεδομένων σχετικά με τις ώρες των ταινιών, τα αθλητικά σκορ, τις τιμές των μετοχών, τις ειδήσεις και τον καιρό. Αυτή η νέα, πρωτοπόρα ικανότητα σηματοδότησε μια σημαντική εξέλιξη στον τομέα της τεχνολογίας, τόσο στο επίπεδο ευφυΐας της μηχανής, όσο και στην αλληλεπίδραση ανθρώπου-υπολογιστή, καθώς τα πληροφοριακά συστήματα είχαν την ικανότητα πρόσβασης μέσω συζήτησης. [68]

Η ανάπτυξη των chatbots τεχνητής νοημοσύνης έκανε το επόμενο βήμα με την δημιουργία έξυπνων προσωπικών βοηθών που αξιοποίησαν την φωνή του χρήστη, ενσωματωμένα σε smartphone ή συσκευές ήχου στο σπίτι, τα οποία είχαν την δυνατότητα αντίληψης των φωνητικών εντολών. Πιο συγκεκριμένα, συνομιλούσαν με ψηφιακές φωνές και χειρίζονταν εργασίες όπως η παρακολούθηση οικιακών αυτοματοποιημένων συσκευών, ημερολογίων, email και άλλα. Οι Apple Siri (Siri),

IBM Watson (Watson Assistant | IBM Cloud, 2020), Google Assistant (Google Assistant, το δικό σας προσωπικό Google, 2019), Microsoft Cortana (Personal Digital Assistant - Cortana Home Assistant - Microsoft, 2019) και Amazon the Alexa αποτελούν τους πιο δημοφιλείς βοηθούς φωνής. Προφανώς, υπήρχαν κι άλλοι, λιγότερο διάσημοι βοηθοί φωνής με άλλα χαρακτηριστικά, αλλά με τις ίδιες βασικές λειτουργίες. Τέλος, λειτουργούν με σύνδεση στο διαδίκτυο και, σε αντίθεση με τους προκατόχους τους, έχουν την ικανότητα γρήγορων και ουσιαστικών απαντήσεων (Hoy, 2018).[2]

Η Siri (Siri), η οποία αναπτύχθηκε από την Apple(2010), πρωτοστάτησε στο δρόμο για την δημιουργία προσωπικών βοηθών. Οι χρήστες είχαν πλέον την δυνατότητα να πραγματοποιούν ερωτήσεις και συνομιλίες μέσω του Messenger και φωνητικών εντολών, ενώ συμπεριλαμβάνει ενσωμάτωση με αρχεία ήχου, βίντεο και εικόνες. Η Siri κάνει συστάσεις για την “προσωπικότητά” της και απαντά σε αιτήματα χρηστών χρησιμοποιώντας διάφορες υπηρεσίες του διαδικτύου, ενώ προσαρμόζεται, μέσω της συνεχούς χρήσης και συνεπώς εκπαίδευσής της, στις χρήσεις, αναζητήσεις και επιθυμίες της γλώσσας των χρηστών (Siri, 2020). Ωστόσο, αν και η Siri ανταποκρίνεται σε αρκετές επιθυμίες του χρήστη λόγω της πολυπλοκότητας της, δεν παύει να έχει αδυναμίες. Λειτουργεί με σύνδεση στο διαδίκτυο και αν και πολύγλωσση, υπάρχουν πολλές γλώσσες που δεν υποστηρίζει, ενώ παράλληλα, οι οδηγίες πλοήγησης υποστηρίζονται μόνο στα αγγλικά. Επίσης, παρουσιάζει δυσκολία παρακολούθησης του συνομιλητή σε περίπτωση έντονης προφοράς ή παρουσίας θορύβου (Soffar, 2019).[46]

Στην αλυσίδα της εξέλιξης των chatbot, ήρθε το 2011 να προστεθεί το Watson (Watson Assistant | IBM Cloud, 2020). Το Watson δημιουργήθηκε από την IBM και είχε την δυνατότητα κατανόησης της φυσικής ανθρώπινης γλώσσας αρκετά καλά, κερδίζοντας έτσι τους δύο προηγούμενους πρωταθλητές στο διαγωνισμό κουίζ “Jeopardy”. Η μορφή του συγκεκριμένου διαγωνισμού ήταν πληροφορίες με τη μορφή απαντήσεων ως είσοδο και αντίστοιχες ερωτήσεις σαν έξοδο από το chatbot. την πορεία του ήρθαν να συμπληρώσουν χρόνια αργότερα εικονικοί βοηθοί επιχειρήσεων, καθώς και βοηθοί διαγνώσεων για γιατρούς στην υγειονομική περίθαλψη. Ωστόσο, ένα μειονέκτημα του Watson είναι ότι υποστηρίζει μόνο αγγλικά.[108]

Το 2012 εμφανίστηκε το Google Now (Google now, 2020), με αρχική χρήση την παροχή πληροφοριών στον χρήστη λαμβάνοντας υπόψη πληροφορίες όπως η ώρα της ημέρας, η τοποθεσία και οι προτιμήσεις του χρήστη. Η εξέλιξή του ήρθε το 2016, όταν δημιουργήθηκε το Google Assistant (Google Assistant, το δικό σας προσωπικός οδηγός Google, 2019), με βαθύτερη τεχνητή νοημοσύνη, περισσότερο φιλική διεπαφή συνομιλίας και παροχή πληροφοριών προς τους χρήστες, οι οποίες προβλέπουν τις απαιτήσεις τους. Ωστόσο, έχει έλλειψη προσωπικότητας και οι ερωτήσεις του ενδέχεται να παραβιάσουν το απόρρητο του χρήστη, καθώς χρησιμοποιεί άμεση σύνδεση με τον λογαριασμό των χρηστών στην Google.

Το 2014 ήρθε να προστεθεί στην λίστα ο προσωπικός βοηθός της Microsoft (Personal Digital Assistant — Cortana Home Assistant — Microsoft, 2019), ο οποίος αναγνωρίζει φωνητικές εντολές και εκτελεί εργασίες όπως ταυτοποίηση χρόνου και

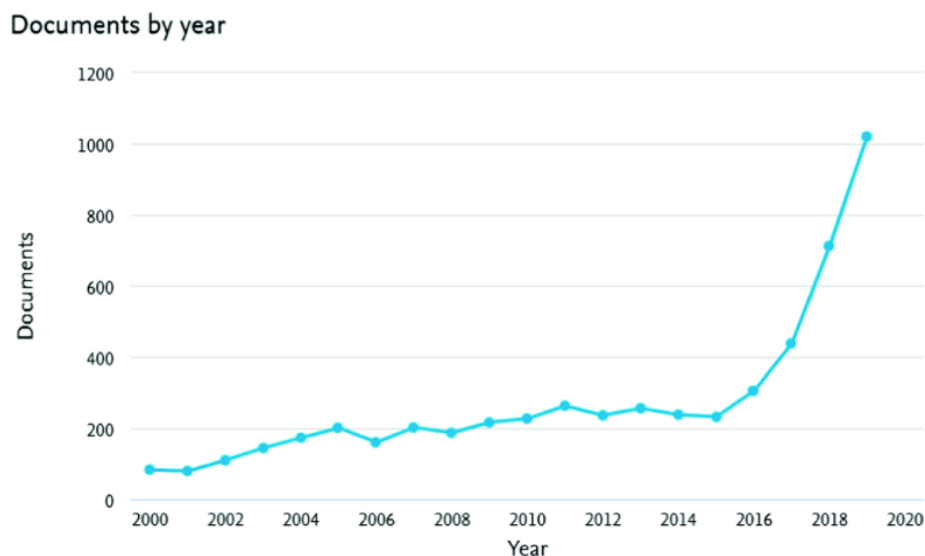
θέσης, υποστήριξη υπενθυμίσεων που βασίζονται σε άτομα, αποστολή μηνυμάτων ηλεκτρονικού ταχυδρομείου και κειμένων, δημιουργία και διαχείριση λιστών, αναπαραγωγή παιχνιδιών και εύρεση πληροφοριών. Παρουσιάζει ωστόσο ένα σημαντικό μειονέκτημα (η Cortana), το οποίο είναι πως υπάρχει πιθανότητα να τρέξει κάποιο πρόγραμμα το οποίο θα μπορούσε να εγκαταστήσει κακόβουλο λογισμικό.

Την ίδια χρονιά, η Amazon παρουσίασε την Alexa. Πρόκειται για ένα chatbot ενσωματωμένο σε συσκευές αυτοματισμού και ψυχαγωγίας στο σπίτι, δημιουργώντας με αυτόν τον τρόπο ένα IoT πιο προσιτό προς τους χρήστες. Μια επιπλέον καινοτομία που πρόσφερε, είναι το γεγονός ότι οι προγραμματιστές μπορούν να χρησιμοποιήσουν το Alexa Skills Kit (ASK) για να δημιουργήσουν και να δημοσιεύσουν - δωρεάν ή επί πληρωμή - δεξιότητες της Alexa. Ωστόσο, παρουσιάζει και αυτό ζητήματα ασφαλείας, λόγω των προσωπικών δεδομένων που χρησιμοποιεί.[37]

Παρόλο που οι προσωπικοί φωνητικοί βοηθοί επιτρέπουν την φωνητική επικοινωνία με τους χρήστες τους, συχνά συμβαίνουν παρεξηγήσεις, καθώς δε μπορούν να κατανοήσουν απόλυτα την ανθρώπινη γλώσσα στον προφορικό λόγο ή δεν κατανοούν ολόκληρο το πλαίσιο της συνομιλίας. Το 2016 ήρθε μία νέα εξέλιξη στον τομέα της Τεχνητής Νοημοσύνης, η οποία θα άλλαζε δραματικά τον τρόπο επικοινωνίας των ανθρώπων με τους κατασκευαστές. Οι πλατφόρμες κοινωνικών μέσων επέτρεψαν στους προγραμματιστές να δημιουργήσουν chatbots για το εμπόριο, ώστε οι πελάτες να μπορούν να εκτελούν ενέργειες μέσω των μέσων κοινωνικής δικτύωσης. Στα τέλη του ίδιου έτους, 34.000 chatbots κάλυψαν ένα ευρύ φάσμα χρήσεων (Wizu, 2018) σε τομείς όπως το μάρκετινγκ, τα υποστηρικτικά συστήματα, την υγειονομική περίθαλψη, την ψυχαγωγία, την εκπαίδευση και την πολιτιστική κληρονομιά. [114]

Επιπλέον, το Internet of Things (IoT) εισήγαγε μια νέα εποχή, όπου η χρήση των chatbots βελτίωσε την επικοινωνία μεταξύ της τεχνολογίας και των χρηστών (Kar & Halder, 2016). Άξιο αναφοράς είναι επίσης το Microsoft XiaoIce, το οποίο είναι ένα AI chatbot που ικανοποιεί την ανθρώπινη ανάγκη για κοινωνικότητα. Διαθέτει δική του προσωπικότητά, συμβάλει στην ανάπτυξη των chatbots και διαθέτει έξυπνο και συναισθηματικό πηλίκιο (IQ-EQ). Δημιουργεί επίσης μακρές συναισθηματικές σχέσεις με τους χρήστες του, λαμβάνοντας υπόψη τις πολιτισμικές ιδιαιτερότητες και τα ηθικά ζητήματα του χρήστη ανάλογα με τις προτιμήσεις του. [37]

Τέλος, ο τρόπος με τον οποίο τα chatbots σήμερα συμμετέχουν σε μια συζήτηση είναι εντελώς διαφορετικός από τον προκάτοχό τους Eliza, ωστόσο μπορούν να μοιραστούν προσωπικές σκέψεις και περίπλοκες συναισθηματικές καταστάσεις, να είναι συναφή αλλά και να μπερδεύουν και να εξαπατούν όπως κάνουν οι άνθρωποι. Όπως φαίνεται στην παρακάτω εικόνα, σύμφωνα με το Scopus (Η μεγαλύτερη βάση βιβλιογραφίας με κριτές επιστημονικά περιοδικά, βιβλία και πρακτικά συνεδρίων), υπήρξε μια ταχεία αύξηση του ενδιαφέροντος για τα chatbots, ειδικά μετά το έτος 2016. Πολλά chatbots αναπτύχθηκαν για βιομηχανικές λύσεις, ενώ υπάρχει ένα ευρύ φάσμα λιγότερο διάσημων chatbots που σχετίζονται με την έρευνα και τις εφαρμογές [2].



Εικόνα 1. Αποτελέσματα αναζήτησης στο Scopus ανά έτος για "chatbot" ή "παράγοντα συνομιλίας" ή "διεπαφή συνομιλίας" ως λέξεις -κλειδιά από το 2000 έως το 2019.[2]

1.2 Τυπολογία των Chatbots

Τα chatbots μπορούν να ταξινομηθούν σε κατηγορίες χρησιμοποιώντας πολλές και διαφορετικές παραμέτρους, όπως το ιστορικό ανάπτυξης τους, ο αριθμός των συνομιλητών που επιτρέπουν, αλλά και πιο σύνθετα, όπως η φιλοσοφία σχεδίασης και ο τομέας γνώσης.

Ο διαχωρισμός των chatbots με βάση το ιστορικό ανάπτυξης τους και με βάση το εύρος και το βάθος εστίασης της συνομιλίας με τους ανθρώπους περιλαμβάνει:

1. **Εικονικούς συντρόφους (Virtual Companions)** - Στην συγκεκριμένη κατηγορία τα chatbots καλύπτουν τις επικοινωνιακές και ψυχαγωγικές ανάγκες των χρηστών, όπως π.χ. το ELIZA, με τις συζητήσεις τους να είναι εκτενείς και με μεγάλη γκάμα θεμάτων.
2. **Ευφυείς βοηθούς (Intelligent Assistants)** - Chatbots με δυνατότητα ενασχόλησης, αλλά και σύντομες συνομιλίες, με σκοπό την άμεση επίλυση του θέματος, όπως π.χ τα Cortana, Siri, Alexa και Google Assistant.
3. **Chatbots επιτέλεσης συγκεκριμένου σκοπού (Task-focused Chatbots)** - Στοχεύουν στην εξυπηρέτηση πελατών, με τα chatbot αυτής της κατηγορίας να διαθέτουν την δυνατότητα συζήτησης μιας μεγάλης ποικιλίας θεμάτων, αλλά σε

συγκεκριμένη θεματολογία κάθε φορά [44].

Επιπροσθέτως, μία ακόμη κατηγοριοποίηση αφορά τα chatbot συγκεκριμένου σκοπού, η οποία τα κατηγοριοποιεί σε δύο επιμέρους κατηγορίες με βάση την αρχιτεκτονική σχεδίασης τους: αυτά που φέρουν αρχιτεκτονική αγωγού (pipeline) και αυτά με αρχιτεκτονική άκρου-εις-άκρον (end-to-end) [25].

Όσον αφορά την πρώτη κατηγορία, ορισμένα βασικά στοιχεία της είναι:

1. Σύστημα κατανόησης γλώσσας - στο οποίο τα εκφωνήματα αντιστοιχίζονται με τις σημασιολογικές τους αναπαραστάσεις (slots), οι οποίες είναι προκαθορισμένες ανά περίπτωση.
2. Σύστημα παρακολούθησης κατάστασης διαλόγου - προβλέπει τον σκοπό του συνομιλητή και αποτελεί τον βασικότερο παράγοντα διατήρησης της εγκυρότητας της δομής διαλόγου στα chatbots (σημασιολογικό πλαίσιο).
3. Εκμάθηση κανόνων λειτουργίας - καθορίζει την επόμενη ενέργεια του συστήματος, την πληροφορία δηλαδή που θα εξάγει βάσει της εισαγωγής του χρήστη.
4. Σύστημα παραγωγής φυσικής γλώσσας - μετατρέπει τις αφηρημένες φράσεις του συστήματος σε φυσική γλώσσα, ώστε η απάντηση του chatbot να είναι κατανοητή από τον χρήστη.

Όσον αφορά την δεύτερη κατηγορία, αυτή της end-to-end αρχιτεκτονικής, ισχύουν τα παρακάτω:

1. Αναπτύχθηκαν με σκοπό την προσαρμογή σε νέους τομείς γνώσης, διαθέτοντας ωστόσο μία σημαντική αλληλεξάρτηση μεταξύ των διαδικασιών τους.
2. Οι έξοδοι του συστήματος εξαρτώνται σημαντικά από τις εισόδους με την προσαρμογή του συστήματος στις διαφορετικές εισόδους να απαιτεί την επανεκπαίδευση με νέα δεδομένα.
3. Στηρίζεται σε μια μονάδα, η οποία λαμβάνει τα δεδομένα μέσω αλληλεπίδρασης με δομημένες εξωτερικές βάσεις δεδομένων, μέσω ενός συστήματος κωδικοποιητή-αποκωδικοποιητή.

Σε ότι αφορά τα chatbot μη συγκεκριμένου σκοπού (ή αλλιώς γενικού), η σχεδίαση τους για την δημιουργία της απόκρισης προς τον χρήστη βασίζεται σε

διάφορα πλαίσια όπως [25] :

1. **Μοντέλα παραγωγής απαντήσεων προς τον χρήστη (Generative models)** - λειτουργούν με νευρωνικά δίκτυα (Neural Generative Models) και βασίζονται σε αρχιτεκτονική τύπου κωδικοποιητή-αποκωδικοποιητή και σε εκμάθηση sequence to sequence (seq2seq).
2. **Μοντέλα βάσει κανόνων (Rule-based)** - οι τεχνικές που χρησιμοποιούνται για την συγκεκριμένη κατηγορία είναι οι εξής:
 - a. Τεχνική Single-turn - η απάντηση βασίζεται αποκλειστικά στο τελευταίο μήνυμα του χρήστη για την εύρεση της κατάλληλης απάντησης.
 - b. Τεχνική multi-turn - οι απαντήσεις δίνονται λαμβάνοντας υπόψη το σύνολο του διαλόγου, επιλέγοντας μια απάντηση που να ταιριάζει στο περιεχόμενο της συνομιλίας με το χρήστη που έχει αναπτυχθεί έως τότε.
3. **Υβριδικά μοντέλα** - μοντέλα που συνδυάζουν και τις δύο παραπάνω τεχνικές λειτουργίας προσφέροντας έτσι τη μέγιστη δυνατή αποτελεσματικότητα στην παραγωγή απαντήσεων.

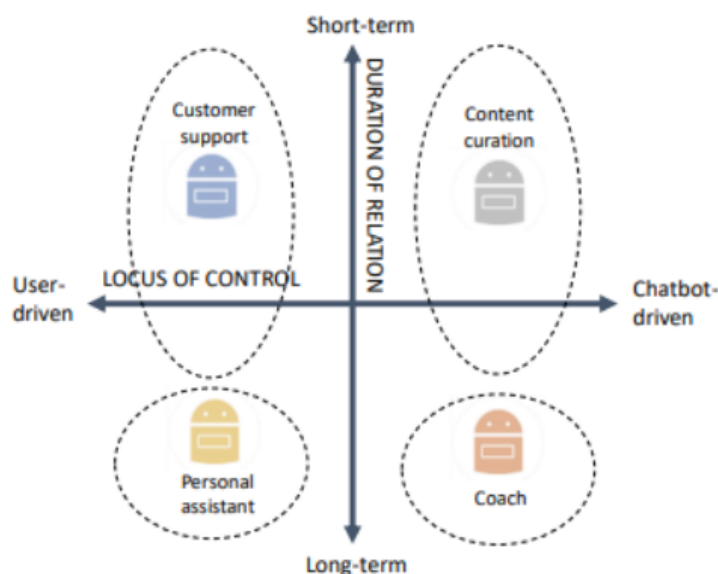
Μια επιπλέον τυπολογία για τα chatbot με βάση δύο νέα κριτήρια, είναι η έδρα ελέγχου (locus of control) και η διάρκεια της σχέσης. Πιο συγκεκριμένα, το πρώτο κριτήριο, αναφέρεται στον βαθμό ανθρωποκεντρισμού του συστήματος. Με βάση αυτό το κριτήριο, εντοπίζονται οι παρακάτω διακρίσεις:

1. **Ο διάλογος κατευθύνεται από το σύστημα (chatbot-driven)** - το chatbot κατευθύνει την συζήτηση σε συγκεκριμένες ερωτήσεις - απαντήσεις, δίνοντας ελάχιστα περιθώρια επιλογών στο χρήστη σχετικά με την κατεύθυνση της συζήτησης.
2. **Ο διάλογος κατευθύνεται από τον χρήστη (user-driven)** - η συνομιλία καθοδηγείται κυρίως από τον χρήστη, βοηθώντας έτσι το chatbot να εκπαιδευτεί ώστε να διαχειρίζεται ένα μεγαλύτερο φάσμα ερωτήσεων και απαντήσεων.

Όσον αφορά το κριτήριο της διάρκειας της σχέσης, αφορά το είδος της σχέσης μεταξύ συστήματος – χρήστη σχετικά με την διατήρηση πληροφοριών του χρήστη για τη μελλοντική διαχείρισή του και αλληλεπίδραση. Πιο συγκεκριμένα:

1. **Βραχυπρόθεσμων σχέσεων** - όπου η αλληλεπίδραση μεταξύ χρήστη-συστήματος γίνεται μια φορά, χωρίς την διατήρηση πληροφοριών εκ μέρους του συστήματος.

2. **Μακροπρόθεσμων σχέσεων** - όπου τα δεδομένα της αλληλεπίδρασης συλλέγονται για την δημιουργία ενός προφίλ του χρήστη, για την καλύτερη δυνατή αντιμετώπιση του χρήστη στο μέλλον.
3. **Υβριδικών σχέσεων** - Συνδυάζοντας τα δύο παραπάνω κριτήρια, δημιουργώντας έτσι ένα δισδιάστατο πλαίσιο, από το οποίο προκύπτουν τέσσερις επιμέρους κατηγορίες chatbot. Μπορούμε να αντιληφθούμε καλύτερα την συγκεκριμένη κατηγορία παρατηρώντας την παρακάτω εικόνα [40].



Εικόνα 2. Κατηγοριοποίηση chatbot συνδυαστικά με τα κριτήρια της έδρας ελέγχου (οριζόντιος άξονας) και του είδους της σχέσης (κάθετος άξονας)

Οι επιμέρους κατηγορίες που δημιουργούνται από τις υβριδικές σχέσεις είναι:

1. **Chatbots υποστήριξης πελατών (Customer Support)** - στόχος των οποίων είναι η παροχή υποστήριξης - συμβουλευτικής και μη - προς τον χρήστη εκ μέρους μιας επιχείρησης, με την συζήτηση να καθοδηγείται από τον χρήστη με βάση τα ερωτήματα που θέτει.
2. **Chatbots – προσωπικοί βοηθοί (Personal Assistants)** - σχεδιασμένα για την εξυπηρέτηση καθημερινών αναγκών του χρήστη, όπως π.χ. η αναζήτηση πληροφοριών ψυχαγωγίας κ.α., με την συζήτηση να καθοδηγείται από τον χρήστη.
3. **Chatbots επιμέλειας περιεχομένου (Content Curation)** - παρέχουν στον χρήστη την δυνατότητα πρόσβασης σε πληροφορίες, όπως ειδήσεις και πολυμέσα ψυχαγωγίας, με τον διάλογο να καθοδηγείται από το σύστημα με τον

χρήστη να περιορίζεται στην αποδοχή ή μη του προτεινόμενου περιεχομένου.

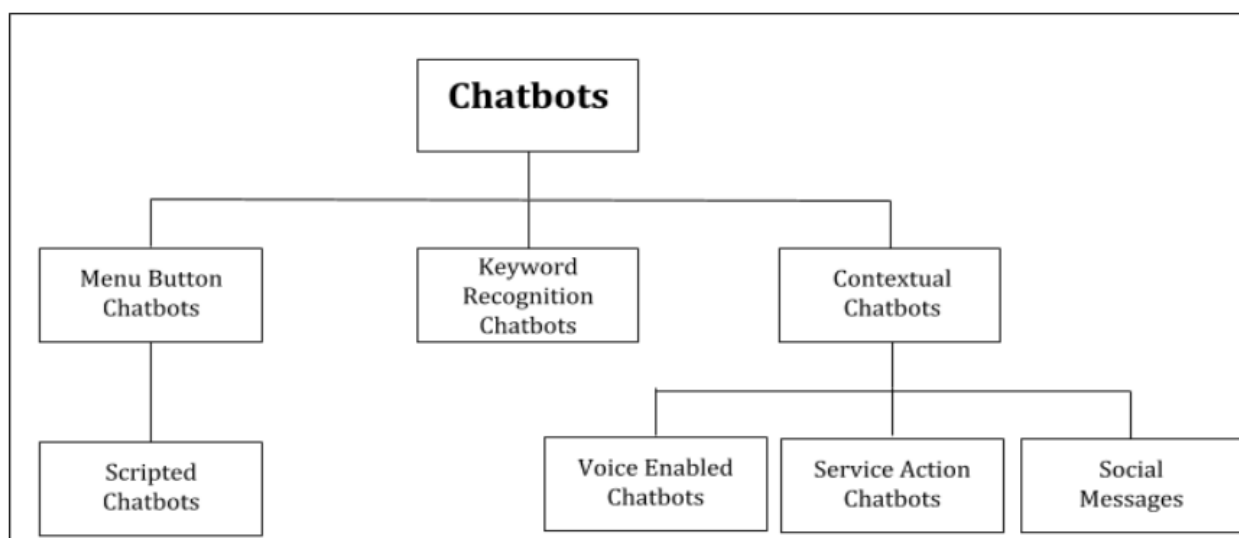
4. **Chatbots προπονητικής (Coaching)** - αναλαμβάνουν καθήκοντα προπονητή, με σκοπό την βοήθεια προς τον χρήστη ώστε να επιτύχει συγκεκριμένα αποτελέσματα, αποτελώντας έτσι ένα κλασικό μοντέλο μακροπρόθεσμων σχέσεων, καθώς η μακροχρόνια σχέση με τον χρήστη απαιτεί απαραίτητη προϋπόθεση για την σωστή λειτουργία ενός τέτοιου είδους [40].

Επιπλέον, τα chatbots μπορούν να ταξινομηθούν με βάση την ευκολία της διεπαφής τους με τον χρήστη, τους αλγόριθμους και τις βασικές τεχνολογίες που χρησιμοποιούν, όπως φαίνεται παρακάτω:

1. **Menu/Button-Based Chatbots** - Τα πιο συνηθισμένα και απλούστερα είδη chatbots στην αγορά σήμερα είναι τα chatbots που βασίζονται στο μενού. Έχουν τη μορφή κουμπιών και μενού από πάνω προς τα κάτω, ενώ χρησιμοποιούν τις αποφάσεις του χρήστη για να δώσουν τις τελικές απαντήσεις, χρησιμοποιώντας την τεχνική δέντρου απόφασης. Ο χρήστης μπορεί να λάβει απαντήσεις ανάλογα με τις επιλογές που του δίνονται και να εμβαθύνει προς την κατάλληλη απάντηση μέσω του δέντρου. Ωστόσο, παρατηρείται μία σχετική καθυστέρηση στην ταχύτητα και την απόδοση και μία έλλειψη αξιοπιστίας, όσον αφορά την απάντηση που θα δώσουν σε σχέση με τα ζητούμενα του χρήστη.
2. **Keyword Recognition-Based Chatbots** - Το συγκεκριμένο είδος chatbot λειτουργεί με την αναγνώριση συγκεκριμένων λέξεων-κλειδιών, προκειμένου να παράξει το επιθυμητό αποτέλεσμα. Δίνει ιδιαίτερη προσοχή στην είσοδο του χρήστη και ανταποκρίνεται ανάλογα. Με την βοήθεια της τεχνολογίας AI και της προσαρμοσμένης λίστας λέξεων-κλειδιών, το συγκεκριμένο είδος chatbot, μπορεί να καθορίσει μια κατάλληλη απάντηση για τον χρήστη χρησιμοποιώντας τους κατάλληλους αλγόριθμους. Τέλος, συνηθίζει να αποτυγχάνει όταν υπάρχουν εμφανίσεις λέξεων-κλειδιών μεταξύ πολλών σχετικών ερωτήσεων. Για παράδειγμα, αν ένας χρήστης κάνει μία ερώτηση όπως: "Πώς μπορώ να δημιουργήσω έναν έλεγχο ταυτότητας αυτόματης σύνδεσης στο τηλέφωνό μου;", το chatbot πιθανότατα θα χρησιμοποιούσε τις λέξεις-κλειδιά όπως "αυτόματης", "σύνδεσης", για να καθορίσει ποια απάντηση είναι η καταλληλότερη.
3. **Contextual Chatbots** - Τα Contextual Chatbots ή αλλιώς Συμφραζόμενα Chatbots, είναι ένα είδος από τα πιο προηγμένα τεχνολογικά chatbots στην αγορά αυτή την στιγμή. Εκμεταλλεύονται τεχνολογίες όπως Μάθηση Μηχανής και Τεχνητή Νοημοσύνη και με την αναγνώριση φωνής, τους αλγόριθμους μετατροπής ομιλίας σε κείμενο κλπ, έχουν την δυνατότητα ερμηνείας

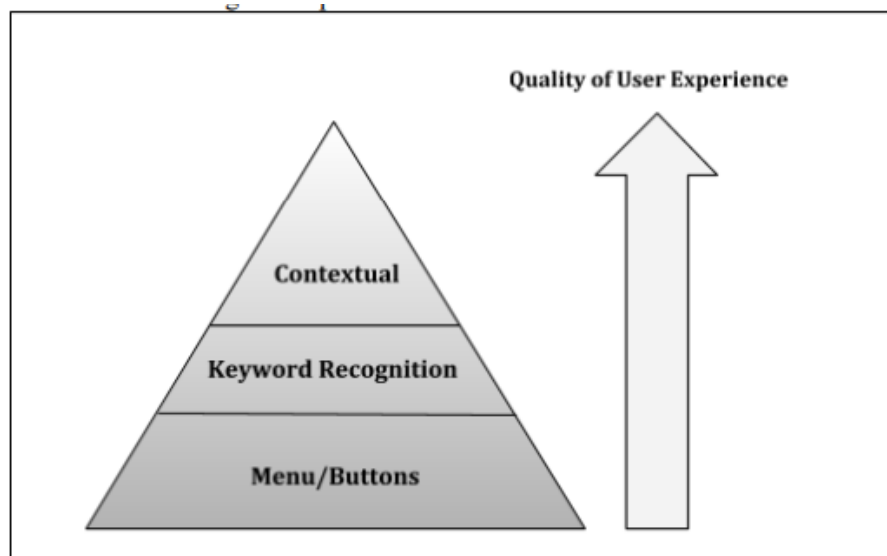
συναισθημάτων του χρήστη. Η αντίληψη των προθέσεων του χρήστη και η δημιουργία μιας στοχαστικής απάντησης, αποτελούν την βασική ιδέα του συγκεκριμένου τύπου. Σκοπός του είναι η αποκρυπτογράφηση του μοτίβου στην βάση δεδομένων, καθώς και η εκπαίδευση και ανάπτυξη του με την πάροδο του χρόνου, μέσω της συχνής επαφής με νέες ερωτήσεις και της συχνής χρήσης του. Ένα παράδειγμα για την συγκεκριμένη περίπτωση chatbot μπορεί να υπάρξει σε μια εφαρμογή παράδοσης τροφίμων. Το προηγούμενο ιστορικό παραγγελιών, μαζί με τις επιλογές πληρωμής και την διεύθυνση παράδοσης του χρήστη αποθηκεύονται στην βάση δεδομένων, ώστε να αναλυθεί η προοπτική του χρήστη και να δημιουργηθούν προτάσεις με βάση τις διαδοχικές παραγγελίες και τις προτιμήσεις του χρήστη [4].

Παρακάτω μπορούμε να παρατηρήσουμε το διάγραμμα των κατηγοριών των Chatbots :



Εικόνα 3. Κατηγορίες και είδη Chatbot. [4]

Η ποιότητα της εμπειρίας χρήστη με βάση τον τύπο chatbot εμφανίζεται στην εικόνα παρακάτω:

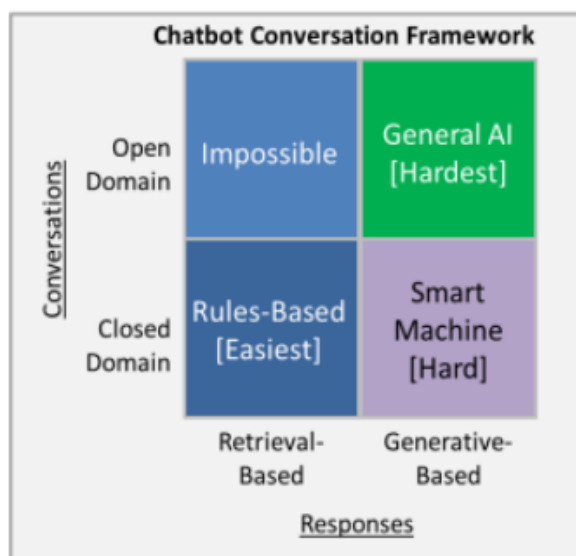


Εικόνα 4. Προτίμηση Chatbots [4]

Τέλος, σύμφωνα με την τεχνική πολυπλοκότητα, τα chatbots μπορούν να ταξινομηθούν σε τρεις ομάδες - απλά chatbots, έξυπνα chatbots και υβριδικά chatbots:

1. **Απλά chatbots** - κάνουν ερωτήσεις με προκαθορισμένες απαντήσεις και ο χρήστης πρέπει να επιλέξει επανειλημμένα μια επιλογή έως ότου λάβει την απαιτούμενη απάντηση στην ερώτηση.
2. **Έξυπνα chatbots** - διαθέτουν τεχνητή νοημοσύνη και μπορούν να αποκρυπτογραφήσουν την γλώσσα, το συναίσθημα και την πρόθεση που ο χρήστης επιθυμεί να επικοινωνήσει.
3. **Υβριδικά chatbots** - πρόκειται για τον συνδυασμό των δύο παραπάνω.

Επιπροσθέτως, τα chatbots θα μπορούσαν να ταξινομηθούν με διαφορετικές παραμέτρους, όπως μπορούμε να παρατηρήσουμε και στον παρακάτω πίνακα, με τις κατηγορίες: γνωστικός τομέας, παρεχόμενη υπηρεσία, στόχοι, επεξεργασία εισόδου και παραγωγή απάντησης, ανθρώπινη προσφορά και αρχιτεκτονική μέθοδος [48]. Τέλος, παρακάτω μπορούμε να παρατηρήσουμε τις κατηγορίες που προαναφέραμε πιο αναλυτικά στον πίνακα, καθώς και την ταξινόμηση με βάση την φιλοσοφία σχεδίαση και τον τομέα γνώσης στην παρακάτω εικόνα:



Εικόνα 5. Ταξινόμηση Chatbot, με βάση την φιλοσοφία σχεδίασης (οριζόντια) και τον τομέα γνώσης (κάθετα)[44]

Ταξινόμηση	Υποκατηγορία	Περιγραφή
Γνωστικός Τομέας	Ανοιχτός Τομέας	Σε αυτόν τον τομέα ανήκουν τα chatbots τα οποία θα μπορούσαν να συνεχίσουν την συζήτηση για γενικά θέματα και να ανταποκριθούν κατάλληλα.
	Κλειστός Τομέας	Σε αυτόν τον τομέα ανήκουν τα chatbots τα οποία εστιάζουν σε ένα συγκεκριμένο θέμα και σε αντίθεση με την προηγούμενη κατηγορία, δεν είναι σε θέση να απαντήσουν σε γενικές ερωτήσεις.
Υπηρεσίες που παρέχονται	Διαπροσωπικές	Τα chatbots τα οποία παρέχουν τέτοιου είδους λειτουργίες, στοχεύουν στην επικοινωνία και στην παροχή υπηρεσιών όπως η κράτηση εστιατορίου, η κράτηση πτήσεων και η απάντηση σχετικών συχνών ερωτήσεων.
	Διαπροσωπικές	Τα chatbots τα οποία παρέχουν τέτοιου

		είδους λειτουργίες, διαθέτουν “συντροφικό” ρόλο για τον χρήστη και τα συναντάμε σε εφαρμογές συνομιλίας, όπως το Messenger, το Slack και το Whatsapp.
	Ανθρωποκεντρικές	Τα chatbots τα οποία παρέχουν τέτοιου είδους λειτουργίες, προσφέρουν υπηρεσίες οι οποίες βασίζονται στις ανάγκες και τις προτιμήσεις του χρήστη με σκοπό την παροχή υπηρεσιών σε αυτόν μέσω της επικοινωνίας του με αυτό. Σε αυτόν τον τομέα ανήκουν chatbots όπως π.χ. η Alexa και η Google Home.
Στόχοι	Πληροφοριακοί	Ο στόχος αυτών των chatbots είναι η παροχή πληροφοριών προς τους χρήστες. Πληροφορίες οι οποίες έχουν αποθηκευτεί προηγουμένως ή είναι διαθέσιμες από σταθερές πηγές, όπως για παράδειγμα πηγές που διαθέτουν συχνές ερωτήσεις για chatbots.
	Επικοινωνιακοί	Ο στόχος αυτών των chatbots είναι η συνομιλία με τους χρήστες, να ήταν αληθινά άτομα με συναισθηματική νοημοσύνη και σκοπός τους είναι να απαντήσουν με ακρίβεια στην πρόταση που είναι επιζητά ο χρήστης.
	Εργασιακοί	Ο στόχος αυτών των chatbots είναι η παροχή υπηρεσιών συγκεκριμένου σκοπού προς τον χρήστη όπως η κράτηση ενός εισιτηρίου για μία πτήση ή η παροχή βοήθειας σε κάποιον.
Επεξεργασία εισόδου και παραγωγή απάντησης	Βασισμένο σε κανόνες	Τα συγκεκριμένα chatbots επιλέγουν μια απόκριση συστήματος με βάση ένα καθορισμένο σύνολο κανόνων οι οποίοι βασίζονται στον προσδιορισμό της λεξιλογικής μορφής του εισαγόμενου κειμένου αντί να αναπτύσσουν νέες απαντήσεις κειμένου.

	Βασισμένο στην ανάκτηση	Τα συγκεκριμένα chatbots εφαρμόζουν στρατηγική αντιστοίχισης στην επιλογή απόκρισης, λαμβάνοντας πολλές επιλογές απόκρισης από ένα ευρετήριο.
	Παραγωγικό	Τα συγκεκριμένα chatbots χρησιμοποιούν αλγόριθμους μηχανικής μάθησης και τεχνικές βαθιάς εκμάθησης με σκοπό να δημιουργηθούν πιο ανθρωπομορφικά chatbot.
Ανθρώπινη Προσφορά	Ανθρωποκεντρικό	Τουλάχιστον ένα μέρος από τα chatbots τροφοδοτείται από άνθρωπινους υπολογισμούς. Ως αποτέλεσμα αυτού, επιβραδύνεται η διαδικασία και μειώνονται τα αιτήματα που θα μπορούσαν να διεκπεραιωθούν λόγω ανθρώπινων περιορισμών.
Αρχιτεκτονική Μέθοδος	Ανοιχτή Πηγή	Η συγκεκριμένη αρχιτεκτονική παρέχει την δυνατότητα στον δημιουργό του chatbot να παρέμβει στις περισσότερες πτυχές της υλοποίησης.
	Κλειστή Πηγή	Η συγκεκριμένη αρχιτεκτονική λειτουργεί ως μαύρο κουτί στις περισσότερες περιπτώσεις, κάτι που μπορεί να είναι μεγάλο μειονέκτημα ανάλογα με τις απαιτήσεις του έργου.

1.3 Συνεισφορά της Διπλωματικής

Η βιβλιογραφία που έχει δημιουργηθεί μέχρι στιγμής, καθώς και τα αμέτρητα πειράματα σε chatbots με την χρήση ιατρικών και μη δεδομένων, δημιουργούν ολοένα και πιο καθαρή εικόνα σχετικά με τα chatbots και τις εφαρμογές τους, καθώς και την σχέση τους με την βιοϊατρική. Ωστόσο, τα περιθώρια έρευνας και ανάπτυξης είναι ατελείωτα και η ανάγκη για έρευνα αυξάνεται συνεχώς, όσο αυξάνονται οι ανάγκες της

καθημερινότητας, πόσο μάλλον οι ανάγκες της ιατρικής. Στην συγκεκριμένη διπλωματική, θα προσπαθήσουμε να δούμε τι επιδράσεις μπορεί να έχει ένα chatbot με δεδομένα εισόδου βιοϊατρικής φύσεως, σε συνδυασμό με την χρήση και μη αλγοριθμικών μοντέλων BERT, καθώς και παρατηρήσουμε τι επιδράσεις μπορεί να επιφέρει ο κάθε αλγόριθμος και εάν τελικά συμφέρει η χρήση του, σε σχέση με τη μη χρήση αλγορίθμου και την χρήση του CountVectorizer και ερωτήσεις τόσο γενικές, όσο και ειδικές σε ιατρικά δεδομένα εισόδου.

Πιο συγκεκριμένα, η συνεισφορά της διπλωματικής μας αποτελείται από:

- Την υλοποίηση ενός chatbot με χρήση βιοϊατρικών δεδομένων και την παροχή του κώδικα σε ανοιχτή βιβλιοθήκη Github : <https://github.com/Evdokia97/MediBot>
- Την χρήση δεδομένων PubMed διαφόρων κλάδων της ιατρικής.
- Την σύγκριση της “συμπεριφοράς” τεσσάρων μοντέλων BERT σε chatbot με χρήση βιοϊατρικών δεδομένων.
- Την σύγκριση της “συμπεριφοράς” του αλγορίθμου BERT και του CountVectorizer σε chatbot με χρήση βιοϊατρικών δεδομένων.

2

Chatbot στην Βιοϊατρική

2.1 Η Ιστορία των Chatbots στην Βιοϊατρική

Η απόκτηση βιοϊατρικής γνώσης αποτελεί τα τελευταία χρόνια ένα μείζον έργο στην ανάκτηση πληροφοριών και διαχείριση γνώσης. Οι επαγγελματίες καθώς και το ευρύ κοινό χρειάζονται αποτελεσματική βοήθεια για πρόσβαση, κατανόηση και κατανάλωση πολύπλοκων βιοϊατρικών εννοιών. Παράδειγμα αποτελεί το γεγονός ότι οι γιατροί επιθυμούν να παραμείνουν ενήμεροι για τα επικαιροποιημένα κλινικά στοιχεία, ώστε να εξελίσσονται στο κομμάτι της διάγνωσης και θεραπείας των ασθενειών οι οποίες βασίζονται σε ιατρικά δεδομένα. Επιπλέον, τα τελευταία χρόνια έχει παρατηρηθεί μια αύξηση του ενδιαφέροντος του ευρύ κοινού ως προς την προσωπική του έρευνα σχετικά με την υγεία του και τα ιατρικά δεδομένα μέσω του διαδικτύου. Έως τώρα, έχουν αναπτυχθεί συστήματα ανάκτησης πληροφοριών (IR), όπως το PubMed, τα οποία έχουν χρησιμοποιηθεί για την κάλυψη τέτοιων αναγκών.

Τα chatbots αποτελούν μία από τις πιο δημοφιλείς τεχνολογίες ΑΙ και μπορούμε να πούμε με σιγουριά πως η τεχνητή νοημοσύνη παίζει καθοριστικό ρόλο στην πρόοδο της τεχνολογίας των πληροφοριών, καθώς και στην βελτίωση της ποιότητας και αποτελεσματικότητας των υπηρεσιών υγείας. Το σύστημα λογισμικού των chatbot προσφέρει μια διαδραστική διεπαφή με ασθενείς ή και ιατρούς, με στόχο την παροχή μιας σειράς εργασιών εξαγωγής γνώσης και εξατομικευμένων σχολίων σε πραγματικό χρόνο. Οι τεχνολογίες των Chatbots σημειώνουν τα τελευταία χρόνια γρήγορη ανάπτυξη, ειδικά στον ιατρικό τομέα, και πολλά ιατρικά συστήματα chatbot έχουν προταθεί όλα αυτά τα χρόνια με στόχο την βέλτιστη ποιότητα συστήματος που μπορεί να δημιουργηθεί για την ιατρική κοινότητα και όχι μόνο.[81]

Το σύστημα αυτόματης απάντησης στις ερωτήσεις του χρήστη (QA) εφαρμόστηκε με επιτυχία σε διάφορους τομείς, όπως σε μηχανές αναζήτησης και chatbots. Το Biomedical QA (BQA) αποτέλεσε μια καινοτομία στον τομέα του, ως ένα αναδυόμενο έργο QA, και επιτρέπει στις εφαρμογές να αντιλαμβάνονται αποτελεσματικά, καθώς και να έχουν πρόσβαση και να κατανοούν περίπλοκα βιοϊατρικά δεδομένα ή παρόμοιες γνώσεις. Τις τελευταίες δύο δεκαετίες σημειώθηκαν ραγδαίες εξελίξεις στον τομέα του BQA, οι οποίες διαφοροποιούνται ανάλογα με την προσέγγισή τους σε 5 κατηγορίες: κλασική, ανάκτηση πληροφοριών, κατανόηση μηχανικής ανάγνωσης, βάση γνώσεων και προσεγγίσεις που συνεπάγονται σε ερωτήσεις. Παρά τις εξελίξεις,

τα συστήματα BQA είναι ακόμα ανώριμα και σπάνια χρησιμοποιούνται σε πραγματικές συνθήκες.[124]

Οι Russell-Rose και Chamberlain ανέφεραν ότι απαιτούνται 4 ώρες εμπειρογνομόνων για να απαντήσουν σύνθετα ιατρικά ερωτήματα χρησιμοποιώντας μηχανές αναζήτησης, καθιστώντας το κλασσικό IR μη αποτελεσματικό. Σε αντίθεση με τα IR, τα συστήματα ανάκτησης συνήθως επιστρέφουν μια λίστα σχετικών εγγράφων στον χρήστη και συστήματα όπως τα Απάντησης-Ερωτήσεων (QA) παρέχουν άμεσες απαντήσεις στις ερωτήσεις των χρηστών είναι πιο απλά και διαισθητικά. Σε γενικές γραμμές, το QA αποτελεί πρόκληση όσον αφορά την χρήση Φυσικής Γλώσσας (NLP) για την αξιολόγηση των ικανοτήτων των ευφυών συστημάτων για την κατανόηση μιας ερώτησης, την ανάκτηση και την χρήση σχετικών υλικών, καθώς και την παραγωγή της απάντησής της. [125]

Με την πάροδο του χρόνου και την ταχεία ανάπτυξη του υπολογιστικού υλικού, τα σύγχρονα μοντέλα QA - ειδικά εκείνα που βασίζονται στην βαθιά μάθηση - επιτυγχάνουν συγκρίσιμες ή και καλύτερες επιδόσεις από τις ανθρώπινες σε πολλά σύνολα δεδομένων αναφοράς και έχουν υιοθετηθεί επιτυχώς σε μηχανές αναζήτησης γενικού τομέα και βοηθούς συνομιλιών. Αργότερα, προτάθηκαν πολλά κλασσικά συστήματα BQA, όπως το EPoCare[126], το PICO, όπως και συστήματα BQA που βασίζονται στην εξαγωγή γνώσης, όπως τα MedQA, Terol et al., Weiming et al., Health on the Net QA (HONQA), AskHERMES κ.λπ. Τέτοια συστήματα χρησιμοποιούν περίπλοκους αγωγούς με πολυάριθμες μονάδες επεξεργασίας ερωτήσεων, εγγράφων και απαντήσεων, οι οποίες αντικατοπτρίζονται συνήθως από το σύστημα IBM Watson.

Το 2013 δημιουργήθηκε το BioASQ, το οποίο αποτέλεσε πρόκληση στον τομέα της αξιολόγησης βιοϊατρικών συστημάτων και της κατανόησης φυσικής γλώσσας. Πρότεινε επίσης μια ποικιλία συστημάτων BQA, βελτιώνοντας έτσι την απόδοση των QA από περίπου 20% κορυφαία μέση αμοιβαία βαθμολογία και λίστα F-μέτρων στο BioASQ1 σε περίπου 50% στο BioASQ8.

Επίσης, αξίζει να σημειωθεί ότι τα μοντέλα που συμμετέχουν στο BioASQ έχουν διαμορφωθεί εκ νέου από διάφορες μεθοδολογικές επαναστάσεις του NLP:

1. Η εισαγωγή κατανεμημένων αναπαραστάσεων λέξεων.
2. Μοντέλα QA βασισμένα στην βαθιά μάθηση, όπως το Bi-Directional Attention Flow (Bi-DAF).
3. Μεγάλης κλίμακας Προ-εκπαιδευμένα Μοντέλα Γλωσσών (PLM) που αντιπροσωπεύονται από Ενσωμάτωση για Μοντέλα Γλώσσας (ELMo) και αμφίδρομες αναπαραστάσεις κωδικοποιητών από μετασχηματιστές (BERT).[81]

Επί του παρόντος, σχεδόν όλα τα συστήματα BQA κορυφαίας απόδοσης χρησιμοποιούν τα βιοϊατρικά PLM (π.χ. BioBERT[127]) στα συστήματά τους και εισάγουν διάφορες προκλήσεις και σύνολα δεδομένων, ώστε να διευκολυνθεί περαιτέρω η έρευνα BQA σε διαφορετικές κατευθύνσεις, π.χ. VQA-Med και PathVQA για οπτικό BQA. Παρά την πρόοδο που έχει σημειώσει έως τώρα, το BQA είναι ακόμα ανώριμο και αντιμετωπίζει αρκετές βασικές προκλήσεις όπως:

1 Κλίμακα δεδομένων:

Τα περισσότερα μοντέλα BQA που υπάρχουν αυτή την στιγμή, χρησιμοποιούν την τεχνική βαθιάς εκμάθησης, με αποτέλεσμα να μην λαμβάνονται τα επιθυμητά αποτελέσματα. Ωστόσο, ο σχολιασμός βιοϊατρικών δεδομένων μεγάλης κλίμακας ή βάσεων γνώσης είναι απαγορευτικά ακριβός για τον συγκεκριμένο τύπο μοντέλου, έχοντας ως αποτέλεσμα τα σύνολα δεδομένων να είναι μικρού μεγέθους, με μόνο εκατοντάδες έως μερικές χιλιάδες παρουσίες ποιότητας. Για την δημιουργία συνόλων δεδομένων μεγάλης κλίμακας, πολλά έργα προσπάθησαν να συλλέξουν αυτόματα σύνολα δεδομένων BQA, αλλά η χρησιμότητά τους είναι περιορισμένη και η ποιότητα σχολιασμού τους δεν είναι εγγυημένη. Τέλος, οι ερωτήσεις των περισσότερων σύγχρονων συνόλων δεδομένων BQA δεν απαιτούν πολύπλοκο σκεπτικό για να απαντηθούν.

2 Γνώση Περιοχής που δεν χρησιμοποιείται πλήρως:

Υπάρχουν πλούσιοι βιοϊατρικοί πόροι που περικλείουν διαφορετικούς τύπους γνώσεων τομέα, συμπεριλαμβανομένων μεγάλων κλιμάκων, διάφορων βιοϊατρικών Knowledge Base (KB), εργαλείων NLP και εικόνες. Έχει παρατηρηθεί πως τα περισσότερα μοντέλα BQA αποτυγχάνουν στην χρήση τους προκαλώντας ως αποτέλεσμα τη μη ολοκληρωμένη χρήση της βιοϊατρικής γνώσης, γεγονός που θα μπορούσε να λυθεί δυνητικά με την συγχώνευση διαφορετικών προσεγγίσεων BQA.

3 Έλλειψη επεξηγήσεων:

Δεδομένου ότι η βιοϊατρική είναι ένας εξαιρετικά εξειδικευμένος τομέας, τα ιδανικά συστήματα δεν πρέπει μόνο να επιστρέφουν τις ακριβείς απαντήσεις (π.χ.: «ναι»/«όχι»), αλλά και να παρέχουν επαρκείς αιτιολογήσεις και λεπτομέρειες για τις απαντήσεις που θα δώσουν. Ωστόσο, δεν υπάρχουν πολλά συστήματα BQA που καλύπτουν αυτή την ανάγκη.

4 Θέματα αξιολόγησης:

Ποιοτικά, οι τρέχουσες αξιολογήσεις επικεντρώνονται κυρίως σε συγκεκριμένες ενότητες, π.χ. κατανόηση ανάγνωσης μηχανών (MRC), εντός ενός πλήρους αγωγού QA και δεν παρουσιάζουν πλούσιες, ποσοτικά, βιοϊατρικά συνώνυμες σχέσεις.

5 Δικαιοσύνη και προκατάληψη:

Τα περισσότερα συστήματα BQA που βασίζονται σε μηχανική εκμάθηση μαθαίνουν από ιστορικά δεδομένα, όπως επιστημονική βιβλιογραφία και ηλεκτρονικά ιατρικά αρχεία, τα οποία μπορεί να είναι δυνητικά προκατειλημμένα και ξεπερασμένα. Ωστόσο, οι μελέτες σχετικά με την δικαιοσύνη του BQA και την διαφάνεια του μοντέλου είναι αρκετά αραιές[81].

2.2 Σχετικό Έργο των Chatbots στην Βιοϊατρική

Ο χρόνος των γιατρών μπορεί να εξοικονομηθεί σε μεγάλο βαθμό, καθώς και η έρευνα όσον αφορά το ιατρικό κομμάτι να γίνει πιο εύκολη και γρήγορη, μέσω της χρήσης chatbot. Για την εφαρμογή ενός chatbot, χρησιμοποιούνται συνήθως τεχνολογίες αλληλεπίδρασης ανθρώπου-υπολογιστή (HCI) που επιτρέπουν την επικοινωνία μεταξύ χρηστών και υπολογιστών χρησιμοποιώντας φυσική γλώσσα. Η χρήση της επεξεργασίας φυσικής γλώσσας (NLP) προσφέρει στην κατανόηση του χρήστη και την δημιουργία μιας απάντησης περισσότερο κατανοητής προς αυτόν. Σύμφωνα με τα τρέχοντα σενάρια, το 80% των εργασιών στο γραφείο, θα μπορούσε να αντικατασταθεί κάλλιστα από chatbots στο εγγύς μέλλον.

Όσον αφορά τον ιατρικό τομέα, έχει παρατηρηθεί ότι οι άνθρωποι δεν είναι συνεπείς στις επισκέψεις τους στους γιατρούς, είτε από αμέλεια, είτε λόγω αποφυγής. Ωστόσο, περνούν ένα σημαντικό χρονικό διάστημα στην περιήγηση στα τηλέφωνα τους, ενώ παράλληλα είναι επιφυλακτικοί όσον αφορά τα αυτόματα ή μηχανικά προσανατολισμένα περιβάλλοντα και προτιμούν να αλληλεπιδρούν με έναν άλλο άνθρωπο, παρά με ένα chatbot. Τα Chatbots θα μπορούσαν να δημιουργήσουν ένα περιβάλλον όπου οι άνθρωποι δεν θα κρίνονται, αλλά θα βρίσκουν λύσεις στα προβλήματά τους με βάση τις γνώσεις που διαθέτουν σε ένα θετικό περιβάλλον για τους ασθενείς, δίχως αποθάρρυνση.

Η συνεχής βελτίωση και ανάπτυξη των chatbots έχει προκαλέσει αύξηση στην προτίμησή τους από τις επιχειρήσεις για τις υπηρεσίες εξυπηρέτησης πελατών τους. Τα ιατρικά chatbots παρέχουν κατευθυντήριες γραμμές για την σωστή διατροφή, αλλαγές στον τρόπο ζωής και περαιτέρω παθολογικές εξετάσεις με βάση τα συμπτώματα των ασθενών που παρέχονται στην είσοδό τους και με την χρήση της τεχνητής νοημοσύνης, θα μπορούσαν να συνταγογραφήσουν επίσης κατάλληλη φαρμακευτική αγωγή, η οποία μπορεί να εγκριθεί περαιτέρω από τον γιατρό του ασθενή. [103]

Τα τελευταία χρόνια, έχουν γίνει αρκετές προσπάθειες για την δημιουργία πολυποίκλων ειδών chatbots στον τομέα της βιοϊατρικής. Παρακάτω υπάρχουν ορισμένες συγκρίσεις τους όσον αφορά τις τεχνικές και τα αποτελέσματά τους:

1. A Self Diagnosis Medical Chatbot using Artificial Intelligence

Πρόκειται για ένα chatbot με γραμμικό σχεδιασμό που περιέχει τεχνητή νοημοσύνη, μοτίβο αντιστοίχισης και τεχνικές επεξεργασία ερωτήματος. Παρέχει τον διάλογο του chatbot με τον χρήστη σε ένα γράφημα πεπερασμένης κατάστασης και προσφέρει λειτουργική αρχιτεκτονική. Έχει επίσης την δυνατότητα αναγνώρισης των αντίστοιχων συμπτωμάτων και μπορεί να παρέχει συμβουλές και προσωποποιημένες διαγνώσεις για την αντίστοιχη ασθένεια, ασχέτως της σοβαρότητας της. Αντιλαμβανόμαστε λοιπόν ότι πρόκειται για ένα σύστημα φιλικό προς το χρήστη και εφαρμογή του μπορεί να προσφέρει

μία επαρκή θεραπεία σε σύντομο χρονικό διάστημα. Τέλος, μπορεί να εντοπίσει όχι μόνο μικρές, αλλά και μεγάλες ασθένειες και να δώσει την άδεια για το αν ο χρήστης είναι αναγκαίο να επισκεφθεί γιατρό ή όχι.[34]

2. Disease Prognosis with Symptoms Endorsement using Combination al Classifier

Το συγκεκριμένο chatbot, δίνει ένα μοντέλο για πρόβλεψη ασθένειας από συμπτώματα ομαδοποιημένα και με συνδυαστική ταξινόμηση, με χρήση διαφόρων αλγορίθμων όπως το K-Nearest Neighbor, Decision Tree, Naive Bayes. Για την αναγνώριση των συμπτωμάτων και την πρόβλεψη της ασθένειας χρησιμοποιεί σχεσιακούς κανόνες και χρησιμοποιεί αλγόριθμο απόφασης δέντρου και Bayes Θεώρημα.

Επιπλέον, μπορεί να βελτιωθεί εξαιτίας του συνδυασμού των ταξινομήσεων, σε αντίθεση με τα συστήματα που είναι βασισμένα σε κανόνες και ακρίβεια των προβλέψεων. Τέλος, ένα αρνητικό του είναι ότι εάν ο αριθμός των συμπτωμάτων είναι μικρότερος, τότε η ακρίβεια εντοπισμού ασθενειών μπορεί να υποβαθμιστεί.[16]

3. Disease Prediction using Machine Learning

Προσφέρει ένα διάγραμμα ροής για την πρόβλεψη της ασθένειας του χρήστη, ακολουθούμενο από σύνολα δεδομένων, επιλογή χαρακτηριστικών, προεπεξεργασία δεδομένων, δέντρο απόφασης, ταξινόμηση, επεξεργασμένο κείμενο και αντιστοίχιση με λέξη-κλειδί. Στην βάση της βιβλιοθήκης δεδομένων από τα συμπτώματα, το σύστημα προβλέπει την ασθένεια με ακρίβεια και είναι εύκολο σε πρόσβαση, γεγονός που καθιερώνει πραγματικούς χρόνους επικοινωνίας χρησιμοποιώντας μοντέρνες και ανανεωμένες τεχνολογίες.

Επιπλέον, πρόκειται για ένα διαδραστικό και φιλικό προς το περιβάλλον chatbot, όσον αφορά την πρόβλεψη της νόσου ενός ασθενούς με τους αλγόριθμους δέντρων αποφάσεων να χρησιμοποιούνται για καλύτερη ακρίβεια πρόβλεψης ασθενειών. Τέλος, δεν είναι σε θέση να διαχειριστεί σωστά κάθε νέο σύμπτωμα που δίνεται από ένα ασθενή.[59]

4. System based on ML and NLP for supporting Women and Families during Pregnancy

Το συγκεκριμένο chatbot διαθέτει αλγόριθμο Support Vector Machine (SVM), αλγόριθμο Porter Stemming και ομοιότητα λέξεων. Διαθέτει λεπτομέρειες φαρμάκων και δοσολογιών, καθώς και προβλέπει την πιθανή ασθένεια του χρήστη. Χρησιμοποιεί τέσσερα σύνολα δεδομένων για διαφορετικά μεγέθη, ώστε να δώσει πληροφορίες για την καρδιακή ασθένεια.

Κατα την δημιουργία του, δοκιμάστηκαν τρεις αλγόριθμοι οι KNN, Naive Bayes

και SVM, με τον SVM να έχει υψηλότερη ακρίβεια. Το συγκεκριμένο chatbot είναι χρήσιμο για τον χρήστη που αναζητά ερωτήσεις σχετικά με ιατρική δοσολογία, φωνητικά καθώς και σε μορφή κειμένου. Το σύστημα λαμβάνει έξοδο για το API ιατρικής, μιλάει και εμφανίζει όλα τα ονόματα φαρμάκων. Παρέχει επίσης πρόβλεψη ασθενειών χρησιμοποιώντας αλγόριθμο SVM και η συνομιλία φωνής-κειμένου είναι δυνατή. Ωστόσο, δεν διαθέτει λειτουργικότητα πλοήγησης.[15]

5. MamaBot

Πρόκειται για ένα Microsoft Bot Framework, LUIS(Language Understanding intelligent Service), το οποίο αναζητά φαρμακεία και νοσοκομεία σε κοντινή εμβέλεια, χρησιμοποιώντας την τοποθεσία του χρήστη. Επιπλέον, δίνει συμβουλές για σωστή διατροφή για τα παιδιά, διαχειρίζεται περιστατικά έκτακτης ανάγκης για παιδιά, προσφέρει υποστήριξη και βοήθεια εγκύων γυναικών, μητέρων και οικογενειών με μικρά παιδιά για τυχόν αμφιβολίες ή προβλήματα που μπορεί να προκύψουν κατά την διάρκεια της εγκυμοσύνης. Τέλος, είναι βασισμένο σε AI, όμως δεν πραγματοποιεί συνομιλίες φωνητικού κειμένου ή επαλήθευση χρήστη.[104]

6. Developing Disease Classification System based on Keyword Extraction and Supervised Learning.

Το συγκεκριμένο χρησιμοποιεί τον αλγόριθμο K-mean, LAMSTA, συλλογή ιατρικού συνόλου δεδομένων, συγκρίσεις συμπτωμάτων χρησιμοποιώντας επαναληπτική αναζήτηση και επικεντρώνεται κυρίως στην επεξεργασία κειμένου, εξαγωγή χαρακτηριστικών και στην ταξινόμηση, ώστε να προτείνει την πρόληψη ασθενειών. Τα αποτελέσματα του αξιολογούνται από την ακρίβεια, τις F1-βαθμολογίες της ασθένειας και με βάση αυτά προκύπτουν ακρίβειες Bayes.

Συμπεραίνεται ότι το σύστημα έχει δημιουργηθεί περισσότερο για την επεξεργασία κειμένου και εξαγωγής χαρακτηριστικών, χρησιμοποιώντας NLP και τεχνικές μάθησης μηχανών για την πρόληψη ασθενειών, καθώς και για τον υπολογισμό των αποτελεσμάτων πολλαπλών λέξεων -κλειδιών και φράσεων-κλειδιών, με καλά αποτελέσματα, σε μικρά σύνολα δεδομένων. Δυστυχώς όμως, στον πραγματικό χρόνο και στα προβλήματα που σχετίζονται με το τοπικό περιβάλλον, τα σύνολα δεδομένων δεν έχουν ληφθεί υπόψη, ενώ παράλληλα πρόκειται για σύνολα δεδομένων.[19]

2.3 Ηθική και Ασφάλεια στον Τομέα των Chatbots

Την τελευταία δεκαετία, οι ιατροί ηθικολόγοι προσπάθησαν να σκιαγραφήσουν αρχές και πλαίσια για την ηθική ανάπτυξη των αναδυόμενων τεχνολογιών, ιδίως της AI, στην υγειονομική

περίθαλψη. Τα chatbots έχουν αποκτήσει ραγδαία δημοτικότητα κατά την διάρκεια της πανδημίας του COVID-19, προκαλώντας έτσι τους ιατρικούς εμπειρογνώμονες να ανταποκριθούν πιο γρήγορα στις νομικές και ηθικές πτυχές των chatbots. Δημιουργήθηκε έτσι μια σειρά από ηθικές προκλήσεις που έχουν προκύψει σε σχέση με τα chatbots στην υγειονομική περίθαλψη, συμπεριλαμβανομένων των: ασφάλεια των ασθενών, εμπιστοσύνη, διαφάνεια μεταξύ όλων των συμμετεχόντων, πηγές περιεχομένου των συστάσεων των chatbots, απειλές κυβερνοασφάλειας, χρήση δεδομένων, ιδιωτικό απόρρητο, ολοκλήρωση, προκατάληψη και ισότητα της υγείας.

Αν προσπαθήσουμε να αναθεωρήσουμε τα προηγούμενα ερευνητικά ευρήματα σχετικά με τα chatbots, ώστε να απαντήσουμε στο ερώτημα ποια είδη κλινικών καθηκόντων πρέπει ή δεν πρέπει να αυξηθούν ή να αυτοματοποιηθούν από τα chatbots, θα καταλήξουμε πως το κλειδί είναι να τονιστεί η ικανότητα του chatbot και να διαβεβαιωθούν οι χρήστες ότι τους παρέχεται η ίδια ποιότητα υπηρεσιών με τους ανθρώπινους πράκτορες.

2.3.1 Ηθικές Επιπτώσεις των Chatbots προς τους Χρήστες

Τα chatbots ενσωματώνουν όλο και περισσότερο το σχεδιασμό μηχανισμών με την χρήση του AI (όπως θεωρία παιγνίων, εξόρυξη δεδομένων και απόψεων, τεχνικές βελτιστοποίησης) σε διαδικτυακούς ιστότοπους κοινωνικής δικτύωσης. Αυτό συνεπάγεται ότι θα πρέπει να συμμορφώνονται με τους κανόνες και την δυναμική των διαδικτυακών κοινωνικών δικτύων, καθώς και να μπορούν να ανταποκριθούν σε πραγματικές συνομιλίες πολλαπλών παραγόντων που απαιτούν τεχνικούς πόρους, ειδικές γνώσεις και ικανότητες επικοινωνίας. Μία επιπλέον χρήση της συγκεκριμένης διαδικασίας, είναι συνήθως η δημιουργία συμφωνιών, η διευκόλυνση των διαδικασιών λήψης αποφάσεων, η βελτίωση της ποιότητας των αλληλεπιδράσεων, καθώς και η χρήση της ως μέσο συλλογής πληροφοριών και ενσωμάτωσης δεδομένων από τους οργανισμούς. Η συγκεκριμένη χρήση βασίζεται σε ένα επίπεδο εμπιστοσύνης, ακόμη και αν αυτή η εμπιστοσύνη βασίζεται σε μια αυστηρά συναλλακτική κατανόηση. Οι ανθρώπινες ιδιότητες των chatbots είναι τέτοιες ώστε να ενθαρρύνουν τους ψηφιακούς χρήστες να τα εμπιστευτούν, όχι μόνο για την συναλλαγή που πραγματοποιούν, αλλά και για να αποκαλύψουν πρόσθετες προσωπικές πληροφορίες, οι οποίες στην συνέχεια θα παραμείνουν μόνιμα αποθηκευμένες, προσβάσιμες και ενσωματωμένες, έτσι ώστε να μπορεί να αναπτυχθεί με την πάροδο του χρόνου μια ευρύτερη εικόνα μεμονωμένων καταναλωτών. Επιπλέον, αυτά τα μεμονωμένα δεδομένα μπορούν να αξιοποιηθούν για την δημιουργία ολοκληρωμένων προφίλ καταναλωτών για τον εκάστοτε οργανισμό.

Πιο συγκεκριμένα, τα chatbots δεν είναι ηθικοί και ανεξάρτητοι πράκτορες με ηθικές ικανότητες συλλογισμού. Προκειται για μια παροχή υπηρεσιών που στοχεύει στην βελτίωση των εμπειριών των πελατών και της αλληλεπίδρασής τους με αυτούς, βάσει δεδομένων. Για παράδειγμα, μπορούν να βελτιώνουν την ικανότητα των οργανισμών, παρέχοντας εξατομικευμένες εμπειρίες πελατών, προσαρμοσμένες στις ανάγκες τους με βάση πληροφορίες που συλλέγονται αυτόνομα κατά την διάρκεια αλληλεπιδράσεων τους με τον χρήστη. Ωστόσο, ακόμη και αν οι οργανισμοί βελτιώσουν την προσωπικότητα των chatbots, παρέχοντας ένα φιλικό προς το χρήστη λογισμικό, τα chatbots υστερούν στον τομέα των ανθρώπινων ιδιοτήτων, όπως

στην κρίση, την ενσυναίσθηση και την διακριτικότητα και συνεπώς δεν είναι σε θέση να ανιχνεύσουν περιπτώσεις όπου οι ενέργειες τους θα πρέπει να αλλάξουν λόγω αυτών των παραγόντων. Επιπλέον, σε πιο ακραίες περιπτώσεις, κινδυνεύουν να «διαδώσουν» φήμες και παραπληροφόρηση ή να δημοσιεύσουν τις σκέψεις και τις απόψεις τους στο διαδίκτυο.

Με μία καλύτερη ματιά, μπορούμε να παρατηρήσουμε ένα κενό που αποκαλύπτει την ανάγκη διερεύνησης των ηθικών επιπτώσεων που σχετίζονται με τον προοδευτικό αυτοματισμό διαδικτυακών διαδικασιών συνομιλίας και την ενσωμάτωσή τους με τις αρχές της τεχνητής νοημοσύνης. Συνεπώς, ειδικά στην αγορά, οι άνθρωποι έχουν χρέος να έχουν επίγνωση της φύσης και των χαρακτηριστικών των chatbots, καθώς και να ενημερώνονται για τις επιπτώσεις της αλληλεπίδρασης μαζί τους. Τα τελευταία αποτελούν μείζον ζήτημα για τους χρήστες, καθώς με κάθε αλληλεπίδραση, παρέχονται στα chatbots πρόσθετες καταναλωτικές, προσωπικές ή/και ιδιωτικές πληροφορίες, οι οποίες με την σειρά τους εμπλουτίζουν τα αποθετήρια δεδομένων τους και μπορούν στην συνέχεια να χρησιμοποιηθούν με τρόπο που να ωφελεί τα ίδια και τους εκπροσώπους τους. Ωστόσο, οι συνομιλίες δεν αποτελούν μόνο μέσα μεταφοράς δεδομένων, αλλά και στρατηγικές διαδικασίες που παίζουν καθοριστικό ρόλο στην δημιουργία σχέσεων εμπιστοσύνης, στην μείωση των αβεβαιοτήτων και στην αύξηση της γνώσης στην διαδικασία αλληλεπίδρασης με τον χρήστη.

Ένας τρόπος εγγύησης εξατομικευμένων πληροφοριών για την ικανοποίηση των αναγκών και προσδοκιών των πελατών είναι η εκμετάλλευση των δυνατοτήτων των τεχνολογιών που βασίζονται σε κοινωνικά δίκτυα και τεχνητή νοημοσύνη. Όπως αναφέρουν οι Vargo και Akaka (2009)[105], ο κύριος στόχος των συστημάτων υπηρεσιών είναι «να παρέχουν εισροή στις διαδικασίες δημιουργίας αξίας άλλων συστημάτων υπηρεσιών και έτσι να αποκτήσουν αμοιβαία απόδοση». Πρόκειται λοιπόν για συστήματα υπηρεσιών, η συμβολή των οποίων είναι η συλλογή, η ενσωμάτωση και ο μετασχηματισμός δεδομένων παρέχοντας πολύτιμες αλληλεπιδραστικές εμπειρίες, τόσο για τους πελάτες, όσο και για τις επιχειρήσεις.

Ωστόσο, οι μηχανές δεν διαθέτουν ανθρώπινες δυνατότητες σύνεσης ή ερμηνευτική ικανότητα να διορθώνουν λάθη και λανθασμένες αποφάσεις. Ως αποτέλεσμα της προσφοράς των chatbots για αυτοδιάγνωση, οι γιατροί μπορεί να αντιμετωπίσουν προβλήματα σχετικά με την εμπιστοσύνη που είχαν χτίσει με τους ασθενείς τους σε κάποια πιθανή προκαταρκτική λανθασμένη διάγνωση που προέρχεται από το chatbot. Αυτό το επίπεδο πειθούς και διαπραγματεύσεως αυξάνει τον φόρτο εργασίας των επαγγελματιών και δημιουργεί νέες εντάσεις μεταξύ ασθενών και ιατρών, καθώς και προκαλεί πιθανή βλάβη στον ιατρικό τομέα όσον αφορά την σωστή συνεργασία του με τους ασθενείς, σε περίπτωση που οι τελευταίοι αγνοήσουν την άποψη του γιατρού. Είναι σημαντικό λοιπόν, να τονιστεί ότι τα chatbots ή τα πιο προηγμένα συστήματα που βασίζονται στην τεχνητή νοημοσύνη συμπληρώνουν - και όχι αντικαθιστούν - ιατρικούς επαγγελματίες (Powell 2019)[77].

Από την πλευρά του πελάτη, όσον αφορά την αξία της χρήσης των chatbots, η υψηλή ταχύτητα, η ποσότητα και η ποικιλία πληροφοριών που μπορούν να παρέχουν σε σύγκριση με τους παραδοσιακούς αντιπροσώπους υπηρεσιών, είναι αδιαμφισβήτητα. Με την βοήθεια τεχνικών όπως αυτή της αντιστοίχισης μοτίβου, τα chatbots έχουν γρήγορη πρόσβαση σε ένα «σύνολο εννοιών που είναι αλληλένδετες μεταξύ τους» και μπορούν να χρησιμοποιήσουν αυτό το ευρύ φάσμα γνώσεων ώστε να παρέχουν γρήγορες και άμεσες υπηρεσίες στον χρήστη. Από την άλλη

πλευρά, οι εταιρείες μέσω της χρήσης των chatbots και της βοήθειας τεχνικών, όπως μηχανική μάθηση, αποκτούν πληροφορίες, μαθαίνουν και τροποποιούν τις δικές τους συμπεριφορές προκειμένου να βοηθήσουν τους πελάτες να χρησιμοποιούν τον χρόνο τους πιο αποτελεσματικά και να τους καθοδηγούν να αναπτύξουν μια βαθύτερη κατανόηση των προϊόντων με θετικές συνέπειες, όσον αφορά τις οικονομικές επιδόσεις.

Έχοντας διαπιστώσει ότι υπάρχουν ηθικά στοιχεία στην χρήση των chatbots, μπορεί να αναφερθεί επίσης ότι υπάρχουν τρία βασικά τα οποία βρίσκονται παρακάτω:

1. Το πρώτο αφορά τον κίνδυνο της ασύμμετρης ανακατανομής της ισχύος. Έχοντας ασυμμετρία πληροφοριών στις αλληλεπιδράσεις ανθρώπου-μηχανής, τα chatbots καλούνται να αυξήσουν την ικανότητα των οργανισμών να συλλέγουν εμπεριστατωμένες, ολοκληρωμένες και συγκεντρωτικές πληροφορίες για τους πελάτες τους, προκειμένου να αποφευχθεί η ασύμμετρη κατανομή ισχύος, ακόμη και σε θέματα που δεν εμπίπτουν στο αντικείμενο της άμεσης συναλλαγής (Parvatiyar & Seth, 2001)[73]. Ωστόσο, δημιουργείται ανισορροπία στην πληροφοριακή ισχύ, καθώς τα chatbots διαχειρίζονται και επεξεργάζονται μεγάλο όγκο πληροφοριών μέσα σε σύντομο χρονικό διάστημα, καθώς να προβλέπουν παράλληλα πιθανές απαντήσεις. Ο ηθικός κίνδυνος στο συγκεκριμένο ζήτημα, βρίσκεται στην λανθασμένη ευθυγράμμιση των συμμετεχόντων που βρίσκονται σε διαφορετικό επίπεδο πληροφοριακής γνώσης. Όπως επισημαίνει ο West (2019)[110], τα chatbots και οι χρήστες αλληλεπιδρούν σε ένα «σύστημα στο οποίο η εμπορευματοποίηση των δεδομένων μας επιτρέπει μια ασύμμετρη ανακατανομή της δύναμης που σταθμίζεται στους φορείς που έχουν πρόσβαση και την ικανότητα να έχουν νόημα στην πληροφόρηση».
2. Η δεύτερη ηθική πρόκληση αφορά τις στρατηγικές διαχείρισης που θα ακολουθηθούν για τον εξανθρωπισμό των μηχανών, προκειμένου να αυξηθεί η αξιοπιστία και η εμπιστοσύνη των χρηστών προς τα chatbots, ενώ παράλληλα να μειωθεί η όποια αμφισβήτηση προς αυτά. Τα οφέλη που μπορούν να προκληθούν από τα παραπάνω για τους οργανισμούς, είναι αδιαμφισβήτητα. Συγκεκριμένα, αφορούν την ταχύτητα, το κόστος, την εξοικονόμηση και τα έσοδα (Wilson & Daugherty, 2018)[113]. Για το λόγο αυτό, ο εξανθρωπισμός τους αφορά ακόμη και αυτά που δεν χαρακτηρίζονται ακόμη από αυτογνωσία (Kaplan & Haenlein, 2019)[56], καθώς υπάρχει άμεση ανάγκη για μείωση της αντίληψης των χρηστών για τον κίνδυνο και την αύξηση της αξιοπιστίας των chatbots χρησιμοποιώντας πιο εξελιγμένα δεδομένα για τους χρήστες, όπως το εύρος των γνωστικών, συναισθηματικών και κοινωνικών ενδείξεων (Kaplan & Haenlein, 2019)[56]. Ο ηθικός κίνδυνος στο συγκεκριμένο ζήτημα είναι ότι εάν επιτευχθεί ο εξανθρωπισμός τους, οι καταναλωτές μπορεί να υποθέσουν ότι έχουν την ίδια νοημοσύνη, ηθικές αξίες, κρίση και κοινωνικά πρότυπα με τις μηχανές (Mengis & Eppler, 2008)[67].
3. Το τρίτο ηθικό ζήτημα αφορά την ανάγκη διαχείρισης της ασφάλειας του δικτύου και την προστασία του απορρήτου των χρηστών. Τα τελευταία χρόνια, έχει αυξηθεί η συλλογή προσωπικών και απρόσωπων δεδομένων, ώστε να συγκεκριμενοποιηθούν οι μεμονωμένες

συμπεριφορές εντός της ψηφιακής αγοράς, σε βαθμό που εγείρουν ανησυχίες σχετικά με συγκεκριμένες ηθικές ανησυχίες, όπως η ιδιωτικότητα των πληροφοριών, η προστασία δεδομένων, η έλλειψη ελέγχου των χρηστών πάνω στα προσωπικά τους δεδομένα και η πιθανή υποδούλωση στις τεχνολογικές συσκευές, δίχως την συνειδητή απόφασή τους. Κατά συνέπεια, η εφαρμογή μέτρων απορρήτου κρίνεται απαραίτητη, ώστε να διασφαλιστεί το απόρρητο, η ακεραιότητα και η διαφάνεια στην σχέση μεταξύ χρήστη και chatbot.

Τέλος, έχει παρατηρηθεί ότι εάν τα chatbots-σύμβουλοι υγείας θεωρούνται έμπιστα από τους χρήστες, οι τελευταίοι δείχνουν πρόθυμοι να εμπιστευτούν προσωπικά δεδομένα σχετικά με τις ιατρικές πληροφορίες τους (Pasquale 2020, σελ. 51)[74]. Επιπλέον, ο Palanica et al. (2019)[72] κατέληξε στο συμπέρασμα ότι η πλειοψηφία των γιατρών δεν εμπιστεύονται τα chatbots για μια αποτελεσματική φροντίδα των ασθενών τους και δεν θεωρούν πως μπορούν να αναπτύξουν οποιαδήποτε “ανθρώπινη πλευρά”. Επίσης, πολλοί γιατροί δήλωσαν ότι τα chatbots υγειονομικής περίθαλψης σχετίζονται με τον κίνδυνο οι ασθενείς να κάνουν αυτοδιάγνωση πολύ συχνά, να μην κατανοούν τις διαγνώσεις ή να μην αισθάνονται επαρκώς συνδεδεμένοι με τον κύριο γιατρό τους (Palanica et al. 2019)[72].[42][50]

2.3.2 Ασφάλεια Προσωπικών Δεδομένων των Χρηστών

Δυστυχώς, οι χρήστες ως επί το πλείστον δεν γνωρίζουν πώς αντιμετωπίζονται, χρησιμοποιούνται, αποθηκεύονται ή ακόμη και κοινοποιούνται οι ευαίσθητες προσωπικές τους πληροφορίες ταυτοποίησης (PII) όταν χρησιμοποιούν chatbots που συλλέγουν προσωπικά τους δεδομένα.

Η ασφάλειά τους στον συγκεκριμένο τομέα, μπορεί να χωριστεί σε δύο τομείς. Ο πρώτος τομέας αφορά την ασφάλεια των δεδομένων που μεταφέρονται μέσω μηνυμάτων, φωνής και εικόνων στον διακομιστή του chatbot. Ο δεύτερος τομέας αφορά τα δεδομένα που θα επεξεργαστούν, αποθηκευτούν ή ακόμη και θα καταλήξουν σε κοινή χρήση. Με την κάλυψη των δύο αυτών τομέων, μπορούν να αποφευχθούν αρκετές απειλές, ωστόσο υπάρχουν μέθοδοι για την περαιτέρω αύξηση της ασφάλειας αυτής.

Οι ακόλουθες μέθοδοι καλύπτουν κάθε επικοινωνία με το chatbot:

1. Έλεγχος ταυτότητας δύο παραγόντων για τα στοιχεία του χρήστη, για την διασφάλιση πως το σωστό άτομο έχει πρόσβαση στα σωστά δεδομένα και υπηρεσίες. Αυτό μπορεί να συμβεί με μέσα όπως email και μήνυμα κειμένου. Υπάρχουν επίσης υψηλότερα σχήματα ελέγχου ταυτότητας πολλαπλών παραγόντων, όμως δεν συνηθίζονται να χρησιμοποιούνται σε τέτοιου είδους συστήματα.
2. Η κρυπτογράφηση από άκρο σε άκρο (E2EE) είναι ένα σύστημα επικοινωνίας στο οποίο τα μέρη που επικοινωνούν είναι και αυτά που μπορούν να διαβάσουν τα μηνύματα. Με

αυτό τον τρόπο, επιτρέπεται στον παραλήπτη να είναι ο μοναδικός που θα μπορεί να αποκρυπτογραφήσει το μήνυμα. Δυστυχώς όμως, τα μεταφερόμενα δεδομένα μπορούν να παραποιηθούν από τρίτο μέρος και ως εκ τούτου, η διασφάλιση των εμπλεκόμενων μερών στην πρόσβαση των κρυπτογραφικών κλειδιών, αποτελεί μείζον ζήτημα για την συνομιλία.

3. Το άρθρο 32 του GDPR απαιτεί από τις εταιρείες να λαμβάνουν αυστηρά μέτρα όσον αφορά τα προσωπικά δεδομένα των χρηστών. Υπάρχει, ωστόσο, μία σύγκρουση «νόμος εναντίον απορρήτου» και οι εταιρείες σε ορισμένα κράτη δίνουν πρόσβαση στις κυβερνήσεις τους, έχοντας σχεδιάσει κρυπτογραφημένα τα προϊόντα και τις υπηρεσίες τους.

Επιπροσθέτως, τα δεδομένα του χρήστη αν και ιδιωτικά, πρέπει να αποθηκεύονται ώστε να μπορέσουν να αποτρέψουν πιθανές απειλές ασφαλείας. Οι ιστορικές συνομιλίες αποτελούν μία αποθηκευμένη συνομιλία, η οποία θα χρησιμοποιηθεί για ανάλυση και προστασία με τις περισσότερες εταιρείες να αποθηκεύουν την επικοινωνία και να υλοποιούν ειδικές ενέργειες για την ασφαλή αποθήκευση δεδομένων, όταν αυτές χειρίζονται PII σύμφωνα με τον ορισμό του GDPR.

Σύμφωνα με το GDPR η επικοινωνία με ένα chatbot μπορεί να αποθηκευτεί μόνο εάν δεν υπάρχει σύνδεση, αλλιώς θα πρέπει να προστατεύεται. Στην πρώτη περίπτωση, η βάση δεδομένων περιέχει μόνο την επικοινωνία και σε περίπτωση που τα δεδομένα δεν προσδιορίζουν ένα άτομο, μπορεί να μοιραστεί με την ομάδα ανάπτυξης. Ωστόσο, η προσβασιμότητα σε πολλές πηγές δεδομένων από τις εταιρίες, αποτελεί δυνητικό κίνδυνο ασφάλειας στα chatbots, θέτοντας έτσι την διαμόρφωση του ελέγχου χρήστη και των δικαιωμάτων πρόσβασης του απαραίτητα. Αντίθετα, στην περίπτωση των ανοιχτών domain chatbots, οι πάροχοι τους αποθαρρύνουν τους χρήστες να συνομιλούν για προσωπικά στοιχεία, καθώς τα συγκεκριμένα chatbots είναι ως επί το πλείστον για διασκέδαση και δεν έχουν ακόμη την δυνατότητα κάποιας σοβαρής συζήτησης για οποιοδήποτε θέμα. Για παράδειγμα, ο παρακάτω πίνακας συνοψίζει τις πληροφορίες που μοιράζονται με τρίτα μέρη τρία δημοφιλή chatbots - Replika, Kuki και CleverBot.

Stored data /chatbots	Replika	Kuki	CleverBot
Name	S*		S
Other profile information birthday, gender, work status, and so forth	N		N
Email address, phone number	S*	S	
Facts about you and your life, hobbies and interests	N		
People mentioned in the chat	N		
Images sent, voice messages	S**		
Text chat	N	S**	S ('starred')
Usage data—button clicks, search queries, and similar	S	S	S
Cookies	S	S	S
Backend infrastructure	Azure, AWS	AWS	UK company servers
E2EE	Standart SSL	HTTPS	

Εικόνα 6. Σημείωση: Το "S" χρησιμοποιείται για κοινή χρήση και "N" για μη κοινόχρηστο. Τα στοιχεία που επισημαίνονται με «*» είναι υποχρεωτικά και «**» είναι στοιχεία που μοιράζονται χωρίς κανένα ΠΠ. Η σημείωση "με αστέρι" σημαίνει ότι μπορούν να κοινοποιηθούν δημοφιλή μηνύματα κειμένου. Τα κενά κελιά δεν καθορίζονται στην πολιτική απορρήτου.

Τέλος, αρκετές εταιρείες διατηρούν τα δεδομένα του χρήστη για όσο το δυνατόν μεγαλύτερο χρονικό διάστημα, ανάλογα με τις ανάγκες τους. Σύμφωνα με τον GDPR οποιαδήποτε δεδομένα μπορούν να οδηγήσουν στην έμμεση ή άμεση αναγνώριση του χρήστη, αποτελούν προσωπικά, καθώς υπάρχουν και ειδικές κατηγορίες προσωπικών δεδομένων στο (π.χ. εθνικότητα ή σεξουαλικός προσανατολισμός). Σύμφωνα με τα παραπάνω, ένα chatbot δεν είναι απαραίτητο να σέβεται τέτοιες πτυχές της συνομιλίας, καθώς θα καθιστούσε δύσκολη την απάντηση δίχως τέτοιου είδους πληροφορίες. Ωστόσο, μια λύση που θα μπορούσε να δοθεί, είναι τα πρότυπα συνομιλίας, τα οποία βασίζονται σε άτομα και παρέχουν συνεπείς απαντήσεις.[64]

3

Αλγόριθμος Επεξεργασίας Φυσικής Γλώσσας BERT

3.1 Μοντέλα Επεξεργασίας Φυσικής Γλώσσας Κειμένου

Τον τελευταίο καιρό, από την άνοδο του AI, συγκεκριμένα επεξεργασία φυσικής γλώσσας (NLP) και υπολογιστική γλωσσολογία, τα chatbots γίνονται σύγχρονα τεχνολογική πτυχή στην επόμενη γενιά συνομιλιών. Τα chatbots είναι τα συστήματα διαλόγου ανθρώπου-υπολογιστή που αλληλεπιδρούν με χρήστες σε φυσική γλώσσα. Η χρήση της φυσικής γλώσσας παρέχει έναν διαισθητικό, ευέλικτο και εύκολο τρόπο στην αλληλεπίδραση ανθρώπου-μηχανής. Το πρόβλημα με την ικανότητα του AI να αντικατοπτρίζει την ανθρώπινη φυσική γλώσσα είναι ότι δεν έχει καμία πραγματική κατανόηση της γλώσσας που επεξεργάζεται, αλλά μπορεί να αναγνωρίσει μοτίβα και χαρακτηριστικά κειμένου που βασίζονται στην επαναλαμβανόμενη τροφοδοσία της δοκιμής και στην ταξινόμηση μιας λέξης και μιας πρότασης σε ένα σχήμα σχεδιασμένο από τον χρήστη.

Η αναγνώριση προτύπων έχει εξελιχθεί σε βαθιά εκμάθηση προτύπων που αναγνωρίζονται στο πλαίσιο της εκπαίδευσης που έχει πραγματοποιήσει ένας χρήστης. Μοντέλα βαθιάς μάθησης έχουν διαπρέψει σε εργασίες επεξεργασίας φυσικής γλώσσας (NLP) όπως η απάντηση ερωτήσεων από το κείμενο τα τελευταία χρόνια. Η ευρεία επιτυχία τους οδήγησε σε γρήγορη υιοθέτηση στον πραγματικό κόσμο μέσω βιομηχανικών εφαρμογών. Ωστόσο, αυτά τα μοντέλα εξαρτώνται από την αναγνώριση και πρόβλεψη προτύπων τεχνικών που συχνά αποτυγχάνουν να ορίσουν σωστά τα συμφραζόμενα κομμάτια εισόδου, ειδικά με ανοιχτού τύπου εισροές που βασίζονται περισσότερο στην φυσική ομιλία παρά σε μια εύκολα επεξεργασμένη συντακτική απάντηση. Κατά την ανάλυση εισαγωγής του χρήστη, τα νευρωνικά δίκτυα καθορίζουν την σημασία κάθε λέξης που αφορά την άλλη και συνοψίζονται σε μοτίβα που αναγνώρισε το δίκτυο και θα συνεχίσει να παρατηρεί σε παρακάτω εισαγωγή. Το ζήτημα που προκύπτει από αυτό είναι όταν το δίκτυο αναγνωρίζει και χρησιμοποιεί μοτίβα που είναι συχνά περιττά και μερικές φορές ακόμη και επιζήμια για τον σκοπό της τεχνητής νοημοσύνης.

Σημαντικά ελαττώματα της μάθησης καθιστούν εξαιρετικά δύσκολη την χρήση σε περιβάλλον αλληλεπίδρασης που έχει υψηλές απαιτήσεις, όπως αυτό της ψυχικής υγείας, καθώς η ασφάλεια του ασθενούς είναι υψίστης σημασίας και, κατά συνέπεια, δεν υπάρχει περιθώριο σφάλματος στο συγχρονισμό της εισόδου χρήστη και διατύπωσης των απαντήσεων. Ωστόσο, η

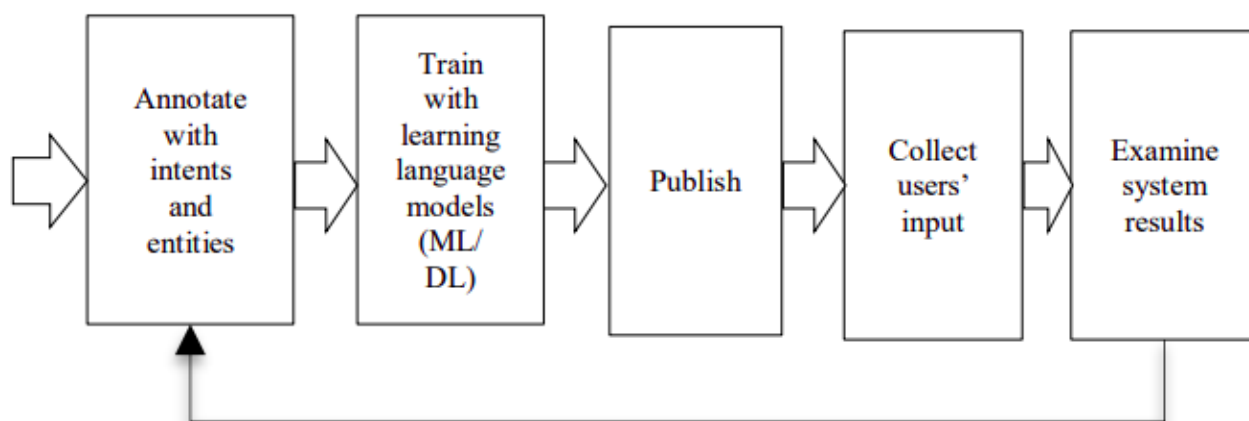
τεχνητή νοημοσύνη υγείας απουσιάζει από την βαθιά μάθηση και καθιστά την αλληλεπίδραση των ασθενών μονότονη και στερείται οποιασδήποτε εξατομικευμένης εμπειρίας. Αυτή η μορφή αλληλεπίδρασης των ασθενών αντικατοπτρίζει περισσότερο μια ερώτηση και απάντηση με προκαθορισμένα δεδομένα τα οποία ο ασθενής μπορεί να θεωρήσει άχρηστα ή πιθανότατα θα μπορούσε απλώς να τα βρει μόνος του στο διαδίκτυο. Επιπλέον, η βαθιά μάθηση είναι χρήσιμη για την αναγνώριση μοτίβων σε δεδομένα που δε μπορούν να υποβληθούν σε συμβολική επεξεργασία βασίζεται μόνο στην γλωσσολογία και το προκαθορισμένο σχήμα. Επομένως, βαθιά μάθηση σε συνδυασμό με συμβολική επεξεργασία ΑΙ θα έδινε στους χρήστες την καλύτερη δυνατή εξατομικευμένη εμπειρία με συνομιλία ΑΙ.

Τα γλωσσικά μοντέλα δημιουργούν μια κατανομή πιθανότητας σε μια ακολουθία μαρκών, δεδομένης μιας εισόδου και έναν αλγόριθμο εκμάθησης. Υπάρχουν επίσης γλωσσικά μοντέλα που εκπαιδεύονται χρησιμοποιώντας τον μηχανισμό αυτοεκτίμησης. Κάθε διακριτικό της πρότασης εισόδου μαθαίνει να προσέχει το ένα διακριτικό του άλλου στην πρόταση εισόδου ενώ παράγει την απόκριση εξόδου. Έτσι, ουσιαστικά αναπαράγονται προτάσεις που βασίζονται σε επαναλαμβανόμενα μοτίβα συσχέτισης μεταξύ των μαρκών. Η ευχέρεια συνομιλίας ενός chatbot ποικίλλει σημαντικά με την ποσότητα της εκπαίδευσης που είχε και τις εισόδους με τις οποίες έχει εκπαιδευτεί. Τα Chatbots εκπαιδεύονται να εργάζονται σε έναν συγκεκριμένο τομέα της γνώσης έχουν αναπτύξει περισσότερη αναγνώριση προτύπων προς αυτόν τον τομέα, ενώ γενικά αφήνοντας έξω διάφορες γνώσεις που μπορεί να μην είναι χρήσιμες σε αυτό το περιβάλλον συνομιλίας.

Τα Chatbots παρέχουν απαντήσεις ή εκτελούν ενέργειες από την επεξεργασία κειμένου ή ομιλίας, που έστειλαν οι άνθρωποι ψηφιακά και μπορούν να μιμούνται στενά τις ανθρώπινες συνομιλίες μέσω τεχνικών NLP και ανάλυση συναισθημάτων. Η φυσική κατανόηση της γλώσσας (NLU) είναι ο κινητήρας και ο πυρήνας της εργασίας NLP ενός chatbot. Είναι μια διαδικασία υλοποίησης φυσικών διεπαφών χρήστη. Το NLU επιτρέπει την αναγνώριση για υποκείμενες προθέσεις, εξαγωγή πλαισίου, σημασιών και συγκεκριμένων τομέων σχετικά με οντότητες στις εκφράσεις των χρηστών. Τα βήματα που κάνει ένα chatbot όταν έχει συγκεκριμένες προθέσεις ενεργοποιούνται από τις δηλώσεις χρηστών είναι οι ενέργειες και έχουν παραμέτρους που αναφέρουν λεπτομερείς πληροφορίες σχετικά με αυτό. Τρεις όψεις σε μια πρόταση προσδιορίζονται συνήθως από το NLU, συγκεκριμένα, πρόθεση που χαρτογραφεί την έκφραση των χρηστών σε μια συγκεκριμένη κλάση που επιτρέπει στους ψηφιακούς εικονικούς βοηθούς για να αποφασίσουν για μια απάντηση ή ενέργεια, οντότητες που απεικονίζουν σημαντικές πληροφορίες όπως ημερομηνία, ώρες και τοποθεσίες και τα συμφραζόμενα αντιστοιχούν στο πλαίσιο του αντικειμένου που είναι ο χρήστης.

Το NLU είναι ένα υποσύνολο NLP και συνομιλίας ΑΙ. Είναι μια μηχανή κατανόησης φυσικής γλώσσας που ταξινομεί την έκφραση κατά προθέσεις και εξαγωγή σχετικών πληροφοριών από δηλώσεις που ονομάζονται οντότητες. Ο κύριος ρόλος της NLU είναι να προσπαθήσουμε να κατανοήσουμε την είσοδο του χρήστη και την πρόθεση της συνομιλίας και να εξαγάγουμε οντότητες που απαιτούνται για την εκτέλεση μιας ενέργειας χρήστη ή επιχειρησιακής λογικής. Μπορούμε να σκεφτούμε τις υπηρεσίες NLU ως σύνολο API υψηλού επιπέδου για την δημιουργία του δικού μας αναλυτή γλώσσας χρησιμοποιώντας υπάρχοντα NLU και Βιβλιοθήκες μηχανικής μάθησης. Βοηθά τους υπολογιστές να κατανοήσουν ανθρώπινη γλώσσα κατανοώντας,

αναλύοντας και ερμηνεύοντας βασικό μήνυμα ή λόγομερη. Εκπαιδεύεται με φυσικές εκφράσεις χρηστών που επισημαίνονται με οντότητες και διαχέονται με την χρήση συνωνύμων. Η έννοια της χρήσης της τεχνητής νοημοσύνης και της ανθρώπινης γλώσσας στην επικοινωνία με μηχανές δεν είναι καινούργια. Ωστόσο, οι ερευνητές είχαν υποτιμήσει την πολυπλοκότητα των ανθρώπινων γλωσσών για χρόνια, ακόμη και δεκαετίες. Ακόμα και στις μέρες μας, ενώ πρόσφατα έχουν προχωρήσει στην τεχνητή νοημοσύνη και το NLP, έχουν διεξαχθεί αρκετές μελέτες που δείχνουν ότι οι χρήστες δεν μπορούν ξεγελαστούν και πάντα έχουν στο μυαλό τους ότι επικοινωνούν με έναν υπολογιστή. Πρόσφατα, τα chatbots με γνώμονα την ΑΙ έχουν χρησιμοποιηθεί για διάφορες εργασίες από διαφορετικούς τομείς εφαρμογών, όπως ιατρική και αυτοδιαγνωστικά και υποστήριξη εξυπηρέτησης φοιτητών πανεπιστημίου. Στην εικόνα που παρατίθεται παρακάτω, μπορούμε να δούμε και την διαδικασία υλοποίησης με χρήση NLP.



Εικόνα 7. Διαδικασία Υλοποίησης με χρήση NLP.

Γενικά, τα chatbots ταξινομούνται σε δύο μεγάλες ομάδες που ονομάζονται προσανατολισμένη στην εργασία και συνομιλητική χωρίς εργασία. Τα chatbots που δεν προσανατολίζονται σε εργασίες είναι τα συνομιλιακά που αφορούν τους χρήστες απαντώντας στις ερωτήσεις τους. Αυτά τα chatbots βασίζονται στο NLP για εξαγωγή πληροφοριών χρήστη και αντίδραση με υψηλή αντιστοίχιση απόκριση και μπορεί να ρυθμιστεί είτε να έχει μικρότερη είτε μεγαλύτερη συνομιλίες. Πιο πρόσφατες εργασίες σχετικά με την εφαρμογή του chatbot έχουν χωρίσει τα chatbots σε τέσσερις ομάδες, συγκεκριμένα, βάσει στόχων, που βασίζονται στην γνώση, βασίζονται στις υπηρεσίες και βασίζονται στην ανταπόκριση. Αυτές οι κατηγορίες εμπίπτουν επίσης στην προηγούμενη ταξινόμηση των εργασιών και chatbots που δεν προσανατολίζονται στην εργασία. Τα chatbots που βασίζονται σε στόχους βασίζονται στο στόχο για επίτευξη που αποτελείται από δραστηριότητες, συζητήσεις και ενημερωτικούς τύπους. Τα chatbots της γνώσης βασίζονται σε γνώσεις που αντλούν από τις πηγές δεδομένων ή τα σύνολα δεδομένων που έχουν εκπαιδευτεί για να περιλαμβάνουν ανοικτό και κλειστό τομέα ή συγκεκριμένο τομέα. Τα chatbots που βασίζονται σε υπηρεσίες βασίζονται στις ευκολίες που παρέχονται στους χρήστες, συμπεριλαμβανομένων των διαπροσωπικών, ενδοπροσωπικών και

διαπροσωπικών σκοπών. Τέλος,,υπάρχουν δύο τύποι μοντέλων που επικρατούν στην ανάπτυξη chatbot με βάση την τεχνητή νοημοσύνη:

1. Μοντέλα που βασίζονται στην ανάκτηση: αυτά χρησιμοποιούν ένα αποθετήριο προκαθορισμένων απαντήσεων ή χρησιμοποιούν ορισμένους τύπους ευρετικής για να επιλέξουν την κατάλληλη απάντηση με βάση την έρευνα και το πλαίσιο.Μπορούν να δώσουν πιο αξιόπιστες και πιο γραμματικά ακριβείς απαντήσεις. Είναι πιο εύκολο να διδαχθούν καθώς απαιτούν λιγότερες πληροφορίες, αλλά δεν είναι σε θέση να ανταποκριθούν σε ερωτήσεις πέρα από την βάση γνώσεων τους.
2. Μοντέλα παραγωγής: αυτά δεν βασίζονται σε προκαθορισμένες απαντήσεις. Δημιουργούν νέες απαντήσεις από την αρχή καθώς μπορούν να απαντήσουν σε διαφορετικές ερωτήσεις. Αυτά τα chatbots γίνονται πιο έξυπνα με την πάροδο του χρόνου και μπορούν να μάθουν από προηγούμενες ερωτήσεις, απαντήσεις και συζητήσεις. Ωστόσο, είναι δύσκολο να εκπαιδευτούν καθώς απαιτούν μεγάλο ποσοστό των δεδομένων και τα μοντέλα ανάκτησης συχνά απολαμβάνουν καλύτερο έλεγχο της ποιότητας απόκρισης από ό,τι γεννητικά μοντέλα.

Επιπλέον, τα chatbots ως επιπλέον πτυχή που σχετίζεται με τα δικαιώματα που παρέχει η ανάπτυξη της πλατφόρμας καθώς και το στοιχείο του chatbot στο οποίο προτιμούμε να προσαρμοζόμαστε στην περίπτωση της χρήσης μας. Τέτοιες πλατφόρμες εμπίπτουν σε δύο κατηγορίες, συγκεκριμένα,ανοιχτού κώδικα(όπως το RASA), ή κλειστές ή ιδιόκτητες πλατφόρμες, τυπικά από τις μεγάλες εταιρείες και τεχνολογικούς γίγαντες που είναι κυρίως μαύρα κουτιά και εκθέτουν μόνο μερικές από τις πτυχές τους στην υλοποίηση προκειμένου να ενσωματώσουμε στην ανάπτυξη το περιβάλλον μας. Ένα μεγάλο πλεονέκτημα με τις υπηρεσίες σε μεγάλες εταιρείες είναι ο μεγάλος όγκος δεδομένων για τα οποία συλλέγουν συνεχή βελτίωση των υποκείμενων μοντέλων τους.

Σε ένα ανεξάρτητο chatbot NLU η αρχιτεκτονική ορίζεται από τρία κύρια μέρη:

- 3.2 Ένα γενικό μεταμοντέλο που ορίζει τις γενικές έννοιες της NLU και των σχέσεών τους,
- 3.2 Ένα γενικό μεταμοντέλο συγκεκριμένης NLU υπηρεσία (π.χ., Dialogflow, RASA) και
- 3.2 Ένα σύνολο κανόνων χαρτογράφησης που χαρτογραφούν κάθε έννοια συγκεκριμένης υπηρεσίας NLU στην έννοια του γενικού μεταμοντέλου.[82][99][31]

3.2 Αλγόριθμος BERT

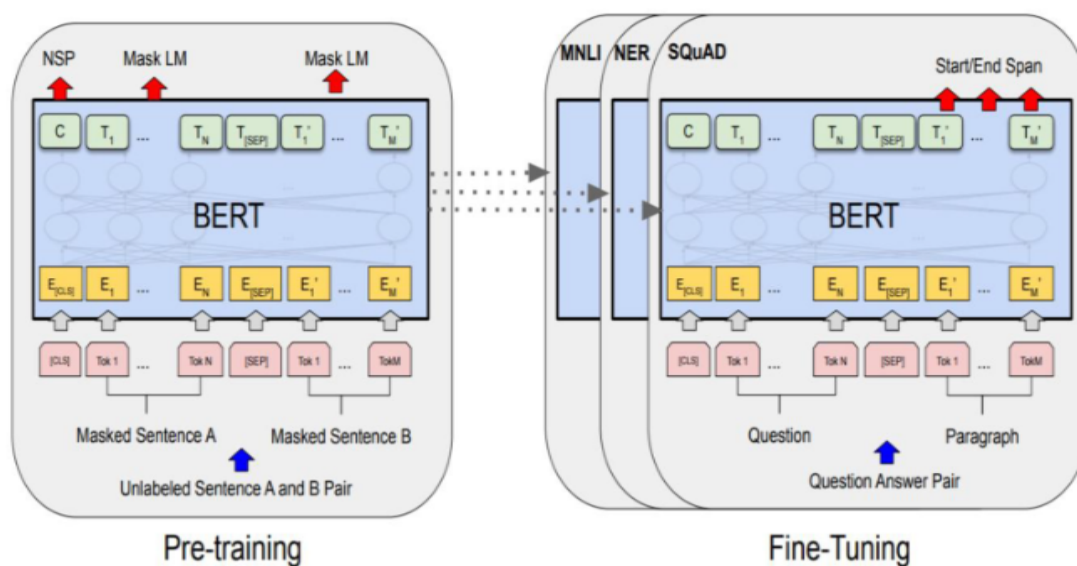
Από την εισαγωγή τους το 2017, τα Transformers (Vaswani et al., 2017) έχουν καταργήσει την επεξεργασία φυσικής γλώσσας NLP, προσφέροντας βελτιωμένη παραλληλοποίηση και καλύτερη μοντελοποίηση εξαρτήσεων μεγάλης εμβέλειας. Τα προεκπαιδευμένα γλωσσικά μοντέλα όπως το BERT (Bidirectional Encoder Representations from Transformers) έχουν αποδειχθεί εξαιρετικά αποτελεσματικά για εργασίες επεξεργασίας φυσικής γλώσσας (NLP). Πρόκειται για ένα σχετικά νέο μοντέλο αναπαράστασης γλώσσας, το οποίο έρχεται να φέρει μία νέα πρόταση με τα παλαιότερα μοντέλα αναπαράστασης γλώσσας (Peters et al., 2018a; [7] Radford et al., 2018). Επιπλέον, πρόκειται για ένα μοντέλο σχεδιασμένο για εκ των προτέρων εκπαίδευση, με βαθιές, αμφίδρομες αναπαραστάσεις κειμένου, χωρίς ετικέτες, με ρύθμιση πλαισίων από κοινού σε όλα τα επίπεδα. Ως αποτέλεσμα, ένα τέτοιου είδους μοντέλο, μπορεί να προσφέρει την βελτίωσή του, με μόλις ένα επιπλέον επίπεδο εξόδου, προσφέροντας έτσι μοντέλα τελευταίας τεχνολογίας τα οποία μπορούν να εφαρμοστούν σε διάφορους τομείς, χωρίς ουσιαστικές τροποποιήσεις αρχιτεκτονικής.

Ωστόσο, η προσφορά του συγκεκριμένου μοντέλου, δεν σταματάει εκεί. Μέσω της προεκπαίδευσης του γλωσσικού μοντέλου, μπορεί να επιτευχθεί σημαντική βελτίωση μοντέλων επεξεργασίας φυσικών γλωσσών. Τα συγκεκριμένα μοντέλα, περιλαμβάνουν εργασίες όπως π.χ. συμπέρασμα φυσικής γλώσσας (Bowman et al., 2015[87]; Williams et al., 2018[3]) και παράφραση (Dolan και Brockett, 2005[87]), που στοχεύουν στην πρόβλεψη των σχέσεων μεταξύ προτάσεων, αναλύοντάς τις ολιστικά, καθώς και εργασίες σε συμβολικό επίπεδο -όπως αναγνώριση ονομαστικής οντότητας και απάντηση ερωτήσεων- όπου παράγεται έξοδος με συμβολική γλώσσα (Tjong Kim Sang και De Meulder, 2003; [11] Rajpurkar et al., 2016[78]).

Όσον αφορά τον τρόπο λειτουργίας του, το BERT χρησιμοποιεί “μονή κατεύθυνση” με γλώσσα κάλυψης (“masked language model” (MLM)), με σκοπό την κάλυψη ορισμένων τυχαίων διακριτικών από την είσοδο, ώστε να προβλέψει το αρχικό αναγνωριστικό λεξιλογίου της μάσκας της εκάστοτε λέξης, που βασίζεται μόνο στο περιεχόμενό της. Σε αντίθεση με τα μοντέλα που χρησιμοποιούν προεκπαίδευση γλωσσικού μοντέλου “αριστερά-δεξιά”, όπου ο στόχος MLM επιτρέπει στην αναπαράσταση να συγχωνεύσει το αριστερό και το δεξί πλαίσιο, που μας επιτρέπει να εκπαιδεύσουμε εκ των προτέρων έναν βαθύ αμφίδρομο μετασχηματιστή.[18]

Επιπλέον, έχει την δυνατότητα για εκ των προτέρων εκπαίδευση με βαθιές αμφίδρομες αναπαραστάσεις. Ουσιαστικά, πρόκειται για μια στοίβα στρωμάτων κωδικοποιητή Transformer που αποτελούνται από πολλαπλές «κεφαλές» αυτο-προσοχής.[86] Το κείμενο που λαμβάνει είναι χωρίς ετικέτα-σε κάθε επίπεδο- διαμορφώνοντας από κοινού το περιβάλλον αριστερά και δεξιά. Μπορεί επιπλέον να ρυθμιστεί με ακρίβεια, με την βοήθεια ενός επιπλέον επιπέδου εξόδου, το οποίο έχει σκοπό την δημιουργία ενός μοντέλου προηγμένης τεχνολογίας για την ανάληψη εργασιών NLP, για εργασίες όπως η απάντηση ερωτήσεων και η ταξινόμηση. [49] Πιο συγκεκριμένα, χρησιμοποιεί μετασχηματιστές με πολλαπλά στρώματα, όπου κάθε στρώμα περιέχει πολλαπλές κεφαλές προσοχής, με κάθε κεφαλή προσοχής να διαθέτει την ικανότητα να μπορεί να δεχτεί ως είσοδο μια ακολουθία διανυσμάτων $h = [h_1, \dots, h_n]$, η οποία αντιστοιχεί στα n

διακριτικά της πρότασης εισόδου. Έπειτα με την σειρά του, κάθε διάνυσμα h_i μετατρέπεται σε διάνυσμα ερωτήματος-κλειδιού και τιμές q_i , k_i , v_i , μέσω χωριστών γραμμικών μετασχηματισμών. Ο τρόπος που μία κεφαλή υπολογίζει τα βάρη προσοχής μεταξύ των ζευγών-λέξεων είναι η κανονικοποίηση με softmax κουκκίδες του ερωτήματος και των διανυσμάτων/κλειδιών, δημιουργώντας μία έξοδο με ένα σταθμισμένο άθροισμα των διανυσμάτων τιμών[111], με τις εξόδους όλων των κεφαλών που βρίσκονται στο ίδιο στρώμα να συνδυάζονται και να διατρέχουν ένα πλήρως συνδεδεμένο στρώμα, όπου κάθε στρώμα τυλίγεται με μια σύνδεση παράλειψης και ακολουθείται από κανονικοποίηση του στρώματος.[86]



Εικόνα 8. Pre-training και Fine-tuning μέθοδοι.[49]

Η συμβατική ροή εργασίας για το BERT αποτελείται από δύο στάδια: προ-εκπαίδευση και τελειοποίηση. Η προεκπαίδευση χρησιμοποιεί δύο αυτο-εποπτευόμενες εργασίες: μοντελοποίηση καλυμμένης γλώσσας (MLM) και πρόβλεψη επόμενης πρότασης (NSP) με τις αναπαραστάσεις εισόδου υπολογίζονται ως εξής: κάθε λέξη στην είσοδο αρχικά μετατρέπεται σε λεκτικά κομμάτια (Wu et al., 2016[120]) και στην συνέχεια τρία επίπεδα ενσωμάτωσης (κουπόνι, θέση και τμήμα) συνδυάζονται για να ληφθεί ένα διάνυσμα σταθερού μήκους[86]. Πιο συγκεκριμένα, για κάθε διακριτικό εισόδου σε μια ακολουθία, κάθε κεφαλή υπολογίζει τα διανύσματα κλειδιού, τιμές και ερωτήματος, που χρησιμοποιούνται για την δημιουργία μιας σταθμισμένης αναπαράστασης. Οι έξοδοι όλων των κεφαλών στο ίδιο στρώμα συνδυάζονται και διατρέχουν ένα πλήρως συνδεδεμένο στρώμα με το κάθε στρώμα τυλίγεται με μια σύνδεση παράλειψης και να ακολουθεί η κανονικοποίηση του στρώματος.

Όσον αφορά την προεκπαίδευσή του, είναι προεκπαιδευμένο σε 3,3 δισεκατομμύρια tokens αγγλικού κειμένου για να εκτελέσει δύο εργασίες, την «μοντελοποίηση μάσκας γλώσσας» και την «πρόβλεψη επόμενης πρότασης». Στην πρώτη περίπτωση, το μοντέλο προβλέπει τις ταυτότητες των λέξεων που έχουν καλυφθεί από το εισαγόμενο κείμενο, ενώ στην δεύτερη

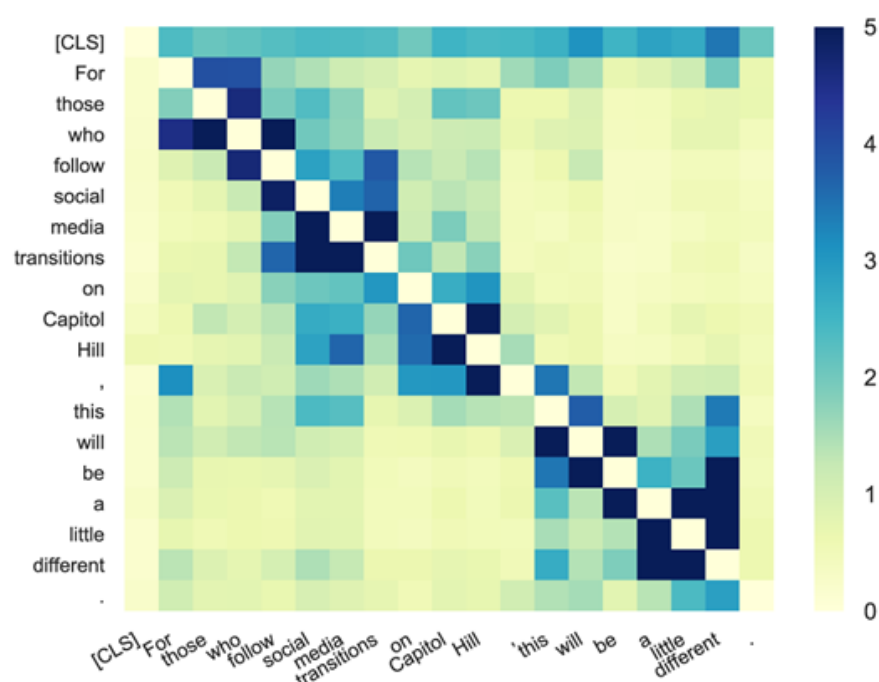
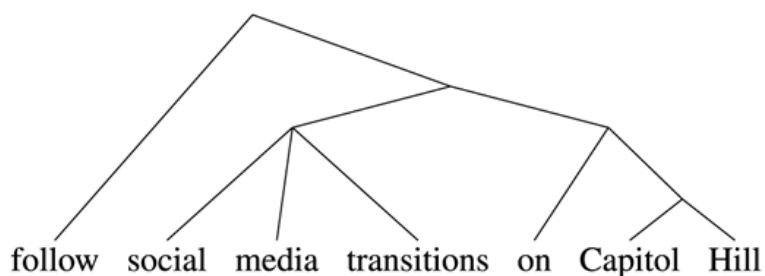
προβλέπει εάν το δεύτερο μισό της εισαγωγής ακολουθεί το πρώτο μισό της εισαγωγής στο κείμενο ή είναι μια τυχαία παράγραφος. Επιπλέον, άξιο αναφοράς αποτελεί ότι οι κεφαλές προσοχής του Bert δίνουν μεγάλη προσοχή είτε στις σταθερές μετατοπίσεις θέσης, είτε σε ευρέως ολόκληρη πρόταση. Ένα εκπληκτικά μεγάλο μέρος της προσοχής του BERT εστιάζει στο ειδικό διακριτικό [CLS] που προστίθεται στην αρχή του κειμένου και σε ένα άλλο διακριτικό [SEP] προστίθεται είτε στο τέλος, είτε έπειτα από κάθε κείμενο, εάν η είσοδος αποτελείται από πολλά ξεχωριστά κείμενα. Τέλος, ενώ καμία μεμονωμένη κεφαλή δεν έχει καλή απόδοση σε πολλές σχέσεις, διαπιστώνεται ότι οι κεφαλές με συγκεκριμένους σκοπούς, αντιστοιχούν εξαιρετικά καλά σε αντίστοιχες συγκεκριμένες σχέσεις. [111]

Τέλος, αν και είναι σαφές ότι το BERT λειτουργεί εξαιρετικά καλά, είναι λιγότερο σαφές το γιατί, γεγονός που περιορίζει την περαιτέρω βελτίωση της αρχιτεκτονικής βάσει υποθέσεων. Παραλλαγές του BERT μπορούμε να βρούμε στο Google1 και στο HuggingFace (Wolf et al., 2020), συμπεριλαμβανομένων των αρχικών εκδόσεων του - "base" και "large"- οι οποίες διαφέρουν ως προς τον αριθμό των κεφαλών, επιπέδων και το μέγεθος κρυφής κατάστασης.[86]

3.2.1 Συντακτική γνώση Μοντέλου

Έρευνες έδειξαν πως οι αναπαραστάσεις BERT είναι ιεραρχικές και όχι γραμμικές. Πιο συγκεκριμένα, απέδειξαν ότι το BERT μπορεί να διαθέσει κάτι παρόμοιο με την δενδρική δομή, χωρίς να διαθέτει πληροφορίες σχετικά με την σειρά των εκάστοτε λέξεων [121]. Επιπλέον, οι ενσωματώσεις BERT δοκιμάστηκαν στο εάν μπορούν να διαθέτουν δυνατότητα κωδικοποίησης πληροφοριών σχετικά με μέρη του λόγου και συντακτικά κομμάτια [69], προσφέροντας έτσι την δυνατότητα γνώσης συντακτικών πληροφοριών σχετικά με κοινές ενσωματώσεις διακριτικών, με σκοπό την δημιουργία συντακτικών δέντρων. Ωστόσο, στο συγκεκριμένο πείραμα, οι ταξινομητές ανίχνευσης δεν κατάφεραν να ανακτήσουν τις ετικέτες των απομακρυσμένων γονικών κόμβων στο συντακτικό δέντρο, καθιστώντας έτσι αδύνατη την συγκεκριμένη δυνατότητα [47].

Όσον αφορά τον τρόπο αναπαράστασης της προαναφερθείσας σύνταξης, διεξήχθησαν αρκετές έρευνες σχετικά με το θέμα, ωστόσο δεν κατάφεραν να δημιουργήσουν μία πλήρη ανάλυση δέντρων για τον BERT [76]. Ωστόσο, υπήρξε δυνατή η απεικόνιση (αν και σύντομη) ενός δέντρου εξάρτησης, το οποίο μπορεί να εξαχθεί απευθείας χρησιμοποιώντας τα βάρη αυτο-προσοχής, χωρίς όμως να καταφέρουν να επιτεύξουν κάποια αξιολόγηση ποσοτική αξιολόγηση. Σχετικά με πιο πρόσφατες έρευνες, υπήρξε μία πρόταση σχετικά με την προσέγγιση της παραπάνω μεθόδου, χωρίς την χρήση παραμέτρων. Η συγκεκριμένη μέθοδος, έχει την δυνατότητα να αξιοποιεί εις βάθος τη μέτρηση του αντίκτυπου που θα μπορούσε να έχει μία λέξη σχετικά με την πρόβλεψη της επόμενης, μέσα σε μια ακολουθία στην εργασία MLM (Εικόνα 9), καταλήγοντας έτσι στο συμπέρασμα ότι ο BERT μαθαίνει «φυσικά» κάποιες συντακτικές πληροφορίες.



Εικόνα 9. Ανιχνευτής χωρίς παραμέτρους για συντακτική γνώση: λέξεις που μοιράζονται συντακτικά υπο δέντρα έχουν μεγαλύτερο αντίκτυπο το ένα στο άλλο στην πρόβλεψη MLM. [123]

Τέλος, οι ανιχνευτές συμπλήρωσης του κενού του MLM έδειξαν πως ο BERT έχει την δυνατότητα, κατά την διάρκεια εκτέλεσης της εργασίας κλεισίματος, να λαμβάνει υπόψη την συμφωνία υποκειμένου-κατηγορήματος, ακόμη και εάν η πρόταση δεν έχει ιδιαίτερο νόημα ή διαθέτει ρήτρες που αποσπούν την προσοχή μεταξύ τους. Σε μία έρευνα σχετικά με στοιχεία αρνητικής πολικότητας (NPIs), απέδειξε πως ο BERT έχει πλεονέκτημα όσον αφορά την ανίχνευση παρουσίας NPIs. Ωστόσο, μία άλλη έρευνα ήρθε να αποδείξει πως οι παραπάνω ισχυρισμοί δεν ισχύουν. Πιο συγκεκριμένα, το γεγονός ότι ο BERT δεν «καταλαβαίνει» την άρνηση και δεν είναι ευαίσθητος σε κακώς μορφοποιημένη εισαγωγή, διαψεύδει το γεγονός πως η

ανίχνευση παρουσίας NPIs αποτελεί πλεονέκτημα για τον BERT. Πιο αναλυτικά, οι προβλέψεις του δε μπορούν να αλλοιωθούν, ακόμη και αν ανακατευτεί η σειρά των λέξεων, εάν περιέχει περικομμένες προτάσεις, εάν αφαιρέσουμε τα υποκείμενα ή τα αντικείμενα. Το συγκεκριμένο γεγονός, θα μπορούσε να σημαίνει ότι είτε η συντακτική γνώση του BERT είναι ελλιπής, είτε δεν χρειάζεται να βασίζεται σε αυτήν για την επίλυση των καθηκόντων του, με το τελευταίο να φαίνεται πιο πιθανό, καθώς έχει παρατηρηθεί πως ένα ενδιαμέσο βήμα λεπτομέρειας με εποπτευόμενη ανάλυση δεν έχει μεγάλη διαφορά για την απόδοση των εργασιών [12].

3.2.2 Σημασιολογική Γνώση

Έχει παρατηρηθεί ότι ο BERT εμφανίζει μία προτίμηση σχετικά με τα λανθασμένα συμπληρώματα για σημασιολογικούς ρόλους, τα οποία σχετίζονται σημασιολογικά με τους σωστούς σημασιολογικούς ρόλους, σε αντίθεση με τους όρους οι οποίοι δεν σχετίζονται. Επιπλέον, υπήρξε έρευνα σχετικά με την κωδικοποίηση πληροφοριών του BERT, σχετικά με τύπους οντοτήτων, σχέσεις και σημασιολογικούς ρόλους, λόγω του γεγονότος ότι τέτοιου είδους πληροφορίες μπορούν να ανιχνευθούν με ταξινομητές ανίχνευσης.

Όσον αφορά τους αριθμούς, παρατηρήθηκε μία έλλειψη διαχείρισης αριθμών κινητής υποδιαστολής εκ μέρους του BERT. Το συγκεκριμένο γεγονός το απέδειξαν εργασίες πρόσθεσης και αποκωδικοποίησης, κατά τις οποίες ο BERT απέτυχε να γενικεύσει τα δεδομένα εκπαίδευσης, καθώς οι αριθμοί παρόμοιων τιμών μπορούν να χωριστούν σε διαφορετικά κομμάτια λέξεων.

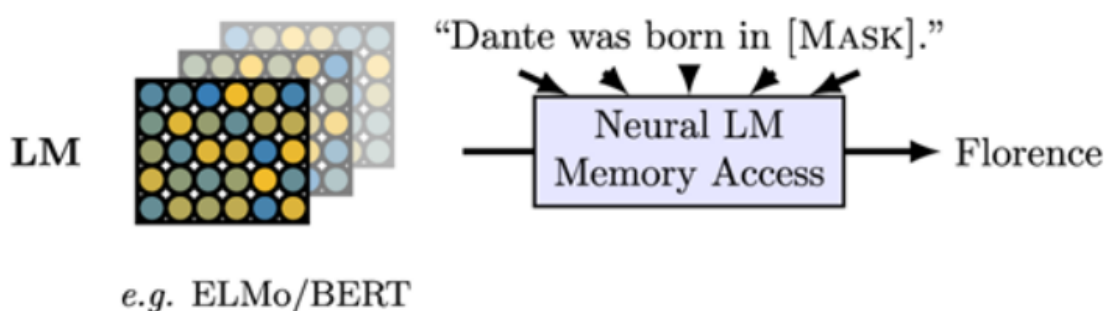
Τέλος, παρατηρήθηκε ότι ο BERT μπορεί να γίνει υπερβολικά εύθραυστος σε περίπτωση αντικατάστασης κάποιας επώνυμης οντότητας. Πιο συγκεκριμένα, σε περιπτώσεις όπως αυτή της αντικατάστασης ονομάτων σε εργασίες βασικής αναφοράς, αποδείχτηκε πως έχουν την ικανότητα να αλλάξουν μέχρι και το 85% των προβλέψεων, αποδεικνύοντας έτσι πως ακόμη και εάν υπάρχουν υψηλές βαθμολογίες F1 στις εργασίες ανίχνευσης NER, το μοντέλο δεν καταφέρνει να επιτύχει στον σχηματισμό μιας γενικής ιδέας για επώνυμες οντότητες. [12]

3.2.3 Παγκόσμια γνώση

Έχει παρατηρηθεί πως το BERT δυσκολεύεται σε περιπτώσεις που αφορούν ρεαλιστικά συμπεράσματα[9]. Πιο συγκεκριμένα, αυτές οι περιπτώσεις μπορεί να είναι περιπτώσεις όπου υπάρχουν αφηρημένα χαρακτηριστικά αντικειμένων και ιδιότητες οι οποίες δεν αναφέρονται, αλλά εννοούνται. Ωστόσο, υπάρχει δυνατότητα προσαρμογής στην επαγωγή γνώσης, λαμβάνοντας την βοήθεια του στοιχείου MLM και συμπληρώνοντας τα κενά, με τον Davison et al. (2019) [53] να υποστηρίζει ότι ο BERT έχει την δυνατότητα να γενικεύει καλύτερα σε μη ορατά δεδομένα.

Τέλος, έχει παρατηρηθεί πως ο BERT δεν έχει την δυνατότητα συλλογισμού με βάση τις παγκόσμιες γνώσεις που διαθέτει. Πιο συγκεκριμένα, έχει αποδειχθεί ο BERT έχει την

δυνατότητα να «μαντέψει» τις δυνατότητες και τις ιδιότητες αρκετών αντικειμένων, χωρίς όμως να μπορεί έχει την δυνατότητα να αιτιολογήσει την σχέση μεταξύ ιδιοτήτων και προσφορών (Forbes et al. (2019) [66]). Επιπλέον, έρευνες έχουν δείξει πως έδειξαν ότι η απόδοση του BERT μειώνεται αναλογικά με τον αριθμό των βημάτων τα οποία κρίνονται απαραίτητα για την δημιουργία συμπερασμάτων (Zhou et al. (2020)[116], Richardson και Sabharwal (2019) [58]). Ωστόσο, παρατηρείται ότι η επιτυχία του BERT, στηρίζεται στην εκμάθηση του αλγορίθμου με βάση ορισμένους στερεοτυπικούς συσχετισμούς, όπως για παράδειγμα πως ένα Ελληνικό όνομα χαρακτηρίζει έναν Έλληνα πολίτη (Poerner et al., 2019) [70]).



Εικόνα 10. BERT Παγκόσμια Γνώση

3.2.4 BERT Embeddings

Ο όρος Embedding - ή αλλιώς ενσωμάτωση - αναφέρεται στην έξοδο ενός στρώματος μετασχηματιστή, με τις περισσότερες φορές να αναφέρεται στο τελικό στρώμα. Σημαντικό αποτελεί το γεγονός ότι μπορούν να θεωρηθούν από την άποψη της αμοιβαίας μεγιστοποίησης πληροφοριών (Kong et al., 2019 [61]) οι συμβατικές στατικές ενσωματώσεις (Mikolov et al., 2013[97]) και κυρίως οι ενσωματώσεις τύπου BERT, οι οποίες ενσωματώνονται στο πλαίσιο. Στις συγκεκριμένες περιπτώσεις, κάθε διάνυσμα αντιπροσωπεύει και ένα διακριτικό διατηρώντας εξάρτηση από το συγκεκριμένο πλαίσιο εμφάνισης και περιέχοντας πληροφορίες σχετικές με το συγκεκριμένο πλαίσιο (Miaschi και Dell'Orletta, 2020 [6]).

Επιπροσθέτως, υπάρχουν αρκετές μελέτες οι οποίες αναφέρουν ότι σε περιπτώσεις όπου τα συμφραζόμενα μειώθηκαν όσο το δυνατόν περισσότερο, υπήρξαν ικανοποιητικά αποτελέσματα σχετικά με την κωδικοποίηση των λέξεων οι οποίες περιείχαν σημασιολογικές πληροφορίες, όπως για παράδειγμα περιπτώσεις που αφορούσαν ομοιότητα λέξεων. Διαδικασίες όπως αυτή, της απόσταξης μιας αναπαράστασης με βάση τα συμφραζόμενα, επιτυγχάνεται συγκεντρώνοντας τις πληροφορίες σε πολλαπλά περιβάλλοντα (Akbik et al., 2019 [5]; Bommasani et al., 2020 [85]), κωδικοποιώντας προτάσεις οι οποίες βασίζονται κυρίως στην σημασία μιας δεδομένης λέξης (

May et al., 2019 [23]) και συχνά εκπαιδεύοντας τις στατικές ενσωματώσεις (Wang et al., 2020d [119]).

Ωστόσο, αν και παρατηρούνται ορισμένες ελλείψεις και αρνητικές επιπτώσεις, υπάρχουν ακόμη περιθώρια βελτίωσης. Ο Ethayarajh (2019) [57] παρατήρησε ότι για να πετύχουμε περισσότερες αναπαραστάσεις με βάση το πλαίσιο πρέπει να επικεντρωθούμε στα μεταγενέστερα επίπεδα. Επιπλέον, παρατηρήθηκε ότι οι ενσωματώσεις του αλγορίθμου στον διανυσματικό χώρο, αναπαριστώνται ως ένας στενός κώνος, γεγονός που αυξάνεται από τα προγενέστερα προς τα μεταγενέστερα στρώματα. Για παράδειγμα, έστω δύο τυχαίες λέξεις για τις οποίες θέλουμε να ανακαλύψουμε την ομοιότητά τους. Οι λέξεις αυτές θα έχουν κατά μέσο όρο πολύ μεγαλύτερη ομοιότητα συνημιτόνου από την αναμενόμενη, εάν οι ενσωματώσεις είναι κατευθυντικά ομοιόμορφες (ή αλλιώς ισότροπες).

Τέλος, οι ενσωματώσεις του BERT, με σκοπό να αντιστοιχίσουν τις γενικεύσεις των λέξεων, σχηματίζουν συχνά διακριτές συστάδες (Wiedemann et al., 2019 [43] · Schmidt and Hofmann, 2020 [39]), γεγονός που καθιστά το BERT επιτυχημένο σε περιπτώσεις που προσπαθούμε να αποσαφηνίσουμε την έννοια κάποιας λέξης. Ωστόσο, οι Mickus et al. (2019) [96] πρόσθεσαν πως οι αναπαραστάσεις της ίδιας λέξης εξαρτώνται από την θέση της στην πρόταση στην οποία εμφανίζεται, πιθανότατα λόγω του στόχου NSP, γεγονός που δεν είναι επιθυμητό από γλωσσικής απόψεως.

3.3 Αλγοριθμικά Μοντέλα BERT

3.3.1 Roberta Model

Στην συγκεκριμένη ενότητα θα ασχοληθούμε με τρία μοντέλα του αλγορίθμου Bert. Το πρώτο μοντέλο που θα ασχοληθούμε είναι το Robustly Optimized BERT Pre-training Approach (RoBERTa) (Liu et al., 2019)[115]. Η συγκεκριμένη βελτιωμένη έκδοση του BERT, βελτιώνεται με βάση τα δημοσιευμένα αποτελέσματα BERT, τόσο στο GLUE, όσο και στο SQuAD. Έχει επιπλέον παρατηρηθεί πως υπάρχει δυνατότητα να επιτύχει την βαθμολογία 88,5 στο δημόσιο leaderboard GLUE, που αντιστοιχεί στο 88,4 που αναφέρεται από τους Yang et al. (2019)[122], όταν εκπαιδεύεται για μεγαλύτερο χρονικό διάστημα σε πρόσθετα δεδομένα, όπως δεδομένα των CC-NEWS [62].

Πρόκειται για μια επέκταση του μοντέλου Bert, η οποία γίνεται σε μεγαλύτερες παρτίδες και ακολουθίες, δίνοντας με αυτό τον τρόπο την δυνατότητα βελτίωσης της ακρίβειας του μοντέλου, σε σύγκριση με τον τρόπο αντιμετώπισης των δεδομένων του απλού BERT. Μία βασική διαφορά μεταξύ αυτών των δύο μοντέλων είναι η διαφορά τους στο δυναμικό μοτίβο κάλυψης του RoBERTa, το οποίο αντιγράφει τα δεδομένα εκπαίδευσης και πραγματοποιεί μια σειρά από στρατηγικές κάλυψης, διαφορετική για κάθε εκτέλεση, σε αντίθεση με το BERT, το οποίο χρησιμοποιεί μια στρατηγική στατικής κάλυψης, η οποία εκτελείται μόνο κατά την

προεπεξεργασία δεδομένων[71]. Επιπλέον, ένα σημαντικό πλεονέκτημα έναντι της στατικής κάλυψης, είναι το γεγονός ότι για κάθε εμφάνιση της ίδιας λέξης, πραγματοποιείται και η εκχώρηση ενός διαφορετικού διανύσματος ενσωμάτωσης, με βάση το πλαίσιο στο οποίο χρησιμοποιείται η λέξη.

Ο παρακάτω πίνακας περιγράφει τα αποτελέσματα αξιολόγησης συγκεκριμένων μοντέλων γλώσσας που βασίζονται σε Transformer κατά την επιστροφή της πιο παρόμοιας πρότασης.

Model	Soft F_1
bert-base-nli-cls-token	0.29
bert-base-nli-max-tokens	0.30
bert-base-nli-mean-tokens	0.31
bert-base-nli-stsb-mean-tokens	0.32
bert-base-wikipedia-sections-mean-tokens	0.25
bert-large-nli-cls-token	0.29
bert-large-nli-max-tokens	0.30
bert-large-nli-mean-tokens	0.30
	0.31
bert-large-nli-stsb-mean-tokens	
roberta-base-nli-mean-tokens	0.28
roberta-base-nli-stsb-mean-tokens	0.29
roberta-large-nli-mean-tokens	0.31
roberta-large-nli-stsb-mean-tokens	0.31

Εικόνα 11. Ονόματα όπως στην βιβλιοθήκη μετασχηματιστών: <https://github.com/huggingface/transformers> [29].

3.3.2 Bert-Large Model

Το BERT-large είναι ένας τύπος μοντέλου BERT. Το συγκεκριμένο μοντέλο (σε αντίθεση με το BERT-base) είναι αξιοσημείωτα καλύτερο στην διάκριση μεταξύ σχετικών αλλά διακριτών οντοτήτων ή εννοιών, ωστόσο, παρατηρείται ότι συχνά δυσκολεύεται να δώσει συγκεκριμένες απαντήσεις για ερωτήσεις που απαιτούν γενικές γνώσεις. Παρόμοια δυσκολία αντιμετωπίζει στο κομμάτι της μοντελοποίησης των αντωνυμιών, ειδικά στα πλαίσια μιας συζήτησης. Επιπλέον, η λειτουργία του Bert-large δεν διαφέρει ιδιαίτερα από το BERT-base μοντέλο, αλλά έχει μεγαλύτερο μέγεθος από την βάση BERT[92]. Επίσης, διαπιστώνεται πως το BERT-large επωφελείται από την χρήση μεγαλύτερων πλαισίων, όπως 384 λέξεων, σε αντίθεση με άλλα μοντέλα, όπως το BERT-base, το οποίο αποδίδει καλύτερα με τον μικρότερο αριθμό των 128 λέξεων. Ωστόσο, παρατηρείται ότι η απόδοσή του μειώνεται όταν αυξήσουμε κι άλλο τα πλαίσια (π.χ. 512) και πως σε περιπτώσεις όπως αυτή της παραλλαγής επικάλυψης, μία μέθοδος η οποία επεκτείνει τεχνητά το πλαίσιο πέρα από το 512, τα tokens δεν παρέχουν καμία βελτίωση. Το

συγκεκριμένο γεγονός, μας οδηγεί στο συμπέρασμα πως η χρήση μεγαλύτερων πλαισίων δε μπορεί να εγγυηθεί απαραίτητα την βελτίωση της εργασίας.[17]

Το BERT-large μπορεί να επιτύχει μία καλή απόδοση στο σύνολο των δεδομένων του, ωστόσο, πρόκειται για ένα σχετικά ακριβό μοντέλο, το οποίο έχει χρησιμοποιηθεί σε αρκετές σημαντικές έρευνες, όπως αυτή του Covid στο κοινωνικό μέσο δικτύωσης tweet, και μεταξύ όλων των αξιολογήσεων που πραγματοποιήθηκαν, το BERT-large είναι το δεύτερο σε μια κλίμακα ίση με 0,781 στο αντίστοιχο F1-score [71]. Επιπλέον, σε μία άλλη περίπτωση πολλαπλών πειραμάτων βαθιάς εκμάθησης και μεταφοράς προσεγγίσεων εκμάθησης για πρόσβαση σε συντακτικές ικανότητες [117], παρατηρήθηκε μία εξίσου εξαιρετική απόδοση σε μετασχηματιστές που βασίζονται στον αλγόριθμο Bert, σε σύγκριση με άλλες προσεγγίσεις [92]. Τέλος, αξιοσημείωτα αποτελούν και τα αποτελέσματα πολλών πειραμάτων, όπως αυτά που αναφέρθηκαν στην προηγούμενη ενότητα στην εικόνα 11 [29].

Επιπροσθέτως, γνωρίζουμε ότι το BERT πρόκειται για έναν αλγόριθμο ταξινόμησης εγγράφων, ο οποίος πραγματοποιεί μία διαδικασία μεγάλου υπολογιστικού κόστους. Στην λύση του συγκεκριμένου προβλήματος, σχετικά με το υπολογιστικό κόστος που σχετίζεται με το συγκεκριμένο αλγόριθμο, η χρήση του BERT large σε μικρό αμφίδρομο LSTM επιτυγχάνει ισοτιμία σε πολλαπλά σύνολα δεδομένων, χρησιμοποιώντας 30×λιγότερες παραμέτρους[35].

3.3.3 Xlm-Large Model

Το XLM πρόκειται για ένα πολύγλωσσο γλωσσικό μοντέλο Bert, προεκπαιδευμένο με δεδομένα από Βικιπαίδειες 15 διαφορετικών γλωσσών, συμπεριλαμβανομένης και της ελληνικής. Το συγκεκριμένο μοντέλο έχει την δυνατότητα παροχής διαγλωσσικών αναπαραστάσεων κειμένου γενικής χρήσης, παρέχοντας παράλληλα αποτελέσματα με παρόμοια λεπτομέρεια με αυτή των μονόγλωσσων γλωσσικών μοντέλων σε εργασίες ταξινόμησης στα αγγλικά. Επιπλέον, παρέχει μη εποπτευόμενους διαγλωσσικούς χώρους ενσωμάτωσης λέξεων με διαφορετικές ιδιότητες και υπερτερεί όσον αφορά την διαγλωσσική ομοιότητα των λέξεων. Είναι επίσης αξιοσημείωτο το γεγονός ότι τα ζεύγη μετάφρασης λέξεων είναι επίσης πολύ πιο κοντά στον διαγλωσσικό χώρο ενσωμάτωσης λέξεων, με το XLM να δίνει την ιδιαιτερότητα να εκπαιδεύονται μαζί με έναν κωδικοποιητή προτάσεων που μπορεί να επιβάλει εγγύτητα. Επιπροσθέτως, το μοντέλο μπορεί να εκπαιδευτεί ώστε να ευθυγραμμίζει τις αναπαραστάσεις μεταξύ των γλωσσών. Συγκεκριμένα, έχει εκπαιδευτεί σε δύο βοηθητικές εργασίες, το MLM και το Translation Language Modeling (TLM). Το TLM είναι μια επέκταση του MLM, όπου κάθε ζεύγος εκπαίδευσης περιέχει δύο προτάσεις, S1 και S2, από δύο διαφορετικές γλώσσες. Αναλυτικότερα, το S1 είναι η μετάφραση του S2 και όταν μια λέξη του S1 (ή S2) είναι καλυμμένη, τότε η αντίστοιχη μετάφραση αυτής της λέξης στο S2 (ή S1) δεν καλύπτεται.

Όσον αφορά την δική μας περίπτωση, χρησιμοποιούμε το xlm-R. Πρόκειται για ένα μοντέλο που εισήγαγε ο Conneau et al. [8], το οποίο βελτίωσε περαιτέρω τα αποτελέσματα του XLM, χωρίς να βασίζεται στο εποπτευόμενο TLM. Χρησιμοποιεί περισσότερα εκπαιδευτικά

δεδομένα από το Common Crawl και ένα μεγαλύτερο λεξιλόγιο 250.000 υπολέξεων, που καλύπτει 100 διαφορετικές γλώσσες. Όπως και στην προηγούμενη περίπτωση, υπάρχει μία κάλυψη του ύψους του 2% όσον αφορά την ελληνική γλώσσα (4.862 υπολέξεις, περίπου 2%). Παρακάτω, στην εικόνα 12, μπορούμε να παρατηρήσουμε ένα χαρακτηριστικό παράδειγμα της αναλογίας κατακερματισμού λέξεων, που μετράται ως το μέσος αριθμός διακριτικών υπολέξεων ανά λέξη, σε τρία τυχαία σύνολα δεδομένων για όλα τα μοντέλα γλώσσας που βασίζονται σε Transformer. Συμπεραίνουμε ότι τα πολύγλωσσα μοντέλα τείνουν να κατακερματίζουν τις λέξεις περισσότερο από το GREEK-BERT, ειδικά σε περιπτώσεις που διαθέτουν παραλλαγές M-BERT, των οποίων ο λόγος κατακερματισμού είναι περίπου διπλάσιος. Το XLM-R έχει χαμηλότερο λόγο κατακερματισμού από M-BERT, καθώς έχει τέσσερις φορές περισσότερες υπολέξεις που καλύπτουν την ελληνική. [55]

Model	GUDT (PoS)	NER	XNLI
M-BERT-UNCASED [10]	2.38	2.43	2.22
M-BERT-CASED [10]	2.58	2.65	2.40
XLM-R [7]	1.82	1.92	1.64
GREEK-BERT (ours)	1.35	1.33	1.23

Εικόνα 12. Η αναλογία κατακερματισμού λέξεων των μοντέλων που βασίζονται στο BERT υπολογίζεται με βάση τα δεδομένα ανάπτυξης όλων των συνόλων δεδομένων. Θεωρούμε πως οι χαμηλότερες αναλογίες είναι καλύτερες με τα καλύτερα αποτελέσματα να εμφανίζονται με έντονους χαρακτήρες.

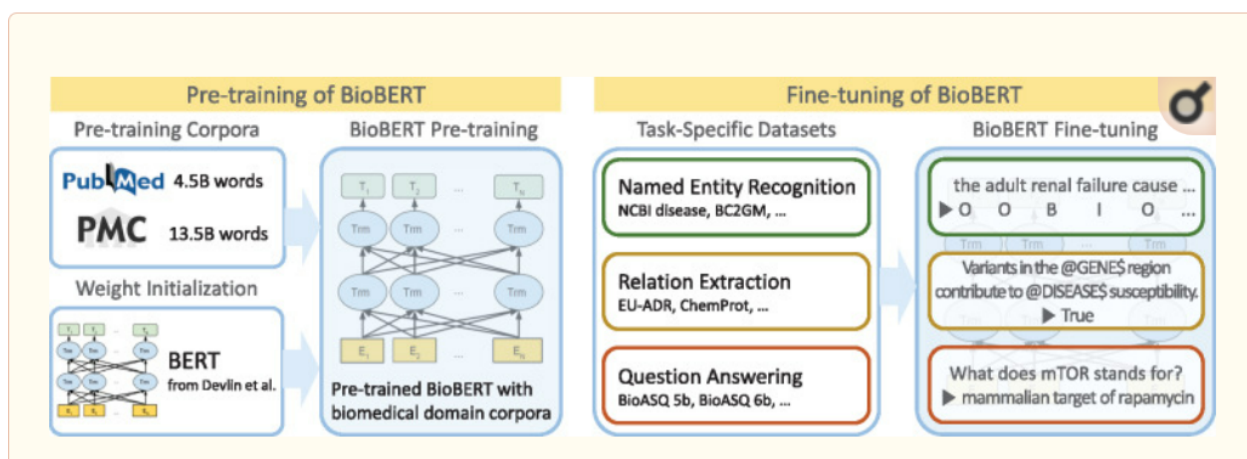
3.3.4 BioBert Model

Με την πρόοδον των τεχνικών βαθιάς εκμάθησης που υφίσταται η επεξεργασία φυσικής γλώσσας (NLP), επωφελείται η πρόοδος των βιοϊατρικών μοντέλων εξόρυξης κειμένου. Για παράδειγμα, τα Long Short-Term Memory (LSTM) και Conditional Random Field (CRF) έχουν σημειώσει αξιοσημείωτη βελτίωση στην απόδοση της αναγνώρισης οντοτήτων με όνομα στον βιοϊατρικό τομέα (NER) τα τελευταία χρόνια (Giorgi and Bader, 2018[98]· Habibi et al., 2017[32]· Wang et al., 2018[30]· Yoon et al., 2019[28]). Επιπλέον, αξίζει να σημειωθεί η βελτίωση των μοντέλων που βασίζονται σε βαθιά μάθηση σε εργασίες βιοϊατρικής εξόρυξης κειμένου όπως η εξαγωγή σχέσεων (RE) (Bhasuran and Natarajan, 2018[13]· Lim and Kang, 2018[24]) και η απάντηση ερωτήσεων (QA) (Wiese et al., 2017[112]).

Όσον αφορά τα μοντέλα με κατανομές λέξεων - γενικών και βιοϊατρικών δεδομένων-, οι οποίες είναι αρκετά διαφορετικές, μπορεί συχνά να αποτελούν πρόβλημα για τα βιοϊατρικά μοντέλα εξόρυξης κειμένου. Ως αποτέλεσμα, τα πρόσφατα μοντέλα στην βιοϊατρική εξόρυξη κειμένου βασίζονται σε μεγάλο βαθμό σε προσαρμοσμένες εκδόσεις αναπαραστάσεων λέξεων (Habibi et al., 2017[32]; Pyysalo et al., 2013[80]). Ενώ ο BERT έχει αποδείξει την

αποτελεσματικότητα των αναπαραστάσεων λέξεων με βάση τα συμφραζόμενα, δε μπορεί να επιτύχει υψηλή απόδοση σε βιοϊατρικά κείμενα, καθώς είναι προεκπαιδευμένο μόνο σε κείμενα γενικού τομέα.

Το BioBERT, πρόκειται για ένα προεκπαιδευμένο μοντέλο αναπαράστασης γλώσσας για τον βιοϊατρικό τομέα με αρχικοποίηση των βαρών του γίνεται με βάση τα βάρη του BERT. Ενώ το BERT είναι προεκπαιδευμένο σε κείμενο γενικού τομέα (αγγλική Wikipedia και BooksCorpus), το BioBERT εκπαιδεύεται σε κείμενα βιοϊατρικού τομέα (περιλήψεις PubMed και άρθρα πλήρους κειμένου PMC). Για το tokenization, το BioBERT χρησιμοποιεί το WordPiece tokenization (Wu et al., 2016[120]), το οποίο μετριάζει το πρόβλημα εκτός λεξιλογίου. Με το WordPiece tokenization, οποιεσδήποτε νέες λέξεις μπορούν να αντιπροσωπεύονται με συχνές υπολέξεις. [60]



Εικόνα 13. Η συνολική διαδικασία προεκπαίδευσης και λεπτομέρειας του BioBERT.[60]

4

Υλοποίηση Υποδομής Chatbot με Χρήση Βιοϊατρικής Πληροφορίας

4.1 Δομή Συστήματος Chatbot

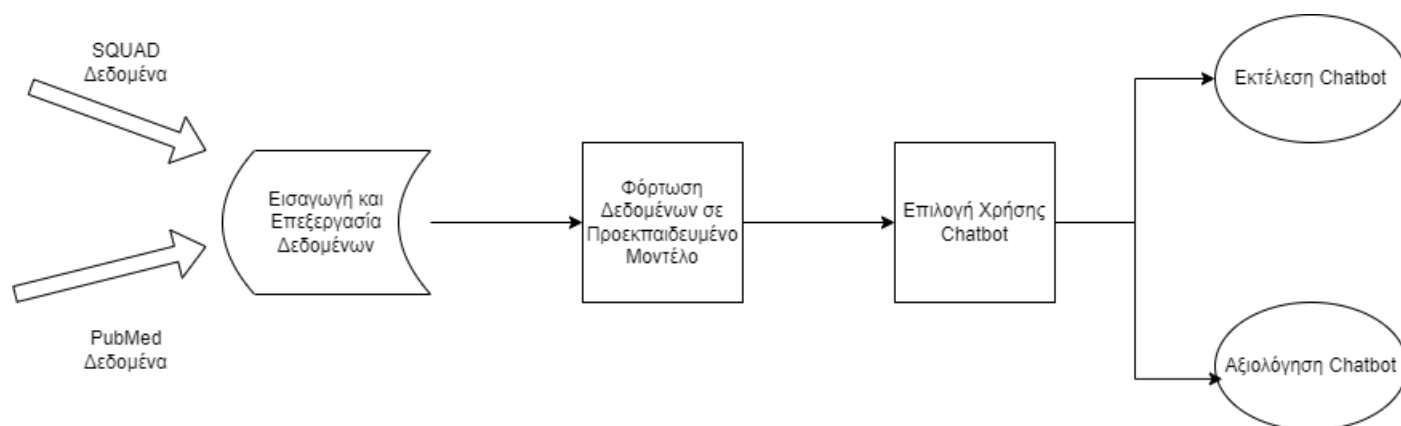
Αρχικά, για την εκτίμηση του chatbot μας, χρειάστηκε να δημιουργηθούν δύο εκδοχές του. Η μία αφορά την περίπτωση όπου το chatbot μας περιέχει τον αλγόριθμο BERT και η δεύτερη, η οποία είναι σχεδόν ίδια με την πρώτη, με την διαφορά ότι δεν περιέχει αλγόριθμο, αλλά τον CountVectorizer, ενώ διαθέτει και ορισμένες διαφορές σχετικά με το κομμάτι της υλοποίησης στο response. Περισσότερες λεπτομέρειες για την δεύτερη εκδοχή θα δοθούν στο Κεφάλαιο 4.3, όπου θα υπάρχουν διευκρινίσεις σε ορισμένα κομμάτια κώδικα και για τις δύο εκδοχές. Όσον αφορά τον βασικό μας κώδικα, τα δεδομένα μας αφού πρώτα εισαχθούν και επεξεργαστούν κατάλληλα, ώστε να αντιμετωπιστούν ξεχωριστά και όχι σαν παράγραφοι, φορτώνονται στο εκάστοτε μοντέλο (διαφορετικό κάθε φορά) και αποθηκεύονται με αυτό σε τοπικό περιβάλλον για μελλοντική χρήση. Στην συνέχεια το chatbot αφού ζητήσει το όνομα του χρήστη, ξεκινάει την συνομιλία μαζί του ώστε να μάθει την πρόθεσή του για run ή evaluation και να πράξει ανάλογα.

Σχετικά με την επιλογή run, το chatbot ζητά από τον χρήστη να καταχωρήσει μία πρόταση, εισάγοντας τα δεδομένα του και ζητώντας να λάβει μία απάντηση. Η απάντησή του chatbot εξαρτάται από το εάν τα δεδομένα του χρήστη αφορούν κάποια λέξη καλωσορίσματος, την πρόθεσή του για τερματισμό της συνομιλίας ή την πρόθεσή του για ερώτηση. Στην πρώτη περίπτωση, η απάντηση προς τον χρήστη θα δοθεί από κάποιες προκαθορισμένες προτάσεις, ενώ στην περίπτωση που ο χρήστης δείξει πρόθεση για τερματισμό, το chatbot θα τον αποχαιρετήσει και θα τερματίσει. Όσον αφορά την περίπτωση της ερώτησης, τα δεδομένα εισόδου του χρήστη θα εισαχθούν στο μοντέλο, ώστε να πραγματοποιηθεί η σύγκρισή τους με τα ήδη υπάρχοντα δεδομένα και να δοθεί απάντηση προς τον χρήστη - ανάλογα με το εάν βρέθηκαν δεδομένα που ικανοποιούν το εκάστοτε ποσοστό κατωφλίου ή όχι - και να συνεχιστεί η συνομιλία.

Στην περίπτωση του evaluation, η διαδικασία απάντησης παραμένει η ίδια, με την διαφορά ότι μόλις δοθεί η ερώτηση από τον χρήστη, το chatbot ζητάει από τον χρήστη να εισάγει την πιθανώς σωστή απάντηση που θα έπρεπε να δοθεί από αυτό. Στην συνέχεια, το chatbot συγκρίνει την απάντησή του με αυτή του χρήστη και εκτυπώνει την απάντησή του, μαζί με τα ποσοστά επιτυχίας ή αποτυχίας της εύρεσής του. Όσον αφορά τα παραπάνω ποσοστά, καλούμε την

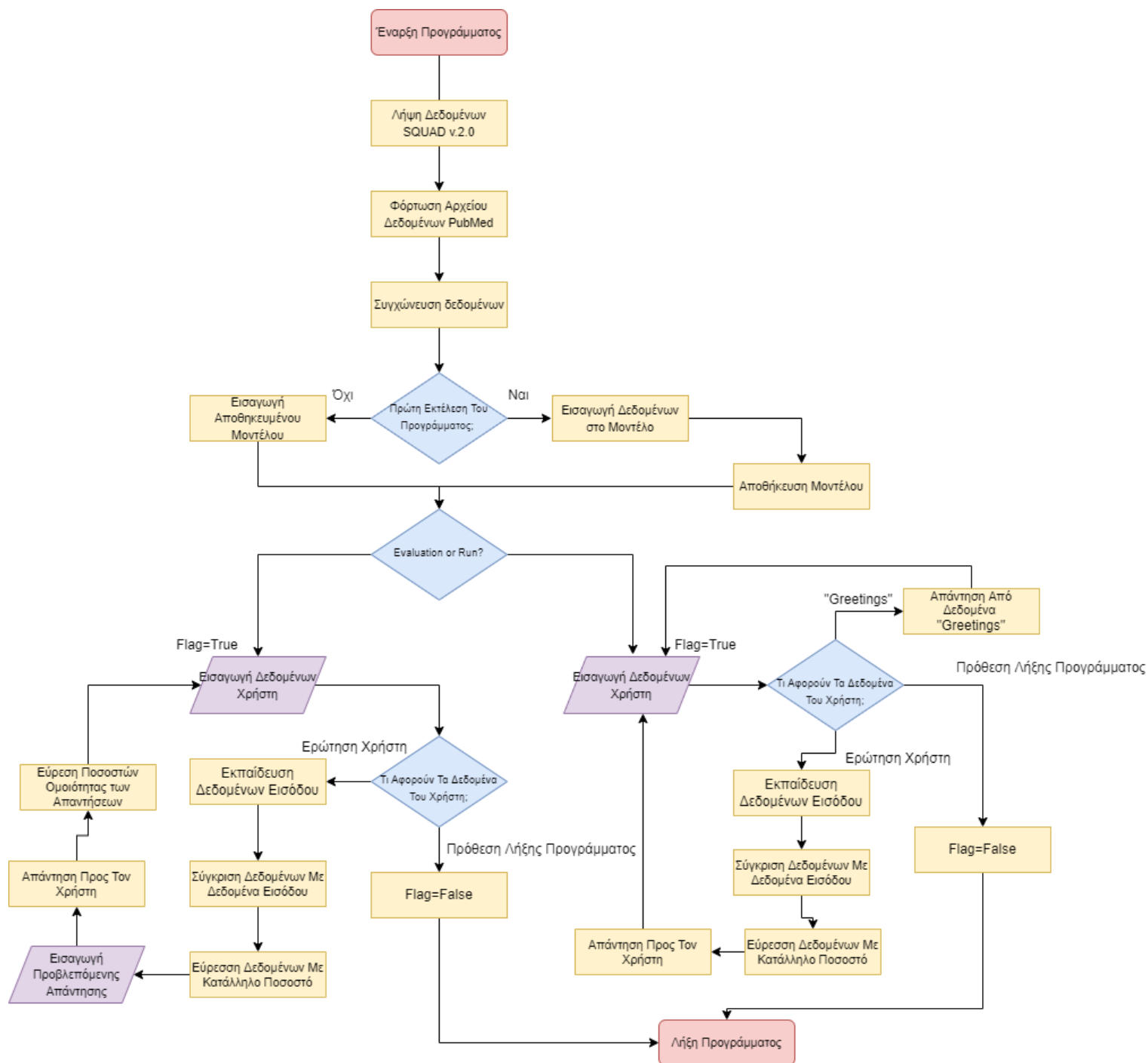
compute_f1, στην είσοδό της οποίας αναθέτουμε τις προτάσεις που θέλουμε να συγκρίνουμε, δηλαδή την απάντηση του chatbot (answer) και την απάντηση που θεωρούμε εμείς σωστή (prediction) και τις συγκρίνουμε ώστε να πάρουμε τις μετρικές Precision, Recall και F1.

Παρακάτω μπορούμε να δούμε την γενική δομή του συστήματός μας στην εικόνα 14:



Εικόνα 14. Διάγραμμα υλοποίησης του Chatbot μας.

Τέλος, περισσότερες λεπτομέρειες σχετικά με τον τρόπο αξιολόγησης, θα δοθούν στο Κεφάλαιο 5, ενώ παρακάτω μπορούμε να δούμε το διάγραμμα ροής του κώδικά μας με περισσότερες λεπτομέρειες:



Εικόνα 15. Διάγραμμα ροής του κώδικα.

Όσον αφορά τον τύπο του chatbot μας, επιλέξαμε να ακολουθήσουμε τα chatbots τύπου QA, τα οποία αφορούν συστήματα ανάκτησης πληροφοριών που στοχεύουν στην απάντηση για ερωτήματα φυσικής γλώσσας με επιστροφή απαντήσεων σχετικών με τα δεδομένα εισόδου, με την δυνατότητα απάντησης σε ερωτήσεις που ανήκουν σε ένα κοινό θέμα. Τα chatbots τα οποία βασίζονται σε ανάκτηση δεδομένων, λειτουργούν με βάση την αρχή των γραφημάτων ή των κατευθυνόμενων ροών και είναι εκπαιδευμένα να παρέχουν την καλύτερη δυνατή απάντηση από μια βάση δεδομένων προκαθορισμένων απαντήσεων. Οι απαντήσεις βασίζονται σε υπάρχουσες πληροφορίες με τα chatbots να χρησιμοποιούν τεχνικές όπως αντιστοίχιση λέξεων-κλειδιών, για να προσδιορίσουν την πιο κατάλληλη απόκριση. Ανεξάρτητα από την τεχνική, αυτά τα chatbots παρέχουν μόνο προκαθορισμένες απαντήσεις και δεν παράγουν νέα έξοδο.

Σε αντίθεση με τα QA chatbots, τα Generative Chatbots έχουν την δυνατότητα να χρησιμοποιούν μεθόδους παραγωγής που μπορούν να δημιουργήσουν νέους διαλόγους που βασίζονται σε μεγάλες ποσότητες δεδομένων εκπαίδευσης συνομιλίας. Πιο συγκεκριμένα, χρησιμοποιούν έναν συνδυασμό εποπτευόμενης μάθησης, μάθησης χωρίς επίβλεψη, ενισχυτικής μάθησης και μάθησης αντιπάλου για εκπαίδευση σε πολλά βήματα. Η εποπτευόμενη μάθηση δομεί μια συζήτηση ως πρόβλημα αλληλουχίας σε αλληλουχία, ενώ η εκμάθηση αλληλουχίας σε ακολουθία αντιστοιχίζει τις εισροές των χρηστών σε μια απόκριση που δημιουργείται από υπολογιστή. Ωστόσο, αυτός ο τύπος μάθησης έχει την τάση να δίνει προτεραιότητα σε απαντήσεις υψηλής πιθανότητας όπως «I know». Τα εποπτευόμενα συστήματα μάθησης έχουν επίσης πρόβλημα να συμπεριλάβουν σωστά ουσιαστικά στην ομιλία τους, καθώς εμφανίζονται λιγότερο στον διάλογο σε σύγκριση με άλλες λέξεις. Ως αποτέλεσμα, τα εποπτευόμενα chatbot εκμάθησης ακούγονται επαναλαμβανόμενα και δε μπορούν να προωθήσουν σταθερή ανθρώπινη συνομιλία.

Τέλος, ο λόγος που επιλέξαμε για την συγκεκριμένη διπλωματική εργασία τον τύπο QA chatbot έναντι άλλων τύπων, όπως αυτόν του Generative Chatbot, είναι το είδος των δεδομένων που οδηγούμε στην είσοδό μας. Πιο συγκεκριμένα, τα δεδομένα μας αφορούν εμπεριστατωμένα ιατρικά κείμενα με ακριβείς ορισμούς και αριθμητικά δεδομένα. Συνεπώς, σκοπός μας ήταν η πιο δυνατή ακριβής απάντηση εκ μέρους του chatbot και όχι η δημιουργία προτάσεων με πιθανό κίνδυνο ανάμειξης δεδομένων ή παραγωγής λανθασμένων ιατρικών προτάσεων.

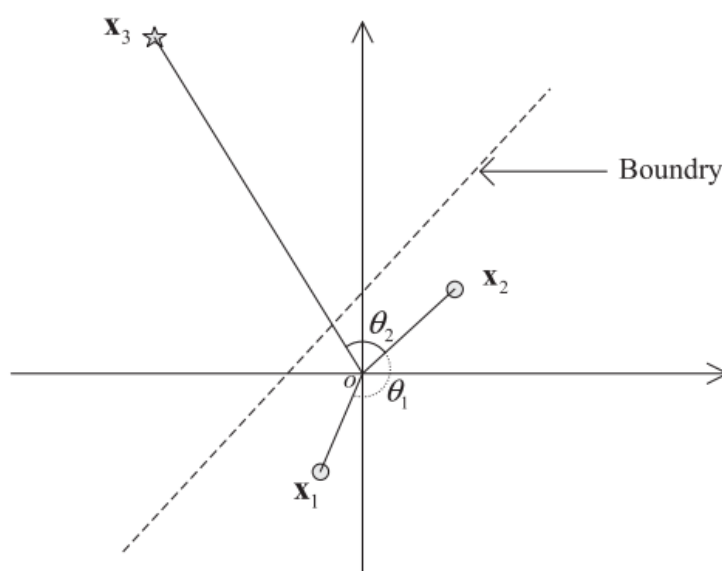
4.2 Μέθοδος Σύγκρισης Δεδομένων Εισόδου

Η ομοιότητα είναι ένα θεμελιώδες ζήτημα στις εργασίες ταξινόμησης και ομαδοποίησης. Ωστόσο, αν και η έννοια της ομοιότητας σχετίζεται με την έννοια της απόστασης, οι δύο αυτές έννοιες δεν είναι ακριβώς οι ίδιες. Στην έννοια της ομοιότητας πραγματοποιείται μέτρηση των κοινών χαρακτηριστικών μεταξύ δύο περιπτώσεων, ενώ στην έννοια της απόστασης αναδεικνύονται οι διαφορές μεταξύ τους. Ωστόσο, μπορεί πρώτα να υπολογιστεί η απόσταση μεταξύ δύο

περιπτώσεων και μετά να οριστεί ένα κατάλληλο όριο για να βγει το πόρισμα εάν αυτές οι δύο περιπτώσεις είναι παρόμοιες ή όχι.

Η ομοιότητα δύο περιπτώσεων εξαρτάται από το πόσο μειώνεται η απόσταση μεταξύ τους. Η Ευκλείδεια απόσταση θεωρείται συνήθως το απλούστερο μέτρο ομοιότητας σε πολλές εργασίες μηχανικής μάθησης και εξόρυξης δεδομένων. Ωστόσο, αυτή η μέτρηση συχνά αποτυγχάνει να δημιουργήσει διακριτικές αναπαραστάσεις, λόγω της υψηλής ευαισθησίας που διαθέτει στα μεγέθη. Σε αντίθεση με την Ευκλείδεια απόσταση, η ομοιότητα συνημιτόνου είναι μια άλλη μετρική που χρησιμοποιείται συνήθως, η οποία μετρά την ομοιότητα έως την γωνία μεταξύ δύο διανυσμάτων και μπορεί να επιτύχει την βέλτιστη απόδοση, καθώς δεν είναι ευαίσθητο στα μεγέθη. Σε αντίθεση με τα μοτίβα που θεωρούνται λιγότερο όμοια καθώς αυξάνεται η Ευκλείδεια απόσταση μεταξύ τους, στην περίπτωση του συνημιτόνου θεωρούνται περισσότερο παρόμοια καθώς αυξάνεται η ομοιότητα μεταξύ τους με συνημίτονο. Ωστόσο, ούτε αυτή η μέθοδος είναι πάντα συμφέρουσα, καθώς παρατηρείται ότι είναι πιθανό δύο μοτίβα με πολύ διαφορετικές τιμές χαρακτηριστικών να έχουν πολύ υψηλό μέτρο ομοιότητας.

Λύση στο συγκεκριμένο πρόβλημα έρχεται να δώσει μια προσαρμοσμένη μέτρηση ομοιότητας συνημιτόνου, η οποία μπορεί να διορθώσει εύκολα αυτό το μειονέκτημα λαμβάνοντας υπόψη τις διαφορετικές κλίμακες μεταξύ των δύο μοτίβων και αφαιρώντας τον αντίστοιχο μέσο όρο από κάθε σχέδιο με βασικό προσανατολισμό ως προς τις μετρήσεις ομοιότητας συνημιτόνου επικεντρώνονται στους προσανατολισμούς. Ωστόσο, η κατανομή δεδομένων είναι συχνά άγνωστη σε πραγματικά προβλήματα, καθώς παρατηρείται ότι όταν τα σχέδια είναι σε πολύ πυκνή κατανομή, οι γωνίες μεταξύ τους μπορεί να είναι πολύ μικρές, με αποτέλεσμα ακόμη και αν δύο μοτίβα είναι ανόμοια, ένας ταξινομητής που βασίζεται στην ομοιότητα συνημιτόνου να είναι πολύ πιθανό να τα ταξινομήσει εσφαλμένα ως παρόμοια. [75]



Εικόνα 16. Ένα παράδειγμα του cosine similarity, όπου το x_2 είναι παρόμοιο του x_1 , και διάφορο του x_3 . Το θ_1 είναι η περιλαμβανόμενη γωνία μεταξύ των x_1 και x_2 , και το θ_2 είναι η περιλαμβανόμενη γωνία μεταξύ των x_2 και x_3 . Η διακεκομμένη γραμμή είναι το όριο απόφασης. [75]

Το cosine similarity αποτελεί μια μέτρηση που εφαρμόζεται ευρέως στην ανάκτηση πληροφοριών και σε σχετικές μελέτες, η οποία μοντελοποιεί ένα έγγραφο κειμένου ως διάνυσμα όρων, με σκοπό τον υπολογισμό της τιμής συνημιτόνου μεταξύ των διανυσματικών όρων δύο εγγράφων [41]. Όσον αφορά τη μηχανή αναζήτησης, η τιμή ομοιότητας μεταξύ του ερωτήματος που μπορεί να θέσει ένας χρήστης και των εγγράφων που υπάρχουν ήδη στη μηχανή, ταξινομείται από την υψηλότερη προς την χαμηλότερη βαθμολογία ομοιότητας μεταξύ του διανύσματος όρου του εγγράφου και του διανύσματος όρου του ερωτήματος.[90]

Όσον αφορά τα ερωτήματα που μπορούν να τεθούν σε ένα έγγραφο, ένα έγγραφο μπορεί να αναπαρασταθεί ως διάνυσμα όρου που οι διαστάσεις του διανύσματος αναφέρονται στους όρους που είναι διαθέσιμοι στο έγγραφο [41].

Ένα έγγραφο μπορεί να περιγραφεί ως διανυσματική μορφή ως εξής:

$\vec{d} = (w_{d0}, w_{d1}, \dots, w_{dk})$ [102] Όπως και το έγγραφο, το ερώτημα του όρου μπορεί να περιγραφεί ως

διανυσματική μορφή ως εξής: $\vec{q} = (w_{q0}, w_{q1}, \dots, w_{qk})$ [20] όπου w_{di} και w_{qi} ($0 \leq i \leq k$) είναι αριθμοί αιώρησης που υποδεικνύουν την συχνότητα κάθε όρου μέσα σε ένα έγγραφο, ενώ η διάσταση κάθε διανύσματος αντιστοιχεί σε έναν όρο που είναι διαθέσιμος στο έγγραφο. Με βάση την ομοιότητα διανυσμάτων, η ομοιότητα μεταξύ δύο διανυσμάτων μπορεί να οριστεί ως εξής:

$$Sim(\vec{q}, \vec{d}) = \frac{\vec{q} \cdot \vec{d}}{|\vec{q}| |\vec{d}|} = \frac{\sum_{k=1}^t w_{qk} \times w_{dk}}{\sqrt{\sum_{k=1}^t (w_{qk})^2} \cdot \sqrt{\sum_{k=1}^t (w_{dk})^2}} \quad [1]$$

4.3 Ενδεικτικός Κώδικας Υλοποίησης

Αρχικά, μπορείτε να βρείτε ολοκληρωμένο τον κώδικα της υλοποίησής μας στο παρακάτω link:

<https://github.com/Evdokia97/MediBot>

Οι βιβλιοθήκες που χρησιμοποιήθηκαν είναι:

```
from sentence_transformers import SentenceTransformer
```

Όσον αφορά το SentenceTransformers, πρόκειται για ένα πλαίσιο της Python με σκοπό τις ενσωματώσεις προτάσεων, κειμένου και εικόνων τελευταίας τεχνολογίας. Δίνει την δυνατότητα υπολογισμού ενσωματώσεων προτάσεων / κειμένου για περισσότερες από 100 γλώσσες, ενσωματώσεις οι οποίες μπορούν στην συνέχεια να συγκριθούν με συναρτήσεις όπως αυτή του cosine similarity, ώστε να βρεθούν προτάσεις με παρόμοια σημασία. Αυτό μπορεί να φανεί χρήσιμο για ένα σημασιολογικό παρόμοιο κείμενο, σημασιολογική αναζήτηση ή εξόρυξη παράφρασης. Το συγκεκριμένο πλαίσιο βασίζεται στα PyTorch και Transformers και προσφέρει

μια μεγάλη συλλογή από προεκπαιδευμένα μοντέλα τα οποία έχουν την δυνατότητα να είναι συντονισμένα για διάφορες εργασίες.

```
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.model_selection import train_test_split
```

Σχετικά με την sklearn (ή αλλιώς Scikit-learn - πρώην scikits.learn) την οποία χρησιμοποιήσαμε, πρόκειται για μια δωρεάν βιβλιοθήκη μηχανικής εκμάθησης λογισμικού της Python. Διαθέτει διάφορους αλγόριθμους ταξινόμησης, παλινδρόμησης και ομαδοποίησης, συμπεριλαμβανομένων μηχανών υποστήριξης-διανυσμάτων, τυχαίων δασών, ενίσχυσης κλίσης, k-means και DBSCAN, και έχει σχεδιαστεί για να λειτουργεί με τις αριθμητικές και επιστημονικές βιβλιοθήκες Python NumPy και SciPy. Επίσης, είναι σε μεγάλο βαθμό γραμμένο σε Python και χρησιμοποιεί εκτενώς το NumPy για λειτουργίες γραμμικής άλγεβρας και πίνακα υψηλής απόδοσης. Ωστόσο, ορισμένοι βασικοί αλγόριθμοι είναι γραμμένοι σε Cython για την βελτίωση της απόδοσης με τις μηχανές διανυσμάτων υποστήριξης να υλοποιούνται από ένα περιτύλιγμα Cython γύρω από το LIBSVM. Τέλος, το Scikit-learn ενσωματώνεται καλά με πολλές άλλες βιβλιοθήκες Python, όπως το Matplotlib και το plotly για σχεδίαση, το NumPy για την διανυσματική διάταξη πίνακα, τα πλαίσια δεδομένων Pandas, το SciPy και πολλά άλλα.

Σχετικά με την υλοποίηση του chatbot μας, υπάρχουν κάποια σημεία του κώδικα που αξίζει να σημειωθούν. Ξεκινώντας από την εισαγωγή, τα δεδομένα εισόδου αφού εισαχθούν, γίνονται split με βάση τις τελείες που περιέχουν, ώστε να μπορεί το chatbot να δώσει συνδυαστικές απαντήσεις και όχι ολόκληρες παραγράφους. Στην συνέχεια, αφού συμπτήξαμε τα δεδομένα PubMed με τα δεδομένα SQUAD, χρησιμοποιήσαμε τα εξής μοντέλα:

```
#model = SentenceTransformer('roberta-large-nli-stsb-mean-tokens')
#model = SentenceTransformer('bert-large-nli-stsb-mean-tokens')
#model = SentenceTransformer('xlm-r-large-en-ko-nli-ststb')
model = SentenceTransformer('monologg/biobert_v1.1_pubmed')

text_embeddings=model.encode(content,batch_size=8,show_progress_bar=True)
```

Όσον αφορά τις πρώτες απαντήσεις του chatbot ή τις τελευταίες που θα δώσει, απαντά αναγνωρίζοντας κάποιες λέξεις που αφορούν καλωσόρισμα ή αποχαιρετισμό και απαντάει αντίστοιχα από ένα σύνολο δεδομένων που ορίσαμε εμείς, μέσω μιας συνάρτησης που ονομάσαμε greeting. Επιπλέον, κάθε φορά που ο χρήστης κάνει μία ερώτηση, τα δεδομένα εισόδου του κωδικοποιούνται με βάση το εκάστοτε μοντέλο, ώστε να βρεθεί το επιθυμητό ποσοστό ομοιότητας και να δοθεί η εκάστοτε απάντηση. Στην περίπτωση που το κατώφλι που δώσαμε δεν ικανοποιείται, επιστρέφεται στον χρήστη μία συγκεκριμένη πρόταση, ώστε να ενημερωθεί πως δεν βρέθηκαν δεδομένα. Η σύγκριση γίνεται με την συνάρτηση cosine_similarity, ενώ τα scores

επεξεργάζονται από την `index_sort`. Στην συγκεκριμένη συνάρτηση πραγματοποιούμε μία σύγκριση των αποτελεσμάτων και δημιουργία μιας λίστας με αυτά ταξινομημένα, ώστε να βρεθούν τα αποτελέσματα που μας ενδιαφέρουν από την συνάρτηση `response`.

```
def response(question):

    quest=[]
    quest.append(question)

    answer=''
    user_embeddings = model.encode(quest, batch_size = 8, show_progress_bar
= False)

    usersim = cosine_similarity(user_embeddings,text_embeddings)
    usersim_sorted = usersim.argsort()
    similarity_scores_list = usersim.flatten()
    index = index_sort(similarity_scores_list)
    index = index[1:]
    response_flag = 0
    j = 0

    for i in range(len(index)):

        if similarity_scores_list[index[i]]>0.3 :
            answer = answer+' '+content[index[i]]+'.'
            response_flag = 1
            j = j+1

        if j > 2:
            break

    if response_flag == 0:
        answer="I think I need to read more about that..."

    return answer
```

Όσον αφορά το `evaluation`, θα κληθεί όταν το επιλέξει ο χρήστης. Αφού επιλεγεί, η διαδικασία είναι παρόμοια με αυτή του `run`, καθώς ζητά από τον χρήστη την ερώτησή του και την πρόβλεψη της απάντησης που θα περίμενε. Στην συνέχεια, τρέχει την διαδικασία που θα έτρεχε και στο `run`, μόνο που αφού βρει την απάντηση που αυτό θεωρεί σωστή και εισάγουμε και εμείς την δική μας πρόβλεψη, χρησιμοποιεί την `compute_f1`, μια συνάρτηση που φτιάξαμε για την αξιολόγηση του chatbot μας. Έπειτα, μαζί με την δική του απάντηση, μας επιστρέφει και τα `precision`, `recall` και `f1` για την απάντηση που εισάγαμε και την απάντηση που μας έδωσε το ίδιο.

Τέλος, ρωτάει ξανά τον χρήστη για την επόμενη ερώτηση που θα ήθελε να ελέγξει, έως ότου αυτός επιλέξει να διακόψει.

```
def evaluation():

    flag=True
    print("\nMediBot: Hello "+name+"! Please enter the question you
        want to test."+"\n\n"+name+":")

    while(flag==True):
        user_response = input()
        user_response=user_response.lower()

        if(user_response!='bye'):
            answer=response(user_response)
            print("\nPlease enter your prediction for this
answer\n\n"+name+":")
            prediction=input()
            print("\nMediBot's answer:",end="")
            print(answer)
            compute_f1(answer,prediction)
            print("\nDo you want to test another question?\n\n"+name+":")

        else:
            flag=False
            print("\nMediBot: Bye"+name)
```

Αξίζει να σημειωθεί, πως για το κομμάτι της αξιολόγησης, καλούμε την `compute_f1` στην είσοδό της οποίας αναθέτουμε τις προτάσεις που θέλουμε να συγκρίνουμε, δηλαδή την απάντηση του chatbot (`answer`) και την απάντηση που θεωρούμε εμείς σωστή (`prediction`). Αρχικά, περνάμε και τις δύο προτάσεις σε μία συνάρτηση που ονομάζουμε `normalize_text`. Στην συγκεκριμένη συνάρτηση “απλοποιούμε” τις προτάσεις αφαιρώντας από αυτές τα άρθρα, τα περιττά κενά και άλλα σημεία της πρότασης που μπορεί να αποτελέσουν αποτρεπτικό παράγοντα στην διαδικασία της σύγκρισης. Στην συνέχεια βρίσκουμε τις κοινές λέξεις και υπολογίζουμε τα Precision, το οποίο αποτελείται από τις κοινές λέξεις προς το μήκος λέξεων της απάντησης του chatbot μας, το Recall το οποίο αποτελείται από τις κοινές λέξεις προς το μήκος λέξεων της απάντησης που δώσαμε ως πρόβλεψη και το F1, το οποίο προέρχεται από τον γνωστό τύπο του. Τέλος, στις περιπτώσεις που δεν υπάρχει κανένα κοινό μεταξύ των προτάσεων, όταν δηλαδή τα Common Tokens είναι 0, τότε θα έχουμε αυτόματα μηδενικά Precision και Recall, άρα και άπειρο F1. Το συγκεκριμένο δε μπορεί να συμβεί, καθώς το F1 παίρνει τιμές μεταξύ του 0 και του 1, άρα στην συγκεκριμένη περίπτωση δεν δίνουμε καμία τιμή και θεωρούμε ότι είναι όλα μηδενικά.

```
def compute_f1(answer, prediction):

    #Our answer
    prediction_tokens=normalize_text(prediction).split()
    #Medibot's answer
    answer_tokens=normalize_text(answer).split()

    if len(answer_tokens) == 0 or len(prediction_tokens) ==0:
        int(answer_tokens == prediction_tokens)

    common_tokens = set(answer_tokens) & set(prediction_tokens)

    if len(common_tokens) == 0:
        print('-----')

    else:
        prec= len(common_tokens)/len(set(answer_tokens))
        rec= len(common_tokens)/len(set(prediction_tokens))
        f1 = 2* (rec*prec)/(rec+prec)
        print('\nPrecision:',prec,'\n')
        print('Recall:',rec,'\n')
        print('F1:',f1,'\n')
```

Τέλος, το chatbot στο οποίο δεν χρησιμοποιήθηκε αλγόριθμος, η κωδικοποίηση πραγματοποιήθηκε με nltk και η σύγκριση διαφέρει από αυτήν που είχαμε στα bert μοντέλα. Πιο συγκεκριμένα, δημιουργήσαμε μία λίστα με tokenizers, στην οποία προσθέτουμε κάθε φορά και την ερώτηση του χρήστη σαν τελευταίο μέλος της λίστας. Στην συνέχεια, συγκρίνουμε το τελευταίο μέλος με όλα τα υπόλοιπα μέσω του cosine_similarity με σκοπό να δοθεί μία απάντηση από το chatbot, ενώ η fit_transform() χρησιμοποιείται στα δεδομένα εκπαίδευσης, ώστε να μπορούμε να τα κλιμακώσουμε και επίσης να μάθουμε τις παραμέτρους κλιμάκωσης τους. [101] Η συνέχεια παραμένει ίδια με την προηγούμενη εκδοχή του κώδικά μας με την διαφορά ότι μετά το πέρας της διαδικασίας, η ερώτηση που προσθέσαμε στην λίστα μας διαγράφεται, ώστε να μην θεωρηθεί κείμενο προς σύγκριση. Τέλος, όσον αφορά την κωδικοποίηση του κειμένου, χρησιμοποιούμε CountVectorizer, το οποίο επιστρέφει ένα κωδικοποιημένο διάνυσμα με το μήκος ολόκληρου του λεξιλογίου και έναν ακέραιο αριθμό για τον αριθμό των φορών που εμφανίστηκε κάθε λέξη στο κείμενο εισόδου χρησιμοποιώντας Bagofwords.

```
def response(user_response):

    robot_response=''
    sentence_list=nltk.sent_tokenize(raw)
    #print(sentence_list)
    sentence_list.append(user_response)
    cm = CountVectorizer().fit_transform(sentence_list)
```

```
similarity_scores = cosine_similarity(cm[-1], cm)
similarity_scores_list = similarity_scores.flatten()
index = index_sort(similarity_scores_list)
index = index[1:]
response_flag = 0
j = 0

for i in range(len(index)):

    if similarity_scores_list[index[i]] > 0.3:
        robot_response = robot_response+' '+sentence_list[index[i]]
        response_flag = 1
        j = j+1

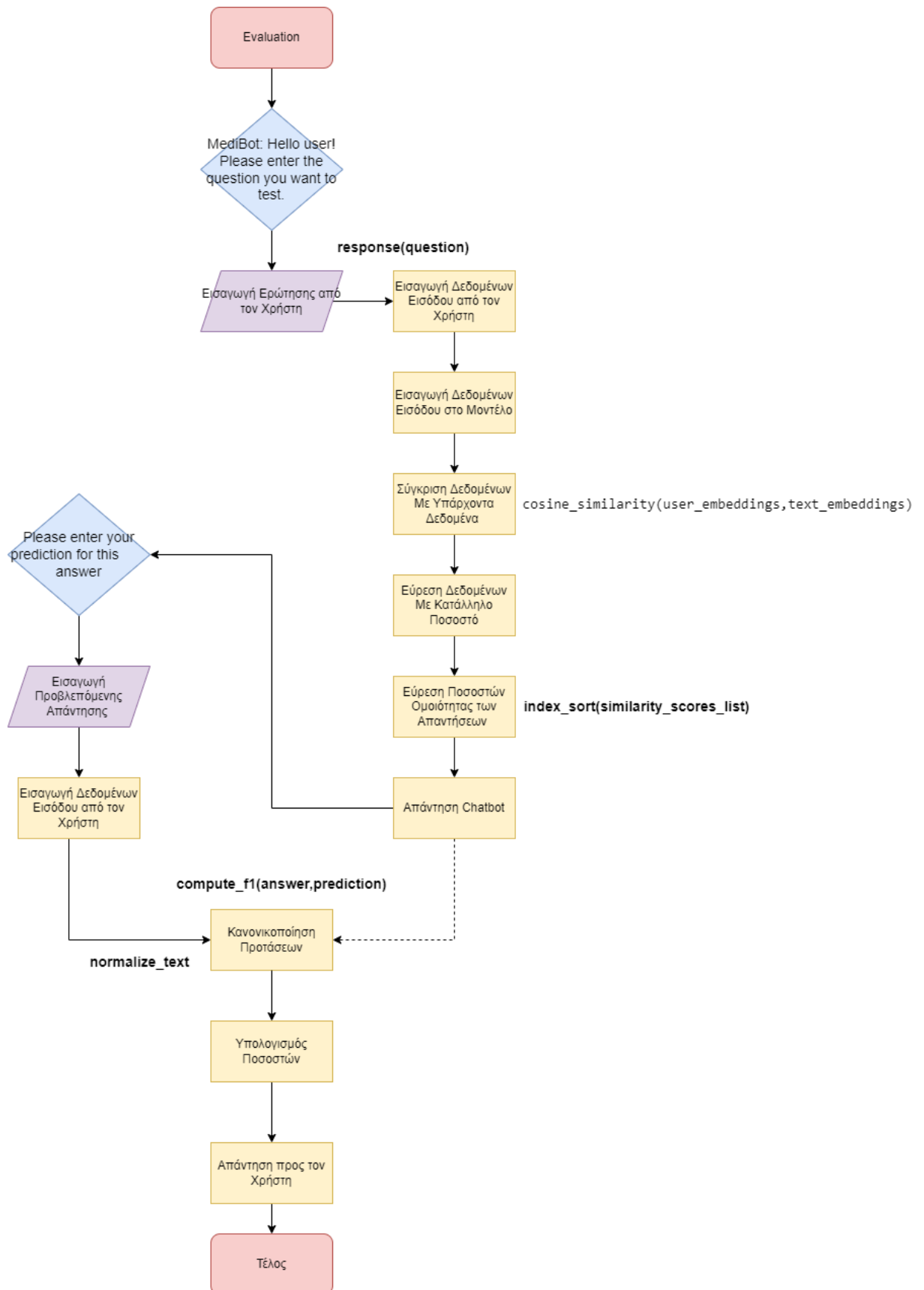
    if j > 2:
        break

if response_flag == 0:
    robot_response=robot_response+"I think I need to read more about
that..."

sentence_list.clear()

return robot_response
```

Τέλος, παρακάτω μπορούμε να παρατηρήσουμε το διάγραμμα ροής της αξιολόγησης του συστήματος, το οποίο αφορά τις response, evaluation και compute_fl :



Εικόνα 17. Διάγραμμα ροής αξιολόγησης.

5

Αξιολόγηση Υποδομής Chatbot με Χρήση Βιοϊατρικής Πληροφορίας

5.1 Μετρικές Σύγκρισης Συστημάτων

Ένας από τους πιο σύνηθες τρόπους μέτρησης της απόδοσης ενός μοντέλου, σχετικά με την ταξινόμηση κειμένου του, είναι η μετρική F1 [22]. Το F1-score βασίζεται στα Precision και Recall ενός μοντέλου πρόβλεψης ($F1 = 2 * (Precision * Recall) / (Precision + Recall)$), όπου η βαθμολογία μπορεί να ποικίλει από 0 έως 1 και χρησιμοποιείται για την αξιολόγηση της διαγνωστικής απόδοσης της πρόβλεψης ενός αλγορίθμου[97]. Ο όρος Recall (ή αλλιώς Sensitivity) πρόκειται για την αναλογία μεταξύ των πραγματικών θετικών περιπτώσεων και των σωστών προβλεπόμενων θετικών περιπτώσεων ($Recall = Sensitivity = tp / tp + fn$). Αντίθετα, το Precision (ή αλλιώς Confidence) αφορά την αναλογία των προβλεπόμενων θετικών περιπτώσεων και των ορθώς πραγματικά θετικών περιπτώσεων ($Precision = Confidence = tp / tp + fp$) (Εικόνα 19) [38].

Η συγκεκριμένη μετρική αποτελεί έναν αρμονικό μέσο ακρίβειας και ανάκλησης και αποτελεί έναν από τους πιο κατατοπιστικούς και χρήσιμους τρόπους μέτρησης της ακρίβειας της ταξινόμησης, λόγω του φαινομένου της ανισορροπίας της τάξης στην ταξινόμηση του κειμένου. Ωστόσο, είναι ευρέως γνωστό πως η F1 είναι ασύμμετρη, τόσο στην θετική, όσο και στην αρνητική κατηγορία και πως δεδομένων των συμπληρωμένων προβλέψεων και των πραγματικών ετικετών, μπορεί να δώσει διαφορετική βαθμολογία[63]. Συνεπώς, κάθε πρόβλεψη εξαρτάται όχι μόνο από την πιθανότητα ότι η ετικέτα ισχύει για αυτό το παράδειγμα, αλλά από την κατανομή των πιθανοτήτων για όλα τα άλλα παραδείγματα των δεδομένων με την συγκεκριμένη εξάρτηση να προκαλεί μεγάλη δυσκολία στην συσχέτιση των προβλέψεων για τις προβλέψεις ενός συστήματος αποτελούμενο από προβλεπόμενες πιθανότητες[95].

Όσον αφορά περιπτώσεις πολλαπλών κλάσεων στην συλλογή εγγράφων (όπως το Reuters-21578 με τις 118 τάξεις του), προτείνεται ο υπολογισμός ενός ενιαίου συνολικού μέτρου, το οποίο θα συνδυάζει τις βαθμολογίες F1, για κάθε κατηγορία. Για την κάλυψη της συγκεκριμένης ανάγκης, υπάρχουν δύο μέθοδοι: ο micro-averaging όρος και ο macro-averaging όρος [63]. Ο πρώτος όρος λειτουργεί συγκεντρώνοντας τα δεδομένα ανά έγγραφο σε όλες τις τάξεις και στην συνέχεια υπολογίζει την συνολική βαθμολογία του F1, σε αντίθεση με τον δεύτερο όρο, ο οποίος υπολογίζει έναν απλό μέσο όρο F1 των κατηγοριών. Οι διαφορές που

παρατηρούνται μεταξύ των δυο μετρικών δεν είναι αμελητέες. Ο micro-averaging δίνει ίση βαρύτητα για κάθε απόφαση ταξινόμησης ανά έγγραφο και επομένως κυριαρχείται από μεγάλες τάξεις, ενώ ο macro-averaging δίνει ίση βαρύτητα σε κάθε κατηγορία[118]. Επιπλέον, ο macro-averaging μέσος όρος αυξάνει το αντίκτυπο στην τελική βαθμολογία απόδοσης σε σπάνιες ετικέτες, ενώ ανά παράδειγμα ο μέσος όρος αυξάνει την σημασία της καλής απόδοσης σε κάθε παράδειγμα [94] και η μετρική micro F1 επηρεάζεται λιγότερο από την απόδοση σε σπάνιες ετικέτες, ενώ το Macro-F1 ζυγίζει το F1 του σε κάθε ετικέτα εξίσου [63]. Τέλος, οι Sokolova και Lapalme [91] έδωσαν έναν εναλλακτικό ορισμό του macro-averaging όρου ως αρμονικό μέσο όρο του απλού μέσου όρου της ακρίβειας και της ανάκλησης στις κατηγορίες. Τόσο ο micro-averaging όρος, όσο και ο macro-averaging όρος έχουν μια απλή ερμηνεία ως μέσο όρο ακρίβειας και ανάκλησης, με διαφορετικούς τρόπους υπολογισμού.

Τέλος, πολλοί ερευνητές πιστεύουν ότι η πιο λογική μέτρηση απόδοσης είναι η αναλογία μεταξύ του αριθμού των σωστά ταξινομημένων δειγμάτων και του συνολικού αριθμού δειγμάτων [89]. Αυτό το μέτρο ονομάζεται average accuracy και, εξ ορισμού, λειτουργεί όταν οι ετικέτες είναι περισσότερες από δύο, όπως για παράδειγμα σε περιπτώσεις όπως αυτή των πολλαπλών κλάσεων. Ωστόσο, όταν το σύνολο δεδομένων δεν είναι ισορροπημένο (ο αριθμός των δειγμάτων σε μια κατηγορία είναι πολύ μεγαλύτερος από τον αριθμό των δειγμάτων στις άλλες κλάσεις), το average accuracy δε μπορεί πλέον να θεωρείται αξιόπιστο μέτρο, καθώς παρέχει μια υπεραισιόδοξη εκτίμηση της ικανότητας του ταξινομητή τάξη πλειοψηφίας [26].

Οι προαναφερθείς μετρικές μπορούν να αναπαρασταθούν ως[93]:

Measure	Formula	Evaluation focus
Accuracy	$\frac{tp+tn}{tp+fn+fp+tn}$	Overall effectiveness of a classifier
Fscore	$\frac{(\beta^2+1)tp}{(\beta^2+1)tp+\beta^2fn+fp}$	Relations between data's positive labels and those given by a classifier
Average Accuracy	$\sum_{i=1}^I \frac{tp_i+tn_i}{tp_i+fn_i+fp_i+tn_i}$	The average per-class effectiveness of a classifier
$Fscore_{\mu}$	$\frac{(\beta^2+1)Precision_{\mu}Recall_{\mu}}{\beta^2 Precision_{\mu}+Recall_{\mu}}$	Relations between data's positive labels and those given by a classifier based on sums of per-text decisions
$Fscore_M$	$\frac{(\beta^2+1)Precision_MRecall_M}{\beta^2 Precision_M+Recall_M}$	Relations between data's positive labels and those given by a classifier based on a per-class average

Εικόνα 18. Τύποι μετρικών αξιολόγησης

	Actual Positive	Actual Negative
Predicted Positive	tp	fp
Predicted Negative	fn	tn

όπου:

Εικόνα 19. Περιεχόμενα μετρικών αξιολόγησης[95]

Ωστόσο, υπάρχει πάντα κάποιου βαθμού αβεβαιότητα σχετικά με τις μέσες βαθμολογίες της F1, οι οποίες αντικατοπτρίζουν μόνο την απόδοση του ταξινομητή κειμένου στα δεδομένα δοκιμής. δε μπορεί να υπάρξει εγγύηση απόδοσης για πεπερασμένο αριθμό αποτελεσμάτων σε πεπερασμένο αριθμό δοκιμών. Για παράδειγμα, ένας ταξινομητής με $miF1$ 0,81 σε 100 έγγραφα δοκιμής δε μπορεί να προσφέρει μεγάλη εμπιστοσύνη για το γεγονός ότι η μελλοντική του απόδοση θα είναι σίγουρα πάνω από 0,80. Σε αντίθεση με έναν ταξινομητή με $miF1$ 0.81 σε 100.000 έγγραφα δοκιμών, όπου μας δίνεται η δυνατότητα να αποκτήσουμε περισσότερη αυτοπεποίθηση σε σχέση με την προηγούμενη περίπτωση[33].

Επίσης, σχετικά με την εφαρμογή των συγκεκριμένων μετρικών στον κωδικά μας, εισάγουμε στην συνάρτηση δύο ειδών προτάσεις, την απάντηση του chatbot και την απάντηση που θεωρούμε εμείς σωστή με βάση το κείμενο. Πιο συγκεκριμένα, η μετρική F1 είναι μια μετρική αξιολόγησης για προβλήματα ταξινόμησης και χρησιμοποιείται ευρέως σε περιπτώσεις όπως η δική μας, στις οποίες χρησιμοποιούνται QA Chatbots. Πρόκειται για μία μετρική κατάλληλη για πληροφορίες όπως η ακρίβεια και η ανάκληση και σε περιπτώσεις όπως η δική μας, υπολογίζεται με βάση τις μεμονωμένες λέξεις στην πρόβλεψη έναντι εκείνων που έχουν δοθεί ως “Πραγματική Απάντηση”. Έτσι λοιπόν καταλήγουμε να λαμβάνουμε την ακρίβεια ως την αναλογία του αριθμού των κοινών λέξεων προς τον συνολικό αριθμό των λέξεων στην πρόβλεψη και την ανάκληση ως την αναλογία του αριθμού των κοινών λέξεων στον συνολικό αριθμό των λέξεων στην “αληθινή” πρόταση [83]. Τις συγκεκριμένες μετρικές τις δημιουργούμε εμείς στον κωδικά μας και δεν χρησιμοποιούμε κάποια συνάρτηση από βιβλιοθήκες, όπως αυτήν της sklearn, καθώς παρατηρήθηκε πως οι συγκεκριμένες συναρτήσεις αντιλαμβάνονται τις λέξεις ως ξεχωριστές κλάσεις και τις συγκρίνουν με τις αντίστοιχες λέξεις-κλάσεις της ίδιας θέσης στην εκάστοτε πρόταση που θέλουμε να συγκριθεί. Συνεπώς, επιλέξαμε να αξιολογήσουμε τις προτάσεις μας με βάση την συχνότητα εμφάνισης της εκάστοτε λέξης και όχι την απόλυτη ομοιότητά τους.

Ωστόσο, το ποσοστό αυτό θα μπορούσε να είναι ίσο με το 0, ενώ υπάρχει απάντηση. Το συγκεκριμένο γεγονός μπορεί να συμβεί στην περίπτωση που δεν δοθεί κάποια σχετική απάντηση και το F1 δεν δέχεται κάποια πραγματική τιμή, καθώς τα Recall και Precision του είναι ίσα με το 0 λόγω των μηδενικών κοινών τους λέξεων. Τις συγκεκριμένες απαντήσεις στην δική μας περίπτωση θα τις ονομάσουμε Not Related (Κεφάλαιο 4.6).

Ένα παράδειγμα που πέρασε το κατώφλι, αλλά έδωσε μηδενικές μετρικές είναι το παρακάτω:

Question:

Do you know about Groupamatic 360 ?

Our Prediction:

Groupamatic 360 can be adapted to the quantitative measurement of erythrocyte agglutination reactions.

Chatbot's Answer:

The 35S was incorporated into glycoprotein, mucopolysaccharide and sulfatide.

Έπειτα από το `normalize_text` που έχουμε εισάγει τις προτάσεις μας, οι λέξεις που θα συγκριθούν για κάθε περίπτωση είναι:

Our Prediction:

['groupamatic', '360', 'can', 'be', 'adapted', 'to', 'quantitative', 'measurement', 'of', 'erythrocyte', 'agglutination', 'reactions']

Chatbot's Answer:

Tokens of Medibot's answer are: ['35s', 'was', 'incorporated', 'into', 'glycoprotein', 'mucopolysaccharide', 'and', 'sulfatide']

Common Tokens: 0

The length of our answer is: 12

The length of Medibot's answer is: 8

Συνεπώς, τα Common Tokens είναι 0, καθώς δεν δίνεται καμία σωστή πρόβλεψη σχετικά με την πρόταση που του δώσαμε, ενώ τα μήκη των προτάσεων είναι 12 και 8. Άρα έχουμε τα παρακάτω:

Precision = $0/8=0$

Recall = $0/12 = 0$

F1 = $2*(\text{Precision}*\text{Recall})/(\text{Precision}+\text{Recall})= \text{Nan}$

Ωστόσο, γνωρίζουμε πως το F1 δε μπορεί να δεχτεί διαφορετικές τιμές εκτός του πλαισίου των τιμών 0 και 1. Για τον συγκεκριμένο λόγο, στις περιπτώσεις όπου δεν υπάρχει κανένα κοινό, όπως στο παράδειγμα που προαναφέραμε, για λόγους ευκολίας δίνουμε την τιμή "--" όταν δεν βρίσκουμε κανένα κοινό, γνωρίζοντας πως η συγκεκριμένη τιμή θα φέρει τα ανάλογα αποτελέσματα και θα μετρηθεί εκ μέρους μας ως μηδενική και Not Related.

5.2 Πειραματική Αξιολόγηση Συστημάτων

Αρχικά, όσον αφορά τα δεδομένα που χρησιμοποιήθηκαν, χρησιμοποιήσαμε 1100 Abstracts από ιατρικά papers, καθώς και 250 SQuAD v2.0 δεδομένα, λόγω περιορισμένης υπολογιστικής ισχύος. Τα παραπάνω δείγματα, διαχωρίστηκαν σε προτάσεις, μετά την είσοδό τους στο πρόγραμμα, προσφέροντας 8737 προτάσεις για σύγκριση.

Πιο συγκεκριμένα, όσον αφορά τα δεδομένα PubMed, η πλατφόρμα η οποία διαθέτει τα συγκεκριμένα papers είναι το PubMed, το οποίο διαθέτει δεδομένα κυρίως από το MEDLINE - κορυφαία βιβλιογραφική βάση δεδομένων του NLM που καλύπτει τομείς όπως της ιατρικής, της νοσηλευτικής, της οδοντιατρικής, της κτηνιατρικής, του συστήματος υγειονομικής περίθαλψης και των προκλινικών επιστημών, όπως η μοριακή βιολογία. Επιπλέον, παρέχει πρόσβαση και σε άλλους πολλούς πόρους, όπως αναφορές εκτός πεδίου εφαρμογής, καθώς και "αναφορές μόνο για το PubMed". Οι συγκεκριμένες αναφορές αφορούν κυρίως περιοδικά γενικής επιστήμης και χημείας που περιέχουν άρθρα βιοεπιστημών με ευρετήριο για το MEDLINE. Επιπροσθέτως, Ως απόκριση στις νέες προσεγγίσεις στην ηλεκτρονική δημοσίευση, το PubMed μπορεί πλέον να φιλοξενήσει και άρθρα πριν από την δημοσίευση. Αναφερόμαστε σε αυτές τις παραπομπές ως αναφορές "εκ των προτέρων" ή "erub". [79] Επιπλέον, τα δεδομένα που δανειστήκαμε από το PubMed, αφού τα φέραμε σε μορφή text, αφαιρέσαμε τους τίτλους που διέθεταν και κρατήσαμε μόνο το Abstract κομμάτι, όπου στο συγκεκριμένο κομμάτι, αφαιρέσαμε τις τελείες από τα αριθμητικά δεδομένα, ώστε να μην επηρεάσουν το split που θα τους εφαρμοστεί στην πορεία. Τέλος, όσον αφορά τις θεματικές τις οποίες καλύπτει το αρχείο που δανειστήκαμε από το PubMed, είναι ποικίλες. Πιο συγκεκριμένα αφορούν διάφορες ασθένειες, διάφορα γονίδια και μικροοργανισμούς, καθώς και τις ιδιότητές τους ή τα πειράματα στα οποία συμμετείχαν. Επιπλέον, περιλαμβάνει ασθένειες που μπορεί να προσβάλλουν ζώα, καθώς και πειράματα που έχουν πραγματοποιηθεί με αυτά. Τέλος, αναφέρονται τεχνικές, όπως αυτή της Gomi, γενικότερες τεχνικές αντιμετώπισης ασθενειών και τεχνικές στις οποίες συμπεριλαμβάνεται η αξιοποίηση του αίματος.

Όσον αφορά τα δεδομένα SQuAD v2.0, πρόκειται για μία επέκταση του ορισμού του SQuAD v1.1, επιτρέποντας στη νέα έκδοση να διατηρεί την πιθανότητα να μην υπάρχει σύντομη απάντηση στην παρεχόμενη παράγραφο, καθιστώντας το πρόβλημα πιο ρεαλιστικό. Πιο συγκεκριμένα, το σύνολο των δεδομένων της συγκεκριμένης έκδοσης αποτελείται από 536 δειγματοληπτικά κορυφαία Wikipedia άρθρα, τα οποία καλύπτουν ένα ευρύ φάσμα θεμάτων. Το σύνολο δεδομένων κατασκευάστηκε από ερωτήσεις και απαντήσεις των συγκεκριμένων κειμένων και μπορεί να επεκταθεί ανάλογα την απάντηση. Η απάντηση της εκάστοτε ερώτησης μπορεί να είναι ημερομηνία, αριθμητική τιμή, όνομα προσώπου, τοποθεσία ή οποιαδήποτε άλλη πληροφορία που το κείμενο μπορεί να δίνει συγκεκριμένα ή να δίνεται από τα συμφραζόμενα[88].

Τα πειράματα πραγματοποιήθηκαν σε ένα chatbot με τρία διαφορετικά μοντέλα bert και σε ένα chatbot χωρίς αλγόριθμο, στο οποίο και χρησιμοποιήθηκε ο CountVectorizer. Όσον αφορά την πρώτη περίπτωση, η φόρτωση των δεδομένων - στους ήδη εκπαιδευμένους αλγορίθμους -

χρειάστηκε περίπου 40 λεπτά σε κάθε αλγόριθμο, ενώ η πραγματοποίηση κάθε πειράματος χρειάστηκε 2 ώρες σε κάθε περίπτωση. Επιπλέον, τα πειράματα πραγματοποιήθηκαν για τρία διαφορετικά κατώφλια (για >0.3 , >0.4 , >0.5), αρά συνολικά πραγματοποιήσαμε 12 πειράματα για να καταλήξουμε στα αποτελέσματα που θα αναλύσουμε στο Κεφάλαιο 5.3.

Σχετικά με τις ερωτήσεις που κάναμε στα chatbots, επιλέχθηκαν 100 ερωτήσεις που βασίζονται στα δεδομένα που εισάγαμε, οι οποίες αφορούσαν τόσο γενικές, όσο και συνδιαστικές ή απόλυτα συγκεκριμένες ερωτήσεις. Όσον αφορά τις απαντήσεις που χαρακτηρίστηκαν ως “κατάλληλες”, επιλέχθηκαν αυτούσιες μέσα από τα δεδομένα εισόδου, χωρίς να πραγματοποιηθεί καμία απολύτως αλλαγή, ώστε να γίνει απόλυτη σύγκριση με τις απαντήσεις που θα δοθούν από το εκάστοτε chatbot. Πιο συγκεκριμένα, η καταγραφή της αξιολόγησης είχε τη μορφή:

Question	Prediction	Answer
What is prune belly syndrome?	Prune-belly syndrome is a rare congenital disease involving a deficiency of abdominal muscles, genitourinary abnormalities, and cryptorchidism. Patients with prune belly syndrome present a spectrum of abnormality, both in the abdominal wall and the urinary tract. The definition, history, and clinical presentation of the prune belly syndrome are outlined.	In the past decade, seven children with the prune-belly syndrome have been seen. Patients with prune belly syndrome present a spectrum of abnormality, both in the abdominal wall and the urinary tract. Urodynamic studies have been carried out in 10 boys with the prune belly syndrome.

Precision	Recall	F1
0,5666666667	0,5151515152	0,5396825397

Εικόνα 20. Μέθοδος καταγραφής αξιολόγησης.

Τέλος, ο τρόπος που πραγματοποιήθηκε η εκτίμηση του chatbot, ήταν η επιλογή του χρήστη να εκτιμήσει και όχι να χρησιμοποιήσει το chatbot, το οποίο οδηγούσε σε ερώτηση από τον χρήστη και εισαγωγή της προτεινόμενης απάντησης, η οποία με την σειρά της θα συγκριθεί με αυτή του chatbot και θα δώσει τα ποσοστά επιτυχίας ή αποτυχίας της συγκεκριμένης απάντησης. Ωστόσο, υπήρξαν απαντήσεις οι οποίες δεν διέθεταν κανένα ποσοστό επιτυχίας, καθώς περιείχαν περιεχόμενο που κρίθηκε μη σχετικό και χαρακτηρίστηκαν ως Not Related, ενώ υπήρξαν και περιπτώσεις που δεν δόθηκε καμία απάντηση η οποία να πληρεί τα ποσοστά του κατωφλίου που είχαμε θέσει. Συνεπώς, οι συνολικές μετρικές υπολογίστηκαν αποκλειστικά για τα

δεδομένα που έδωσαν ποσοστά μεγαλύτερα του μηδενός. Πιο συγκεκριμένα, έστω ότι δόθηκαν x Related απαντήσεις από τις 100 ερωτήσεις που κάναμε με y το άθροισμα της εκάστοτε μετρικής, τότε το συνολικό ποσοστό z υπολογίζεται από τον τύπο $z = (y/x) * 100$.

Παρακάτω παρατηρούμε ένα παράδειγμα αξιολόγησης όπως θα το τρέχαμε σε πραγματικό χρόνο:

Medibot : Hi, I am Medibot. I am an expert on medical staff. What s your name?

Evdokia

Medibot : Hello Evdokia! Do you want to evaluate or run the chatbot?

Evdokia :
I would like to evaluate.

Medibot : Please enter the question you want to test.

Evdokia :
What does the use of 15 channels blood grouping autoanalyzer with automated recording prove?

Medibot : Please enter your prediction for this answer

Evdokia :
The use of the 15 channels blood grouping autoanalyzer with automated recording and interpretation proved its fiability through more than 30 000 tests for ABO, Rh and Kell determinations.

Medibot s answer: The use of the 15 channels blood grouping autoanalyzer with automated recording and interpretation proved its fiability through more than 30 000 tests for ABO, Rh and Kell determinations. This was compared with intermittent determinations of the pH in the capillary blood from the heel of the infant. The meaning of the general blood-gas analysis in the artery umbilicalis for the morbidity of acidity and the check of the obstetrical activity of the clinic is shown.

Precision: 0.52
Recall: 1.0
F1: 0.6842105263157895

Εικόνα 21. Αναπαράσταση εκτέλεσης της αξιολόγησης του Chatbot μας.

Τέλος, με σκοπό να επιβεβαιώσουμε τα παραπάνω αποτελέσματα, μπορούμε να παρατηρήσουμε την ανάλυσή τους παρακάτω:

Question:

What does the use of 15 channels blood grouping autoanalyzer with automated recording prove?

Our Prediction:

The use of the 15 channels blood grouping autoanalyzer with automated recording and

interpretation proved its fiability through more than 30 000 tests for ABO, Rh and Kell determinations.

Chatbot's Answer:

The use of the 15 channels blood grouping autoanalyzer with automated recording and interpretation proved its fiability through more than 30 000 tests for ABO, Rh and Kell determinations. This was compared with intermittent determinations of the pH in the capillary blood from the heel of the infant. The meaning of the general blood-gas analysis in the artery umbilicalis for the morbidity of acidity and the check of the obstetrical activity of the clinic is shown.

Έπειτα από το `normalize_text` που έχουμε εισάγει τις προτάσεις μας, οι λέξεις που θα συγκριθούν για κάθε περίπτωση είναι:

Our Prediction:

['use', 'of', '15', 'channels', 'blood', 'grouping', 'autoanalyzer', 'with', 'automated', 'recording', 'and', 'interpretation', 'proved', 'its', 'fiability', 'through', 'more', 'than', '30', '000', 'tests', 'for', 'abo', 'rh', 'and', 'kell', 'determinations']

Chatbot's Answer:

['use', 'of', '15', 'channels', 'blood', 'grouping', 'autoanalyzer', 'with', 'automated', 'recording', 'and', 'interpretation', 'proved', 'its', 'fiability', 'through', 'more', 'than', '30', '000', 'tests', 'for', 'abo', 'rh', 'and', 'kell', 'determinations', 'this', 'was', 'compared', 'with', 'intermittent', 'determinations', 'of', 'ph', 'in', 'capillary', 'blood', 'from', 'heel', 'of', 'infant', 'meaning', 'of', 'general', 'bloodgas', 'analysis', 'in', 'artery', 'umbilicalis', 'for', 'morbidity', 'of', 'acidity', 'and', 'check', 'of', 'obstetrical', 'activity', 'of', 'clinic', 'is', 'shown']

Common Tokens: 26

The length of our answer is: 26

The length of Medibot s answer is: 50

Precision = $26/50 = 0.52$

Recall = $26/26 = 1$

$$F1 = \frac{2 * (Precision * Recall)}{(Precision + Recall)} = \frac{2 * (1 * 0.52)}{(1 + 0.52)} = \frac{1.04}{1.52} = 0.6842105263157895$$

5.3 Συγκρίσεις Αποτελεσμάτων Αξιολόγησης

Έπειτα από τα πειράματα που πραγματοποιήθηκαν, συγκεντρώθηκαν στους παρακάτω πίνακες ποσοστά σχετικά με τα αποτελέσματα του κάθε πειράματος για το chatbot με χρήση τριών μοντέλων Bert, το chatbot χωρίς αλγόριθμο, καθώς και για κάθε διαφορετικό κατώφλι. Όσον αφορά τους πίνακες, περιέχουν τα ποσοστά των ερωτήσεων που δεν κατάφεραν να απαντηθούν (Not Answered), ποσοστά των ερωτήσεων που απαντήθηκαν -χωρίς καμία επιτυχία (ποσοστά precision, recall και f1 ίσα με το μηδέν) -, καθώς και τα ποσοστά των απαντήσεων που απαντήθηκαν με ορισμένα ποσοστά επιτυχίας.

Παρατηρούμε λοιπόν, πως για την περίπτωση όπου το κατώφλι (cosine similarity) τίθεται μεγαλύτερο του 0.5, το Roberta αν και διαθέτει τις λιγότερες απαντημένες ερωτήσεις ανάμεσα στα τέσσερα μοντέλα Bert, δίνει τα καλύτερα ποσοστά επιτυχίας με το Bert Large να ακολουθεί με τη μικρή διαφορά του ~3%, αλλά με μία μεγάλη διαφορά στις απαντημένες σχετικές ερωτήσεις. Ακολουθεί το Biobert με τα καλύτερα ποσοστά απαντημένων ερωτησεων και μικρή διαφορά επιτυχίας και το Xlm με παρόμοια ποσοστά απαντήσεων, αλλά τα χαμηλότερα ποσοστά επιτυχίας. Τέλος, στην περίπτωση που δεν υπήρξε μοντέλο, παρατηρούνται αξιοθαύμαστα ποσοστά επιτυχίας των απαντήσεων, ωστόσο οι περισσότερες ερωτήσεις δεν είχαν καμία απάντηση, έστω και λανθασμένη. Όσον αφορά την συγκεκριμένη περίπτωση, παρατηρείται ότι το CountVectorizer διαθέτει τις καλύτερες τιμές σε κάθε μετρική, ενώ μεταξύ των μοντέλων BERT, το Roberta διαθέτει το καλύτερο Precision, αρα και τα λιγότερα περιττά στοιχεία στις απαντήσεις του με το καλύτερο Recall, αρα και τις πιο εύστοχες απαντήσεις, με τα Bert Large, Biobert και Xml να ακολουθούν.

	Models	Not answered	Not Related	Related	Related		
					Precision	Recall	F1
Cosine Similarity 0.5	roberta-large-nli-stsb-mean-tokens	18%	2%	80%	25.65%	39.27%	29.34%
	xlm-r-large-en-ko-nli-stsb	3%	1%	96%	17.78%	28.91%	20.31%
	bert-large-nli-stsb-mean-tokens	0%	2%	98%	22.46%	38.58%	26.31%
	biobert_v1.1_pubmed	0%	2%	98%	21.17%	34.93%	24.88%
	CountVectorizer	64%	0%	36%	80.45%	90.29%	82.76%

Όσον αφορά την περίπτωση που θέσαμε το κατώφλι στο 0,4, το Roberta, αν και μειώνει τα ποσοστά επιτυχίας του, αυξάνει κατά 16% τις σχετικές απαντήσεις που δίνει, ενώ δεν δίνει καμία απάντηση που να μην έχει καμία ομοιότητα. Ωστόσο, στα υπόλοιπα τρία μοντέλα, παρατηρούνται πολύ μικρές αρνητικές διακυμάνσεις με τα Xlm και Biobert να σημειώνουν αλλαγές των 1% σχετικά με τις απαντήσεις που έδωσαν. Τέλος, όσον αφορά το chatbot που δεν είχε κάποιον αλγόριθμο, αν και παρουσιάζει μείωση των ποσοστών επιτυχίας, παρουσιάζει μεγάλη θετική διαφορά στα ποσοστά των απαντήσεων που δίνει, αντιστρέφοντας έτσι τα “αρνητικά” αποτελέσματα επιτυχίας σε “θετικά”. Όσον αφορά τα Precision και Recall, οι θέσεις που κατείχαν παρέμειναν οι ίδιες, ενώ παρατηρείται πως το Roberta και το CountVectorizer είχαν τη μεγαλύτερη μείωση σχετικά με τις συγκεκριμένες μετρικές, ενώ τα υπόλοιπα είχαν μία σχετικά ανεπαίσθητη μείωση, γεγονός που υποδηλώνει πως τα πρώτα επηρεάστηκαν περισσότερο από τη μείωση του κατωφλίου σε σχέση με τα δεύτερα.

		Related					
		Not answered	Not Related	Related	Precision	Recall	F1
Cosine Similarity 0.4	roberta-large-nli-stsb-mean-tokens	4%	0%	96%	21.50%	35.32%	25.29%
	xlm-r-large-en-ko-nli-ststb	1%	1%	98%	17.50%	28.59%	20.09%
	bert-large-nli-stsb-mean-tokens	0%	2%	98%	22.42%	38.43%	25.97%
	biobert_v1.1_pubmed	0%	1%	99%	21.13%	34.69%	24.76%
	CountVectorizer	40%	1%	59%	65.99%	81.03%	68.92%

Τέλος, όσον αφορά την τελευταία μας περίπτωση, αυτή του 0,3 κατωφλίου, τα ποσοστά των απαντήσεων δεν άλλαξαν σε καμία περίπτωση, εκτός αυτής του chatbot με CountVectorizer, το οποίο αυξάνει κατά 19% τις σχετικές του ερωτήσεις και κατα 2% τις μη σχετικές. Στις περιπτώσεις των Roberta και Bert Large παρατηρείται μικρή πτώση όσον αφορά τα ποσοστά επιτυχίας, ενώ στα Xlm και Biobert μία μικρή αύξηση. Τέλος, το chatbot με CountVectorizer κινείται στα ίδια πλαίσια με την προηγούμενη περίπτωση, καθώς μαζί με την αύξηση των ποσοστών απάντησης, σημειώνεται και μείωση των ποσοστών επιτυχίας. Όσον αφορά τις τιμές Precision και Recall, τα Bert Large, Xlm και Biobert σημείωσαν πολύ μικρή αλλαγή, όπως και στην προηγούμενη περίπτωση, γεγονός που μας υποδηλώνει πως στις συγκεκριμένες περιπτώσεις, το κατώφλι δεν φέρει ιδιαίτερες αλλαγές. Σχετικά με το Roberta, υπήρξε μία μείωση μικρότερη από αυτήν της προηγούμενης φοράς, επομένως η μείωση του κατωφλίου μεταξύ του 0.3 και 0.4 δεν προκαλεί μεγάλες αλλαγές σε αντίθεση με τη μετατόπιση μεταξύ του 0.5 και 0.4. Ωστόσο,

παρατηρείται πως ο CountVectorizer πραγματοποίησε μία εξίσου μεγάλη μείωση σε σχέση με την προηγούμενη περίπτωση και συνεπώς η επιρροή της μείωσης του κατώφλιου διατηρεί την επιρροή της πάνω του.

					Related		
	Models	Not answered	Not Related	Related	Precision	Recall	F1
Cosine Similarity 0.3	roberta-large-nli-stsb-mean-tokens	4%	0%	96%	20.52%	34.25%	24.21%
	xlm-r-large-en-ko-nli-stsb	1%	1%	98%	17.54%	28.75%	20.21%
	bert-large-nli-stsb-mean-tokens	0%	2%	98%	22.46%	38.58%	26.31%
	biobert_v1.1_pubmed	0%	1%	99%	21.12%	34.78%	24.81%
	CountVectorizer	19%	3%	78%	46.04%	72.82%	53.12%

Εν κατακλείδι, παρατηρείται ότι όσο μειώνεται το κατώφλι, τόσο μειώνεται και η απόδοση του chatbot φαινομενικά, ωστόσο σε ορισμένες περιπτώσεις αυξάνονται οι ερωτήσεις που δίνονται και συνεπώς θα μπορούσε να θεωρηθεί πως δίνεται καλύτερο και όχι χειρότερο αποτέλεσμα. Ωστόσο, εάν συγκρίνουμε και τους τρεις πίνακες, θα παρατηρήσουμε ότι η κατάταξη σχετικά με το cosine similarity έχει ως εξής : 0.5, 0.3, 0.4, ενώ είναι ξεκάθαρο πως το μοντέλο που δεν περιείχε κανέναν αλγόριθμο, έχει τη μεγαλύτερη επιτυχία, ενώ ανάμεσα στα μοντέλα, το roberta είναι αυτό που κάνει την διαφορά με τα bert-large, xlm και biobert να ακολουθούν στην περίπτωση του 0,5, ενώ στις υπόλοιπες περιπτώσεις οι θέσεις αλλάζουν με το Bert Large να κυριαρχεί και τα Biobert, Roberta και Xlm να ακολουθούν. Ο λόγος που συμβαίνει αυτό είναι το γεγονός ότι, όπως αναφέραμε, το roberta δεν χρησιμοποιεί στατική κάλυψη, αλλά στρατηγική κάλυψης, διαφορετική για κάθε εκτέλεση, με κάθε εμφάνιση της ίδιας λέξης να συνοδεύεται από την εκχώρηση ενός διαφορετικού διανύσματος ενσωμάτωσης. Σε συνέχεια, το bert-large έχει το θετικό της διάκρισης μεταξύ σχετικών, αλλά διακριτών οντοτήτων ή εννοιών, ενώ το xlm έχει την τάση να κατακερματίζει τις λέξεις σε ικανοποιητικό βαθμό. Τέλος, το biobert χρησιμοποιεί διαφορετικό tokenization από το απλό Bert, σε αντίθεση με τα προηγούμενα. Παρ' όλα αυτά, παρατηρείται πως όσο μειώνεται το κατώφλι, τόσο μειώνεται και το Precision και το Recall και συνεπώς αυξάνεται η “φλυαρία” και μειώνεται η πιθανότητα να βρεθεί η σωστή πρόταση αντίστοιχα. Ωστόσο, παρατηρείται πως στις περιπτώσεις που χρησιμοποιήθηκε κάποιο μοντέλο του αλγορίθμου, οι διακυμάνσεις είναι μικρές, σε αντίθεση με την περίπτωση που δεν χρησιμοποιήθηκε κανένας αλγόριθμος, όπου οι διακυμάνσεις γίνονται φανερά αισθητές.

Γενικεύοντας την περίπτωση του BERT, γνωρίζουμε πως μπορεί να δημιουργήσει διαφορετικές ενσωματώσεις λέξεων για μια λέξη που αποτυπώνει το πλαίσιο μιας λέξης, καθιστώντας αναγκαία την χρήση ενός μοντέλου που χρησιμοποιήθηκε για την εκπαίδευση των

διανυσμάτων ακόμη και μετά την εκπαίδευση. Ωστόσο, δεδομένου ότι τα αποτελέσματά του εξαρτώνται από το περιβάλλον του, αντιπροσωπεύει την είσοδο ως υπολέξεις και μαθαίνει να δημιουργεί ενσωματώσεις για αυτές. Αντίθετα, το CountVetorizer χρησιμοποιείται για τη μετατροπή μιας συλλογής εγγράφων κειμένου σε ένα διάνυσμα μετρήσεων όρων. Επιτρέπει την προεπεξεργασία δεδομένων κειμένου πριν από την δημιουργία της διανυσματικής αναπαράστασης, καθιστώντας το έτσι μια εξαιρετικά ευέλικτη ενότητα αναπαράστασης χαρακτηριστικών για κείμενο.

Συνοψίζοντας, παρατηρείται ότι αν και φαινομενικά, ένα προεκπαιδευμένο μοντέλο θα περιμέναμε να δώσει καλύτερα αποτελέσματα, το chatbot που δεν περιέχει κάποιον αλγόριθμο, αλλά τον CountVectorizer παρέχει μακράν καλύτερα αποτελέσματα, ενώ στις περιπτώσεις των Bert, περισσότερες προσπάθειες απάντησης. Το συγκεκριμένο γεγονός προέρχεται από το γεγονός πως ο CountVectorizer επιστρέφει ένα κωδικοποιημένο διάνυσμα με το μήκος ολόκληρου του λεξιλογίου και έναν ακέραιο αριθμό για τον αριθμό των φορών που εμφανίστηκε κάθε λέξη στο κείμενο εισόδου, δίνοντας έμφαση σε κάθε λέξη ξεχωριστά και συλλέγοντας τα ποσοστά των λέξεων και αντίστοιχα των προτάσεων, σε αντίθεση με το Bert το οποίο αναλύει τις προτάσεις ολιστικά. Πιο συγκεκριμένα, πραγματοποιεί προβλέψεις σχετικά με το αρχικό αναγνωριστικό λεξιλογίου της μάσκας της εκάστοτε λέξης, που βασίζεται μόνο στο περιεχόμενό της και οδηγεί στην έξοδο ένα σταθμισμένο άθροισμα των διανυσμάτων τιμών, δίνοντας μεγάλη προσοχή είτε στις σταθερές μετατοπίσεις της θέσης, είτε σε ευρέως ολόκληρη πρόταση, με ιδιαίτερη προσοχή στην αρχή και το τέλος του κειμένου, σε αντίθεση με τον CountVectorizer ο οποίος δίνει περισσότερη προσοχή σε ολόκληρη την πρόταση.

6

Συμπεράσματα και Προεκτάσεις

6.1 Συμπεράσματα

Ολοκληρώνοντας την συγκεκριμένη εργασία, αποκόμισα μία πιο ολοκληρωμένη εικόνα των δυνατοτήτων που θα μπορούσαν να έχουν τα chatbots στην ζωή μας, καθώς και την πολυποίκιλη φύση των μορφών τους. Η επιλογή των δομικών στοιχείων κατά την δημιουργία ενός chatbot αποτελεί την πιο σημαντική απόφαση. Απόφαση η οποία μπορεί να παίζει καθοριστικό παράγοντα στην επιτυχία ή την αποτυχία του. Είναι πολύ σημαντικό, πριν από αυτήν την απόφαση, να αναπτυχθεί μία όσο το δυνατόν πιο καθαρή και συγκεκριμένη εικόνα του σκοπού και της χρήσης του chatbot, καθώς και των επιλογών που υπάρχουν για την εκάστοτε περίπτωση.

Πιο συγκεκριμένα, για να έχει αξία μία πειραματική αξιολόγηση ή και μια πιο ουσιαστική δημιουργία ενός chatbot, θα πρέπει να υπάρξει πλήρης ανάλυση των βασικών στοιχείων που θα επιλέξουμε, όπως το Bert ή ο CountVectorizer στην δική μας περίπτωση. Επιπλέον, θα πρέπει να υπάρχει ανάλυση των δυνατοτήτων τους και φυσικά δοκιμές για το καλύτερο δυνατό αποτέλεσμα. Ωστόσο, την δική μας περίπτωση, οι δοκιμές δεν είχαν το μέγεθος που θα θέλαμε, καθώς υπήρχε χαμηλή υπολογιστική ισχύς, γεγονός όμως που δε μας απέτρεψε από το να βγάλουμε ένα προσεγγιστικό αποτέλεσμα για κάθε περίπτωση που εξετάσαμε.

Σημαντικό ρόλο έπαιξε και το είδος των δεδομένων που εισάγαμε, καθώς στην περίπτωση μας επιθυμούσαμε να υπάρχει ακρίβεια και κυρίως έγκυρα δεδομένα λόγω του κλάδου που επιλέξαμε να συνδυάσουμε, αυτόν της βιοϊατρικής. Σε δοκιμές τέτοιου είδους, απαιτείται μεγάλη προσοχή στο είδος του chatbot που θα επιλεγεί, καθώς όπως προαναφέραμε η επιλογή ενός QA Chatbot δεν ήταν τυχαία. Προσπαθήσαμε όσο το δυνατόν περισσότερο να συνδυάσουμε μία πιο διαδραστική εμπειρία, με έγκυρα δεδομένα και προτάσεις που δεν θα αλλοίωναν έρευνες χρόνων και δεδομένα υψηλής σημασίας και ακρίβειας.

Τέλος, η αξία της συγκεκριμένης εργασίας δεν βρίσκεται τόσο στην χρήση του chatbot που δημιουργήθηκε, αλλά στις κατευθύνσεις που μπορεί να πάρει κάποια επόμενη αντίστοιχη έρευνα. Τα αποτελέσματά μας, αν και μικρής κλίμακας, σκιαγραφούν μία αρκετά ξεκάθαρη εικόνα για κάποιον που θα επιθυμούσε να πάει την έρευνά του ένα βήμα παραπέρα, γνωρίζοντας εκ των προτέρων ποια στοιχεία θα τον ωφελήσουν και ποια όχι, ενώ παράλληλα θα μπορούσαν να αξιοποιηθούν ο κώδικας και οι πηγές που χρησιμοποιήθηκαν στην συγκεκριμένη εργασία.

6.2 Μελλοντική Έρευνα

Στα πλαίσια της εξέλιξης της συγκεκριμένης διπλωματικής, προτείνεται η εξέλιξη της συγκεκριμένης έρευνας, τόσο σε επίπεδο αρχιτεκτονικής, όσο και σε επίπεδο δεδομένων και αλγορίθμων. Πιο συγκεκριμένα, προτείνεται ο πειραματισμός με την αρχιτεκτονική του chatbot, καθώς και η φόρτωση σε αυτό μεγαλύτερου όγκου δεδομένων, γεγονός που δεν πραγματοποιήθηκε στην συγκεκριμένη διπλωματική λόγω έλλειψης υπολογιστικής ισχύος και θα μπορούσε να δώσει περισσότερα και πιο ακριβή δεδομένα αξιολόγησης. Επίσης, όσον αφορά τα δεδομένα, θα μπορούσε να παρατηρηθεί και η συμπεριφορά του μοντέλου σε δεδομένα τα οποία θα αφορούν ένα συγκεκριμένο ιατρικό ζήτημα και τις αποδόσεις του σε τέτοιου είδους περιπτώσεις. Σχετικά με το κοινό που απευθύνεται, η χρήση μεγαλύτερου όγκου δεδομένων ή δεδομένων που αφορούν έναν συγκεκριμένο κλάδο, θα μπορούσε να διαμορφώσει ένα chatbot το οποίο θα μπορούσε να αξιοποιηθεί από επαγγελματίες και όχι απαραίτητα από το ευρύ κοινό. Επιπλέον, σχετικά με τα μοντέλα του αλγορίθμου που χρησιμοποιήθηκαν, προτείνεται ο πειραματισμός με άλλου είδους αλγορίθμους, καθώς η συγκεκριμένη διπλωματική παρέχει συγκρίσεις με έναν ικανοποιητικό αριθμό μοντέλων του συγκεκριμένου αλγόριθμου και η επέκταση σε άλλου είδους αλγορίθμους θα βοηθούσε σε μία πιο επεκτατικής φύσεως έρευνα και σε ακόμη περισσότερα δεδομένα σύγκρισης ή η δημιουργία ενός πιθανού μοντέλου με την χρήση CountVectorizer. Το συγκεκριμένο μοντέλο θα μπορούσε να διατηρήσει τα πλεονεκτήματα του CountVectorizer, ενώ παράλληλα θα μπορούσε να αποκτήσει μνήμη για καλύτερες επιδόσεις. Τέλος, θα μπορούσε να αναπτυχθεί μια διαδραστική πλατφόρμα, τόσο για την οπτική βελτίωση, όσο και για την διευκόλυνση πιθανών χρηστών και τη μεγαλύτερη οικειότητα που θα μπορούσε να αναπτυχθεί με τον χρήστη.

Βιβλιογραφία - Αναφορές

- [1]. A. Madylova and S.G. Oguducu, "A taxonomy based semantic similarity of documents using the cosine measure," in *Proceeding of International Symposium on Computer and Information Sciences*, 2009, pp. 129–134
- [2]. Adamopoulou Eleni and Moussiades Lefteris, Maglogiannis Ilias, Iliadis Lazaros and Pimenidis Elias, "An Overview of Chatbot Technology", *Artificial Intelligence Applications and Innovations*, 2020, Springer International Publishing, Cham, 373--383
- [3]. Adina Williams, Nikita Nangia, and Samuel R Bowman. 2018. A broad-coverage challenge corpus for sentence understanding through inference. In *NAACL*.
- [4]. Aishwarya Gupta, Divya Hathwar, "Introduction to AI Chatbots", *International Journal of Engineering Research and Technology (IJERT)*, 2020
- [5]. Alan Akbik, Tanja Bergmann, and Roland Vollgraf. 2019. Pooled Contextualized Embeddings for Named Entity Recognition. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 724–728, Minneapolis, Minnesota. Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/N19-1078>
- [6]. Alessio Miaschi and Felice Dell'Orletta. 2020. Contextual and Non-Contextual Word Embeddings: An in-depth Linguistic Investigation. In *Proceedings of the 5th Workshop on Representation Learning for NLP*, pages 110–119. DOI: <https://doi.org/10.18653/v1/2020.repl4nlp-1.15>
- [7]. Alex Wang, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel Bowman. 2018a. Glue: A multi-task benchmark and analysis platform for natural language understanding. In *Proceedings of the 2018 EMNLP Workshop BlackboxNLP: Analyzing and Interpreting Neural Networks for NLP*, pages 353–355
- [8]. Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Unsupervised Cross-lingual Representation Learning at Scale. *arXiv:cs.CL/1911.02116*
- [9]. Allyson Ettinger. 2019. What BERT is not: Lessons from a new suite of psycholinguistic diagnostics for language models. *arXiv: 1907.13528 [cs]*. DOI: https://doi.org/10.1162/tacl_a_00298
- [10]. Alsentzer, Emily and Murphy, John R. and Boag, Willie and Weng, Wei-Hung and Jin, Di and Naumann, Tristan and McDermott, Matthew B. A., *Computation and Language (cs.CL)*, FOS: Computer and information sciences, FOS: Computer and information sciences, Publicly Available Clinical BERT Embeddings
- [11]. Andriy Mnih and Geoffrey E Hinton. 2009. A scalable hierarchical distributed language model. In D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, editors, *Advances in Neural Information Processing Systems 21*, pages 1081–1088. Curran Associates, Inc.

- [12]. Anna Rogers, Olga Kovaleva, Anna Rumshisky; *A Primer in BERTology: What We Know About How BERT Works*. *Transactions of the Association for Computational Linguistics* 2020; 8 842–866. doi: https://doi.org/10.1162/tacl_a_00349
- [13]. Automatic extraction of gene-disease associations from literature using joint ensemble learning, Bhasuran B, Natarajan J, *PLoS One*. 2018; 13(7):e0200699.
- [14]. Balaji, M., & Yuvaraj, N. (2019, November). *Intelligent Chatbot Model to Enhance the Emotion Detection in social media using Bi-directional Recurrent Neural Network*. In *Journal of Physics: Conference Series* (Vol. 1362, No. 1, p. 012039). IOP Publishing.
- [15]. Bayan Abu Shawar, Eric Atwell, “Chatbots: are they really useful?”, *LDVForum* 2007.
- [16]. Bayu Setiaji, Ferry Wahyu Wibowo, “Chatbot Using A Knowledge in Database”, *7th International Conference on Intelligent Systems, Modelling and Simulation*, 2016.
- [17]. *BERT for Coreference Resolution: Baselines and Analysis*, Mandar Joshi, Omer Levy, Daniel S. Weld, Luke Zettlemoyer, 2019
- [18]. *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding*, Jacob Devlin and Ming-Wei Chang and Kenton Lee and Kristina Toutanova, 2019
- [19]. Bozic and F. Wotawa. *Plan It! Automated Security Testing Based on Planning*. In *Proceedings of the 26th IFIP International Conference on Testing Software and Systems (ICTSS'14)*, pages 48–62, September 2014
- [20]. C. Corley and R. Mihalcea, “Measuring the Semantic Similarity of Text,” in *Proceedings of the ACL workshop on Empirical Modeling of Semantic Equivalence and Entailment*, 2005, pp. 13–18.
- [21]. C. D. Manning, P. Raghavan, and H. Schütze. *Introduction to Information Retrieval*. Cambridge University Press, 2008.
- [22]. C. J. van Rijsbergen. *Information Retrieval*. Butterworths, London, UK, 2nd edition, 1979.
- [23]. Chandler May, Alex Wang, Shikha Bordia, Samuel R. Bowman, and Rachel Rudinger. 2019. *On Measuring Social Biases in Sentence Encoders*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 622–628, Minneapolis, Minnesota. Association for Computational Linguistics.
- [24]. Chemical-gene relation extraction using recursive neural network., Lim S, Kang J, *Database (Oxford)*. 2018 Jan 1; 2018():.
- [25]. Chen, H., Liu, X., Yin, D., & Tang, J. (2019). *A survey on dialogue systems: Recent advances and new frontiers*. *Acm Sigkdd Explorations Newsletter*, 19(2), pp. 25-35.
- [26]. Chicco, D., Jurman, G. The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics* 21, 6 (2020). <https://doi.org/10.1186/s12864-019-6413-7>, 21, Pages 1135-1142, ISSN 1529-9430, <https://doi.org/10.1016/j.spinee.2021.02.007>.
- [27]. Colby KM, Weber S, Hilf FD. Artificial paranoia. *Artif. Intell.* 1971;2:1–25.
- [28]. CollaboNet: collaboration of deep neural networks for biomedical named entity recognition, Yoon W, So CH, Lee J, Kang J, *BMC Bioinformatics*. 2019 May 29; 20(Suppl 10):249

- [29]. *Contract Discovery: Dataset and a Few-Shot Semantic Retrieval Challenge with Competitive Baselines* Łukasz Borchmann, Dawid Wiśniewski, Andrzej Gretkowski, Izabela Kosmala, Dawid Jurkiewicz, Łukasz Szalkiewicz, Gabriela Palka, Karol Kaczmarek, Agnieszka Kaliska, Filip Graliński
- [30]. *Cross-type biomedical named entity recognition with deep multi-task learning.*, Wang X, Zhang Y, Ren X, Zhang Y, Zitnik M, Shang J, Langlotz C, Han J, *Bioinformatics*. 2019 May 15; 35(10):1745-1752.
- [31]. D. Inupakutika, M. Nadim, G. R. Gunnam, S. Kaghyan, D. Akopian, “Integration of NLP and Speech-to-text Applications with Chatbots “, *The University of Texas at San Antonio; San Antonio, P. Chalela, A. G. Ramirez; University of Texas Health Science Center at San Antonio, TX, USA*, 2021
- [32]. *Deep learning with word embeddings improves biomedical named entity recognition.*, Habibi M, Weber L, Neves M, Wiegandt DL, Leser U, *Bioinformatics*. 2017 Jul 15; 33(14):i37-i48.
- [33]. Dell Zhang, Jun Wang, and Xiaoxue Zhao. 2015. Estimating the Uncertainty of Average F1 Scores. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval (ICTIR '15)*. Association for Computing Machinery, New York, NY, USA, 317–320. DOI:<https://doi.org/10.1145/2808194.2809488>
- [34]. Divya, S., et al. "A self-diagnosis medical chatbot using artificial intelligence. " *Journal of Web Development and Web Designing* 3.1 (2018): 1-7.
- [35]. DocBERT: BERT for Document Classification, Ashutosh Adhikari, Achyudh Ram, Raphael Tang, Jimmy Lin, 2019
- [36]. Dr. Sbaitso (2019). Wikipedia. https://en.wikipedia.org/w/index.php?title=Dr._Sbaitso&oldid=922071070.
- [37]. Eleni Adamopoulou and Lefteris Moussiades, "Chatbots: History, technology, and applications", *Machine Learning with Applications*, 2020, 100006
- [38]. *Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation*, David M. W. Powers, 2020
- [39]. Florian Schmidt and Thomas Hofmann. 2020. BERT as a Teacher: Contextual Embeddings for Sequence-Level Reward. *arXiv preprint arXiv:2003.02738*.
- [40]. Følstad, A., & Skjuve, M. (2019, August). Chatbots for customer service: user experience and motivation. In *Proceedings of the 1st International Conference on Conversational User Interfaces* (pp. 1-9).
- [41]. G. Salton and C. Buckley, "Term-weighting Approaches in Automatic Text Retrieval," *Information Processing and Management*, vol.24, no.5, 1988, pp.513–523.
- [42]. Grazi Murtarelli, Anne Gregory Stefania Romenti, "A conversation-based perspective for shaping ethical human-machine interactions: The particular challenge of chatbots.", *Journal of Business Research*, 2021, 927-935
- [43]. Gregor Wiedemann, Steffen Remus, Avi Chawla, and Chris Biemann. 2019. Does BERT Make Any Sense? Interpretable Word Sense Disambiguation with Contextualized Embeddings. *arXiv preprint arXiv:1909.10430*.
- [44]. Grudin, J., & Jacques, R. (2019, May). Chatbots, humbots, and the quest for artificial general intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in*

Computing Systems (pp. 1-11).

- [45]. Guokun Lai, Qizhe Xie, Hanxiao Liu, Yiming Yang, and Eduard Hovy. 2017. Race: Large-scale reading comprehension dataset from examinations. *arXiv preprint arXiv:1704.04683*.
- [46]. Hoy, M. (2018). Alexa, siri, cortana, and more: An introduction to voice assistants. *Medical Reference Services Quarterly*, 37, 81–88. <http://dx.doi.org/10.1080/02763869.2018.1404391>.
- [47]. Ian Tenney, PatrickXia, BerlinChen, AlexWang, AdamPoliak, R.Thomas McCoy, NajoungKim, Benjamin VanDurme, Samuel R.Bowman, DipanjanDas, and ElliePavlick. 2019b. What do you learn from context? Probing for sentence structure in contextualized word representations. In *International Conference on Learning Representations*.
- [48]. IMPROVING EMPLOYEE ENGAGEMENT WITH THE HELP OF ENGY CHATBOT, Rozenberg, Aleksei
- [49]. J Briskilal, C.N. Subalalitha, An ensemble model for classifying idioms and literal texts using BERT and RoBERTa, *Information Processing & Management*, Volume 59, Issue 1, 2022, 102756, ISSN 0306-4573
- [50]. Jaana Parviainen & Juho Rantala, “Chatbot breakthrough in the 2020s? An ethical reflection on the trend of automated consultations in health care.”, 2021, *Medicine, Health Care and Philosophy*
- [51]. Jabberwacky (2019). Wikipedia. Retrieved from <https://en.wikipedia.org/w/index.php?title=Jabberwacky&oldid=926273345>.
- [52]. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *North American Association for Computational Linguistics (NAACL)*.
- [53]. Jeff Da and JungoKasai. 2019. Cracking the Contextual Commonsense Code: Understanding Commonsense Reasoning Aptitude of Deep Contextual Representations. In *Proceedings of the First Workshop on Commonsense Inference in Natural Language Processing*, pages 1–12, Hong Kong, China. Association for Computational Linguistics.
- [54]. Joe Davison, JoshuaFeldman, and AlexanderRush. 2019. Commonsense Knowledge Mining from Pretrained Models. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 1173–1178, Hong Kong, China. Association for Computational Linguistics. DOI: <https://doi.org/10.18653/v1/D19-1109>
- [55]. John Koutsikakis, Ilias Chalkidis, Prodromos Malakasiotis, and Ion Androutsopoulos. 2020. GREEK-BERT: The Greeks visiting Sesame Street. In *11th Hellenic Conference on Artificial Intelligence (SETN 2020)*. Association for Computing Machinery, New York, NY, USA, 110–117. <https://doi.org/10.1145/3411408.3411440>
- [56]. Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25. <https://doi.org/10.1016/j.bushor.2018.08.004>
- [57]. Kawin Ethayarajh. 2019. How Contextual are Contextualized Word Representations? Comparing the Geometry of BERT, ELMo, and GPT-2 Embeddings. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 55–65, Hong Kong, China. Association for Computational Linguistics. DOI:

- <https://doi.org/10.18653/v1/D19-1006>
- [58]. Kyle Richardson and AshishSabharwal. 2019. *What Does My QA Model Know? Devising Controlled Probes using Expert Knowledge*. arXiv:1912.13337 [cs]. DOI: https://doi.org/10.1162/tacl_a_00331
- [59]. Lalwani, Tarun, et al. "Implementation of a Chat Bot System using AI and NLP." *International Journal of Innovative Research in Computer Science & Technology (IJIRCST)* 6.3 (2018): 26-30.
- [60]. Lee J, Yoon W, Kim S, Kim D, Kim S, So CH, Kang J. BioBERT: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics*. 2020 Feb 15;36(4):1234-1240. doi: 10.1093/bioinformatics/btz682. PMID: 31501885; PMCID: PMC7703786.
- [61]. Lingpeng Kong, Cyprien de Masson d'Autume, LeiYu, WangLing, ZihangDai, and DaniYogatama. 2019. *A mutual information maximization perspective of language representation learning*. In *International Conference on Learning Representations*
- [62]. Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D. et al. (2019). *Roberta: A robustly optimized bert pretraining approach*. arXiv preprint arXiv:1907.11692.
- [63]. Manning, C., Raghavan, P., Schütze, H.: *Introduction to information retrieval*, vol. 1. Cambridge University Press (2008)
- [64]. Martin Hasal, Jana Nowaková, Khalifa Ahmed Saghair, Hussam Abdulla, Václav Snášel, Lidia Ogiela, "Chatbots: Security, privacy, data protection, and social aspects.", *Concurrency and Computation: Practice and Experience*, 2021, 6426
- [65]. Mauldin, M. L. (1994). *CHATTERBOTS, TINYMUDS, and the turing test: entering the loebner prize competition*. AAAI.
- [66]. Maxwell Forbes, AriHoltzman, and YejinChoi. 2019. *Do Neural Language Representations Learn Physical Commonsense?* In *Proceedings of the 41st Annual Conference of the Cognitive Science Society (CogSci 2019)*, page 7
- [67]. Mengis, J., & Eppler, M. J. (2008). *Understanding and managing conversations from a knowledge perspective: An analysis of the roles and rules of face-to-face conversations in organizations*. *Organization Studies*, 29(10), 1287–1313. <https://doi.org/10.1177/0170840607086553>
- [68]. Molnár, G., Zoltán, S.: *The role of chatbots in formal education*. Presented at the 15 September 2018
- [69]. Nelson F. Liu, MattGardner, YonatanBelinkov, Matthew E.Peters, and Noah A.Smith. 2019a. *Linguistic Knowledge and Transferability of Contextual Representations*. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 1073–1094, Minneapolis, Minnesota. Association for Computational Linguistics.
- [70]. Nina Poerner, UlliWaltinger, and HinrichSchütze. 2019. *BERT is not a knowledge base (yet): Factual knowledge vs. name-based reasoning in unsupervised qa*. arXiv preprint arXiv:1911.03681.
- [71]. P. R. M. Odel, T. O. A. Nalyse, C.-C. O. On, M. Müller, and M. Salathe, "bert: a n," 2020

- [72]. Palanica, A., P. Flaschner, A. Thommandram, M. Li, and Y. Fossat. 2019. Physicians' perceptions of chatbots in health care: cross-sectional web-based survey. *Journal of Medical Internet Research* 21 (4): e12887.
- [73]. Parvatiyar, A., & Sheth, J. N. (2001). Customer relationship management: Emerging practice, process, and discipline. *Journal of Economic & Social Research*, 3(2), 1–34.
- [74]. Pasquale, F. 2020. *New laws of robotics. Defending human expertise in the age of AI*. Cambridge: Harvard University Press.
- [75]. Peipei Xia, Li Zhang, Fanzhang Li, Learning similarity with cosine similarity ensemble, *Information Sciences*, Volume 307, 2015, Pages 39-52, ISSN 0020-0255
- [76]. Phu Mon Htut, JasonPhang, ShikhaBordia, and Samuel R.Bowman. 2019. Do attention heads in BERT track syntactic dependencies?arXiv preprint arXiv:1911.12246.
- [77]. Powell, J. 2019. Trust Me, I'm a Chatbot: How artificial intelligence in health care fails the turing test. *Journal of Medical Internet Research* 21 (10): e16222.
- [78]. Pranav Rajpurkar, Jian Zhang, Konstantin Lopyrev, and Percy Liang. 2016. Squad: 100,000+ questions for machine comprehension of text. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 2383–2392.
- [79]. PubMed: The Bibliographic Database by Kathi Canese, Jennifer Jentsch, and Carol Myers
- [80]. Pyysalo S. et al. (2013) Distributional semantics resources for biomedical text processing. In: *Proceedings of the 5th International Symposium on Languages in Biology and Medicine*, Tokyo, Japan, pp. 39–43. <https://academic.oup.com/bioinformatics/article/33/14/i37/3953940>.
- [81]. Qiao Jin and Zheng Yuan and Guangzhi Xiong and Qianlan Yu and Huaiyuan Ying and Chuanqi Tan and Mosha Chen and Songfang Huang and Xiaozhong Liu and Sheng Yu, "Biomedical Question Answering: A Survey of Approaches and Challenges", 2021
- [82]. Rade Matic 1, Milos Kabiljo 1,2, Miodrag Zivkovic 2, and Milan Cabarkapa 3, "Extensible Chatbot Architecture Using Metamodels of Natural Language Understanding.", *Electronics*, 2021, 18
- [83]. Rajpurkar, Pranav and Zhang, Jian and Lopyrev, Konstantin and Liang, Percy, *Computation and Language (cs.CL)*, FOS: Computer and information sciences, FOS: Computer and information sciences, SQuAD: 100,000+ Questions for Machine Comprehension of Text, 2016
- [84]. Ralyté J., B. Roelens, and S. Demeyer (Eds.):DoEM 2017 Doctoral Consortium and Industry Track Papers, pp. 70-78, 2017.
- [85]. Rishi Bommasani, KellyDavis, and ClaireCardie. 2020. Interpreting Pretrained Contextualized Representations via Reductions to Static Embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4758–4781. DOI: <https://doi.org/10.18653/v1/2020.acl-main.431>
- [86]. Rogers, Anna and Kovaleva, Olga and Rumshisky, Anna, *A Primer in BERTology: What We Know About How BERT Works*, *Transactions of the Association for Computational Linguistics*, 842-866, 2021
- [87]. Samuel R. Bowman, Gabor Angeli, Christopher Potts, and Christopher D. Manning. 2015. A large annotated corpus for learning natural language inference. In *EMNLP. Association for Computational Linguistics*.

- [88]. *savchukquestion, Question Answering with SQuAD v2. 0, Savchuk, Vladislav and Shults, Vyacheslav and Galeev, Farit*
- [89]. *Seering, J., Flores, J. P., Savage, S., & Hammer, J. (2018). The social roles of bots: evaluating impact of bots on discussions in online communities. Proceedings of the ACM on Human-Computer Interaction, 2(CSCW), 1-29.*
- [90]. *Semantic Cosine Similarity Faisal Rahutomo*, Teruaki Kitasuka, and Masayoshi Aritsugi Graduate School of Science and Technology, Kumamoto University*
- [91]. *Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. Inf Process Manage 45:427-437*
- [92]. *Taiar, Redha, Qasim, Rukhma, Bangyal, Waqas Haider, Alqarni, Mohammed A.Ali Almazroi, Abdulwahab, 2022, A Fine-Tuned BERT-Based Transfer Learning Approach for Text Classification.*
- [93]. *Takahashi, K., Yamamoto, K., Kuchiba, A. et al. Confidence interval for micro-averaged F1 and macro-averaged F1 scores. Appl Intell (2021). <https://doi.org/10.1007/s10489-021-02635-5>*
- [94]. *Tan, S.: Neighbor-weighted k-nearest neighbor for unbalanced text corpus. Expert Systems with Applications 28, 667-671 (2005)*
- [95]. *Thresholding Classifiers to Maximize F1 Score, Zachary Chase Lipton and Charles Elkan and Balakrishnan Narayanaswamy, 2014, 1402.1892*
- [96]. *Timothee Mickus, Denis Paperno, Mathieu Constant, and Kees van Deemeter. 2019. What do you mean, BERT? assessing BERT as a distributional semantics model. arXiv preprint arXiv:1911.05758.*
- [97]. *Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S. Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In Advances in Neural Information Processing Systems 26 (NIPS 2013), pages 3111-3119.*
- [98]. *Transfer learning for biomedical named entity recognition with neural networks., Giorgi JM, Bader GD, Bioinformatics. 2018 Dec 1; 34(23):4087-4094.*
- [99]. *Triyasha Dastidarb and Kaushik Roy, "NLP is Not enough - Contextualization of User Input in Chatbots Nathan Dolbira", 2021*
- [100]. *Turing AM. Computing machinery and intelligence. Mind. 1950;59:433-460.*
- [101]. *UBIS: Unigram Bigram Importance Score for Feature Selection from Short Text, Expert Systems with Applications, Volume 195, 2022, 116563, ISSN 0957-4174.*
- [102]. *V.C. Bhavsar, H. Boley, L. Yang, "A Weighted-Tree Similarity Algorithm for Multi-agent System in E-Business Environments," Computational Intelligence, vol. 20, no. 4, 2004, pp. 584-602.*
- [103]. *Vaibhav Joshi, Atharv Khare, Rahul Gawali, Asmita Manna, "A Comparative Study of Medical Chatbots Jitendra Chaudhary", International Research Journal of Engineering and Technology (IRJET), 2021*
- [104]. *Vaira, Lucia, et al. "MamaBot: a System based on ML and NLP for supporting Women and Families during Pregnancy." Proceedings of the 22nd International Database Engineering & Applications Symposium. 2018*

- [105]. Vargo, S. L., & Akaka, M. A. (2009). *Service-dominant logic as a foundation for servescience: Clarifications*. *Service Science*, 1(1), 32–41. <https://doi.org/10.1287/serv.1.1.32>
- [106]. Wallace RS. *The anatomy of A.L.I.C.E.* In: Epstein R, Roberts G, Beber G, editors. *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer*. Cham: Springer; 2009. pp. 181–210
- [107]. Wang L, Chu F, Xie W. *Accurate cancer classification using expressions of very few genes*. *IEEE/ACM Trans Comput Biol Bioinforma*. 2007;4(1): 40–53]
- [108]. Watson assistant | ibm cloud. (2020). <https://www.ibm.com/cloud/watson-assistant/>. (Retrieved 29 February 2020).
- [109]. Weizenbaum J. *ELIZA—a computer program for the study of natural language communication between man and machine*. *Commun. ACM*. 1966;9:36–45.
- [110]. West, S. M. (2019). *Data capitalism: Redefining the logics of surveillance and privacy*. *Business & Society*, 58(1), 20–41. Westerman, D., Cross, A. C., & Lindmark, P. G. (2019).
- [111]. *What Does BERT Look At? An Analysis of BERT's Attention*, Kevin Clark and Urvashi Khandelwal and Omer Levy and Christopher D. Manning, 2019
- [112]. Wiese G. et al. (2017) *Neural domain adaptation for biomedical question answering*. In: *Proceedings of the 21st Conference on Computational Natural Language Learning (CoNLL 2017)*, Vancouver, Canada pp. 281–289. Association for Computational Linguistics. <https://www.aclweb.org/anthology/K17-1029>.
- [113]. Wilson, H. J., & Daugherty, P. R. (2018). *Collaborative intelligence: Humans and AI are joining forces*. *Harvard Business Review*. Available at <https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces> Accessed 12 January 2019
- [114]. Wizu(2018). *A visual history of chatbots*. Medium website: <https://chatbotsmagazine.com/a-visual-history-of-chatbots-8bf3b31dbfb2>. (Retrieved 24 February 2020).
- [115]. Xiaodong Liu, Pengcheng He, Weizhu Chen, and Jianfeng Gao. 2019b. *Multi-task deep neural networks for natural language understanding*. *arXiv preprint arXiv:1901.11504*
- [116]. Xuhui Zhou, YueZhang, LeyangCui, and DandanHuang. 2020. *Evaluating Commonsense in Pre-Trained Language Models*. In *AAAI 2020*. DOI: <https://doi.org/10.1609/aaai.v34i05.6523>
- [117]. Y. Goldberg, “Assessing BERT’s syntactic abilities,” pp. 2–5, 2018
- [118]. Y. Yang and X. Liu. *A re-examination of text categorization methods*. In *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR)*, pages 42–49, Berkeley, CA, USA, 1999.
- [119]. Yile Wang, LeyangCui, and YueZhang. 2020d. *How Can BERT Help Lexical Semantics Tasks?* *arXiv:1911.02929 [cs]*.
- [120]. Yonghui Wu, MikeSchuster, ZhifengChen, Quoc V.Le, MohammadNorouzi, WolfgangMacherey, MaximKrikun, YuanCao, QinGao, KlausMacherey, et al 2016. *Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation*.
- [121]. Yongjie Lin, YiChern Tan, and RobertFrank. 2019. *Open Sesame: Getting inside BERT’s Linguistic Knowledge*. In *Proceedings of the 2019 ACL Workshop BlackboxNLP: Analyzing and*

- Interpreting Neural Networks for NLP*, pages 241–253.
- [122]. Zhilin Yang, Zihang Dai, Yiming Yang, Jaime Carbonell, Ruslan Salakhutdinov, and Quoc V Le. 2019. *Xlnet: Generalized autoregressive pretraining for language understanding*. *arXiv preprint arXiv:1906.08237*
- [123]. Zhiyong Wu, YunChen, BenKao, and QunLiu. 2020. *Perturbed Masking: Parameter-free Probing for Analyzing and Interpreting BERT*. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4166–4176, Online. Association for Computational Linguistics.
- [124]. Ming Zhu, Aman Ahuja, Wei Wei, and Chandan K. Reddy. 2019. *A Hierarchical Attention Retrieval Model for Healthcare Question Answering*. In *The World Wide Web Conference (San Francisco, CA, USA) (WWW '19)*. Association for Computing Machinery, New York, NY, USA, 2472–2482
- [125]. Tony Russell-Rose and Jon Chamberlain. 2017. *Expert search strategies: the information retrieval practices of healthcare information professionals*. *JMIR medical informatics* 5, 4 (2017), e33
- [126]. Yun Niu, Graeme Hirst, Gregory McArthur, and Patricia Rodriguez-Gianolli. 2003. *Answering Clinical Questions with Role Identification*. In *Proceedings of the ACL 2003 Workshop on Natural Language Processing in Biomedicine*. Association for Computational Linguistics, Sapporo, Japan
- [127]. Jinhyuk Lee, Wonjin Yoon, Sungdong Kim, Donghyeon Kim, Sunkyu Kim, Chan Ho So, and Jaewoo Kang. 2020. *BioBERT: a pre-trained biomedical language representation model for biomedical text mining*. *Bioinformatics* 36, 4 (2020), 1234–1240