



Δράση «Εμβληματικές δράσεις σε διαθεματικές επιστημονικές περιοχές με ειδικό ενδιαφέρον για την σύνδεση με τον παραγωγικό ιστό» ID 16618

Εθνικό δίκτυο έρευνας για την ανάδειξη της γενετικής βάσης των νευροεκφυλιστικών νόσων Alzheimer και Parkinson, την ανίχνευση αξιόπιστων βιοδεικτών και την ανάπτυξη καινοτόμων υπολογιστικών τεχνολογιών και θεραπευτικών στρατηγικών στη βάση της ιατρικής ακριβείας (BRAIN PRECISION, TAEDR-0535850)

ΤΙΤΛΟΣ ΠΑΡΑΔΟΤΕΟΥ: Παροχή βοήθειας για την χρήση του FEGA μέσω εξειδικευμένου Helpdesk για τους συνεργάτες της εμβληματικής δράσης

ΕΝΟΤΗΤΑ ΕΡΓΑΣΙΑΣ 2: Καταγραφή πασχόντων και δημιουργία τράπεζας βιολογικού υλικού των νευροεκφυλιστικών νόσων Alzheimer και Parkinson και των πρόδρομων μορφών τους

ΥΠΕΥΘΥΝΗ ΕΡΕΥΝΗΤΙΚΗ ΟΜΑΔΑ (ΦΟΡΕΑΣ): MARTIN RECZKO (ΕΚΕΒΕ “Αλέξανδρος Φλέμιγκ”)

Παροχή βοήθειας για την χρήση του FEGA μέσω εξειδικευμένου Helpdesk για τους συνεργάτες της εμβληματικής δράσης

ΠΕΡΙΓΡΑΦΗ

- LLM-based Helpdesk:** Αναπτύχθηκε σύστημα αυτοματοποιημένης τεχνικής υποστήριξης βασισμένο σε Μεγάλα Γλωσσικά Μοντέλα (LLMs) για τη διευκόλυνση των χρηστών της υποδομής.
- Ένα δημοσιευμένο **mini-tutorial** εξηγεί το κίνητρο για το FEGA ως λύση για τη συμμόρφωση με τον GDPR για ευαίσθητα ανθρώπινα δεδομένα

ΠΑΡΑΔΟΤΕΑ

A. Υλοποίηση Αυτοματοποιημένου Helpdesk με χρήση Τεχνητής Νοημοσύνης

Το **ELIXIR FEGA** αποτελεί μια σύνθετη υποδομή για την αποθήκευση και διαμοίραση ευαίσθητων βιοϊατρικών δεδομένων. Η πολυπλοκότητα των διαδικασιών (υποβολή δεδομένων, κρυπτογράφηση, διαχείριση αδειών πρόσβασης) δημιουργεί την ανάγκη για συνεχή και εξειδικευμένη υποστήριξη χρηστών. Προκειμένου να βελτιωθεί η ταχύτητα απόκρισης και να μειωθεί ο φόρτος της χειροκίνητης υποστήριξης, υλοποιήθηκε ένα **αυτοματοποιημένο σύστημα Helpdesk** βασισμένο σε τεχνολογίες αιχμής.

Αντί για ένα απλό chatbot, επιλέχθηκε η αρχιτεκτονική **Retrieval-Augmented Generation (RAG)**, η οποία επιτρέπει σε ένα Μεγάλο Γλωσσικό Μοντέλο (LLM) να απαντά με βάση συγκεκριμένες, έγκυρες πηγές πληροφοριών.

- Το Μοντέλο (LLM):** Χρησιμοποιήθηκε το μοντέλο **Gemma2**, το οποίο προσφέρει υψηλή απόδοση στην κατανόηση τεχνικών κειμένων και μπορεί να φιλοξενηθεί σε ιδιωτικές υποδομές, διασφαλίζοντας την προστασία της πληροφορίας.
- Τροφοδοσία Δεδομένων (Knowledge Base):** Το σύστημα τροφοδοτήθηκε με το επίσημο documentation, οδηγούς χρήσης και τεχνικά έγγραφα του ELIXIR FEGA.
- Μηχανισμός RAG:** Όταν ένας χρήστης υποβάλλει μια ερώτηση, το σύστημα αναζητά τα πιο σχετικά αποσπάσματα από τα έγγραφα (Retrieval) και στη συνέχεια το Llama συνθέτει μια ακριβή και κατανοητή απάντηση (Generation) βασισμένη αποκλειστικά σε αυτά τα δεδομένα.

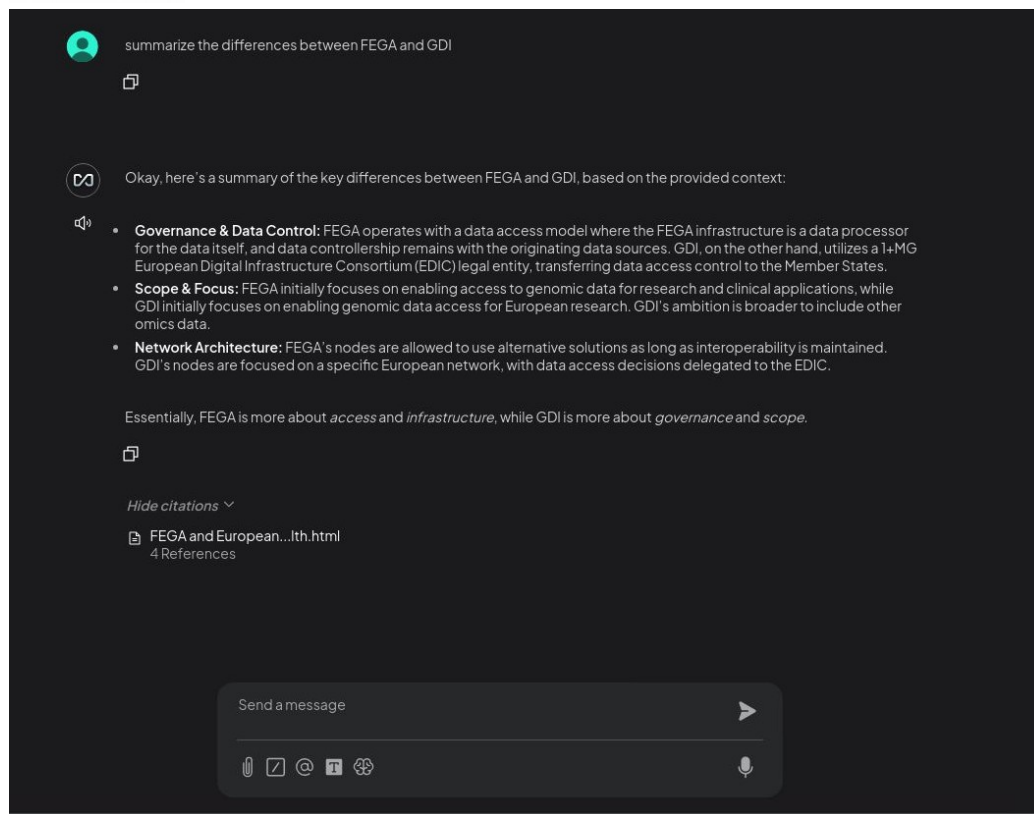
Πλεονεκτήματα της Προσέγγισης

- Διαφάνεια, Ακρίβεια και Αξιοπιστία:** Κάθε απάντηση έχει υποδείξεις στα πρωτότυπα έγγραφα που περιέχουν τις σχετικές λεπτομέρειες. Η χρήση του RAG εξαλείφει τις πιθανότητες εμφάνισης «παραισθήσεων» (hallucinations) μοντέλου, καθώς οι απαντήσεις περιορίζονται στο πλαίσιο επίσημων εγγράφων.
- Άμεση Υποστήριξη 24/7:** Οι ερευνητές λαμβάνουν απαντήσεις για περίπλοκα ζητήματα (π.χ. «Πώς λειτουργεί το crypt4gh;») σε δευτερόλεπτα, οποιαδήποτε στιγμή. Οι απαντήσεις δημιουργούνται μέσα σε λίγα δευτερόλεπτα, εξαλείφοντας την ανάγκη για ticketing σύστημα.
- Επεκτασιμότητα:** Το σύστημα μπορεί εύκολα να επικαιροποιηθεί προσθέτοντας νέα έγγραφα ή οδηγίες, χωρίς την ανάγκη επανεκπαίδευσης του μοντέλου.

Συμπέρασμα

Η ενσωμάτωση ενός LLM-based Helpdesk στο οικοσύστημα του έργου *Brain Precision* αποδεικνύει πώς η Τεχνητή Νοημοσύνη μπορεί να λειτουργήσει ως επιταχυντής για την ανοιχτή επιστήμη. Διευκολύνοντας την πρόσβαση στην πληροφορία, μειώνοντας τα τεχνικά εμπόδια για τους ερευνητές, ενισχύοντας παράλληλα την αποδοτικότητα της υποδομής.

Στην εικόνα που ακολουθεί φαίνεται η απάντηση σε μια ερώτηση χρήστη για τον εντοπισμό διαφορών μεταξύ FEGA και GDI. Εκτός της απάντησης, δίνονται στον τελικό χρήστη και τα citation που οδήγησαν στην διαμόρφωση της συγκεκριμένης απάντησης.



Β. Συμμόρφωση με τον GDPR και το Ομοσπονδιακό EGA: Κίνητρα για Ερευνητές στις Βιοεπιστήμες

Στην συγκεκριμένη αναφορά εξηγείται ότι τα ανθρώπινα γονιδιωματικά δεδομένα είναι ιδιαίτερα ευαίσθητα και αυστηρά προστατευμένα από τον GDPR, γεγονός που δυσκολεύει τη διεθνή ερευνητική συνεργασία. Παρουσιάζεται το FEGA ως μια ομοσπονδιακή λύση που επιτρέπει την ασφαλή αποθήκευση δεδομένων εντός εθνικών συνόρων και την ελεγχόμενη παγκόσμια πρόσβαση μέσω μεταδεδομένων, διευκολύνοντας τη συμμόρφωση και τη συνεργασία. <https://doi.org/10.5281/zenodo.18553028>

ΑΝΑΛΥΣΗ ΤΗΣ ΠΟΡΕΙΑΣ ΥΛΟΠΟΙΗΣΗΣ ΤΗΣ ΕΕ

Στο πλαίσιο της ΕΕ2 αναπτύχθηκε ένα καινοτόμο σύστημα αυτοματοποιημένης τεχνικής υποστήριξης βασισμένο σε Μεγάλα Γλωσσικά Μοντέλα (LLMs) και αρχιτεκτονική Retrieval-Augmented Generation (RAG). Χρησιμοποιήθηκε το μοντέλο Gemma2, το οποίο τροφοδοτήθηκε με την επίσημη τεχνική τεκμηρίωση του ELIXIR FEGA. Το σύστημα παρέχει άμεσες απαντήσεις σε τεχνικά ερωτήματα των χρηστών, συνοδευόμενες από παραπομπές στις αρχικές πηγές πληροφοριών για διασφάλιση της ακρίβειας. Παράλληλα, εκπονήθηκε εξειδικευμένο mini-tutorial με τίτλο «Συμμόρφωση με τον GDPR και το Ομοσπονδιακό EGA: Κίνητρα για Ερευνητές στις Βιοεπιστήμες». Η αναφορά αυτή αναλύει την πολυπλοκότητα της διαχείρισης των ιδιαίτερα ευαίσθητων γονιδιωματικών δεδομένων υπό το πρίσμα του GDPR και παρουσιάζει το FEGA ως την ιδανική ομοσπονδιακή λύση για την υπέρβαση των εμποδίων στη διεθνή συνεργασία.

Αναμενόμενα Αποτελέσματα – Οφέλη Κάθε Φορέα

Για το Φλέμιγκ: Η ανάπτυξη του LLM-based Helpdesk τοποθετεί το Φλέμιγκ στην αιχμή της εφαρμογής Τεχνητής Νοημοσύνης για την υποστήριξη βιοπληροφορικών υποδομών, μειώνοντας δραστικά τον φόρτο της χειροκίνητης τεχνικής υποστήριξης.

Για τους Λοιπούς Φορείς: Οι ερευνητές του δικτύου έχουν πλέον πρόσβαση σε άμεση (24/7) υποστήριξη για πολύπλοκα ζητήματα (π.χ. κρυπτογράφηση Crypt4GH). Αυτό μειώνει τα τεχνικά εμπόδια εισόδου και αυξάνει την αποδοτικότητα της χρήσης της υποδομής FEGA από όλους τους συνεργάτες. Επιπλέον, μέσω του νέου mini-tutorial, οι ερευνητές αποκτούν σαφή εικόνα για το πώς η χρήση του FEGA διευκολύνει τη συμμόρφωση με τον GDPR, επιτρέποντας την ασφαλή αποθήκευση δεδομένων εντός εθνικών συνόρων και την ελεγχόμενη παγκόσμια πρόσβαση μέσω μεταδεδομένων, προωθώντας έτσι τη διεθνή ερευνητική συνεργασία