BACHELORS THESI

CONTEXT-AWARE ASSISTIVE DIFFERENTIAL DIAGNOSIS SYSTEM

INTRODUCTION & SALIENT FEATURES

Medical professionals and patients often struggle with managing and interpreting vast amounts of medical information, like medical reports, and current symptoms for 100s of patients daily.
Our project aims to address this by developing a medical software solution that serves doctors and patients.

- Conversion of medical reports into compressed token-based embeddings to form a **patient's knowledge base.**
- On the go, **text-to-graph generator** based on reports of the user.
- **Context-aware DDx tool** for doctors, using the knowledge base and transparent reasoning for each deduction.

PROBLEM STATEMENT

Development of a medically comprehensive AI toolkit for higher practical utility

01 LACK OF CENTRALIZED KNOWLEDGE BASE

A single location where all medically relevant information and reports are stored can lead to a more comprehensive diagnosis tool.

02 POOR INTERPRETATION OF REPORTS

Most patients aren't medically trained, and thus not able to understand the reports leaves them in the dark. A visual report generator can help.

03 DOCTOR-SPECIFIC TOOL

Most AI models are trying to help the patient, but considering the sheer number of patients a doctor has to handle. We need to increase medical diagnosis efficiency.

04 IRRELEVANT INFORMATION

A lot of times patients don't share everything, as they might consider a stomachache that happened 2 weeks ago as irrelevant. These small "irrelevant" details can be crucial in diagnosis.

OUR AIMS & OBJECTIVES

01

Creation of a medical knowledge base per patient, to keep a centralized record of their history, symptoms etc.

02

Developing a robust Text-to-Token model to extract vital medical data from reports.

03

Implementing a Text to Vitals generator, which can be used to visualize information on the go.

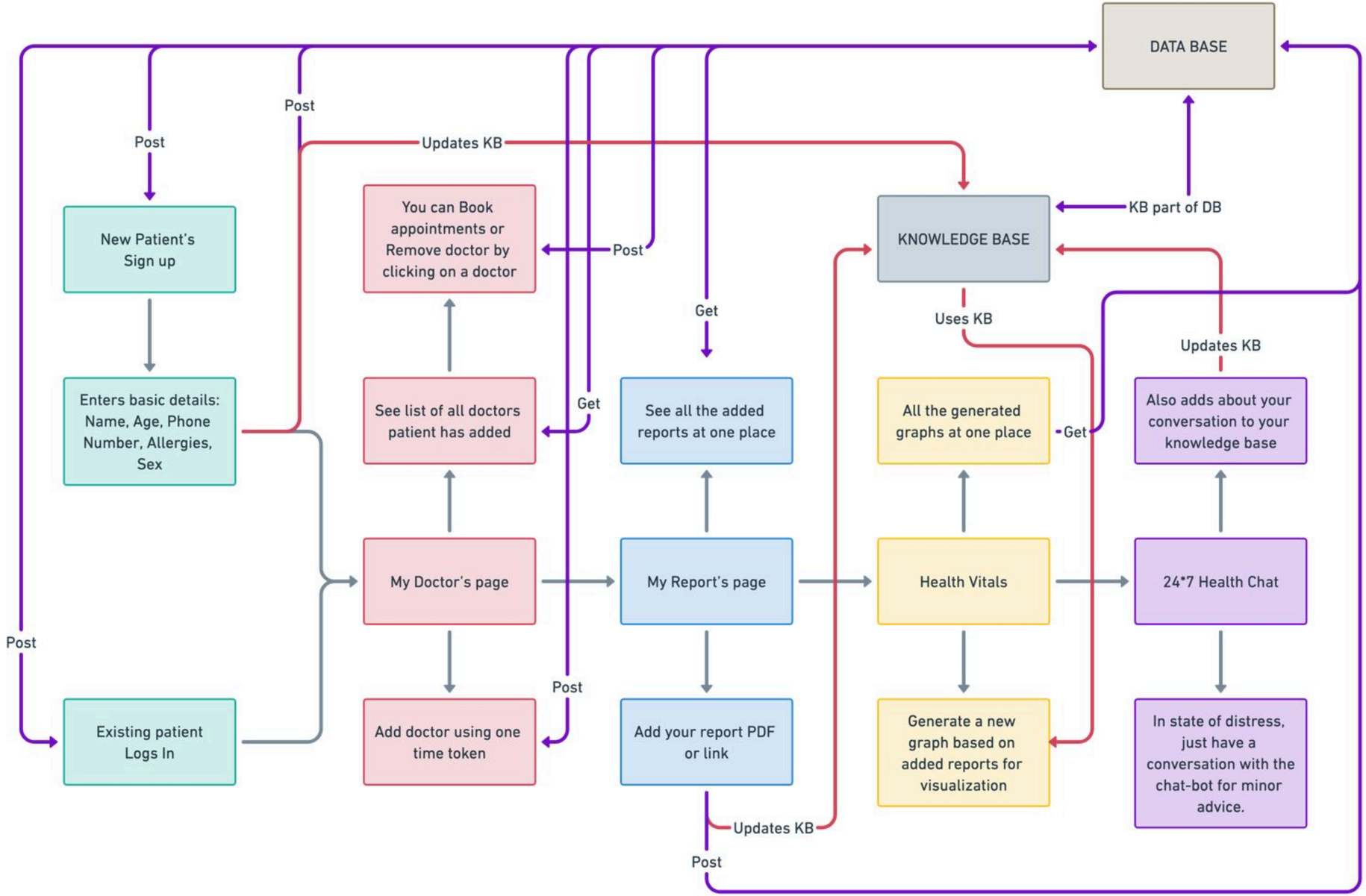
04

Creating a 24*7 Medical System which not only helps the user, but also gets us the regular day to day health status of the user, which is appended to the knowledge base.

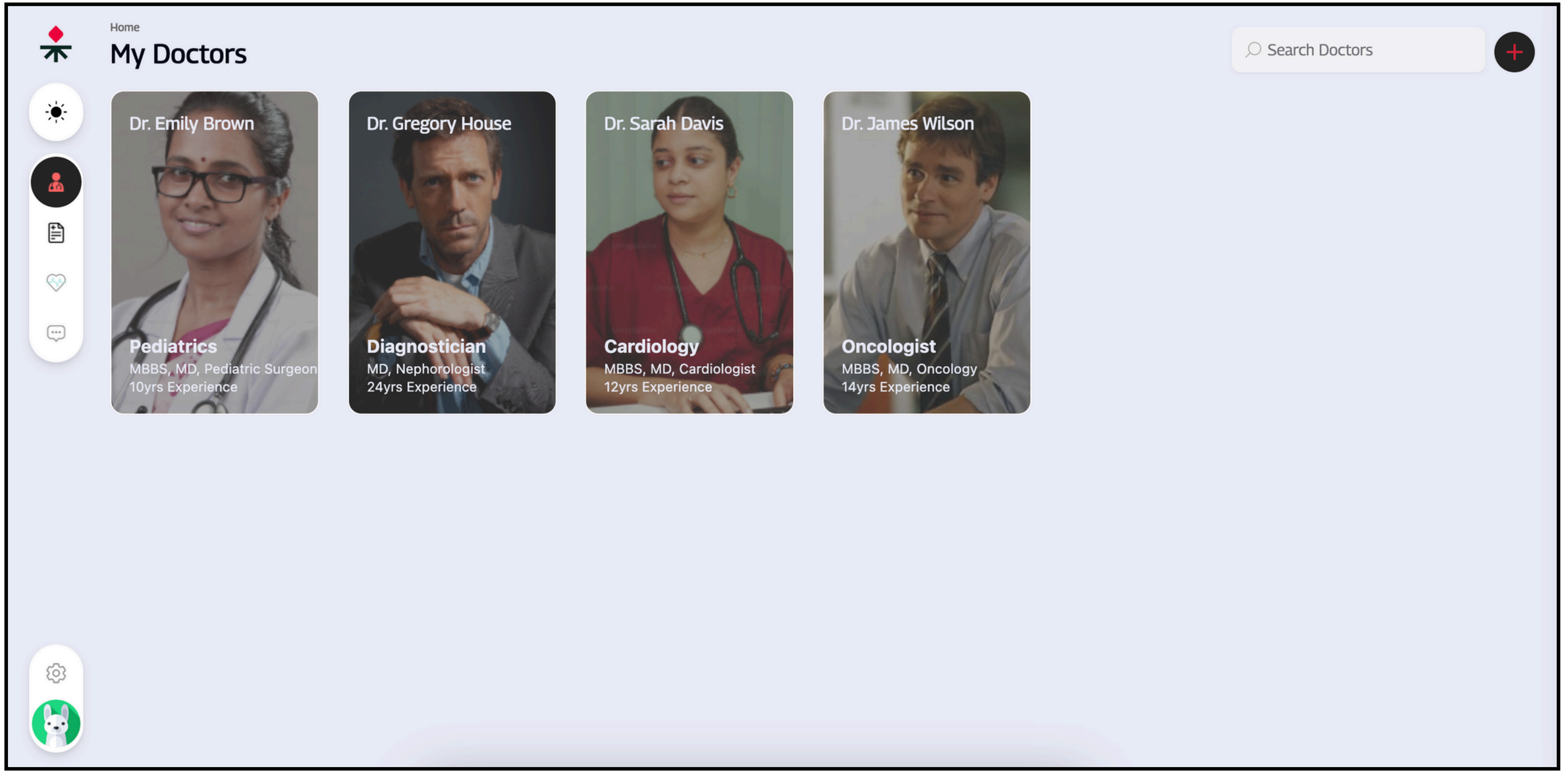
05

Making a Differential Diagnosis tool which can act as an agent to the doctor, helping them with its unmatched memory during diagnosis, saving time and effort.

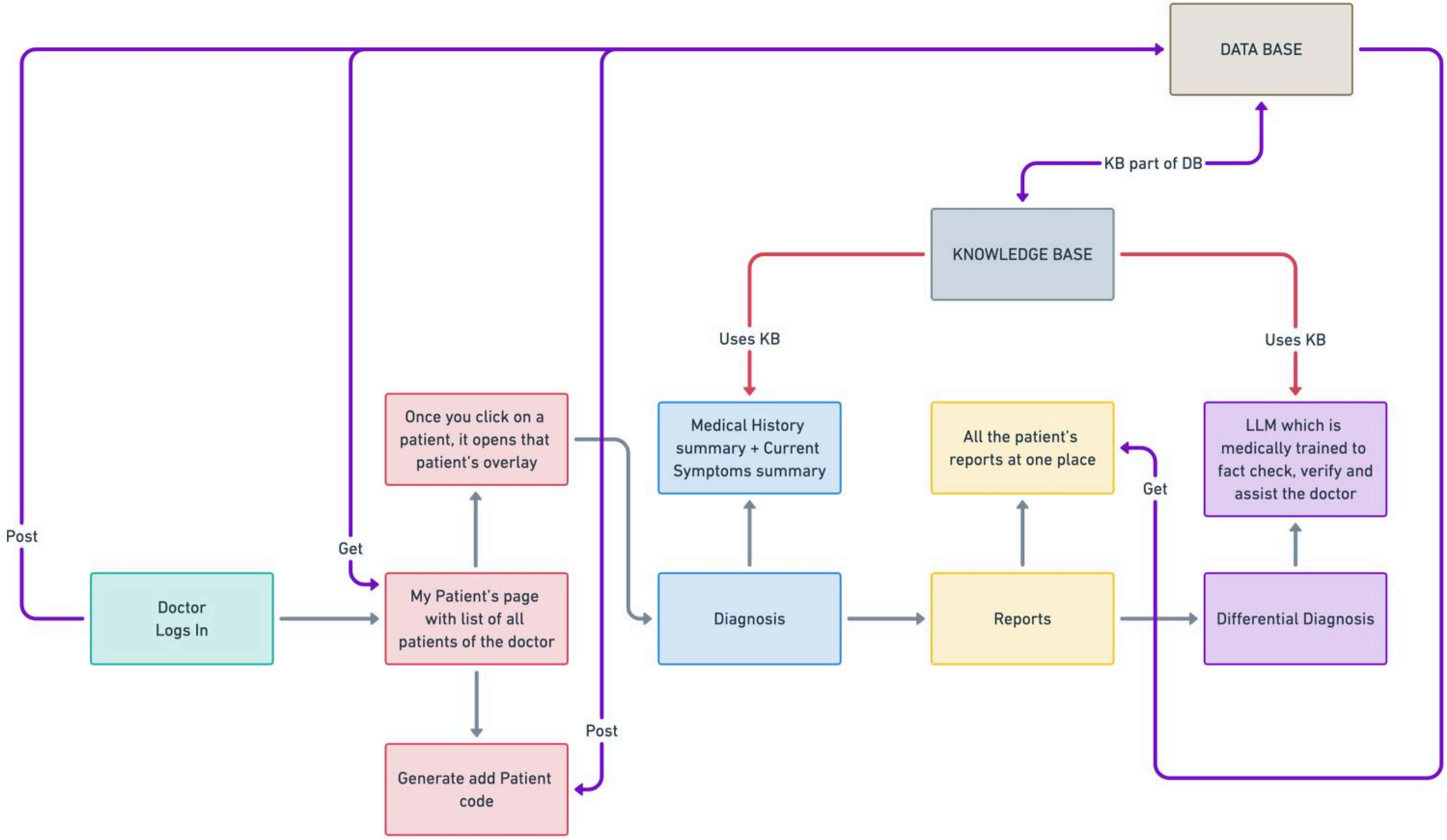
PATIENT'S WORKFLOW



PATIENT'S WORKFLOW



DOCTOR'S WORKFLOW



DOCTOR'S WORKFLOW

Home

My Patients

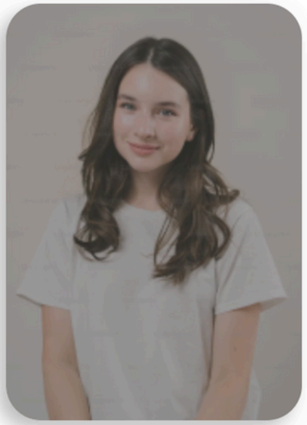
Search Patient

Janie Do

Female

22yo

Ms. Khushi Shah



Reports | Diagnosis | Health Vitals

Sex: Female
Age: 22
Condition: Lupus
Blood Group: O-

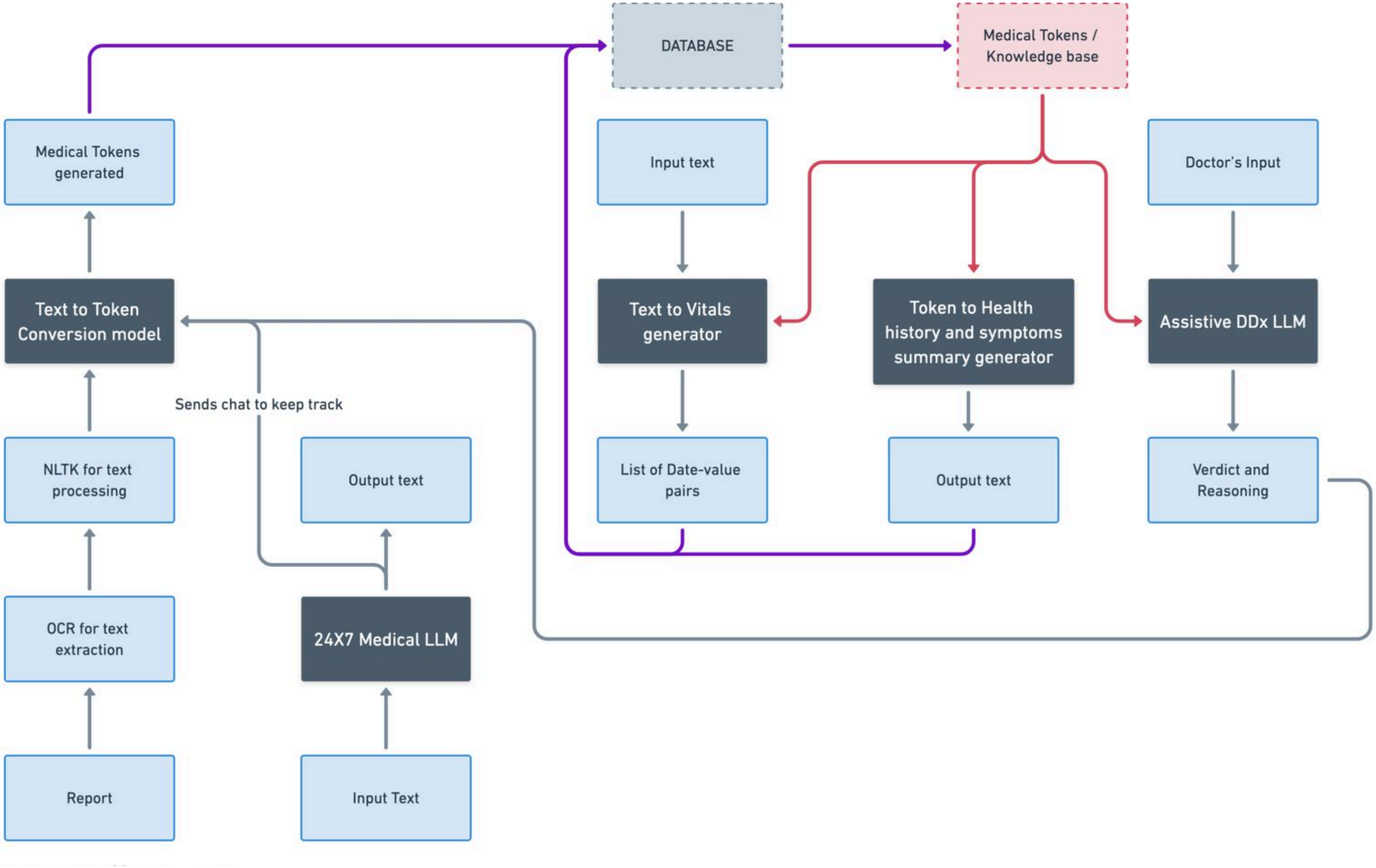
Medical History

22-year-old with a medical history that includes childhood chickenpox, an appendectomy at age 13, mild anemia managed with supplements, and mild asthma controlled with an inhaler. She has seasonal pollen allergies and a known allergy to penicillin. She has a family history of hypertension, diabetes, high cholesterol, osteoporosis, and heart disease. She leads an active lifestyle, enjoying yoga and jogging, and maintains a mostly vegetarian diet. Recently, she has managed mild anxiety related to academic pressures through mindfulness and counseling. Her immunizations are up-to-date, including the COVID-19 vaccine, and she experiences regular menstrual cycles with no significant gynecological concerns.

Current Symptoms

The patient is experiencing a high body temperature ranging from 100-104°F (38-40°C), accompanied by chills and sweating. They report headaches, body aches, and joint pain, leading to a feeling of fatigue and weakness. Respiratory symptoms include a sore throat, cough, and congestion. The patient also has gastrointestinal complaints such as nausea, vomiting, and diarrhea. Additional symptoms include loss of

AI MODELS AND DATAFLOW



LITERATURE OVERVIEW

01 Rapid advancement seen in medical token generation system using Bio-BERT

02 BM25 and Colbert are efficient for text-to-data points searching with high accuracy in the medical domain.

03 MedPaLM, have shown promise in this area, enabling patient interactions that medically aware.

04 Transformer models such as BART and T5 have been applied to medical summarization tasks, offering a way to distill complex medical histories into actionable summaries.

05 PubMedBERT has been trained specifically on medical data to support diagnosis tasks by offering contextually appropriate suggestions

RESEARCH OPPORTUNITIES

TOKENIZATION

Converting Reports, Conversations, and text inputs to Medical tokens to minimize information drain and maximize storage efficiency.

TEXT TO GRAPH GENERATION

Text to Graph creation system by searching in reports using AI can be researched and developed further

DIFFERENTIAL DAIGNOSIS TOOL

A differential diagnosis tool which is highly sensitive to context, which focuses on fact-checking and arguing with the doctor with its memory.

REFERENCES

[1] L. Zheng, X. Liu, Z. Sun, and Y. He, “Improving medical ocr information extrac tion with integrated bert and layoutxlm models,” in Health Information Process ing. Evaluation Track Papers, B. Tang, Q. Chen, H. Lin, F. Wu, L. Liu, T. Hao, Y. Wang, H. Wang, J. Lei, Z. Li, and H. Zong, Eds. Singapore: Springer Nature Singapore, 2023, pp. 164–173.

[2] K. Huang, J. Altosaar, and R. Ranganath, “Clinicalbert: Modeling clin ical notes and predicting hospital readmission,” 2020. [Online]. Available: https://arxiv.org/abs/1904.05342

[3] A. Holtzman, J. Buys, L. Du, M. Forbes, and Y. Choi, “The curious case of neural text degeneration,” 2020. [Online]. Available: https://arxiv.org/abs/1904.09751

[4] M. Barczy, F. K. Nedenyi, and G. Pap, “On simultaneous limits for aggregation ´ of stationary randomized inar(1) processes with poisson innovations,” 2020. [Online]. Available: https://arxiv.org/abs/2001.07127

[5] J. E. Sanches, L. S. Ricco, Y. Marques, W. N. Mizobata, M. de Souza, I. A. Shelykh, and A. C. Seridonio, “Majorana molecules and their spectral fingerprints,” Physical Review B, vol. 102, no. 7, Aug. 2020. [Online]. Available: http://dx.doi.org/10.1103/PhysRevB.102.075128

[6] A. Demlow and M. Neilan, “A tangential and penalty-free finite ele ment method for the surface stokes problem,” 2023. [Online]. Available: https://arxiv.org/abs/2307.01435

[7] M. Patwari, R. Gutjahr, R. Raupach, and A. Maier, JBFnet - Low Dose CT Denoising by Trainable Joint Bilateral Filtering. Springer International Publishing, 2020, p. 506–515. [Online]. Available: http://dx.doi.org/10.1007/978- 3-030-59713-949

[8] M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, “Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension,” 2019. [Online]. Available: https://arxiv.org/abs/1910.13461 18

[9] C. Raffel, N. Shazeer, A. Roberts, K. Lee, S. Narang, M. Matena, Y. Zhou, W. Li, and P. J. Liu, “Exploring the limits of transfer learning with a unified text-to-text transformer,” 2023. [Online]. Available: https://arxiv.org/abs/1910.10683

[10] Y. Gu, R. Tinn, H. Cheng, M. Lucas, N. Usuyama, X. Liu, T. Naumann, J. Gao, and H. Poon, “Domain-specific language model pretraining for biomedical natural language processing,” ACM Transactions on Computing for Healthcare, vol. 3, no. 1, p. 1–23, Oct. 2021. [Online]. Available: http://dx.doi.org/10.1145/3458754

PRESENTATION BY

Avneet Singh	2021UCA1815
Balvinder Singh	2021UCA1845
Sneha Gupta	2021UCA1859

UNDER THE GUIDANCE OF

Dr. Rudresh Dwivedi
Department of Computer Science