EG UK Computer Graphics & Visual Computing (2024)
A. Slingsby and D. Hunter (Editors)

Authoring Visualisation of Routinely Collected Data Using LLMs

A. Hosseini[1,2], J. Wood[1,2], M. Elshehaly[1,2]

[1]City St George's, University of London, Department of Computer Science, United Kingdom
[2]giCentre, City St George's, University of London

Abstract

The integration of routinely collected healthcare data into decision-making processes has the potential to revolutionise patient care and health outcomes. However, the complexity and heterogeneity of these datasets pose significant challenges for effective querying and analysis. Visualisation supports socio-technical processes where data analytics are augmented with human expertise to overcome data complexity. However, the authorship of effective visualisation is a challenging task, especially for users without a technical background, such as commissioners, clinicians and population health experts. This complexity calls for more efforts to develop natural language interfaces (NLIs) to democratise access to and understanding of routine data through visualisation. This short paper presents an innovative approach utilising Large Language Models (LLMs) to facilitate the querying and visualisation of routinely collected healthcare data. We present a preliminary framework for combining natural language queries with visualisation recommendation systems to retrieve and visualise relevant information from electronic health records (EHRs). We propose a human-in-the-loop approach for establishing accurate and efficient LLM-enabled information retrieval. Our preliminary findings suggest that LLMs can significantly streamline the visualisation authoring process, enabling stakeholders and healthcare professionals to access critical information rapidly and accurately. This work underscores the potential of LLM-driven solutions in advancing healthcare data utilisation and paves the way for future research in this promising intersection of artificial intelligence and medical informatics.

CCS Concepts

• ***Human-centered computing*** → ***Information visualization; Natural language interfaces;***

1. Introduction

The exponential growth of digital healthcare data, driven by the widespread adoption of electronic health records (EHRs) and other healthcare information systems has opened new frontiers in clinical research, patient care, and healthcare management. Routinely collected healthcare data encompasses a vast array of information, including patient demographics, clinical notes, diagnostic codes, results, and treatment histories. This wealth of data holds the potential to enhance evidence-based medical practice, support precision medicine initiatives, and improve patient outcomes.

Despite its promise, the effective utilisation of routinely collected healthcare data presents significant challenges. The heterogeneous and complex nature of these datasets make traditional querying and analysis methods labor-intensive and often impractical. Healthcare professionals and researchers require sophisticated tools and technical knowledge to navigate these gaps and extract actionable insights. Furthermore, there is a well-documented gap between insights generated using data-driven tools (e.g., visualisation dashboards) and decision-making in practice [DZTF21]. Decision makers require 'qualitative insights' that provide them with a nuanced understanding of the health system stakeholders' needs and experiences [CBB*22], which contextualise data insights. This requires novel solutions that go beyond traditional insight-driven visual analytics and enable decision-makers to tell data-reinforced human stories [EEMW23].

Recent advancements in natural language processing (NLP) have paved the way for innovative solutions to these challenges. In particular, large language models (LLMs) such as BERT [DCLT19] (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer) have demonstrated remarkable capabilities in leveraging human language for data enhancement [TVDF*23], and generating [YHH*24], refining [SCL*23], and explaining [RWVW23] data visualisations. These models, pre-trained on extensive corpora of text, can be fine-tuned for specific domains, making them highly adaptable to various tasks, including medical text analysis and information retrieval. However, their ability to overcome inconsistencies and gaps in real-world data (e.g., [FMM*23]) and support clinical decision workflows remains largely unexplored.

This short paper outlines a novel approach for the application of LLMs for querying and visualising routinely collected healthcare data, focusing on their ability to interpret and process natural

DOI: 10.2312/cgvc.20241228

language interactions to retrieve relevant information from EHRs and other healthcare data sources. By leveraging the contextual understanding of LLMs and coupling this with our newly proposed human-in-the-loop approach, we aim to guide the LLM in retrieving correct information, relevant to the user's request, facilitating more efficient and accurate data retrieval which is beneficial especially for clinicians and non-technical users. We begin by reviewing the current state of healthcare data utilisation and the challenges associated with querying such data. Next, we discuss the advancements in NLP, particularly the development and application of LLMs in healthcare. We then present our preliminary framework for leveraging LLMs to enhance healthcare data querying, followed by a case study to motivate the usefulness of our approach. Finally, we discuss the implications of our findings, potential challenges, and future directions for research in this area.

2. Related Work

The application of large language models (LLMs) in democratising access to healthcare data is a rapidly evolving field at the intersection of artificial intelligence (AI) and visualisation. This section outlines the state-of-the-art in these two areas.

2.1. Advances in Natural Language Processing (NLP) and LLMs in Healthcare Applications

Recent advancements in NLP have opened new avenues for processing and analysing medical texts. Early efforts for AI in healthcare applications focused on rule-based systems and machine learning models trained on structured datasets [MSKSH08]. However, these approaches often struggled with generalisability and the ability to interpret complex clinical narratives. The advent of LLMs, such as BERT (Bidirectional Encoder Representations from Transformers) and GPT (Generative Pre-trained Transformer), marked a significant breakthrough in NLP capabilities. These models, pre-trained on vast corpora of text data, can be fine-tuned for specific tasks, demonstrating remarkable performance in understanding and generating human language [DCLT19] [RWC*19].

The application of LLMs in healthcare has shown promising results in various tasks, including clinical note summarisation, question-answering, and predictive analytics. Lee et al. [LYK*19] demonstrated the utility of BERT in extracting relevant information from clinical texts, achieving high accuracy in entity recognition and relation extraction. Similarly, Huang et al. [HAR20] utilised GPT-2 for generating coherent and contextually relevant responses to medical queries, showcasing the potential of generative models in supporting clinical decision-making. However, the potential of these tools to support decision-making using real-world data requires a human-in-the-loop visual analytic approach to ensure the validity of their results.

2.2. AI-enabled Generation of Visualisation

The integration of LLMs in authoring, refining, and explaining visualisation is creating new opportunities for the democratisation of data storytelling. One area of exploration is the automatic generation of infographics from natural language statements, which can simplify the creation of effective visual aids for healthcare professionals who may lack design expertise. "Text-to-Viz" by Cui et al. [CZW*19] offers a solution by automatically generating infographics from natural language statements. This system addresses the needs of casual users who occasionally create infographics and may not have advanced design skills. It leverages machine learning models, particularly Conditional Random Fields (CRFs) and Convolutional Neural Networks (CNNs), to understand and extract information from text statements, and then generates corresponding visualisations based on pre-designed styles. Adopting a similar approach in healthcare could facilitate the rapid creation of educational and informational materials. For instance, a physician could input a statement like "30% of patients experience side effects from medication X," and the system would generate an infographic that visually represents this statistic. This can be particularly useful for creating patient-facing materials that need to be both accurate and easily understandable.

Similarly, "Chat2VIS" [MS23] introduces an innovative approach to generating data visualisations from natural language queries using advanced LLMs. The study focused on three primary models from OpenAI: GPT-3's "text-davinci-003," Codex's "code-davinci-002," and ChatGPT. The models demonstrated strong capabilities in generating accurate visualisations from NL queries. The Codex model, in particular, showed proficiency due to its extensive training on GitHub repositories. The study found that the models could handle a wide range of queries, making the visualisation process more accessible to users without programming expertise.

Several limitations have so far been reported in the literature, such as the models' dependency on the quality of the NL input and challenges in handling highly complex queries [MS23]. Both limitations are of interest to our proposed framework, which seeks to demonstrate the potential of LLMs to transform how users interact with data through natural language, thereby lowering the barrier to effective data analysis and visualisation.

3. Proposed Framework

Decision making in health applications begins with an understanding of the population at risk of experiencing adverse health outcomes due to an illness or a combination of interventions, comorbidities, and risk factors [EEMW23]. For example, a clinician or health commissioner may ask *"Within a cohort, which type 2 diabetics need a monthly review, which type 2 diabetics are okay to have six-monthly review?"*. Understanding the characteristics and needs of the diabetic cohort can inform resourcing and service planning decisions. Upon receiving a prompt such as the above, the system aims to create a cohort that *(a)* fits the criteria of "Type 2 diabetes" **and** any other selection criteria intended in the users' prompt that specify what they mean by contextualising statements, such as *"Within a cohort,..."* (we refer to this as direct cohort selection); or *(b)* fits domain-specified risk criteria that augment the cohort of diagnosed diabetics. The latter, which we refer to as indirect selection, is critical to overcoming inconsistencies in data recording practices and wider inequalities that may impact individuals' access to the diagnostic process. We describe the process of generating visualisations that capture both direct and indirect selections in four stages which constitute our proposed framework.

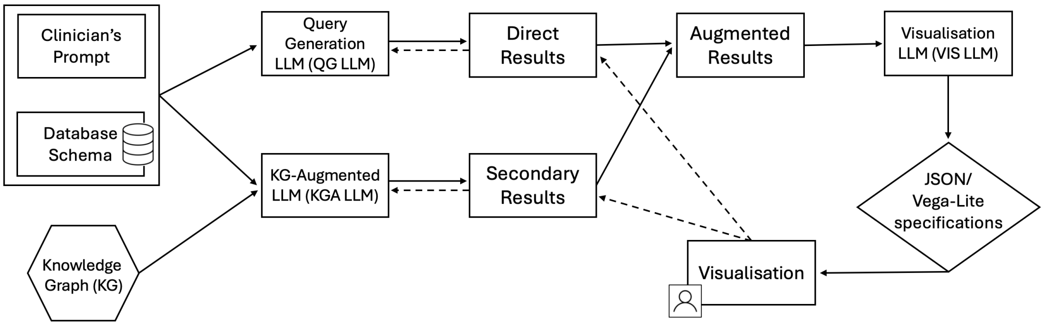

Figure 1: *Creating clinician's "population at risk" cohort using LLMs and Knowledge Graphs to retrieve augmented data and generate JSON/Vega-Lite specifications for visualisation.*

3.1. Generating Criteria for Direct Selection

For cohort creation, the Query Generation LLM (QG LLM) processes a user's prompt along with the database schema. Utilising its contextual and linguistic capabilities, the LLM navigates the database schema to identify the appropriate dataset, tables, and fields, ultimately generating a code list that defines the primary diagnosis to look for, and a number of constraints that define cohort demographics (e.g., residents of a specific locality, who are within a certain age range).

The Systematised Nomenclature of Medicine Clinical Terms (SNOMED-CT) is a global clinical terminology aimed at enhancing data quality and patient safety [LdKLC13]. However, mapping natural language to SNOMED codes is challenging due to the variability and context-dependence of natural language descriptions. This can lead to inaccuracies in cohort selection, as the LLM must interpret and match the user's intent with the exact SNOMED codes, often without the benefit of detailed contextual information. One potential solution is to enhance the LLM with a more extensive medical ontology and a robust training dataset that includes a wide variety of clinical descriptions and their corresponding SNOMED codes. Additionally, integrating feedback loops where clinicians can validate and correct the generated codes could improve accuracy (see dashed lines from results back to LLMs in Figure 1) .

3.2. Generating Criteria for Bias and Risk Mitigation

We propose the incorporation of domain-specific risk criteria as knowledge graphs to retrieve information that is not obtained through direct selection. This approach is critical for addressing inconsistencies in data recording practices and mitigating broader inequalities that may affect individuals' access to diagnostic procedures. Disease risk factors, sources of inequalities, and socio-economic factors can be utilised as a foundation for cohort creation. For instance, when querying the risk factors for Type 2 diabetes, factors such as obesity, a sedentary lifestyle, and family history can be identified. Furthermore, socio-economic determinants, such as income level and demographic details (e.g., residents of specific localities or individuals within certain age ranges), can be highlighted, as these factors contribute to health disparities.

We can conceptualise a specific domain, such as autism care, by integrating existing domain ontologies (e.g., AutismOnt [HM22]) and SNOMED-CT concepts used in health and social care records. This will involve adapting existing methodologies for mapping semantic resources using Knowledge Graph (KG) embeddings (e.g., [CJRH*21]). This approach will enable the creation of a digital representation of concepts, properties, and axioms that characterise risks, needs, and experiences beyond diagnostic labels. The resulting ontology will assist the KG-Augmented LLM (KGA LLM) in extracting pertinent information from the database, resulting in a refined query that captures the essential knowledge constructs used by frontline staff to record routine data about the cohort.

3.3. SQL Query Generation

As illustrated in Figure 1, the clinician's prompt and the database schema are initially processed by QG LLM, which identifies a mapping from the natural language prompt to SNOMED codes, thereby generating an SQL query to query the cohort information using the direct selection approach.

Concurrently, the clinician's prompt and database schema are submitted to the KGA LLM, which is supplemented by a knowledge graph. This domain-specific knowledge graph is developed based on input from expert clinicians, incorporating specified factors associated with the health-related issue and employing a weighted model for easier data retrieval. The knowledge graph effectively guides the LLM towards tables with relevant information. By pruning the number of tables, the complexity of the schema traversal is significantly decreased, which enhances the LLM's ability to generate more accurate SQL queries. This reduction in the number of tables not only simplifies the LLM's task but should also

Proceedings published by Eurographics - The European Association for Computer Graphics.

improve the accuracy of data retrieval enhancing the overall utility of routinely collected healthcare data for research and decision-making. Ultimately, the outputs from both LLMs are integrated to produce the augmented dataset.

3.4. Visualisation Specification

The Visualisation LLM (VIS LLM) will assist in generating JSON specifications for visualising the augmented results that were produced by the QG LLM and the KGA LLM (see Figure 1). By interpreting the clinician's query and the resultant augmented data, the VIS LLM will produce a JSON or Vega-Lite specification that defines *how* the data should be visualised. This includes determining the type of chart, visual mark, channel, etc. that best represent the data. We will feed the resulting specifications to a visualisation recommendation system, such as Draco [MWN*18], to ensure the produced visualisation design aligns with best practices in visualisation design. This human-in-the-loop approach will lead to increased efficiency and accessibility of visual creation for healthcare professionals. However, ensuring the clinical accuracy of the generated visuals and integrating them seamlessly into existing healthcare workflows remains a challenge. Consequently, the system must be capable of handling the complex and nuanced language often found in medical data and research.

4. Case Study: Connected Bradford

Connected Bradford serves as the primary dataset for this investigation. Connected Bradford is a Whole System Data Linkage Accelerator that includes a wide range of data from various sources such as health, education, social care, and environmental data. The data has been linked using pseudonymised identifiers, such as the NHS number, for 600,000 individuals [SMB*22]. This allows for longitudinal analysis and near real-time monitoring of population health. This dataset is a valuable resource for research within the field of Population Health Management (PHM). Specifically, the ambition is to use insights from this data to support public engagement, practitioner and policy integration, and data validity and visualisation.

The incomplete nature of routine data often complicates the manifestation and characterisation of a reference population. For instance, our previous work considers an autism example from the Connected Bradford dataset [EEMW23]. Our analysis identified two relevant constructs: Autistic Spectrum Disorder (ASD) and Developmental Disorder of Motor Function (DDMF). Despite healthcare records for over 600,000 citizens, only 154 cases of DDMF were annotated by clinicians, compared to 6,249 Autism cases, despite an estimated prevalence five times lower than DDMF. This discrepancy can arise from any of the discrepancies listed by McCurdy et al. [MGM19]. Selecting the autistic "Reference population" requires an augmentation step as proposed in our framework.

Following the approach outlined in Figure 1, QG LLM would retrieve this information using the direct querying approach. And with the help of the KGA LLM, we aim to overcome this inherent issue of routine data by augmenting the initial results with the secondary results creating a dataset that is a more comprehensive description of the reality.

5. Conclusion and Future Work

The inherent complexity and heterogeneity of routinely collected data in health and care present substantial challenges for effective querying and analysis. This paper has explored the innovative application of LLMs to facilitate the querying and visualisation of such data, proposing a preliminary framework that combines natural language queries with visualisation recommendation systems to enhance information retrieval and presentation from electronic health records.

Our proposed human-in-the-loop approach leverages LLMs to streamline the visualisation authoring process, enabling healthcare professionals and stakeholders to access and interpret critical information more rapidly and accurately. Preliminary findings suggest that LLMs can significantly reduce the barriers to effective data analysis and visualisation, making these processes more accessible to users without technical expertise. This framework underscores the potential of LLM-driven solutions to advance healthcare data utilisation, highlighting the importance of integrating human expertise with advanced AI capabilities to overcome the challenges posed by complex routinely collected healthcare datasets.

The case study on Connected Bradford illustrates the practical application and potential benefits of our proposed framework. By addressing issues such as data incompleteness and the need for nuanced, context-rich insights, our approach demonstrates a path forward for improving the accuracy and utility of healthcare data analysis. The use of domain-specific knowledge graphs and enhanced SQL query generation techniques further refines the data retrieval process, ensuring that the resulting datasets more accurately reflect the reality of patient populations.

To build upon the current findings and address the identified limitations, several avenues for future research are proposed: **(i) Enhanced Schema Representation:** Providing both QG and KGA LLMs with comprehensive descriptions of routine data tables and relations could improve query accuracy and reduce the likelihood of errors. This would involve developing more detailed schema documentation and incorporating it into the query generation process. **(ii) Comparative Performance Analysis:** Exploring the use of a variety of LLMs, such as Mistral, Groq, Codex and GPT-4o, for querying healthcare data could provide valuable insights into their relative performance. Additionally, specialised models like Google's Tapas and Microsoft's Tapex, designed for table question-answering, could offer superior performance on tabular datasets such as those in the Connected Bradford repository. **(iii) User Interface and Experience Improvements:** Designing intuitive and user-friendly interfaces that enable non-technical users to interact seamlessly with LLM-driven visualisation tools. Conducting usability studies to understand the specific needs and preferences of different user groups, including commissioners, clinicians, and population health experts, to tailor the tools accordingly. **(iv) Advanced Visualisation Techniques:** Exploring advanced visualisation techniques, such as AI-generated visual stories will provide more comprehensive and actionable insights.

By exploring these future directions, the integration of LLMs into the visualisation of healthcare data can be further refined, ultimately improving the accessibility and usability of routinely collected healthcare data for clinical and research purposes.

References

[CBB*22] Chishtie J., Bielska I. A., Barrera A., Marchand J.-S., Imran M., Tirmizi S. F. A., Turcotte L. A., Munce S., Shepherd J., Senthinathan A., et al.: Interactive visualization applications in population health and health services research: systematic scoping review. *Journal of medical Internet research 24*, 2 (2022), e27534. 1

[CJRH*21] Chen J., Jiménez-Ruiz E., Horrocks I., Antonyrajah D., Hadian A., Lee J.: *Augmenting Ontology Alignment by Semantic Embedding and Distant Supervision*. 05 2021, pp. 392–408. doi:10.1007/978-3-030-77385-4_23. 3

[CZW*19] Cui W., Zhang X., Wang Y., Huang H., Chen B., Fang L., Zhang H., Lou J.-G., Zhang D.: Text-to-viz: Automatic generation of infographics from proportion-related natural language statements, 2019. arXiv:1907.09091. 2

[DCLT19] Devlin J., Chang M.-W., Lee K., Toutanova K.: Bert: Pre-training of deep bidirectional transformers for language understanding, 2019. arXiv:1810.04805. 1, 2

[DZTF21] Dimara E., Zhang H., Tory M., Franconeri S.: The unmet data visualization needs of decision makers within organizations. *IEEE Transactions on Visualization and Computer Graphics 28*, 12 (2021), 4101–4112. 1

[EEMW23] Elshehaly M., Eddy L., Mon-Williams M.: Clever: A framework for connecting lived experiences with visualisation of electronic records. pp. 126–130. doi:10.1109/VIS54172.2023.00034. 1, 2, 4

[FMM*23] Fonseca M., MacKenna B., Mehrkar A., Collaborative O., Walters C. E., Hickman G., Pearson J., Fisher L., Inglesby P., Bacon S., et al.: Primary care coding activity related to the use of online consultation systems or remote consulting: an analysis of 53 million peoples' health records using opensafely. *medRxiv* (2023), 2023–01. 1

[HAR20] Huang K., Altosaar J., Ranganath R.: Clinicalbert: Modeling clinical notes and predicting hospital readmission, 2020. arXiv:1904.05342. 2

[HM22] Hassan M. M., Mokhtar H. M.: Autismont: An ontology-driven decision support for autism diagnosis and treatment. *Egyptian Informatics Journal 23*, 1 (2022), 95–103. URL: https://www.sciencedirect.com/science/article/pii/S1110866521000451, doi:https://doi.org/10.1016/j.eij.2021.07.002. 3

[LdKLC13] Lee D., de Keizer N., Lau F., Cornet R.: Literature review of SNOMED CT use. *Journal of the American Medical Informatics Association 21*, e1 (07 2013), e11–e19. URL: https://doi.org/10.1136/amiajnl-2013-001636, arXiv:https://academic.oup.com/jamia/article-pdf/21/e1/e11/17375846/21-e1-e11.pdf, doi:10.1136/amiajnl-2013-001636. 3

[LYK*19] Lee J., Yoon W., Kim S., Kim D., Kim S., So C. H., Kang J.: Biobert: a pre-trained biomedical language representation model for biomedical text mining. *Bioinformatics 36*, 4 (Sept. 2019), 1234–1240. URL: http://dx.doi.org/10.1093/bioinformatics/btz682, doi:10.1093/bioinformatics/btz682. 2

[MGM19] McCurdy N., Gerdes J., Meyer M.: A framework for externalizing implicit error using visualization. *IEEE Transactions on Visualization and Computer Graphics 25*, 1 (2019), 925–935. doi:10.1109/TVCG.2018.2864913. 4

[MS23] Maddigan P., Susnjak T.: Chat2vis: Generating data visualizations via natural language using chatgpt, codex and gpt-3 large language models. *IEEE Access 11* (2023), 45181–45193. doi:10.1109/ACCESS.2023.3274199. 2

[MSKSH08] Meystre S. M., Savova G. K., Kipper-Schuler K. C., Hurdle J. F.: Extracting information from textual documents in the electronic health record: a review of recent research. *Yearbook of Medical Informatics* (2008), 128–144. 2

[MWN*18] Moritz D., Wang C., Nelson G. L., Lin H., Smith A. M., Howe B., Heer J.: Formalizing visualization design knowledge as constraints: Actionable and extensible models in draco. *IEEE transactions on visualization and computer graphics 25*, 1 (2018), 438–448. 4

[RWC*19] Radford A., Wu J., Child R., Luan D., Amodei D., Sutskever I.: Language models are unsupervised multitask learners. URL: https://api.semanticscholar.org/CorpusID:160025533. 2

[RWVW23] Raval S., Wang C., Viégas F., Wattenberg M.: Explain-and-test: An interactive machine learning framework for exploring text embeddings. In *2023 IEEE Visualization and Visual Analytics (VIS)* (2023), IEEE, pp. 216–220. 1

[SCL*23] Shi C., Cui W., Liu C., Zheng C., Zhang H., Luo Q., Ma X.: Nl2color: Refining color palettes for charts with natural language. *IEEE Transactions on Visualization and Computer Graphics* (2023). 1

[SMB*22] Sohal K., Mason D., Birkinshaw J., West J., McEachan R., Elshehaly M., Cooper D., Shore R., McCooe M., Lawton T., Mon-Williams M., Sheldon T., Bates C., Wood M., Wright J.: Connected bradford: a whole system data linkage accelerator. *Wellcome Open Research 7* (11 2022), 26. doi:10.12688/wellcomeopenres.17526.2. 4

[TVDF*23] Thompson W. E., Vidmar D. M., De Freitas J. K., Pfeifer J. M., Fornwalt B. K., Chen R., Altay G., Manghnani K., Nelsen A. C., Morland K., et al.: Large language models with retrieval-augmented generation for zero-shot disease phenotyping. *arXiv preprint arXiv:2312.06457* (2023). 1

[YHH*24] Ye Y., Hao J., Hou Y., Wang Z., Xiao S., Luo Y., Zeng W.: Generative ai for visualization: State of the art and future directions. *Visual Informatics 8*, 2 (2024), 43–66. URL: https://www.sciencedirect.com/science/article/pii/S2468502X24000160, doi:https://doi.org/10.1016/j.visinf.2024.04.003. 1

Proceedings published by Eurographics - The European Association for Computer Graphics.