Leveraging Natural Language Processing and Large Language Models in the Hospital Outpatient Setting

Moorfields Eye Hospital NHS Foundation Trust
NHS Digital Academy
NHS Guy's and St Thomas' NHS Foundation Trust

Rishi Ramessur[1], Nicole Quah[1], Quang Nguyen[1], Nikolas Pontikos[1], Peter Thomas[1]
[1]Moorfields Eye Hospital, London, United Kingdom

The Clinical Problem

- Electronic health records (EHRs) store vast amounts of patient information – although often unstructured and not suitable for computational analysis (Figures 1 and 2). Manual case identification, notes summarisation, referral triage and letter generation are lengthy processes that add to clinicians' workload at Moorfields Eye Hospital (MEH).

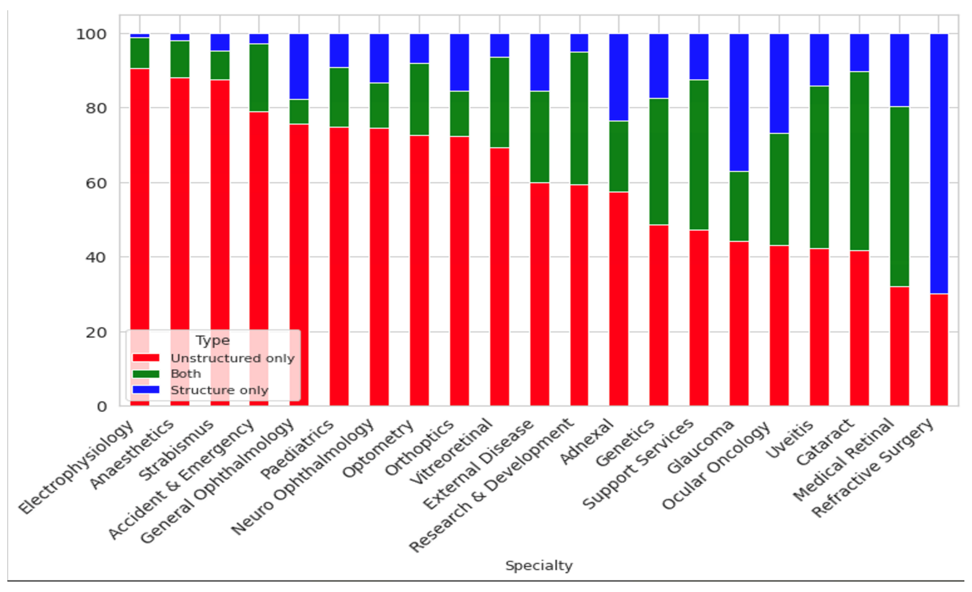


Figure 1. Proportion (%) of 3,820,797 MEH visits where visual acuity (VA) values were recorded in unstructured data only (red), structured only (blue) and in both (green). Unstructured VA accounts for nearly 68% of VA recordings, of which 35% could only be found in the letters.

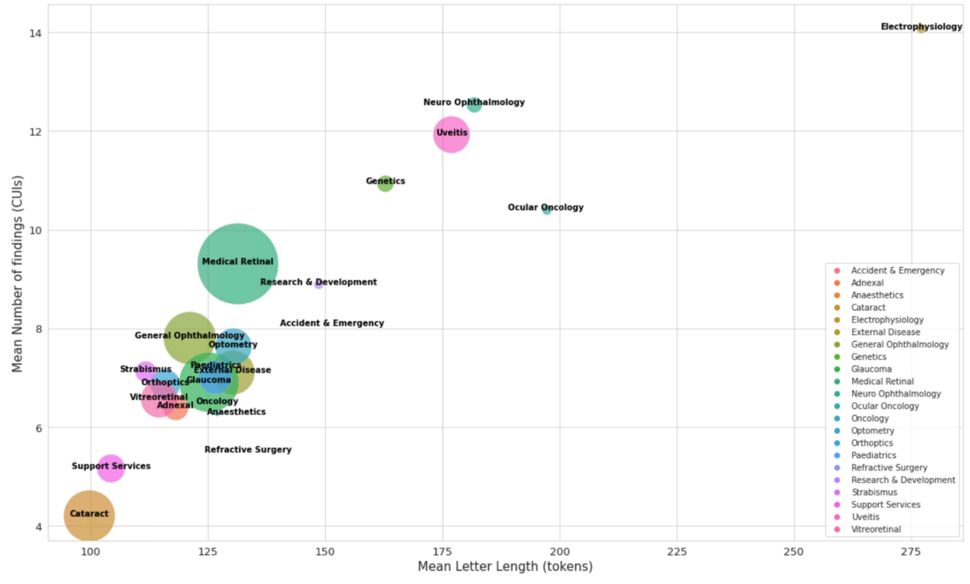


Figure 2. Relationship between letter length and mentioned clinical findings by specialty. Circle size is proportional to number of letters in that specialty. Longer letters appear to contain more clinical information.

The AI Solution

- Natural Language Processing (NLP) is a subfield of Artificial Intelligence (AI) that enables computers to understand and communicate in human language. It combines computational linguistics with machine learning and deep learning to recognise, understand, and generate text and speech. Large Language Models (LLMs) are foundation models trained on vast data, enabling them to understand and generate natural language and other content to perform various tasks.
- NLP and LLMs can be used to create a more effective querying system that allows clinicians to find patient records without having to enter exact words or phrases. They can be used to automatically assign SNOMED-CT codes (a structured clinical vocabulary for electronic health records) for patient visits. They can detect phenotypes, measurements and prescriptions in text and map them to controlled terminologies for further analysis (case identification, genetic diagnosis, etc.). They can also facilitate easier clinical audit and research, automate consult note and letter generation from a set of keywords. Other uses include making patient letters more understandable, linking them to information resources, and assisting with referral management.

AI Solution Deployment and Progress

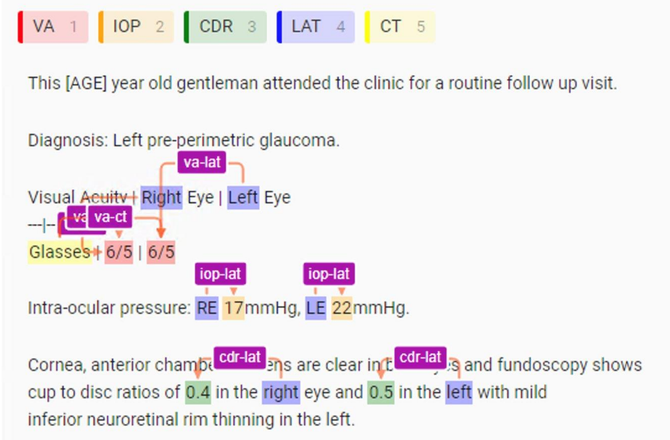


Figure 3. Relation classes and data pre-processing for relation extraction model.

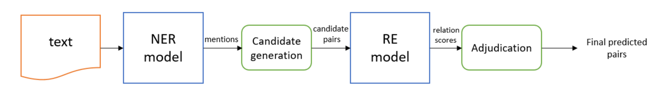


Figure 4. The end-to-end relation extraction pipeline. NER = Named Entity Recognition model. R = Relation extraction model.

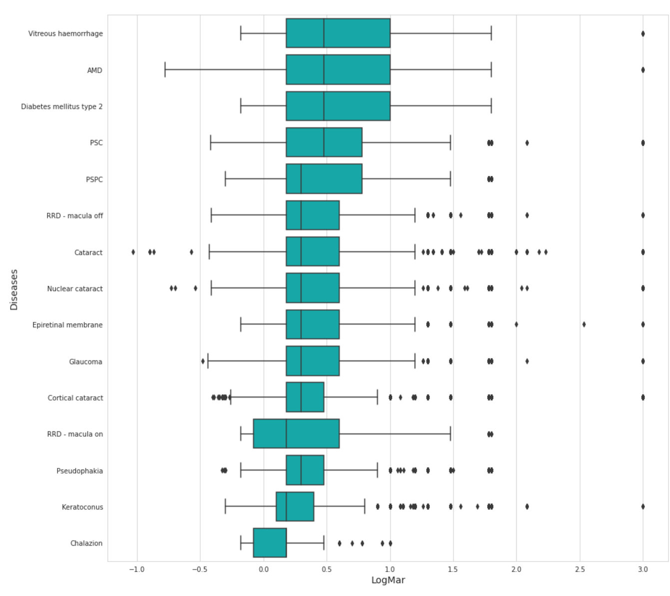


Figure 5. VA extracted from Eyenote for the 15 most frequently seen diseases in MEH EHR, sorted by VA median value.

Eyenote: a locally developed BERT-based model.

A Bidirectional Encoder Representation of Transformers (BERT) based model was pre-trained on 1 million anonymised MEH notes and fine-tuned on a sample of annotated notes (Figure 3).

Named Entity Recognition (NER) involves identifying and classifying named entities in text. This was used to extract key features, such as VA, laterality (lat), correction type (ct) and cup-to-disc ratio from patient letters. Relation extraction can be framed as a classification problem in which the model takes two entities as inputs. It outputs the probability of whether the two entities have a semantic connection, and the category of relation between them. This was used to define relation classes between VA-lat, and VA-ct (Figure 4). Eyenote detected VA much better (F1 95.07%) than the regular expression baseline (F1 61.2%). Relation extraction appeared to be a more challenging task, with Eyenote F1 ranging from 74-88%, similar to baseline model.

VA medians and distributions clinically correlated with diagnosis (Figure 5). The model extracted right VA with 82.78% accuracy and left VA with 72.37% accuracy on an independent test set of 836 letters[1] (Figure 6).

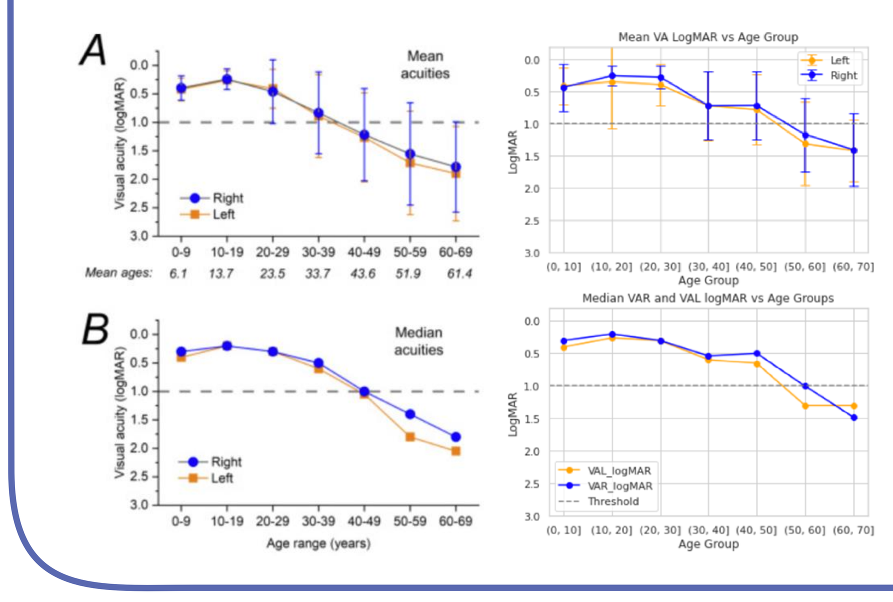
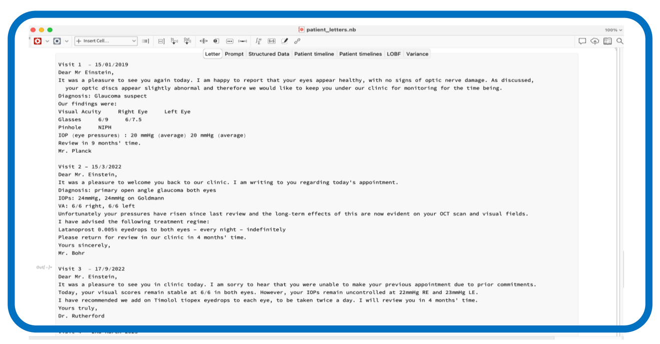


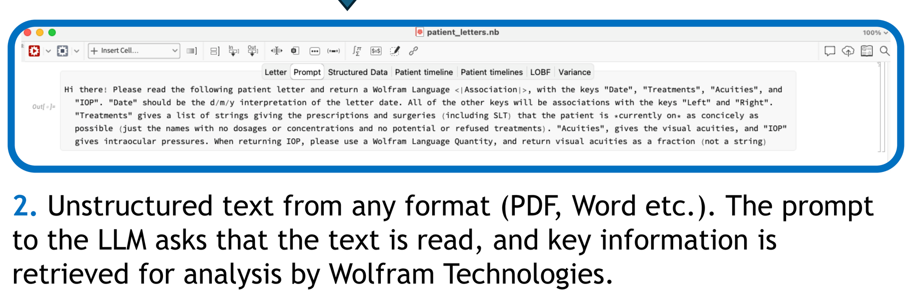
Figure 6. LogMAR VA extracted using Eyenote from an independent test set was plotted against age groups, as previously done by De Silva et. al[1]. VAs acquired by Eyenote (top right and bottom right) show a similar overall shape to those acquired by hand in De Silva et al's paper (top left and bottom left)[1], but VA of age groups above 40 years old appear higher than the ground truth.

Wolfram Alpha: Leveraging novel LLM technology

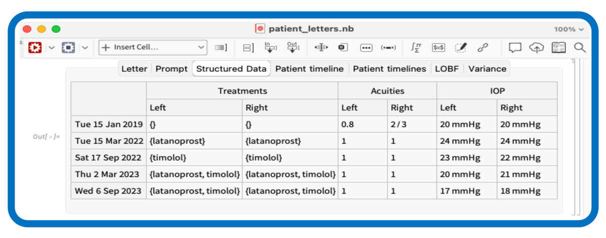
With rapid development in LLM capabilities, we explored whether commercially available or open-source technologies could better achieve our goals. Collaborating with a computational analytics company with early access to ChatGPT, we developed a second proof-of-concept (Figure 7).



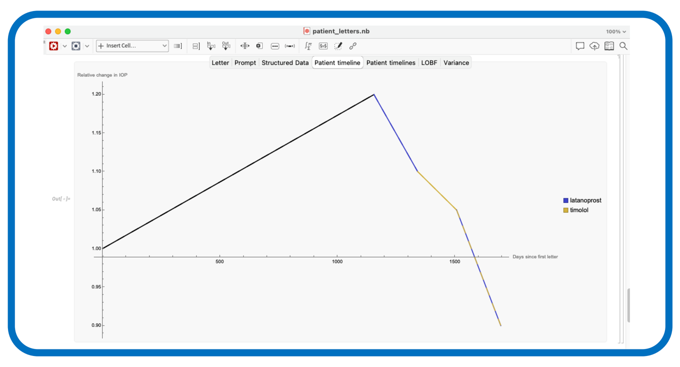
1. Text extracted from a series of PDF patient letters for one patient with Glaucoma. The patient was prescribed two different treatments. Which treatment worked best?



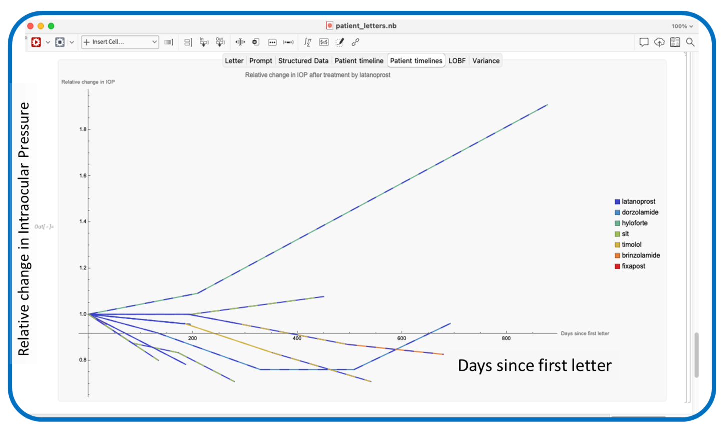
2. Unstructured text from any format (PDF, Word etc.). The prompt to the LLM asks that the text is read, and key information is retrieved for analysis by Wolfram Technologies.



3. Values 'understood' by Wolfram Technology, and therefore can be easily manipulated for example, dates, pressure measurements etc.



4. Data can then be 'computed' according to project needs: e.g. plotted on graphs and maps. For example, the graph (automatically plotted), illustrates the changes in the Intraocular Pressure (IOP) from when it was first measured. Changes in treatment are indicated according to the key.



5. Data can be manipulated from the letters of many patients. In this example to reveal any new, optimal treatment protocols.

6. The full range of Wolfram Technologies (data visualisation, statistical analysis and machine learning, and many more) can reveal data insights.

Figure 7. Even without any Ophthalmology specific training, the Wolfram Alpha proof-of-concept tool was able to structure clinical data from synthetic correspondence, and in fact went substantially beyond our expectations – it was able to reconcile discrepancies in a series of correspondence, negating a medication incorrectly documented in a letter based on a subsequent letter that identified the error.

Current and future work

Our proof-of-concept applications have informed strategic insights into local deployment, including:

- Decisions on whether to **buy or build** solutions: our locally developed Eyenote tool was rapidly superseded by novel LLM technology – an impressive GPT-4 based tool was developed with Wolfram Alpha in 2 months. In a rapidly advancing field, efforts are best focused on preparing the organisation for deployment of commercially available LLMs.
- Tailoring **bespoke use cases**: internal efforts to deploy LLMs were initially met with caution. A carefully defined use case for pre-coding referrals to determine coding variations for financial reimbursement across London Ophthalmology providers offers a low-risk, high-value domain in which to deploy and develop LLM capabilities. Our position as North Central and East London Single Point of Access (SPoA) referral pathway provider – and recent SPoA audit employing NLP techniques, cluster analysis and logistic regression – has helped secure MEH a £2-3 million contract for North Central London Lead Provider of Community Ophthalmology. This includes funding for LLM deployment research.
- **Local vs cloud deployment**: We will test proof-of-concept applications using a two-pronged approach. Firstly, using open-source LLMs on a secure unnetworked device. Secondly, deploying commercially available LLMs via private cloud.
- **Training and communication**: Whilst continuing to work towards deployment, I am developing an organisational strategy to guide AI integration. I have also received Topol Fellowship funding to embed Clinical Informatics training into Ophthalmology curricular delivery, preparing the workforce for safe and responsible AI use once deployed (Figure 8).

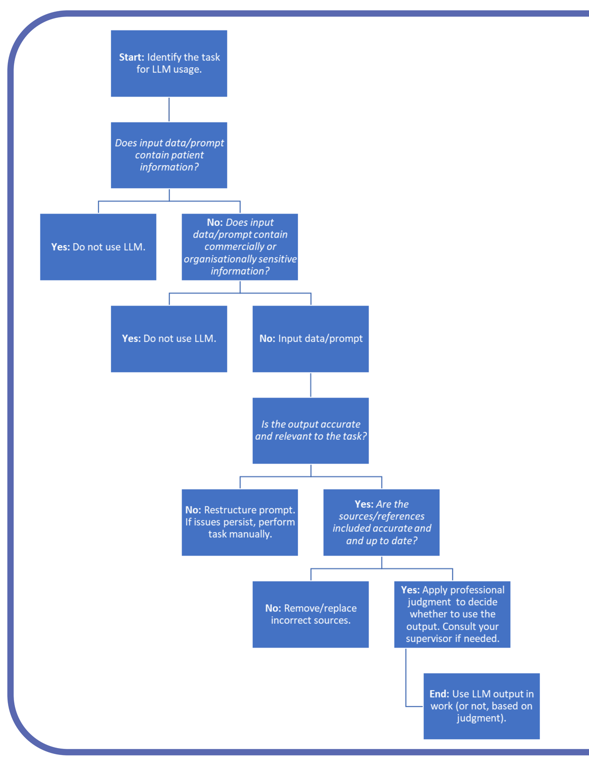


Figure 8. A flow diagram from MEH's LLM staff guidance document aimed at educating the workforce in responsible AI use, as part of a provisional AI strategy.

References

1. De Silva SR, Chan HW, Agarwal A, Webster AR, Michaelides M, Mahroo OA. Visual Acuity by Decade in 139 Males with RPGR-Associated Retinitis Pigmentosa. Ophthalmol Sci. 2023 Jul 24;4(2):100375. doi: 10.1016/j.xops.2023.100375. PMID: 37868789; PMCID: PMC10587616.

Email: Rishi.ramessur@nhs.net
LinkedIn: www.linkedin.com/in/rishi-ramessur