

GLLaucoMed: A Secure LLM-Powered Agentic Workflow for Automated Medication Extraction from Free-Text Glaucoma Clinical Notes

Nicholas Solages BS¹, Rafael Scherer MD, PhD¹, Gustavo A. Samico, MD^{1,2}, Naomi E. Gutkind, MD^{1,3}, Janet Kang BS⁴, Felipe A. Medeiros, MD, PhD¹, Swarup S. Swaminathan, MD¹

1. Bascom Palmer Eye Institute, University of Miami Miller School of Medicine, Miami, FL
2. Department of Ophthalmology and Visual Sciences, Escola Paulista de Medicina Universidade Federal de São Paulo, São Paulo, SP, Brazil
3. Olive View – UCLA Medical Center, Sylmar, CA
4. Miller School of Medicine, University of Miami, Miami, FL

Background

Observational health databases conforming to the OMOP Common Data Model (CDM) aggregate structured electronic health record (EHR) data to enable large-scale, reproducible, federated studies. Drug exposure data in the OMOP CDM is typically sourced from structured medication list fields within the EHR. However, chronic conditions like glaucoma — a leading cause of irreversible blindness affecting approximately 80 million people worldwide — structured medication lists are often unreliable. A manual chart review of 150 glaucoma outpatient visits found medication discrepancies in 32% of encounters,¹ and a separate study found that 75% of patients followed the clinician-stated plan rather than the EHR medication list when discrepancies existed.² These findings suggest that the drug exposure domain in OMOP CDM may systematically misrepresent actual treatment for glaucoma patients, directly confounding analyses of disease progression and treatment effectiveness within OHDSI network studies.

Free-text clinical progress notes are known to contain richer and more accurate treatment information than structured fields; only 13% of concepts extracted from patient notes have corresponding structured counterparts.³ Yet unstructured notes are excluded from most large observational repositories — including the AAO IRIS Registry and NIH All of Us program — primarily due to protected health information (PHI) concerns. Large language models (LLMs) now offer a validated pathway to extract structured medication data from these notes at scale. However, no prior work has demonstrated a secure, end-to-end pipeline that captures not only current medications but also clinician-documented treatment changes from full clinical notes.

Methods

We developed GLLaucoMed, a staged agentic LLM pipeline for medication extraction from free-text EHR notes. The data source was the Bascom Palmer Ophthalmic Repository (BPOR), a single-institution EHR-derived database of over 70,000 glaucoma patients. We extracted 1,242 clinic notes from unique patients seen between 2014 and 2024, representing 20 glaucoma specialists. Two glaucoma specialist graders labeled each note for four tasks — current topical medications (CTM), proposed changes to topical medications (Δ TM), current oral medications (COM), and proposed changes to oral medications (Δ OM) — with a third specialist adjudicating disagreements. Notes were split 10%/10%/80% into training, validation, and test sets stratified by clinicians for prompt engineering and final evaluation.

The GLLaucoMed pipeline consisted of three sequential agentic stages (initial extraction, validation, and conditional revision) followed by a standardization step. The standardization script mapped brand names to generic equivalents via a canonical drug library and Drugs@FDA API calls, normalized frequency expressions, and applied fuzzy matching for typographical errors. Five leading LLMs (Claude Opus 4.6, GPT-5.2, DeepSeek-V3.2, Grok 4.1, and Qwen3.6-35B-A3B) were evaluated within HIPAA-compliant Microsoft Azure AI Foundry containers. Performance was assessed using F1 score for all notes, and exact match accuracy and Jaccard Index (JI) among positive cases — both at the full-label level and decomposed into individual components (medication name, laterality, frequency, dosage for COM/ Δ OM, and change term for Δ TM/ Δ OM) — with 1,000-bootstrap 95% confidence intervals.

Results

Among 992 test notes, 68.2% of eyes had prescribed topical medications (CTM), 20.3% had documented medication changes (Δ TM), 4.7% were on oral medications (COM), and 2.5% had oral medication changes (Δ OM). Intergrader agreement (Gwet AC1) was 0.799, 0.888, 0.985, and 0.988 for CTM, Δ TM, COM, and Δ OM, respectively, and model failures were rare (0–6 notes per task).

Claude Opus 4.6 achieved the highest F1 scores across all tasks — CTM: 0.985 (95% CI: 0.980–0.989), Δ TM: 0.905 (0.883–0.925), COM: 0.923 (0.860–0.972), Δ OM: 0.958 (0.889–1.000) — with other models ranging from 0.968–0.978, 0.826–0.897, 0.887–0.906, and 0.815–0.940, respectively. Among positive cases, Claude also led in exact match accuracy and Jaccard Index (JI) — CTM (0.882, 0.918), Δ TM (0.780, 0.827), COM (0.872, 0.870), Δ OM (0.920, 0.922) — while other models ranged from 0.730–0.830, 0.619–0.733, 0.766–0.851, and 0.583–0.680 for exact match, and 0.809–0.896, 0.668–0.781, 0.765–0.852, and 0.583–0.877 for JI across the same tasks.

Conclusion

GLLaucoMed demonstrates that a secure, HIPAA-compliant agentic LLM workflow can systematically convert unstructured free-text clinical notes into standardized, structured medication data at scale. By extracting not only current medications but also clinician-documented treatment changes — and mapping them to canonical drug identifiers — this pipeline produces a more accurate and complete drug exposure record than structured EHR fields alone. Enriching observational databases such as OMOP CDM with this standardized output has direct implications for the reliability of large-scale cohort studies, particularly analyses of disease progression and treatment effectiveness in glaucoma, where medication misclassification has previously been shown to confound longitudinal findings.

References

1. Chen JS, Lin WC, Kaluzny JV, et al. Discrepancies in ophthalmic medication documentation for glaucoma patients. *Ophthalmol Sci* 2022;2(1):100091.
2. Bacon TS, Fan KC, Desai MA. Electronic medical record and glaucoma medications: connecting the medication reconciliation with adherence. *Clin Ophthalmol* 2016;10:221-5.
3. Seinen TM, Kors JA, van Mulligen EM, Rijnbeek PR. Using structured codes and free-text notes to measure information complementarity in electronic health records: feasibility and validation study. *J Med Internet Res* 2025;27:e66910.