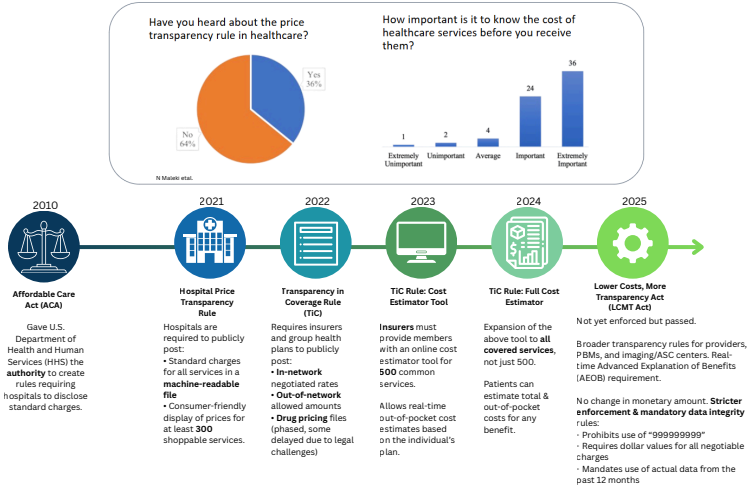




ABSTRACT

Despite major U.S. price transparency initiatives such as the Hospital Price Transparency Rule (2021), the Transparency in Coverage Rule (2022), and the Lower Costs, More Transparency Act (2023), hospital pricing data remain difficult for patients to access and interpret. Compliance is still limited, with only about 24.5 percent of hospitals meeting federal requirements. The data remain fragmented and technical, limiting their usefulness to the public. To address these challenges, we propose an artificial intelligence based navigation tool that interprets and aggregates federally mandated price data into clear, personalized cost insights, helping patients make more informed and cost effective healthcare decisions.



DATASET

We have collected 14 different hospital MRFs, CDMs, and code lists for NDC(National Drug Code), CPT(Current Procedural Terminology), HCPCS(Healthcare Common Procedure Coding System), and RC(Revenue Codes). Each dataset (MRF and CDM) was obtained directly from the respective institution's website. We merged each MRF and CDM into two separate files and created an SQL database to query the data. Each code list was also matched with its corresponding codes in the MRFs. Since CDMs do not contain codes, they were only used when no possible relationships could be found in the MRFs. Regarding the CDMs, the main focus is the description of each service and the straightforward costs that hospitals charge.

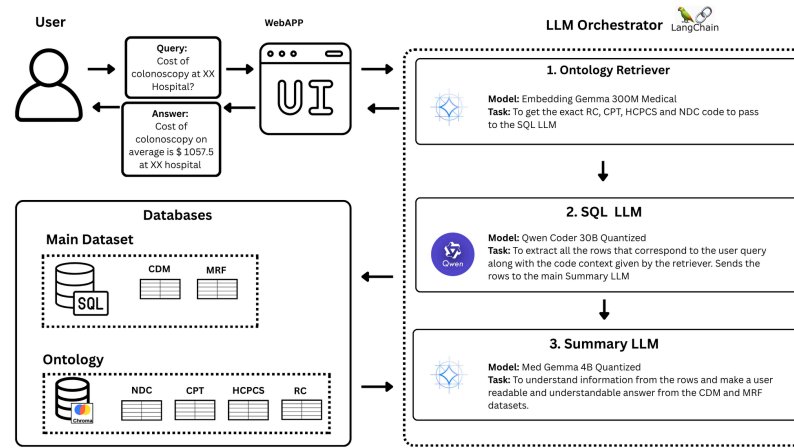
	column	dtype	non_null	nulls	null_%	sample_values
Description	object	75038	0	0.0		BLOCK ANESTHESIA SUPPLIES, CARDOVERSION - INT...
Hospital	object	75038	0	0.0		
Std Charge	object	75038	0	0.0		107, 1360, 3146, 3375, 1179

MRFs, on the other hand, represent the detailed charge history for each patient's payment, with various embedded codes explaining how the total price was determined.

This includes information on insurance, discounts, negotiated rates, related diagnoses, procedures, medications, and more. The table below illustrates the CMS-provided format and dataset instructions..

MRF Information	Standard Charges	Item & Service Information
MRF Date	Gross Charge	General Description
CMS Template Version	Discounted Cash Price	Setting
Affirmation Statement	Payer Name	Drug Unit of Measurement*
	Plan Name	Drug Type of Measurement*
	Standard Charge Method	
	Payer-Specific Negotiated Charge - Dollar Amount	
	Payer-Specific Negotiated Charge - Percentage	
Hospital Information	Payer-Specific Negotiated Charge - Algorithm	Coding Information
Hospital Name	Estimated Allowed Amount*	Billing/Accounting Code
Hospital Location(s)	Additional Generic Notes	Code Type
Hospital Address(es)	Additional Payer-Specific Notes	Modifiers*
Hospital License Information	De-identified Minimum Negotiated Charge	
	De-identified Maximum Negotiated Charge	

MODEL PIPELINE



Explanation of Diagram:

- User Query:** Natural language question via web interface (e.g., cost inquiry).
- Orchestration:** LangChain coordinates three LLM modules sequentially.
- Ontology Retriever:** Embedding-based search uses ChromaDB vector store to fetch medical billing codes from domain ontology.
- SQL LLM:** Quantized model (Qwen Coder 30B) translates ontology retrieval into efficient DuckDB SQL queries for price datasets.
- Summary LLM:** Medical summarizer (Med Gemma 4B) generates plain-language, patient-friendly answers from SQL results.
- Data Storage:** Ontology embeddings (NDC, CPT, HCPCS, RC) in ChromaDB; hospital charge datasets in SQL tables.
- Deployment:** Fully GPU-accelerated pipeline in Google Colab, leveraging PyTorch, DuckDB, ChromaDB, and LangChain.
- Outcome:** Delivers explainable, code-explicit, and robust clinical question answering for hospital pricing transparency.

RESULTS & EVALUATION

Resultant Output:

Query: Which hospitals offer the lowest cost for a colonoscopy?
Answer: Okay, let's break down the cost information for a colonoscopy based on the data provided.

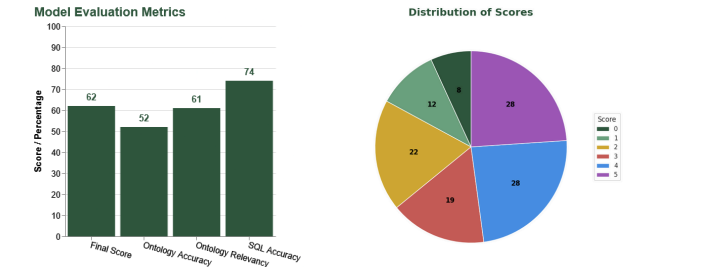
Key Findings:

- Procedure:** The data covers "HC COLONOSCOPY W/ LES REM MD" or "COLONOSCOPY, FLEXIBLE; WITH REMOVAL OF TUMORS)...", which means a colonoscopy that includes the removal of a lesion or tumor.
- Hospitals:** Several [redacted] appear in the list.
- Gross Charges:** The initial listed charges (Gross Charges) for this procedure are \$1483.00 at [redacted].
- Discounted Cash Price:** For these hospitals, the discounted cash price (what you might pay if you pay upfront) is \$889.80.
- Negotiated Price:** The negotiated price (what insurance companies typically pay) varies. For [redacted], it's around \$1124-\$1172. For others like [redacted], the negotiated price is much higher (e.g., \$3669.58).
- Insurance:** The costs shown are for patients with specific insurance plans (Aetna, Cigna, PA Health & Wellness, Tricare, Medicare, Medicaid, Highmark, etc.). Your actual cost will depend heavily on your specific insurance plan and coverage.
- Cost Variability:** There's a significant variation in charges and negotiated prices between different hospitals and insurance plans.

Answer: Based on this data, [redacted] show the lowest initial gross charge (\$1483.00) and the same discounted cash price (\$889.80) for this specific type of colonoscopy. However, the negotiated price (what insurance pays) varies.

Important Note: This information represents charges, not necessarily what you will pay. Your out-of-pocket cost depends on your insurance plan's deductible, co-pays, and co-insurance. To get the most accurate cost estimate, please contact your insurance company to understand your specific coverage and potential out-of-pocket expenses for a colonoscopy at different hospitals. They can provide details tailored to your plan.

Evaluation:



- LLM-as-a-Judge:** GPT-4 was used to assess accuracy, relevance, and SQL code of the pipeline outputs, leveraging its reliability in model evaluation.
- Method:** A custom dataset of 39 representative user queries was constructed. Answers were rated using a modified Likert scale (0-5), where 0 = blank and 5 = most relevant.
- Analysis:** Final scores quantify overall pipeline performance as shown in the above diagram with the overall pipeline achieving **62% accuracy** on the custom varied test set of user queries.

CONCLUSION & FUTURE WORK

One key limitation of this work lies in the inherent complexity and poor data quality of hospital MRF and CDM files. These datasets are often difficult to interpret, containing missing values, inconsistent formats, and vague descriptions. As a result, building an accurate and reliable navigation system has been a significant challenge.

Currently, the legislative efforts intended to promote hospital price transparency remain largely underutilized by the public. Even when accessible, the datasets are difficult to navigate because of their technical structures and the use of specialized billing codes unfamiliar to most patients. Our approach addresses these issues by offering accurate cost predictions paired with transparent, code-based explanations derived directly from federally regulated price data. This design lowers the barrier for non-technical users and aligns with federal policy goals for open and explainable healthcare pricing.

Future Work

- Enhance the pipeline further to reach SOTA model.
- Integrate machine learning models to compare their performance with the current SQL LLM based retrieval system.
- Evaluate which method, symbolic SQL querying or data driven machine learning inference, offers better accuracy, efficiency, and interpretability.
- Collaborate with hospitals and insurance companies to build a publicly accessible patient price navigation platform.
- Conduct field studies to assess how the tool influences patient decision making and cost awareness.

REFERENCE

Centers for Medicare & Medicaid Services. (2021). Hospital Price Transparency Final Rule (CMS-1717-F2). U.S. Department of Health and Human Services. <https://www.cms.gov/hospital-price-transparency>

Centers for Medicare & Medicaid Services. (2022). Transparency in Coverage Rule. U.S. Department of Health and Human Services. <https://www.cms.gov/transparency-coverage>

U.S. House of Representatives. (2023). Lower Costs, More Transparency Act. H.R. 5378, 118th Cong. <https://www.congress.gov/bills/118th-congress/house-bill/5378>

Koos, H. (2025, October 22). Reforming Hospital Price Transparency Enforcement: Closing Loopholes and Empowering Patients. Health Affairs Forefront. <https://www.healthaffairs.org/content/forefront/reforming-hospital-price-transparency-enforcement-closing-loopholes-and-empowering>

Jusko, J. (2025, October 16). Transparency Reveals Health Care Prices – And A Billing And Payments System In Need Of Overhaul. Health Affairs Forefront. https://www.healthaffairs.org/content/forefront/transparency-reveals-health-care-prices-and-billing-and-payments-system-need-overhaul?utm_campaign=forefront&utm_medium=email&utm_source=pmc&utm_term=.0f9c7c2c2gliba3f7gwa98dgnMLUxn0dRVR_1q4EUVF_19kvcv_qHhMnjh3VRfH96wVajGocvCSHwCoP7Eaxg&_hsmi=385522050&utm_source=hat

Sun, E., & Sood, N. (2024). Hospital Price Transparency and Patient Decision-Making. JAMA Health Forum, 5(2), e240145. <https://pmc.ncbi.nlm.nih.gov/articles/PMC11129567/>

Patient Rights Advocate. (2023, February 6). Fourth Semi-Annual Hospital Price Transparency Compliance Report. Retrieved September 2, 2025, from <https://www.patientrightsadvocate.org/fourth-semi-annual-compliance-report>

Kurani, N., Kearney, A., Krzinger, A., & Cox, C. (2021). Few Adults Are Aware of Hospital Price Transparency Requirements. Peterson-KFF Health System Tracker. Retrieved September 2, 2025, from <https://www.healthsystemtracker.org/brief/few-adults-are-aware-of-hospital-price-transparency-requirements/>

Maleki, N., Padmanabhan, B., Dutta, K. Usability of Health Care Price Transparency Data in the United States: Mixed Methods Study. J Med Internet Res. 2024 Mar 29;26:e50629. doi:10.2196/50629.