Understanding the Why: Leveraging Theme-Based Knowledge Graphs for Enhanced Causal Reasoning

Anonymous ACL submission

Abstract

Understanding causality to answer "why" and "how" questions is a formidable challenge for Large Language Models (LLMs), particularly in specialized fields requiring extensive domain knowledge. Although advances in retrieval Augmented Generation (RAG) provide LLMs with access to information beyond their training scope, RAG models struggle to infer implicit causal links and build causal narratives, resulting in incomplete and extremely verbose explanations.

007

009

011

012

013

017

018

019

022

We introduce KG-RAG, a novel framework that integrates theme-based knowledge graphs (KGs) with RAG for enhanced causal inference. KG-RAG leverages GPT-40 to extract explicit and implicit $\langle cause, relation, effect \rangle$ triples from domain-specific corpora. These triples are structured into a directed acyclic graph (DAG) to enable multi-hop causal reasoning.

KG-RAG is evaluated on two datasets: Bitcoin price fluctuations (BP), where financial narratives demand high granularity causal inference, and Gaucher disease (GD), a well-researched medical condition with known causal relations. The results show that KG-RAG outperforms GPT-40 with RAG and GPT-40 baselines, achieving higher readability, chain similarity, and conciseness scores. We also evaluated the results using an LLM-as-a-Judge experiment, combining the expertise of 4 state-of-theart LLMs (GPT-4, GPT-4o, LLaMA 3.1-8B-Instruct, and Mistral-7B-Instruct). The results showed that the performance of the KG-RAG was superior compared to the baselines for both datasets.

038Our findings demonstrate the power of inte-
grating structured knowledge graphs into the
RAG process to improve causal reasoning in
LLMs, paving the way for more interpretable
and reliable AI-driven decision making.

1 Introduction

Understanding and interpreting causal relations embedded within specialized domains, such as healthcare and finance, is crucial to gain actionable insights about the particular domain. Although well-known Large Language Models (LLMs) such as GPT (Achiam et al., 2023), LLaMA (Touvron et al., 2023), and Gemini (Team et al., 2023) excel in tasks like *question answering* (Pichappan et al., 2023; Bahak et al., 2023) and *summarization* (Ravaut et al., 2024), their ability to perform causal inference is poor due to their limited ability to access, understand, and structure domainspecific knowledge (Zečević et al., 2023).

Retrieval-augmented generation (RAG) (Lewis et al., 2020; Mansurova et al., 2024) provides LLMs access to rich knowledge bases whose wideranging and up-to-date information significantly improves the factual accuracy of the LLM's responses. However, the lack of structured causal modeling limits the ability of RAG systems to perform multi-hop causal reasoning.

Classical causal inference methods rely on surface-level correlations via statistical patterns rather than explicitly modeling causal chains, making them insufficient for extracting implicit causal relationships (Weinberg et al., 2024; Cox and Wermuth, 2004; Hassani et al., 2017).

We introduce KG-RAG, a novel framework that integrates theme-based knowledge graphs with RAG for enhanced causal inference. Unlike existing approaches, KG-RAG is able to extract explicit and implicit relations from domain-specific corpora by modeling events as $\langle cause, relation, effect \rangle$ triples and organizing them into a directed acyclic graph (DAG). This structured representation supports both forward and backward chaining, enabling multi-hop causal reasoning and interpretable answer generation.

KG-RAG performance is evaluated on two

043

044

045

047

131

132

133

datasets: Bitcoin price fluctuations (BP) (Wu et al., 2024), where reasoning over financial narratives requires high-granularity causal extraction, and Gaucher disease (GD) (Grabowski, 2008), where well-documented biomedical causal relationships provide a strong testbed. Results show that KG-RAG delivers superior performance compared to the baseline LLM and RAG approaches.

2 Related Work

084

091

097

098

101

102

105

106

107

108

109

110

111

112

113

114

115

116

117

118

119

Causal reasoning has long been a challenging problem in Artificial Intelligence (AI) (Yamin et al., 2024; Joshi et al., 2024). Traditional methods such as *Granger Causality* (Granger, 1969), commonly used in Time Series prediction, and *Propensity Score Matching* (Dehejia and Wahba, 2002) for confounder adjustment in structured datasets, have been widely applied in statistics and econometrics. However, these methods are not well-suited for extracting causal relationships from unstructured text or inferring complex, multi-hop dependencies.

Recent advances in Machine Learning, Large Language Models, and Knowledge Graphs have created new avenues for causal inference.

2.1 Causal Relation Extraction in NLP

Early approaches for causal relation extraction relied on rule-based parsing to detect explicit causal markers (e.g., "because", "due to") (Girju, 2003). Recent Deep Learning approaches leverage pretrained transformers to improve causal relation extraction as in *CausalBERT* (Li et al., 2021) and attention models for relation extraction (Liu et al., 2019; Li et al., 2019). These methods treat causal relations as isolated pairs and do not construct structured causal pathways, inhibiting their ability to model complex causal interactions.

2.2 Machine Learning (ML) Based Causal Inference

Machine Learning methods to learn causal struc-120 tures directly from unstructured text and multi-121 modal data integrate graph neural networks (GNNs) 122 or apply counterfactual reasoning in NLP (Pearl, 123 2009; Schölkopf et al., 2021). Neural network 124 based models, like DECI (Deep End-to-End Causal 125 126 Inference) (Geffner et al., 2022) integrate causal graph discovery with causal effect estimation. 127 However, the reliance of these approaches on high-128 quality data and substantial computational cost limit its suitability for multi-domain contexts. 130

2.3 Knowledge Graph Based Reasoning

Early Knowledge Graph (KG) approaches such as ConceptNet (Speer and Lowry-Duda, 2017) and WordNet (Miller, 1995) focused on hierarchical and semantic relationships but did not explicitly model causal dependencies.

Many recent works (Chen et al., 2023; Fang et al., 2024; Li et al., 2024; Trajanoska et al., 2023; Wei et al., 2024; Zhu et al., 2024; Kau et al., 2024) have explored using LLMs for KG construction. ThemeKG (Ding et al., 2024) addresses challenges of limited information granularity and timeliness in existing KG by constructing theme specific knowledge graphs. The ThemeKG framework processes raw, theme specific corpora to generate high-quality KGs that include salient entities and relations relevant to the theme. This process involves developing an entity ontology from sources like Wikipedia, generating candidate relations using LLMs to construct a relation ontology, and parsing documents to consolidate relations for entity pairs.

However, ThemeKG does not explicitly model causal pathways or support multi-hop causal reasoning. In contrast, KG-RAG extends this paradigm by integrating causal dependencies into the knowledge graph, enabling *forward* (cause \rightarrow effect) and *backward* (effect \rightarrow cause) *chaining* for effective and interpretable causal question answering.

2.4 Retrieval Augmented Generation (RAG)

RAG enhances factual accuracy by incorporating the retrieved document context into the LLM outputs (Lewis et al., 2020). Traditional RAG approaches primarily retrieve independent document chunks without explicitly capturing causal dependencies.

Multi-hop retrieval models (Asai et al., 2019) improve long-form reasoning by retrieving supporting evidence across multiple documents but lack a structured approach to causal modeling, limiting their effectiveness for causal inference.

In summary, existing methods have the following challenges in causal reasoning:

- **Implicit Causality Handling**: Previous models rely on explicit causal markers and struggle with implicit causal reasoning, failing to infer causal links from nuanced textual patterns.
- Lack of Structured Multi-Hop Reasoning: Most causal extraction models process relations

275

228

229

230

as isolated pairs rather than integrating them intostructured causal pathways.

182

183

184

186

187

188

190

191

192

193

194

195

197

198

206

210

211

212

213

214

215

216

218

219

220

221

227

- Reliance on Statistical Correlations: ML causal discovery methods detect statistical patterns rather than explicit causal structures, making them less interpretable.
- Scalability Constraints: Many causal inference models require large labeled datasets or high computational resources, limiting their realworld effectiveness.
- Limited Explanatory Power: Existing approaches generate fragmented or verbose outputs that are less interpretable compared to structured causal narratives.

Our **KG-RAG** framework makes the following key innovations to overcome the above limitations:

- Unsupervised Extraction of (*cause, relation, effect*) Triples. We develop an unsupervised framework that extracts explicit and implicit causal relations from unstructured domain-specific corpora. This method leverages LLM-driven extraction with retrieval-aware constraints to improve the precision of causal knowledge representation without requiring annotated training data.
- 2. Theme-Based Knowledge Graph Construction for Causal Representation. We propose a novel approach to structure extracted causal triples into a Directed Acyclic Graph (DAG) which ensures a coherent and structured representation of causal relationships. This approach enables better organization and retrieval of causal knowledge for specialized domains.
- 3. Multi-hop Causal Reasoning via Forward and Backward Chaining. The above generated DAG is traversed using both forward and backward chaining for multi-hop inference. This structured inference in turn improves the LLM's ability to generate interpretable causal narratives.

These contributions position KG-RAG as a competitive approach that bridges knowledge graph reasoning, causal inference, and retrieval-augmented generation to advance the capabilities of LLM based causal reasoning.

3 Methodology

The KG-RAG framework leverages LLMs for entity extraction to construct causal KGs, chain selection to identify the most relevant causal pathways, and summarization to generate concise, humanreadable explanations from extracted relationships. We outline the key components of our framework below.

3.1 Entity Extraction and DAG Construction

KG-RAG uses GPT-40 to extract $\langle cause, relation, effect \rangle$ triples from our text corpora. We use a prompt based approach described below to identify both explicit (word-based) and implicit (an inferred connection that can be understood contextually) causal links, ensuring a more comprehensive representation of causal relationships. Further, when applicable, the extracted entities and relations are aligned with domain-specific taxonomies (e.g., **MeSH** (National Library of Medicine, 1960) for biomedical datasets) to enhance consistency and interpretability.

Given a document, the model is prompted to extract causal triples, where:

- **Cause** refers to an entity, event, or action that triggers an outcome, even if the causal connection is not explicitly stated.
- **Relation** is a causal verb or phrase (e.g., *caused*, *led to*, *resulted in*, *triggered*, *influenced*), or an inferred connection that is understood contextually.
- Effect represents the resulting entity, event, or action, regardless of whether the causal relationship is directly stated in the text.

For example, given the text: "Bitcoin prices surged after major financial institutions significantly increased their holdings." our method extracts:

\langle Major financial institutions increase holdings, influenced, Bitcoin price surge \rangle

Here, no explicit causal phrase such as "because" or "due to" is present. The LLM identifies that a significant increase in institutional holdings is a plausible causal factor for Bitcoin's price surge even though this causal relationship is not explicitly stated in the text.

The extracted triples are used to construct a **Directed Acyclic Graph (DAG)**, which serves as the foundation for structured causal reasoning. The DAG captures both direct and indirect dependencies by explicitly mapping causal chains as shown in Figure 1.



Figure 1: KG Representing Causal Relationships in Bitcoin Price Fluctuations

3.2 User Query Processing

Given a user query, we extract key entities and concepts using *KeyBERT* (Grootendorst, 2020), a transformer-based keyword extraction to identify the most relevant words and phrases from the query while filtering out stop words and irrelevant terms. These extracted keywords and phrases are matched against nodes in the KG using *Sentence-BERT* (Reimers, 2019) embeddings, allowing semantic similarity-based retrieval of relevant entities.

The user query's causal direction is classified as *forward* (identifying effects) or *backward* (identifying causes) using an LLM prompt. This classification ensures that the retrieval process aligns with the user's intent. The extracted key entities are used to retrieve the relevant causal pathways from the KG via forward or backward chaining depending on the identified causal direction.

3.3 KG-LLM Querying and Graph-Based Reasoning

The LLM enhanced querying strategy on the DAG supports:

- Forward Chaining: Starting from an initial cause, traverse downstream effects to construct a causal sequence.
- **Backward Chaining:** Given an observed effect, trace upstream paths to uncover potential causes.

This bidirectional reasoning allows KG-RAG to generate answers that not only respond to the query

but also provide structured, explainable causal pathways. 305

306

307

308

309

310

311

312

313

314

315

316

317

318

319

320

321

322

323

324

325

326

328

330

331

332

333

335

337

338

339

341

343

344

3.4 Causal Chain Selection

Since many causal chains may exist for a given query, we leverage the capabilities of the LLM to rank candidate chains dynamically. The model evaluates the relevance and coherence of each chain based on the query context, incorporating aspects such as factual alignment, informativeness, and logical progression.

This approach ensures that the selected responses are accurate and easily interpretable. Prior studies have shown that LLMs excel in ranking tasks by leveraging vast contextual embeddings to assess *semantic similarity* and *coherence* (Mann et al., 2020; Raffel et al., 2020). Furthermore, by dynamically adapting to different queries, LLM based ranking mitigates biases introduced by predefined heuristics, resulting in more flexible and contextually appropriate selections.

3.5 Output Generation with Justification

The final step synthesizes the reasoning process into a natural language response using LLMpowered summarization and explanation. Given the selected causal pathway, the LLM generates a structured response that includes:

- A direct answer to the user's query, grounded in the retrieved causal chain.
- A justification explaining the reasoning process, explicitly tracing the multi-hop causal pathway identified in the knowledge graph.

Figure 3 shows example outputs from each step of the process described above. This ensures transparency in causal reasoning, addressing the opacity of traditional LLM outputs. By leveraging an LLM for both structured response generation and justification, our approach improves trustworthiness and interpretability in AI-driven causal analysis.

4 **Experiments**

4.1 Dataset Collection

We constructed two datasets, Bitcoin Price Fluctua-
tions (BP) in the finance domain and Gaucher Dis-
ease (GD) in the medical domain. These datasets
are sourced from a combination of news articles
from Reuters, Financial Times, Bloomberg, etc.345
346
346
347

304



Figure 2: RAG vs. KG-RAG Processes

and scientific papers from PubMed and the National Institute of Health (NIH). The documents were selected based on the following criteria:

351

353

367

- Relevance: The document explicitly discusses causal relationships within the theme.
- Reliability: Priority to sources with established credibility (e.g., peer reviewed papers, major news outlets).
- Granularity: Documents provide sufficient detail to extract meaningful relations.
- Diversity: Documents represent a range of perspectives and subtopics.

We collected and processed 20 high-quality documents in each dataset and extracted entities and causal triples from the text using GPT-40 as described in Section 3.1. The final dataset statistics are shown in Table 1.

4.2 Baseline Comparisons

We compare KG-RAG against two baselines.

Dataset	Documents	Entities	Triples
Bitcoin Price (BP)	20	317	172
Gaucher Disease (GD)	20	573	419

Table 1: Dataset Statistics

Query:	How did public statements from key leaders impact the price of Bitcoin after the 2024 US Presidential Election?
Extracted Keyphrases:	['key leaders impact', 'bitcoin 2024 presidential', 'impact price bitcoin']
Candidate Causal Chains:	Chain 1: President-elect Donaid Trump's election win and subsequent announcements about his second term's priorities around deregulation -> Bitcoin Chain 2: President-elect Trump -> VS crypto capital of the world Chain 3: President-elect Tomp -> Fed Chief Powell's comments on Bitcoin -> hopes that more monetary policy leaders will embrace the space Chain 4: Trump named David Sacks as his crypto and Al czar -> helping formulate crypto and Al policy from the White House Chain 5: Elon Musk post -> surging and tumbling sharply of crypto currencies
Extracted Causal Chain:	President-elect Donald Trump -> Fed Chief Powell's comments on Bitcoin -> hopes that more monetary policy leaders will embrace the space
Justificaton:	Chain 3 is most relevant as it directly involves Jerome Powell's comments impacting the price of Bitcoin after the 2024 US Presidential Election, given the context of Trump's administration and its pro-crypto stance.
Summary:	Following the 2024 US Presidential Election, President-elect Trump's selection of a pro-crypto SEC chief and the public statements from Federal Reserve Chair Jerome Powell had a notable impact on the price of Bitcoin. Powell's comments, particularly those addressing the regulatory stance and monetary policy outlook, influenced market sentiment possitively, contributing to a rise in Bitcoin's value. The anticipation of a more favorable regulatory environment under Trump's administration reinforced this upward trend in the cryptocurrency market.

Figure 3: Bitcoin KG-RAG Example

• GPT-40 with RAG: A retrieval-augmented system that grounds GPT-40 using locally retrieved document chunks. We implemented this by leveraging structured chunking (Lewis et al., 2020) for efficient segmentation and a retrieval mechanism that dynamically selects the most relevant context for each query, ensuring accurate and context-aware responses.

369

370

371

372

373

374

375

376

379

380

381

385

386

387

388

390

391

392

393

394

395

396

397

398

• Vanilla GPT-40: A baseline that employs GPT-40 out of the box without any augmentation.

4.3 Evaluation Metrics

The performance of the three systems was evaluated using three key metrics that capture different aspects of response quality:

• **Readability:** Measured using the Flesch readability formula (Flesch, 1948), this metric quantifies how easy a response is to read by considering factors such as average sentence length and word complexity (measured via syllable count). In our context, a higher Readability score indicates that the text uses simpler, more accessible language, making it easier for a broad audience to understand. The formula is given by:

Readability = $206.835 - 1.015 \times \left(\frac{N_{\text{words}}}{N_{\text{sentences}}}\right) - 84.6 \times \left(\frac{N_{\text{syllables}}}{N_{\text{words}}}\right)$

• Chain Similarity: Computes the cosine similarity between the sentence embeddings of the generated causal chain and a reference (gold standard) causal chain. The reference chains were manually curated based on a thorough review of domain-specific literature. These chains

474

475

476

477

478

479

480

481

482

483

484

485

486

446

447

448

were constructed to reflect the most widely accepted causal relationships pertaining to each 400 question. Chain similarity quantifies the seman-401 tic alignment between the system's reasoning 402 and the expected reasoning steps. A higher chain 403 similarity score implies that the generated chain 404 closely follows the semantic structure of the ref-405 erence, indicating that the system successfully 406 captured the key causal relationships. 407

408

409 410

411

412

413

414

415

416

417

418

419

420

421

422

423

424

425

426

497

428

429

430

431

432

433

434

435

436

437

438

439

440 441

442

443

444

445

• **Conciseness:** Quantifies the succinctness of a response based on the *lexical density*—the ratio of content-rich words (excluding common stop-words) to the total word count—which captures the information density of the text normalized by the logarithm of the word count and scaling by 100. The formula is given by:

Conciseness Score =
$$100 \times \frac{\frac{N_{\text{content}}}{N_{\text{total}}}}{\ln (N_{\text{total}})}$$

where N_{content} is the number of content words and N_{total} is the total word count. In our context, a higher conciseness score indicates that a response is both information dense and succinct, delivering key insights without being needlessly verbose. Conversely, lower scores indicate the answer might be padded with fluff, reducing its overall effectiveness.

Overall Interpretation: High values across these metrics indicate that the system generates responses which are clear, semantically aligned with the intended reasoning, and succinct. Our goal is to ensure that the system not only provides correct and comprehensive information but does so in a manner that is easily interpretable and free of irrelevant information.

4.4 LLM-as-a-Judge

To complement our automated metrics, we conducted an LLM-as-a-Judge (Zheng et al., 2024) evaluation using a panel of four models: GPT-4, GPT-4o, LLaMA 3.1-8B-Instruct, and Mistral-7B-Instruct. We present each model with three answer choices—KG-RAG, LLM w/ RAG, and Vanilla LLM—and ask it to select the best response based on correctness, level of detail, and conciseness.

A response is selected as the preferred answer if at least three out of four models favor it, or if votes are split across three different answers (e.g., 2-1-1), then the answer with the highest number of votes is selected as the winner. This method allows us to incorporate qualitative assessments from multiple LLMs, ensuring a balanced evaluation that accounts for multiple aspects of response quality.

4.5 Infrastructure

All of our experiments were conducted on Google Colab (Google, 2025), leveraging its GPU resources for our work. We estimate that we utilized approximately 50 GPU hours to execute all experiments.

5 Results

In this section, we quantitatively assess framework performance by comparing key evaluation metrics—*readability, chain similarity, and conciseness*—across both datasets.

5.1 Quantitative Metrics

Table 2 summarizes our evaluation metrics for two case studies. For the Bitcoin (BP) dataset, KG-RAG achieves both a higher chain similarity and a higher conciseness score. This indicates that approach provides more relevant responses and that the answers are more focused and to the point. The readability score of KG-RAG is slightly lower than GPT-40 with RAG reflecting more detailed explanations that use complex wording. Notably, all scores for Vanilla GPT-40 are 0 for BP because it was unable to address questions related to Bitcoin price fluctuations post the 2024 US presidential election—queries that fall outside its training scope.

For the GD dataset, KG-RAG outperforms both GPT-40 with RAG and Vanilla GPT-40 across all metrics. These comprehensive improvements indicate that KG-RAG is more effective in providing detailed, precise, and contextually relevant responses.

Dataset	Metric	KG-RAG	GPT-40 with RAG	Vanilla GPT-40
	Readability	15.67	15.75	0.00
BP	Chain Similarity	0.80	0.74	0.00
	Conciseness	16.44	13.94	0.00
	Readability	9.64	3.32	11.55
GD	Chain Similarity	0.70	0.68	0.62
	Conciseness	15.68	13.20	11.66

Table 2: Quantitative Evaluation Metrics

The evaluation unearthed limitations with the *chain similarity* metric wherein the metric at times yields high scores when a generated response closely mirrors the phrasing of the input question rather than providing relevant insight. In such

530

531

542 543

541

544 545

546

547

548

549

550

551

552

553

554

555

556

558

559

560

561

562

565

487 cases, qualitative evaluation by LLM-as-a-Judge
488 proved to be a more reliable indicator of answer
489 quality. This further underscores the need for com490 plementary evaluation methods to ensure robust
491 assessment of causal reasoning performance.

5.2 LLM-as-a-Judge Evaluation Results

492

493

494

495

496

497

498

499

503

504 505

509

510

512

513

514

515

516

517

518

519

520

522

524

525

526

529

Table 3 summarizes the LLM-as-a-Judge results. In the BP dataset, the advantage of KG-RAG is very apparent, achieving 14/18 votes. In contrast, both GPT-40 with RAG and Vanilla GPT-40 scored 0/18. The "Split" score of 4/18 indicates that the judge LLMs were divided in their opinion—two LLMs favored one answer while the other two preferred an alternative. This split underscores the ambiguity in some responses but still highlights the strong performance of KG-RAG.

For the GD dataset, KG-RAG also performed quite well, with a score of 15/25. It significantly outperformed both GPT-40 with RAG (3/25) and Vanilla GPT-40 (2/25), while the "Split" category, with a score of 5/25, again reflected a divided judgment among the LLMs.

Dataset	KG-RAG	GPT-40 with RAG	Vanilla GPT-40	Split
BP	14/18	0/18	0/18	4/18
GD	15/25	3/25	2/25	5/25

Table 3: LLM-as-a-Judge Results

Overall, these results reinforce the capability of KG-RAG and suggest that integrating structured causal reasoning into RAG significantly improves answer interpretability and accuracy, particularly in specialized domains requiring multi-hop causal inference.

6 Ablation Study

We conducted an ablation study in which we replaced the downstream analysis LLM (GPT-40) with LLaMA-3.1-8B-Instruct. In this variant, GPT-40 is still used for high quality entity extraction and causal graph construction, but all subsequent processing—including candidate chain analysis and selection, and chain summarization—is performed using LLaMA. The objective is to assess whether KG-RAG's structured causal reasoning remains effective even with a less powerful model.

6.1 Quantitative Results

Table 4 summarizes the ablation study results using LLaMA instead of GPT-40. Here, we compare three configurations—LLaMA KG-RAG, LLaMA with RAG, and Vanilla LLaMA—across both datasets.

LLaMA KG-RAG achieves higher chain similarity and conciseness than LLaMA with RAG and Vanilla LLaMA indicating that the responses are more relevant and focused. The lower readability score for LLaMa KG-RAG compared to LLaMa with RAG and Vanilla LLaMA are due to KG-RAG's more technically nuanced explanations as discussed in the appendix.

A similar trend is observed in the GD dataset. These results confirm that the KG-RAG architecture effectively improves performance metrics because incorporating structure into the RAG process compensates for a less powerful model like LLaMA.

Dataset	Metric	LLaMA KG-RAG	LLaMA with RAG	Vanilla LLaMA
	Readability	27.85	44.69	50.96
BP	Chain Similarity	0.76	0.63	0.61
	Conciseness	12.04	9.40	8.96
	Readability	17.19	32.40	23.54
GD	Chain Similarity	0.64	0.61	0.60
	Conciseness	14.93	10.28	10.40

Table 4: Dataset Evaluation Metrics

6.2 LLM-as-a-Judge Evaluation

The LLM-as-a-Judge evaluations further support KG-RAG's superior performance even when a less powerful model like LLaMa is used as shown in Table 5.

Dataset	LLaMA KG-RAG	LLaMA with RAG	Vanilla LLaMA	Split
BP	13/25	2/25	8/25	2/25
GD	14/18	1/18	2/18	1/18

Table 5: LLM-as-a-Judge Results

The ablation study highlights the value of integrating a structured Knowledge Graph into an LLM based pipeline and its ability to maintain high performance in resource constrained settings making KG-RAG a promising solution for scalable causal reasoning applications.

7 Future Work

While KG-RAG demonstrates strong performance in structured causal reasoning, the following areas merit future exploration.

- **Multimodal Knowledge Integration**: Extending KG-RAG to incorporate images, graphs, and tabular data would enhance its applicability.
- Adaptive Retrieval Mechanisms: While our approach relies on static knowledge graphs, de-

- veloping dynamic retrieval mechanisms that update causal representations in real-time would
 provide relevant and up-to-date answers in
 rapidly evolving domains.
- Scaling to Broader Domains: While we validated KG-RAG on biomedical and financial datasets, we plan to explore its effectiveness across other fields, including social sciences, legal reasoning, and supply chain analysis.
 - Human-in-the-Loop Evaluation: While the LLM-as-a-Judge evaluation provides a scalable assessment mechanism, incorporating domain experts for human validation would ensure a more rigorous evaluation of KG-RAG's generated causal explanations.

By addressing these, KG-RAG can evolve into a more versatile and scalable framework, bridging the gap between knowledge-driven AI and realworld decision-making.

8 Conclusion

576

578

580

584

585

589

591

593

596

599

602

603

606

607

610

612

In this paper, we introduced **KG-RAG**, a structured causal reasoning framework that integrates knowledge graphs with RAG to improve multi-hop causal inference. Unlike conventional RAG approaches, KG-RAG explicitly models causal pathways using Directed Acyclic Graphs (DAGs), enabling more coherent, interpretable, and structured causal reasoning.

Through evaluation on two datasets—Gaucher Disease and Bitcoin Price Fluctuations—we demonstrate that KG-RAG outperforms standard RAG and vanilla LLM baselines both in terms of quantitative metrics and a qualitative assessment using LLM-as-a-Judge. Our ablation study further confirms that even with less powerful LLM (LLaMA 3.1-8B-Instruct), the structured causal reasoning in KG-RAG remains effective.

KG-RAG has significant implications for realworld decision-making in high-stakes domains:

- Biomedical Research: By structuring relationships between genetic, environmental, and pharmaceutical factors, KG-RAG can support disease progression modeling and drug discovery.
- Financial Market Analysis: KG-RAG can provide a structured understanding of macroeconomic and geopolitical influences, improving risk assessment in volatile markets.

• **Policy and Legal Analysis**: The ability to trace causal responsibility in regulatory compliance, legal disputes, and public policy can enhance transparency and accountability.

613

614

615

616

617

618

619

620

621

622

623

624

625

626

627

628

629

630

631

632

633

634

635

636

637

638

639

640

641

642

643

644

645

646

647

648

649

650

651

652

653

654

655

656

657

658

659

These findings highlight the importance of explicit causal modeling in LLM-driven reasoning. By integrating structured knowledge representations, KG-RAG moves beyond surface-level retrieval to produce explanations that are more interpretable, reliable, and well-suited for real-world decision making. As AI systems continue to evolve, the fusion of causal reasoning and structured knowledge retrieval will be essential for enabling transparent, explainable, and trustworthy AI-driven insights, particularly in domains where understanding "why" and "how" is critical for informed action.

9 Limitations

We recognize that the KG-RAG framework has the following limitations.

- KG-RAG relies on pre-trained large language models (LLMs) for entity and relation extraction, which introduces potential hallucination and bias issues. LLMs may extract spurious causal links, particularly in low-resource domains, and reflect biases from training data, limiting their reliability in emerging topics. Future work could incorporate domain-specific fine-tuning to mitigate these effects.
- As knowledge graphs grow, causal graph traversal can become computationally expensive, particularly for multi-hop reasoning. Optimizations such as graph pruning or embedding-based traversal would need to be implemented to enable scaling to large-scale applications.
- LLM systems are inherently non-deterministic. As a result, executing KG-RAG multiple times even with the same query can yield varying results, potentially producing different answers on each run.
- Although prompting strategies are designed to guide LLM responses toward desired outcomes, during the summarization phase the LLM may deviate from the intended chain, leading to hallucinations or inaccuracies.
- Recent advances in reasoning models such as OpenAI's o1 (OpenAI, 2024) and o3 (OpenAI, 2025), and DeepSeek's R1 (DeepSeek-AI et al.,

6602024) show promise in logical reasoning and661causal inference. We are currently unable to662benchmark KG-RAG against these models due663to the lack of public API access. As these mod-664els become more accessible, we will explore665their effectiveness in knowledge graph integra-666tion and causal inference.

667 Despite these limitations, we believe that KG-RAG
668 represents a significant step forward in structured
669 causal reasoning. Addressing these limitations will
670 further enhance KG-RAG's ability to provide ro671 bust, scalable, and interpretable causal insights.

References

682

694

697

701 702

703

704 705

707

708

710

711

- Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Akari Asai, Kazuma Hashimoto, Hannaneh Hajishirzi, Richard Socher, and Caiming Xiong. 2019. Learning to retrieve reasoning paths over wikipedia graph for question answering. *arXiv preprint arXiv:1911.10470*.
- Hossein Bahak, Farzaneh Taheri, Zahra Zojaji, and Arefeh Kazemi. 2023. Evaluating chatgpt as a question answering system: A comprehensive analysis and comparison with existing models. *arXiv preprint arXiv:2312.07592*.
- Jiao Chen, Luyi Ma, Xiaohan Li, Nikhil Thakurdesai, Jianpeng Xu, Jason HD Cho, Kaushiki Nag, Evren Korpeoglu, Sushant Kumar, and Kannan Achan. 2023. Knowledge graph completion models are few-shot learners: An empirical study of relation labeling in e-commerce with llms. *arXiv preprint arXiv:2305.09858*.
- David R Cox and Nanny Wermuth. 2004. Causality: a statistical view. *International Statistical Review*, 72(3):285–305.
- DeepSeek-AI, Xiao Bi, Deli Chen, Guanting Chen, Shanhuang Chen, Damai Dai, Chengqi Deng, Honghui Ding, Kai Dong, Qiushi Du, Zhe Fu, Huazuo Gao, Kaige Gao, Wenjun Gao, Ruiqi Ge, Kang Guan, Daya Guo, Jianzhong Guo, Guangbo Hao, Zhewen Hao, Ying He, Wenjie Hu, Panpan Huang, Erhang Li, Guowei Li, Jiashi Li, Y. K. Li, Wenfeng Liang, Fangyun Lin, A. X. Liu, Bo Liu, Wen Liu, Xiaodong Liu, Xin Liu, Yiyuan Liu, Haoyu Lu, Shanghao Lu, Fuli Luo, Shirong Ma, Xiaotao Nie, Tian Pei, Yishi Piao, Junjie Qiu, Hui Qu, Tongzheng Ren, Zehui Ren, Chong Ruan, Zhangli Sha, Zhihong Shao, Junxiao Song, Xuecheng Su, Jingxiang Sun, Yaofeng Sun, Minghui Tang, Bingxuan Wang, Peiyi Wang, Shiyu Wang, Yaohui Wang, Yongji Wang,

Tong Wu, Y. Wu, Xin Xie, Zhenda Xie, Ziwei Xie, Yiliang Xiong, Hanwei Xu, R. X. Xu, Yanhong Xu, Dejian Yang, Yuxiang You, Shuiping Yu, Xingkai Yu, B. Zhang, Haowei Zhang, Lecong Zhang, Liyue Zhang, Mingchuan Zhang, Minghua Zhang, Wentao Zhang, Yichao Zhang, Chenggang Zhao, Yao Zhao, Shangyan Zhou, Shunfeng Zhou, Qihao Zhu, and Yuheng Zou. 2024. Deepseek Ilm: Scaling opensource language models with longtermism. *arXiv preprint arXiv:2401.02954*.

713

714

715

716

717

721

722

723

724

725

726

727

728

729

730

731

732

733

734

735

736

737

738

739

740

741

742

743

744

745

746

747

748

749

750

751

752

753

754

755

756

757

758

759

761

762

763

764

765

766

- Rajeev H Dehejia and Sadek Wahba. 2002. Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and statistics*, 84(1):151–161.
- Linyi Ding, Sizhe Zhou, Jinfeng Xiao, and Jiawei Han. 2024. Automated construction of theme-specific knowledge graphs. *arXiv preprint arXiv:2404.19146*.
- Junfeng Fang, Shuai Zhang, Chang Wu, Zhengyi Yang, Zhiyuan Liu, Sihang Li, Kun Wang, Wenjie Du, and Xiang Wang. 2024. Moltc: Towards molecular relational modeling in language models. *arXiv preprint arXiv:2402.03781*.
- Rudolph Flesch. 1948. A new readability yardstick. *Journal of applied psychology*, 32(3):221.
- Tomas Geffner, Javier Antoran, Adam Foster, Wenbo Gong, Chao Ma, Emre Kiciman, Amit Sharma, Angus Lamb, Martin Kukla, Nick Pawlowski, et al. 2022. Deep end-to-end causal inference. *arXiv preprint arXiv:2202.02195*.
- Roxana Girju. 2003. Automatic detection of causal relations for question answering. In *Proceedings of the ACL 2003 Workshop on Multilingual Summarization and Question Answering*, pages 76–83, Sapporo, Japan. Association for Computational Linguistics.
- Google. 2025. Google colaboratory. https://colab. research.google.com/.
- Gregory A Grabowski. 2008. Phenotype, diagnosis, and treatment of gaucher's disease. *The Lancet*, 372(9645):1263–1271.
- Clive WJ Granger. 1969. Investigating causal relations by econometric models and cross-spectral methods. *Econometrica: journal of the Econometric Society*, pages 424–438.
- Maarten Grootendorst. 2020. Keybert: Minimal keyword extraction with bert.
- Hossein Hassani, Xu Huang, and Mansi Ghodsi. 2017. Big data and causality. *Annals of Data Science*.
- Nitish Joshi, Abulhair Saparov, Yixin Wang, and He He. 2024. Llms are prone to fallacies in causal inference. *arXiv preprint arXiv:2406.12158*.
- Amanda Kau, Xuzeng He, Aishwarya Nambissan, Aland Astudillo, Hui Yin, and Amir Aryani. 2024. Combining knowledge graphs and large language models. *arXiv preprint arXiv:2407.06564*.

Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474.

770

777

778

779

781

787

788

789

790

791

794

801

804

810

811

812

813

814

815

816

817

818

819

821

- Dawei Li, Zhen Tan, Tianlong Chen, and Huan Liu. 2024. Contextualization distillation from large language model for knowledge graph completion. *arXiv preprint arXiv:2402.01729*.
- Pengfei Li, Kezhi Mao, Xuefeng Yang, and Qi Li. 2019. Improving relation extraction with knowledgeattention. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP), pages 229–239, Hong Kong, China. Association for Computational Linguistics.
 - Zhongyang Li, Xiao Ding, Kuo Liao, Bing Qin, and Ting Liu. 2021. Causalbert: Injecting causal knowledge into pre-trained models with minimal supervision. *arXiv preprint arXiv:2107.09852*.
- Sijia Liu, Liwei Wang, Vipin Chaudhary, and Hongfang Liu. 2019. Attention neural model for temporal relation extraction. In *Proceedings of the 2nd Clinical Natural Language Processing Workshop*, pages 134– 139, Minneapolis, Minnesota, USA. Association for Computational Linguistics.
- Ben Mann, N Ryder, M Subbiah, J Kaplan, P Dhariwal, A Neelakantan, P Shyam, G Sastry, A Askell, S Agarwal, et al. 2020. Language models are fewshot learners. arXiv preprint arXiv:2005.14165, 1.
- Aigerim Mansurova, Aiganym Mansurova, and Aliya Nugumanova. 2024. Qa-rag: Exploring llm reliance on external knowledge. *Big Data and Cognitive Computing*, 8(9):115.
- George A Miller. 1995. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41.
- National Library of Medicine. 1960. Medical Subject Headings (MeSH). Accessed: February 15, 2025.
- OpenAI. 2024. O1 system card.
- OpenAI. 2025. Openai o3-mini system card.
 - Judea Pearl. 2009. *Causality*. Cambridge university press.
 - Pit Pichappan, M Krishnamurthy, and P Vijayakumar. 2023. Analysis of chatgpt as a question-answering tool. *Journal of Digital Information Management*, 21(2):50–60.
 - Colin Raffel, Noam Shazeer, Adam Roberts, Katherine Lee, Sharan Narang, Michael Matena, Yanqi Zhou, Wei Li, and Peter J Liu. 2020. Exploring the limits of transfer learning with a unified text-to-text transformer. *Journal of machine learning research*, 21(140):1–67.

Mathieu Ravaut, Aixin Sun, Nancy Chen, and Shafiq Joty. 2024. On context utilization in summarization with large language models. In Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers), pages 2764–2781, Bangkok, Thailand. Association for Computational Linguistics. 822

823

824

825

826

827

829

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

851

852

853

854

855

856

857

858

859

860

861

862

863

864

865

866

867

868

869

870

871

872

873

874

- N Reimers. 2019. Sentence-bert: Sentence embeddings using siamese bert-networks. *arXiv preprint arXiv:1908.10084*.
- Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. 2021. Toward causal representation learning. *Proceedings of the IEEE*, 109(5):612–634.
- Robyn Speer and Joanna Lowry-Duda. 2017. Concept-Net at SemEval-2017 task 2: Extending word embeddings with multilingual relational knowledge. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*, pages 85–89, Vancouver, Canada. Association for Computational Linguistics.
- Gemini Team, Rohan Anil, Sebastian Borgeaud, Jean-Baptiste Alayrac, Jiahui Yu, Radu Soricut, Johan Schalkwyk, Andrew M Dai, Anja Hauth, Katie Millican, et al. 2023. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805*.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.
- Milena Trajanoska, Riste Stojanov, and Dimitar Trajanov. 2023. Enhancing knowledge graph construction using large language models. *arXiv preprint arXiv:2305.04676*.
- Yanbin Wei, Qiushi Huang, James T Kwok, and Yu Zhang. 2024. Kicgpt: Large language model with knowledge in context for knowledge graph completion. *arXiv preprint arXiv:2402.02389*.
- Abraham Itzhak Weinberg, Cristiano Premebida, and Diego Resende Faria. 2024. Causality from bottom to top: A survey. *arXiv preprint arXiv:2403.11219*.
- Jingyang Wu, Xinyi Zhang, Fangyixuan Huang, Haochen Zhou, and Rohtiash Chandra. 2024. Review of deep learning models for crypto price prediction: implementation and evaluation. *arXiv preprint arXiv:2405.11431*.
- Khurram Yamin, Shantanu Gupta, Gaurav R Ghosal, Zachary C Lipton, and Bryan Wilder. 2024. Failure modes of llms for causal reasoning on narratives. *arXiv preprint arXiv:2410.23884*.

879

Matej Zečević, Moritz Willig, Devendra Singh Dhami, and Kristian Kersting. 2023. Causal parrots: Large language models may talk causality but are not causal. *arXiv preprint arXiv:2308.13067*.

Lianmin Zheng, Wei-Lin Chiang, Ying Sheng, Siyuan Zhuang, Zhanghao Wu, Yonghao Zhuang, Zi Lin, Zhuohan Li, Dacheng Li, Eric Xing, et al. 2024. Judging llm-as-a-judge with mt-bench and chatbot arena. Advances in Neural Information Processing Systems, 36.

Yuqi Zhu, Xiaohan Wang, Jing Chen, Shuofei Qiao, Yixin Ou, Yunzhi Yao, Shumin Deng, Huajun Chen, and Ningyu Zhang. 2024. Llms for knowledge graph construction and reasoning: Recent capabilities and future opportunities. *World Wide Web*, 27(5):58.

A Appendix

In this section, we provide specific examples of responses from KG-RAG, the LLM with RAG, and the Vanilla LLM from our experiments and the ablation study..

891

892

893

894

895

896

897

898

899

900

901

902

903

904

905

906

907

908

909

910

911

912

913

914

915

916

A.1 Experiment Results and Analysis

Tables 6 and 7 show the responses from the KG-RAG and the baseline approaches to two sample queries from the BP dataset.

Query	What influence did major institutional players
	like BlackRock and Fidelity have on Bitcoin's
	post-election price surge?
KG-RAG	Following Election Day, major institutional
	players such as BlackRock and Fidelity sig-
	nificantly influenced the Bitcoin price surge.
	Their substantial investments and increasing
	acceptance of Bitcoin as a valuable asset
	boosted investor confidence and drew more
	capital into the cryptocurrency market, con-
	tributing to the dramatic increase in Bitcoin's
	price
GPT-40 with RAG	Major institutional players like BlackRock and
	Fidelity significantly influenced Bitcoin's post-
	election price surge by participating in the mar-
	ket through newly approved exchange-traded
	funds (ETEs) that invest directly in the cryn-
	tocurrency. The involvement of these major
	asset managers broadened Bitcoin's potential
	investor base and contributed to its increased
	ministream adoption. This heightened insti-
	tutional participation, coupled with favorable
	regulatory developments, he set ad investor con
	federate and played a major role in the array
	nuence and played a major role in the cryp-
	tocurrency's price surge.
Vanilla GPT-40	N/A

Table 6: Responses from different approaches for BP Query 1

A qualitative analysis of BP Query 1 in Table 6 shows the following:

• Strengths of KG-RAG:

- Clarity and Conciseness: KG-RAG delivers a succinct explanation that directly links institutional investment to Bitcoin's price surge, correctly answering the question.
- Core Causal Factors: It successfully highlights the main causal links, such as increased investor confidence and capital inflow driven by the acceptance of Bitcoin as a valuable asset.

• Weaknesses of KG-RAG:

 Missing Details: KG-RAG's answer doesn't include specific details-such as the role of newly approved ETFs—that are present in the GPT-40 with RAG's output.

Query	Why did Bitcoin's price surge after the 2024
Query	US avaidantial abatian?
	US presidential election?
KG-RAG	Bitcoin's price surged by more than 50 percent
	after the 2024 US presidential election due
	to Donald Trump's victory. His election win
	boosted investor confidence, driving increased
	demand for Bitcoin and resulting in the sub-
	stantial price increase.
GPT-40 with RAG	Bitcoin's price surged after the 2024 US pres-
	idential election due to several factors, pri-
	marily the pro-crypto stance and policies of
	the incoming Trump administration. Donald
	Trump's re-election brought optimism among
	investors about favorable regulatory changes,
	including the potential nomination of Paul
	Atkins, a crypto-friendly former regulator, as
	the chair of the Securities and Exchange Com-
	mission. Trump's campaign promises to make
	the United States the crypto capital of the plan-
	etänd the nomination of crypto-enthusiasts for
	key regulatory posts sent clear signals of po-
	litical support and regulatory latitude for the
	crypto industry. This fostered increased in-
	vestor confidence and inflows into Bitcoin,
	driving its price past significant milestones,
	including the \$100,000 mark.
Vanilla GPT-40	N/A

Table 7: Responses from different approaches for BP Query 2

A qualitative analysis of BP Query 2, found in Table 7 shows the following:

• Strengths of KG-RAG:

917

918

919

920

921

922

923

924

925 926

927

928 929

930

 Clarity and Directness: KG-RAG presents a straightforward explanation by directly linking Bitcoin's price surge to Donald Trump's victory and its effect on investor confidence.

Weaknesses of KG-RAG:

- Oversimplification: The response attributes the surge solely to Trump's win, overlooking additional market factors and policy changes highlighted in the GPT-40 with RAG response. Tables 8 and 9 shows the responses from the KG-RAG and the baseline approaches to two sample queries from the GD dataset.

Query	How does a gene mutation in the GBA gene
	initiate Gaucher disease by impairing gluco-
	cerebrosidase production and leading to gluco-
	cerebroside accumulation?
KG-RAG	A mutation in the GBA gene leads to Gaucher
	disease by impairing the production of the glu-
	cocerebrosidase enzyme. This enzyme is es-
	sential for breaking down glucocerebroside.
	a fatty substance. When glucocerebrosidase
	is deficient or malfunctioning, glucocerebro-
	side accumulates in the body's cells, disrupt-
	ing normal cellular function and leading to the
	symptoms and complications associated with
	Gaucher disease.
GPT-40 with RAG	A gene mutation in the GBA gene initiates
	Gaucher disease by causing variants (or mu-
	tations) that either greatly reduce or eliminate
	the activity of the enzyme lysosomal gluco-
	cerebrosidase. The GBA gene provides in-
	structions for making this enzyme, which is re-
	sponsible for breaking down a fatty substance
	called glucocerebroside into glucose and ce-
	ramide. When there is insufficient activity
	of lysosomal glucocerebrosidase due to the
	gene mutation, glucocerebroside is not prop-
	erly broken down and begins to accumulate
	within cells. This buildup of glucocerebroside
	and related substances reaches toxic levels and
	damages tissues and organs, leading to the
	characteristic features of Gaucher disease.
Vanilla GPT-40	Gaucher disease is a lysosomal storage dis-
	order caused by mutations in the GBA gene,
	which encodes the enzyme glucocerebrosidase.
	Glucocerebrosidase is responsible for break-
	ing down a fatty substance called glucocere-
	broside into glucose and ceramide within the
	lysosomes.When there is a mutation in the
	GBA gene, the production of functional gluco-
	cerebrosidase is impaired. This can lead to a
	reduction in the enzyme's activity or its com-
	plete absence. As a result, glucocerebroside
	is not efficiently broken down and begins to
	accumulate within the lysosomes of various
	cells, particularly macrophages. The accumu-
	lation of glucocerebroside-laden macrophages,
	known as Gaucher cells, occurs primarily in
	the spleen, liver, and bone marrow. The buildup
	of these Gaucher cells leads to the clinical
	manifestations of Gaucher disease, which in-
	clude splenomegaly (enlarged spleen), hep-
	atomegaly (enlarged liver), bone pain and
	tractures, anemia, and thrombocytopenia (low
	platelet count).

Table 8: Responses from different approaches for GD Query 1

A qualitative analysis of GD Query 1, found in	934
Table 8 shows the following:	935
Strengths of KG-RAG:	936
- Conciseness: KG-RAG provides a clear	937

- **Conciseness:** KG-RAG provides a clear 937 and succinct explanation of how a mutation 938

953

954

955

957

Query	How does a genetic change in the GBA gene
	lead to anemia and fatigue, thereby impairing
	daily activities in Gaucher disease?
KG-RAG	In Gaucher disease, a genetic mutation in the
	GBA gene leads to the accumulation of gluco-
	cerebroside in certain cells. This buildup can
	cause damage to the bone marrow, resulting
	in anemia due to reduced red blood cell pro-
	duction. Anemia then leads to fatigue because
	the body's tissues receive less oxygen. The
	resultant fatigue significantly impairs daily ac-
	tivities.
GPT-40 with RAG	A genetic change in the GBA gene leads to
	Gaucher disease by causing insufficient activ-
	ity of the enzyme glucocerebrosidase. This
	enzyme deficiency results in the progressive
	accumulation of glucocerebroside in various
	tissues and organs, particularly in the bone
	marrow spleen liver and lungs. The accu-
	mulation of these fatty substances in the bone
	marrow disrupts normal blood cell production
	and leads to the rapid and premature destruc-
	tion of blood cells (Note: The response
	continues for several more sentences)
Vanilla GPT-40	In Gaucher disease genetic mutations in the
Vanna OI 1-40	GBA gene, which encodes the enzyme gluco-
	carebrosidosa, result in a deficiency or mal
	function of this anzuma. Glucocarabrosidasa
	is amount for breaking down glucocorebrosidase
	is clucial for breaking down glucocerebio-
	the enzyme is faulty or insufficient gluce
	the enzyme is faulty of insufficient, gluco-
	cerebroside accumulates in cens, particularly
	in macrophages, which are part of the im-
	nune system. I nese engorged cells, known as
	Gaucher cells, primarily accumulate in the
	bone marrow, liver, and spleen (Note:
	I he response continues for several more sen-
	tences).

in the GBA gene leads to impaired produc-

tion of glucocerebrosidase and subsequent

- Direct Causal Link: It effectively estab-

lishes the causal connection between en-

zyme deficiency and the buildup of gluco-

cerebroside, which is key to understanding

- Limited Clinical Context: KG-RAG

doesn't mention any of the downstream

clinical manifestations or the broader im-

pact of glucocerebroside accumulation on

tissues and organs. While this isn't neces-

sarily in the scope of the question, it can be

critical for a comprehensive understanding

glucocerebroside accumulation.

Gaucher disease.

• Weaknesses of KG-RAG:

of the disease.

Table 9: Responses from different approaches for GD Query 2

A qualitative analysis of GD Query 2, found in Table 9 shows the following:

Strengths of KG-RAG:

- Direct Causal Link: KG-RAG clearly outlines the sequence from the genetic mutation to the accumulation of glucocerebroside, resulting in bone marrow damage, anemia, and subsequent fatigue that impairs daily activities.

958

959

960

961

962

963

964

965

966

967

968

969

970

971

972

973

974

975

976

977

978

979

980

981

982

983

984

985

986

987

988

989

990

991

992

993

994

995

996

997

998

999

- Conciseness: The explanation is succinct and to the point, making the causal chain easy to follow.

Weaknesses of KG-RAG:

- Lack of Extended Detail: While the core causal pathway is presented, KG-RAG doesn't provide any additional detail, such as broader impacts or a detailed description of the pathological process, which is included by the other responses.

A.2 Ablation Study Results and Analysis

Table 10 and 11 show the responses from the KG-RAG and the baseline approaches for two sample queries from the BP dataset for the ablation study. A qualitative analysis of Ablation BP Query 1, found in Table 10 shows the following:

• Strengths of KG-RAG:

- Focused Explanation: KG-RAG delivers a clear and coherent narrative by directly linking BlackRock's successful launch of its iShares Bitcoin Trust ETF to a massive inflow of institutional investment.
- Causal Clarity: It effectively explains how this inflow, along with Fidelity's subsequent decision to offer Bitcoin trading, contributed to increased demand and the ensuing price surge.

Weaknesses of KG-RAG:

- Narrow Scope: While the response is focused, it overlooks additional market dynamics and external factors that could further contextualize the price surge.

A qualitative analysis of BP Query 2, found in Table 11 shows the following:

Strengths of KG-RAG:

- Causal Clarity: KG-RAG clearly links Donald Trump's 2024 election win to a 1001

1006	ent chain of events from increased investor
1007	optimism to greater market accessibility
1008	and growth in Bitcoin's market capitaliza-
1009	tion.
1010	• Weaknesses of KG-RAG:
1011	- Limited Context: The response focuses
1012	primarily on investor sentiment and acces-
1013	sibility, without addressing other potential
1014	influencing factors such as regulatory or po-
1015	litical signals detailed by other approaches.
1016	Tables 12 and 13 show the responses from the
1017	KG-RAG and the baseline approaches to two sam-
1018	ple queries from the BP dataset for the ablation
1019	study.
1020	A qualitative analysis of Ablation GD Query 1,
1021	found in Table 12 shows the following:
1022	• Strengths of KG-RAG:
1023	- Clear Causal Pathway: The response out-
1024	lines the core mechanism-how a muta-
1025	tion in the GBA gene leads to enzyme defi-
1026	ciency, resulting in glucocerebroside accu-
1027	mulation and cellular toxicity.
1028	- Link to Clinical Manifestations: It con-
1029	nects the biochemical cascade to the clini-
1030	cal symptoms of Gaucher's disease.
1031	Weaknesses of KG-RAG:
1032	- Terminology Errors: There are numer-
1033	ous misspellings (e.g., "Gba," "glucoc-
1034	cerobrosidases") and incorrect disease
1035	names (e.g., "Gauger's disease" instead of
1036	Gaucher's disease), which undermine re-
1037	sponse clarity.
1038	– Overextension: The inclusion of
1039	Parkinson's-like symptoms, while relevant
1040	in some contexts, strays from the primary
1041	focus of the query.
1042	A qualitative analysis of Ablation GD Query 2,
1043	found in Table 13 shows the following:
1044	• Strengths of KG-RAG:
1045	- Clarity and Brovity: Provides a succinct

shift in investor sentiment and the subse-

quent surge in Bitcoin's value, which at a

- Narrative Coherence: It outlines a coher-

high level, is correct.

1002

1003

1004

1005

1046

 Clarity and Brevity: Provides a succinct explanation linking the genetic defect in the glucocerebrosidase gene to glucosyce-
ramide accumulation, and subsequently to
anemia and fatigue.1047
1048
1049Direct Causal Chain: Clearly outlines1050

1051

1052

1053

1054

1055

 Direct Causal Chain: Clearly outlines how the buildup of glucosyceramide due to impaired enzyme function leads to bone marrow infiltration and resultant symptoms.

• Weaknesses of KG-RAG:

Lack of Detail: The explanation is overly brief and does not delve into additional clinical details that are needed for a complete understanding.
 1056
 1057
 1058
 1059

Query	What influence did major institutional players	ſ	Query
	like BlackRock and Fidelity have on Bitcoin's	-	KC BAC
KG-BAG	BlackBock's successful launch of its iShares		KG-KAG
KO-KAO	Bitcoin Trust ETF led to massive inflow of in-		
	stitutional investment into the market This		
	influx of capital attracted more institutions		
	to the space, further increasing demand and		
	contributing to the Bitcoin price surge post-		
	election. Fidelity's subsequent decision to of-		
	fer Bitcoin trading to its institutional clients		
	may have also been influenced by the growing		
	institutional interest sparked by BlackRock.		
	The combined effect of these institutional play-		
	ers' actions likely had a significant impact on		
	the market's direction. As a result, the price of		
	Bitcoin surged, reflecting the growing accep-		
	major institutional investors. This price surge		
	was a direct consequence of the increased de-		
	mand driven by institutional investors who		
	were drawn to the asset class by the success of		
	the iShares trust. The influence of these major		
	institutional actors on the Bitcoin market was		
	a key factor in the price increase, highlight-	ŀ	LLaMA with R
	ing the growing importance of institutional in-		
	vestors in shaping the cryptocurrency's value.		
	The increased institutional involvement also		
	paved the way for further growth and main-		
	stream acceptance of Bitcoin.		
LLaMA with RAG	The text does not mention Fidelity and only		
	mentions BlackRock in a context that indicates		
	they entered the fold of investors in Bitcoin af-		
	in January. The price surge is attributed to		
	other factors such as Trump's nomination of a		
	friendly regulator and the introduction of Bit-		
	coin as a store-of-value option by the Federal		
	Reserve. However, the text does mention that		
	the crypto industry as a whole poured over 119		
	million dollars into influencing elections, pri-		
	marily through a political-action committee.		
	Coinbase is mentioned as having spent over		
	50 million dollars on political contributions,		
	but it denies that its CEO openly supports the		
	president-elect. The CEO of Coinbase report-		
	edly met last month with the president elect,		
	dodot blockchain James Slussar says that the		
	price surge may be due to investors feeling		
	more confident that US regulation would favor		
	them due to Trump's election. He urges in-		
	vestors, however, to approach the market with		
	caution. The text also mentions that Trump has		
	launched his own non-fungible tokens (NFTs)		
	and has called for the US to be the crypto cap-		
	ital of planet earth. The Federal Reserve chair-		
	man, Jerome Powell, has compared Bitcoin to		
	gold, which is seen as a significant endorse-		
	ment of the cryptocurrency as a valuable asset.	Ļ	
	However the text also notes that Bitcoin is still		Vanilla LLaMA
Vanilla I LaMA	nigniy volatile		
vanilla LLAMA	IN/A	L	

Table 10: Responses from different approaches for BP Query 1

Query	Why did Bitcoin's price surge after the 2024
KG-RAG	Donald Trump's 2024 election win led to in-
	creased optimism in the US economy. As in-
	vestors became more confident in the country's
	future, they began to take on more risk. This
	shift in sentiment led to a surge in Bitcoin's
	value as investors sought alternative assets to
	diversify their portfolios. The increased de-
	mand for Bitcoin drove up its price, making
	an attractive option for investors looking to
	market capitalization grew significantly solid-
	if ying its position as a leading cryptocurrency
	The surge in demand also led to an increase
	in the number of exchanges and trading plat-
	forms offering Bitcoin, making the asset more
	accessible to a wider audience. This increased
	accessibility further fueled the growth of Bit-
	coin's user base, creating a self-reinforcing cy-
	cle of growth. The combination of increased
	demand, growing market capitalisation and ac-
	cessibility ultimately led to Bitcoin's surge in
	value.
LLaMA with RAG	Bitcoin's price surged after the election due
	to the pro-Bitcoin promise of Donald Trump
	and his picks for regulatory agencies. The
	seen as favorable to crupto by industry experts
	contributed to the surge Additionally the fi-
	nancial industry's growing acceptance of Bit-
	coin and the promise of deregulation under
	the Trump administration also contributed to
	its price increase. Some experts also believe
	that speculation about the US building up a
	Bitcoin reserve added to the tailwind for the
	cryptocurrency. Overall, the surge was driven
	by a combination of factors, including politi-
	cal support, regulatory latitude, and growing
	legitimacy in the eyes of the financial commu-
	nity. The crypto corporations also played a
	fole by pouring \$119m into the federal elec-
	tions, which helped to bring more politicians
	ton. The increased inflows of money to the
	crypto markets particularly to Bitcoin also
	contributed, as it is the most established and
	widely held cryptocurrency, making it the most
	natural place for new investors to put their
	money. Furthermore, the fact that the price of
	a coin is influenced by its market cap means
	that it takes more money to move the price,
	which also contributed the surge of Bitcoin's.
	It is worth noting that the cryptocurrency mar-
	ket is known for its volatility, and the price can
	of Bitcoin makes it more stable and loss more
	to sudden price changes. Therefore the price
	increase is expected be more stable in the long
	term
Vanilla LLaMA	N/A. I don't have information on the 2024 US
	presidential elections, or the subsequent price
	of Bitcoin.

Table 11:	Responses	from di	ifferent	approaches	for BP
Query 2					

	1
Query	How does a gene mutation in the GBA gene
	initiate Gaucher disease by impairing gluco-
	cerebroside accumulation?
KC DAC	A mutation in one of the Cha gana copies
NG-NAG	leads to a deficiency of glucoccerobrosidases
	which are enzymes that break down a type of
	fat called glucocerbroside. This deficiency re-
	sults in the accumulation of glucoscerbroside
	within cells, which in turn causes Gaucher's
	disease. The accumulation of this fat is toxic
	to cells, causing them to malfunction and
	die, leading to the symptoms of Gauger's dis-
	ease, such as anemia, bone pain, and lung dis-
	ease. In the brain, the accumulation can cause
	Parkinson's disease like symptoms, including
	tremors, rigidity and bradykinesia.
LLaMA with RAG	The GBA mutation impairs the production of
	the glucoceresidase enzyme, which is neces-
	sary for breaking down glucocerbroside. As a
	ing Gaucher's disease. The enzyme breaks glu
	cocerasebroside down into a glucose and a ce-
	ramide. Without the enzyme the glucerbroide
	builds up. The enzyme does not break it down
	into glucose and ceramide, so it stays in the
	cell and builds up, causing damage to the cell.
	Gaucher's disease is a genetic disorder caused
	by mutations in GBA genes. These mutations
	lead to a deficiency in the enzyme glucoccer-
	obrosidasa. Without this enzyme gluccero-
	broside builds and causes cell damage. The
	the body to break down glucerobrode, which
	causes the cells to become damaged. The ac-
	cumulation of glucerobreoside is toxic to cells.
	leading to the symptoms of Gauchers disease.
Vanilla LLaMA	A gene mutation leads to impaired glucocore-
	brosdase production, which in turn leads to the
	accumulation of glucocerbroside. The ques-
	tion was about the GABA gene, but the re-
	sponse is about the gene GBA, which is as-
	(glucocarabosidase) gane mutation impairs the
	production of the glucoccerobrosidasa enzyme
	leading to an accumulation of the substance
	glucococerebside. A mutation in GBA leads
	to a deficiency of the enzyme glucocorere-
	broidase, resulting in an accumulation gluco-
	cercerobrosides. Mutations in the gene that
	codes for glucoccorrebrodase, impair the pro-
	duction and activity of this enzyme, resulting
	glucocorticoid accumulation. Mutation of the
	GCB gene impairs glucoccorebrosidas produc-
	tion, leading accumulation of gluccorebroides.
	Impaired glucocorrebrodasa production due to
	accumulations Gene mutations in the enzyme
	that breaks down glucocorbosides lead to an
	impaired enzyme production and glucocorra-
	boside accumulation

Table 12: Responses from different approaches for Ablation GD Query 1

Query	How does a genetic change in the GBA gene
	lead to anemia and fatigue, thereby impairing
	daily activities in Gaucher disease?
KG-RAG	A genetic defect in the glucocerebrosidase
	gene leads to Gaucher's disease, causing the
	accumulation glucosyceramide in the cells.
	This accumulation leads to various symptoms,
	including anemia due to bone marrow infiltra-
	tion, which subsequently causes fatigue.
LLaMA with RAG	The genetic change leads to insufficient activ-
	ity of the enzyme glucoc erebroside, which
	results in the accumulation of glucos ylce-
	ramide, causing anemia and fatigue due to
	impaired blood cell production, leading to im-
	paired daily activities. The rapid and prema-
	ture destruction of blood certs leads to an in-
	vers poin associated with joints and bones on
	curs frequently presenting in hips and knees
	further impairing daily activities. Summary:
	The genetic change in the GBA gene leads to
	insufficient activity and accumulation gluc ce-
	ramide, resulting in anemia. This results in
	fatigue, impaired daily activity, an increased
	infection risk, bleeding, and severe pain. It
	also causes bone and joint problems, includ-
	ing osteoporosi s, osteopeni a, bone fractures,
	arthritis, joint pain, damage, and weakened
	bones. Neurological symptoms occur in some
	types of Gauch er's, such as impaired olfactory
	and cognition, serious convulsion, hypertonia.
	intellectual disability, apnea, muscle twitche
	s, convulsion. dementia, ocular muscle apraxi
	a. Parkinson's disease is recognized as be-
	their beterozy gous carrier relatives. Vellowish
	brown skin nigmentation also occurs
Vanilla I I aMA	The GBA (glucocerebrosidase) gene encodes
Valina EEalvilly	an enzyme that breaks down glucocerebroside.
	a type of lipid. In Gaucher's disease, a genetic
	mutation in the gene leads to a deficiency of
	this enzyme. As a result, glucocerbroside ac-
	cumulates in the body's cells, particularly in
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue,
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocerbor- bids in the splene activity before
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inclusive to fits out downgood and
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fotigue is caused due to the
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea- son for the anemia in Gauchers disease. The
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea- son for the anemia in Gauchers disease. The combination of anemia & fatigue makes it dif-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea- son for the anemia in Gauchers disease. The combination of anemia & fatigue makes it dif- ficult for the patients to perform daily activities
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea- son for the anemia in Gauchers disease. The combination of anemia & fatigue makes it dif- ficult for the patients to perform daily activities and they may experience more severe symp-
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea- son for the anemia in Gauchers disease. The combination of anemia & fatigue makes it dif- ficult for the patients to perform daily activities and they may experience more severe symp- toms such as dizziness and shortness of breath. The cumptom on the succumulation of succumulation
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea- son for the anemia in Gauchers disease. The combination of anemia & fatigue makes it dif- ficult for the patients to perform daily activities and they may experience more severe symp- toms such as dizziness and shortness of breath. The symptoms are more pronounced in cases where the disease is left untreacted on where the
	cumulates in the body's cells, particularly in the bone marrow, liver, and spleen. This ac- cumulation leads to the breakdown of these cells, which in turn causes the symptoms of Gaucher's disease, including anemia, fatigue, and impaired daily activities. The anemia is caused by the accumulation of glucocercbro- side in the spleen, which leads to its dysfunc- tion and the inability to filter out damaged red blood cells. The fatigue is caused due to the accumulation in the liver and spleens, which impairs the body's ability to produce and re- cycle red blood cell components. The genetic change leads to an accumulation of harmful substances in the cells, leading to the destruc- tion of the cells and the release of inflamma- tory substances into the bloodstream. This leads to further complications and the symp- toms mentioned above. The accumulation of these substances also affects the production of red blood corpuscles, which is the main rea- son for the anemia in Gauchers disease. The combination of anemia & fatigue makes it dif- ficult for the patients to perform daily activities and they may experience more severe symp- toms such as dizziness and shortness of breath. The symptoms are more pronounced in cases where the disease is left untreated or where the genetic mutation is more eavera

Table 13: Responses from different approaches for Ablation GD Query 2