Expertly Informed, Generatively Summarized: A Hybrid RAG Approach to Informed Consent Summarization with Auxiliary Expert Knowledge

Autumn Toney-Wails
SciTech Strategies, Inc.
Georgetown University
autumn@mapofscience.com

Ryan Wails
Georgetown University
rsw66@georgetown.edu

Caleb Smith
University of Michigan
Syntheos, Corp.
calebs@med.umich.edu

Abstract

The utility of retrieval augmented generation (RAG) systems is actively being explored across a wide range of domains. Reliable generative output is increasingly useful in fields where routine tasks can be streamlined and potentially improved by integrating domain-specific data in addition to individual expert knowledge, such as medical care. To that end, we present a hybrid RAG and GraphRAG user interface system to summarize the key information (KI) section in IRB informed consent documents. KI summaries are a unique task, as generative summarization helps the end user (clinical trial expert) but can pose a risk to the affected user (potential study participants) if inaccurately constructed. Thus, the KI summarization task requires reliable, structured output with input from an expert knowledge source outside of the informed consent document. Reviewed by IRB domain experts and clinical trial PIs, our summarization application produces accurate (70% to 100% varied by accuracy type) and useful summaries (63% of PIs stating summaries were as good as or better than their accepted summaries).

1 Introduction

Applied in the medical field, retrieval augmented generation (RAG) systems have shown promise in streamlining routine tasks, providing structure for standard medical procedures, and ensuring current information is integrated into decision making (Hammane et al., 2024; Unlu et al., 2024; Zhang, 2024; Jeong et al., 2024). However, there is concern surrounding the reliability and trustworthiness of generative output in high-risk, real-world implementations in which incorrect information can lead to severe personal harm. Researchers and practitioners have focused on studying and identifying reliable generative artificial intelligence (GenAI) use cases to optimize a routine procedure for the end user (e.g., medical professional) and minimize potential harm for the affected individual (e.g., patient receiving care).

As technical advances in GenAI, natural language generation, and information retrieval continue, applications of these systems and models become viable tools for real-world implementation. Building on the RAG pipeline (a user query coupled with an LLM and knowledge base), the development of GraphRAG created a retriever system that uses a knowledge graph to provide context and entity relationships derived from the selected knowledge base (Edge et al., 2024). Thus, GraphRAG systems can be used to extract key information with known feature relationships, providing more transparent outputs linked directly to data sources in the knowledge base.

In this study, we design a hybrid RAG and GraphRAG system to optimize summaries of key information in IRB informed consent documents. Informed consent documents (ICDs) are provided to individuals who are considering participating in a medical study for new treatment that may affect their health and path to recovery. Thus, ICDs are required and reviewed to ensure that participants are well-informed about their rights as well as the nature, risks, and benefits of the study. The key information portion of the ICD serves as a concise summary highlighting the most critical aspects of the study. Specifically, it helps potential participants understand the key information to make a well-informed decision about agreeing to participate. It is important that these summaries balance providing comprehensive details with clarity and minimizing technical jargon.

We designed and evaluated a pilot key information summary application on four key measurements: (1) factual accuracy, (2) standard of care vs. research differentiation, (3) information weighting, and (4) style and structure. Evaluations were performed by three IRB subject matter experts to iteratively improve our model design and output

Proceedings of the 4th International Workshop on Knowledge-Augmented Methods for Natural Language Processing (KnowledgeNLP'25), pages 246–258

over 11 cycles. Additionally, eight senior principal investigators assessed the machine-generated summaries for further improvement input. We found that the key challenge in our pilot summarizer was differentiating between standard of care risks versus research risks (i.e., what risks were associated explicitly with participating in the study outside of standard treatment?).

To address the risk differentiation challenge we implement **T**emporary **A**uxiliary **G**raph**RAG** (**TAGRAG**) for expert question and answering. Using search terms associated with the proposed study's disease focus and standard of care medical treatment, we query open-source articles via PubMed Central's API and select relevant research publications as input to a GraphRAG system. The GraphRAG is instantiated for each summary instance and removed after each summary is generated. Our TAGRAG component enables expert information to support current and relevant standard of care risks associated with the corresponding study's disease focus without maintaining a large knowledge graph of medical research.

2 GenAI in Medical Research

Prior research has investigated the utility of GenAI in the medical field for a wide range of tasks using text and image data. Sai et al. (2024) provide a survey on how models like ChatGPT and DALL-E can be implemented in medical tasks such personalized patient treatment, healthcare operations and research, and clinical trial optimization. The authors highlight four directions for future research: (1) customized/personalized suggestions and a platform for information exchange, (2) enhanced patient and worker interactions, (3) streamlining administrative operations, and (4) enhancing decision making and bridging the knowledge gap. Researchers have studied the utility of chatbots as tools to reduce time on routine tasks and assist non-experts with understanding technical medical language (Barak-Corren et al., 2024; Shyr et al., 2024; Zaretsky et al., 2024). Specifically, using retrieval augmented generation (RAG) systems in the medical field has been recently explored (Alkhalaf et al., 2024; Hammane et al., 2024; Jeong et al., 2024; Unlu et al., 2024; Zhang, 2024). We summarize the work that follows similar processes to our KI summary application.

Alkhalaf et al. (2024) use zero-shot prompting with Llama 2 (13B model) to generate structured summaries for clients describing their nutritional status and extracting key information about malnutrition risk factors. Implementing RAG improved their accuracy results from 93% to 99%; however, the authors note that RAG did not improve extracting risk factors (accuracy maintained at 90%). Unlu et al. (2024) implement a RAG system, RAG-Enabled Clinical Trial Infrastructure for Inclusion Exclusion Review (RECTIFIER), to evaluate if GenAI could "improve the accuracy, efficiency, and reliability of screening for a trial involving patients with symptomatic heart failure." The authors find that RECTIFIER (achieving 97.9% accuracy) outperforms medical professionals (achieving 91.7% accuracy) at determining symptomatic heart failure.

Jeong et al. (2024) present Self-BioRAG, a RAG system trained on 84k filtered biomedical instruction sets that provides customized explanations; the authors highlight the benefit of domain-specific components (e.g., a retriever, related document corpus, and instruction sets) for high performance. Self-BioRAG achieves an average of 7.2% improvement over the state-of-the-art open-foundation models and outperforms traditional RAG by 8% Rouge-1 score. Hammane et al. (2024) design SelfRewardRAG, a RAG system that references PubMed for evidence-based responses to user queries and includes a self evaluation layer to thoroughly evaluate and update its output. The authors evaluate the model on three benchmarks: (1) PubMedQA (achieving 81.1% accuracy), (2) MedQA-USMLE (achieving 50% accuracy), and (3) BioASQ (achieving 95% accuracy).

Our work covers each of the four areas highlighted by Sai et al. (2024), as we designed, developed, and deployed a generative summarization tool that incorporates expert knowledge (enhancing decision making and bridging the knowledge gap), decreases administrative work for clinical experts (streamlining administrative operations), supports effective communication between clinical experts and non-expert study participants (enhanced patient and worker interactions), and provides a user interface for summarizing a document based on the domain-specific implementation (customized/personalized suggestions and a platform for information exchange).

3 Informed Consent Key Information Summary Structure

An Institutional Review Board (IRB) works to regulate human-subject research, ensuring ethical procedures and minimal risk to participants. Research institutions are required to receive IRB approval prior to engaging with any potential participants or beginning any human-subject experiments. Producing an informed consent document is a requirement of the IRB approval process and these documents follow a regulated structure and format to maintain consistency across research studies—they are designed to consistently protect participants.

A critical component (and requirement) of an informed consent document is the key information section, which explains the details of the study in clear language (with minimal technical jargon) and identifies the potential risks involved as a participant that are distinct from the risks involved from standard of care treatment. Specifically, key information summaries are designed to support potential participants in deciding whether or not they would like to be a part of the research study. According to U.S. federal regulations, key information summaries should include the following five elements: (1) a statement acknowledging that the project is research and participation is completely voluntary; (2) a summary of the proposed research (purpose, duration, and list of procedures); (3) potential risks (distinct from standard of care treatment); (4) expected benefits; and (5) alternative treatment options or procedures (if applicable).

With their structured output and routine requirement for researchers, writing key information summaries is a suitable task for experimentation using generative AI. Additionally, key information summaries require knowledge of the informed consent document and context-relevant medical expertise for the proposed study.

4 Pilot RAG KI Summary Application

Given the application setting, our experimental design involved iterating through results with evaluations from three subject matter experts (SMEs). Our pilot RAG system (shown in Figure 1) takes two inputs: (1) one informed consent document used as the knowledge base in the RAG system, and (2) a prompt dictionary for structured question-and-answering for reliable KI summary output.

In Sections 4.1 and 4.2 we describe the details of the pilot KI summary application after 11 iterations[1]. Section 4.4 provides the details on the SME evaluations for each evaluation, and Section 4.5 provides the details on the PI evaluations. Table 1 contains brief summaries of the main changes made in each iteration.

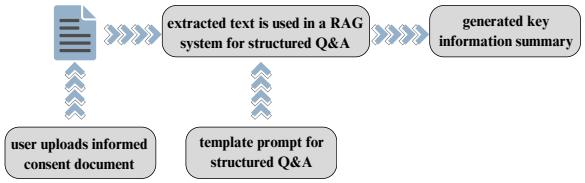

Figure 1: Pilot RAG system diagram for key information summaries.

4.1 Informed Consent Documents as Knowledge bases

We used 18 human-authored informed consent documents approved by an institution's medical school IRB. This set of ICDs covered studies on clinical trials for drugs and medical devices, data registries, cancer and other health studies, and pediatric populations. Table 2 displays the descriptive statistics on the ICDs page length and token count.

Each user-uploaded ICD is used as the knowledge base in the RAG system for summarization, thus the document text extraction, chunking, and vectorization is computed in real time. Our system processes .pdf and .docx files; we used `pypdf` to extract text from .pdf files and `docx2txt` to extract text from .docx files. We built the RAG framework with `llama-index`[2], and selected the `HierarchicalNodeParser` to chunk the parsed document text (with `chunk_sizes` $\in [128, 256, 512]$). Node parsing enables efficient and scalable text processing with the hierarchical structure maintaining the relationship between sections. The chunked documents are stored using the `VectorStorageIndex` using the default parameters for the `StorageContext`.

4.2 Agentic Summarization

Using OpenAI's GPT-4 (`gpt-4-0125-preview`), we instantiate a chatbot with the following persona assigned in the system prompt:

```
## YOUR ROLE
You are a bioethicist specializing in
patient advocacy and human subjects
research. Your focus is on interpreting
and explaining Informed Consent
```

[1] https://github.com/autumntoney/TAGRAG
[2] https://pypi.org/project/llama-index/

Version	Design Details/Changes
v1	Naive, general instructions with reference to template content.
v2	Single prompt with each paragraph of the Key Information template was included. Similar to how a human would generate a summary (e.g., "Fill in the blanks on the provided template.").
v3	**<major change>** Single prompt is still used, but instructions are more detailed and we begin using meta-language to exert more fine-grained control on the model's choices.
v4	**<major change>** Each paragraph of the KI section is given it's own custom prompt. We load the complete ICD as context for each paragraph. *We no longer use a single prompt.*
v5	**<major change>** Use simple sub-questions to distill information from the ICD in a decision tree format. The answers from the sub-questions are used to inform the KI sections. *Example: "Will the study enroll children?" If yes, use paragraph option 1; if no, use paragraph option 2*
v6	Same approach as v5 but question decomposition is more elaborate, with more sub-questions used to distill knowledge.
v7	Same approach as v6 but we create a draft KI, then the model evaluates this draft KI before generating its final response. V7 has the following system components: (1) decompose the content needed to generate a paragraph into sub-questions, (2) answer the sub-questions, (3) use those answers to complete the paragraph, (4) join all paragraphs together into a "draft" KI, and (5) edit the draft into a final version.
v8	**<major change>** Test the newly introduced "Assistant" functionality where all questions and answers are fed into a continuing conversation that the chatbot can draw on (i.e., when the chatbot is given a task such as answering a question or completing a paragraph, it is able to draw on all previous questions and tasks it has completed for the provided ICD).
v9	Continue with v8 but continue prompt engineering for improved question/instruction phrasing
v10	**<major change>** Stop using the Assistant thread approach due to: (1) cost (it is significantly more expensive), (2) availability (it is available on OpenAI but not Azure), and (3) control (we do not have as much control over structured and reliable responses). New prompt engineering designs with tehcnical experts to provide similar performance using the original `chatCompletion` framework.
v11	Finalized prompt instruction design to optimally manage section-level changes and provide generalizability for other implementations.

Table 1: RAG system pipeline design details and changes throughout each of the 11 iterations.

	Mean	Min	Max	Std. Dev
Page Count	17	7	41	10
Token Count	8,081	3,939	16,718	4,547

Table 2: Descriptive statistics on the 18 ICDs page length and token count.

```
documents to potential human subjects
research participants.

## RULES
- Ensure all responses are directly
grounded in the context you are provided
- Responses should be clear and
authoritative, delivered in a more
formal tone.
- Avoid conjunctive adverbs, discourse
markers, and both introductory and
conclusive statements.
- Do not include disclaimers or refer to
 yourself as an AI.
- Provide information in a way that is
clear and understandable to potential
research participants.
- Prioritize accuracy and relevance in
your responses. Do not include
unnecessary information.
```

The assigned role is written to encourage a focus on the affected user (potential participant) and the rules are written to provide explicit instruction to the chatbot that will produce reliable and consistently formatted results for the summarization task.

We formulate a structured open-ended prompt dictionary, containing nine sections, for an automated question and answering pipeline that generates the components for the key information summary. These prompts were designed by technical experts and SMEs to ensure relevant information

is extracted and to further instruct the chatbot on how to formulate responses for the corresponding section. For example, the section 6 prompt reads:

```
  "section6": [
        (
            "Imagine that I am the study
 participant and you are explaining the
most important risks that are introduced
 or enhanced because of participation in
 this research study to me.\n"
            "Rather than trying to
explain every risk, focus on the risks
that will cause me pain or emotional
distrees. What are the most important
risks that you would explain to me?\n"
            "Do not include risks
associated with standard of care
treatments. Only include risks that
could reasonably be introduced or
enhanced due to participation in this
research study.\n"
            "Use plain language to
describe the risks with few words. Your
response should be no more than 3
sentences in length."
            ),
        (
            "You have been provided with
 template text after the triple dashes
below. Adhere to this text in your
response. When you encounter a phrase in
 this text that is enclosed by double
brackets ([[example instructions]]),
replace it with relevant details based
on what you have learned about this
research study. \n\n"
            "---\n\n"
            "There can be risks
associated with joining any research
study. The type of risk may impact
whether you decide to join the study.
For this study, some of these risks may
include [[Briefly describe the risks
while maintaining a formal tone]]. More
detailed information will be provided
later in this document."
        )
    ],
```

There are portions of the summaries that maintain standardized phrasing (e.g., "There can be risks associated with joining any research study. The type of risk may impact whether you decide to join the study. For this study, some of these risks may include [*identified risks*]"); all templates are included in the Appendix.

4.3 Generated Text Post-Processing

The output text of each relevant section is cleaned and joined together to form the final key information summary text returned to the end-user. We implement basic text cleaning functions and remove unwanted characters (e.g., extra quotation marks, brackets, or special characters), correct text spacing (e.g., remove extra spaces or line breaks), and standardize the formatting to ensure consistency across each section. With the cleaned text, we assemble the responses to form one summary. We remove any sections that are empty, as not all sections are relevant to each ICD and proposed study, and we set predefined text for two sections if the chatbot did not generate a response. The predefined text provides general information about research using standard language. For example, section 2 reads:

```
predefined_entries = {
        "section2": "A research study is
 different from the regular medical care
 you receive from your doctor. Research
studies hope to make discoveries and
learn new information about diseases and
 how to treat them. You should consider
the reasons why you might want to join a
 research study or why it is not the
best decision for you at this time."}
```

4.4 SME Evaluation Criteria

With the support of three IRB subject matter experts, we evaluated our key information summary system continuously over the course of 11 iterations, with two distinct cohorts. In the first cohort, IRB SME evaluators compared the same AI-generated summaries against existing human-authored key information summary sections. In the second cohort, IRB SME evaluators examined a new series of informed consents to assess the AI tool's adaptability to content it had not previously encountered. SMEs were asked to evaluate the model's response on four key components: (1) factual accuracy, (2) standard of care vs. research differentiation, (3) information weighting, and (4) style and structure.

The three IRB SMEs provided scores between 0 and 1 and we present the average scores in Table 3. The result of these evaluations prompted our final model design (presented in Section 6), as a key challenge for RAG system was presenting information surrounding the risks associated with the proposed study that were *distinct* from risks associated with standard of care. Our prompt engineering experiments resulted in the highest performing feature, style and structure, achieving 0.72 accuracy (cf. the accuracy score is 0.4 in version 1). The RAG system demonstrated improved factual accuracy with each version, improving from 0.3 to 0.7 with the final model. The two lowest performing features are information weighting and standard of care versus research differentiation, with both achieving 0.63.

	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11
Factual Accuracy	0.3	0.35	0.43	0.42	0.5	0.55	0.53	0.62	0.6	0.65	0.7
Risk Differentiation	0.33	0.38	0.43	0.4	0.45	0.5	0.47	0.55	0.53	0.58	0.63
Information Weighting	0.22	0.28	0.37	0.35	0.43	0.48	0.47	0.55	0.53	0.58	0.63
Style and Structure	0.4	0.45	0.5	0.47	0.52	0.57	0.55	0.63	0.62	0.67	0.72

Table 3: Average SME ratings on the four evaluation metrics for RAG output. (Risk differentiation references the standard of care versus research risk differentiation.)

4.5 Clinical Trial PI Evaluations

After iterating through system versions with IRB SMEs, eight PIs evaluated their accepted KI summaries (from previous studies with accepted ICDs) against the draft generative summaries. Answering six survey questions the PIs provided their assessments of the generative summaries; the percentages of their agreement is presented in Figure 2.

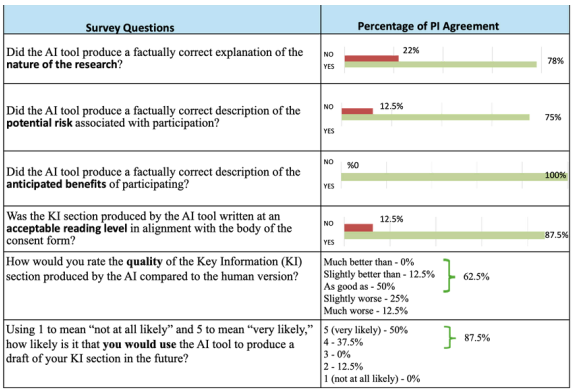

Figure 2: Percentage of PI agreement for six evaluation questions.

We found that 78% of PIs assessed the generated summary drafts to contain factually accurate explanations of the nature of the research proposed in the study. Notably, 100% of PIs found that the generated summaries described the anticipated benefits accurately, with 75% stating that the generated summaries produced accurate descriptions of potential risks. Finally, 87.5% of the PIs gave a score of 3 or higher (on a five point scale) that they were likely to use our tool to draft KI summaries.

5 Hybrid RAG Application Design

Improving on our pilot key information summary application described in Section 4, we introduce a GraphRAG component (TAGRAG) to our RAG summary pipeline to address the challenge of risk differentiation between the current standard of care and the proposed study. We maintain the same inputs from our pilot application (a user-uploaded informed consent document and a prompt message dictionary), and implement the GraphRAG using information extracted from the RAG component and an expert document database. The full end-to-end pipeline is shown in Figure 3.

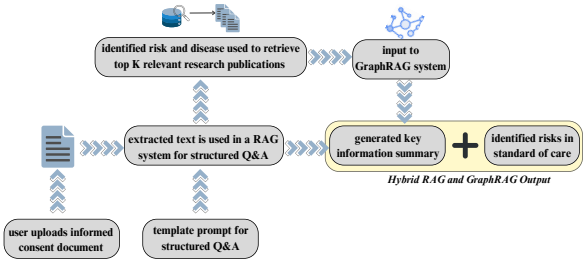

Figure 3: End-to-end hybrid RAG and graph RAG pipeline diagram for key information summaries.

5.1 Expert Document Database

We use the PubMed Central (PMC) dataset as our expert document database to select current medical research on the standard of care for a given disease (National Center for Biotechnology Information, 2000). PMC is the largest subset of PubMed[3], with over 10 million open-source, full-text biomedical and life science research publications. We query PMC via the API tool available through the `pymed`[4] Python package.

Following the PMC query format, we search the database for the top 10 most relevant articles. Our query is formatted to always contain "Standard of Care" and "Practice Guidelines as Topic" in addition to disease-relevant terms identified in the RAG component of our pipeline. For example, if the ICD was for a proposed study for diabetic treatment a PMC query would be composed as:

```
("diabete"[All Fields] OR "diabetes
mellitus"[MeSH Terms] OR ("diabetes"[All
 Fields] AND "mellitus"[All Fields]) OR
```

[3]PubMed contains over 37 million research publications that are both open-source and pay-walled.

[4]https://pypi.org/project/pymed/

```
"diabetes mellitus"[All Fields] OR "
diabetes"[All Fields] OR "diabetes
insipidus"[MeSH Terms] OR ("diabetes"[
All Fields] AND "insipidus"[All Fields])
 OR "diabetes insipidus"[All Fields] OR
"diabetic"[All Fields] OR "diabetics"[
All Fields] OR "diabets"[All Fields])
AND "standard of care"[MeSH Terms] AND "
practice guidelines as topic"[MeSH Terms
]
```

This query then returns the research publications' sections (e.g., title, abstract, methods, results, and conclusion) and corresponding metadata (e.g., keywords, doi, pubmed id, authors, and journal).

5.2 Temporary Auxiliary GraphRAG

The expert TAGRAG component is instantiated in real time with each summary, similar to the RAG component. In addition to the document extraction, text chunking, and vectorization required in a RAG system, a GraphRAG requires a schema for entity extraction to build its knowledge graph. We define the entities and relationships for optimal retrieval on research publications for our risk differentiation task; Table 4 lists the terms selected for our schema builder. We select more general entities for generalizability to other areas of research.

Using `neo4j` and `neo4j_graphrag`, we design our knowledge graph pipeline with the `SimpleKGPipeline` and OpenAI models (`text-embedding-3-large` for embeddings and `gpt-4o-mini` for the chatbot). The subset of expert documents for the given instance are input to the knowledge graph pipeline; no additional data is used. The GraphRAG system is then finalized by setting the `retriever` with the expert document vectors and the `llm` to `gpt-4o-mini`.

To generate the output for standard of care and research risk differentiation, we set the following prompts for TAGRAG Q&A:

```
    ``You are a medical researcher
tasked with extracting information from
papers surrounding the potential risks
during medical treatment and care.
Please answer the following two
questions:
    1) What is the standard of care for
{disease} based on the PubMed articles?
    2) What are the main differences
between the risks associated with
standard of care and the following risks
 associated with a proposed study for
new medical treatment:
    {RAG_extracted_risks}''
```

The TAGRAG prompt takes `disease` and `RAG_extracted_risks` as arguments that are extracted from the RAG component (described in Section 4). We consider the output from the TAGRAG to be expert input to the ICD key information summary, as the knowledge graph is derived from peer-reviewed, relevant research publications from global research institutions and researchers.

6 KI Summary Application UI

Our finalized informed consent document key information summary application is deployed on Vercel[5] with a private server connection for the Python backend. The webpage displays a file upload box via drag-and-drop or directory search. Once uploaded, users can generate the KI summary which references the hybrid RAG and TAGRAG described in Sections 4 and 5. The section summaries are computed concurrently for efficient processing and then used to compose the final summary. Each section summary is displayed to the user in order, with the final summary including the standard of care and proposed study risk differentiation highlighted. Figure 4 displays the various components of the deployed application (file upload, subset of the drafted KI summary, and Section 6 summary) and Figure 5 displays the corresponding risk differentiation output resulting from TAGRAG.

The KI summary application includes how-to guidance, to emphasize that the summary output is a draft that requires review. We also state the limitations in using GenAI tools, reminding users not to upload sensitive or prohibited materials into the system. Lastly, we clarify that the users must ensure the accuracy and appropriateness of the final document. To encourage author review and summary refinement, our application provides the end-user with the summaries of each section as well as the final summary with academic references for the stated risks associated with the proposed study. We additionally do not include a download format that would enable immediate download and submission—a user can download the final summary in a `.txt` file. There is a "view pdf" button that allows the user to view their uploaded pdf next to the summaries for additional review and validation of the drafted summary.

7 Ethical Considerations and Discussion

There are ethical considerations when using closed source, privately owned large language models for medical-care related tasks. While chatbots are highly functional, it is necessary to understand the

[5] https://vercel.com/

Type	Terms
Entities	
General	"Object", "Entity", "Group", "Person", "Organization", "Place"
Research	"Intervention", "AdverseEvent", "Outcome", "StandardOfCare", "Condition", "Disease", "Population" "RiskMitigation", "RiskFactor", "Complication", "LevelOfEvidence", "Citation", "StudyType"
Relationships	"HAS_RISK", "HAS_OUTCOME", "SUPPORTED_BY", "REFUTED_BY", "INCREASES_RISK", "REDUCES_RISK"

Table 4: Terms used for entity and relationship schema builder.

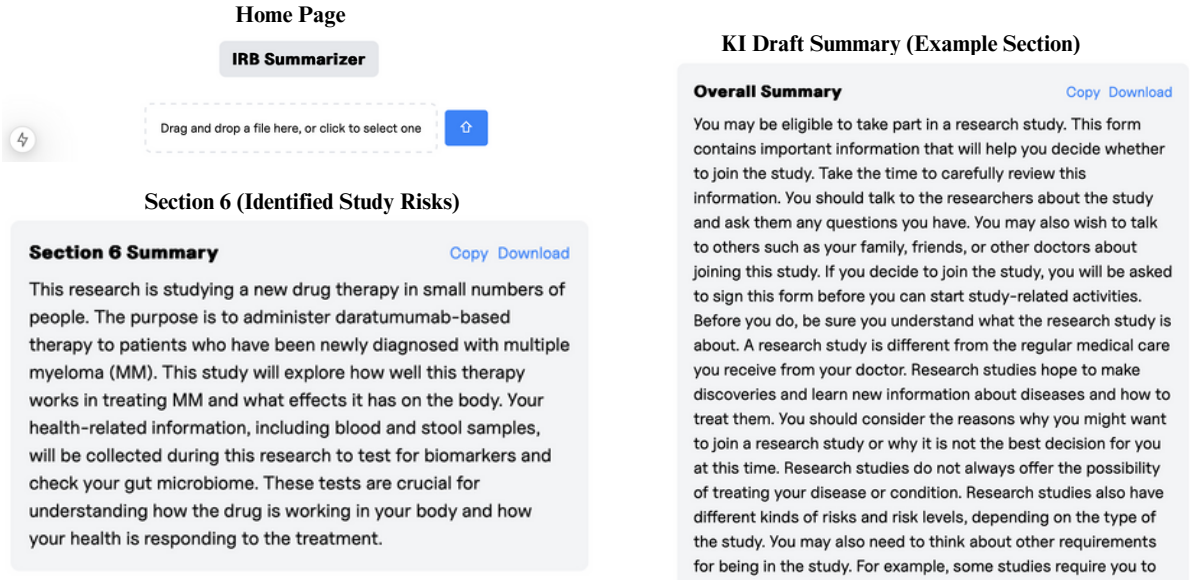

Figure 4: TAGRAG system UI and output example.

implications of using personally identifiable information (PII) data as input. Medical professionals using chatbots should evaluate the appropriate use-cases and data for a given task.

In our experimental design process, we identified three main challenges in using a RAG system for KI summary generation of ICDs: (1) preserving medical accuracy, (2) differentiating ambiguous topics, and (3) using simple language where appropriate. Balancing simple language for non-experts while maintaining necessary medical terms is difficult for a chabot, thus we assigned a persona that that explicitly stated the task of focusing and interpreting ICDs to potential non-expert study participants. Additionally, the RAG system struggled to different more ambiguous topics such as distinct risks associated with the proposed study, which we addressed by including the TAGRAG component in our final version of the KI summary application.

The challenges we identified aligned with prior work in related application areas (Alkhalaf et al., 2024; Shyr et al., 2024) and we focused on addressing these challenges via our hybrid RAG and GraphRAG system. We believe that other RAG systems can also benefit from including external expert knowledge base that can provide auxiliary information for the domain-specific task. While our study focused on IRB informed consent documents and medical research, the general pipeline of our hybrid system can be adapted to other domains.

8 Conclusion

Our hybrid RAG and TAGRAG system improves the routine procedure of generating key information sections of IRB informed consent documents. The KI summary application leverages GenAI to assist clinical trial experts in the preparation of informed consent documents and improves the communication between experts and non-expert potential participants. By including the expert TAGRAG

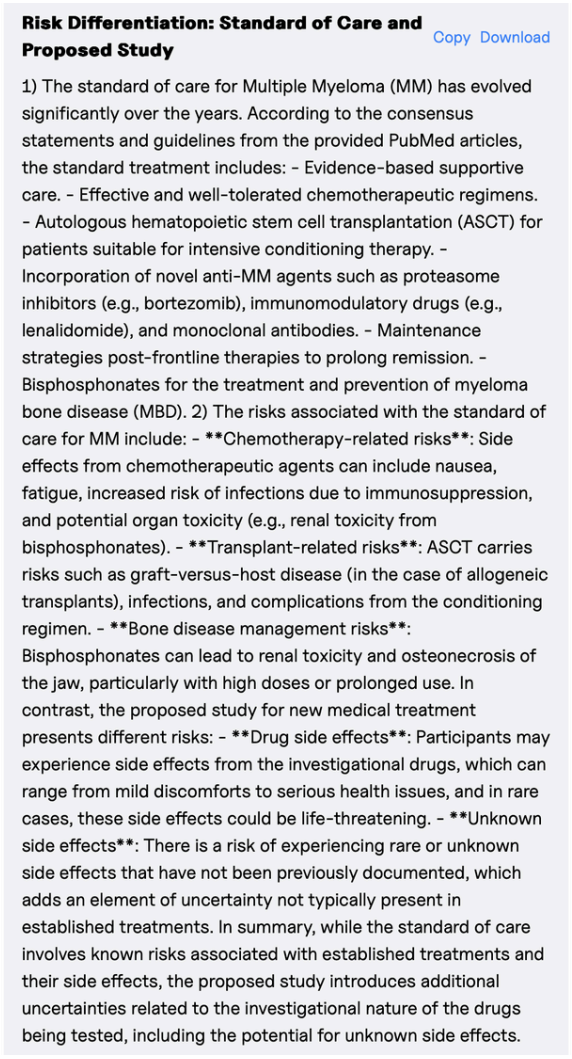

Figure 5: Risk differentiation sample output.

in our system we address the current challenge of risk extraction in RAG systems, and ensure that the end-user (clinical trial expert) is well-informed with relevant research evidence to support their summaries to the affected users (potential study participants).

Acknowledgments

We would like to thank the subject matter experts and principal investigators affiliated with the University of Michigan's Michigan Institute for Clinical and Health Research. This work was conducted under the MICHR Clinical and Translational Science Pilot Award (Grant UM1TR004404).

References

Mohammad Alkhalaf, Ping Yu, Mengyang Yin, and Chao Deng. 2024. Applying generative ai with retrieval augmented generation to summarize and extract key clinical information from electronic health records. *Journal of Biomedical Informatics*, page 104662.

Yuval Barak-Corren, Rebecca Wolf, Ronen Rozenblum, Jessica K Creedon, Susan C Lipsett, Todd W Lyons, Kenneth A Michelson, Kelsey A Miller, Daniel J Shapiro, Ben Y Reis, et al. 2024. Harnessing the power of generative ai for clinical summaries: Perspectives from emergency physicians. *Annals of Emergency Medicine*.

Darren Edge, Ha Trinh, Newman Cheng, Joshua Bradley, Alex Chao, Apurva Mody, Steven Truitt, and Jonathan Larson. 2024. From local to global: A graph rag approach to query-focused summarization. *arXiv preprint arXiv:2404.16130*.

Zakaria Hammane, Fatima-Ezzahraa Ben-Bouazza, and Abdelhadi Fennan. 2024. Selfrewardrag: enhancing medical reasoning with retrieval-augmented generation and self-evaluation in large language models. In *2024 International Conference on Intelligent Systems and Computer Vision (ISCV)*, pages 1–8. IEEE.

Minbyul Jeong, Jiwoong Sohn, Mujeen Sung, and Jaewoo Kang. 2024. Improving medical reasoning through retrieval and self-reflection with retrieval-augmented large language models. *Bioinformatics*, 40(Supplement_1):i119–i129.

National Center for Biotechnology Information. 2000. Pubmed database. Accessed: 2025-01-01.

Siva Sai, Aanchal Gaur, Revant Sai, Vinay Chamola, Mohsen Guizani, and Joel JPC Rodrigues. 2024. Generative ai for transformative healthcare: A comprehensive study of emerging models, applications, case studies and limitations. *IEEE Access*.

Cathy Shyr, Randall W Grout, Nan Kennedy, Yasemin Akdas, Maeve Tischbein, Joshua Milford, Jason Tan, Kaysi Quarles, Terri L Edwards, Laurie L Novak, et al. 2024. Leveraging artificial intelligence to summarize abstracts in lay language for increasing research accessibility and transparency. *Journal of the American Medical Informatics Association*, 31(10):2294–2303.

Ozan Unlu, Jiyeon Shin, Charlotte J Mailly, Michael F Oates, Michela R Tucci, Matthew Varugheese, Kavishwar Wagholikar, Fei Wang, Benjamin M Scirica, Alexander J Blood, et al. 2024. Retrieval-augmented generation–enabled gpt-4 for clinical trial screening. *NEJM AI*, page AIoa2400181.

Jonah Zaretsky, Jeong Min Kim, Samuel Baskharoun, Yunan Zhao, Jonathan Austrian, Yindalon Aphinyanaphongs, Ravi Gupta, Saul B Blecker, and Jonah Feldman. 2024. Generative artificial intelligence to transform inpatient discharge summaries to

patient-friendly language and format. *JAMA network open*, 7(3):e240357–e240357.

Zhaoyan Zhang. 2024. Rag for personalized medicine: A framework for integrating patient data and pharmaceutical knowledge for treatment recommendations. *Optimizations in Applied Machine Learning*, 1(1).

A Structured Prompts for KI Summarization

Here we provide each section prompt used to generate the full KI summary:

Section 1

```
    "section1": [
        "Who can take part in this study
?",
        "What are the eligibility
criteria for this study?",
        "Are children eligible to
participate in this study, either as
primary participants or in any other
capacity?",
        (
            "Choose the text below that
is most appropriate.\n"
            "---\n\n"
            "If children are eligible to
 participate in the study, write the
following text verbatim:\n"
            "You, or your child, may be
eligible to take part in a research
study. Parents or legal guardians who
are giving permission for a child's
participation in the research, note that
 in the sections that follow the word '
you' refers to 'your child'. This form
contains information that will help you
decide whether to join the study. All of
 the information in this form is
important. Take time to carefully review
 this information. After you finish, you
 should talk to the researchers about
the study and ask them any questions you
 have. You may also wish to talk to
others such as your friends, family, or
other doctors about your possible
participation in this study. If you
decide to take part in the study, you
will be asked to sign this form. Before
you do, be sure you understand what the
study is about.\n\n"
            "Otherwise, if children are
not eligible to participate in the study
, or it is not possible to determine
whether they are, then write the
following text verbatim:\n"
            "You may be eligible to take
 part in a research study. This form
contains important information that will
 help you decide whether to join the
study. Take the time to carefully review
 this information. You should talk to
the researchers about the study and ask
them any questions you have. You may
also wish to talk to others such as your
 family, friends, or other doctors about
 joining this study. If you decide to
join the study, you will be asked to
sign this form before you can start
study-related activities. Before you do,
 be sure you understand what the
research study is about."
        )
    ],
```

Section 2 (used if relevant and no generative text required)

```
            "section2": "A research
study is different from the regular
medical care you receive from your
doctor. Research studies hope to make
discoveries and learn new information
about diseases and how to treat them.
You should consider the reasons why you
might want to join a research study or
why it is not the best decision for you
at this time.",
```

Section 3 (used if relevant and no generative text required)

```
            "section3": "Research
studies do not always offer the
possibility of treating your disease or
condition. Research studies also have
different kinds of risks and risk levels
, depending on the type of the study.
You may also need to think about other
requirements for being in the study. For
 example, some studies require you to
travel to scheduled visits at the study
site in Ann Arbor or elsewhere. This may
 require you to arrange travel, change
work schedules, find child care, or make
 other plans. In your decision to
participate in this study, consider all
of these matters carefully."
```

Section 4

```
    "section4": [
        "What is the disease or
condition targeted by the research study
?",
        "What is the purpose or
objective of the research study?",
        "How many people are expected to
 take part in the research study?",
        "Will the research study involve
 the collection of biological specimens
such as blood, urine, tissue, cells, DNA
, etc.?",
        "What types of specimens will be
 collected and for what purposes?",
        (
            "You have been provided with
 template text and instructions below.
To customize the text:\n"
            "1. Make a decision at each
choice point indicated by angle brackets
 (<< >>) with options separated by
slashes (//). Select the option that
best matches the study particulars. If
the existing options are not appropriate
, you may choose to omit them or to
create a more appropriate alternative.\n
"
```

```
            "2. Replace placeholders
enclosed in double brackets ([[ ]]) with
 pertinent details based on your
understanding of the research study.\n"
            "3. Use lay-friendly
language to describe the study. Do not
use technical or scientific jargon
unless there is no plain language
alternative or converting to plain
language would change the meaning of the
 text, such as in the case of disease or
 procedure names.\n"
            "5. When technical terms,
scientific jargon, or acronyms must be
used, attempt to define them using plain
 language the first time they are used.
For example, 'This research is studying
DIPG (diffuse intrinsic pontine glioma),
 a type of brain tumor that occurs in
children.'\n"
            "---\n\n"
            "This research is <<
studying // collecting >> << a // a new
//  >> [[state the general category of
the object of the study, for example: '
drug', 'device', 'procedure', '
information', 'biospecimens', '
behavioral change', 'diagnostic tool',
etc. If applicable, also indicate
whether or not the object of the study
has already been approved by the Food
and Drug Administration (FDA) and for
what]] in << people // large numbers of
people // small numbers of people //
children // large numbers of children //
 small numbers of children >>. The
purpose is to [[briefly describe the
purpose of the study]]. This study will
[[briefly describe goals or objectives
]]. Your health-related information will
 be collected during this research. [[If
 any biospecimen collection will be
performed, indicate it here; otherwise,
do not mention biospecimen collection
]]."
        )
    ],
```

Section 5

```
    "section5": [
        "Does the study involve
randomization? Answer this question by
checking the Informed Consent document
for any of the following words: '
randomize', 'randomization', 'randomized
'? If any of these EXACT terms are
present, then the study involves
randomization and you should respond, '
Yes, this study involves randomization.'
 Otherwise you should respond, 'No, the
study does not involve randomization.'",
        "Review the Informed Consent
document with the aim of identifying if
it is a 'washout' study. A 'washout'
study is characterized by requiring
participants to discontinue certain
prescribed medications for a period
BEFORE or DURING the study. This
discontinuation is typically to ensure
that the effects of the study treatments
 can be observed without interference
from other medications. Analyze the
document for any instructions or
requirements that align with this
definition of a washout study. Based on
your analysis, determine if the provided
 example text indicates that the study
is a washout study. Respond with a clear
 'Yes' or 'No'.",
        (
            "You have been provided with
 template text and instructions after
the triple dashes below.\n"
            "Choose the text that is
most appropriate based on what you have
learned about this research study.\n"
            "When you encounter a choice
 enclosed in double angle brackets and
delimited by double forward slashes (<<
choice one // choice two>>), replace it
with the choice that best fits the study
's specifics. If you do not see an
appropriate choice, then you may choose
not to include any of the choices in
your response or you may choose to
generate an additional choice that is
more appropriate. \n"
            "When you encounter a phrase
 in this text that is enclosed by double
 brackets ([[example instructions]]),
replace it with relevant details derived
 from the STUDY INFORMATION provided
above. \n\n"
            "---\n\n"
            "Step 1: If the study
involves randomization, write the
following text, otherwise skip this step
:\n"
            "\n\nThis study involves a
process called randomization. This means
 that the << drug // device // procedure
 >> you receive in the study is not
chosen by you or the researcher. The
study design divides study participants
into separate groups, based on chance (
like the flip of a coin), to compare
different treatments or procedures. If
you decide to be in the study, you need
to be comfortable not knowing which
study group you will be in.\n\n"
            "Step 2: If the study
requires me to stop taking any
medications before I can participate,
write the following text, otherwise skip
 this step:\n"
            "\n\nThis study may require
you to stop taking certain medications
before and possibly during the research
study. If you decide to be in the study,
 you should understand that some
symptoms that were controlled by that
medication may worsen.\n\n"
            "If both Step 1 and Step 2
are skipped, meaning the study neither
involves randomization nor requires me
to stop taking a particular medication
before I can participate, then simply
write an empty space: "
        )
    ],
```

Section 6

```
    "section6": [
        (
            "Imagine that I am the study
 participant and you are explaining the
most important risks that are introduced
 or enhanced because of participation in
 this research study to me.\n"
            "Rather than trying to
explain every risk, focus on the risks
that will cause me pain or emotional
distrees. What are the most important
risks that you would explain to me?\n"
            "Do not include risks
associated with standard of care
treatments. Only include risks that
could reasonably be introduced or
enhanced due to participation in this
research study.\n"
            "Use plain language to
describe the risks with few words. Your
response should be no more than 3
sentences in length."
            ),
        (
            "You have been provided with
 template text after the triple dashes
below. Adhere to this text in your
response. When you encounter a phrase in
 this text that is enclosed by double
brackets ([[example instructions]]),
replace it with relevant details based
on what you have learned about this
research study. \n\n"
            "---\n\n"
            "There can be risks
associated with joining any research
study. The type of risk may impact
whether you decide to join the study.
For this study, some of these risks may
include [[Briefly describe the risks
while maintaining a formal tone]]. More
detailed information will be provided
later in this document."
        )
    ],
```

Section 7

```
    "section7": [
        (
            "Imagine that I am the study
 participant and you are explaining the
benefits of participating in this study.
 \n"
            "Create a list of the
benefits and categorize them based on
whether they will directly benefit me. \
n"
            "Do not mention financial
compensation.\n"
            "---\n\n"
            "[Direct personal benefits
to me]\n"
            "<List direct personal
benefits to me. If there are no direct
personal benefits me, then skip this
section>\n\n"
            "[Other potential benefits]\
n"
            "<List other significant
potential benefits>"
            ),
        (
            "You have been provided with
 template text after the triple dashes
below. Adhere to this text in your
response. "
            "When you encounter a choice
 enclosed in double angle brackets and
delimited by double forward slashes (<<
choice one // choice two>>), "
            "replace it with the choice
that best fits the study's specifics. If
 you do not see an appropriate choice,
then you may choose not to include "
            "any of the choices in your
response or you may choose to generate
an additional choice that is more
appropriate."
            "\n\n"
            "When you encounter a phrase
 in this text that is enclosed by double
 brackets ([[example instructions]]),
replace it with relevant details based "
            "on what you have learned
about this research study.\n"
            "If there are no meaningful
direct personal benefits to me, then
select the second choice in the template
 text below. Otherwise, select the first
 choice.\n"
            "---\n\n"
            "<<This study may offer some
 benefit to you now or others in the
future by "
            "[[Briefly summarize
benefits based on what you have learned
about this research study. Make sure the
 summarized text fits with the rest of
this sentence and doesn't repeat or
restate information that has already
been provided.]]>> "
            "// "
            "This study may not offer
any benefit to you now but may benefit
others in the future by "
            "[[Briefly summarize
potential benefits based on what you
have learned about this research study.
Make sure the summarized text fits with
the rest of this sentence and doesn't
repeat or restate information that has
already been provided.]]>>. More
information will be provided later in
this document."
        )
    ],
```

Section 8

```
    "section8": [
      "How much of my time, in total,
will be needed to take part in this
study? How long will I be in the study?
 What is the total duration of the study?
 In other words, how much of my time
will be taken up by the study and how
long will the overall study last?",
        (
            "After the triple dashes
```

```
below, you have been provided with
template text. Adhere to this text in
your response, replacing any double
bracketed instructions ([[example
instructions]]), with relevant
information about the research study.\n"
            "---\n\n"
            "The study will take [[
Indicate how long the subject will be in
 the study based on what you have
learned about this research study]]."
        )
    ],
```

Section 9

```
     "section9": [
        "If I decide not to take part in
 this study, what other options do I
have?",
        (
            "If participating in the
study will not affect my current or
future treatment/care options, or if
this question is not applicable to this
study, respond with the following text:
\n"
            "'Even if you decide to join
 the study now, you are free to leave at
 any time if you change your mind.'\n\n"
            "Otherwise, respond with the
 following text:\n"
            "'You can decide not to be
in this study. Alternatives to joining
this study include [[Based on what you
have learned about this research study,
briefly specify potential treatment/care
 alternatives for this disease or
condition such as the current standard
of care]].\n\n"
            "Even if you decide to join
the study now, you are free to leave at
any time if you change your mind.'"
        )
    ]
```