

Revolutionizing Informed Consent Form Analysis with Generative AI: Enhancing Efficiency in Drug Development and Data Re-use

Presenter: Hangyu (Cedric) Liu

Key Contributors: Andrew Borgman, Jake Gagnon, Dan Boisvert, Haleh Valian

Biogen

Feb 2024, Bethesda, MD

Agenda

- The Problem and Motivation
- The Solution - Large Language Models (LLMs)
- A Step-by-Step Guide for Applying Trustworthy LLMs in the Pharmaceutical Industry
- Key Takeaways and Future Plans

Inform Consent Form (ICF) Review

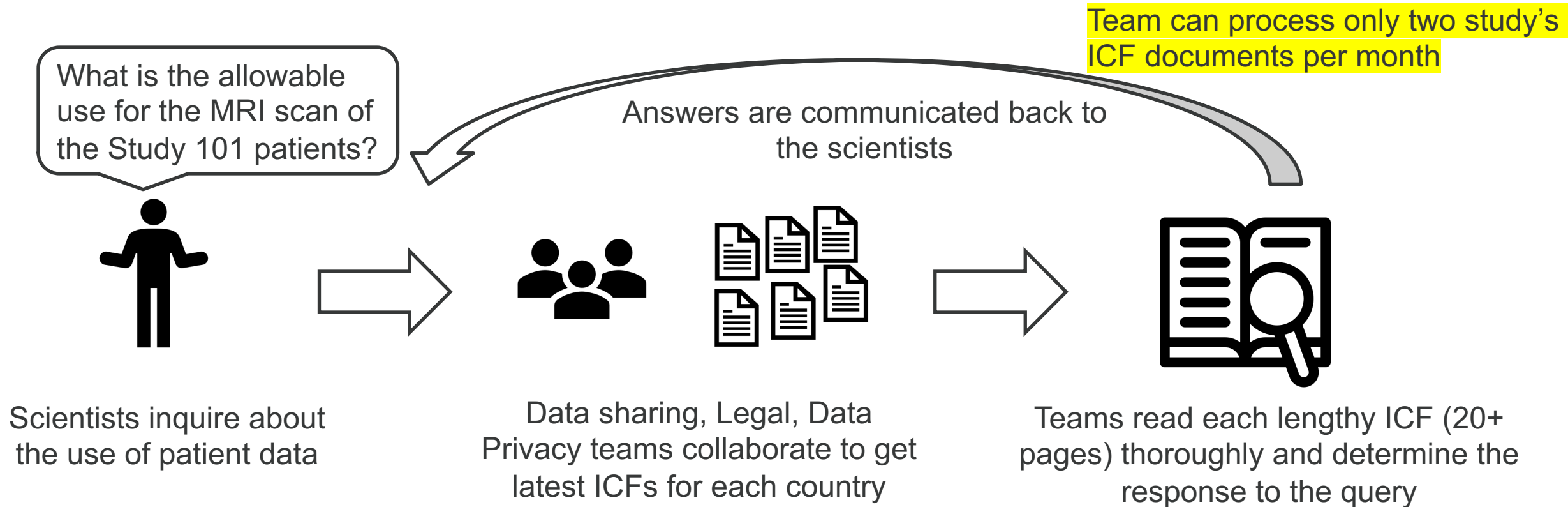
Challenges:

- Reuse of personal data must comply with consent as basis for processing
- Each study has LOTS of ICFs (e.g. Main, Future Scientific Research, Genetic, per country, per site)
- Data usage rules differ from country to country



Motivation - Researchers Often Face Challenges With Rules For Re-using Data

In biomarker and translational science, confirming patient consent for sample use in ICFs is important, but also challenging due to the amount and complexity of ICFs. This leads to delays in finding important information, slowing down data analysis and research progress.



Unable to Fully Reuse all Available Data

Too complicated to review all ICFs (we have two bad choices)

Risk Adverse

Don't use any data/samples/images beyond the initial study = limits research, is not fully utilizing data to the full extent we told patients

Risk Forward

Potentially use data outside of legal basis conflicting with GDPR and other data privacy laws. Reputational damage if seen as not using data appropriately

The right answer is to find a way to improve the efficiency of ICF review, and use the data up to its full legal limit but no further

The Solution, Benefits, and the Next Problem

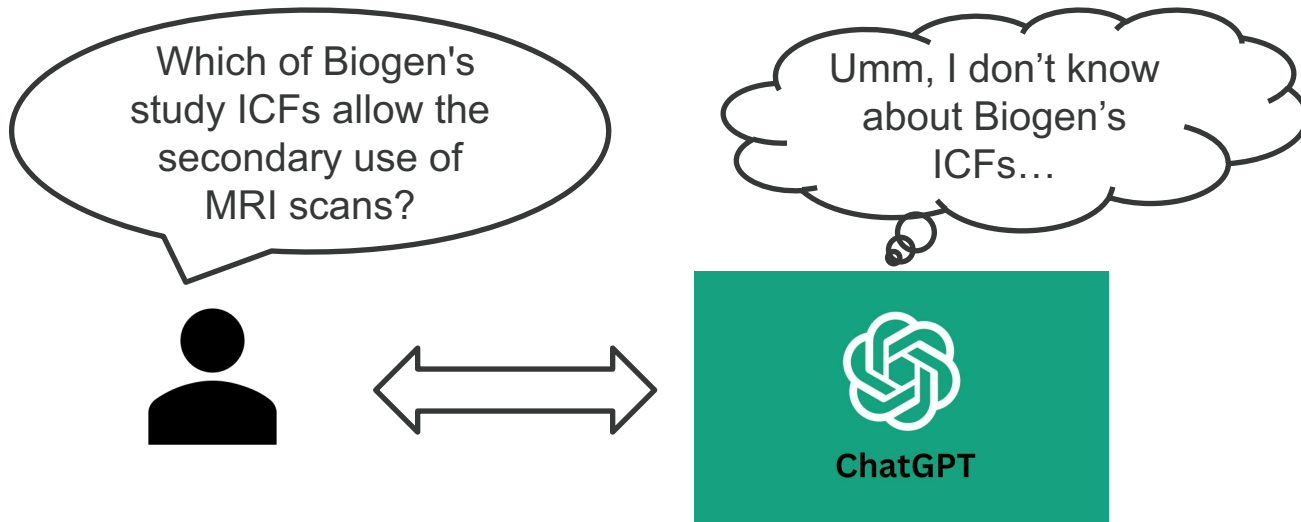
Large Language Models (LLMs) provides a scalable solution for reviewing unstructured data, such as ICF documents

Benefits:

- Reduce turn-around time of data-reuse queries from scientists
- Reduce cost of manual effort
- Improve the efficiency, scalability and consistency in ICF analysis

But naively using generic LLMs in the pharmaceutical industry poses problems with data protection, accuracy, reliability and trustworthiness

Limitations of Large Language Models (LLMs)



Limitations of generic LLMs:

- Lack knowledge of internal private data
- Static knowledge with a cutoff date of their training data (e.g. 2021 Sep for GPT 3.5)
- Hallucinations – generate inaccurate or untrue information
- Biased training data

How to Apply Trustworthy LLMs in the Pharmaceutical Industry

Step 1: Make it Safe

Step 2: Make it Factual

Step 3: Make it Reproducible

Step 4: Make it Trustable

Step 5: Make it Easy for the User

Step 1: Make it Safe – Technical Infrastructure and Data Protection

Safeguard our data with robust infrastructure

Technical Infrastructure and Data Protection

Data loss and unintended data leakage are critical risks of using LLMs

Data Privacy

- Deploy open-source LLMs internally
- For commercial close-source LLMs, establish the rigorous data protection agreement with 3rd party vendors, isolate the data storage, encrypt data, and set up the schedule for auto-deletion of data

Open-source LLMs (LLaMa 2, WizardLM, etc.):

- Deploy internally with vLLM hosting framework on High-Performance Computing (HPC) infrastructure

Closed-source Commercial LLMs (GPT 3.5, GPT 4, etc.):

- Deploy on Azure cloud, and establish a rigorous data sharing agreement with Microsoft

Deemed out of scope of 21CFR Part 11 as the activity is not Good Clinical Practice (GCP)

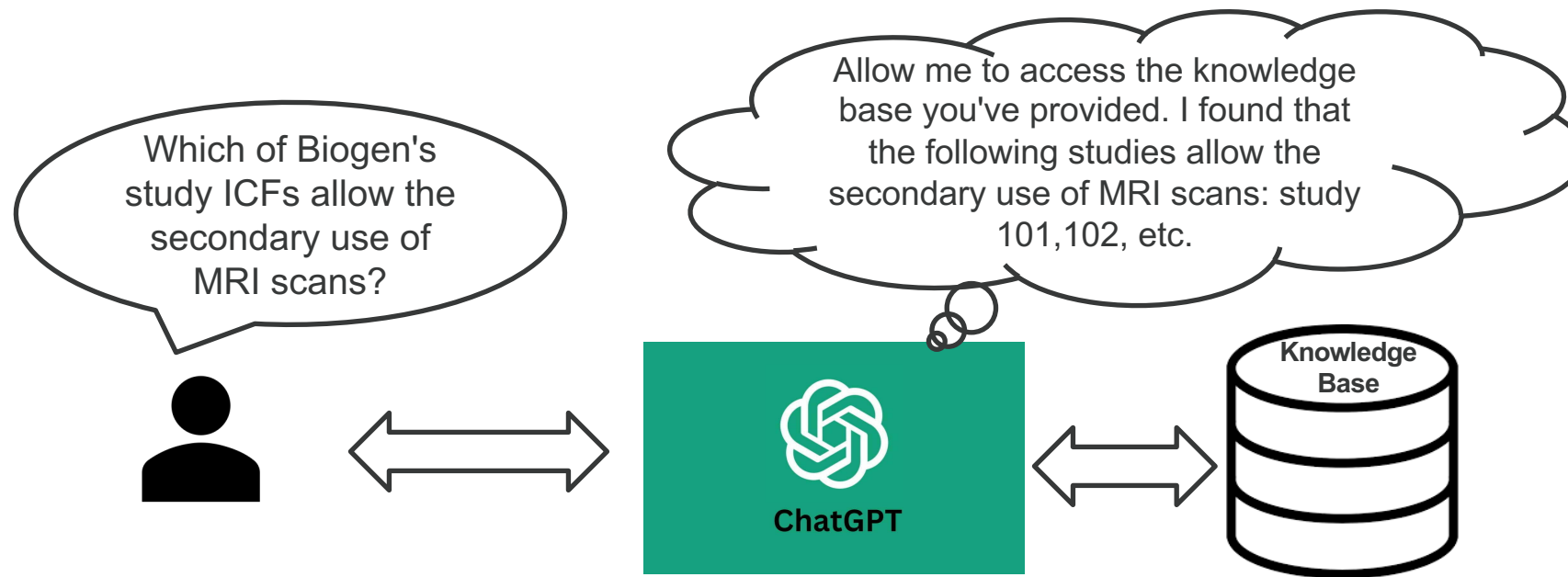
Step 2 : Make it Factual – Retrieval-Augmented Generation (RAG) Framework

Reduce hallucinations and ingest internal knowledge

Retrieval-Augmented Generation (RAG)

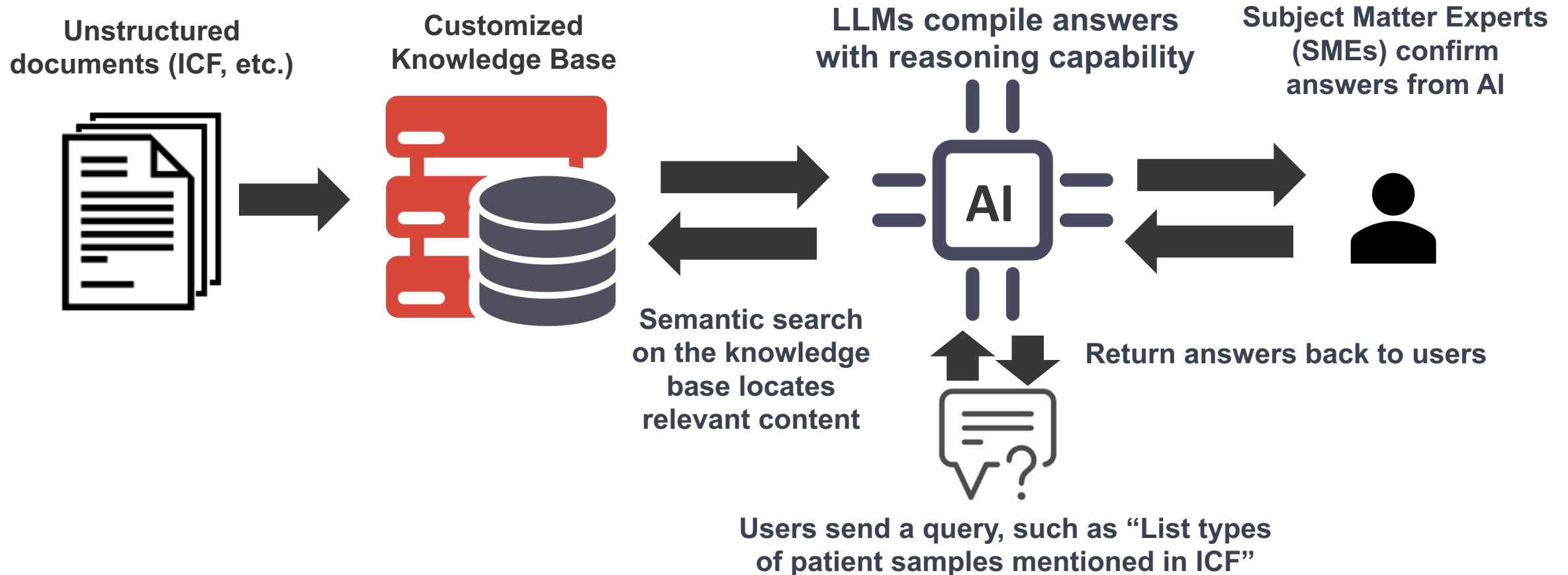
In the context of ICF analysis, the need for accurate and factual responses is imperative. Due to limitations of generic LLMs, we opt for the Retrieval-Augmented Generation (RAG) framework to address these constraints

RAG retrieves relevant contents from documents and then uses them as context for generating responses, it ensures that LLMs can access to real-time and domain-specific data from a knowledge base, minimizing the likelihood of hallucinations of LLMs



Overall Technical Framework

At a high level, there are 3 major components in our technical framework: (1) Ingestion of ICF documents as a knowledge base; (2) Context-based semantic search; (3) LLM-powered question answering

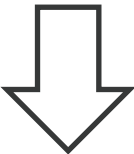


Step 3: Make it Reproducible

Derive a way to create consistency from the probabilistic nature of LLMs

Considerations To Improve Robustness And Reproducibility Of LLMs

Same Questions, Varied Responses



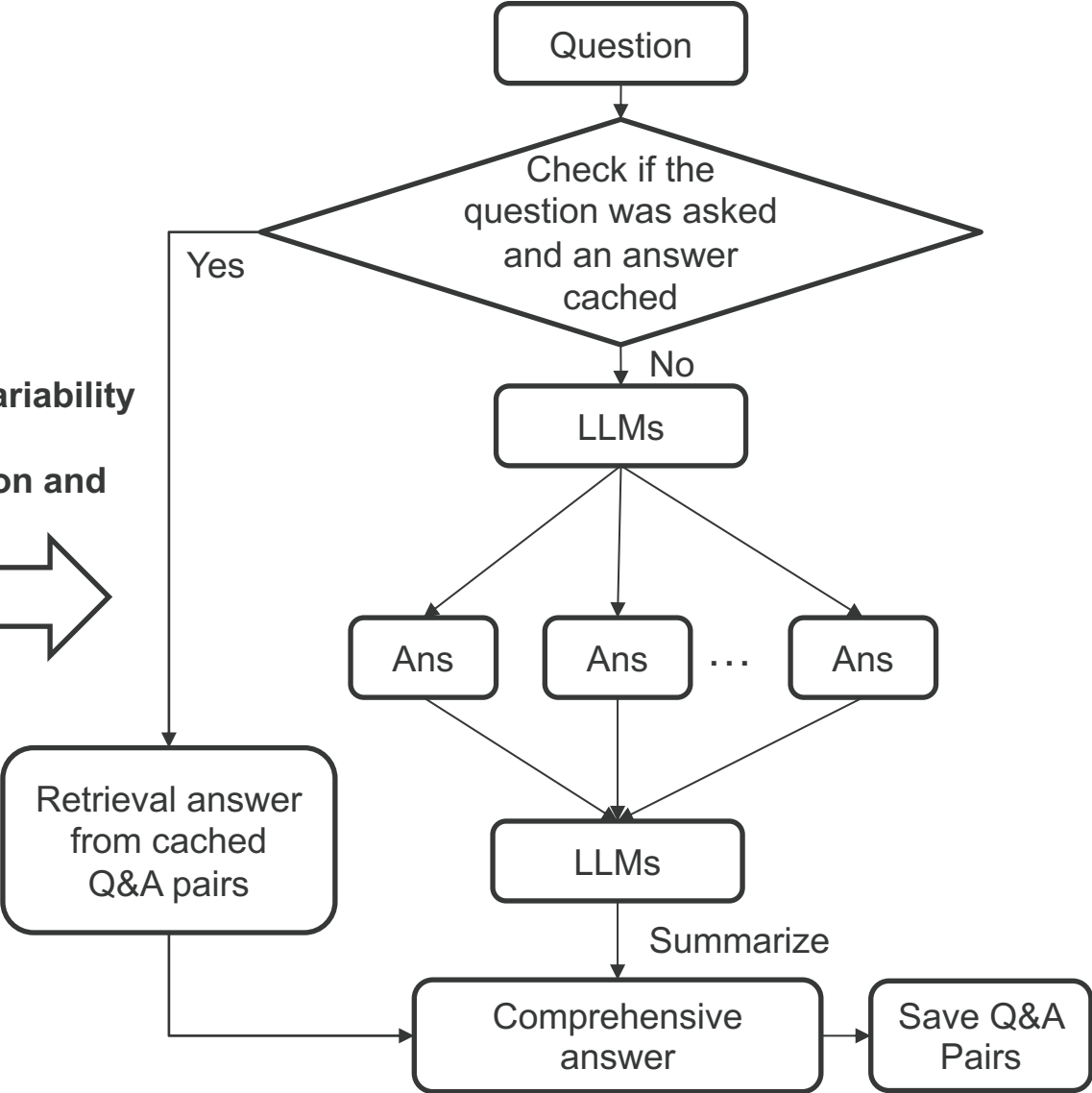
- You**
explain what is inform consent form in one sentence

ChatGPT
An informed consent form is a document that explains the details of a medical treatment or procedure and requires the patient's signature to acknowledge understanding and agreement to proceed.
- You**
explain what is inform consent form in one sentence

ChatGPT
An informed consent form is a document that outlines the risks, benefits, and alternatives of a particular medical procedure or treatment, ensuring that a patient agrees to it with full understanding of its implications.

Minimize Answer Variability

Text Summarization and Q&A Cache



Step 4: Make it Trustable - Verification Framework

Getting it right and building trust

Verification Framework

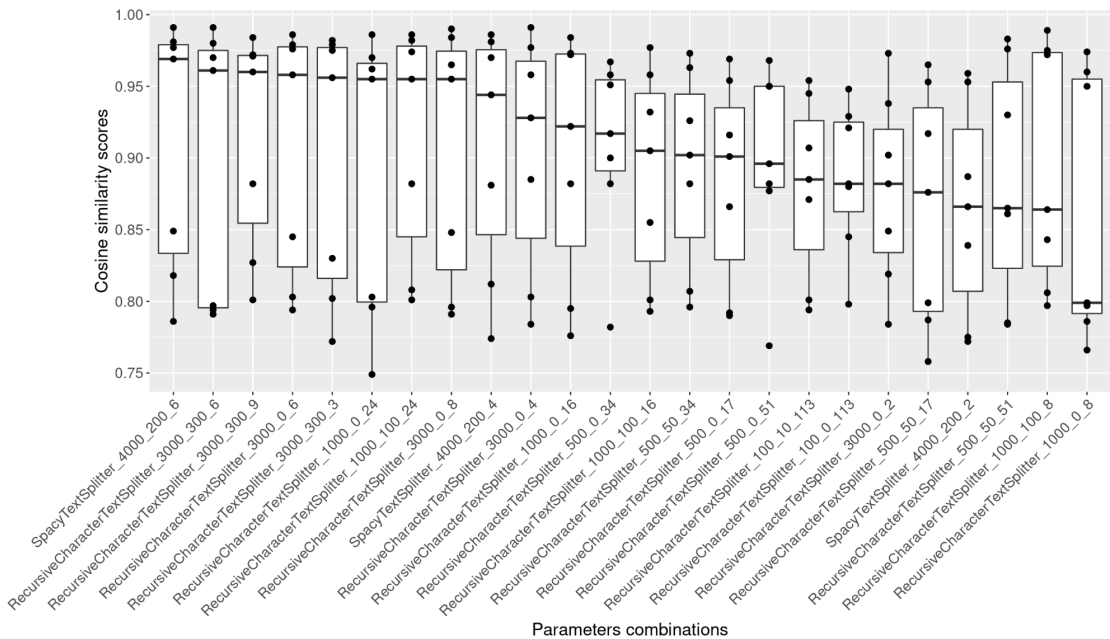
The best performance model and parameters

The following combination of LLM model and parameters has the best performance (median similarity score: **0.969**) with the median total cost **\$0.004**:

- **model**: gpt-35-turbo
- **text splitter**: SpacyTextSplitter
- **chunk size**: 4000
- **overlap**: 200
- **K**: 6

Parameters tested Answers Plot of cosine similarity scores

The following plot shows the cosine similarity scores.



We sampled 150 pre-existing ICF question-answer pairs prepared by Subject Matter Experts (SMEs) as our validation dataset. To verify the model's output, we employed two methods:

1. Calculated the cosine similarity score between the model's generated answers and the human-provided answers. The best-performing model demonstrated a **96%** similarity to the SMEs' answers
2. Additionally, SMEs manually reviewed the model's answers and double-confirmed the accuracy of the model

Step 5: Make it Easy for the User

Make the interaction seem obvious



Welcome to the ICF Q&A Chatbot! 🤖

The ICF Q&A Chatbot uses Artificial Intelligence (AI) algorithms to help you answers questions about ICF documents.

Upload a PDF of an ICF (Informed Consent Form)



Drag and drop file here

Limit 200MB per file • PDF, PDF

Browse files

* This is a publicly available example ICF, not a real Biogen's ICF

Key Takeaways

Achievements:

- This framework can help any company looking to improve their management and understanding of ICFs or data use agreements. It speeds up the decision-making process in data sharing and reuse
- Established a customized verification framework with both quantitative measures and SME validation
- Privacy considerations of LLMs, methods to mitigate hallucinations, interactive Chatbot, etc.

Value:

- Reduce ICF review time **from ~2 studies' ICFs per month to ~2 studies' ICFs per hour**
- Significantly increase the efficiency of clinical data re-use and speed up the overall research progress

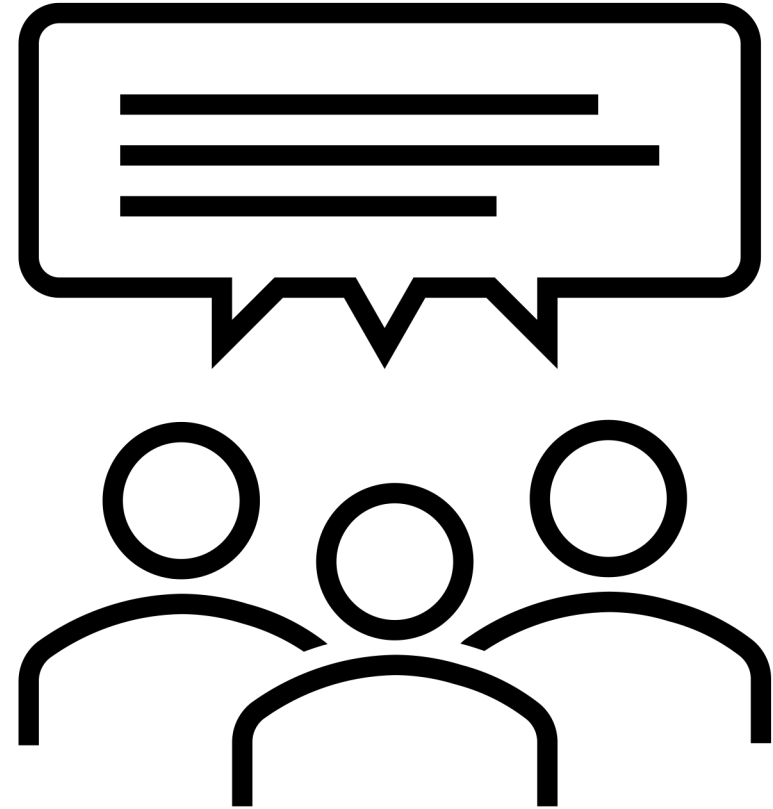
Risks:

- Acknowledges risks associated with LLMs, such as the potential for misinformation, and highlights the importance of **human oversight** to verify outputs and maintain ethical standards in research practices

Future Possibilities

- **Future Plan to open-source Python Code**
 - Collaborate industry-wide via GitHub, enable pharmaceutical companies of all sizes to utilize, contribute to, and enhance the framework
- **Broad Application of Our LLM Framework**
 - Applying natural language questions to other unstructured data sources for industry and regulators (e.g. CSR, SAP, Protocols)
 - Patient facing applications – Enable a patient to ‘chat’ with an ICF or a protocol to understand risks of participating in a clinical trial
 - Summarizing documentation for patients and other lay-readers (e.g. Lay language summaries)
- **More Advanced AI Algorithms To Enhance Performance**
 - GPT-4-Turbo and similar advanced LLMs for their larger context windows and better performance
 - Explore advanced methods like Query Transformation, Hypothetical Document Embeddings (HyDE), and reranking to boost RAG framework performance.
 - Create a pipeline to batch process numerous ICFs to enable the scalability of the framework.

Questions / Discussion



Thank you

If you have any questions, please contact

Hangyu (Cedric) Liu hangyu.liu@biogen.com

Andrew Borgman andrew.borgman@biogen.com

Jake Gagnon jake.gagnon@biogen.com

Dan Boisvert daniel.boisvert@biogen.com

Haleh Valian haleh.valian@biogen.com