



Applying Large Language Models to Interactive Information Retrieval: A Practical Exploration

Part III: Future of LLMs with IIR



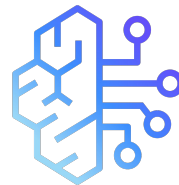
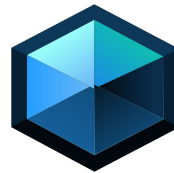
ZUVA



RMIT
UNIVERSITY

What we didn't talk about

- All the tools
 - Ollama, MSTY.app, HuggingFace (and local models)
- All the APIs
 - Azure, OpenAI, Google, Amazon Bedrock
- All the training
 - PEFT, (Q)LORA



The screenshot displays a chat window with a sidebar on the left containing various icons for navigation. The main chat area shows a conversation with a user and an AI assistant. The user's message is: "Give me an overview of interactive information retrieval research that uses LLMs". The AI's response is: "Here's an overview of interactive information retrieval (IIR) research that utilizes Large Language Models (LLMs):". Below this, the AI provides a detailed overview, starting with a "Background" section, followed by an "Introduction of LLMs in IIR" section, and then "Key Research Directions". The directions listed are: 1. LLMs as Retrieval Models, 2. Interactive Text Generation, 3. Conversational Interfaces, and 4. Personalized Information Retrieval. The chat concludes with a "Key Challenges" section, listing "Scalability and Efficiency" as a primary challenge. At the bottom, there is a text input field with the placeholder "Type your message here..." and a "Sync" button.

Model Instructions ^

You

Give me an overview of interactive information retrieval research that uses LLMs

Llama 3.1

< 2 / 2 > Here's an overview of interactive information retrieval (IIR) research that utilizes Large Language Models (LLMs):

Background

Interactive Information Retrieval is a field of study that focuses on designing systems and algorithms to help users interactively search, browse, and explore large datasets. IIR aims to provide relevant results in real-time by leveraging user feedback, such as queries, clicks, or ratings.

Introduction of LLMs in IIR

Recently, the advent of Large Language Models (LLMs) has revolutionized various applications, including natural language processing (NLP), text generation, and information retrieval. LLMs are pre-trained on massive datasets to learn patterns and relationships between words, enabling them to generate coherent and context-specific responses.

Key Research Directions

Several research directions have emerged that incorporate LLMs into IIR:

- LLMs as Retrieval Models:** Some studies explore using LLMs as a replacement for traditional retrieval models (e.g., TF-IDF) in information retrieval systems.
- Interactive Text Generation:** Researchers investigate the use of LLMs to generate text responses based on user queries, taking into account contextual and semantic relationships between words.
- Conversational Interfaces:** IIR studies using LLMs as a conversational interface for exploring datasets or providing recommendations.
- Personalized Information Retrieval:** Some work focuses on leveraging LLMs to build personalized information retrieval models that adapt to individual users' preferences.

Key Challenges

While the incorporation of LLMs into IIR has shown promise, several challenges need to be addressed:

- Scalability and Efficiency:** Supporting large-scale query processing

Type your message here...

Sync

The screenshot shows a chat application interface. At the top, there's a header with a document icon and the text 'Model Instructions'. Below this, the chat history is displayed. The first message is from 'You' asking for an overview of interactive information retrieval research. The second message is from 'gemma2:latest' providing a detailed response. The response starts with a paragraph about traditional IR, followed by a paragraph about LLMs' capabilities, and then a list of four key research areas: 1. Conversational Search, 2. Query Reformulation, 3. Document Summarization & Extraction, and 4. Personalized Search. Each area has a list of bullet points. At the bottom, there's a text input field with the placeholder 'Type your message here...' and a row of icons for various functions like search, settings, and a sync button.

Model Instructions

You

Give me an overview of interactive information retrieval research that uses LLMs

gemma2:latest

Interactive Information Retrieval with LLMs: A Revolution in Search

Traditional IR relies on keyword matching and ranking, often resulting in irrelevant or incomplete results.

LLMs (Large Language Models) are changing this paradigm by enabling **interactive** and **conversational search**, leading to a more natural and human-like interaction with information.

Here's an overview of key research areas:

- 1. Conversational Search:**
 - LLMs can understand complex queries expressed in natural language, allowing users to ask follow-up questions, refine their search based on previous results, and engage in a dialogue to obtain precise answers.
 - Examples include **chatbots integrated with search engines**, providing conversational guidance and summarizing relevant information from various sources.
- 2. Query Reformulation:**
 - LLMs can analyze user queries and suggest alternative formulations that are more effective for retrieval.
 - This helps overcome the limitations of keyword-based searches and retrieves more relevant results even when users struggle to articulate their needs precisely.
- 3. Document Summarization & Extraction:**
 - LLMs excel at summarizing large documents and extracting key information, enabling users to quickly grasp the essence of complex texts without reading everything.
 - This can be invaluable for research, news consumption, and understanding legal or technical documents.
- 4. Personalized Search:**
 - LLMs can learn user preferences and past search history to personalize

Type your message here...

Sync

The image is a screenshot of a web-based chat application. At the top right, there is a header with a plus icon, a speech bubble icon, and the text "Add Split Chat". Below this, on the left, is a calendar icon and the text "Model Instructions". The main chat area shows a conversation. On the left, a small person icon is next to the text "You". On the right, a larger speech bubble contains the text "Give me an overview of interactive information retrieval research that uses LLMs". Below this, another small person icon is next to the text "OpenAI: ChatGPT-4o". The main response from the AI is a detailed overview of Interactive Information Retrieval (IIR) research, structured with numbered sections and bullet points. The sections are: 1. Conversational Search, 2. Query Reformulation & Expansion, 3. Personalized & Context-Aware Retrieval, 4. Explainable & Transparent Search, and 5. Interactive Summarization & Result Presentation. Each section contains several bullet points detailing research trends and challenges. At the bottom of the chat area, there is a text input field with the placeholder "Type your message here..." and a "Send" button. To the right of the input field, there are icons for voice input, image input, and a "Sync" button.

Gemini Deep Research example

Give me an overview of interactive information retrieval (IIR) research that uses LLMs



Here's the plan I've put together. Let me know if you need any changes before I start researching.

LLMs in Interactive Retrieval



Research Websites

Give me an overview of interactive information retrieval (IIR) research that uses LLMs

(1) Search for academic papers and research articles on "interactive information retrieval" and "large language models".

(2) Identify research that specifically explores the integration or application of LLMs within IIR systems.

(3) Summarize the different ways LLMs are being used in IIR, such as query reformulation,...

[More](#)



Analyze Results



Create Report



Ready in a few mins

<https://g.co/gemini/share/2f5125c8ae36>

Where are LLMs being used now?

- Personalization
 - Wang et al. [Task Supportive and Personalized Human-Large Language Model Interaction: A User Study](#)
- Querying and Feedback
 - Dhole et al. [An Interactive Query Generation Assistant using LLM-based Prompt Modification and User Feedback](#)
- Simulation
 - Azzopardi et al. [SimIIIR 3: A Framework for the Simulation of Interactive and Conversational Information Retrieval](#)
- Query Sensitivity/Performance Prediction
 - Rezavi et al. [Benchmarking Prompt Sensitivity in Large Language Models](#)

Now is your chance!

- Soundboard your ideas!
- Ever wondered if LLMs are a good fit? Let's discuss!

Before we go



Some interactive questions...

Feedback

- If we do this again, what would you like to see?
 - Using LLMs to do research (e.g., analysis)?
 - More concrete problems?
 - More examples of prompting techniques?
 - Training LLMs?