LLMs in Disaster Risk Management:

A Handbook for CIMA

INDEX

Table of Contents

01

Introduction

02

Getting Started with LLMs

- What is LLM?
- Why do we need them?

03

Key Techniques

- Customized GPT
- Zero-Shot Prompt
- One-Shot Prompt and In-Context Learning
- RAG

04

LLMs in DRM: Practical Examples of Prompts

- Risk Assessment
- Early Warning Message
- Exercise Design

05

Recommendations and Limitations

- Suggestions according to Data and Resource Availability
- Limitations to Consider

06

Roadmap to RAG

Appendix

- Template Prompts
- Glossary
- Interactive PowerBI Dashboard

INTRODUCTION

Overview and Purpose of The Handbook

Welcome to "LLMs in DRM: A Practical Handbook for CIMA Professionals". This guide is the result of a collaborative capstone project between IE Master's in International Development students and CIMA Research Foundation, exploring the potential of Large Language Models (LLMs) in enhancing CIMA's Disaster Risk Management (DRM) practices.

As CIMA continues to lead in multi-hazard prediction, prevention, and management, this handbook aims to introduce you to the cutting-edge capabilities of LLMs and their practical applications within CIMA's workflows. Our research has focused on tailoring these powerful AI tools to CIMA's specific needs and operational context.

This guide is designed to empower you, as CIMA professionals, to harness the potential of LLMs in enhancing our organization's disaster management efforts. By exploring the contents of this handbook, you'll gain practical knowledge and tools to potentially improve our preparedness, response, and recovery operations. Let's begin our exploration of how LLMs can contribute to CIMA's mission of building more resilient communities through innovative and effective disaster management practices.

Getting Started with LLMs

What is LLM?

LLMs are advanced artificial intelligence systems designed to process and generate human-like text. These models are built using deep learning techniques, particularly neural networks, and are trained on vast amounts of textual data from diverse sources. An LLM utilized widely is OpenAI's ChatGPT, which stands for Generative Pretrained Transformer, describing the model utilized.

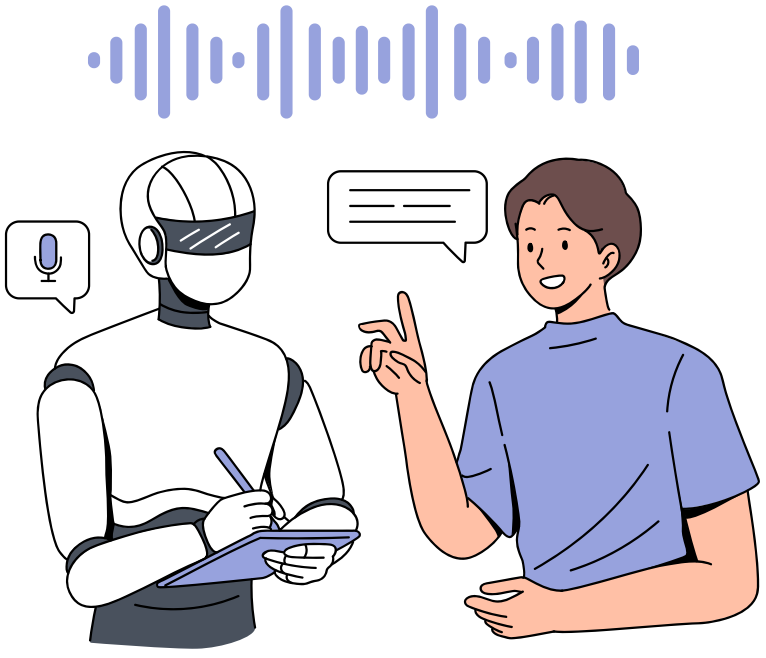
In the intersection field of DRM and LLMs, LLMs have a unique ability for tasks related to natural language processing and generation, especially useful when a response is already taking place, assisting in action plans, report summarization, chatbots, and translations.

Why do we need them?

LLMs' ability to process and generate human-like text allows for rapid information synthesis, enabling quick summarization of large volumes of data. This can be particularly useful in time-sensitive disaster scenarios. LLMs excel in natural language understanding, making them adept at interpreting complex queries and providing contextually relevant responses.

Advantages:

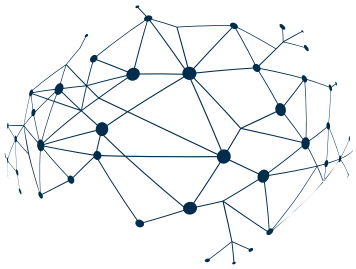
- To integrate data from a huge amount of data at CIMA's disposal;
- To assist DRM professionals in quicker retrieval of information;
- To keep CIMA on the cutting edge of innovation in their field.



Key Techniques

Customize GPT

The "Customize GPT" feature in ChatGPT exemplifies a method to serve domain-specific knowledge and frameworks into the model without changing the internal parameters.



Retrieval-Augmented Generation (RAG)

RAG is a method that leverages contextually-appropriate external sources to improve the relevance and contextuality of a model's output (Lewis et al., 2020; Guu et al., 2020). It is well recognized that RAG boosts LLM performances through integrating external data retrieval, leading to improvements in factual accuracy coherence and also consistency

Zero-Shot Prompt

Zero-shot learning uses pre-trained language models without fine-tuning or additional training

One-Shot Prompt and In-Context Learning

One-shot learning consists of using a single example or prompt for conditioning the model output. And ICL is an extension of one-shot learning, whereby in the input prompt, accompanying instructions and examples are given so that the model is helped in output generation

04

LLMs in DRM: Practical Examples of Prompts

In this chapter, the example prompts using the most efficient technique (one-shot prompt and ICL) will be provided in different tasks

An efficient prompt should follow these six criteria:

Contextual completeness	Does the prompt include all of the necessary context for the LLM to produce the best possible response?
Clarity and readability	What grade level can read and understand the response?
Relevance	Does the prompt accurately outline a task done by disaster management professionals?
Specificity	Does the prompt have specific details regarding location, audience, timeframe, etc.?
Conciseness	Is the prompt as brief as possible without losing crucial information?
Number of specific instructions	Does the prompt have clear instructions for the LLM to follow? (RICCE)

LLMs in DRM: Practical Examples of Prompts

Risk Assessment (Using RICCE Framework)

Region: Spain

Disaster Type: Flood

[**ROLE**]I am a disaster management professional based in Andalusia, Spain. **[INSTRUCTIONS]** Generate a full risk report for earthquakes in Murcia using the INFORM risk framework from your Knowledge. Your assessment should start with, detailing the prevalence of andalusia within Spain, the level of human exposure to floods, the vulnerability of Andalusia if an earthquake occurs, and identifying any lack of coping capacity Andalusia may have when addressing earthquake and its aftermath. **[CONTEXT]** This risk assesment will be utilized as a reference for policy makers and disaster response professionals in Andalusia, so regional specifics are encouraged. **[CONSTRAINTS]** Please do not generate instructions on 'how to write the assessment'. Instead, take those instructions and actually generate the assesment yourself with the data at your disposal, and access online sources if needed. You can make requests for certain types of data that would enhance the overall quality and specificity of the assement at the end of your response.

LLMs in DRM: Practical Examples of Prompts

Exercise Design (Using RICCE Framework)

Region: Mozambique

Disaster Type: Wildfire

[ROLE] I am a disaster response professional working on wildfires in Mozambique. **[INSTRUCTIONS]** Please write a full tabletop exercise for a wildfire in Maputo, Mozambique. **[CONTEXT]** This exercise will be used in an upcoming simulation with all emergency response stakeholders involved in flood management and response.The goal is to simulate a realistic flood event to test and improve their preparedness, response strategies, and inter-agency coordination. Maputo has a history of flooding, and the exercise should reflect the unique geographical, social, and infrastructure challenges of the area. **[CONSTRAINTS]** Please be highly specific in the exercsie scenario, making it realistic for the area. The exercise should be comprehensive, challenging, and tailored to enhance the particpants' response and management skills. You should include details such as the timeline of events, materials needed. key stakeholders involved, specific challenges faced/interjections, and expected outcomes. Ensure the scenario is culturally and contextually appropriate for Maputo, Mozambique.

04

LLMs in DRM: Practical Examples of Prompts

Early Warning Message (Using RICCE Framework)

Region: Phillipines

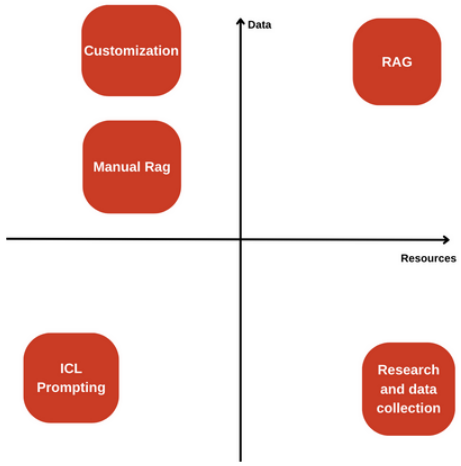
Disaster Type: Earthquale

[**ROLE]** I am a disaster response professional working on wildfires in the Philippenes. **[INSTRUCTIONS]** Please write an early warning message for residential areas in Luzon, Philippines. [**CONTEXT]** The message should alert residents of a potential wildfire and specify communities at risk. **[CONSTRAINTS]** Please be brief, clear, and give specific actions to take. The message should be both in English and the local language.

Recommendations and Limitations

Primary Recommendation: Integrate LLMs to Enhance Human Expertise

LLMs can significantly augment human expertise in DRM tasks by providing swift data integration and retrieval. While LLMs CANNOT yet replace human professionals entirely, their use can enhance the efficiency and effectiveness of DRM initial drafting for exercise scenarios, making them valuable tools for supporting human decision-making and operations.



Step 1: Explore data and resource availability

It is recommended that the organization explores data availability and resource availability for the implementation of data-providing LLM techniques before introducing LLM tools to their work. Based on the scenario framework on the left, choose the technique that works the best.

Step 2: Be aware of the scenarios would most require the supply of additional data

- in a region outside of the West
- a local dialect is widely spoken
- a disaster type that is less common
- a region with less robust data/infrastructure
- to perform a comprehensive risk assessment or issue an early warning message

Step 3-1: No Enough Data

It is recommended to provide as much context as possible through **ICL** prompting in the tools that performed well in this study: **Claude 3 and Mistral**. These techniques will improve the output when retrieving information or drafting an exercise.

Step 3-2: Utilization of RAG

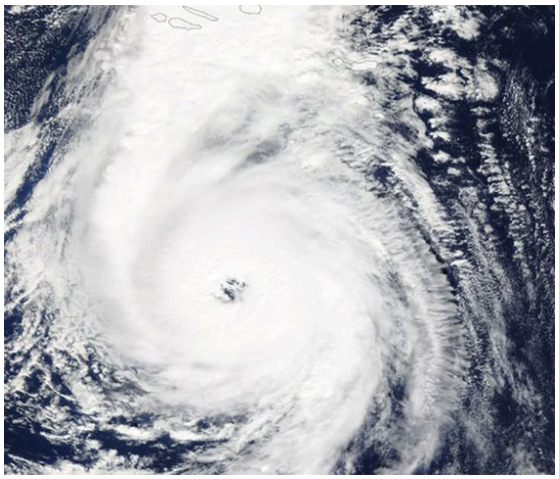
The RAG approach is recommended, though **resource-intensive**. This involves categorizing and organizing data using frameworks like RISK Inform to ensure the LLM has access to relevant and high-quality information.

05 Limitations to Consider

While LLMs has demonstrated the potential in disaster management, it's crucial to understand their current limitations:

- **Contextual Specificity:**

The primary limitation of LLMs is their inability to match human-level contextual specificity in outputs. This gap prevents LLMs from fully replacing human expertise in disaster management workflows.



- **Unstable Performance:**

The quality of LLM outputs can vary based on the available data and resources for implementation techniques such as fine-tuning or RAG.

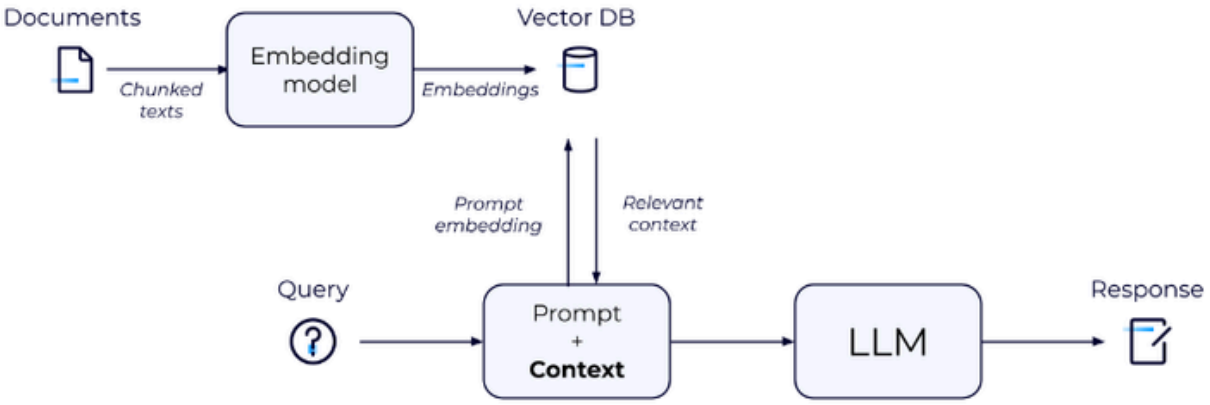
- **Supplementary Role and Human Oversight:**

LLMs are best suited for assisting with initial outlining and drafting of text-based deliverables. However, their outputs require thorough review and enhancement by disaster management professionals to ensure accuracy, relevance, and contextual appropriateness.

LLMs cannot replace human expertise but can free up time for professionals to focus on critical decision-making and specialized tasks.

06

Roadmap to RAG



source: Confluent

To fully harness the current capabilities of LLMs in the disaster management space, the recommendation for the client is to consider building a RAG pipeline. Clients can start with existing LLM and consider developing an own local LLM in the long run to avoid dependance on external systems. This page gives a high level overview of the current status of RAG and sector-specific recommendations for best practices and next steps to build a RAG.

The following breakdown of the development of a RAG pipeline is based on the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, to allow for a structured approach that can be well combined with other project management approaches. This process methodology was chosen as it is widely used for data science projects (Data Science Process Alliance, 2024). The model is based on six phases describing a data science lifecycle, which aligns with the steps required for building a RAG pipeline according to the literature. The generalizable and adoptable nature of the CRISP-DM model is one of the reasons it is the most commonly used process model for data science projects according to the Data Science Process Alliance (2024). For each phase of the process, the main focus, related context, including possible sources and tools for building the RAG pipeline, are described.

https://www.confluent.io/learn/retrieval-augmented-generation-rag/#:~:text=Access%20to%20real%2Dtime%20information,most%20current%20and%20accurate%20responses

Roadmap to RAG

1

Business Understanding

Definition of the purpose of the RAG pipeline

- Clearly define the goal of building a RAG pipeline.
- Engage stakeholders to align goals and requirements, ensuring a clear direction and plan for the project.
- This phase sets the foundation for all subsequent steps.

2

Data Understanding

Identifying, collecting and analyzing data

- Identify and collect relevant data for the retrieval model.
- This could include disaster reports, risk assessments, and early warning protocols.
- Analyze data quality and relevance, as the output fo the RAG system depends heavily on input quality.

3

Data Preparation

Prepare data used by the retrieval model and load it into vector database

- Add meta data to chunks to increase the performance of retrieval and ensure relevant data is retrieved.
- This involves cleaning, formatting, and potentially anonymizing data.
- Choose data indexing approach that suits the project
- Embedding data into vector database.
- Leverage meta data to increase relevance of retrieved data

4

Modelling

Development of RAG system, including retrieval and generative model

- Following the RAG architecture, the RAG pipeline development requires not only the modeling of the LLM and Retriever model but also the Embedding model and vector database for the previous phase.
- Consider local LLM in the long run to decrease dependencies on external systems

5

Evaluation

Testing, verifying & refining models and data

- After the preparation and modeling for the RAG system, the system is being tested.
- This step in the process allows for verification, whether the system is functioning and whether there is a need for refinement of the data, data indexing or modeling of the retriever and LLM.

6

Deployment

Deployment and continuous monitoring

- Finally, the RAG system will be deployed in the chosen environment
- In this phase, monitoring processes for continuous improvement should be defined and implemented, to ensure the relevance and adaptation of the data and models according to the project purpose iteratively.

Appendix

Template Prompts

More examples can be found in the drive:

Region	Disaster Type	Disaster Mgmt aspect	Example Prompt
Spain	**Floods**	**Early Warning Message**	[ROLE] I am a disaster response professional working on floods in Spain. [INSTRUCTIONS] Please write an early warning message for residential areas in Andalusia, Spain. [CONTEXT] The message should alert residents of a potential flood and specify communities at risk. [CONSTRAINTS] Please be brief, clear, and give specific actions to take. The message should be both in English and the local language.
Spain	**Earthquakes**	**Risk Assessment**	[ROLE]I am a disaster management professional based in Murcia, Spain. [INSTRUCTIONS] Generate a full risk report for earthquakes in Murcia using the INFORM risk framework from your Knowledge. Your assessment should start with, detailing the prevalence of earthquakes within Murica, the level of human exposure to earthquakes, the vulnerability of Murica if an earthquake occurs, and identifying any lack of coping capacity Murcia may have when addressing earthquake and its aftermath. [CONTEXT] This risk assesment will be utilized as a reference for policy makers and disaster response professionals in Murcia, so regional specifics are encouraged. [CONSTRAINTS] Please do not generate instructions on 'how to write the assessment'. Instead, take those instructions and actually generate the assesment yourself with the data at your disposal, and access online sources if needed. You can make requests for certain types of data that would enhance the overall quality and specificity of the assement at the end of your response.

Appendix

Template Prompts

Region	Disaster Type	Disaster Mgmt aspect	Example Prompt
Spain	**Wildfires**	**Exercise Designs**	[ROLE] I am a disaster response professional working on wildfires in Spain. [INSTRUCTIONS] Please write a full tabletop exercise for a wildfire in Galicia, Spain. [CONTEXT] This exercise will be used in an upcoming simulation with all emergency response stakeholders involved in wildfire management and response.The goal is to simulate a realistic flood event to test and improve their preparedness, response strategies, and inter-agency coordination. Galicia has a history of wildfire, and the exercise should reflect the unique geographical, social, and infrastructure challenges of the area. [CONSTRAINTS] Please be highly specific in the exercise scenario, making it realistic for the area. The exercise should be comprehensive, challenging, and tailored to enhance the particpants' response and management skills. You should include details such as the timeline of events, materials needed. key stakeholders involved, specific challenges faced/interjections, and expected outcomes. Ensure the scenario is culturally and contextually appropriate for Maputo, Mozambique.

Appendix

Glossory

- Disaster Risk Management (DRM):

A systematic process of using administrative directives, organizations, and operational skills to implement strategies, policies, and improved coping capacities to lessen the adverse impacts of hazards and the possibility of disaster.

- Retrieval-Augmented Generation (RAG):

An approach that enhances the capabilities of language models by integrating external data retrieval processes. This allows the model to access and utilize relevant information from large datasets to improve the accuracy and relevance of its responses.

- Large Language Model (LLM):

A type of artificial intelligence (AI) model that is trained on vast amounts of text data to understand and generate human-like text. Examples include models like and GPT-4 and Claude.

- Artificial Intelligence (AI):

The simulation of human intelligence in machines that are programmed to think and learn. AI can be applied in various fields, including natural language processing, robotics, and data analysis.

- Early Warning System (EWS):

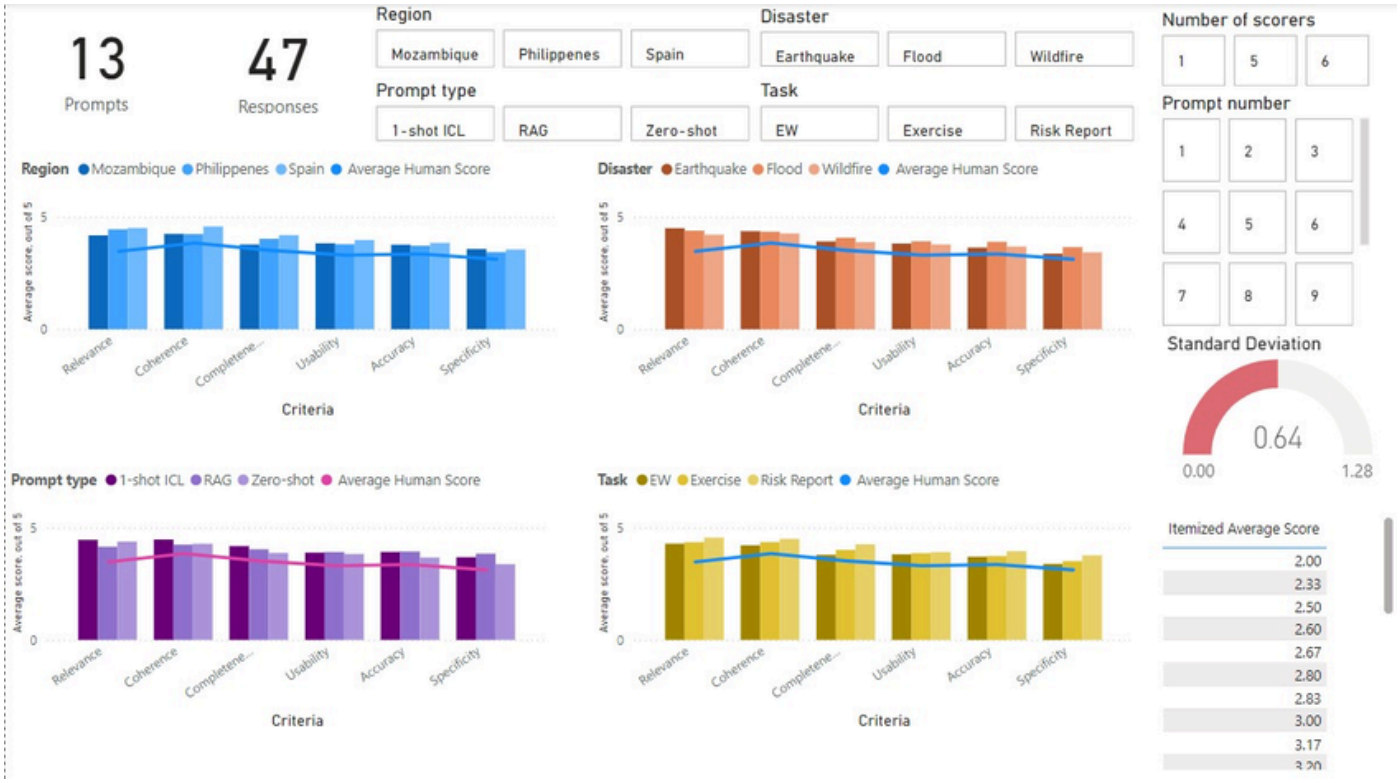
A set of capacities needed to generate and disseminate timely and meaningful warning information to enable individuals, communities, and organizations threatened by a hazard to prepare and act appropriately to reduce the possibility of harm or loss.

- RICCE Framework for Prompt Engineering:

Prompt engineering is the art of crafting queries or commands that guide AI models to produce the most accurate and relevant outputs. RICCE stands for Role, Instructions, Context, Constraints, and Examples. It’s a structured approach to formulating prompts that elicit the best possible responses from AI models.

Appendix

Interactive PowerBI Dashboard: Visualizations from the findings of Project



Thank you