

August 2024

WATERMARK



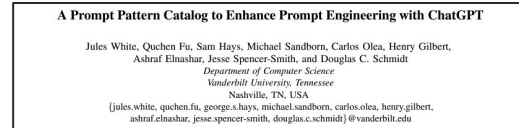
Project Overview: WatermarqAI

Contextual data is vital for informing water management strategies, but it is not easily available to those who need it.

- We aim to demonstrate that the **vast amounts of unstructured textual data** relevant to water management can be productively harnessed through the optimised use of Large Language Models (LLMs).
- In this project, we aim to create two outputs:
 1. A tool creating **narrative outputs** from contextual data, based on user queries on specific water basins, countries and companies;
 2. A framework for the **ranking and scoring** of different contextual factors pertaining to water management, creating a value-based signal from unstructured textual data.

Work completed to date

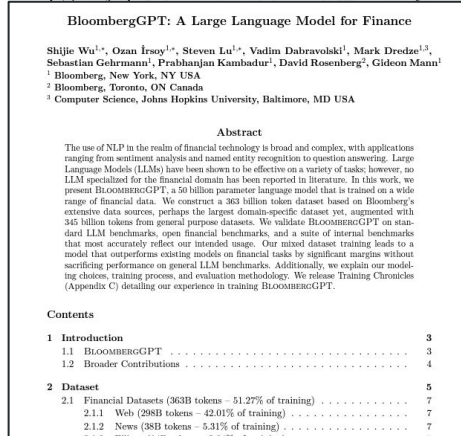
1. Understanding native capacity of LLMs to respond to water security-related queries
2. Research on approaches to achieve domain-specific augmentation of LLMs
3. Identification of contextual factor source documentary databases
4. Defining a universe of contextual factor 'classifiers' that a user may be interested in gaining information on relating to water security
5. Prompt testing and definition of necessary output quality



Abstract—Prompt engineering is an increasingly important skill set needed to converse effectively with large language models (LLMs), such as ChatGPT. Prompts are instructions given to an LLM to enforce rules, automate processes, and ensure specific qualities (and quantities) of generated output. Prompts are also a form of prompt interactions with LLMs. This paper defines unique presented common problems are a knowledge since they provided in a particular of when working on. This paper prompt engineering development tasks, patterns for structure so that they can presents a catalog



Abstract deep learning in NLP allows for impressive results, outperforming traditional methods by large margins (Vainri et al. 2020). Large-based approaches make use of language models (LMs), which are trained on large amounts of unlabeled data. This training on unlabeled data leads to the model learning representations of words and patterns of common language. This language model is called BERT (Bidirectional Encoder Representations from Transformers) and its successors RoBERTa (Liu et al. 2019) and ELECTRA (Kang et al. 2020). These models have been trained on huge amounts of data. However, it was crawled from an unprecedented sources. In the pre-training phase, most LMs are trained on additional data. For the downstream tasks, the models are fine-tuned on the pre-training data. The models are especially large on downstream tasks and the amount of samples is difficult and, thus, the cost of training is high (hundreds or few thousand samples).



Core Findings & Users

The **native capacity** of LLMs to generate high quality narrative outputs detailed enough to inform domain-specific decision making is **limited**.

Evidence suggests that it is possible to **greatly improve output quality through augmentation** of existing generative AI models.

The ability to derive insights from large volumes of contextual information has a **very wide set of use cases**. We narrow our focus to two possible users:



Development finance – providing deep contextual detail to inform investment decisions in areas of key importance to addressing water insecurity



Private sector – providing a rapid understanding of contextual dynamics impacting water use and stewardship decisions in water-scarce basins

Our Value Add

We aim for WatermarqAI to become the go to resource for generating best quality contextual insights relevant to water resource management, by:

1

Source data - making specific corpuses of documentary data available for the derivation of narrative outputs

2

Classifiers - developing optimal prompt engineering and making it relevant to the context of the water sector

3

Retrieval augmentation
Ensuring the correct contextual data is available for specific queries

Our ask

We are seeking to partner with organisations who:

1. See the value in the harnessing of contextual information to address the challenge of water security;
2. Are willing to fund novel AI applications and who believe these tools create new routes to solving some of the greatest environmental and development challenges facing humanity;
3. Want to work with a dynamic team with a track record of addressing water security through innovative models, and with a close relationship with world-leading AI and environmental researchers.

We have established a separate nonprofit (Watermarq Public CIC) to facilitate philanthropic support of this work.