

People's Democratic Republic of Algeria  
Ministry of Higher Education and Scientific Research  
National Polytechnic School of Algiers

## Document-based Question Answering

Hachem Betrouni

National Polytechnic school of Algiers, Industrial Engineering Department, Data Science  
and AI, Algiers, hachem.betrouni@g.enp.edu.dz

2023

<b>President</b>	Mme. Latifa DEBBI	Prof.	ENP, Alger
<b>Examiner</b>	Mr. Salah Eddine TACHI	MCA.	ENP, Alger
<b>Supervisor</b>	Mr. Hakim FOURAR LAIDI	MCA.	ENP, Alger

# Table of Contents

- 1 Introduction
- 2 BIGmama technology
- 3 Document-based question answering
- 4 Implementation and results
- 5 Conclusion

# Introduction

In today's AI consulting market, creating value goes beyond providing standard solutions. The arrival of Large Language Models (LLMs) and other fundamental models raises the level of requirements. To create value, bespoke AI solutions are now required, tailored to specific needs and surpassing basic requirements.

## The requirements

The requirements are no longer to predict a number, but to explain the prediction, it's no longer to recommend a blog title but to generate its whole content.

## Market shift

Today we are well aware that answering individual or organizations needs for AI consulting has shifted from traditional approaches (e.g. written reports) to building tangible assets that concretize the added value. Today, clients are expecting working solutions to their problems in short time.

# Table of Contents

- 1 Introduction
- 2 BIGmama technology**
- 3 Document-based question answering
- 4 Implementation and results
- 5 Conclusion

# BIGmama technology



BIGmama is a French registered startup with an Algerian extension, specialized in data science and AI. They have been developing predictive applications for more than 8 years.



## Mission

BIGmama's mission is to democratize the access to bespoke AI applications, With an original methodology they are able to identify, extract expertise and combine it with machine learning system (Hybrid AI).

## Vision

Data science will soon become a commodity, the arrival of large language models mark but a start of a new wave of foundational models. In the near future, we can expect the emergence of generalist models that are equivalent to ChatGPT but specialized in computer vision tasks, forecasting, pattern recognition, and other domains.

The future is going to become in the hands of those who are capable of harnessing the power of these generalist models, and are capable of combining it with human expertise.

# Unique methodology

## AI starts with *problematization*

Problematization is going from a subject → problems.  
It's an iterative work.

## AI is a tool, not an end-goal

Current AI approaches are not always the best way to solve **every** problem, certain classical approaches are more effective, simpler and more explainable to solve certain problems.

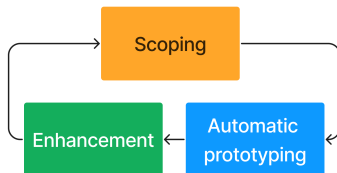
## AI should be hybridized with expertise

By extracting expertise from our clients and combining it with AI systems, we are able to produce tools that are more effective, easier to maintain and that are far less expensive.



# Hyko

BIGmama is currently working on Hyko, a software that intrinsically embeds its mission of democratizing AI and its unique methodology.



**Figure:** Iterative process of developing an AI asset in Hyko.

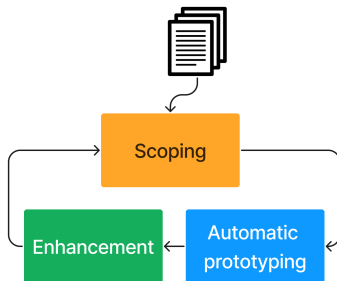
Scoping

Automatic prototyping

Prototype enhancement

# Knowledge representation

A very important aspect of the Scoping phase is the ability to represent already existing knowledge (e.g. technical documents) and make it digestible by Hyko chatbot.



# Knowledge representation

Previous approaches :

- Ontologies.
- Knowledge graphs.

These methods require expertise in the targeted field, and demand careful curation of knowledge, making it an un-scalable option.

With the last advances in language modeling and information retrieval systems, **document-based question answering** become a more appealing option for knowledge representation.

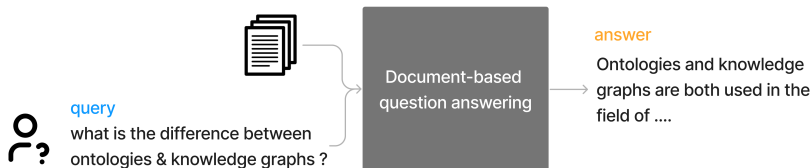
# Table of Contents

- 1 Introduction
- 2 BIGmama technology
- 3 Document-based question answering**
- 4 Implementation and results
- 5 Conclusion

# Document-based question answering

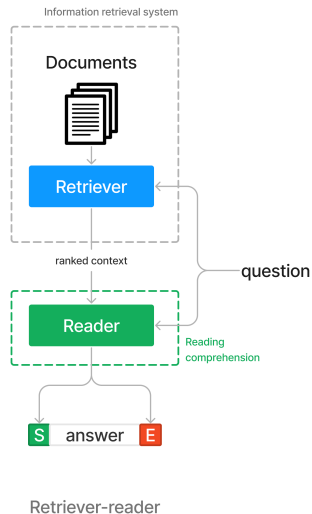
## Definition

It's a machine learning task that focuses on extracting accurate and relevant answers from a given document and respond in natural language to a user query. It involves representing and manipulating content within textual documents, such as technical documentation in order to provide precise answers to user's queries.



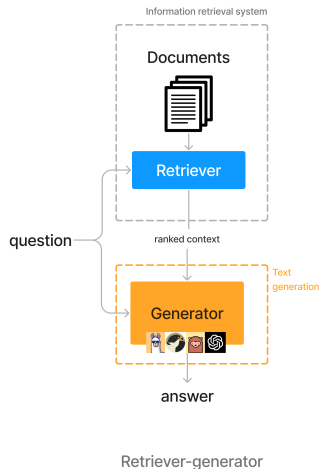
# Retriever-reader

- 1 finding the related context in the documentation.
- 2 processes the retrieved context to **extract** the start/end positions of an answer.



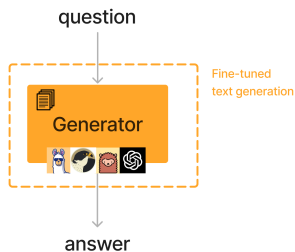
# Retriever-generator

Unlike the reader model, the generator model generates free text conditioned with the retrieved context to answer the question.





# Closed-book QA



Closed-book QA

With a substantial number of parameters, Large language models possess the ability to memorize factual information within their weight parameters. Consequently, one is able to employ this property for question-answering tasks without relying on explicit context.

Retriever reader models require structured data, it doesn't take advantage of unsupervised training (as LLM do).

Closed-book approach require substantial amount of computation and time to finetune the model on new documentation.

Retriever-generator approach seems promising for our use-case.

# Table of Contents

- 1 Introduction
- 2 BIGmama technology
- 3 Document-based question answering
- 4 Implementation and results**
- 5 Conclusion

# Table of Contents

- 1 Introduction
- 2 BIGmama technology
- 3 Document-based question answering
- 4 Implementation and results
- 5 Conclusion**