

Give me an idea
Improving Retrieval System in RAG Architecture to
Expand LLM Context |
Applied Data Science Project 2024 |

Project name*


ADSP – P9 – RAG MARCO |

Continue

OR CONTACT

 Homayoun Afshari

 Arash Daneshvar

 Hossein Khodadadi



What can I help with?



Message ChatGT3



Create Image



Code



Summarize



Get advice

More





What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



Hey! Would you please explain the basics of the project explained in this file? |



Create Image



Code



Summarize



Get advice

More





Today

Project definition



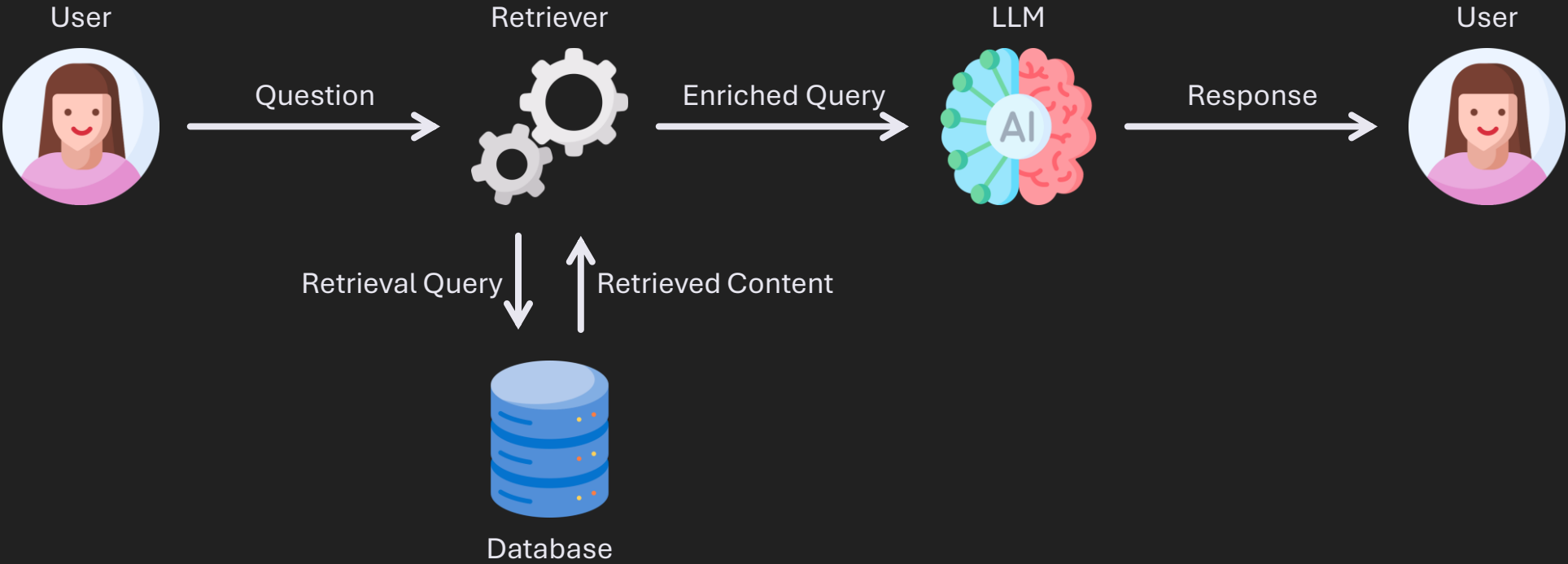
ADSP - P9 - RAG MARCO.pdf
PDF

Hey! Would you please explain the basics of the project explained in this file?



Well, for that question, as mentioned in the provided PDF file, we can use RAGs!

RAG, or Retrieval-Augmented Generation, is a system that combines retrieving relevant information from a database with generating responses using an LLM to provide accurate, context-rich answers.



Message ChatGT3





Today

Project definition



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



Please describe the objectives of the project explained in this file. |



Create Image



Code



Summarize



Get advice

More





Today

Project definition



Project objectives



ADSP - P9 - RAG MARCO.pdf
PDF

Please describe the objectives of the project explained in this file.



Let's dive into optimizing information quality in a RAG system, with a spotlight on retrieval and response generation.

When it comes down to it, the quality of information is everything.

Based on insights from the PDF, here's a breakdown of the main objectives:

- **Objective 1:** Enhance the Retrieval Process Using LLMs
- **Objective 2:** Elevate the Quality of Context for the Response Generator

And here's a potential **research question** to dig deeper:

“How optimizing the retrieval stage in a RAG system ensures high-quality and contextually accurate responses?”

Also, for this project, the **value proposition** could be:

“Boosting Accuracy through Advanced Retrieval”

This approach keeps the focus sharp - enhancing accurate retrieval and delivering relevant.



Message ChatGT3





ChatGT3



Explore GT3s

Today

Project definition



Project objectives



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



How can our project contribute to a sustainable and equitable future? |



Create Image



Code



Summarize



Get advice

More



Upgrade plan





Today

Project definition



Project objectives



Project relevance to SDGs



ADSP - P9 - RAG MARCO.pdf
PDF

How can our project contribute to a sustainable and equitable future??



Based on the provided reference and the alignment with the united nations Sustainable Development Goals (SDGs), the project could relate to the following:

- **SDG 4: Quality Education:** The project improves information access, supporting quality education through enhanced knowledge retrieval.
- **SDG 9: Industry, Innovation, and Infrastructure:** By advancing retrieval technology, the project promotes innovation and strengthens information infrastructure.

These connections underscore how improved knowledge retrieval can facilitate educational access and technological progress, aligning with the goals of equitable knowledge distribution and enhanced data infrastructure.

4 QUALITY
EDUCATION



9 INDUSTRY, INNOVATION
AND INFRASTRUCTURE



Message ChatGT3





ChatGT3



Explore GT3s

Today

Project definition



Project objectives



Project relevance to SDGs



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



Who are we designing this experience for? |



Create Image



Code



Summarize



Get advice

More



Upgrade plan





Today

Project definition



Project objectives



Project relevance to SDGs



Design: user personas



ADSP - P9 - RAG MARCO.pdf
PDF

Who are we designing this experience for?



Below, I will suggest several persona ideas based on common use cases for RAG systems:

- **Emma Carter**, data analyst, 28 years old, works at a research institute in the EU.
 - Behaviors and Habits: Uses search engines and databases for literature reviews and trend analyses.
 - Needs and Goals: Needs reliable summaries and relevant insights to compile reports efficiently.
- **Michael Foster**, AI software manager, 38 years old, works at a startup in San Francisco.
 - Behaviors and Habits: Keeps up with the latest AI research to drive innovation.
 - Needs and Goals: Needs efficient, scalable, and reliable AI systems to be integrated into his existing software development projects.



Message ChatGT3





ChatGT3



Explore GT3s

Today

Project definition



Project objectives



Project relevance to SDGs



Design: user personas



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



What steps do users take from curiosity to complete satisfaction? |



Create Image



Code



Summarize



Get advice

More



Upgrade plan





Today

- Project definition ...
- Project objectives ...
- Project relevance to SDGs ...
- Design: user personas ...
- Design: customer journey ...

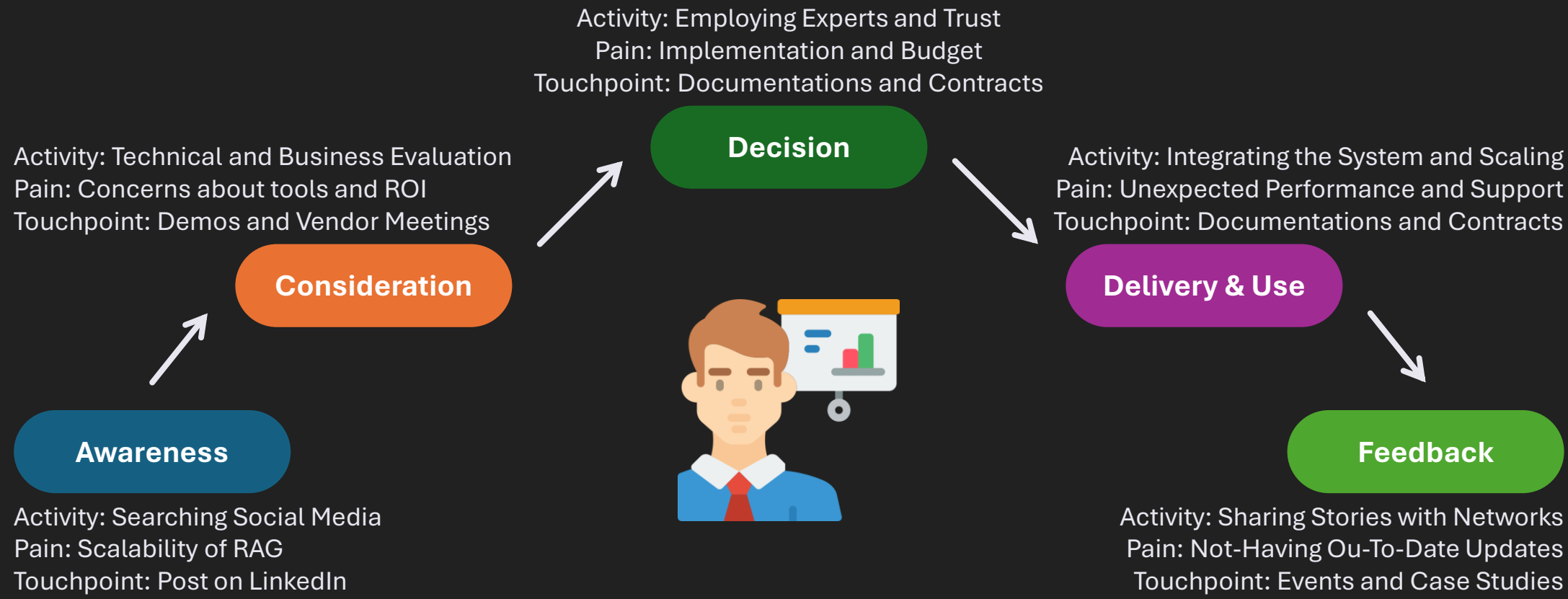


ADSP - P9 - RAG MARCO.pdf
PDF

What steps do users take from curiosity to complete satisfaction?



Let's focus on Michael Foster. Here's his journey:



Message ChatGT3





ChatGT3



Explore GT3s

Today

Project definition



Project objectives



Project relevance to SDGs



Design: user personas



Design: customer journey



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



How do all the pieces of our system work together? |



Create Image



Code



Summarize



Get advice

More



Upgrade plan





Today

- Project definition
- Project objectives
- Project relevance to SDGs
- Design: user personas
- Design: customer journey
- Design: functional diagram



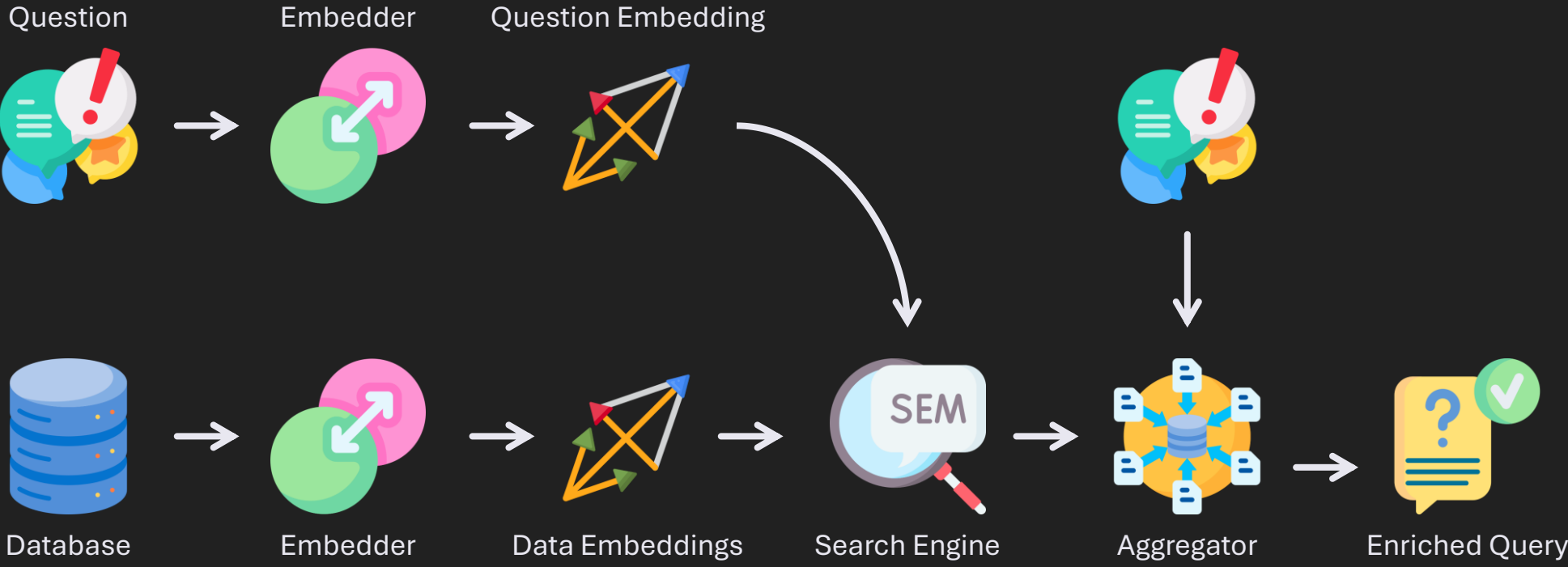
ADSP - P9 - RAG MARCO.pdf
PDF

How do all the pieces of our system work together?



In a RAG system, the query and retrieved context work together: relevant info is pulled from the database, combined with the query, and then transformed by the LLM into a precise, context-rich answer!

This project incorporates the following **functional diagram**:



Message ChatGT3





Today

Project definition



Project objectives



Project relevance to SDGs



Design: user personas



Design: customer journey



Design: functional diagram



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



How can we break down this project into clear, manageable stages and tasks? |



Create Image



Code



Summarize



Get advice

More





Today

- Project definition
- Project objectives
- Project relevance to SDGs
- Design: user personas
- Design: customer journey
- Design: functional diagram
- Manage: WBS



ADSP - P9 - RAG MARCO.pdf
PDF

How can we break down this project into clear, manageable stages and tasks?



By breaking the project into focused stages and clear tasks!

For instance, we can tackle each piece step-by-step, making complex goals achievable and keeping progress on track!

The following table describes the **WBS** considered for this project:

WP No.	WP Title	Lead Name	PW	Start Week	End Week
1	Phase 1 - Design and Manage	Arash	2	1	3
2	Phase 2 - Develop	Hossein	3	3	6
3	Phase 3 - Dry Run	Homayoun	6	6	11
4	Phase 4 - Final Presentation and Exam	Arash	3	11	14
Total	-	-	14	-	-



Message ChatGT3





ChatGT3



Explore GT3s

Today

Project definition



Project objectives



Project relevance to SDGs



Design: user personas



Design: customer journey



Design: functional diagram



Manage: WBS



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



What's our timeline for turning this vision into reality? |



Create Image



Code



Summarize



Get advice

More



Upgrade plan





Today

- Project definition ...
- Project objectives ...
- Project relevance to SDGs ...
- Design: user personas ...
- Design: customer journey ...
- Design: functional diagram ...
- Manage: WBS ...
- Manage: Gantt ...



ADSP - P9 - RAG MARCO.pdf
PDF

What's our timeline for turning this vision into reality?



Based on the file you provided, you could use a **Gantt chart**!

This tool can break down our ambitious vision into manageable, action-oriented steps, helping to align your team, set milestones, and track progress dynamically.

By laying out each phase of your project, from ideation and resource allocation to development and launch, you can visualize the entire journey from inception to completion. Here's an overview of a **Gantt chart** suitable for this project:

WP Title	October	November	December	January
Phase 1 - Design and Manage	28	14		
Phase 2 - Develop		15	5	
Phase 3 - Dry Run			6	9
Phase 4 - Final Presentation and Exam				10



Message ChatGT3





Today

- Project definition ...
- Project objectives ...
- Project relevance to SDGs ...
- Design: user personas ...
- Design: customer journey ...
- Design: functional diagram ...
- Manage: WBS ...
- Manage: Gantt ...



ADSP - P9 - RAG MARCO.pdf
PDF

What's our timeline for turning this vision into reality?



For Phase 1 (Design and Manage), tasks are as follows:

Task Name (Phase 1 - Design and Manage)	October	November	December	January
Project Understanding	28	3		
Design	30	7		
Manage	4	14		
Initializing Presentation #1	4	6		
Meeting with Mentor		7		
Developing Presentation #1		6	14	
Checkpoint #1		14	14	



Message ChatGT3





Today

- Project definition ...
- Project objectives ...
- Project relevance to SDGs ...
- Design: user personas ...
- Design: customer journey ...
- Design: functional diagram ...
- Manage: WBS ...
- Manage: Gantt ...



ADSP - P9 - RAG MARCO.pdf
PDF

What's our timeline for turning this vision into reality?



For Phase 2 (Develop), tasks are as follows:

Task Name (Phase 2 - Develop)	October	November	December	January
Reading Related Resources		15 21		
Hypothesis and Research Question		16 21		
Meeting with the Mentor		21 21		
Building a Baseline		22 27		
Brainstorming New Ideas		21 22		
Developing Presentation #2		2 5		
Checkpoint #2		5 5		



Message ChatGT3





Today

- Project definition
- Project objectives
- Project relevance to SDGs
- Design: user personas
- Design: customer journey
- Design: functional diagram
- Manage: WBS
- Manage: Gantt



ADSP - P9 - RAG MARCO.pdf
PDF

What's our timeline for turning this vision into reality?



For Phase 3 (Dry Run), tasks are as follows:

Task Name (Phase 3 - Dry Run)	October	November	December	January
Literature Review and Strategy Development			611	
Database Restructuring			1316	
Experimenting New Ideas			1723	
Meeting with Mentor			1919	
MVP Production			198	
Developing Presentation #2			48	
Checkpoint #3				99



Message ChatGT3





Today

- Project definition
- Project objectives
- Project relevance to SDGs
- Design: user personas
- Design: customer journey
- Design: functional diagram
- Manage: WBS
- Manage: Gantt



ADSP - P9 - RAG MARCO.pdf
PDF

What’s our timeline for turning this vision into reality?



For Phase 4 (Final Presentation and Exam), you could bring all your hard work to fruition by preparing a compelling final presentation that highlights the journey and impact of your project. This is your chance to showcase the results, share insights, and demonstrate the full scope of what you’ve accomplished. With each detail polished and every achievement highlighted, this phase will set the stage for a successful presentation. So, the tasks are as follows:

Task Name (Phase 4 - Final Presentation and Exam)	October	November	December	January
Applications and Use Cases				1017
Writing Report				1319
Developing the Final Presentation				2026
Developing the Final Product				10
Exam				31



Message ChatGT3





ChatGT3



Explore GT3s

Today

Project definition



Project objectives



Project relevance to SDGs



Design: user personas



Design: customer journey



Design: functional diagram



Manage: WBS



Manage: Gantt



What can I help with?



ADSP - P9 - RAG MARCO.pdf
PDF



What are the other steps needed to be taken in the future? |



Create Image



Code



Summarize



Get advice

More



Upgrade plan





Today

Project definition



Project objectives



Project relevance to SDGs



Design: user personas



Design: customer journey



Design: functional diagram



Manage: WBS



Manage: Gantt



Conclusion



ADSP - P9 - RAG MARCO.pdf
PDF

What are the other steps needed to be taken in the future?



The expected outcome of this project is a high-performance RAG retrieval system that

- **Significantly enhances retrieval quality** by incorporating advanced processing techniques (e.g., summarization, NER, and topic extraction) and optimized scoring mechanisms.
- **Fulfills LINKS' objective of advancing RAG capabilities for real-world applications** through providing more contextually accurate and relevant responses from the RAG system.


To achieve these outcomes, we will


- Execute the tasks outlined in our **WBS** and timeline, as represented in the **Gantt chart**.
- Proceed in phases, beginning with focused **development** efforts to build the system's core functionalities.
- Conduct a thorough **dry run** to assess performance and make iterative improvements.
- Produce an **MVP** that embodies the core capabilities of the enhanced RAG system, positioning us to meet both project objectives and LINKS' real-world application goals.












Message ChatGT3



 ChatGT3

 Explore GT3s

Today

- Project definition 
- Project objectives 
- Project relevance to SDGs 
- Design: user personas 
- Design: customer journey 
- Design: functional diagram 
- Manage: WBS 
- Manage: Gantt 
- Conclusion 



ADSP - P9 - RAG MARCO.pdf

PDF

What are the other steps needed to be taken in the future?



The expected outcome of this project is a high-performance RAG retrieval system that

- **Significantly enhances retrieval quality** by incorporating advanced processing techniques (e.g., summarization, NER, and topic extraction) and optimized scoring mechanisms.
- **Fulfills LINKS’ objective of advancing RAG capabilities for real-world applications** through providing more contextually accurate and relevant responses from the RAG system.

To achieve these outcomes, we will

- Execute the tasks outlined in our **WBS** and timeline, as represented in the **Gantt chart**.
- Proceed in phases, beginning with focused **development** efforts to build the system’s core functionalities.
- Conduct a thorough **dry run** to assess performance and make iterative improvements.
- Produce an **MVP** that embodies the core capabilities of the enhanced RAG system, positioning us to meet both

You’ve hit the Free plan limit for GT3.

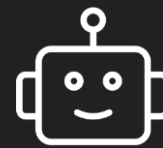
You need GT3-4o to continue this chat because there’s an attachment.
Your limit resets after December 5th.

Get Plus



Message ChatGT3





Oops!

Thank you for your attention, but our systems are busy at the moment.
Please take a break and ask your questions.



Homayoun Afshari



Arash Daneshvar



Hossein Khodadadi