

Master thesis
Toward Optimizing
a Retrieval
Augmented
Generation Pipeline
using Large
Language Model

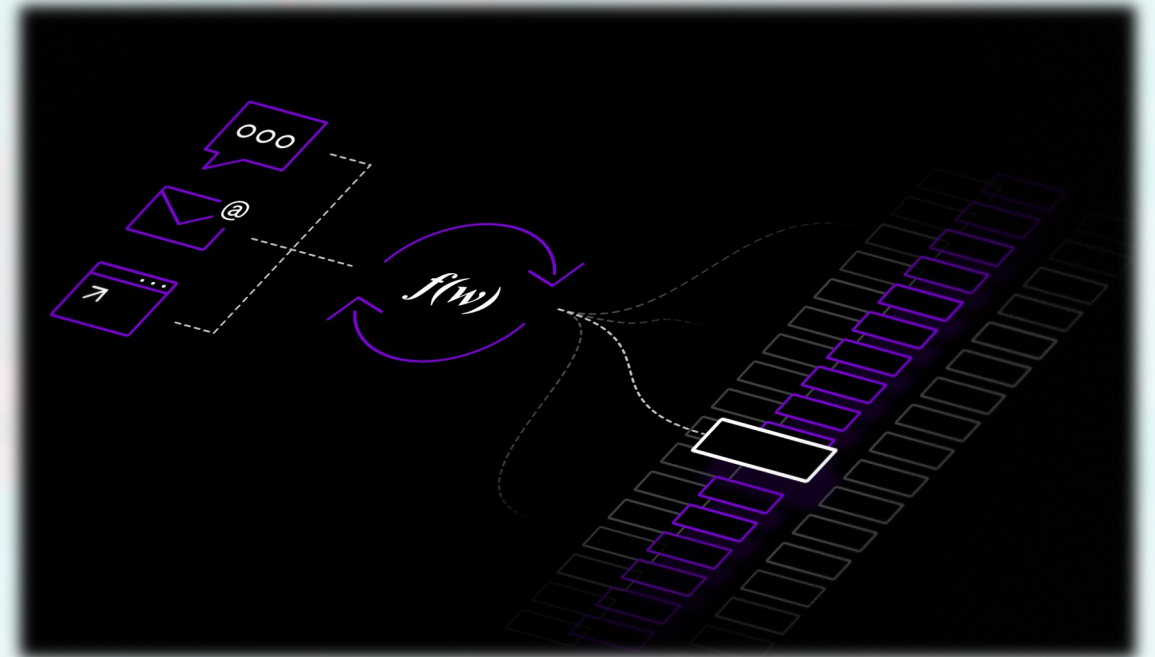
Student: Gentrit Fazlija

Supervisor: Anum Afzal

Professor: Prof. Dr. Florian Matthes

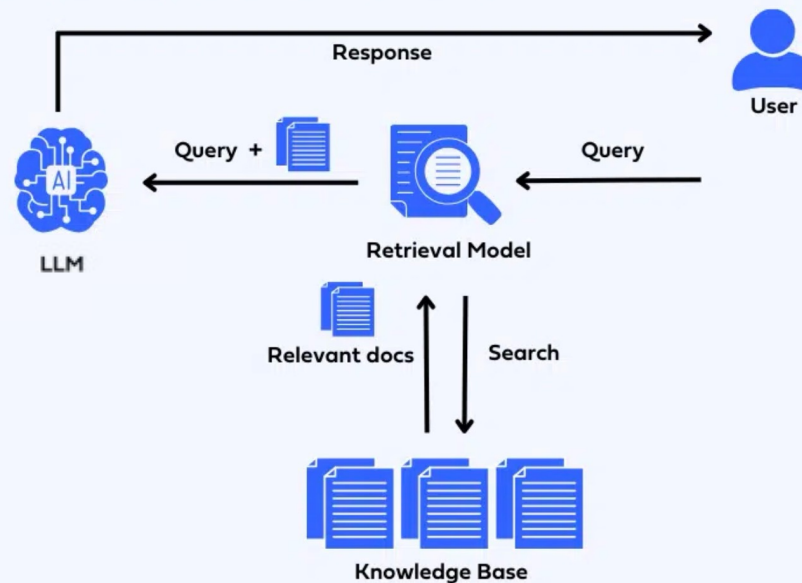
Table of content

- i. Key Components & Motivation
- ii. Knowledge Base
- iii. Research Question
 - i. Challenge
 - ii. Solution Idea
- iv. Testing & Evaluation
- v. Outlook



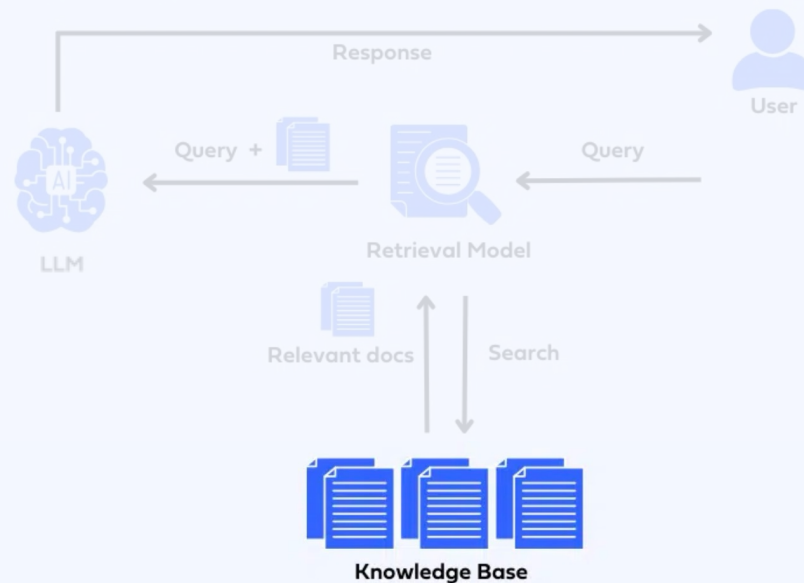
Key Components & Motivation

Retrieval Augmented Generation



Key Components & Motivation

Retrieval Augmented Generation



Knowledge Base

Technical University of Munich

NEWS AND EVENTS STUDIES LIFELONG LEARNING RESEARCH INNOVATION COMMUNITY ABOUT TUM DE | EN

Whether you are pursuing an academic degree, want to take your expertise to a new level, or are looking for the right skills to put your entrepreneurial ideas into practice – TUM is your partner for lifelong learning. Discover our range of courses:

Select degree program or enter keyword

▼ Narrow results

168 results found

A

Bachelor of Science (B.Sc.)
Aerospace

Start your personal "Mission Earth" with the Bachelor's Program Aerospace. This study degree program of 6 semesters (3 years) is fully taught in English and is the ideal entry to the global topic of aeronautics and astronautics.

[read more](#)

Master of Science (M.Sc.)
Aerospace

Cleared for take off! It is your goal to design aircraft which fly faster, higher and further? Do you want to sound out what is technically possible in space travel? In our aerospace degree course you will learn everything you need to know to make your fascination and passion to your profession.

[read more](#)

Master of Science (M.Sc.)
Aerospace Engineering

Awarded by TUM, the program is conducted in Singapore and serves to provide graduates with an in-depth knowledge in the field of aerospace engineering, focusing in the areas of aeronautical design, space design and research.

[read more](#)

Master of Science (M.Sc.)
Agricultural Biosciences

KEY SKILL PROGRAMS

Additional, extracurricular courses for students

STUDENT ADVISING AND INFORMATION SERVICES

+ 49 89 289 22245
studium@tum.de

Please observe the e-mail etiquette.

Personal advising sessions with General Student Advising by appointment

SERVICE DESK

Campus Munich
Arcisstraße 21, Room 0144
80333 München

Monday, 9 a.m. – 12 p.m.
Wednesday, 9 a.m. – 12 p.m.
Friday, 9 a.m. – 12 p.m.

OTHER COURSES AND PROGRAMS

- FAQs
- Orientation
- Part-Time Degrees
- Doctoral studies
- Continuing education
- Additional qualifications

<https://www.tum.de/en/studies/degree-programs>

Knowledge Base

What does this program cover?

How is the program structured?

How do I enroll?

How long does the master's thesis take?

Master of Science (M.Sc.)
Mathematics in Data Science

The master's degree program Mathematics in Data Science combines a high-profile education in mathematics with an emphasis on the burgeoning area of Big Data.

[TUM School of Computation, Information and Technology](#)

Key Data

Type of Study Full Time	Standard Duration of Studies 4 (fulltime)	Credits 120 ECTS
Main Locations Garching	Application Period Winter semester: 01.01. – 31.05. Summer semester: 01.09. – 30.11.	Admission Category Aptitude Assessment for Master
Start of Degree Program Possible for both winter and summer semester	Costs Student Fees: 85.00 €	Required Language Proficiency English

Information on Degree Program

- What does this program cover? +
- How is the program structured? +
- What is the language of instruction? +
- Which further expertise and skills will I acquire? +

Binding Regulations for Progression of Studies, Examinations and Application

Knowledge Base

What does this program cover?

How is the program structured?

Master Mathematics in Data Science

[Beratung](#) ▾

[Das Wichtigste zum Masterstudium](#) ▾

[Voraussetzungen für den Master Mathematik](#) ▾

[Bewerbung für den Master](#) ▾

[Das Wichtigste zur Masterarbeit](#) ▾

Sie jonglieren gerne mit Daten wie sie etwa in den sozialen Medien generiert werden? Sie finden Verfahren zur Datenerfassung, Datenaufbereitung und Datenanalyse richtig spannend und wollen diese bei komplexen Daten in ökologischen Systemen anwenden? Das Masterstudium "Mathematics in Data Science" an der Technischen Universität München (TUM) macht Sie fit für den zukunftsträchtigen Arbeitsbereich Big Data. Als Daten-Experte sind Sie eine gefragte Person in Forschung und Entwicklung, in der Finanzindustrie, Biotechnologie und Logistik, im Gesundheitswesen, bei Versicherungen und für IT-Sicherheit.

Fachstudienberater



PD Peter Massopust,
Ph.D.
[mscapp_datascience](mailto:mscapp_datascience@ma.tum.de)
[\(at\) ma.tum.de](mailto:m(at)ma.tum.de)

Studienschwerpunkte

Im Masterprogramm "Mathematics in Data Science" stehen die Bereiche Data Engineering, Data Analytics, Data Analysis, Machine Learning und Data Science im Mittelpunkt. Studierende konzentrieren sich dabei auf Techniken der Datenhaltung und -auswertung. Sie lernen, diese an konkrete Problemstellungen anzupassen, sie zu kombinieren oder neu zu entwickeln und daraus Vorhersage- und Klassifikationsmodelle abzuleiten. Daneben beschäftigen sie sich auch mit der Weiterentwicklung von Algorithmen zur Problemlösung. Überfachliche Lehrveranstaltungen, die sich mit gesellschaftlichen und politischen Implikationen von Big Data beschäftigen sowie juristisches Grundwissen und Fremdsprachenkenntnisse vermitteln, sind ebenso wichtige Bestandteile des Masterprogramms.

Nicht vergessen!

Hier finden Sie wichtige Termine, Deadlines und andere relevante Hinweise.



Bewerbungsfristen

Wer mit dem Masterstudium beginnen will, sollte sich möglichst frühzeitig bewerben. Alle notwendigen Unterlagen müssen Sie innerhalb der Fristen in Ihrem Bewerberaccount im [Online-System der TUM](#) vollständig hochladen.

Bewerbungsfrist für das Wintersemester: 1. Januar bis 31. Mai

Bewerbungsfrist für das Sommersemester: 1. September bis 30. November

How do I enroll?

How long does the master's thesis take?

Knowledge Base

What does this program cover?

How is the program structured?

1

Fachprüfungs- und Studienordnung für den Masterstudiengang Mathematics in Data Science an der Technischen Universität München

Vom 17. August 2023

Aufgrund von Art. 9 Satz 1 und 2 in Verbindung mit Art. 80 Abs. 1 Satz 1, Art. 84 Abs. 2 Satz 1 sowie Art. 90 Abs. 1 Satz 2 des Bayerischen Hochschulinnovationsgesetzes (BayHIG) erlässt die Technische Universität München folgende Satzung:

Inhaltsverzeichnis:

- § 34 Geltungsbereich, akademischer Grad
- § 35 Studienbeginn, Regelstudienzeit, ECTS
- § 36 Qualifikationsvoraussetzungen
- § 37 Modularisierung, Modulprüfung, Lehrveranstaltungen, Unterrichtssprache
- § 37 a Berufspraktikum, Projekt, Auslandsaufenthalt
- § 38 Prüfungsfristen, Studienfortschrittskontrolle, Fristversäumnis
- § 39 Prüfungsausschuss
- § 40 Anrechnung von Studienzeiten, Studien- und Prüfungsleistungen
- § 41 Studienbegleitendes Prüfungsverfahren, Prüfungsformen
- § 42 Zulassung und Anmeldung zur Masterprüfung
- § 43 Umfang der Masterprüfung
- § 44 Wiederholung, Nichtbestehen von Prüfungen
- § 45 Studienleistungen
- § 45 a Multiple-Choice-Verfahren
- § 46 Master's Thesis
- § 47 Bestehen und Bewertung der Masterprüfung
- § 48 Zeugnis, Urkunde, Diploma Supplement

How do I enroll?

How long does the master's thesis take?

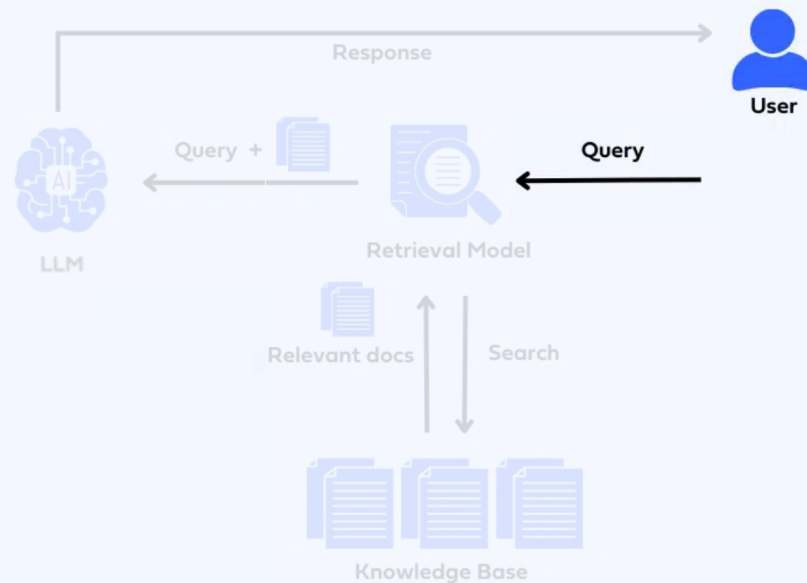
Solution

I want to create a state-of-the-art information retrieval model, capable of querying TUM course data. This will help existing students understand their own degree program better and at the same time support applicants.

TUM GPT

Key Components & Motivation

Retrieval Augmented Generation



Challenge & Solution

Challenge:

Students do not formulate the questions exactly



Students ask the same question in different manners



They might not even know exactly what they are looking for

Solution:



Challenge & Solution

Challenge:

Students do not formulate the questions exactly



Students ask the same question in different manners



They might not even know exactly what they are looking for

Solution:



Research Question

RQ 1: Would a multi-query formulation system improve the performance?

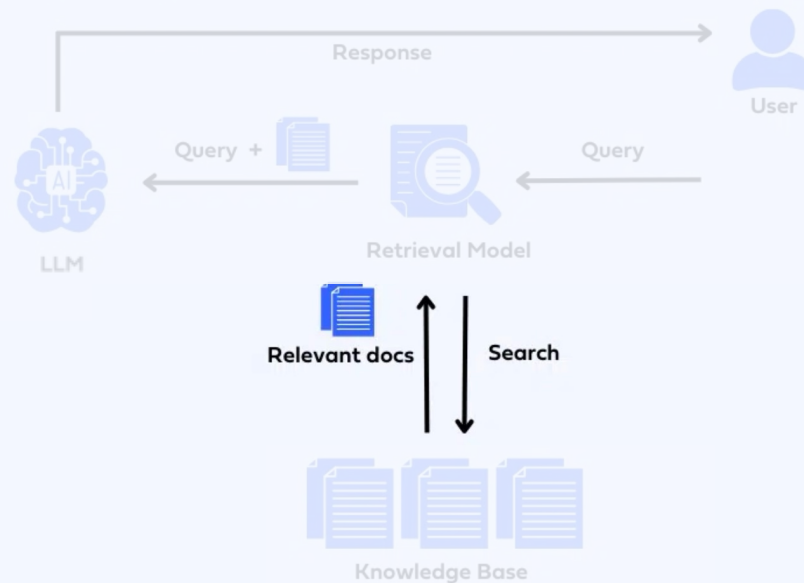
RQ 2: Would an optimization approaches, such as ensemble retriever in combination with a child-parent chunking improve the performance of the passage retriever?

RQ 3: Would few-shot learning enhance the performance of the system as compared to the Zero-Shot of the system?

RQ 4: How does the performance change when using an open-source model compared to a paid closed source model? How can open-sourced models be optimized?

Key Components & Motivation

Retrieval Augmented Generation



Challenge & Solution

Challenge:

Small Chunks: Does not capture the whole meaning



Big Chunks: Has difficulty with similarity search

Solution:

1

Child-Chunk

Feld, Feldübungen etc. mit dem Ziel der Durchführung, Auswertung und Erkenntnisgewinnung. ²Bestandteil können z. B. sein: die Beschreibung der Vorgänge und die jeweiligen theoretischen Grundlagen inkl. Literaturstudium, die Vorbereitung und praktische Durchführung, ggf. notwendige Berechnungen, ihre Dokumentation und Auswertung sowie die Deutung der Ergebnisse hinsichtlich der zu erarbeitenden



Parent-Chunk

b) ¹Eine **Laborleistung** beinhaltet je nach Fachdisziplin Versuche, Messungen, Arbeiten im Feld, Feldübungen etc. mit dem Ziel der Durchführung, Auswertung und Erkenntnisgewinnung. ²Bestandteil können z. B. sein: die Beschreibung der Vorgänge und die jeweiligen theoretischen Grundlagen inkl. Literaturstudium, die Vorbereitung und praktische Durchführung, ggf. notwendige Berechnungen, ihre Dokumentation und Auswertung sowie die Deutung der Ergebnisse hinsichtlich der zu erarbeitenden Erkenntnisse. ³Die Laborleistung kann durch eine Präsentation ergänzt werden, um die kommunikative Kompetenz bei der Darstellung von wissenschaftlichen Themen vor einer Zuhörerschaft zu überprüfen.

Challenge & Solution

Challenge:

Data is very similar in between
different study programs

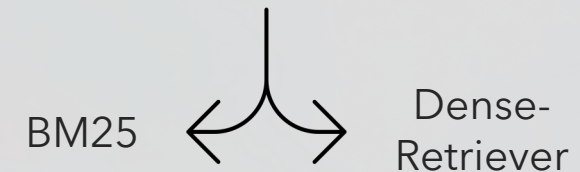
+

Data within one study program
has specific words being specific
things

Solution:

2

Retrieval-System



Research Question

RQ 1: Would a multi-query formulation system improve the performance?

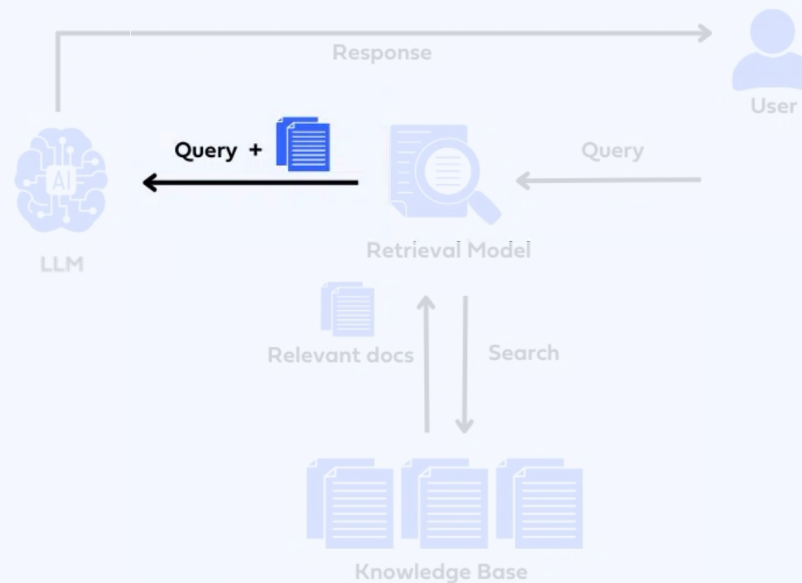
RQ 2: Would an optimization approaches, such as ensemble retriever in combination with a child-parent chunking improve the performance of the passage retriever?

RQ 3: Would few-shot learning enhance the performance of the system as compared to the Zero-Shot of the system?

RQ 4: How does the performance change when using an open-source model compared to a paid closed source model? How can open-sourced models be optimized?

Key Components & Motivation

Retrieval Augmented Generation



Challenge & Solution

Challenge:

LLMs tend to hallucinate

+

LLMs tend to give highly divergent outputs

Solution:

"What is ..."



ICL

+

"What is ..."

Research Question

RQ 1: Would a multi-query formulation system improve the performance?

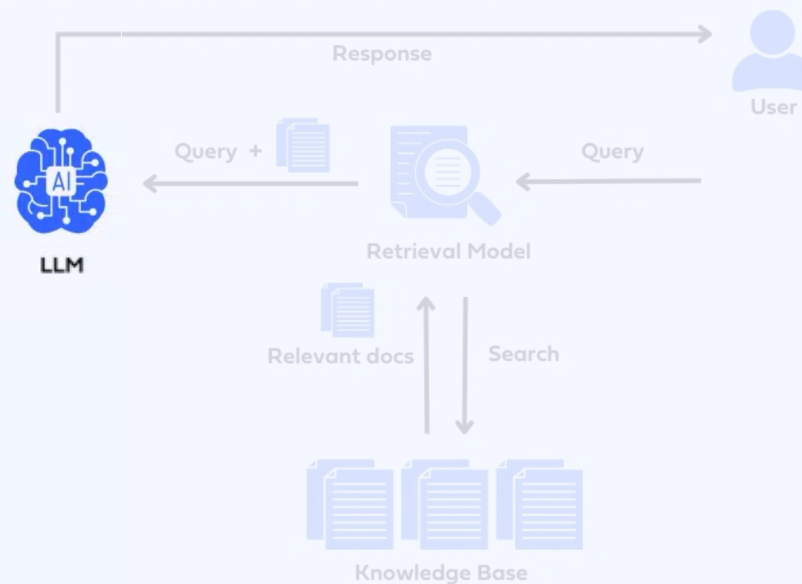
RQ 2: Would an optimization approaches, such as ensemble retriever in combination with a child-parent chunking improve the performance of the passage retriever?

RQ 3: Would few-shot learning enhance the performance of the system as compared to the Zero-Shot of the system?

RQ 4: How does the performance change when using an open-source model compared to a paid closed source model? How can open-sourced models be optimized?

Key Components & Motivation

Retrieval Augmented Generation



Challenge & Solution

Challenge:

Data is valuable

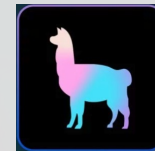


API call costs

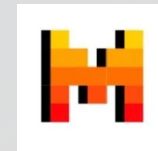


Sensitive Data should not go to a third party

Solution:



Llama 2



Mistral



Hugging Face LLM

Research Question

RQ 1: Would a multi-query formulation system improve the performance?

RQ 2: Would an optimization approaches, such as ensemble retriever in combination with a child-parent chunking improve the performance of the passage retriever?

RQ 3: Would few-shot learning enhance the performance of the system as compared to the Zero-Shot of the system?

RQ 4: How does the performance change when using an open-source model compared to a paid closed source model? How can open-sourced models be optimized?

Testing & Evaluation

Binary Classification:

"How much time do I have for the master thesis?"

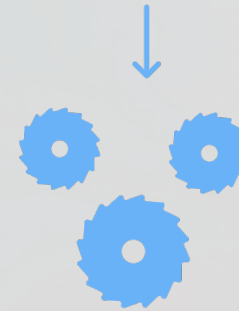
"When is the application deadline for Data Science?"

"Do I need a Proof of English language skills?"

Human Evaluation:

1

"Tell me about the Conditions I have to fulfill when I want to study <x>?"

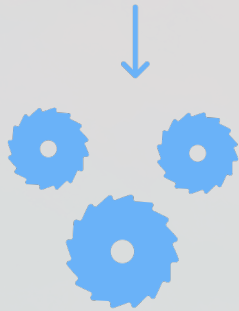


Testing & Evaluation

Human Evaluation:

1

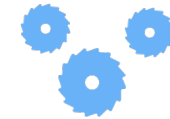
"Tell me about the Conditions I have to fulfill when I want to study <x>?"



Self-Evaluation:

2

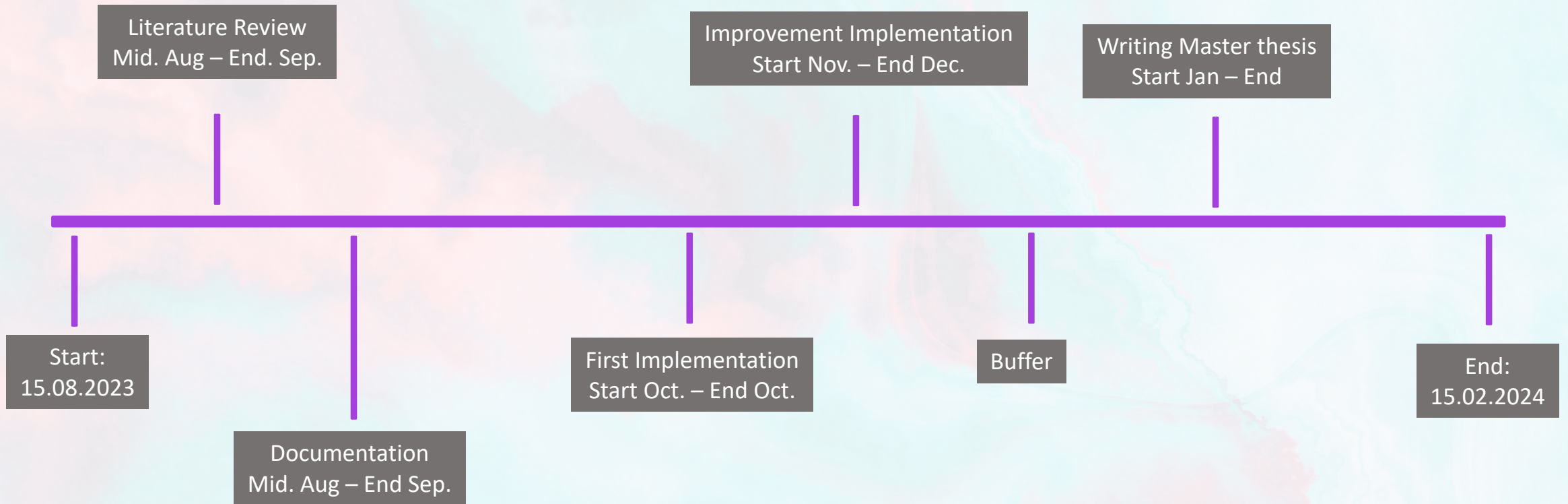
" If we want to study <x> we have the following conditions:



Is that right?"



Outlook



<Thanks for the attention>



Gentrit Fazlija

MSc Student Mathematics in Data Science

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.
gentrit.fazlija@tum.de
www.matthes.in.tum.de

