UU Guest Lecture
LLMs
Sep 9, 2024

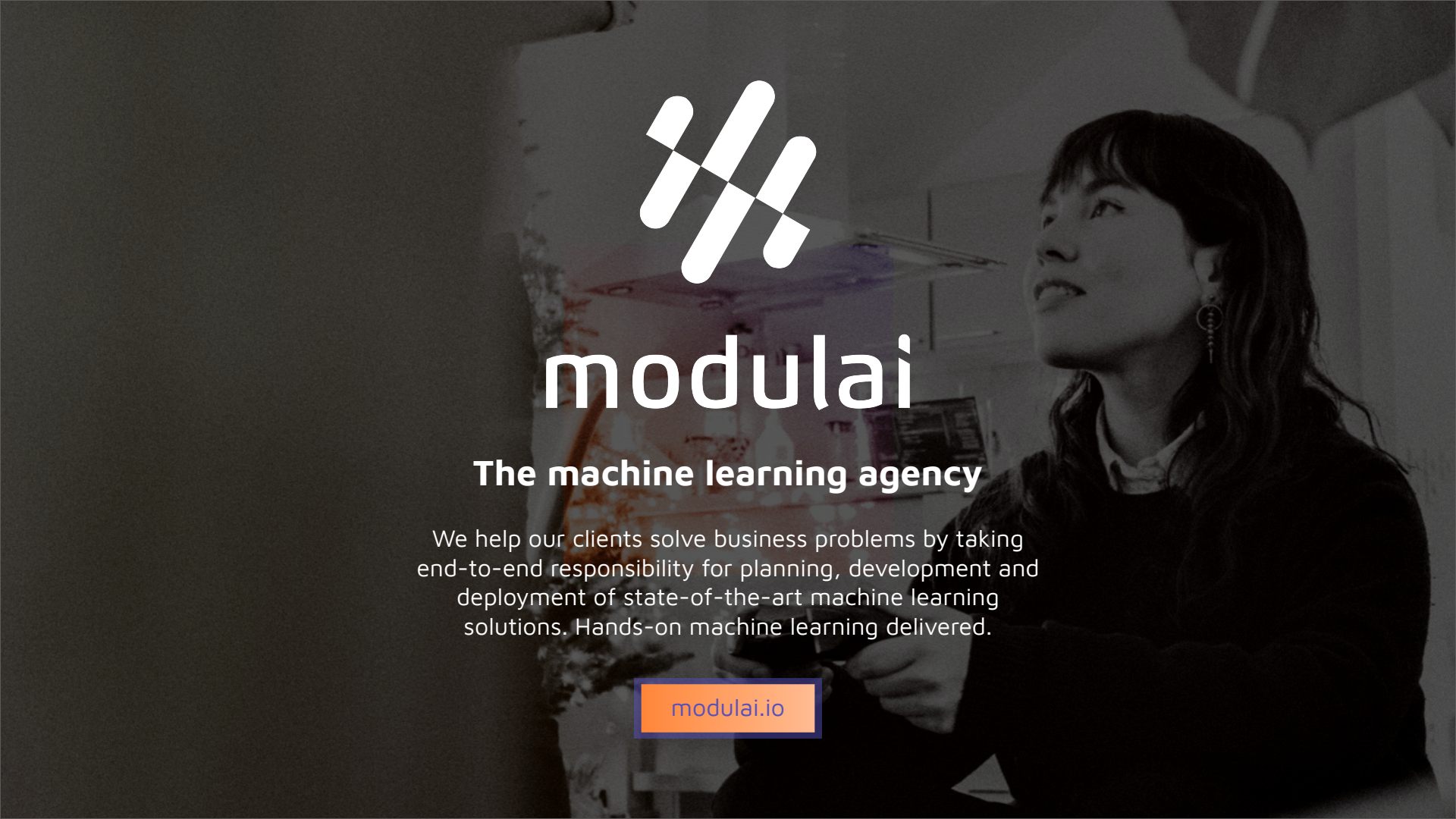
modulai

Agenda

- About me and Modulai
- Use case: NL2SQL
- Use case: RAG
- A few important basic concepts

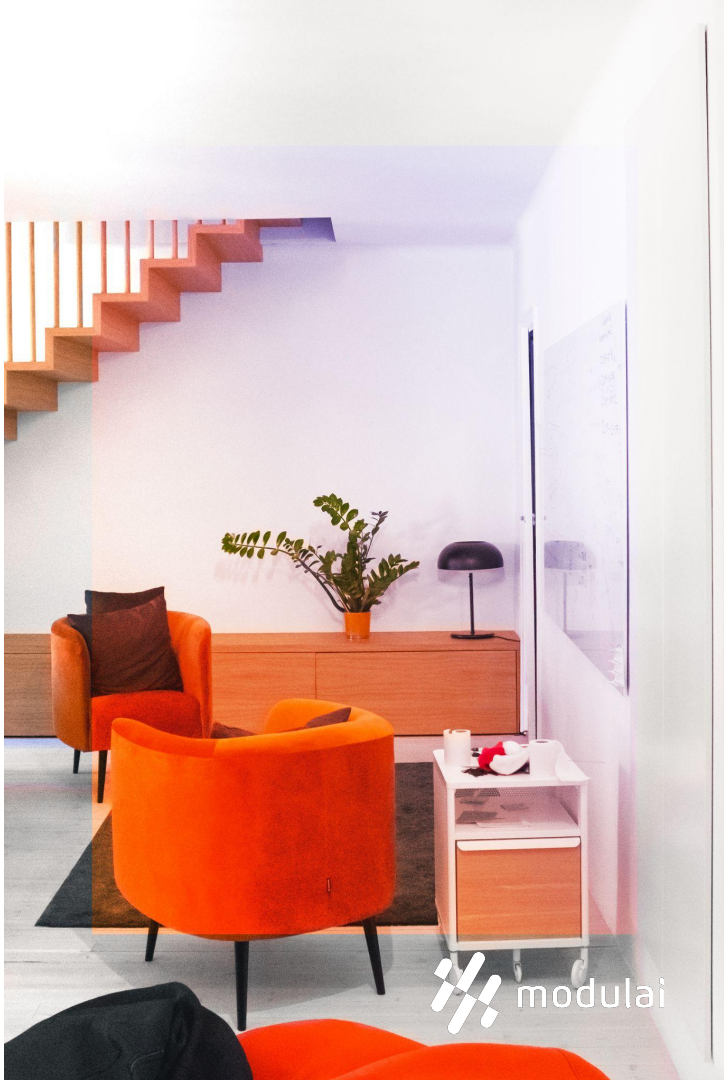
Me (Dmitrijs Kass)

- From Riga, Latvia
- Moved to Sweden in 2021
- 1st batch of UU Data Science MSc 2022
- Banks > FinTech > ML consulting
- Hobbies: being a dad
- Find me on LinkedIn


modulai
The machine learning agency
We help our clients solve business problems by taking end-to-end responsibility for planning, development and deployment of state-of-the-art machine learning solutions. Hands-on machine learning delivered.
modulai.io

Who we are

- We help our clients to solve some of their most important problems by developing **end-to-end Machine Learning (ML) products**.
- We offer project based consulting, joint venture models and develop own products, teaching & workshops.
- A **track record** of +80 models in production in the areas of churn prediction, campaign personalisation, recommendation engines, time series forecasting, natural language processing (NLP) and more.
- Modulai was founded 2018 by a team of **ex-Klarna machine learning engineers** with physics/math background. Team of 25 based in Stockholm & Gothenburg. Profitable from start, owned by employees.
- We support and help our clients leverage their ML competence long-term by assisting with recruitment (when needed) and working closely with the technical team.



Computer vision

Use images and video streams for improved decision-making or autonomous intelligence

Natural language processing

Leverage the value in natural language data produced by and for humans. LLM, RAG, LangChain

Tabular ML and time series

Leverage machine learning on data almost all companies produce

Personalization and AI for retail

AI models for increased sales, relevant content and cost saving in retail

AI fo Medtech

Solutions for treating disease, improving health, and supporting professionals

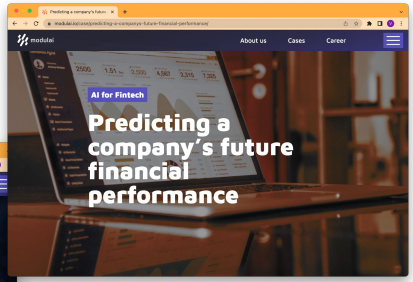
AI fo Fintech

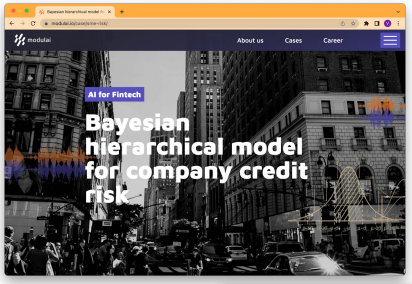
Better financial decisions by assessing risks and accurate forecasting

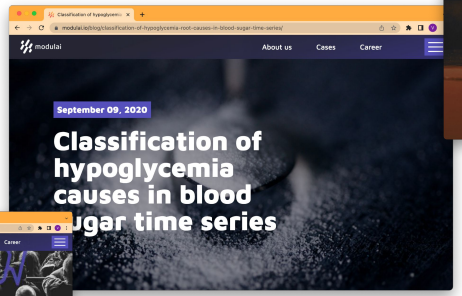
modulai

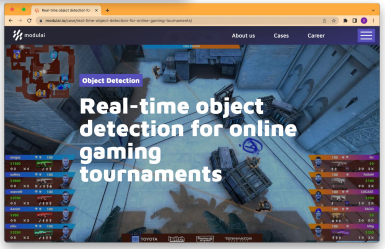
Cases

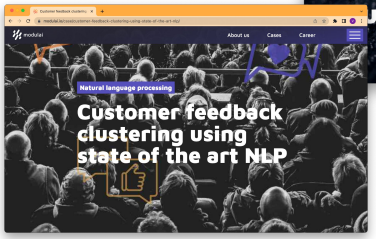
Gathered

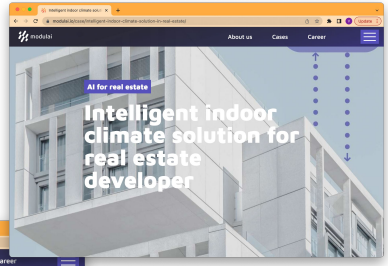


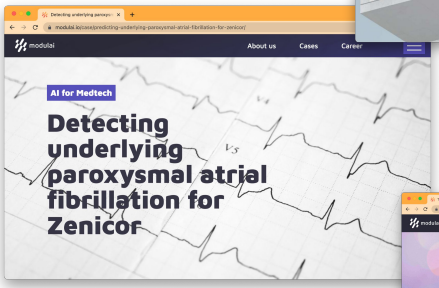


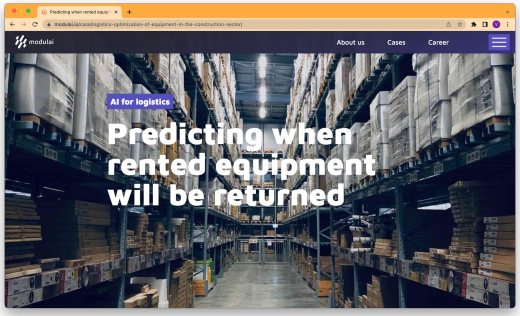


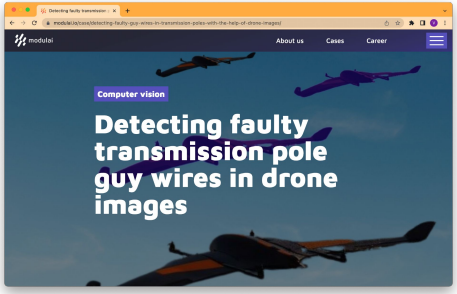


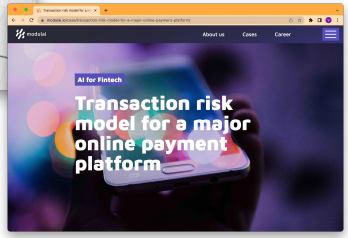










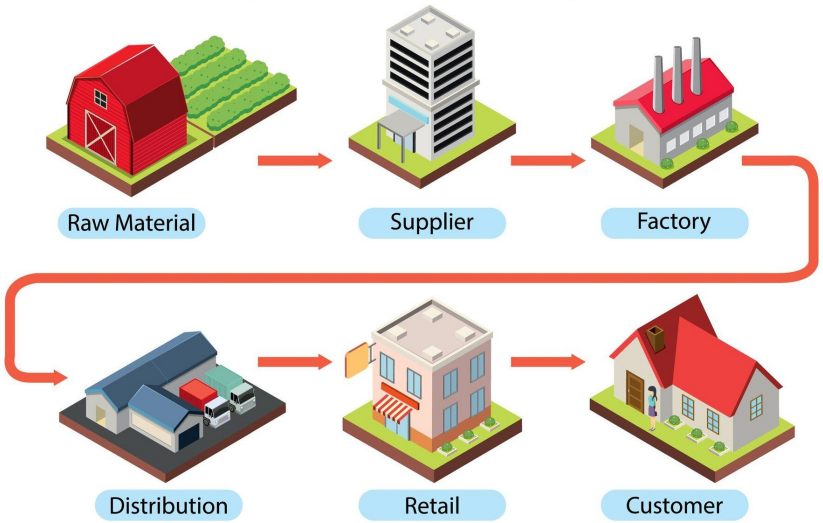


modulai

Who we work with

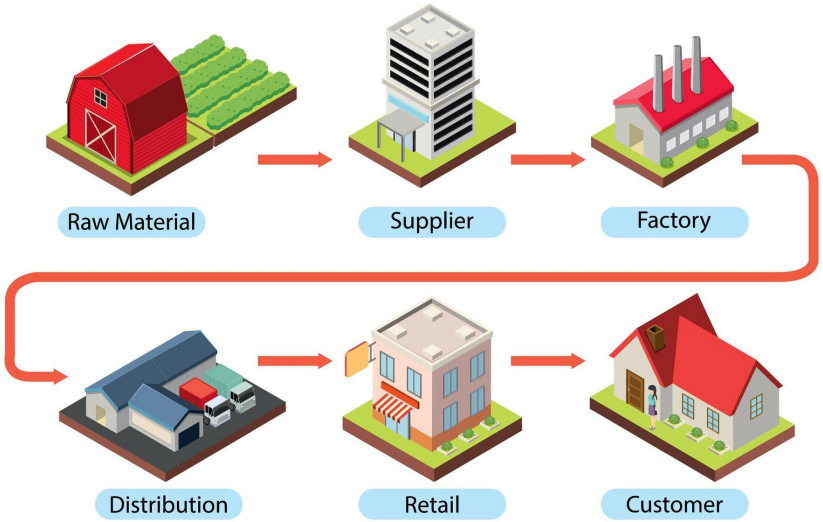

Shallow
Deep-Dive into
a Use Case
modulai

Supply Chain Management (SCM)



modulai

Supply Chain Management (SCM)



How to make it efficient?

Tools like

- Scenario analysis
- Optimization

require

- Large volumes of data
- Database-like structure

modulai

Business problem

modulai

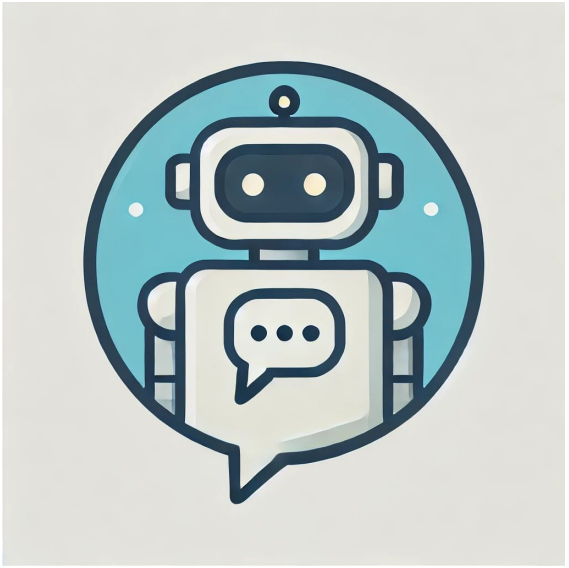
Images: DALLE

Business problem

modulai

Images: DALLE

Solution



modulai

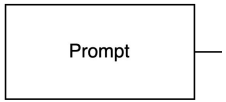
Images: DALLE

Optilogic

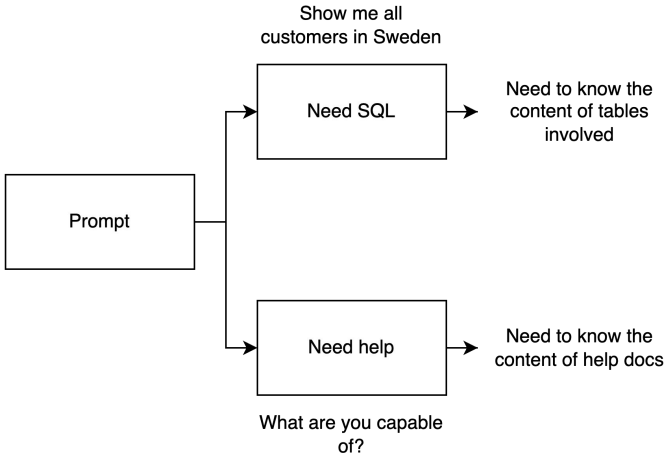
- Optilogic is a supply chain solutions company.

- Main product: Cosmic Frog
- Live demo of the platform

- Enables customers to build, edit, and test supply chain models.

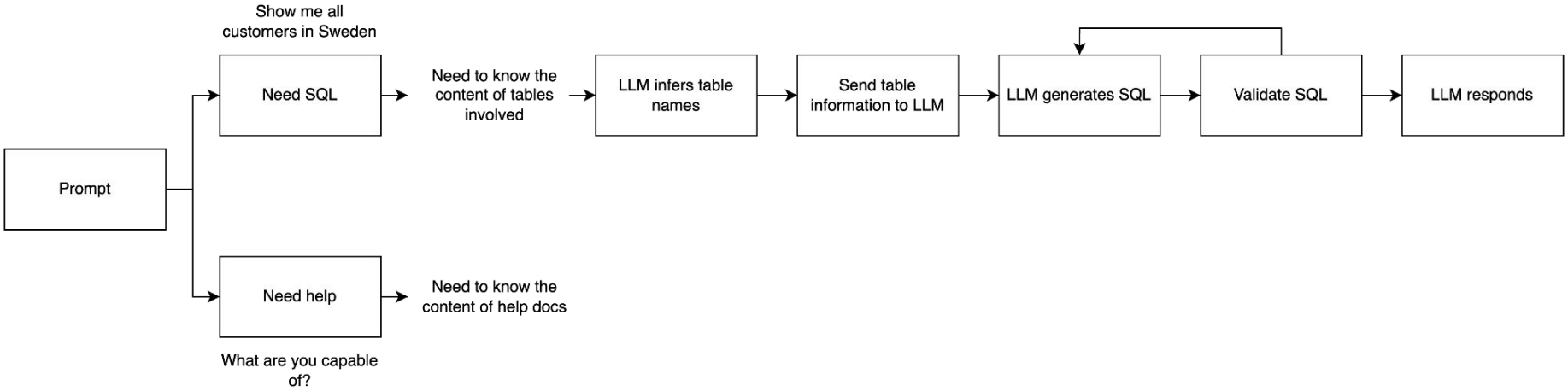


modulai



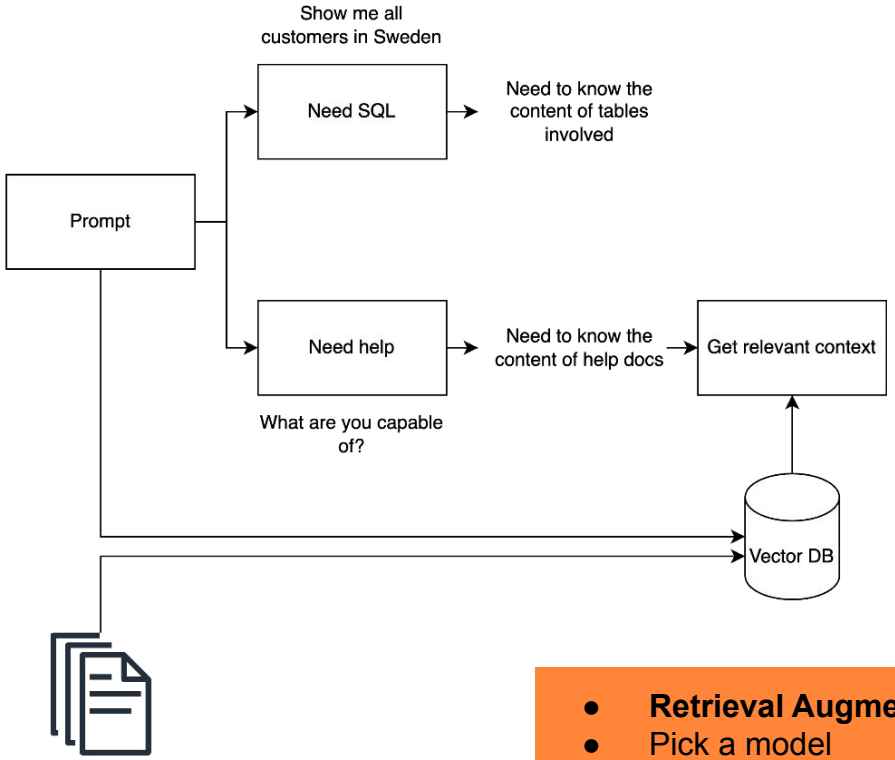
- Tool calling
- Limited context window
- Even if you can
  - Cost
  - Latency
  - Lost in the middle

modulai



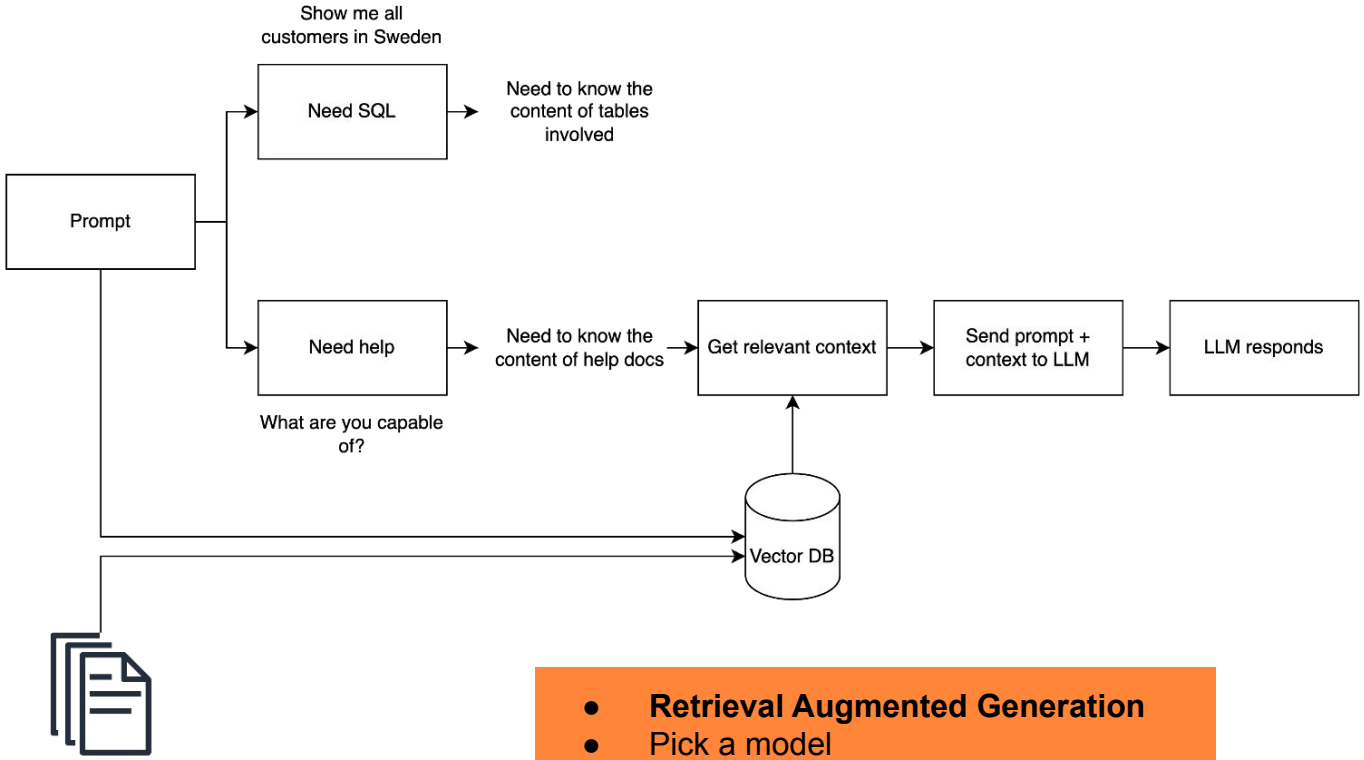
- **Structured output**
- Pick a model
- Complex SQL and style
  - **Few-shot learning**
  - Fine-tuning

modulai



- **Retrieval Augmented Generation**
- Pick a model
- Domain adaptation

modulai



- **Retrieval Augmented Generation**
- Pick a model
- Domain adaptation

modulai

Structured output

Notebook

modulai

Few-shot learning (1/2)

<question>
Select all customers from Sweden
</question>

<instructions>
Generate a SQL statement for the prompt above
</instructions>

modulai

Few-shot learning (2/2)

<question>
Select all customers from Sweden
</question>

<instructions>
Generate a SQL statement for the prompt above using examples below. Use the database schema and column values exactly as presented in examples.
</instructions>

<examples>
Q: How many customers do I have?
A: SELECT COUNT(DISTINCT customer_name) FROM customers

Q: What are Swedish products?
A: SELECT product_name FROM products WHERE country = "swe"
</examples>

modulai

RAG

<question>
How to solve A?
</question>

<instructions>
Please provide a concise answer to the question above using only the context below. If context doesn't contain the necessary information, reply with "Don't know"
</instructions>

<context>
To solve A, do B
To solve C, to D
C cannot be solved
</context>

modulai

RAG

<question>
How to solve A?
</question>

<instructions>
Please provide a concise answer to the question above using only the context below. If context doesn't contain the necessary information, reply with "Don't know"
</instructions>

<context>
To solve A, do B
To solve C, to D
C cannot be solved
</context>

} Imagine, you have 1,000,000,000 of examples

modulai

Ingestion

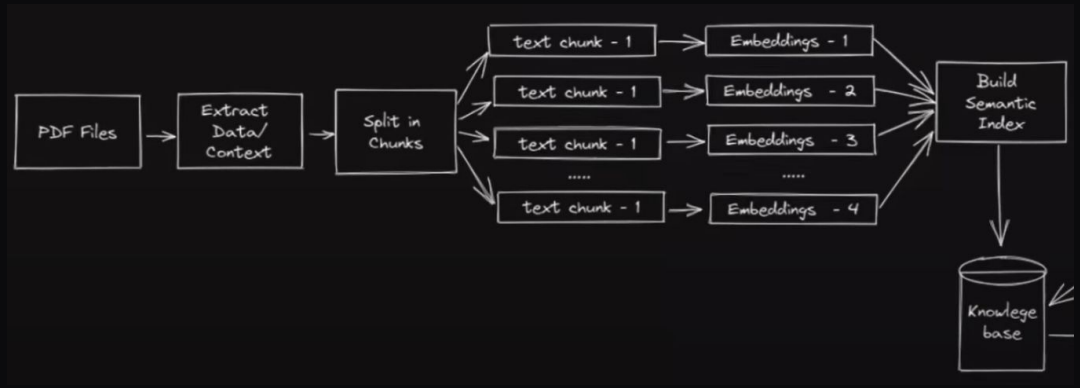


Image credits: https://www.youtube.com/watch?v=TLf90ipMzfE

Querying

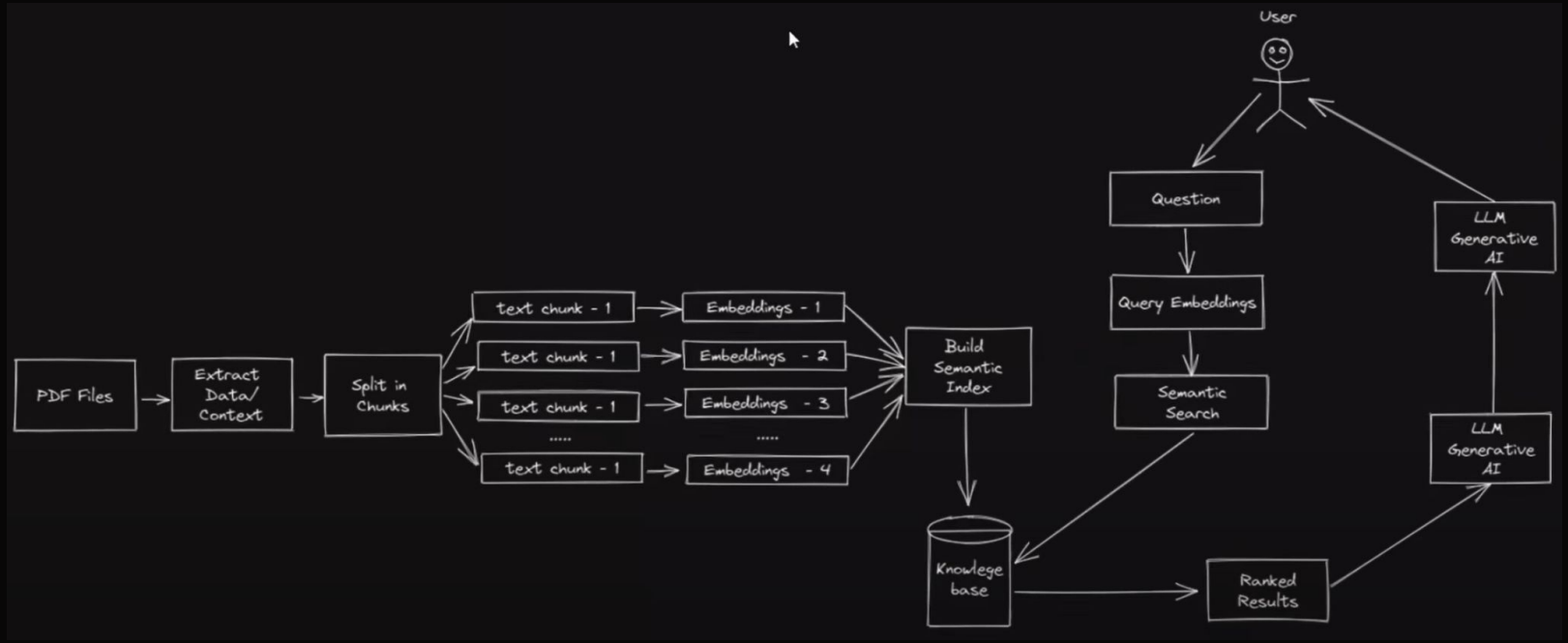
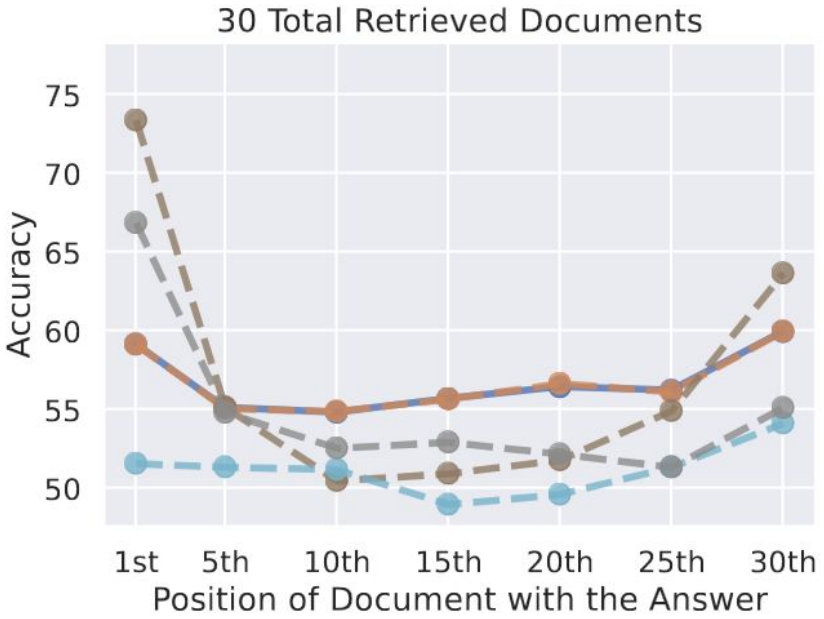


Image credits: https://www.youtube.com/watch?v=TLf90ipMzfE

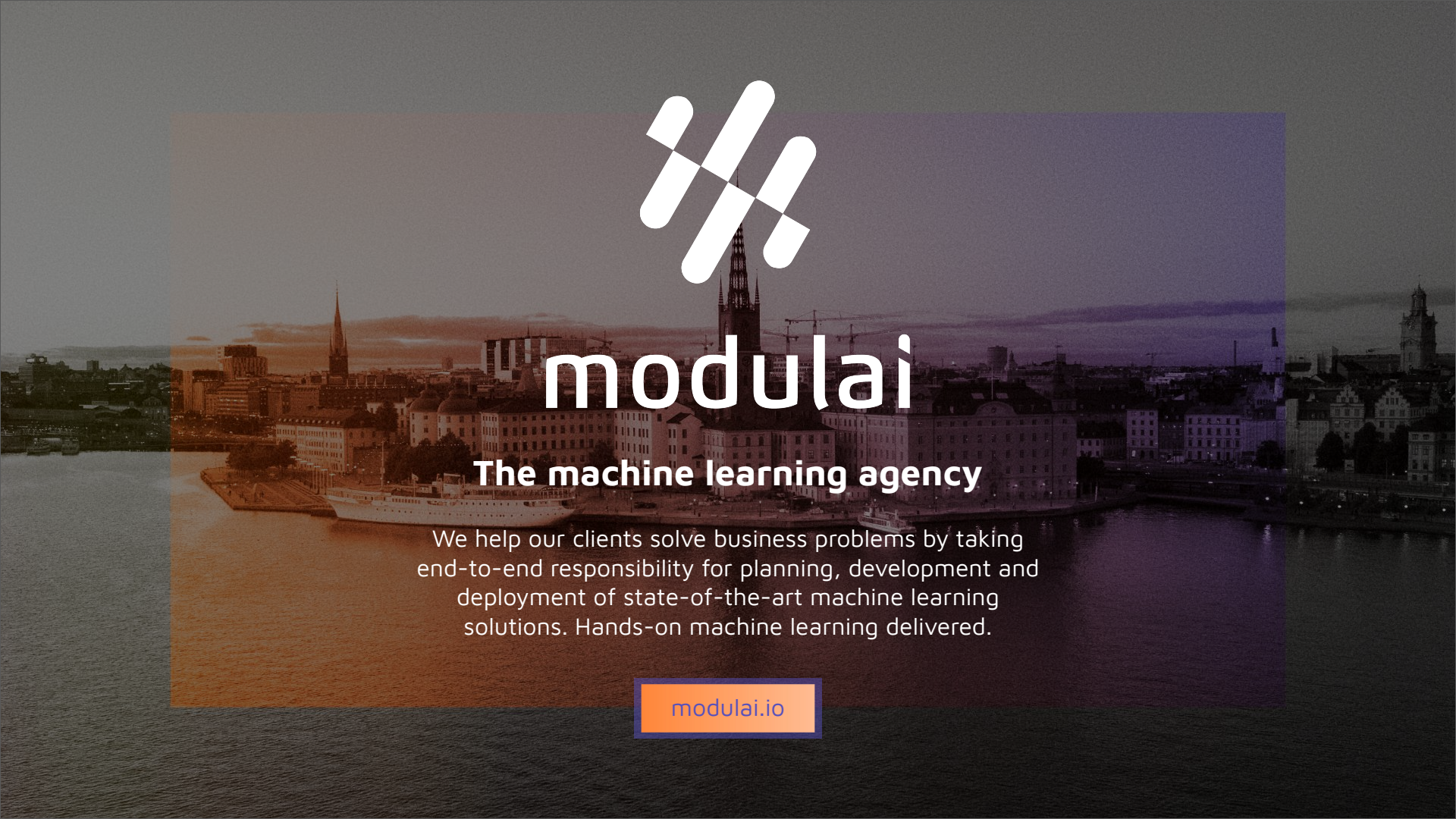
Lost in the middle

[Lost in the Middle: How Language Models Use Long Contexts]

- LLMs have issues with long context windows, not capturing the information in the middle of the context.
- More accurate retrieval and document *ranking* → more accurate answers



modulai


modulai
The machine learning agency
We help our clients solve business problems by taking end-to-end responsibility for planning, development and deployment of state-of-the-art machine learning solutions. Hands-on machine learning delivered.
modulai.io