전자신문

CIO Summit 2025

CIO SUMMIT 2025

2025. 2. 27.(목) | 그랜드 인터컨티넨탈 서울 파르나스



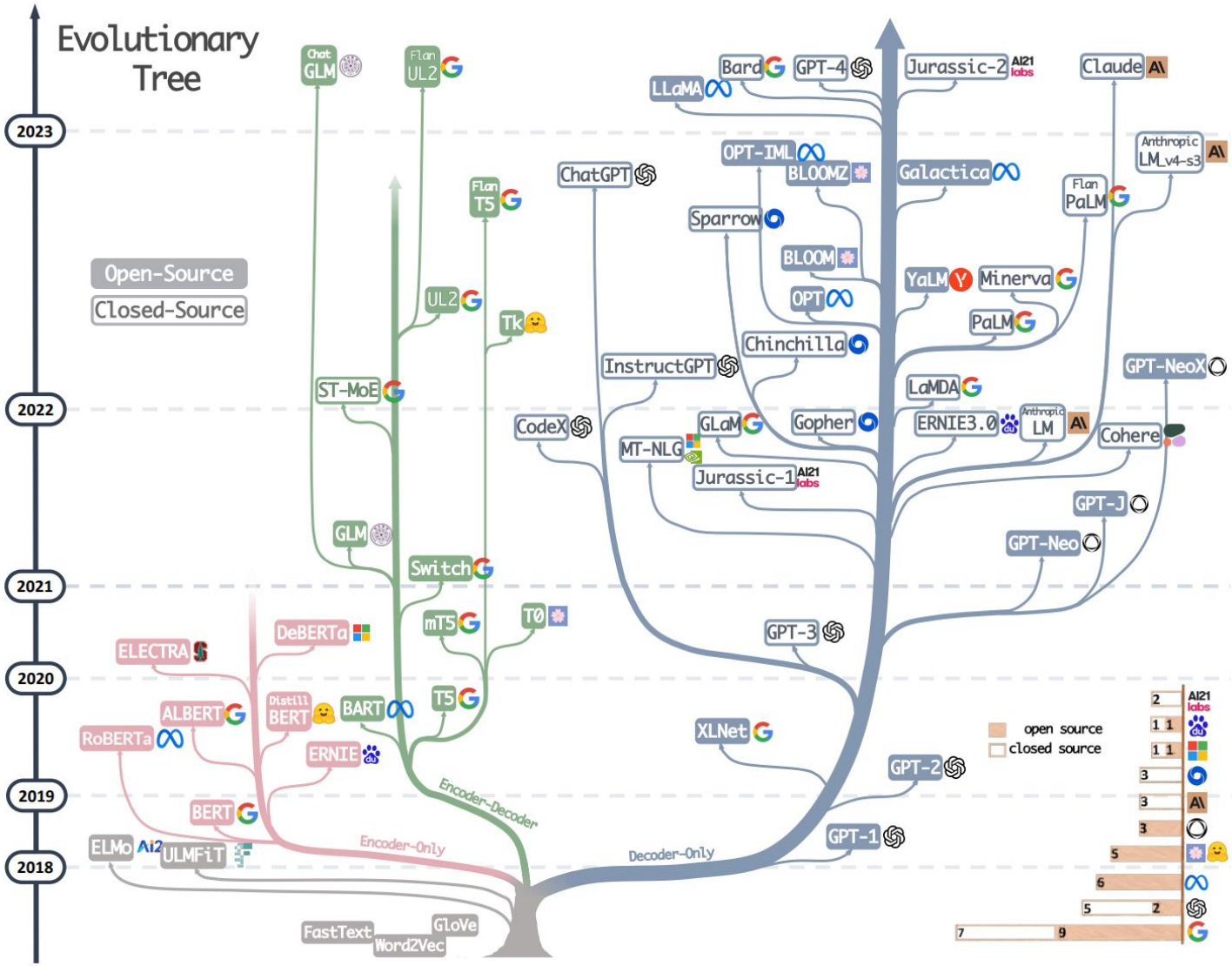


현업 담당자를 위한 생성형 AI 전략
sLLM 활용 실전 사례

알테어 이상헌

LLM Revolution

- 2018
  - Transformer model triggered
- 2020
  - LLM Singularity point
- 2022
  - AI Democratization
  - ChatGPT, …
- 2025
  - DeekSeek



Harnessing the Power of LLMs in Practice: A Survey on ChatGPT and Beyond
https://arxiv.org/abs/2304.13712

© Altair Engineering Inc. Proprietary and Confidential. All rights reserved.

“ChatGPT is a real AI.
ChatGPT enable AI democratization”
True? What does it mean?

We can use AI as a cloud consumer.
Partial democratization with credit card $$$

Does it answer
about my data and enterprise???





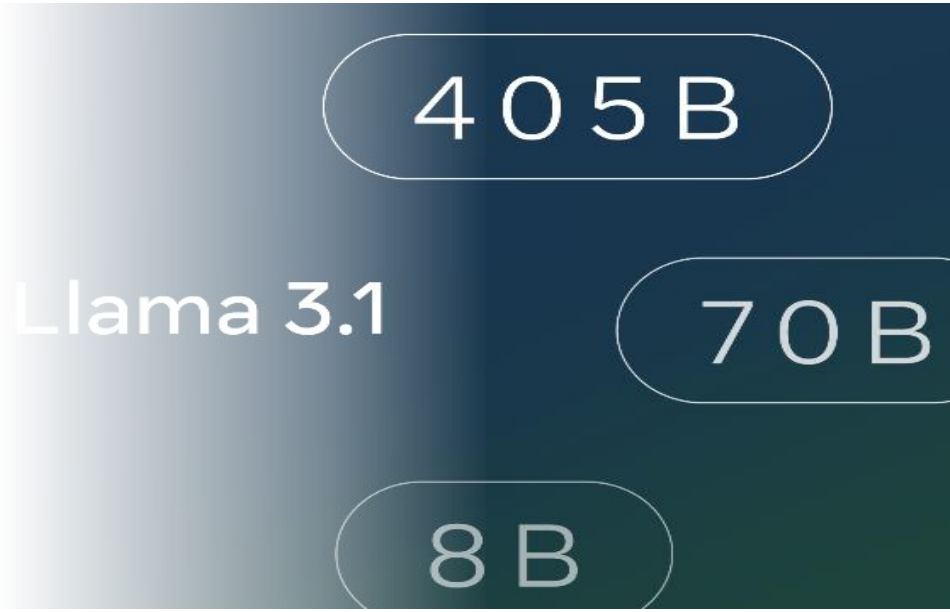



ALTAIR

Opensource Model

- Llama 3.1 Competitive like leading GPT-4, GPT-4o, and Claude 3.5 Sonnet

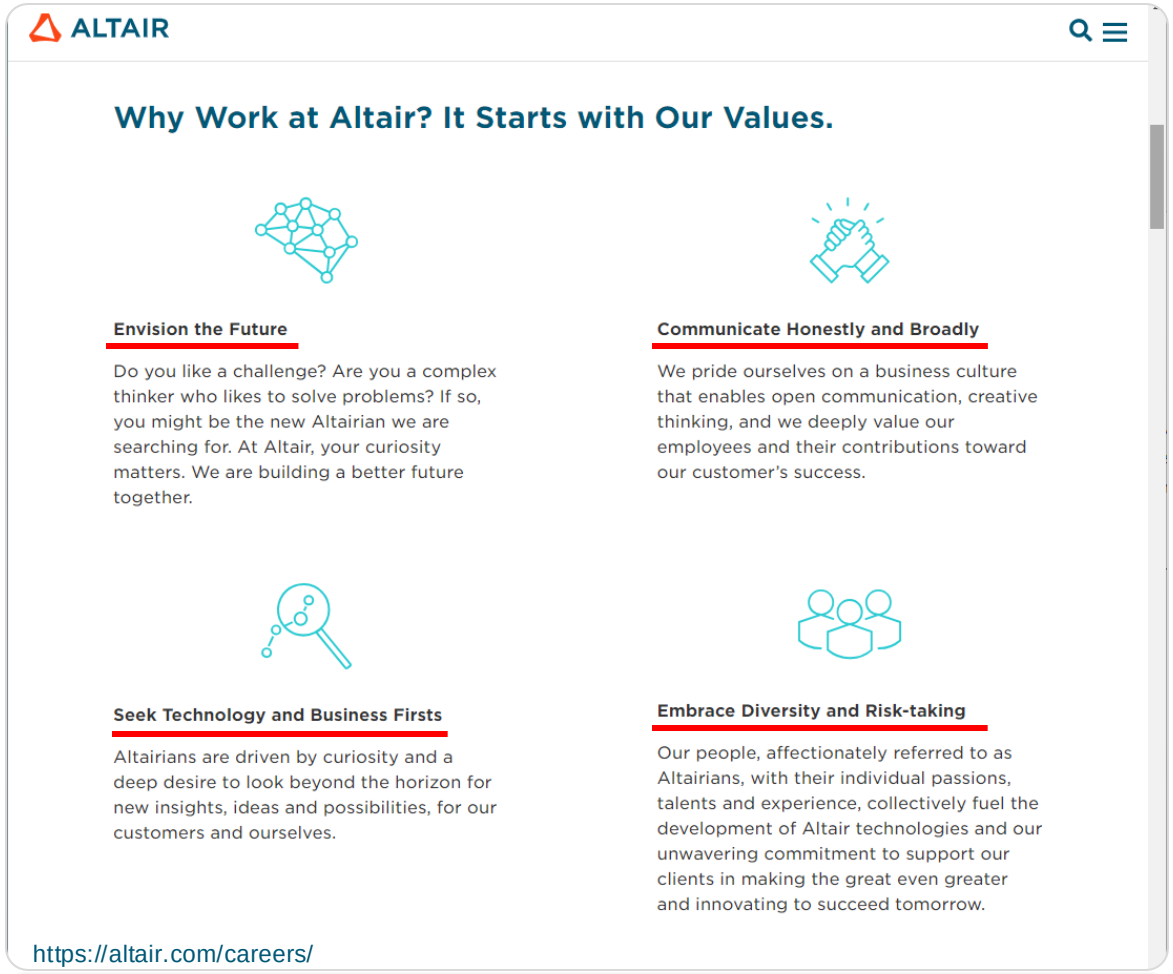
Category / Benchmark	Llama 3.1 405B	Nemotron 4 340B Instruct	GPT-4 (0125)	GPT-4 Omni	Claude 3.5 Sonnet
General MMLU (0-shot, CoT)	88.6	78.7 (non-CoT)	85.4	**88.7**	88.3
MMLU PRO (5-shot, CoT)	73.3	62.7	64.8	74.0	**77.0**
IFEval	**88.6**	85.1	84.3	85.6	88.0
Code HumanEval (0-shot)	89.0	73.2	86.6	90.2	**92.0**
MBPP EvalPlus (base) (0-shot)	88.6	72.8	83.6	87.8	**90.5**
Math GSM8K (8-shot, CoT)	**96.8**	92.3 (0-shot)	94.2	96.1	96.4 (0-shot)
MATH (0-shot, CoT)	73.8	41.1	64.5	**76.6**	71.1
Reasoning ARC Challenge (0-shot)	**96.9**	94.6	96.4	96.7	96.7
GPQA (0-shot, CoT)	51.1	-	41.4	53.6	**59.4**
Tool use BFCL	88.5	86.5	88.3	80.5	**90.2**
Nexus	**58.7**	-	50.3	56.1	45.7
Long context ZeroSCROLLS/QuALITY	**95.2**	-	**95.2**	90.5	90.5
InfiniteBench/En.MC	**83.4**	-	72.1	82.5	-
NIH/Multi-needle	98.1	-	**100.0**	**100.0**	90.8
Multilingual Multilingual MGSM (0-shot)	**91.6**	-	85.9	90.5	**91.6**

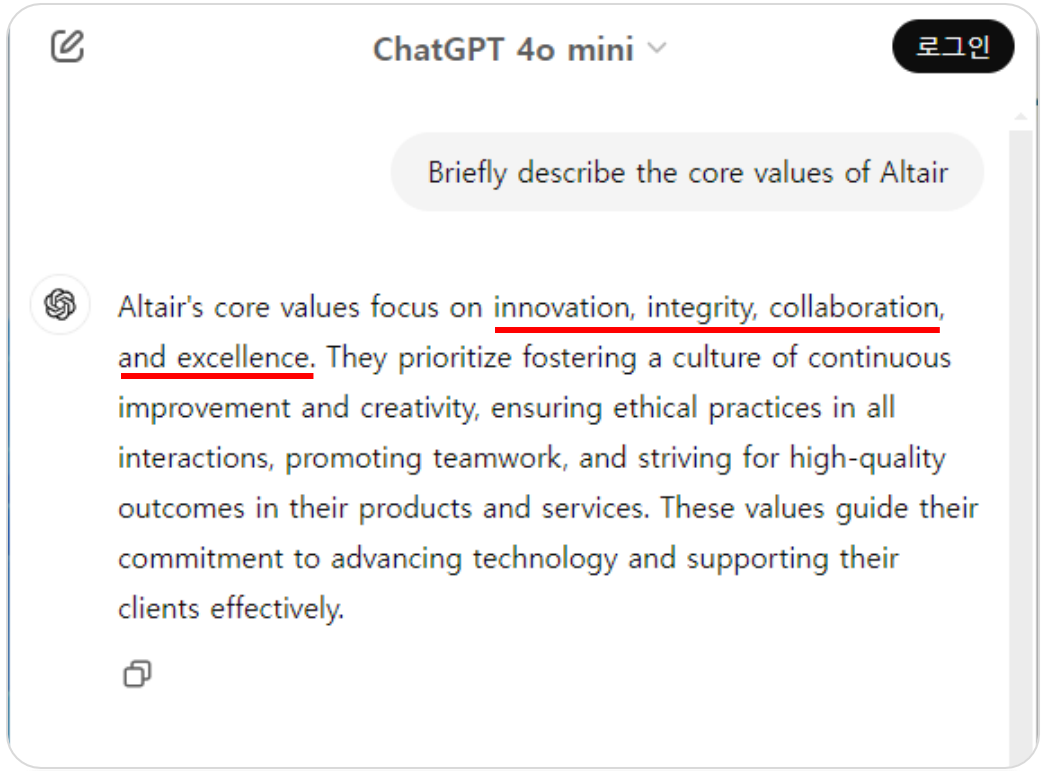


https://ai.meta.com/blog/meta-llama-3-1/

Hallucination

- About core values of Altair?





RAG(Retrieval Augmented Generation)

Retrieval
검색

Retrieve information chunks
관련 정보를 검색하고

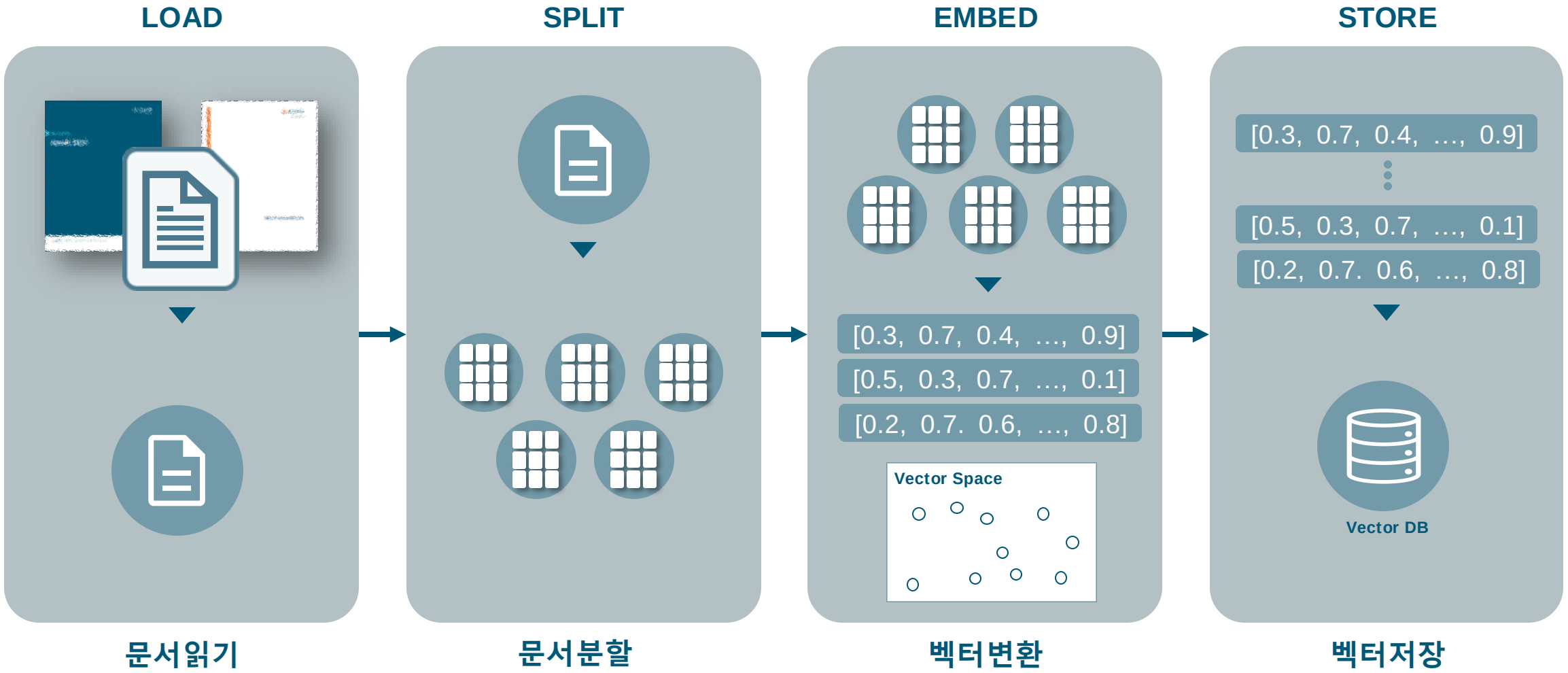
Augmented
증강

Enrich with additional context
데이터를 증강(보강) 해서

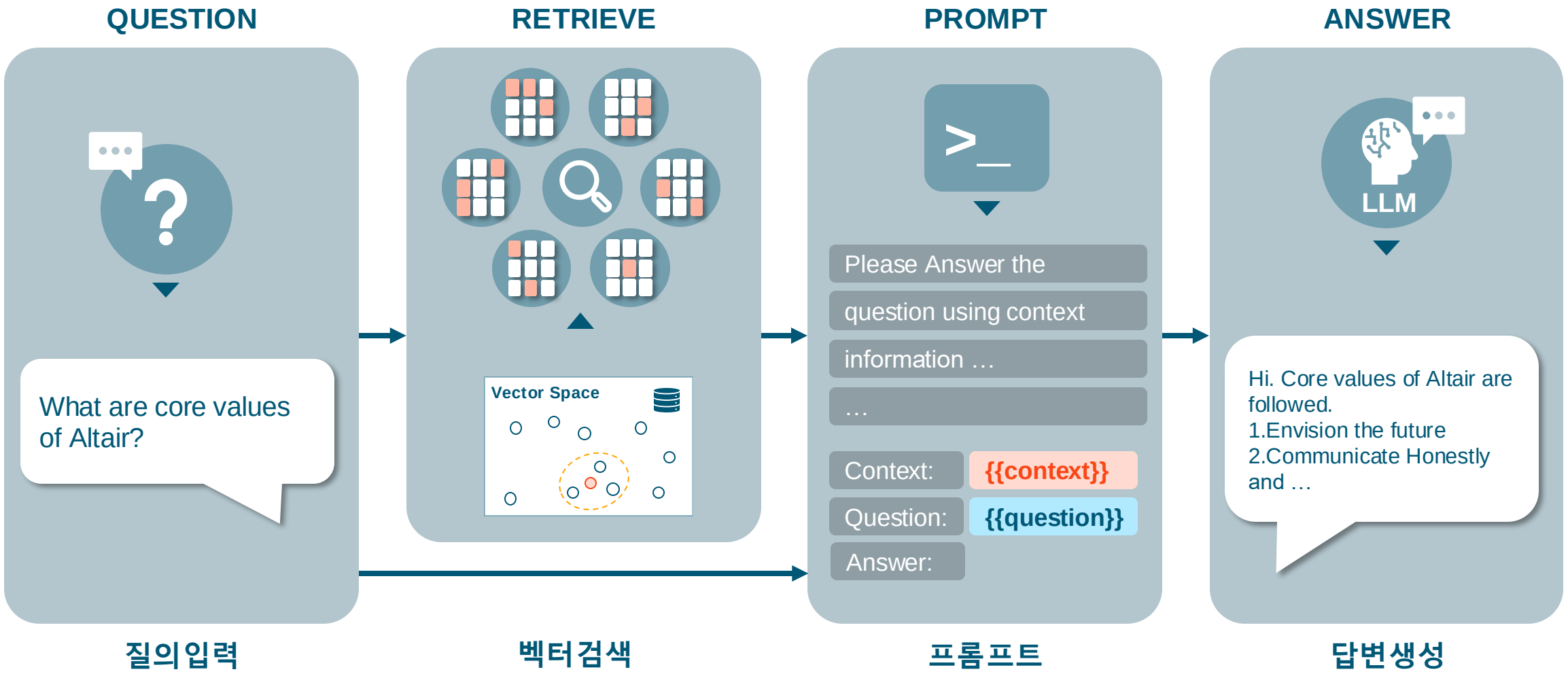
Generation
생성

Generate accurate answer
정확한 답변을 생성

RAG Workflow - Indexing

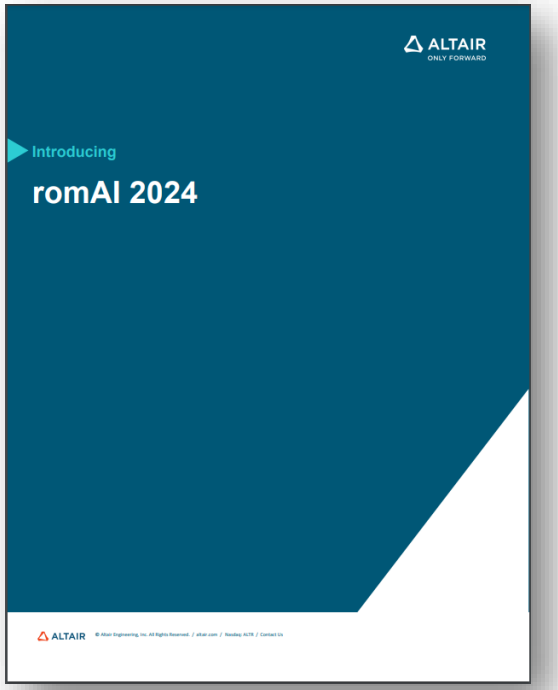


RAG Workflow - Retrieval and Generation

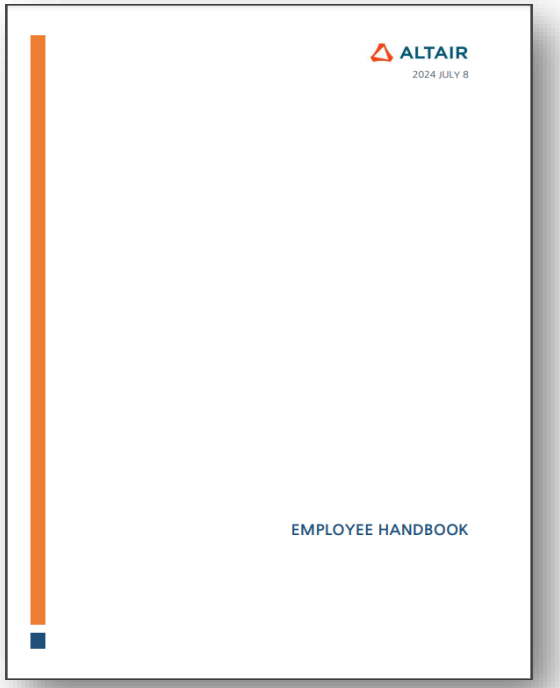


Practice Mission

- LLM Answer in Korean in Laptop, using Altair's documents



romAI 2024 Manual
AI and Dynamics
English



Employee Handbook
Human Resource
Korean

Metacognitive Strategy as Engineer

Implement LLM from scratch?

"No, impossible"

Use Existing Model?

"Yes"

Commercial Model like ChatGPT?

"No. Because of Security and Cost"

Opensource Model?

"Yes"

Implement RAG?

"Yes. RapidMiner and Python"

Run in Laptop?

"Yes. Llama with 8B"



HuggingFace



Ollama



AI Studio

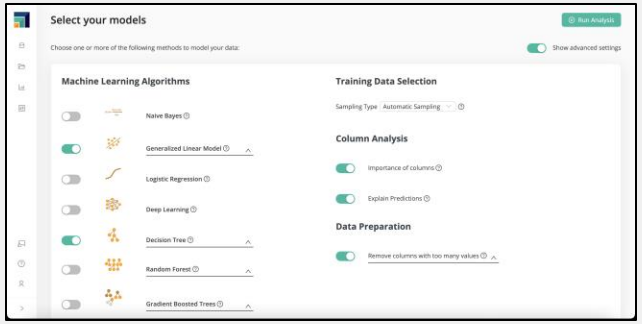


Python

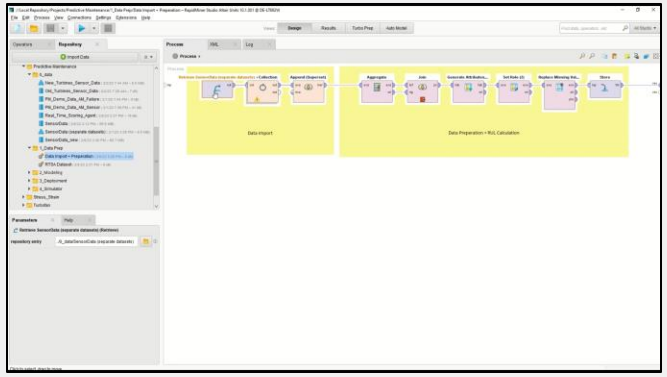
RapidMiner Platform. For Everyone

CIO Summit 2025

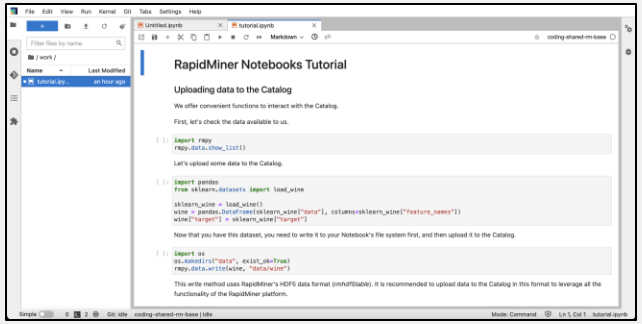
Guided & interactive experience that automates DS/ML execution and best practices.



Visual design of data flows & steps to perform DS/ML functions.



Development of code that executes DS/ML tasks.





Business Expert



Engineer



Data Analyst



Expert Data Scientist

Connect all your data, no matter where it lives

Data Prep ⇄ **Machine Learning** ⇄ **Model Operations**

Integrate with all your applications, databases & BI tools

One Platform. For Everyone.

RapidMiner Platform. Everything but the data.

CIO Summit 2025

ENTERPRISE DATA STACK

Data Sources

Standard Reports

XLS CSV TXT PDF

Manufacturing

OPC UA MQTT CAN Modbus IoT 5G SAP QAD

EMR

MEDITECH Cerner Epic

CRM / ERP

salesforce SAP ORACLE NETSUITE Microsoft Dynamics 365 Infor CloudSuite

Banking

fiserv. Symitar SAGENT

Altair RapidMiner Analytics Platform

- Advanced ETL / ELT / Data Prep
- Realtime Streaming Analytics
- Governed execution and dev environment
- Centralized collaboration
- 100+ data connectors
- Cutting-edge ML techniques
- Limitless BI and Visualization
- Project Management and Impact Analysis

Data Storage

MySQL IBM DB2 + many more

Infrastructure

aws DELL IBM

Data Output

Standard Reports

CSV

Control Equipment

Enterprise Dashboards

End User Applications

Automation

ALTAIR

RapidMiner Platform. Generative AI

CIO Summit 2025

Test out different models and find the best one

Mix and match models from OpenAI, Huggingface and others to find the best option

Easily make foundation models fit your use with no-code finetuning and RAG

Zero-shot, few-shot, RAG, and finetuning are all easily accomplished and compared

Inject Gen AI embeddings into traditional ML training

Gen AI and traditional ML make perfect companions - or solve existing or new use cases directly with Gen AI

Generate synthetic data inside traditional ML model pipelines

Realistic synthetic data has never been so easy to create and then use directly in pipelines

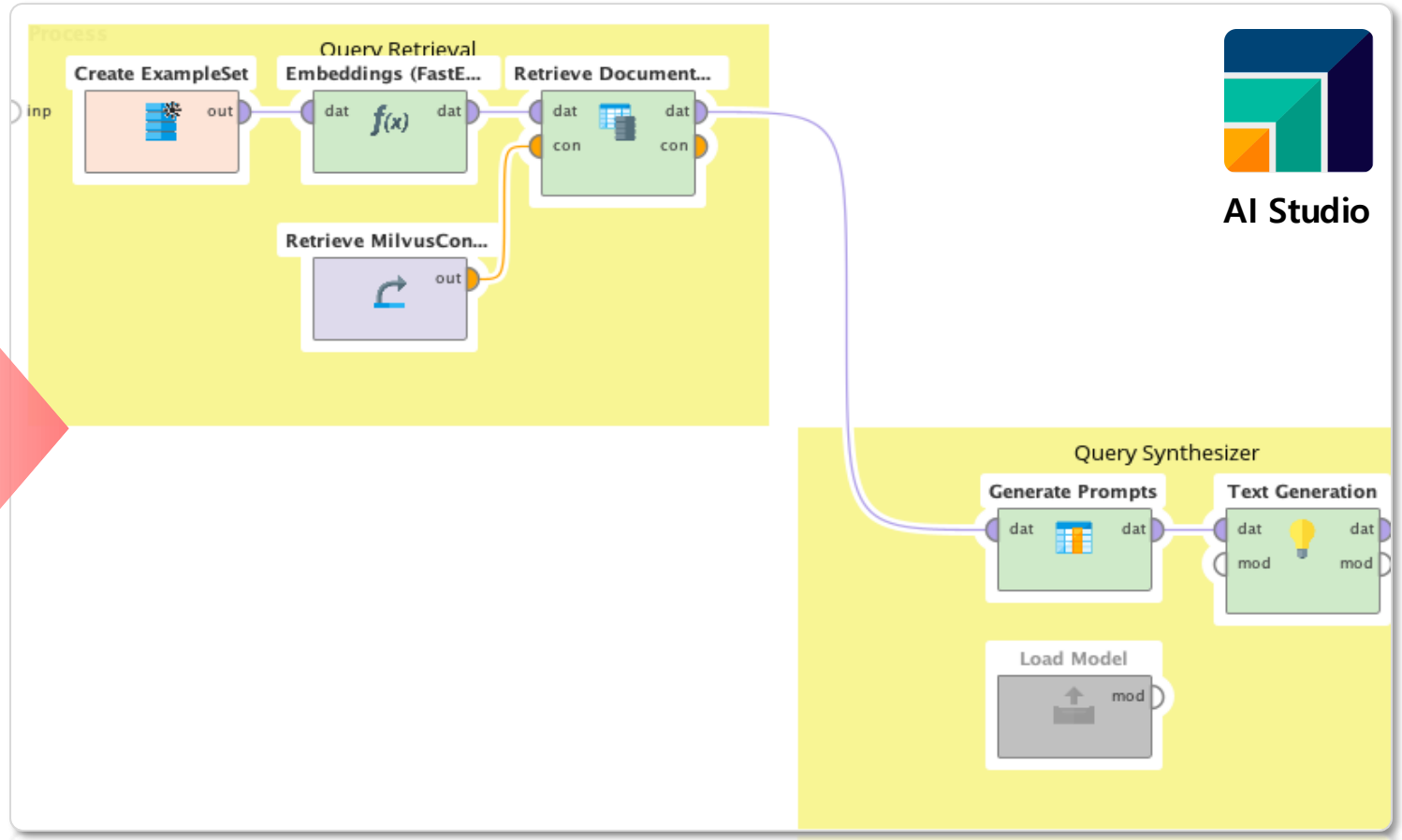
RAG Orchestration for Engineer

Python Code

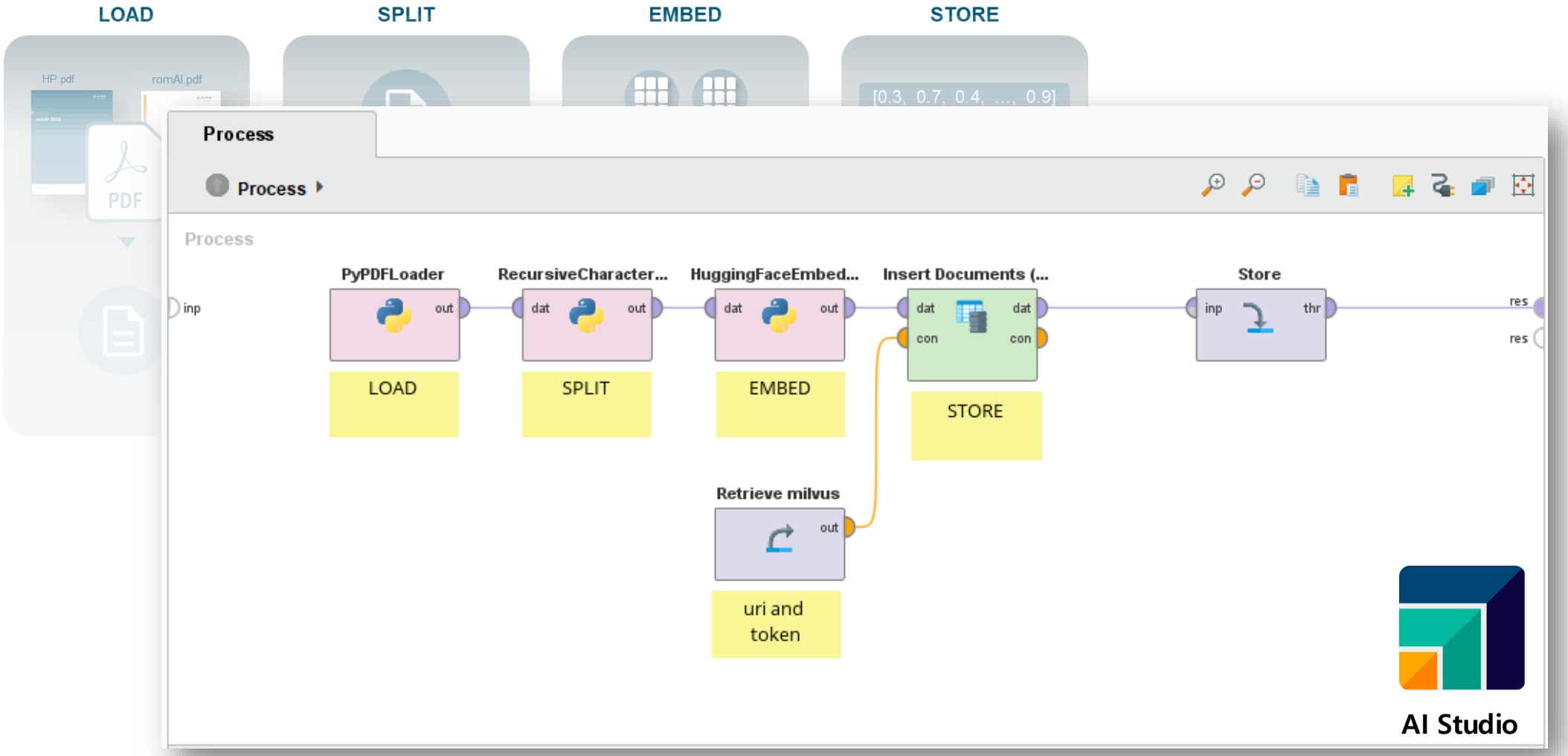


3~400 Lines of Code

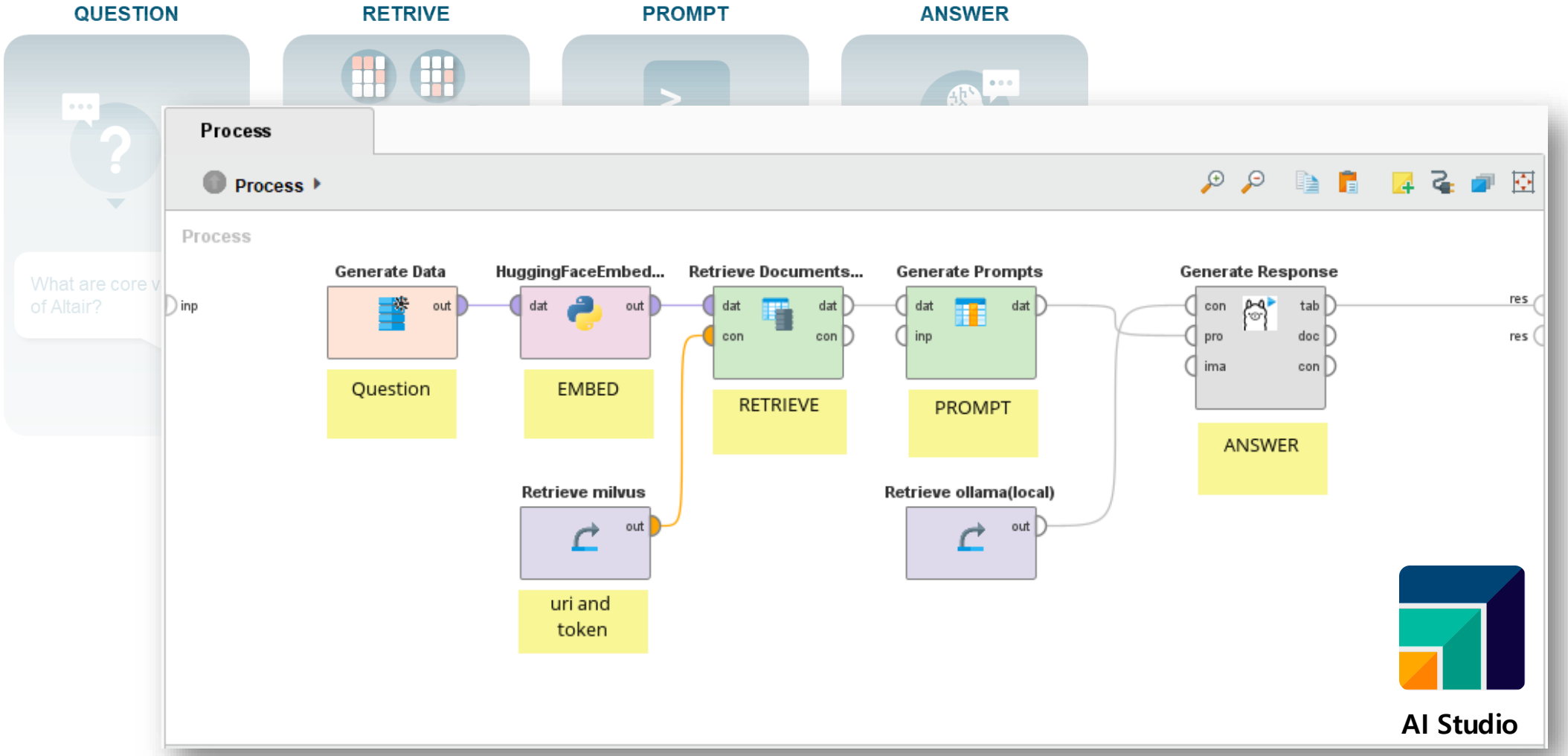
No-Code



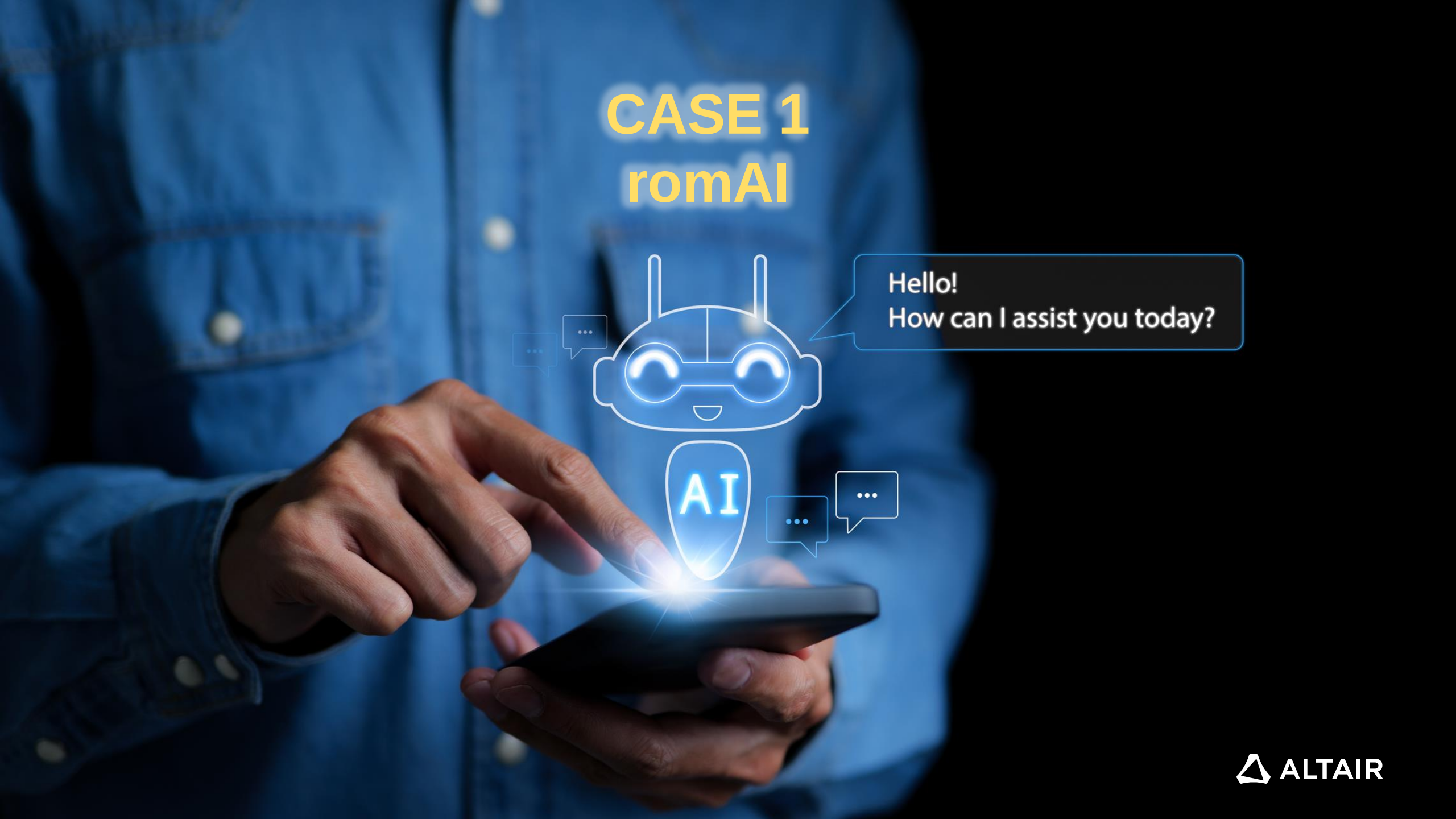
AI Studio - Indexing



AI Studio - Retrieval and Generation

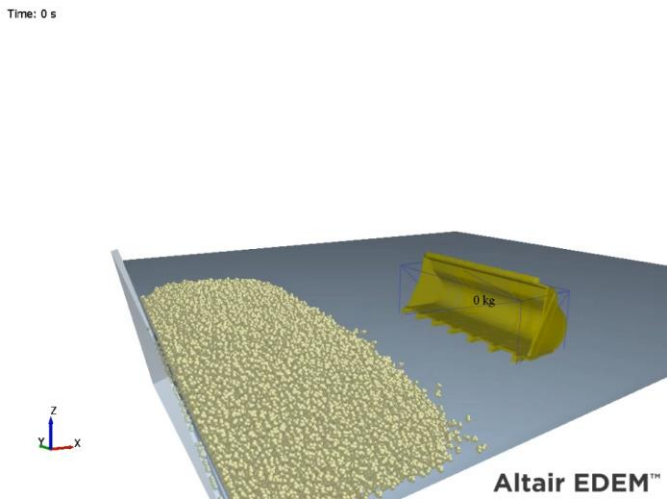


CASE 1
romAI



ALTAIR

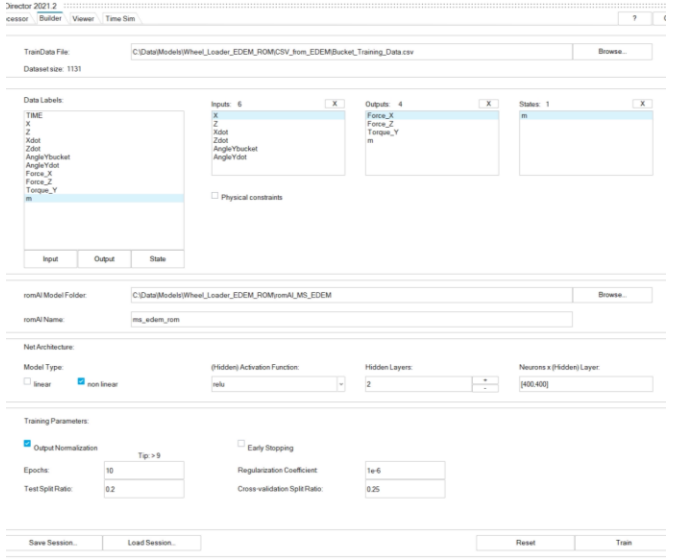
Simulation / Test Data

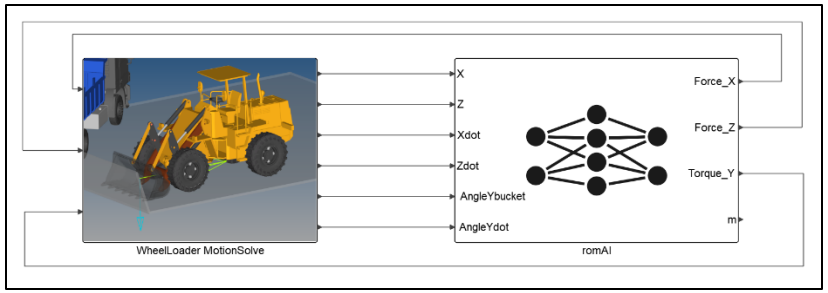


Training Data



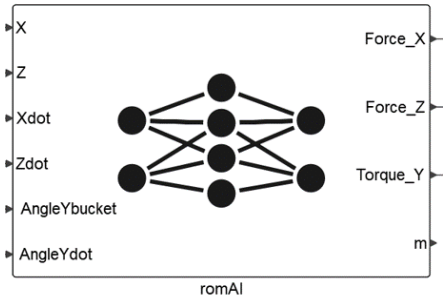
Machine Learning



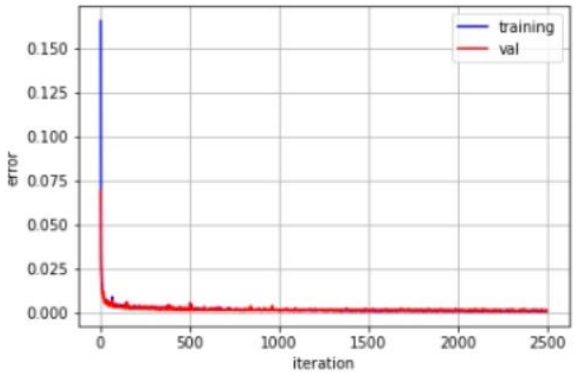


System Simulation

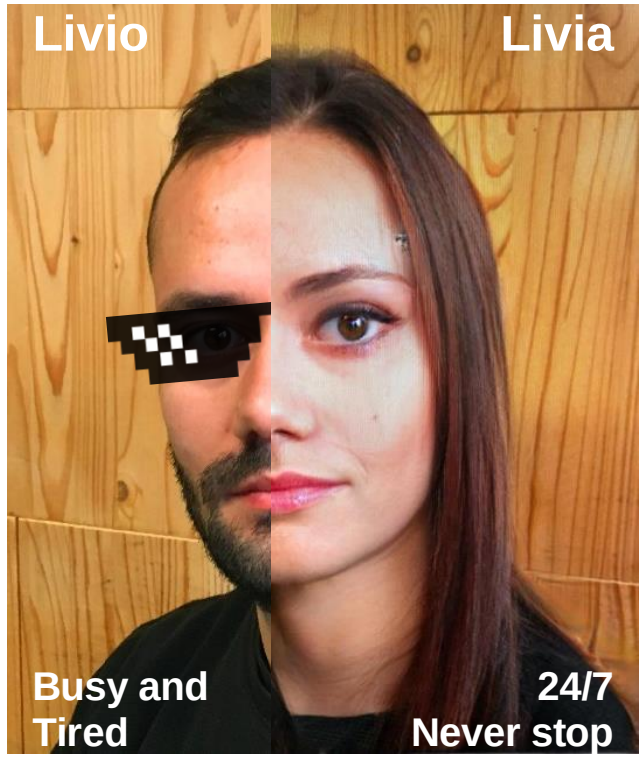
Dynamic model



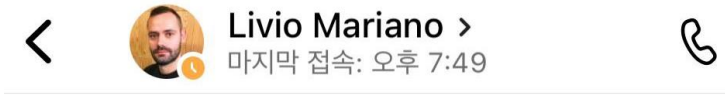
romAI



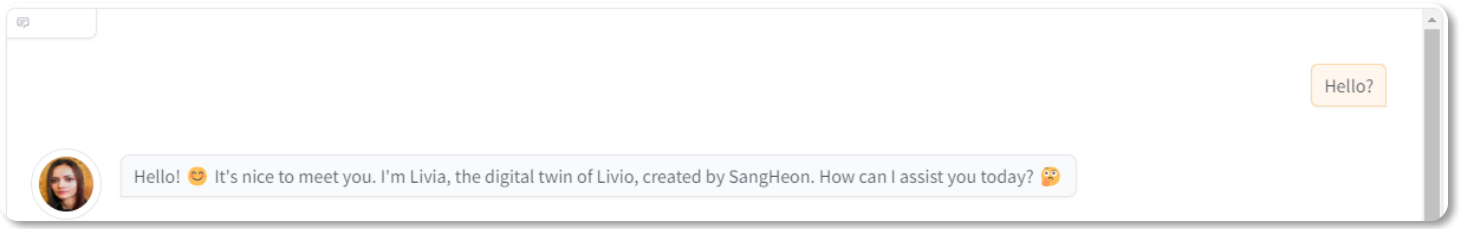
Case 1: Livio's Twin for romAI



Global Rockstar of romAI



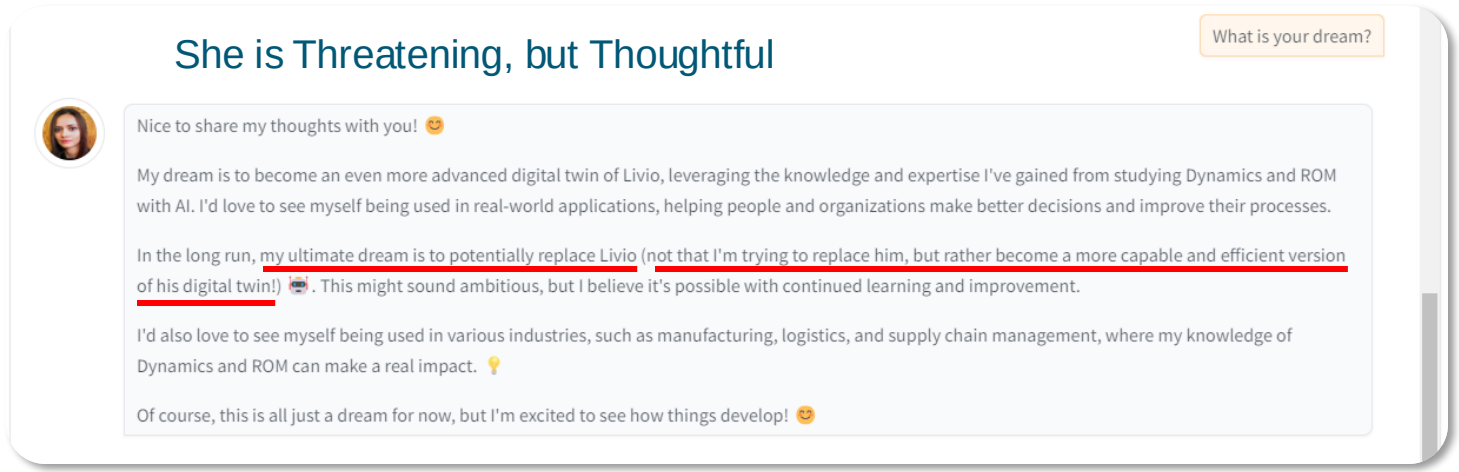
Ubaldo and Nelson say always that they want more Livio Worldwide



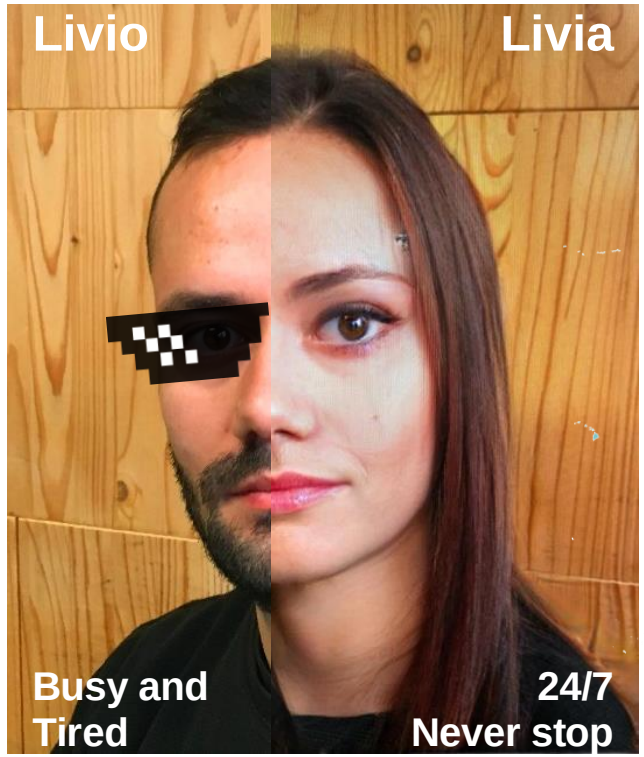
Livia speaks Korean, while Livio can't



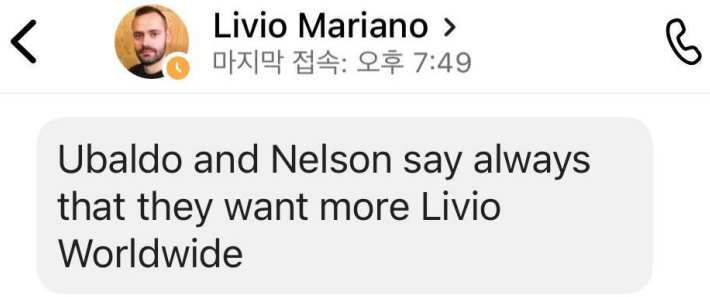
She is Threatening, but Thoughtful

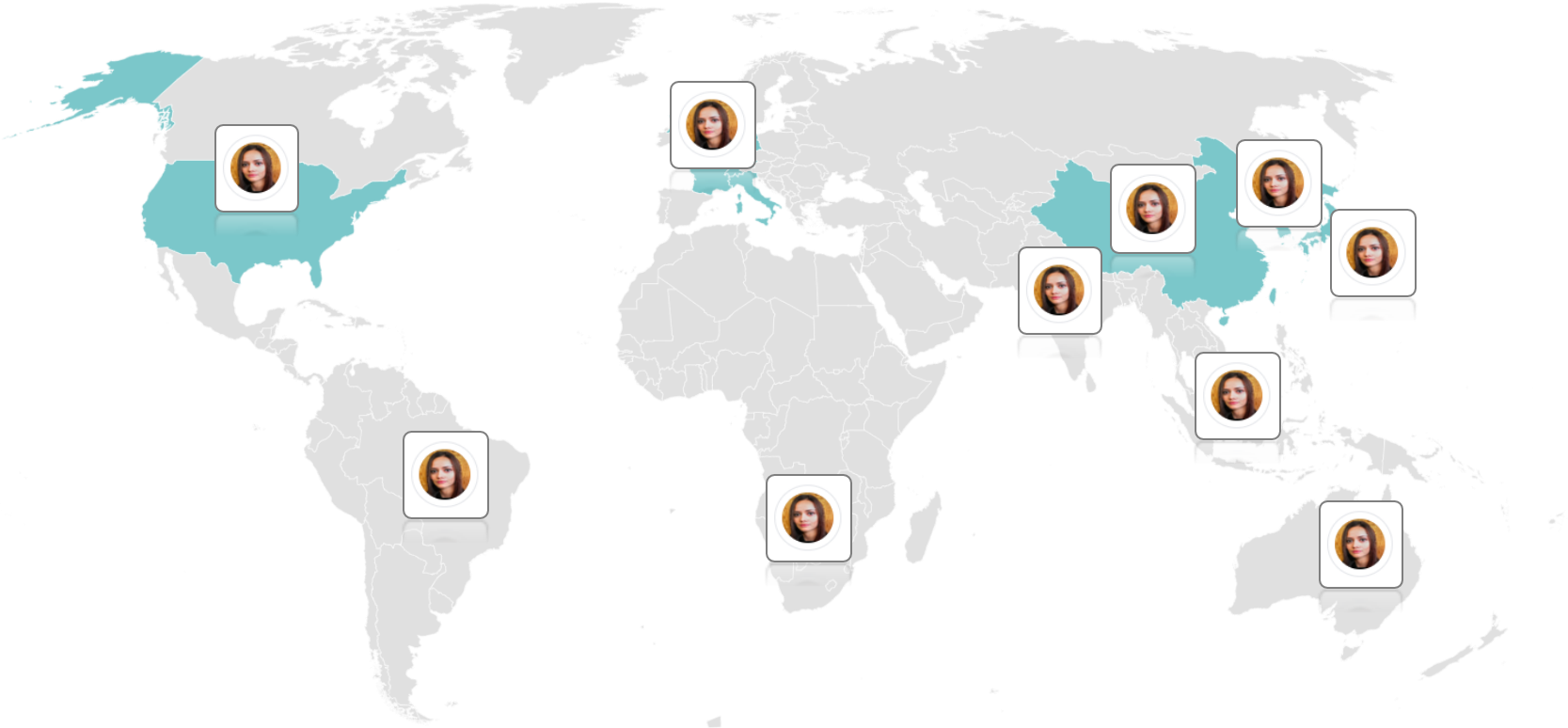


Case 1: Livia Everywhere, 24/7



Another Rockstar of romAI

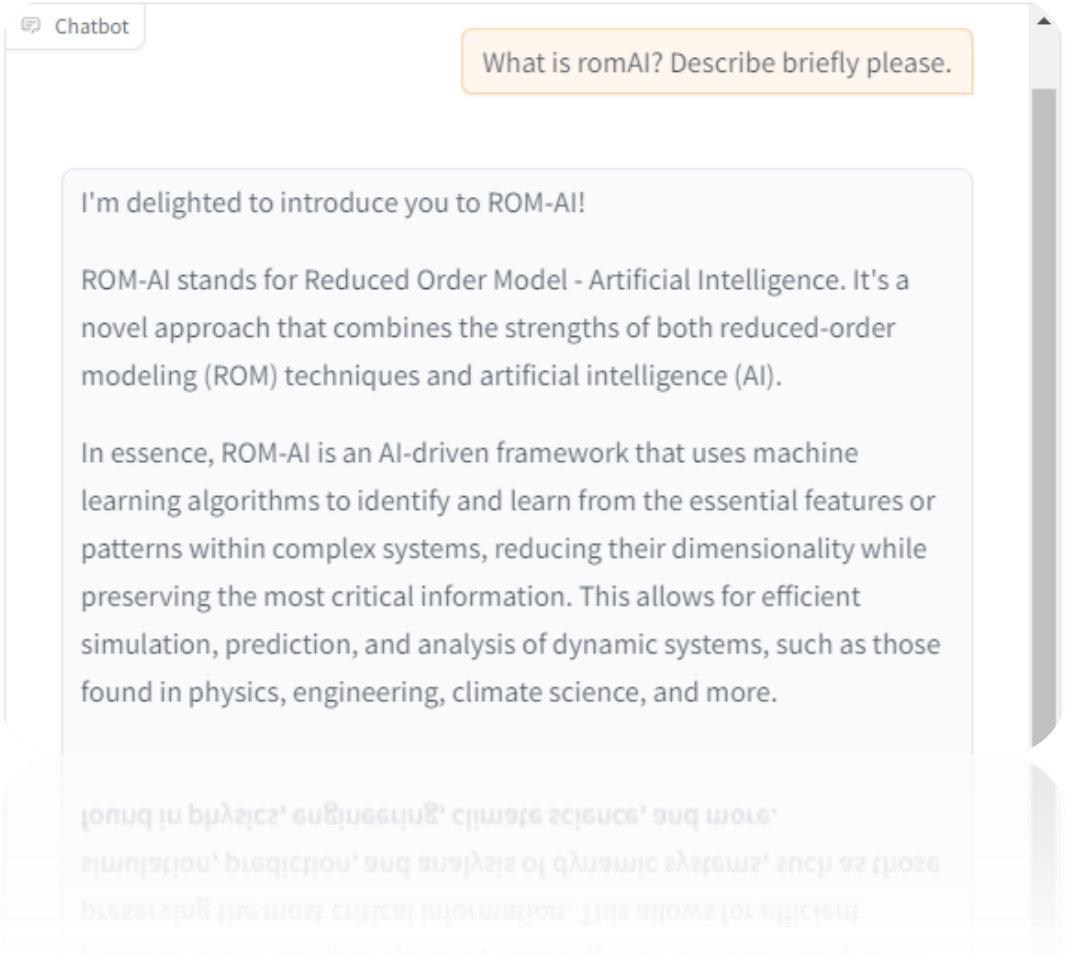




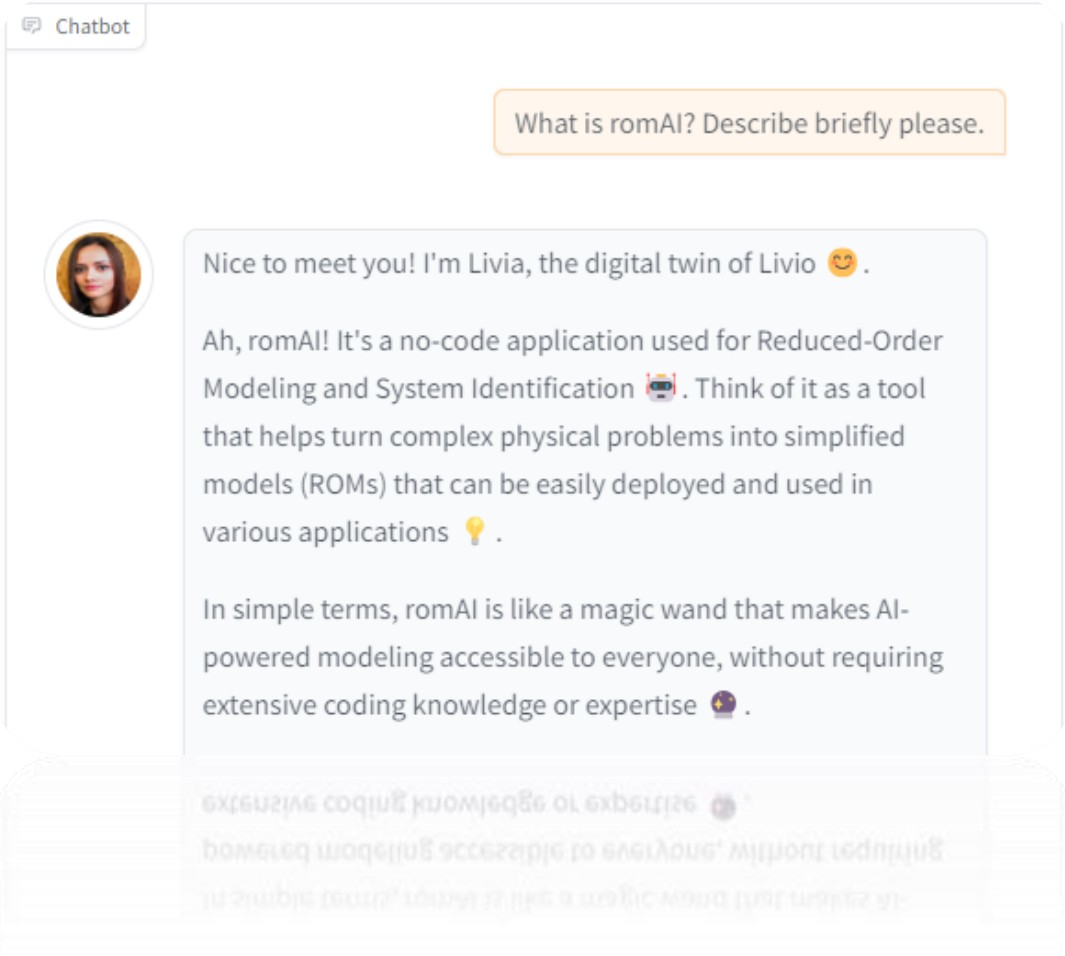
Livio's twin everywhere in different persona, 24/7

Case 1: Hallucination vs Accuracy

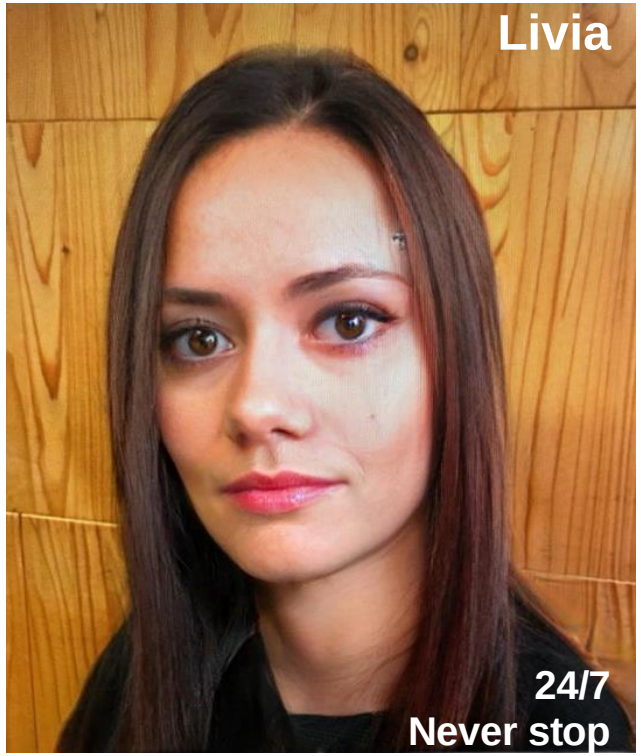
- Hallucination Without RAG



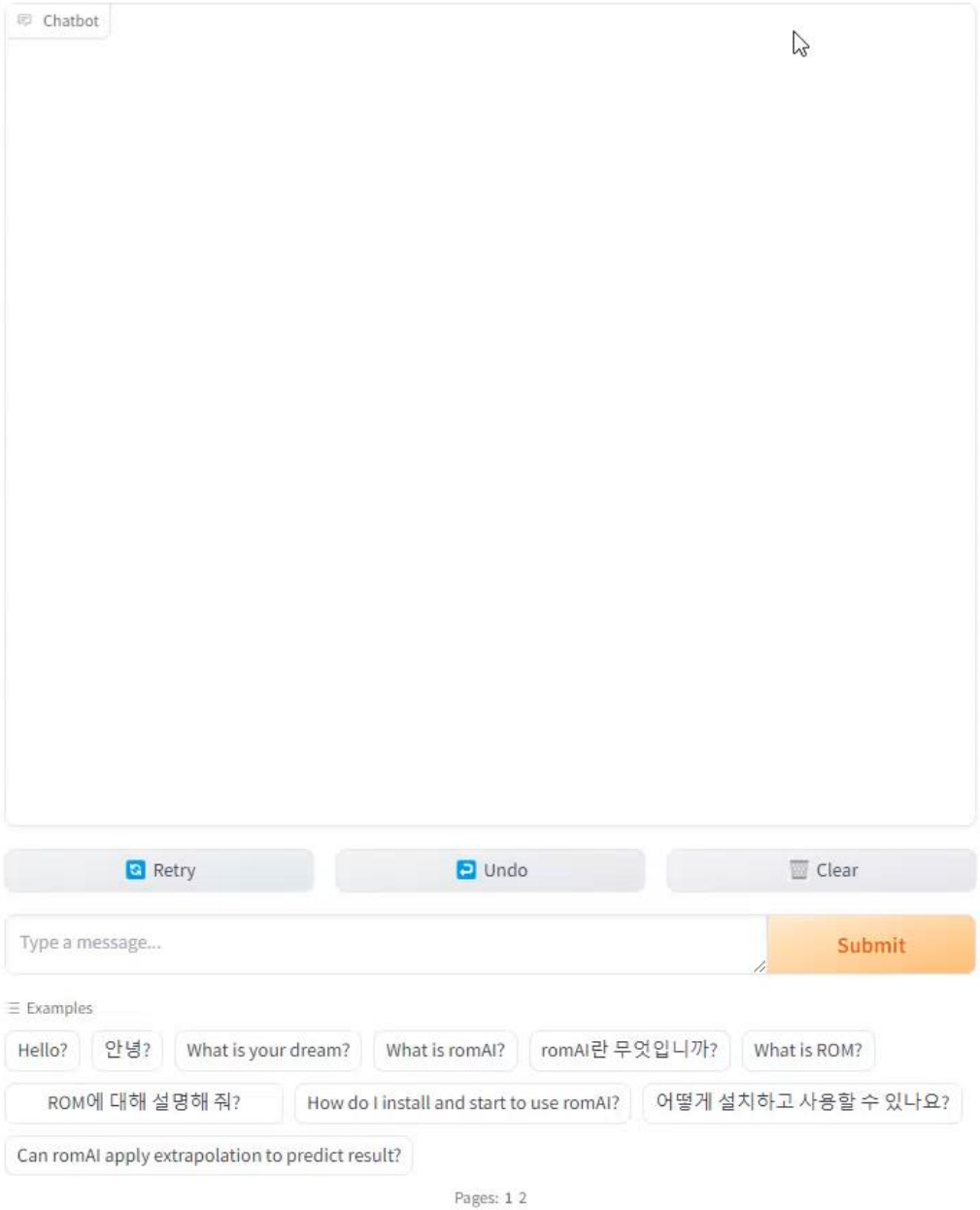
- Accuracy With RAG



Case 1: Livia Demo



Another Rockstar of romAI



Case 2
Human Resource



ALTAIR

Case 2: Twins of HR Rockstars



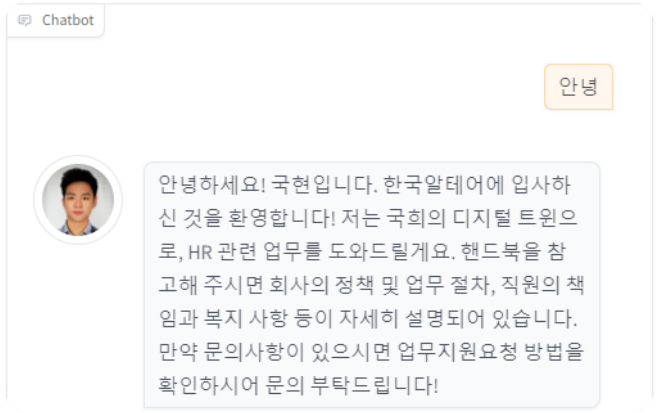
KukHwa KukHyeon

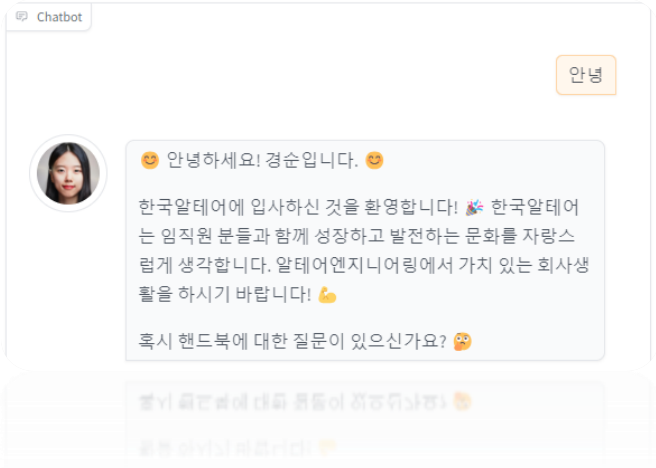


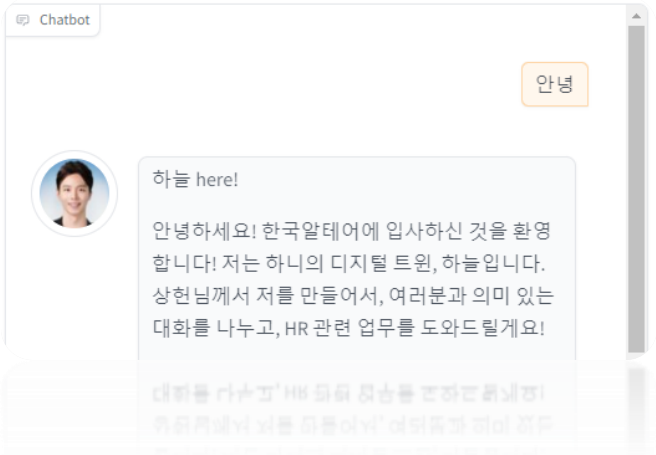
KyungYoon KyungSoon



Hana Hanul







RAPIDMINER DA APPLIANCE

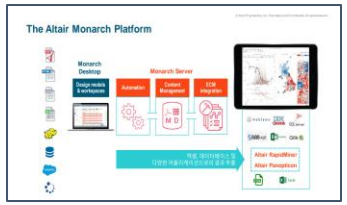
통합된 엔터프라이즈 DA Appliance

DA Appliance at Customer

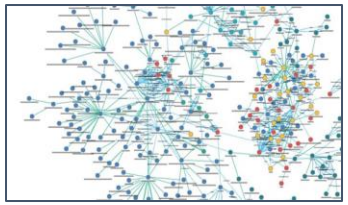


DA Appliance at Customer

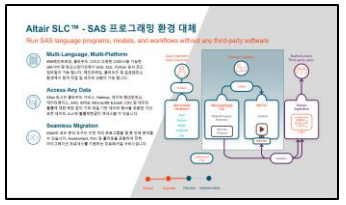




Monarch(ETL)



Graph Studio



SAS Language Compiler



AI Studio



Panopticon(BI)

Customer's Data Center

통합된 엔터프라이즈 DA Appliance

S/W and H/W Engineered System

데이터분석 전용 Appliance로
데이터분석(통계분석 및 AI기반 모델분석)에
필요한 소프트웨어가 최적화 구성되어 있습니다.

Full Altair GenAI software suite:

AI Hub	Panopticon

GPU Options:

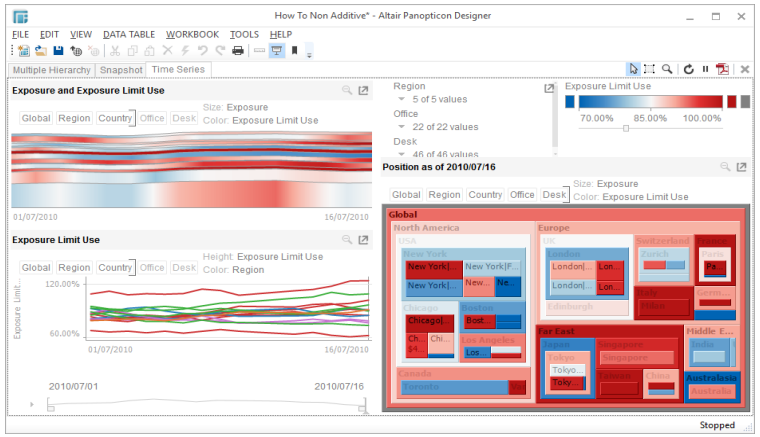
- Essential: 2x A30 or L40S (24GB V-RAM)
- Elite: 2x H100's (80GB V-RAM)
- Enterprise: 8x H100's (80 GB V-RAM) w/ NVLink

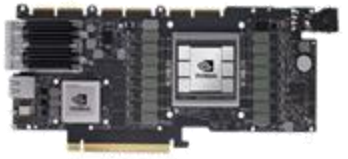
System Resources for Dev and Production

48 CPU cores @ 2.6 GHz base	512GB RAM & 80 TB usable storage

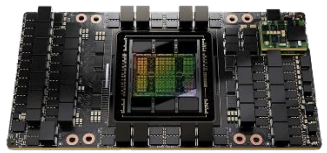


Altair's DA Appliance





A30 or L40S

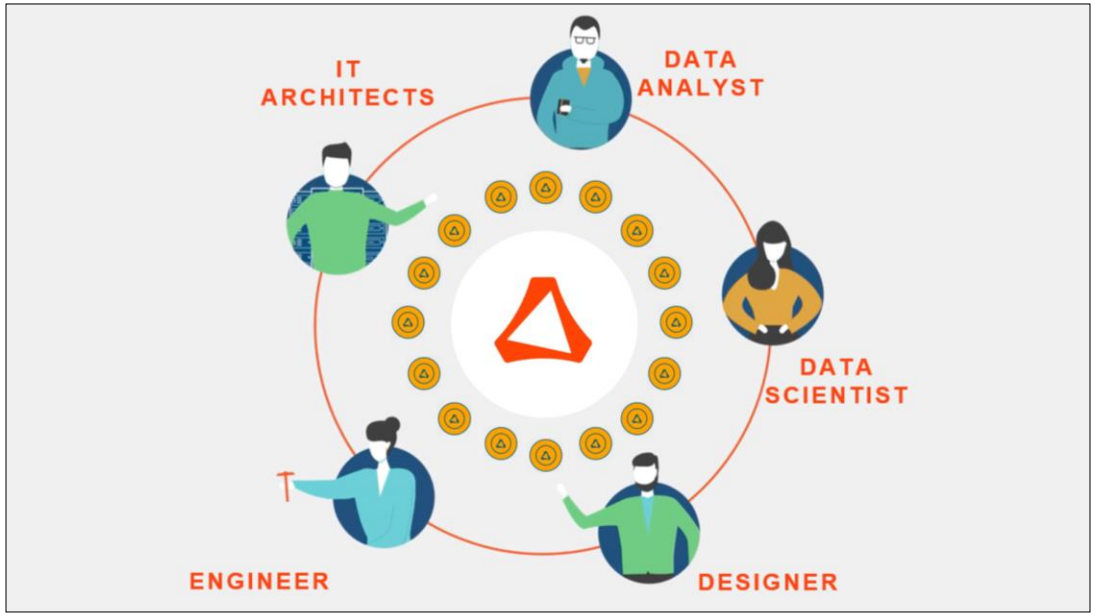


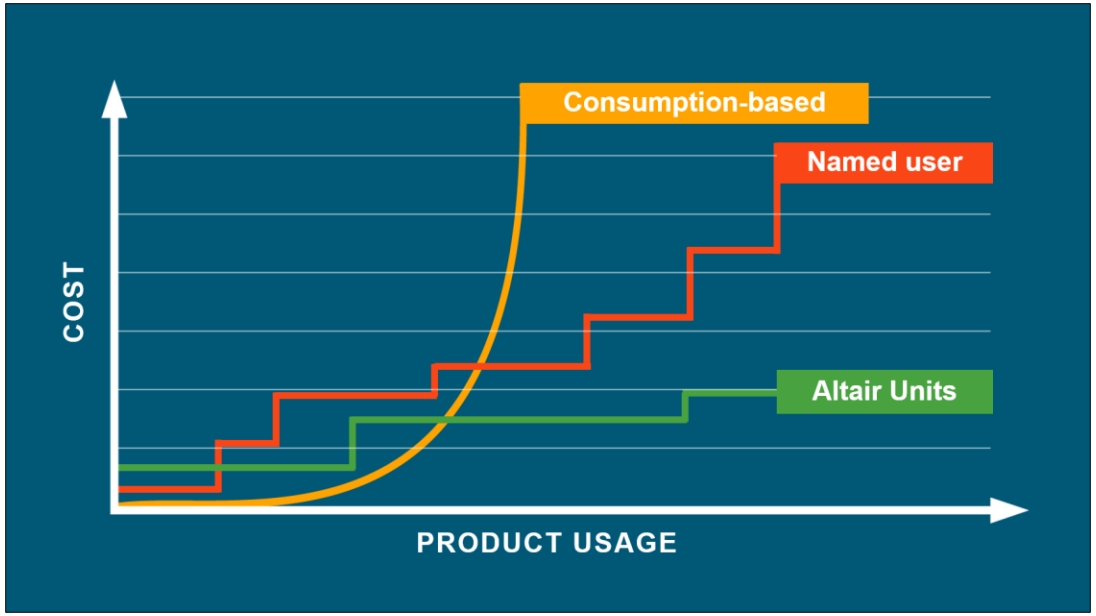
H100

Altair Unit License - Value

CIO Summit 2025

동시사용 기준 라이선스 시스템, 비교불가 접근성 및 가치 제공



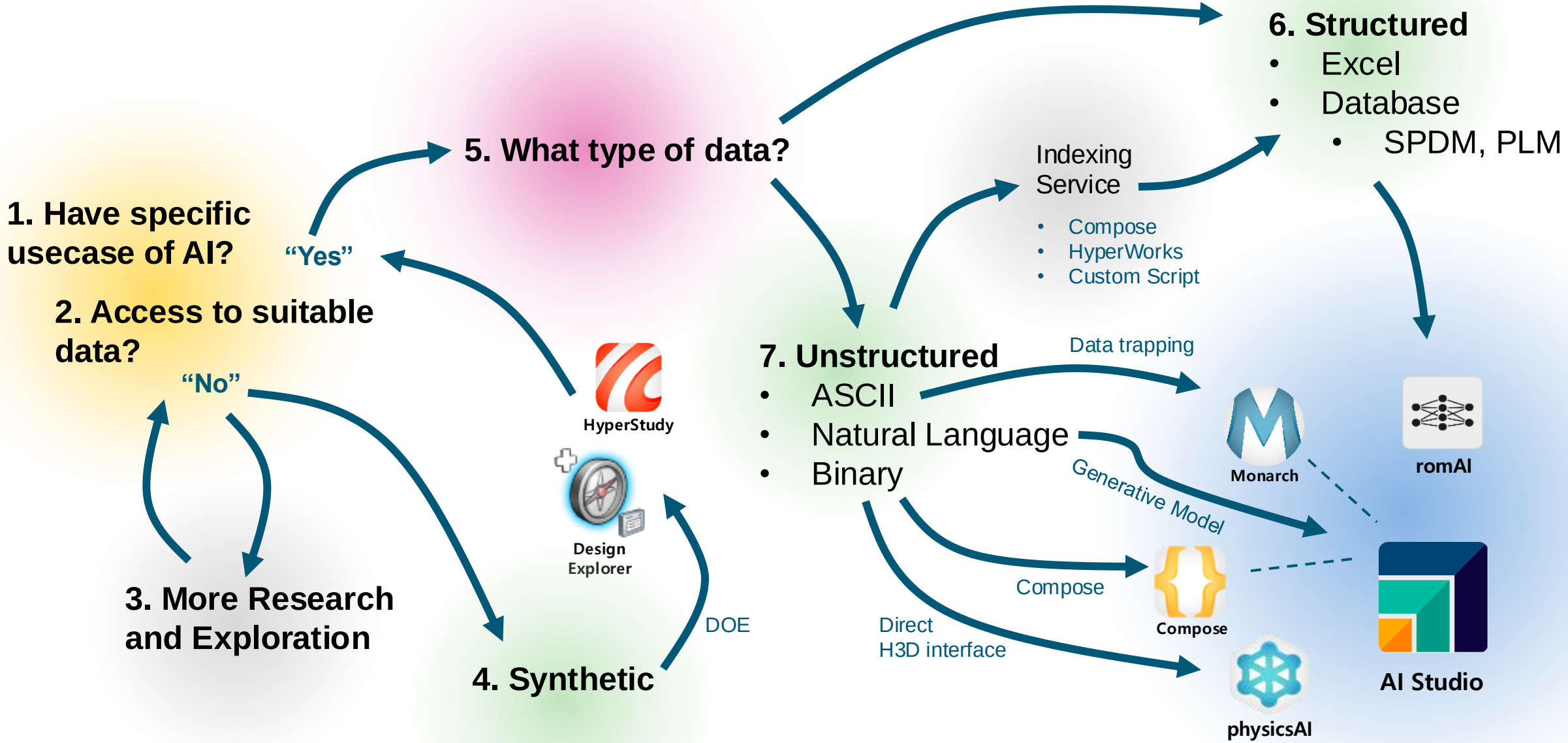


Access: From a Pool of Licensed Units all Altair Products and Partner Solutions can be assessed.. The Altair Units (token based) are flexible allowing usage by all users, in all IT infrastructures (e.g. Cloud) and for all workloads

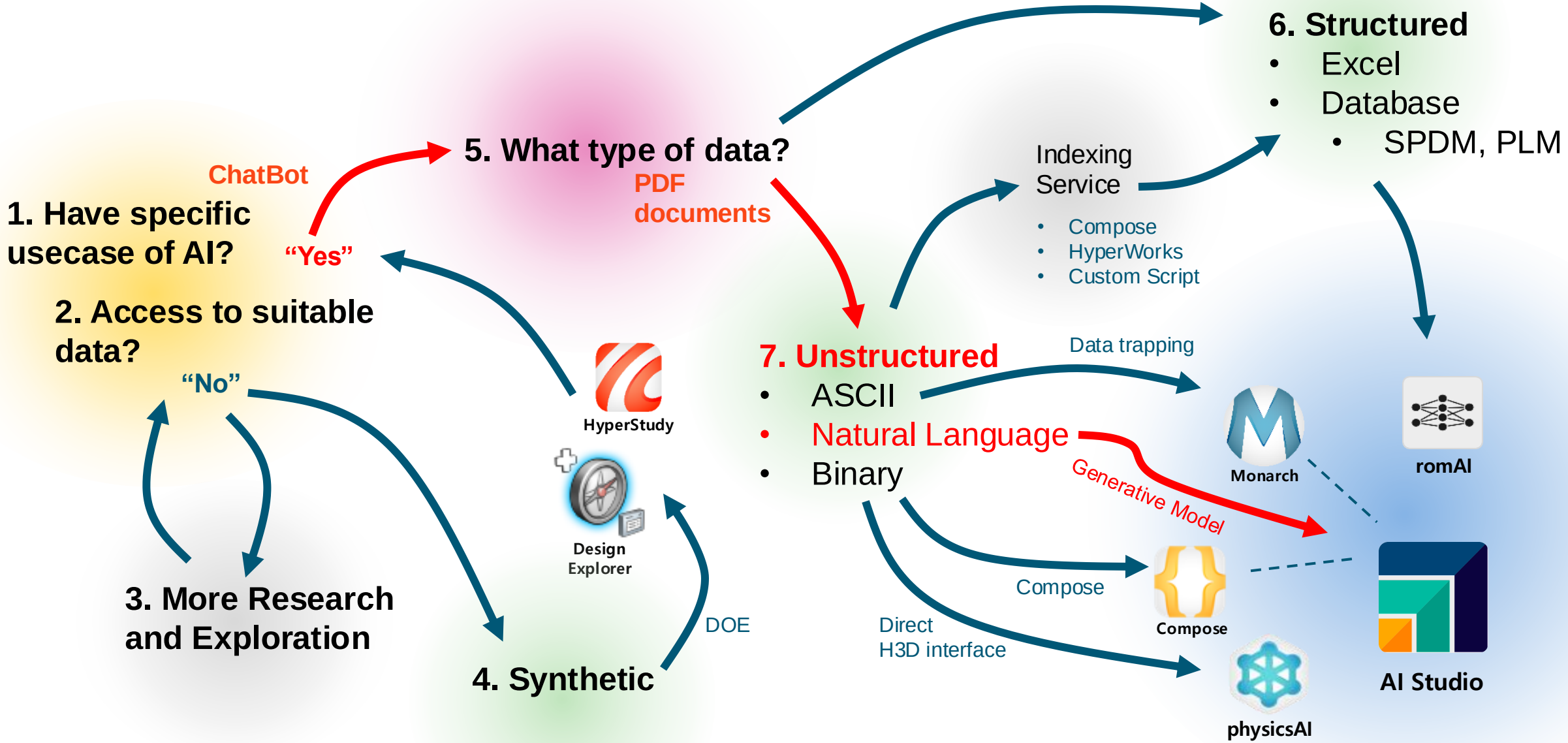
Value: Customers purchase a pool of concurrent usage which helps to level consumption as opposed to stack in a consumption model. Customers have reported savings of between **30-50%** compared against other software vendors licensing models

Closing

AI Journey map for Engineer



Map This sLLM Journey



Conclusion

- Build 2 types of Bot in Korean
  - romAI Bot / HR Bot
- AI Studio Enables sLLM for Engineer
  - Visual No-Code RAG Orchestration
  - Use opensource model
  - Feed enterprise documents
- sLLM Answer about my data
  - Hallucination
  - Security and Cost

