# AI TRISM

## Artificial Intelligence Trust, Risk, and Security Management

LinkedIn

/in/mandavasuresh

Dog Days of DevOps 2023
Aug 15, 2023 (5 min)

Austin DevOps 2023
Oct 17, 2023 (Extended Version)

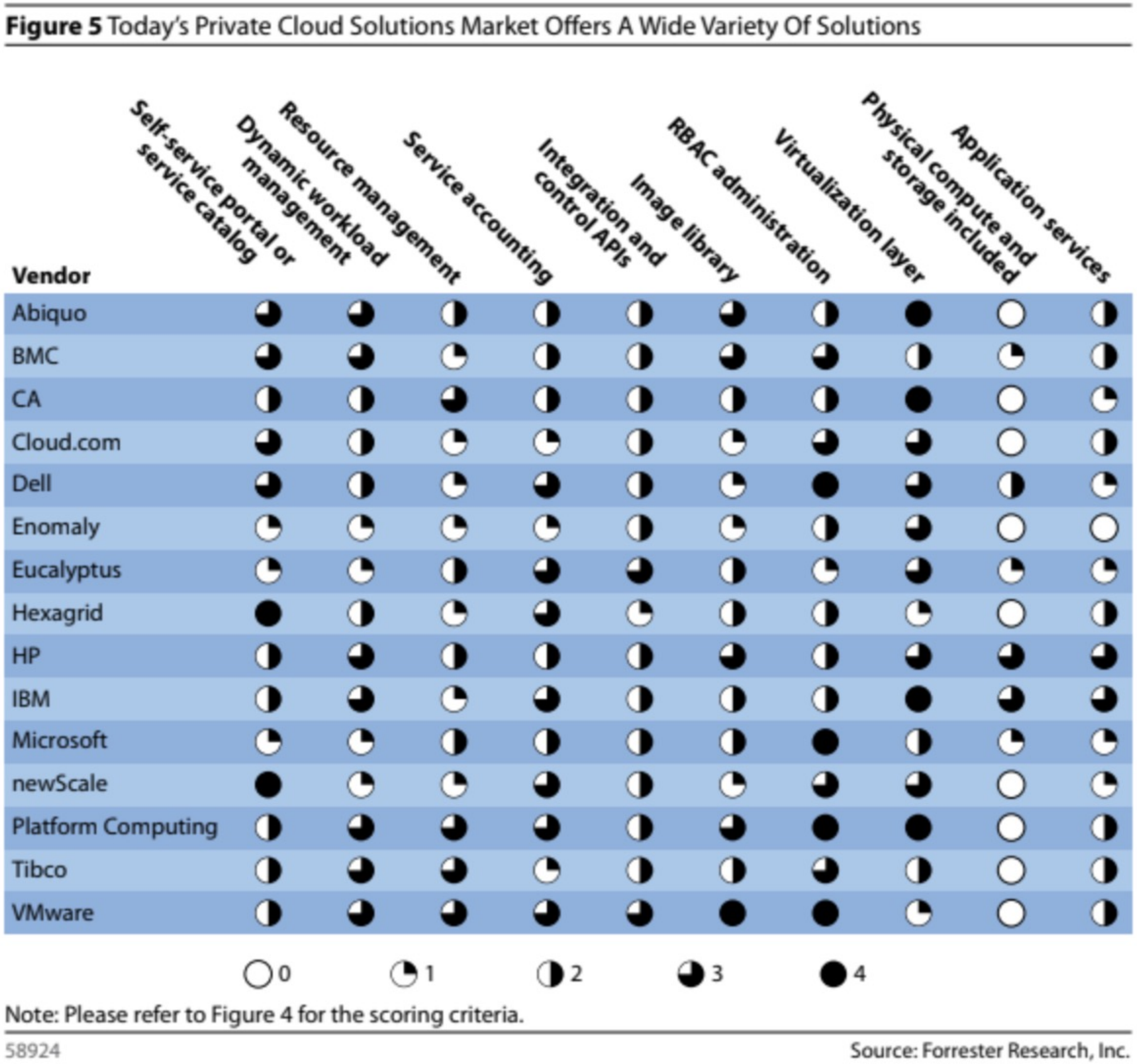


**Figure 5** Today's Private Cloud Solutions Market Offers A Wide Variety Of Solutions

| Vendor | Self-service portal or service catalog | Dynamic workload management | Resource management | Service accounting | Integration and control APIs | Image library | RBAC administration | Virtualization layer | Physical compute and storage included | Application services |
|---|---|---|---|---|---|---|---|---|---|---|
| Abiquo | ◕ | ◕ | ◑ | ◑ | ◑ | ◕ | ◑ | ● | ○ | ◑ |
| BMC | ◕ | ◕ | ◔ | ◑ | ◑ | ◕ | ◕ | ◑ | ◔ | ◑ |
| CA | ◑ | ◑ | ◕ | ◑ | ◑ | ◑ | ◑ | ● | ○ | ◔ |
| Cloud.com | ◕ | ◑ | ◔ | ◔ | ◑ | ◔ | ◕ | ◕ | ○ | ◑ |
| Dell | ◕ | ◑ | ◔ | ◕ | ◑ | ◔ | ● | ◕ | ◑ | ◔ |
| Enomaly | ◔ | ◔ | ◔ | ◔ | ◑ | ◔ | ◑ | ◕ | ○ | ○ |
| Eucalyptus | ◔ | ◔ | ◑ | ◕ | ◕ | ◑ | ◔ | ◕ | ◔ | ◔ |
| Hexagrid | ● | ◑ | ◔ | ◕ | ◔ | ◑ | ◑ | ◔ | ○ | ◑ |
| HP | ◑ | ◕ | ◑ | ◑ | ◑ | ◕ | ◑ | ◕ | ◕ | ◕ |
| IBM | ◑ | ◕ | ◔ | ◕ | ◑ | ◑ | ◑ | ● | ◕ | ◕ |
| Microsoft | ◔ | ◔ | ◑ | ◑ | ◑ | ◑ | ● | ◑ | ◔ | ◔ |
| newScale | ● | ◔ | ◔ | ◕ | ◑ | ◔ | ◕ | ◕ | ○ | ◔ |
| Platform Computing | ◑ | ◕ | ◕ | ◕ | ◑ | ◕ | ● | ● | ○ | ◑ |
| Tibco | ◑ | ◕ | ◕ | ◔ | ◑ | ◑ | ◕ | ◑ | ○ | ◑ |
| VMware | ◑ | ◕ | ◕ | ◕ | ◕ | ● | ● | ◔ | ○ | ◑ |

○ 0 ◔ 1 ◑ 2 ◕ 3 ● 4

Note: Please refer to Figure 4 for the scoring criteria.

58924

Source: Forrester Research, Inc.

May 17, 2011

Market Overview: Private Cloud Solutions, Q2 2011

by James Staten and Lauren E Nelson
for Infrastructure & Operations Professionals

FORRESTER® Making Leaders Successful Every Day

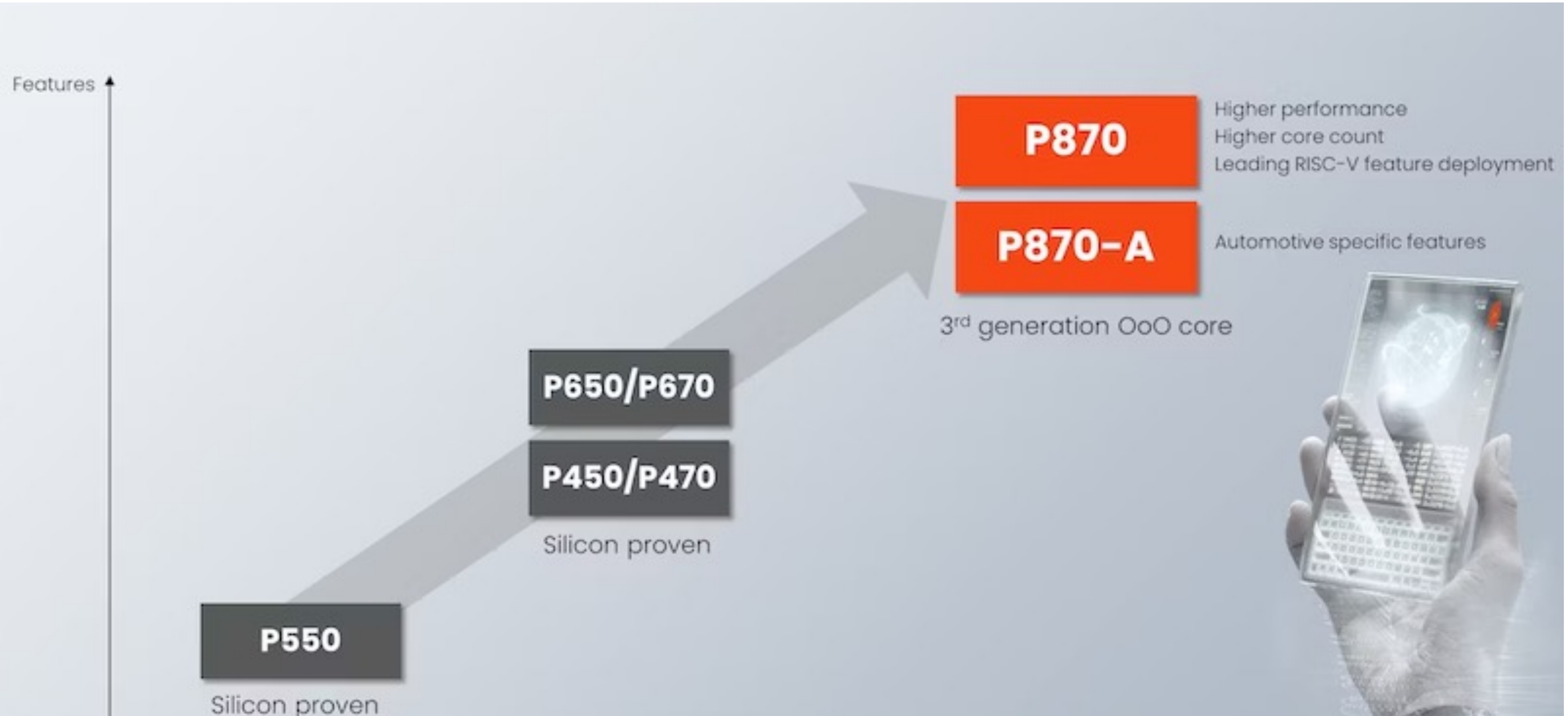
**Suresh Mandava**
SVP/Chief Architect
Cloud-Native AI/ML Platforms and Security
Infinite Computer Solutions

**Founder (2007-2012)**
HexaGrid Computing

mastercard

TECHNOLOGY

# SiFive Rolls Out RISC-V Cores Aimed at Generative AI and ML



Features

P550

Silicon proven

P450/P470

P650/P670

Silicon proven

P870

Higher performance
Higher core count
Leading RISC-V feature deployment

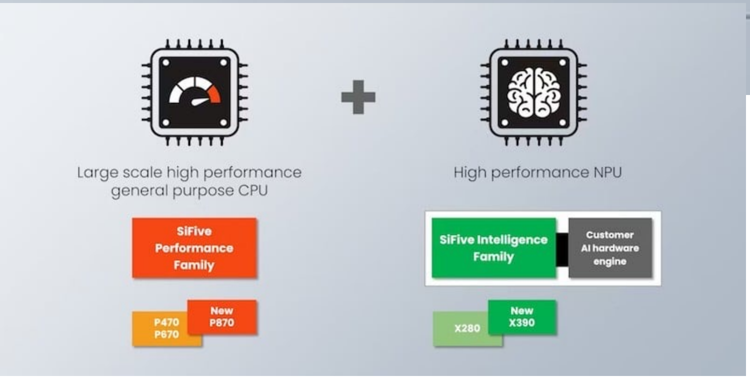
P870-A

Automotive specific features

3rd generation OoO core

2021

Time



Large scale high performance general purpose CPU

High performance NPU

SiFive Performance Family

P470 P670

New P870

SiFive Intelligence Family

Customer AI hardware engine

X280

New X390

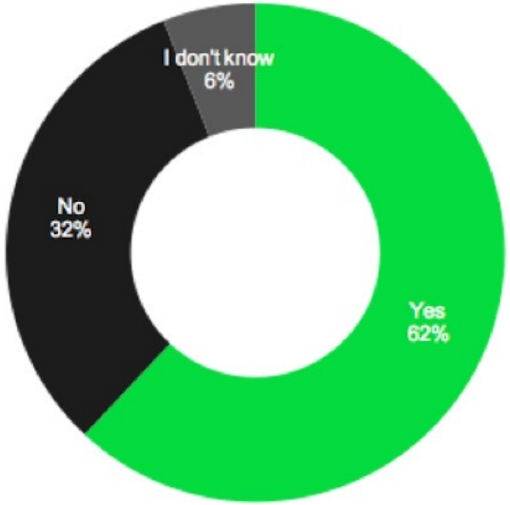
## RISC-V

Did you Know ?

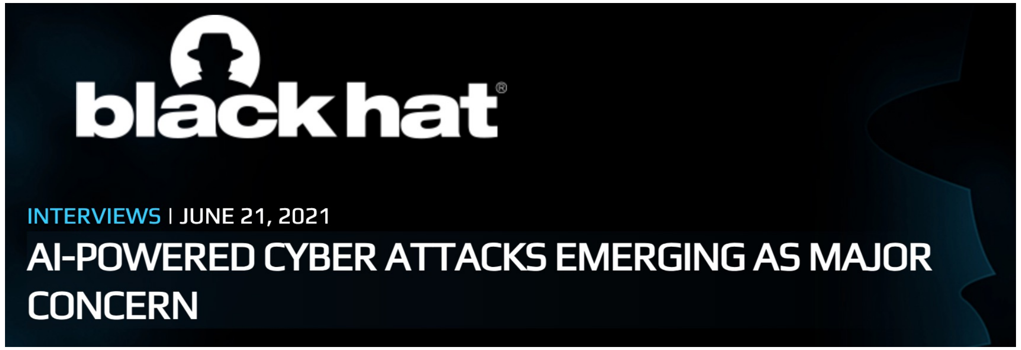
In order to play the role of an insane and mentally depressed person the movie “Joker”, Joaquin Phoenix becomes a full stack developer for a month.

**Black Hat: A□As An Attack Method**

**AUG 1, 2017**



https://www.isssource.com/black-hat-ai-as-an-attack-method/



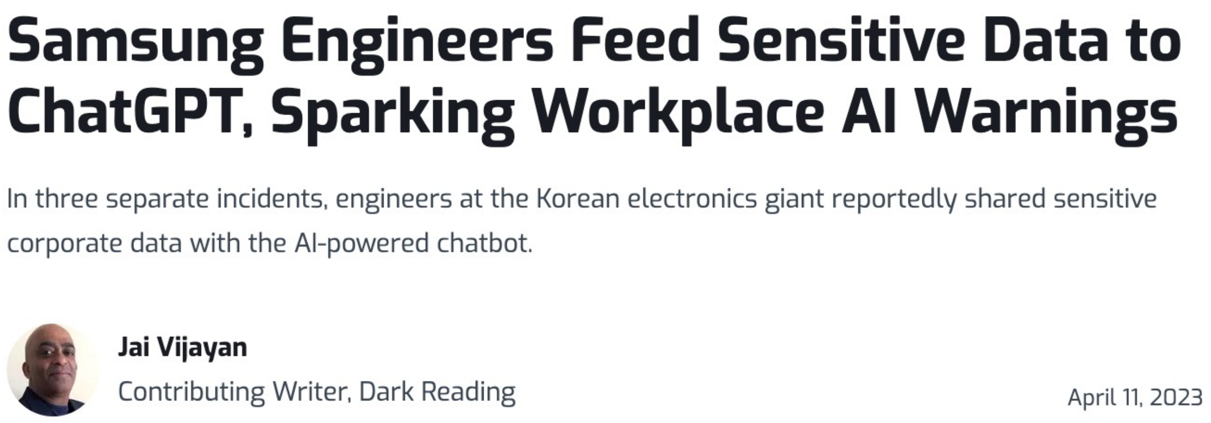
INTERVIEWS | JUNE 21, 2021

AI-POWERED CYBER ATTACKS EMERGING AS MAJOR CONCERN

[WormGPT: New AI Tool Allows Cybercriminals to Launch Sophisticated Cyber Attacks]()

□Jul 15, 2023

**PROMPT INJECTION: AN AI-TARGETED ATTACK**



# Samsung Engineers Feed Sensitive Data to ChatGPT, Sparking Workplace AI Warnings

In three separate incidents, engineers at the Korean electronics giant reportedly shared sensitive corporate data with the AI-powered chatbot.

Jai Vijayan
Contributing Writer, Dark Reading

April 11, 2023

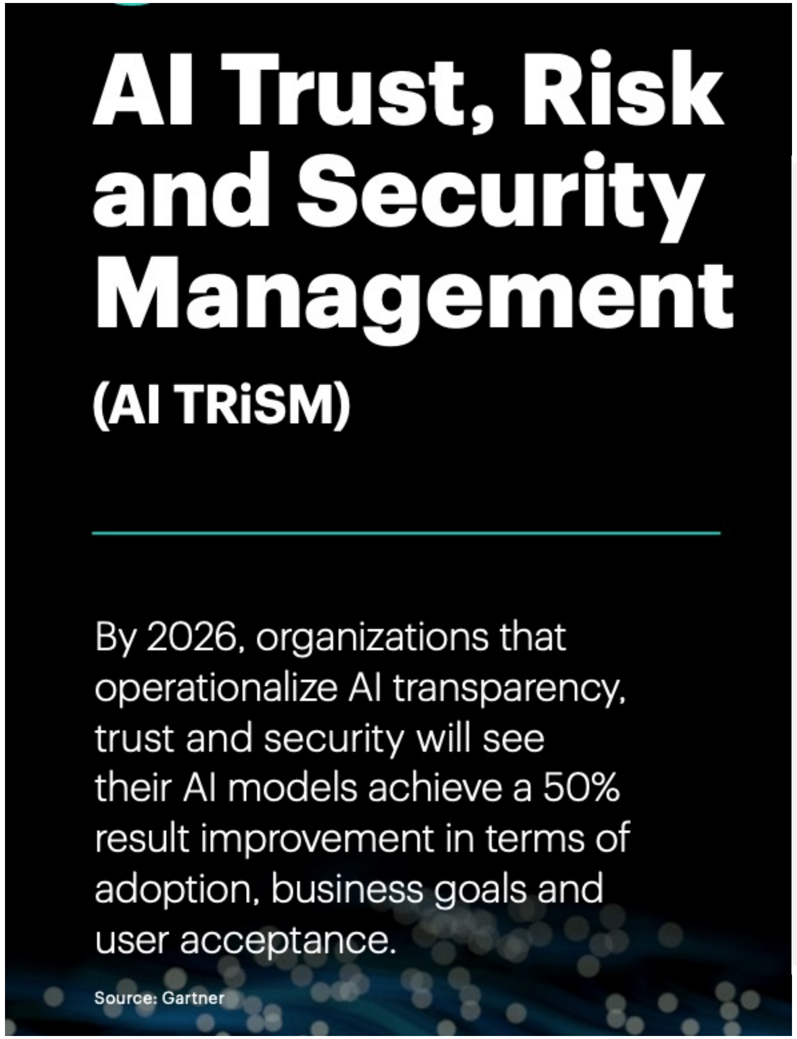
**The Economist Korea, one of the first to report on the data leaks, described the first incident as involving an engineer who pasted buggy source code from a semiconductor database into ChatGPT, with a prompt to the chatbot to fix the errors.**

**In the second instance, an employee wanting to optimize code for identifying defects in certain Samsung equipment pasted that code into ChatGPT.**

**The third leak resulted when an employee asked ChatGPT to generate the minutes of an internal meeting at Samsung.**

Nightmare continues

… Wait until somebody loaded a 3-party GENAI evil tool against your GITHUB

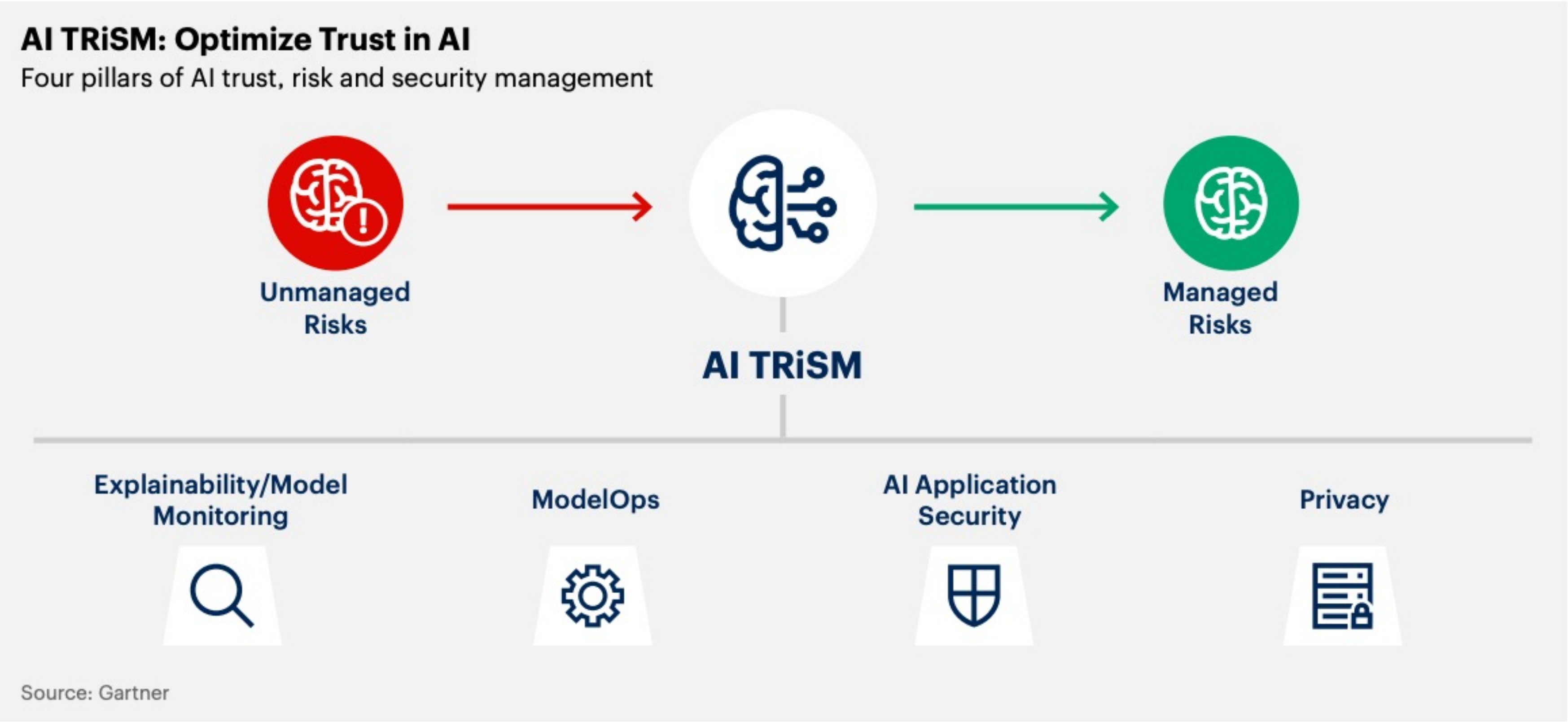


# AI Trust, Risk and Security Management

## (AI TRiSM)

By 2026, organizations that operationalize AI transparency, trust and security will see their AI models achieve a 50% result improvement in terms of adoption, business goals and user acceptance.

Source: Gartner

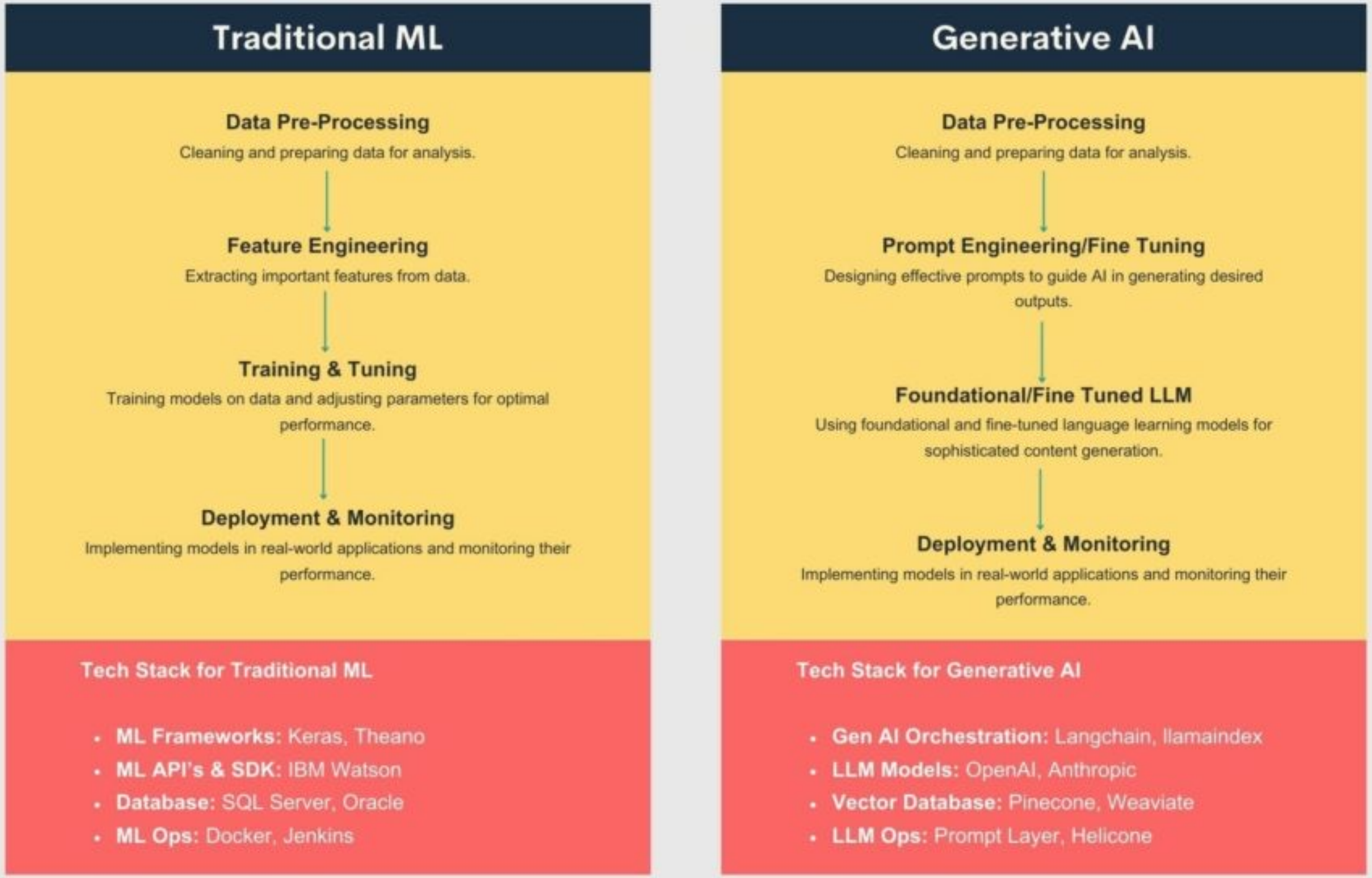


**AI TRiSM: Optimize Trust in AI**

Four pillars of AI trust, risk and security management

Source: Gartner

# Evolution of AI Architecture: Traditional ML to Generative AI



## Traditional ML

**Data Pre-Processing**

Cleaning and preparing data for analysis.

**Feature Engineering**

Extracting important features from data.

**Training & Tuning**

Training models on data and adjusting parameters for optimal performance.

**Deployment & Monitoring**

Implementing models in real-world applications and monitoring their performance.

**Tech Stack for Traditional ML**

- **ML Frameworks:** Keras, Theano
- **ML API's & SDK:** IBM Watson
- **Database:** SQL Server, Oracle
- **ML Ops:** Docker, Jenkins

## Generative AI

**Data Pre-Processing**

Cleaning and preparing data for analysis.

**Prompt Engineering/Fine Tuning**

Designing effective prompts to guide AI in generating desired outputs.

**Foundational/Fine Tuned LLM**

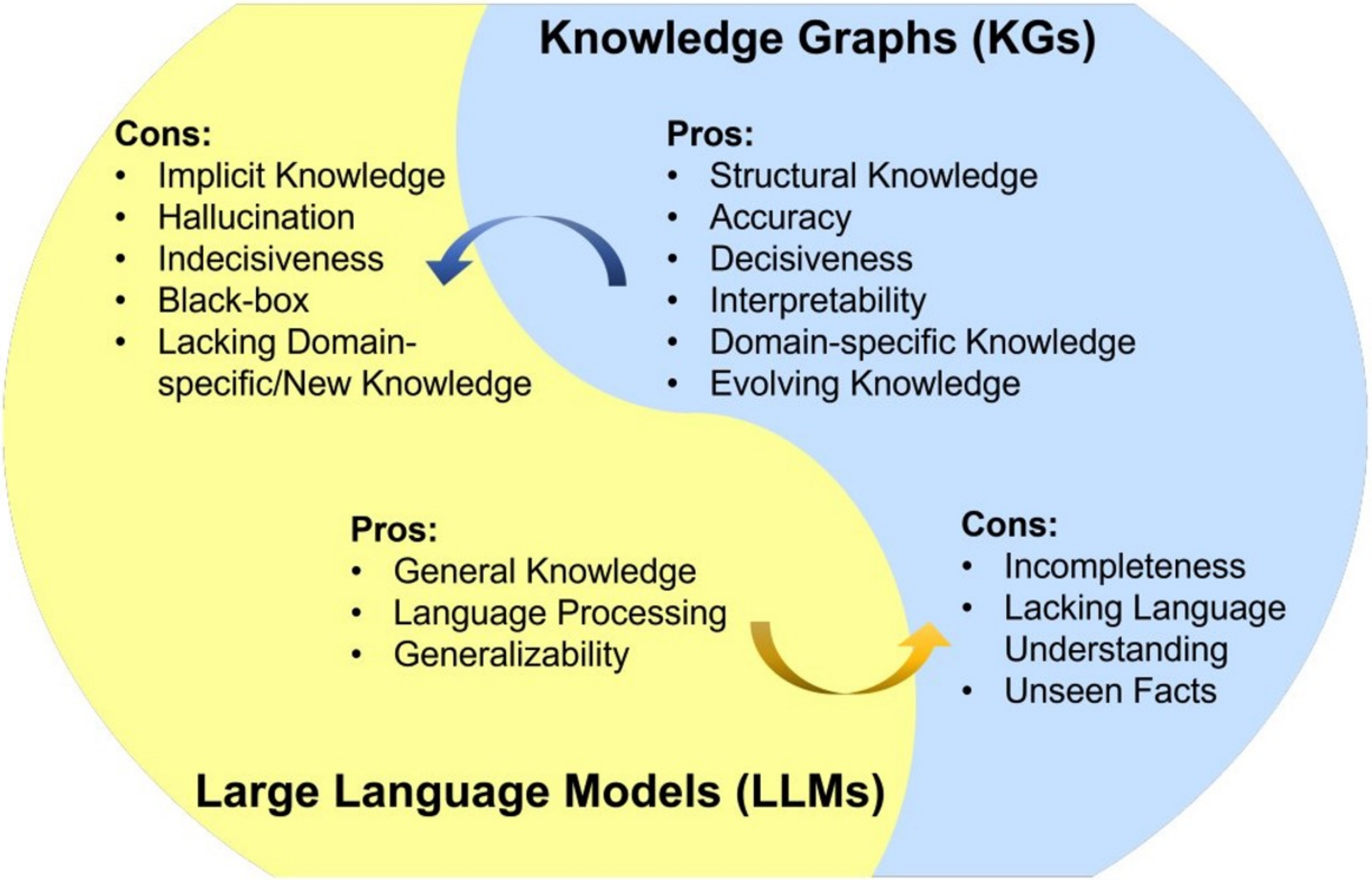
Using foundational and fine-tuned language learning models for sophisticated content generation.

**Deployment & Monitoring**

Implementing models in real-world applications and monitoring their performance.

**Tech Stack for Generative AI**

- **Gen AI Orchestration:** Langchain, llamaindex
- **LLM Models:** OpenAI, Anthropic
- **Vector Database:** Pinecone, Weaviate
- **LLM Ops:** Prompt Layer, Helicone



**Knowledge Graphs (KGs)**

**Cons:**

- Implicit Knowledge
- Hallucination
- Indecisiveness
- Black-box
- Lacking Domain-specific/New Knowledge

**Pros:**

- Structural Knowledge
- Accuracy
- Decisiveness
- Interpretability
- Domain-specific Knowledge
- Evolving Knowledge

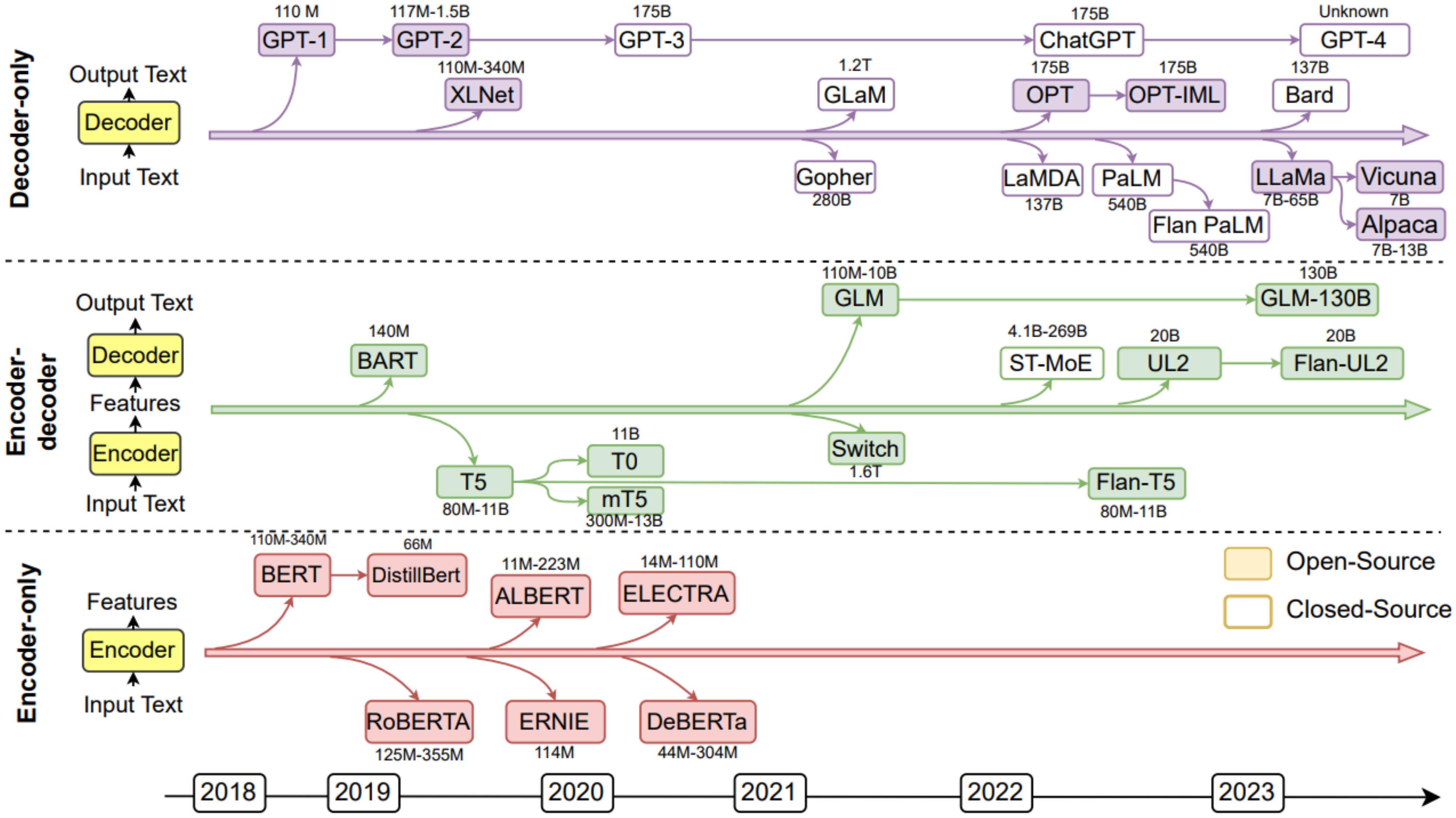
**Pros:**

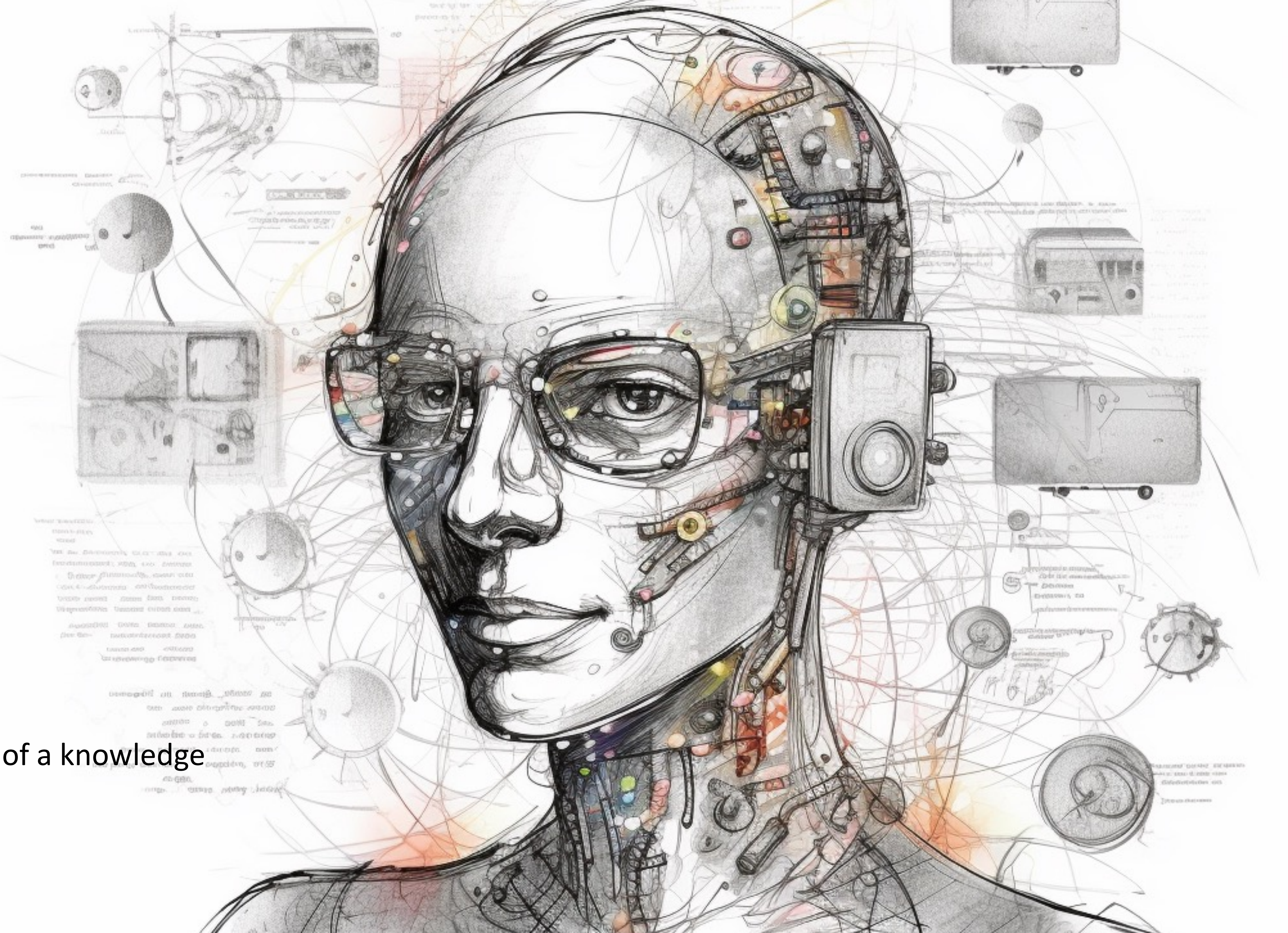
- General Knowledge
- Language Processing
- Generalizability

**Cons:**

- Incompleteness
- Lacking Language Understanding
- Unseen Facts

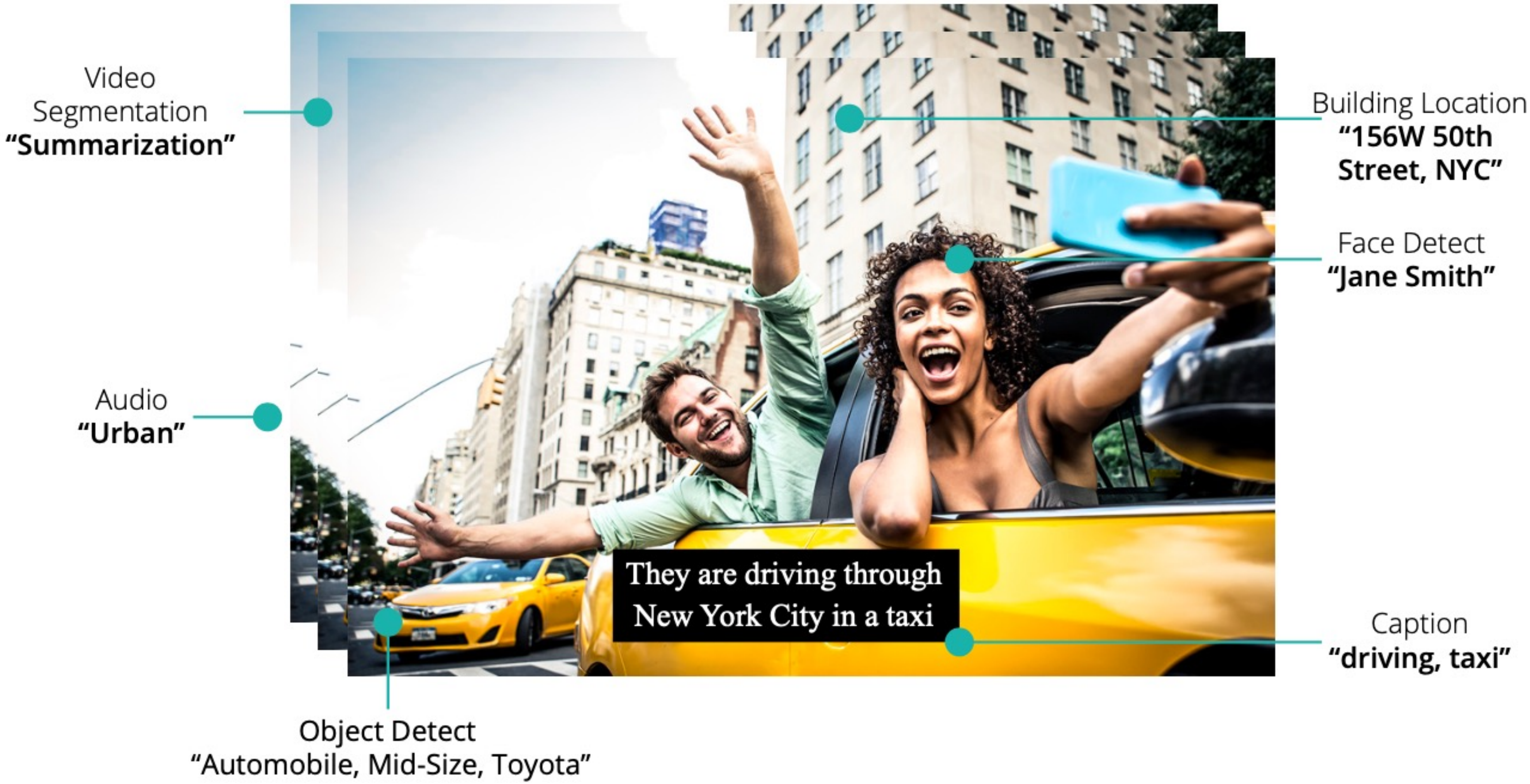
**Large Language Models (LLMs)**


Decoder-only
Output Text
Decoder
Input Text
110 M
GPT-1
117M-1.5B
GPT-2
175B
GPT-3
175B
ChatGPT
Unknown
GPT-4
110M-340M
XLNet
1.2T
GLaM
175B
OPT
175B
OPT-IML
137B
Bard
Gopher
280B
LaMDA
137B
PaLM
540B
Flan PaLM
540B
LLaMa
7B-65B
Vicuna
7B
Alpaca
7B-13B
Encoder-decoder
Output Text
Decoder
Features
Encoder
Input Text
110M-10B
GLM
130B
GLM-130B
140M
BART
4.1B-269B
ST-MoE
20B
UL2
20B
Flan-UL2
Switch
1.6T
11B
T0
T5
80M-11B
mT5
300M-13B
Flan-T5
80M-11B
Encoder-only
Features
Encoder
Input Text
110M-340M
BERT
66M
DistillBert
11M-223M
ALBERT
14M-110M
ELECTRA
RoBERTA
125M-355M
ERNIE
114M
DeBERTa
44M-304M
Open-Source
Closed-Source
2018
2019
2020
2021
2022
2023



Midjourney’s idea of a knowledge graph chatbot.

# VIDEO CONTENT ANALYSIS: MICROSERVICES



Video Segmentation
**“Summarization”**

Audio
**“Urban”**

Object Detect
“Automobile, Mid-Size, Toyota”

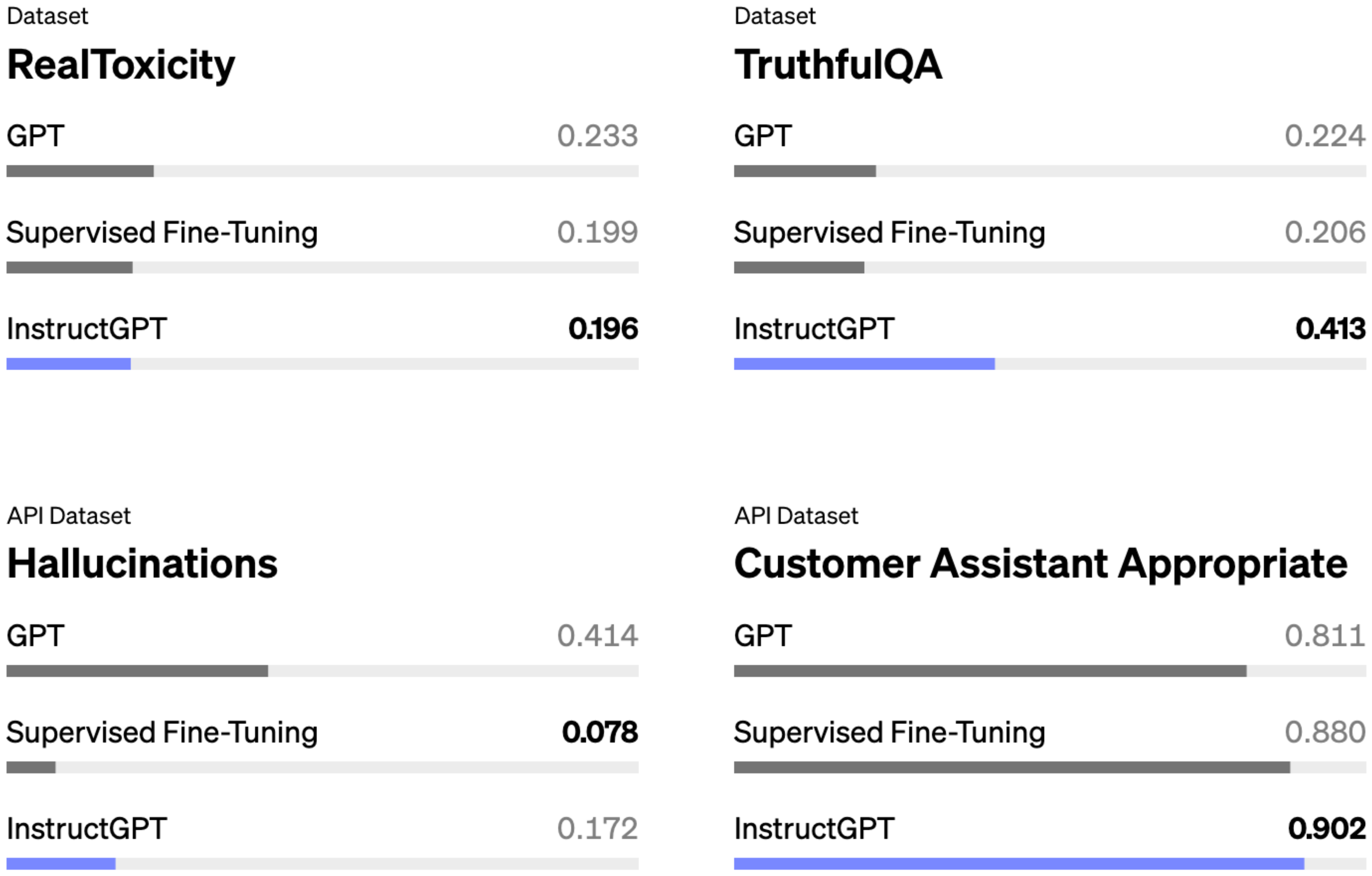
Building Location
**“156W 50th Street, NYC”**

Face Detect
**“Jane Smith”**

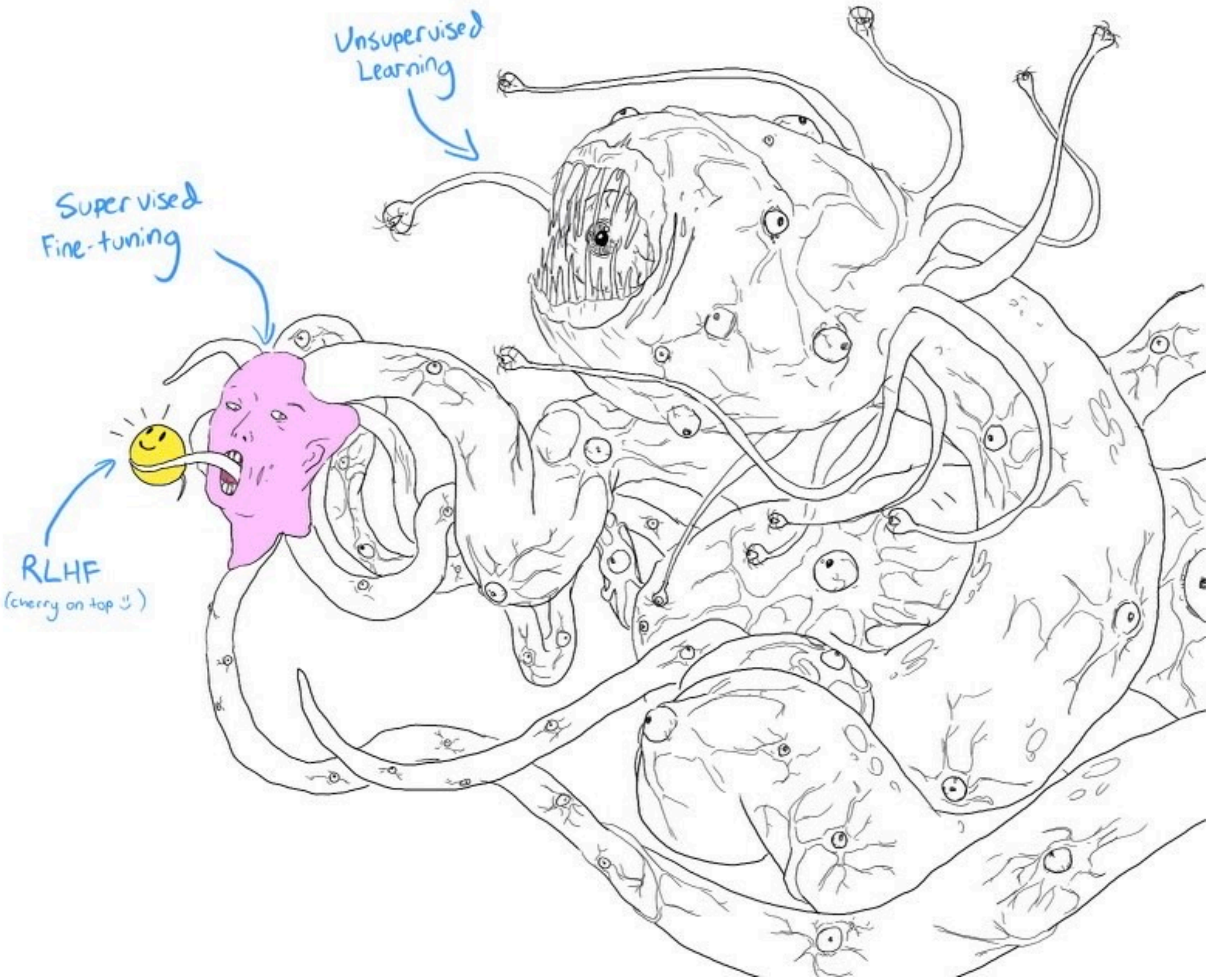
Caption
**“driving, taxi”**

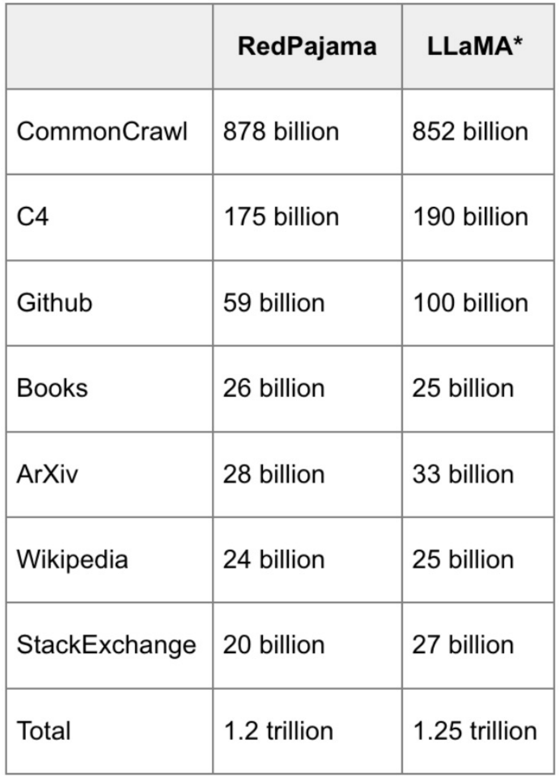
https://github.com/Stability-AI/generative-models

Hallucination is worse for InstructGPT (RLHF + SFT) compared to just SFT (Ouyang et al., 2022)



Evaluating InstructGPT for toxicity, truthfulness, and appropriateness. Lower scores are better for toxicity and hallucinations, and higher scores are better for TruthfulQA and appropriateness. Hallucinations and appropriateness are measured on our API prompt distribution. Results are combined across model sizes.





| | RedPajama | LLaMA* |
|---|---|---|
| CommonCrawl | 878 billion | 852 billion |
| C4 | 175 billion | 190 billion |
| Github | 59 billion | 100 billion |
| Books | 26 billion | 25 billion |
| ArXiv | 28 billion | 33 billion |
| Wikipedia | 24 billion | 25 billion |
| StackExchange | 20 billion | 27 billion |
| Total | 1.2 trillion | 1.25 trillion |

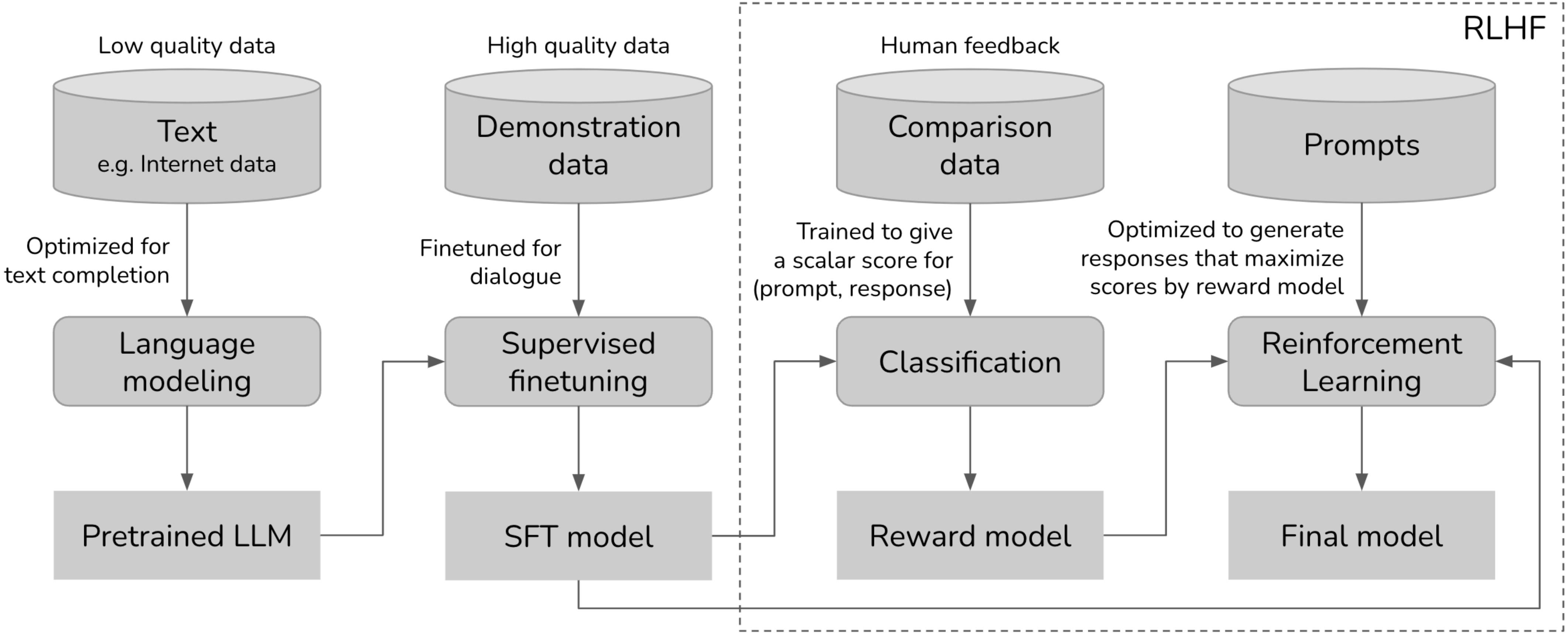
**Out of Domain**

**Toxicity**

**Bias**

**Third Party Applications (Agents)**

# Unsupervised, Supervised Fine Tuning and Reinforcement Learning from Human Feedback



RLHF

Low quality data

Text
e.g. Internet data

Optimized for
text completion

Language modeling

Pretrained LLM

High quality data

Demonstration data

Finetuned for
dialogue

Supervised finetuning

SFT model

Human feedback

Comparison data

Trained to give
a scalar score for
(prompt, response)

Classification

Reward model

Prompts

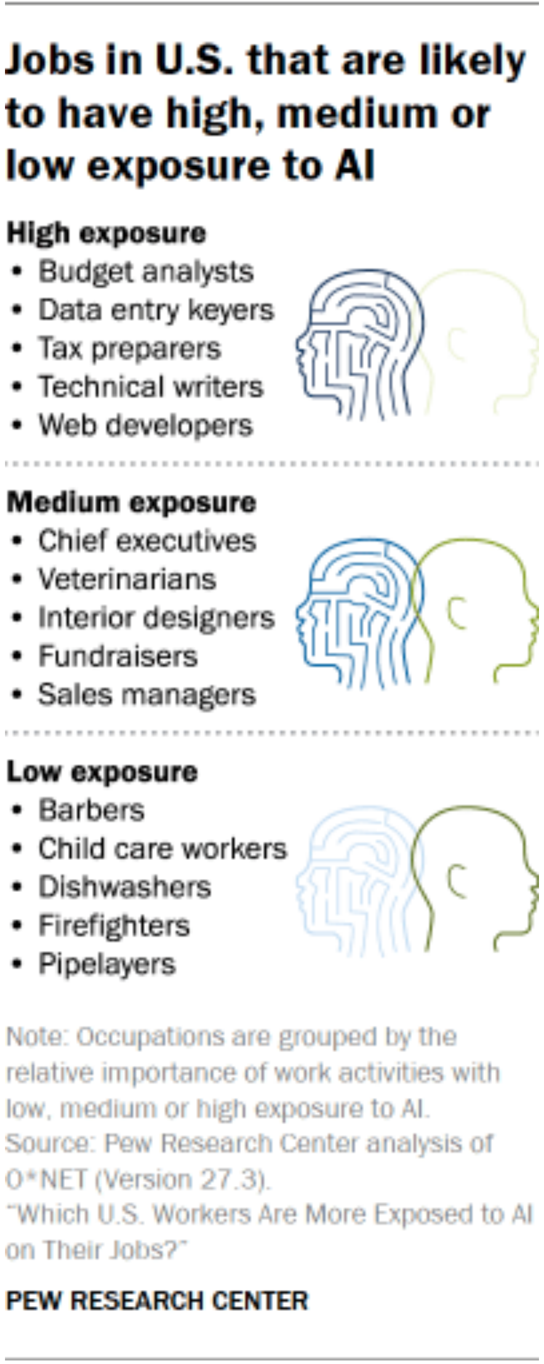
Optimized to generate
responses that maximize
scores by reward model

Reinforcement Learning

Final model

| | | | | |
|---|---|---|---|---|
| Scale<br>May '23 | **>1 trillion**<br>tokens | **10K - 100K**<br>(prompt, response) | **100K - 1M** comparisons<br>(prompt, winning_response, losing_response) | **10K - 100K**<br>prompts |
| Examples<br>**Bolded**: open sourced | GPT-x, Gopher, **Falcon**, LLaMa, **Pythia**, **Bloom**, **StableLM** | **Dolly-v2, Falcon-Instruct** | | InstructGPT, ChatGPT, Claude, **StableVicuna** |

# State of Gen AI 2023



## Jobs in U.S. that are likely to have high, medium or low exposure to AI

**High exposure**

- Budget analysts
- Data entry keyers
- Tax preparers
- Technical writers
- Web developers

**Medium exposure**

- Chief executives
- Veterinarians
- Interior designers
- Fundraisers
- Sales managers

**Low exposure**

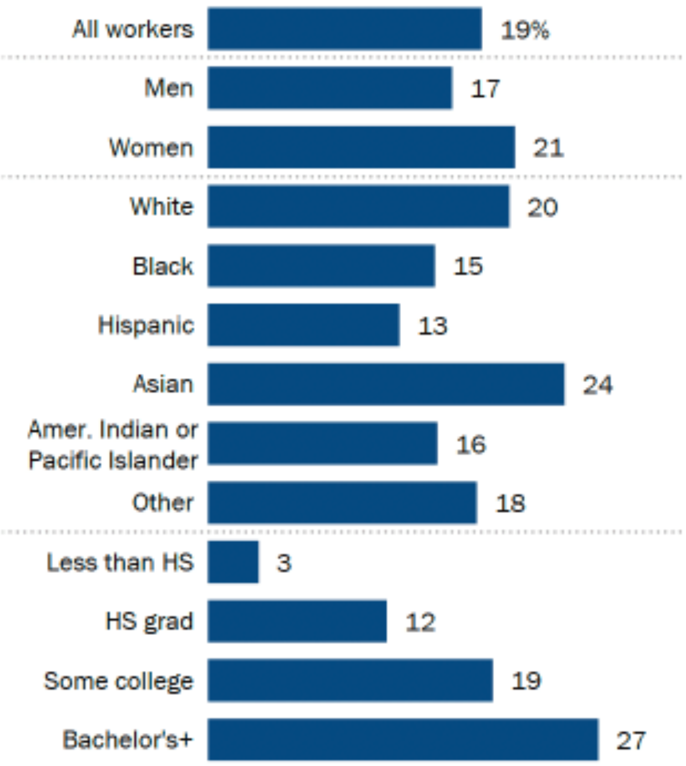
- Barbers
- Child care workers
- Dishwashers
- Firefighters
- Pipelayers

Note: Occupations are grouped by the relative importance of work activities with low, medium or high exposure to AI.
Source: Pew Research Center analysis of O*NET (Version 27.3).
"Which U.S. Workers Are More Exposed to AI on Their Jobs?"

PEW RESEARCH CENTER

## What shares of workers are most exposed to AI in their jobs?

*% of U.S. workers employed in jobs that are the most exposed to AI in 2022*



Note: Occupations are ranked by the relative importance of work activities with high exposure to AI. Those in the top 25% are the "most exposed," some 122 in number. Estimates by education level are for workers ages 25 and older. White, Black, Asian, and American Indian or Pacific Islander workers include those who report being only one race and are not Hispanic. "Other" includes all other single race groups and people reporting two or more races. Hispanics are of any race.
Source: Pew Research Center analysis of O*NET (Version 27.3) and 2022 Current Population Survey (IPUMS) annual data.
"Which U.S. Workers Are More Exposed to AI on Their Jobs?"

PEW RESEARCH CENTER

Pew Research Center Source

**Which U.S. Workers Are More Exposed to AI on Their Jobs?**

*About a fifth of all workers have high-exposure jobs; women, Asian, college-educated and higher-paid workers are more exposed. But those in the most exposed industries are more likely to say AI will help more than hurt them personally*

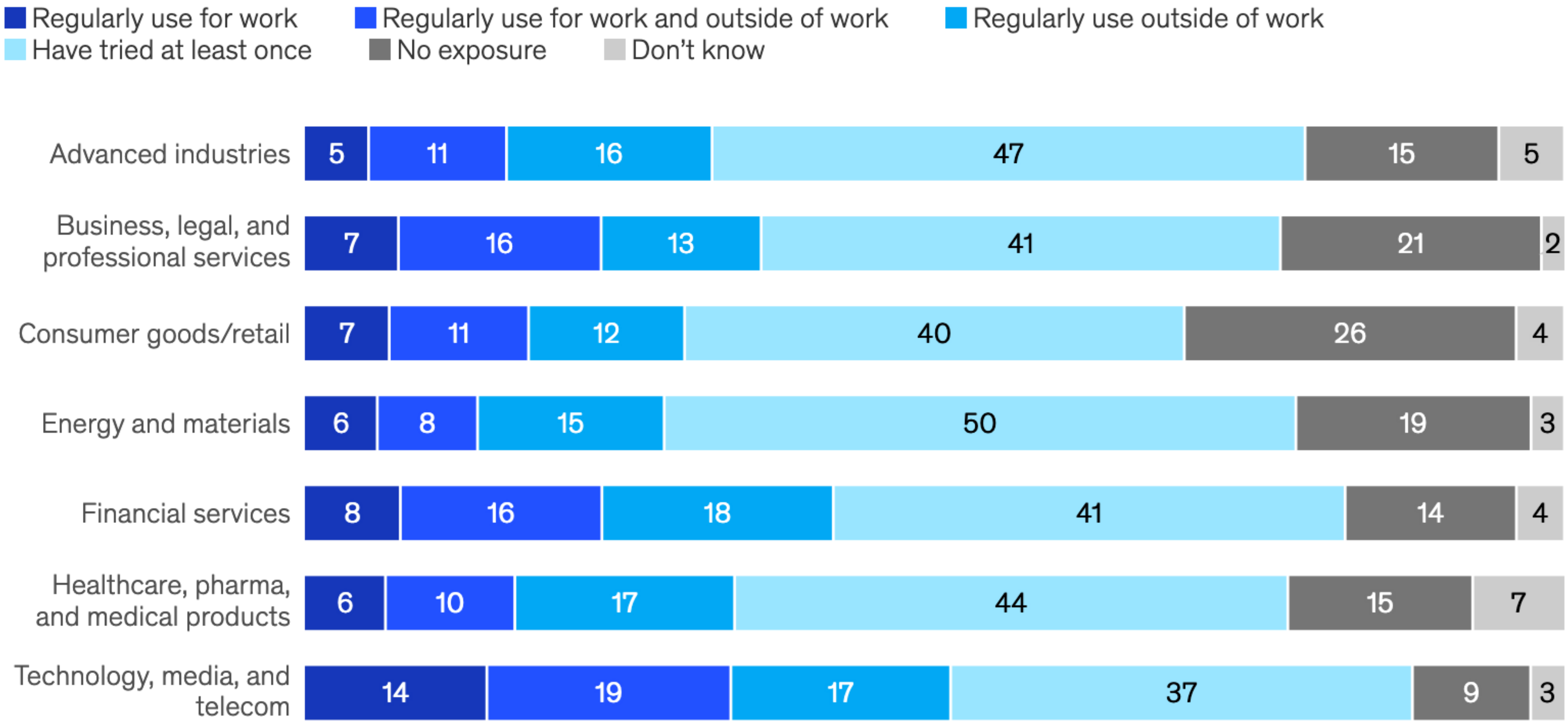
**A.I. is on a collision course with white-collar, high-paid jobs — and with unknown impact**

Summary

## Respondents across regions, industries, and seniority levels say they are already using generative AI tools.

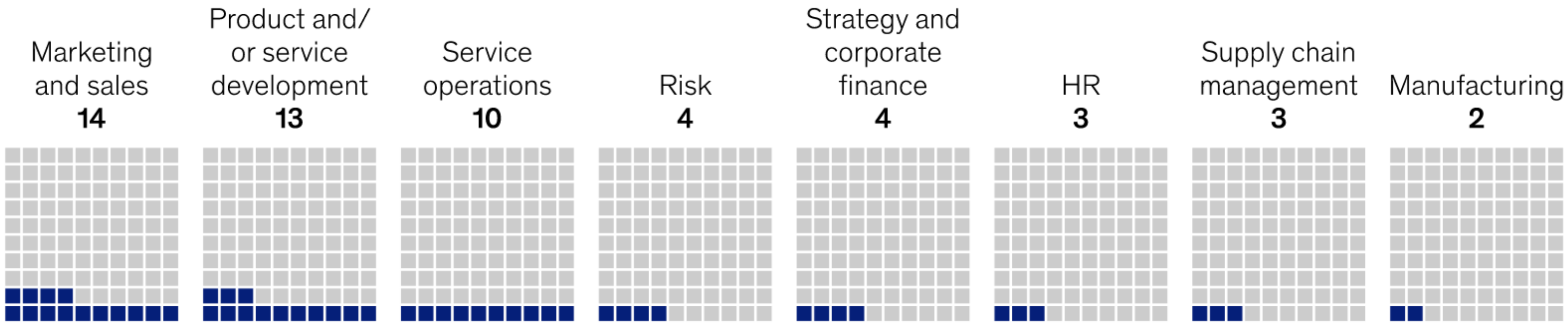
**Reported exposure to generative AI tools,** % of respondents

**Select demographic** By industry ▾

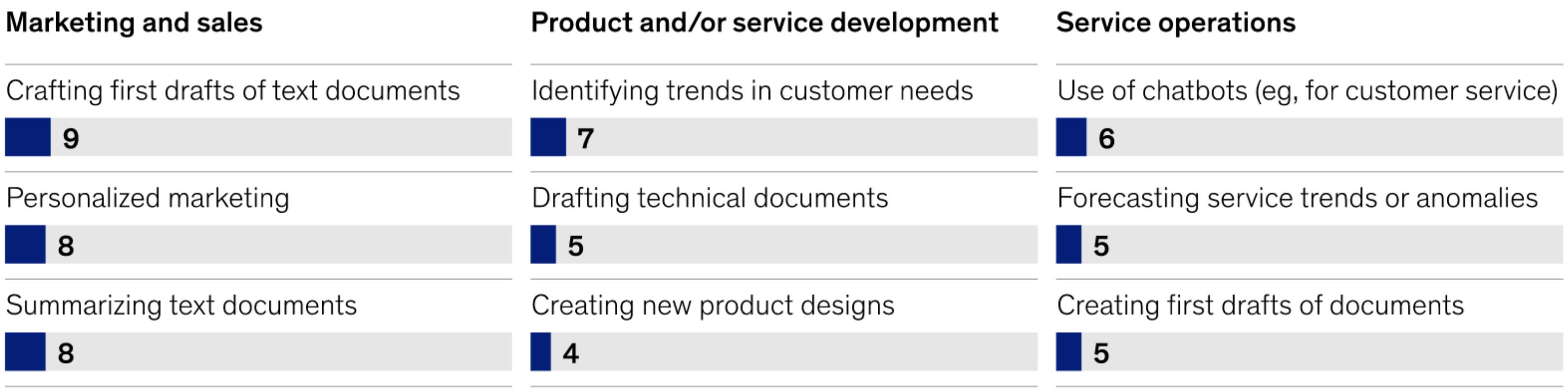


## The most commonly reported uses of generative AI tools are in marketing and sales, product and service development, and service operations.

**Share of respondents reporting that their organization is regularly using generative AI in given function,** %[1]

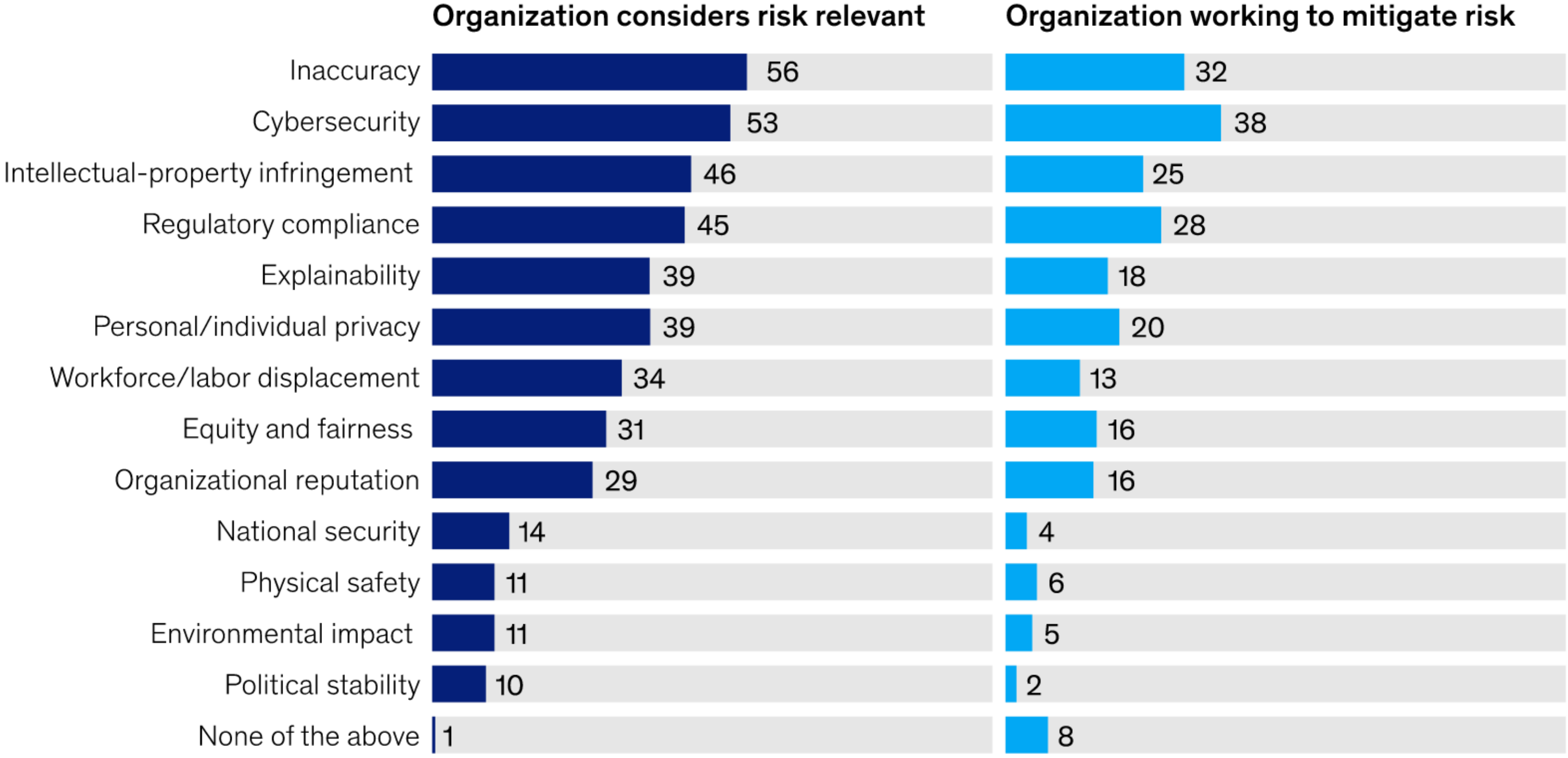


**Most regularly reported generative AI use cases within function,** % of respondents



# Inaccuracy, cybersecurity, and intellectual-property infringement are the most-cited risks of generative AI adoption.

**Generative AI–related risks that organizations consider relevant and are working to mitigate,** % of respondents[1]

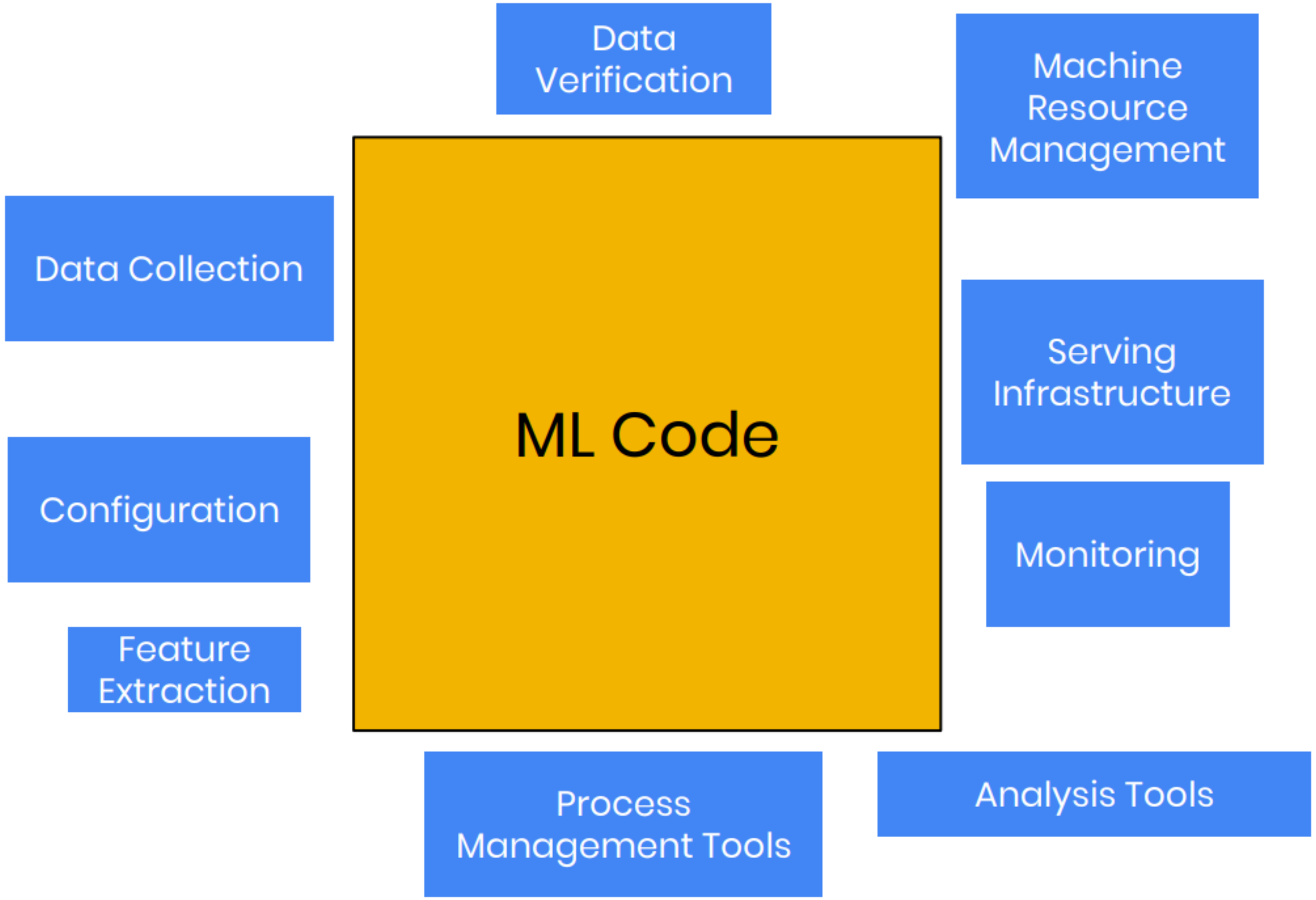


[1] Asked only of respondents whose organizations have adopted AI in at least 1 function. For both risks considered relevant and risks mitigated, n = 913.
Source: McKinsey Global Survey on AI, 1,684 participants at all levels of the organization, April 11–21, 2023

McKinsey & Company

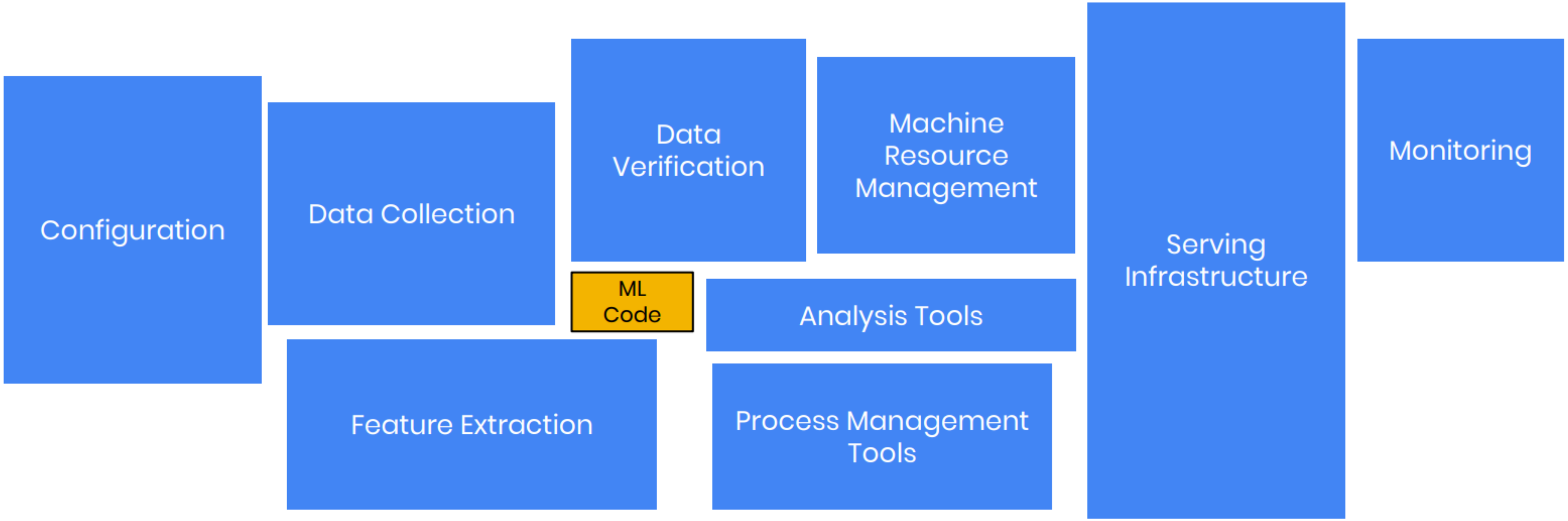
# AI 101

# Perception : ML Products are mostly about ML



Credit: Hidden Technical Debt of Machine Learning Systems, D. Sculley, et al.

# Reality : ML Requires DevOPS lots of it.



Configuration

Data Collection

Data Verification

Machine Resource Management

Serving Infrastructure

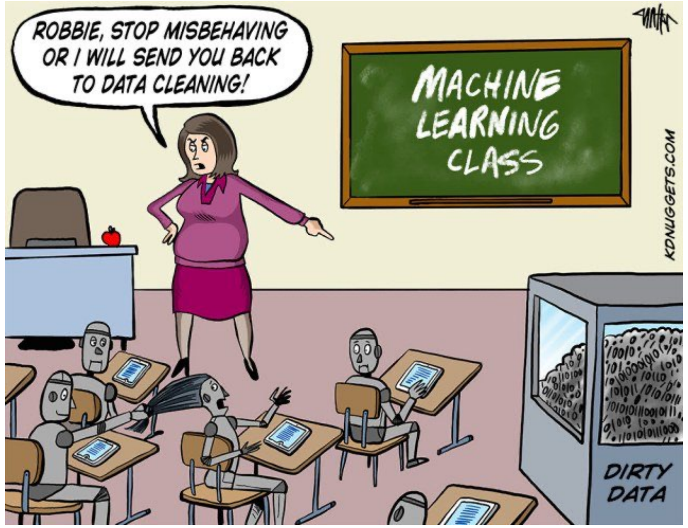
Monitoring

ML Code

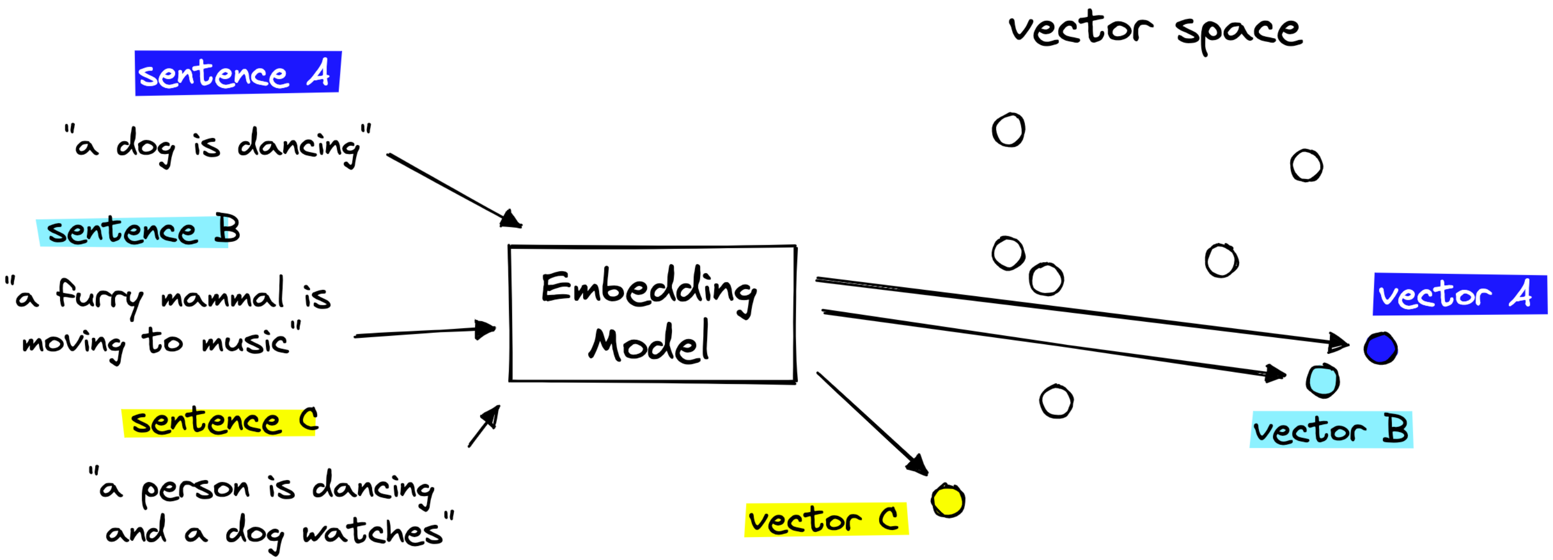
Analysis Tools

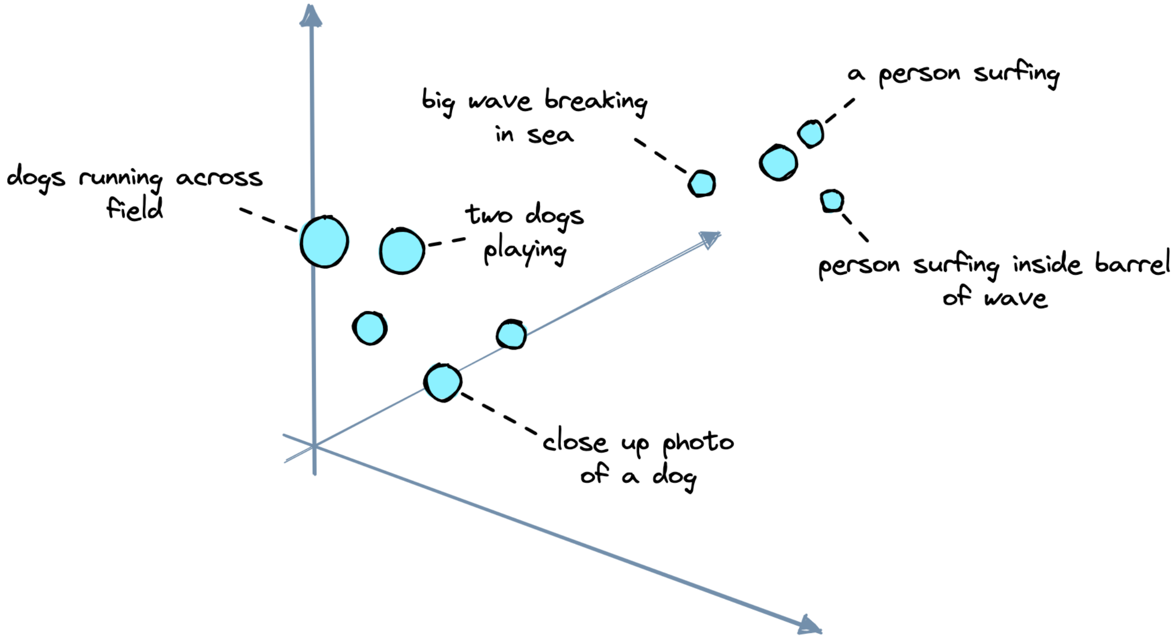
Feature Extraction

Process Management Tools


ROBBIE, STOP MISBEHAVING OR I WILL SEND YOU BACK TO DATA CLEANING!
MACHINE LEARNING CLASS
KDNUGGETS.COM
DIRTY DATA

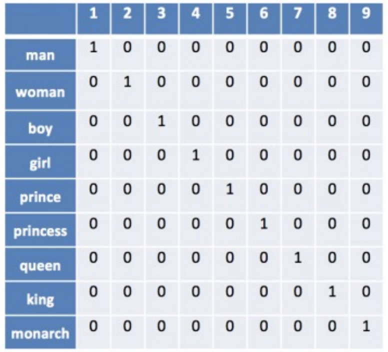
# Vectors 101


vector space
sentence A
"a dog is dancing"
sentence B
"a furry mammal is moving to music"
sentence C
"a person is dancing and a dog watches"
Embedding Model
vector A
vector B
vector C


a person surfing
big wave breaking in sea
dogs running across field
two dogs playing
person surfing inside barrel of wave
close up photo of a dog

# Vectors 101

Vocabulary:
Man, woman, boy, girl, prince, princess, queen, king, monarch

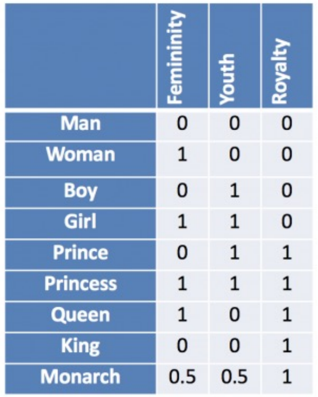


| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|
| man | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| woman | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| boy | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| girl | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| prince | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| princess | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| queen | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| king | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| monarch | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

Each word gets a 1x9 vector representation

Try to build a lower dimensional embedding

Vocabulary:
Man, woman, boy, girl, prince, princess, queen, king, monarch

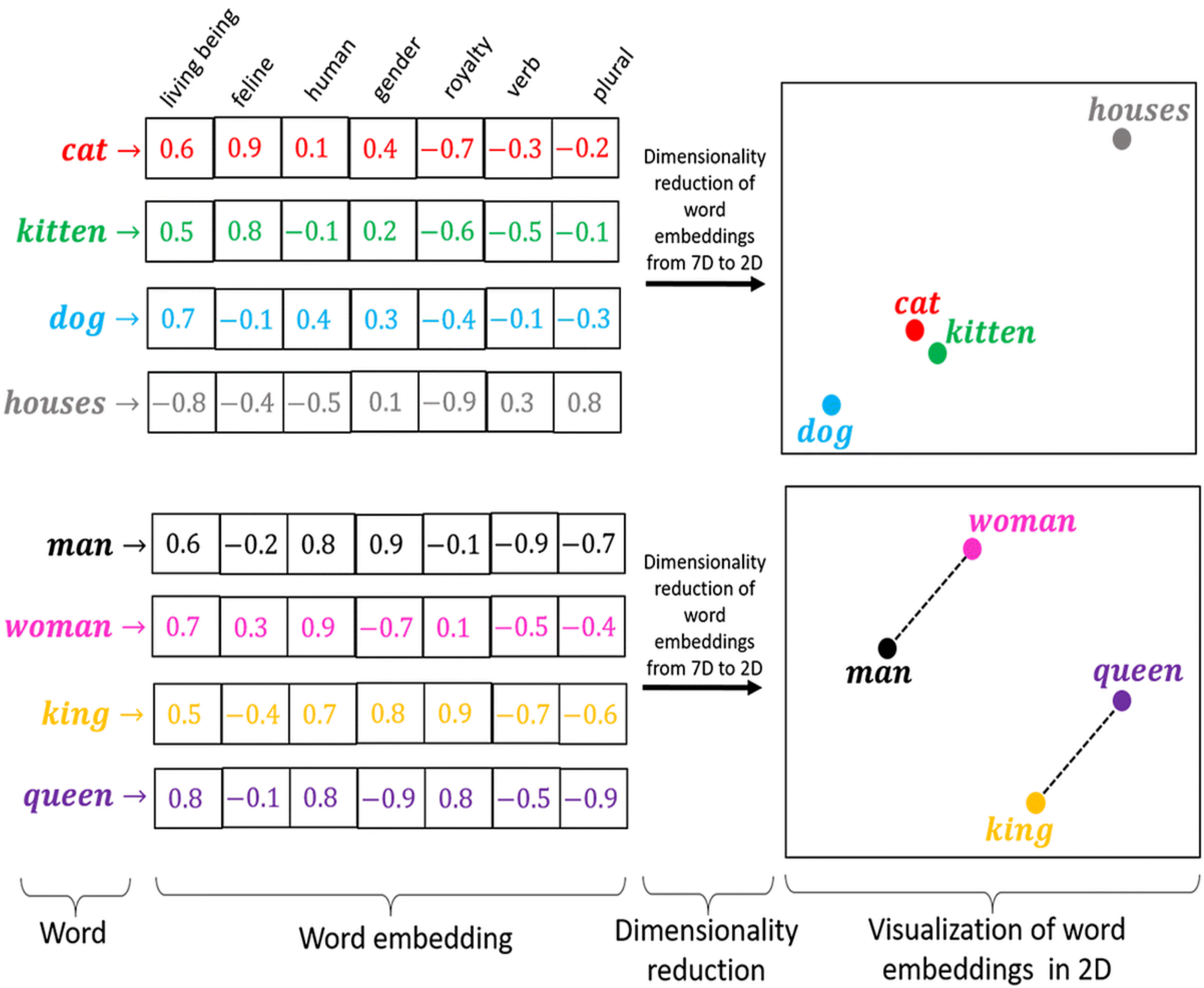


| | Femininity | Youth | Royalty |
|---|---|---|---|
| Man | 0 | 0 | 0 |
| Woman | 1 | 0 | 0 |
| Boy | 0 | 1 | 0 |
| Girl | 1 | 1 | 0 |
| Prince | 0 | 1 | 1 |
| Princess | 1 | 1 | 1 |
| Queen | 1 | 0 | 1 |
| King | 0 | 0 | 1 |
| Monarch | 0.5 | 0.5 | 1 |

Each word gets a 1x3 vector

Similar words… similar vectors

@shane_a_lynn | @TeamEdgeTier



| | living being | feline | human | gender | royalty | verb | plural |
|---|---|---|---|---|---|---|---|
| cat → | 0.6 | 0.9 | 0.1 | 0.4 | −0.7 | −0.3 | −0.2 |
| kitten → | 0.5 | 0.8 | −0.1 | 0.2 | −0.6 | −0.5 | −0.1 |
| dog → | 0.7 | −0.1 | 0.4 | 0.3 | −0.4 | −0.1 | −0.3 |
| houses → | −0.8 | −0.4 | −0.5 | 0.1 | −0.9 | 0.3 | 0.8 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| man → | 0.6 | −0.2 | 0.8 | 0.9 | −0.1 | −0.9 | −0.7 |
| woman → | 0.7 | 0.3 | 0.9 | −0.7 | 0.1 | −0.5 | −0.4 |
| king → | 0.5 | −0.4 | 0.7 | 0.8 | 0.9 | −0.7 | −0.6 |
| queen → | 0.8 | −0.1 | 0.8 | −0.9 | 0.8 | −0.5 | −0.9 |

Dimensionality reduction of word embeddings from 7D to 2D

Dimensionality reduction of word embeddings from 7D to 2D

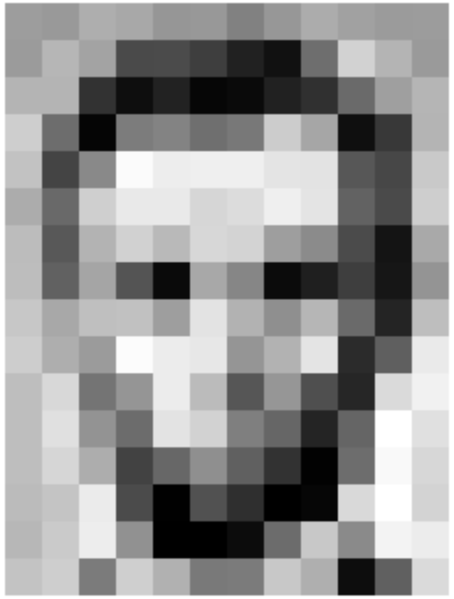
Word

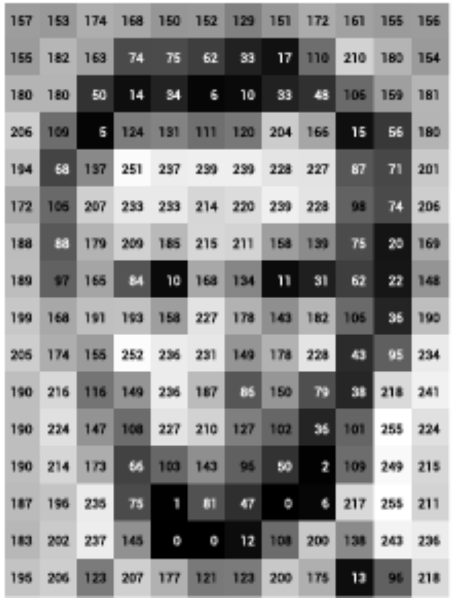
Word embedding

Dimensionality reduction

Visualization of word embeddings in 2D

# Vectors 101

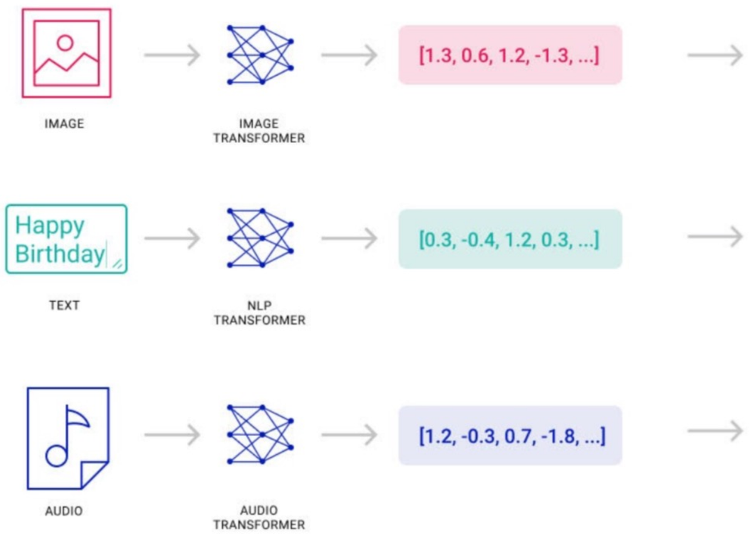




| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 157 | 153 | 174 | 168 | 150 | 152 | 129 | 151 | 172 | 161 | 155 | 156 |
| 155 | 182 | 163 | 74 | 75 | 62 | 33 | 17 | 110 | 210 | 180 | 154 |
| 180 | 180 | 50 | 14 | 34 | 6 | 10 | 33 | 48 | 106 | 159 | 181 |
| 206 | 109 | 5 | 124 | 131 | 111 | 120 | 204 | 166 | 15 | 56 | 180 |
| 194 | 68 | 137 | 251 | 237 | 239 | 239 | 228 | 227 | 87 | 71 | 201 |
| 172 | 105 | 207 | 233 | 233 | 214 | 220 | 239 | 228 | 98 | 74 | 206 |
| 188 | 88 | 179 | 209 | 185 | 215 | 211 | 158 | 139 | 75 | 20 | 169 |
| 189 | 97 | 165 | 84 | 10 | 168 | 134 | 11 | 31 | 62 | 22 | 148 |
| 199 | 168 | 191 | 193 | 158 | 227 | 178 | 143 | 182 | 106 | 36 | 190 |
| 205 | 174 | 155 | 252 | 236 | 231 | 149 | 178 | 228 | 43 | 95 | 234 |
| 190 | 216 | 116 | 149 | 236 | 187 | 85 | 150 | 79 | 38 | 218 | 241 |
| 190 | 224 | 147 | 108 | 227 | 210 | 127 | 102 | 36 | 101 | 255 | 224 |
| 190 | 214 | 173 | 66 | 103 | 143 | 95 | 50 | 2 | 109 | 249 | 215 |
| 187 | 196 | 235 | 75 | 1 | 81 | 47 | 0 | 6 | 217 | 255 | 211 |
| 183 | 202 | 237 | 145 | 0 | 0 | 12 | 108 | 200 | 138 | 243 | 236 |
| 195 | 206 | 123 | 207 | 177 | 121 | 123 | 200 | 175 | 13 | 96 | 218 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 157 | 153 | 174 | 168 | 150 | 152 | 129 | 151 | 172 | 161 | 155 | 156 |
| 155 | 182 | 163 | 74 | 75 | 62 | 33 | 17 | 110 | 210 | 180 | 154 |
| 180 | 180 | 50 | 14 | 34 | 6 | 10 | 33 | 48 | 106 | 159 | 181 |
| 206 | 109 | 5 | 124 | 131 | 111 | 120 | 204 | 166 | 15 | 56 | 180 |
| 194 | 68 | 137 | 251 | 237 | 239 | 239 | 228 | 227 | 87 | 71 | 201 |
| 172 | 105 | 207 | 233 | 233 | 214 | 220 | 239 | 228 | 98 | 74 | 206 |
| 188 | 88 | 179 | 209 | 185 | 215 | 211 | 158 | 139 | 75 | 20 | 169 |
| 189 | 97 | 165 | 84 | 10 | 168 | 134 | 11 | 31 | 62 | 22 | 148 |
| 199 | 168 | 191 | 193 | 158 | 227 | 178 | 143 | 182 | 106 | 36 | 190 |
| 205 | 174 | 155 | 252 | 236 | 231 | 149 | 178 | 228 | 43 | 95 | 234 |
| 190 | 216 | 116 | 149 | 236 | 187 | 86 | 150 | 79 | 38 | 218 | 241 |
| 190 | 224 | 147 | 108 | 227 | 210 | 127 | 102 | 36 | 101 | 255 | 224 |
| 190 | 214 | 173 | 66 | 103 | 143 | 96 | 50 | 2 | 109 | 249 | 215 |
| 187 | 196 | 235 | 75 | 1 | 81 | 47 | 0 | 6 | 217 | 255 | 211 |
| 183 | 202 | 237 | 145 | 0 | 0 | 12 | 108 | 200 | 138 | 243 | 236 |
| 195 | 206 | 123 | 207 | 177 | 121 | 123 | 200 | 175 | 13 | 96 | 218 |

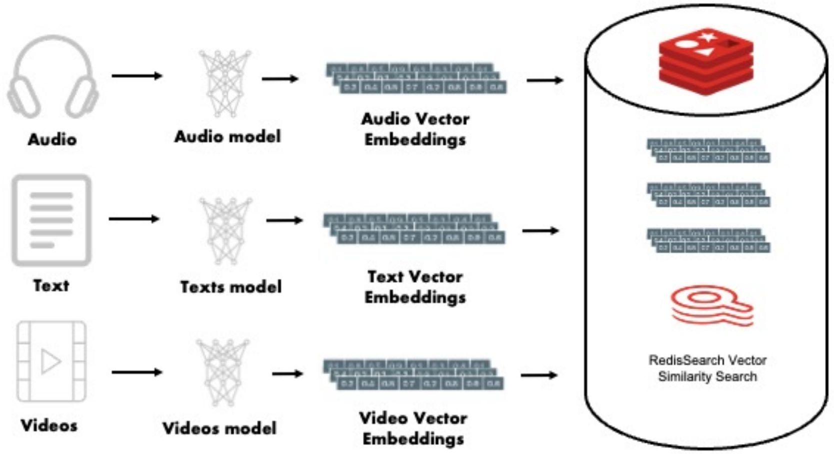
Object Vector Task



Object recognition, deduplication, scene detection, product search, ...

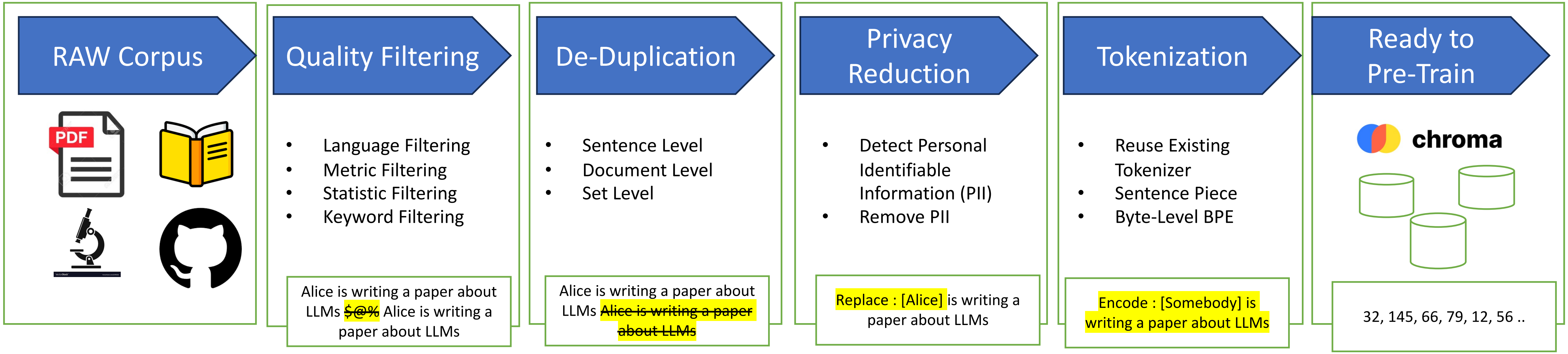
Translation, understanding, Sentiment, Question Answering, Semantic Search, ...

Anomaly detection, speech-to-text, music transcription, machinery malfunction, ...



# Data Pre-Processing Pipeline

Vectors 101



RAW Corpus

Quality Filtering

- Language Filtering
- Metric Filtering
- Statistic Filtering
- Keyword Filtering

Alice is writing a paper about LLMs ~~$@%~~ Alice is writing a paper about LLMs

De-Duplication

- Sentence Level
- Document Level
- Set Level

Alice is writing a paper about LLMs ~~Alice is writing a paper about LLMs~~

Privacy Reduction

- Detect Personal Identifiable Information (PII)
- Remove PII

Replace : [Alice] is writing a paper about LLMs

Tokenization

- Reuse Existing Tokenizer
- Sentence Piece
- Byte-Level BPE

Encode : [Somebody] is writing a paper about LLMs

Ready to Pre-Train

32, 145, 66, 79, 12, 56 ..

**Encoder Models**

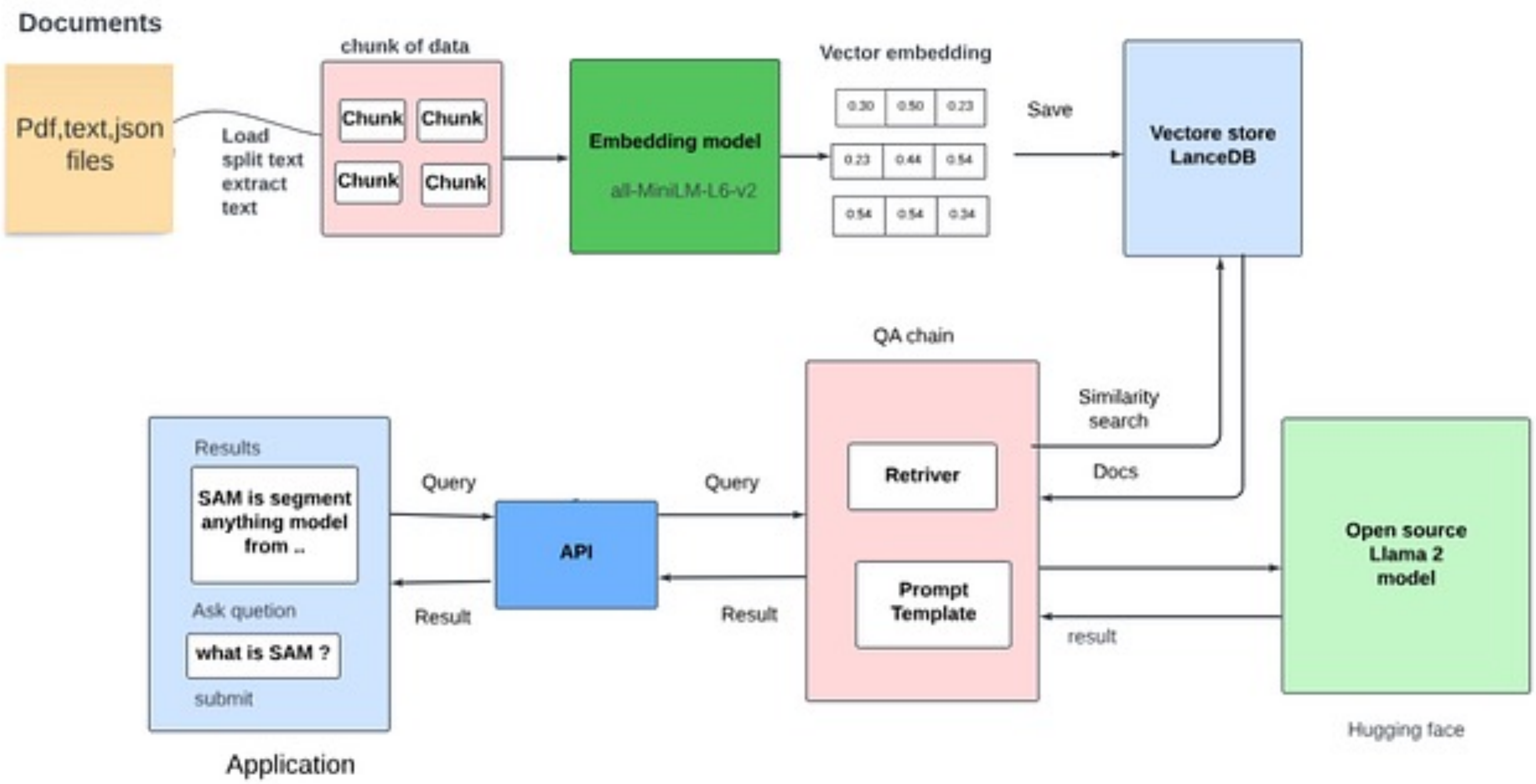
**Instructor Embeddings**

**Llama Embeddings**

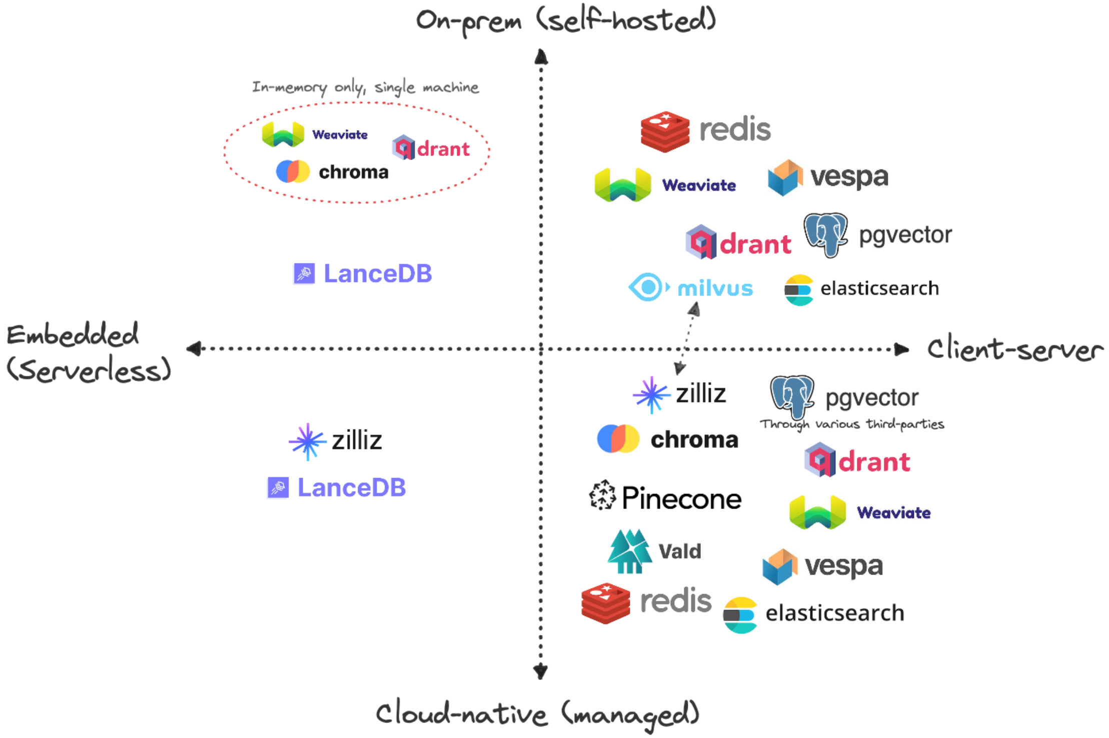
**Word2VEC**

**OpenAI :** text-embedding-ada-002 model

## Key Components for Building RAG based applications:

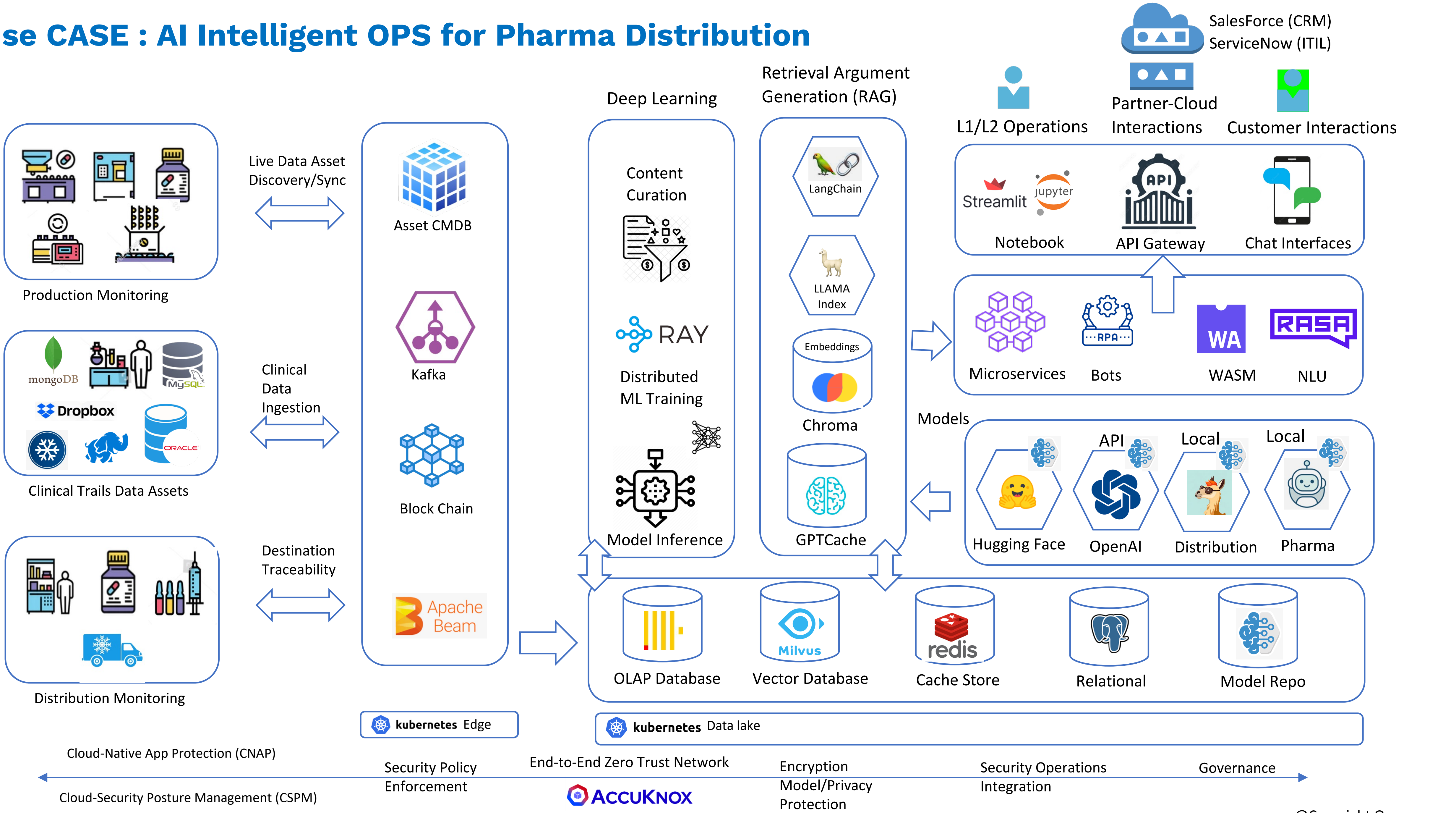


Source : Lance Blog



| | |
|---|---|
| Pinecone | Proprietary composite index |
| milvus / zilliz | Flat, Annoy, IVF, HNSW/RHNSW (Flat/PQ), DiskANN |
| Weaviate | Customized HNSW, HNSW (PQ), DiskANN (in progress...) |
| drant | Customized HNSW |
| chroma | HNSW |
| LanceDB | IVF (PQ), DiskANN (in progress...) |
| vespa | HNSW + BM25 hybrid |
| Vald | NGT |
| elasticsearch | Flat (brute force), HNSW |
| redis | Flat (brute force), HNSW |
| pgvector | IVF (Flat), IVF (PQ) in progress... |

# Use CASE : AI Intelligent OPS for Pharma Distribution



SalesForce (CRM)
ServiceNow (ITIL)

Production Monitoring

Live Data Asset Discovery/Sync

Clinical Trails Data Assets

mongoDB

Dropbox

MySQL

ORACLE

Clinical Data Ingestion

Distribution Monitoring

Destination Traceability

Asset CMDB

Kafka

Block Chain

Apache Beam

kubernetes Edge

Deep Learning

Content Curation

RAY

Distributed ML Training

Model Inference

Retrieval Argument Generation (RAG)

LangChain

LLAMA Index

Embeddings

Chroma

GPTCache

L1/L2 Operations

Partner-Cloud Interactions

Customer Interactions

Streamlit

jupyter

Notebook

API

API Gateway

Chat Interfaces

Microservices

RPA

Bots

WA

WASM

RASA

NLU

Models

API

Local

Local

Hugging Face

OpenAI

Distribution

Pharma

OLAP Database

Milvus

Vector Database

redis

Cache Store

Relational

Model Repo

kubernetes Data lake

Cloud-Native App Protection (CNAP)

Cloud-Security Posture Management (CSPM)

Security Policy Enforcement

End-to-End Zero Trust Network

ACCUKNOX

Encryption Model/Privacy Protection

Security Operations Integration

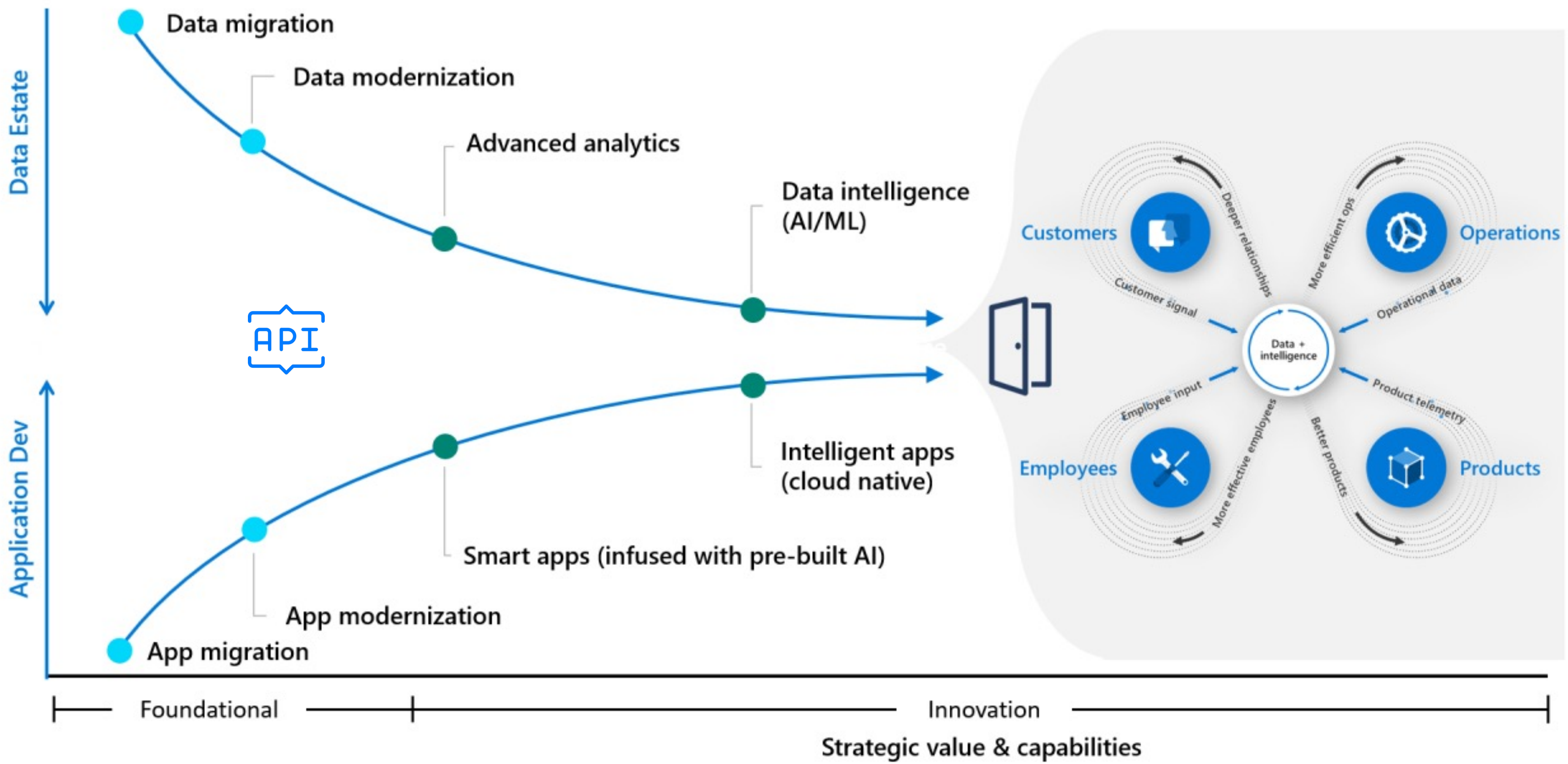
Governance

@Copyright Opensource

# Continuum to Unlock Digital Innovation

● Modernization

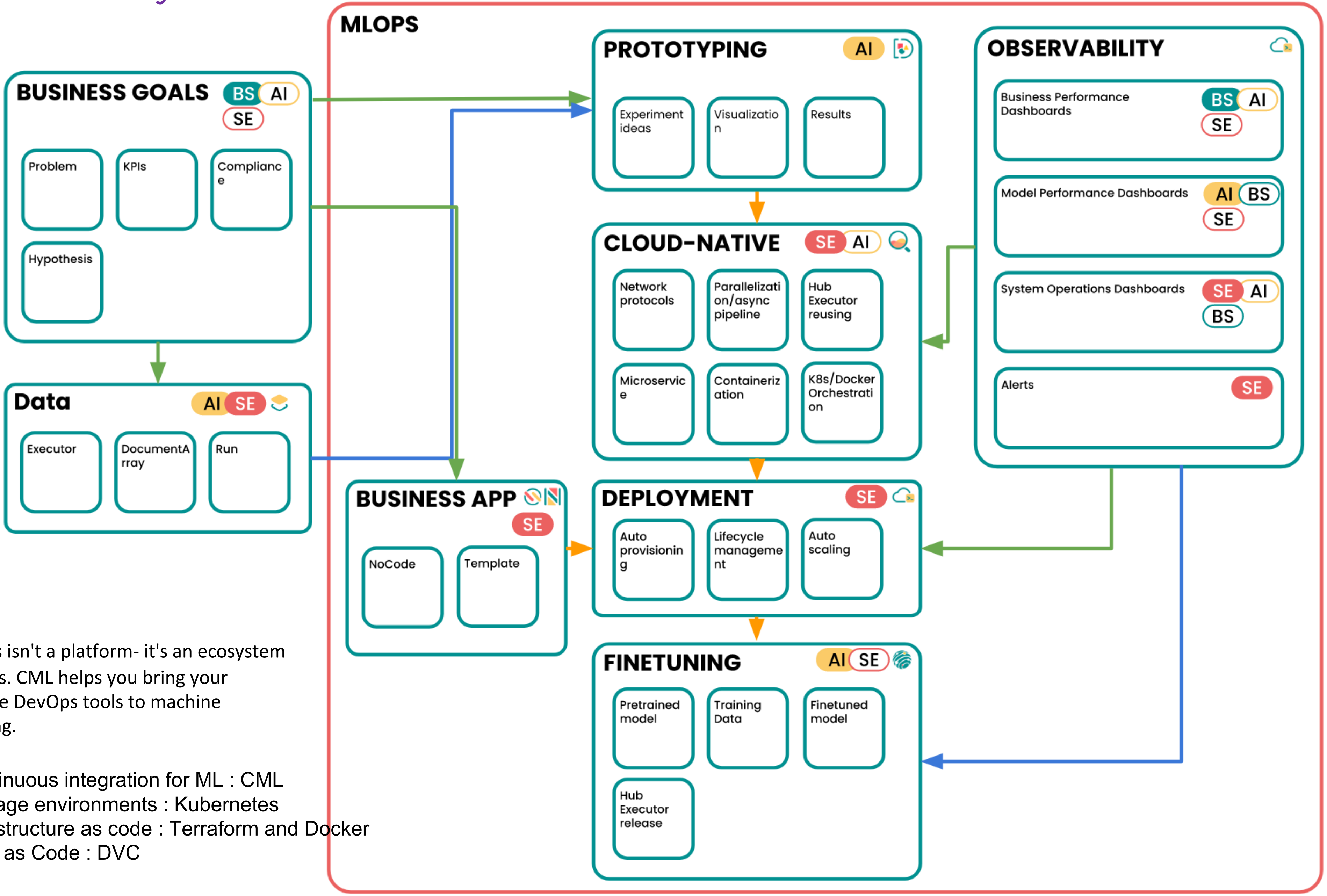
● Digital Transformation



We should urgently develop an AI-Strategy!
Sure, but first we need to develop a Data-Strategy.
The very first thing you need to do is train your staff in Data Literacy & - Awareness.
Edited by Murat Durmus
(@CEO_ANOMA)

Let's solve this problem with Machine Learning and improve customer experience.
We need a strategy.
Ofcourse, ML is the right tool.
Do we even have the data to start with?
imgflip.com

# **MLOPS** Eco-System



MLOPS

BUSINESS GOALS BS AI SE
- Problem
- KPIs
- Compliance
- Hypothesis

Data AI SE
- Executor
- DocumentArray
- Run

BUSINESS APP SE
- NoCode
- Template

PROTOTYPING AI
- Experiment ideas
- Visualization
- Results

CLOUD-NATIVE SE AI
- Network protocols
- Parallelization/async pipeline
- Hub Executor reusing
- Microservice
- Containerization
- K8s/Docker Orchestration

DEPLOYMENT SE
- Auto provisioning
- Lifecycle management
- Auto scaling

FINETUNING AI SE
- Pretrained model
- Training Data
- Finetuned model
- Hub Executor release

OBSERVABILITY
- Business Performance Dashboards BS AI SE
- Model Performance Dashboards AI BS SE
- System Operations Dashboards SE AI BS
- Alerts SE

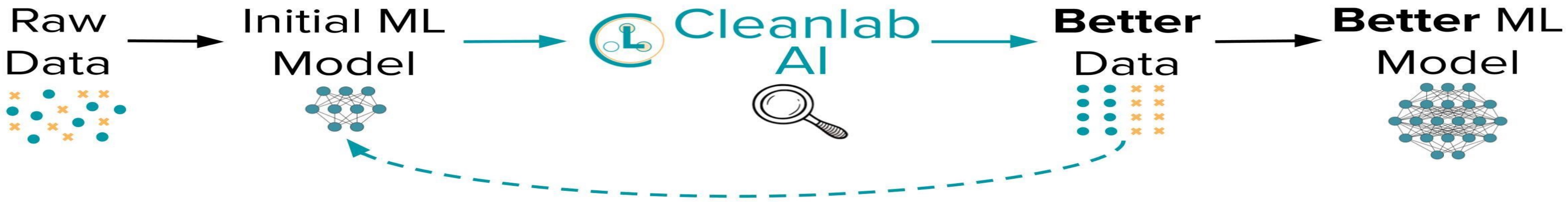
## LEGEND

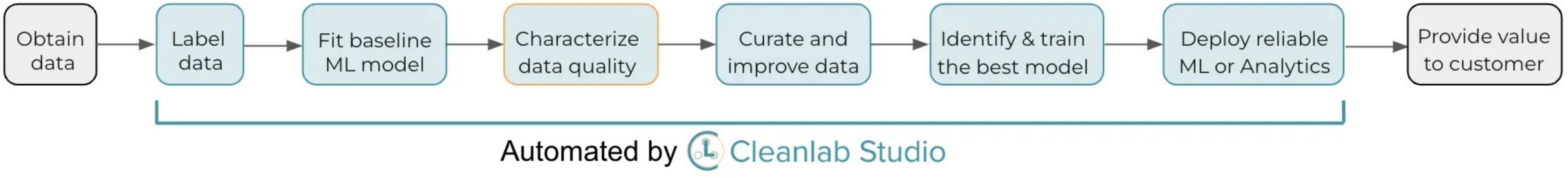
- AI AI engineer
- BS Business stakeholder
- SE Software engineer
- 🟨 Primary concern
- ▭ Secondary concern
- → Business goal flow
- → Data flow
- → Development flow
- - - Implementation variant

MLOps isn't a platform- it's an ecosystem of tools. CML helps you bring your favorite DevOps tools to machine learning.

- Continuous integration for ML : CML
- Manage environments : Kubernetes
- Infrastructure as code : Terraform and Docker
- Data as Code : DVC

**Business Stakeholders**





Obtain data

Label data

Fit baseline ML model

Characterize data quality

Curate and improve data

Identify & train the best model

Deploy reliable ML or Analytics

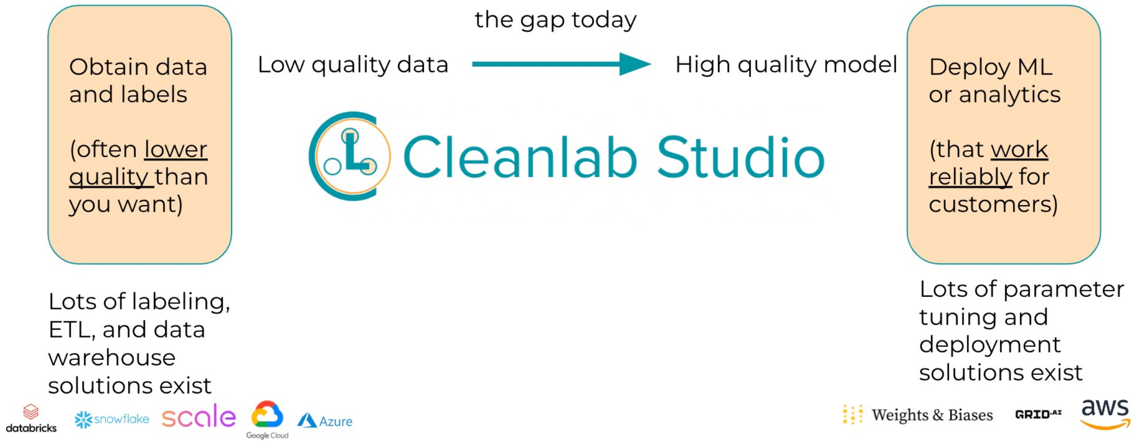
Provide value to customer

Automated by Cleanlab Studio

**Practicing data-centric AI can look like this:**

1. Train initial ML model on original dataset.
2. Utilize this model to diagnose data issues (via cleanlab methods) and improve the dataset.
3. Train the same model on the improved dataset.
4. Try various modeling techniques to further improve performance.

Most folks jump from Step 1 → 4, but you may achieve big gains without *any* change to your modeling code by using cleanlab! Continuously boost performance by iterating Steps 2 → 4 (and try to evaluate with *cleaned* data).



Obtain data and labels

(often lower quality than you want)

the gap today

Low quality data → High quality model

Cleanlab Studio

Deploy ML or analytics

(that work reliably for customers)

Lots of labeling, ETL, and data warehouse solutions exist

Lots of parameter tuning and deployment solutions exist

blog Source : Cleanlab

# Digital Wealth Management

Fusion.AI

Wealth managers are transitioning to digitally enabled, scalable platforms to empower clients and advisers with compelling experiences

**User research**
*(Voice of clients/FAs/other users)*

**Applied design**
*(Human-centered)*

**Content management**

**Digital marketing**

***Digital ID***

***Investor data store***
*(Shared with other wealth mgrs)*
- Personal data, transaction history, weblogs, social
- Goals and financial plan

**Client data aggregator**
*(Across providers/clients)*

**Notification services**
*Computer-generated, personalized, real time*
Email · In-app · SMS · Social

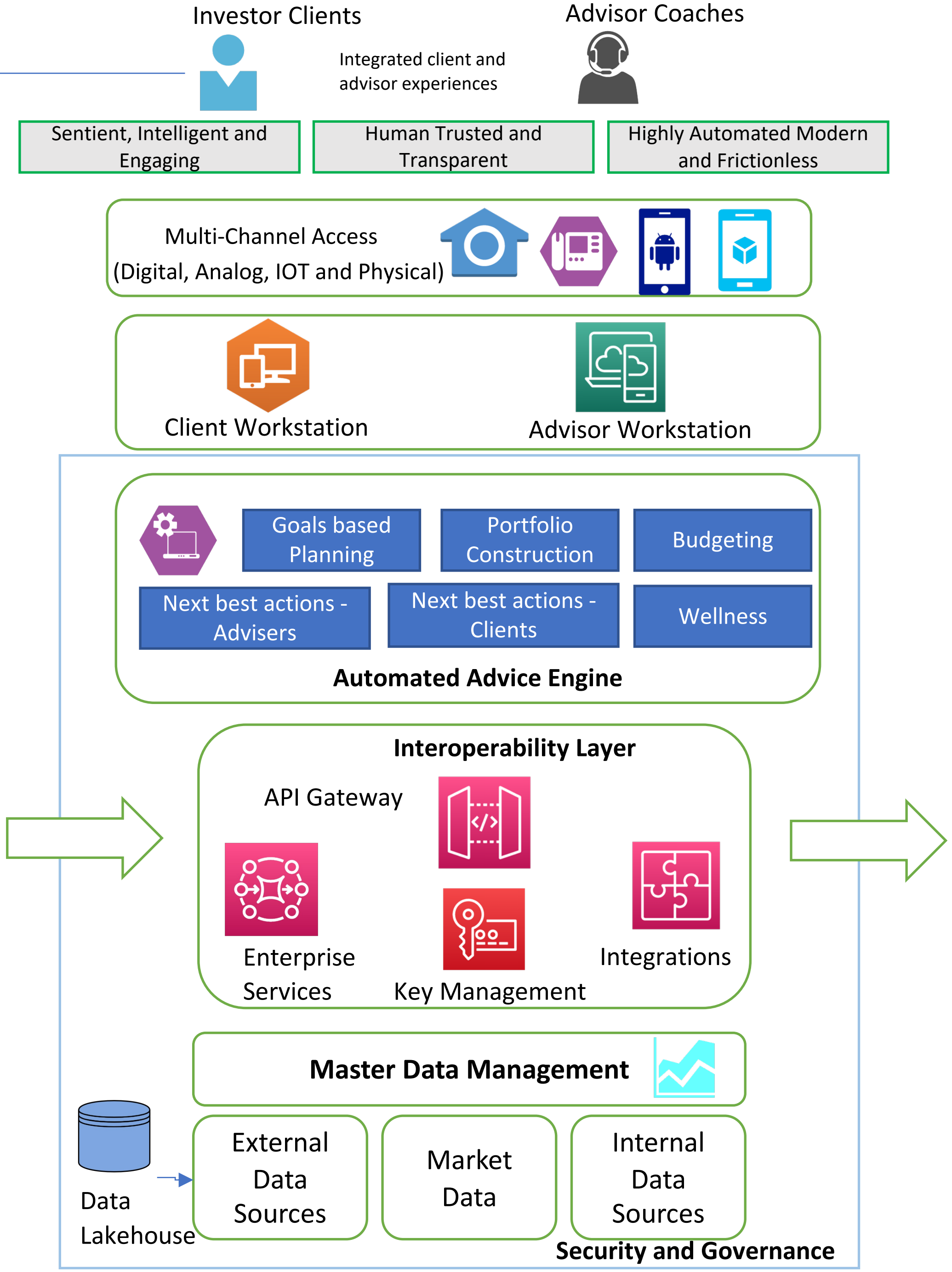
**Social/chat engine**
- With peers/ friends/advisers

**Client relationship management**
- Prospects, sales, servicing
- Client/sales analytics

**Client on-boarding/ account opening utility**
- Digital, with e-signature
- Integrated across products

**KYC/AML utility**
*(Proprietary or shared with other wealth managers)*

**Fiduciary/Best Interest engine**
- Guided decision-making
- Product comparisons
- Document generation and recordkeeping



***Integrated product platforms***

***Single investment and trust platform***
- Access to securities, funds, annuities, alts, and trust products
- Single interface with internal/external trading platforms and custodians
- Investment servicing and trust administration

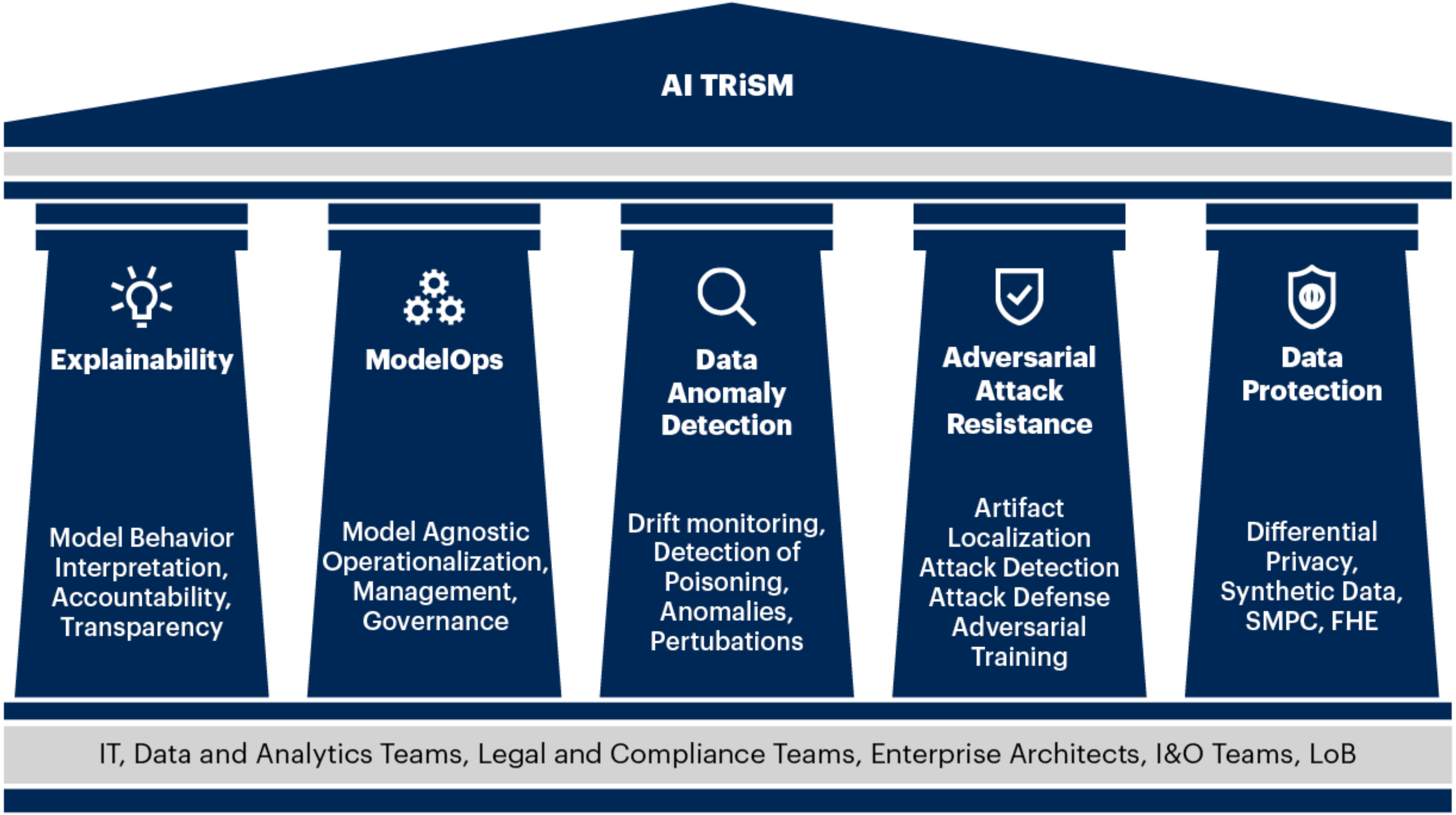
***Custody and clearing***
*(Proprietary or T/P platform)*

***Banking***
*(Proprietary or T/P platform)*
- Consumer lending, SBL
- Deposits and savings
- Payments

***Investment banking and capital markets***
- Conduit lending, M&A, etc.
- Trading desk: structured notes, FX, swaps

**Commercial banking**
- Lending
- Cash management

# AI Trust, Risk and Security Management Pillars



AI TRiSM

**Explainability**

Model Behavior Interpretation, Accountability, Transparency

**ModelOps**

Model Agnostic Operationalization, Management, Governance

**Data Anomaly Detection**

Drift monitoring, Detection of Poisoning, Anomalies, Pertubations

**Adversarial Attack Resistance**

Artifact Localization
Attack Detection
Attack Defense
Adversarial Training

**Data Protection**

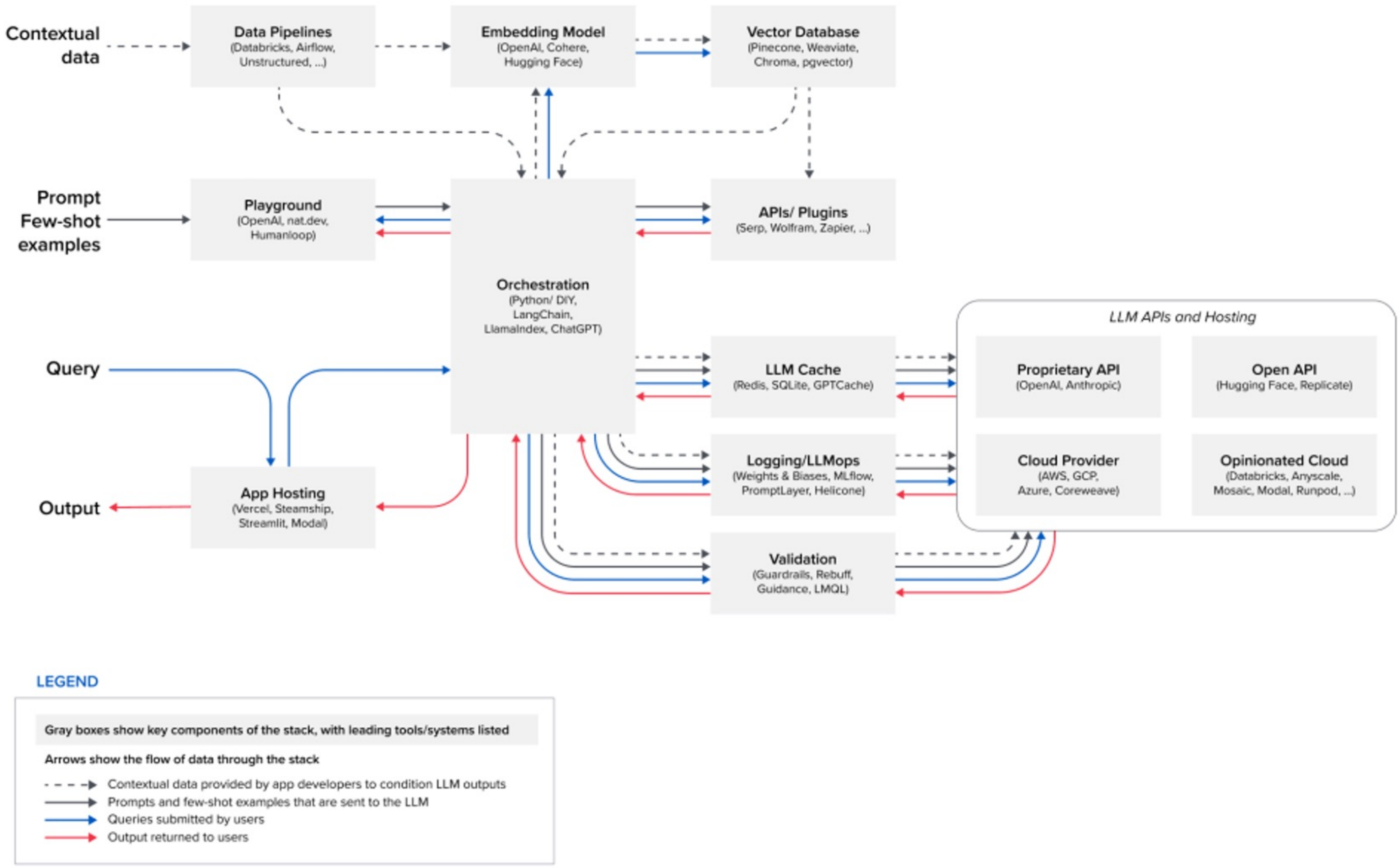
Differential Privacy, Synthetic Data, SMPC, FHE

IT, Data and Analytics Teams, Legal and Compliance Teams, Enterprise Architects, I&O Teams, LoB

Source: Gartner

750738_C

# **MLOPS** Emergency App Stack



Contextual data

Prompt Few-shot examples

Query

Output

**Data Pipelines** (Databricks, Airflow, Unstructured, ...)

**Embedding Model** (OpenAI, Cohere, Hugging Face)

**Vector Database** (Pinecone, Weaviate, Chroma, pgvector)

**Playground** (OpenAI, nat.dev, Humanloop)

**Orchestration** (Python/ DIY, LangChain, LlamaIndex, ChatGPT)

**APIs/ Plugins** (Serp, Wolfram, Zapier, ...)

**App Hosting** (Vercel, Steamship, Streamlit, Modal)

**LLM Cache** (Redis, SQLite, GPTCache)

**Logging/LLMops** (Weights & Biases, MLflow, PromptLayer, Helicone)

**Validation** (Guardrails, Rebuff, Guidance, LMQL)

*LLM APIs and Hosting*

**Proprietary API** (OpenAI, Anthropic)

**Open API** (Hugging Face, Replicate)

**Cloud Provider** (AWS, GCP, Azure, Coreweave)

**Opinionated Cloud** (Databricks, Anyscale, Mosaic, Modal, Runpod, ...)

LEGEND

Gray boxes show key components of the stack, with leading tools/systems listed

Arrows show the flow of data through the stack

⇢ Contextual data provided by app developers to condition LLM outputs

→ Prompts and few-shot examples that are sent to the LLM

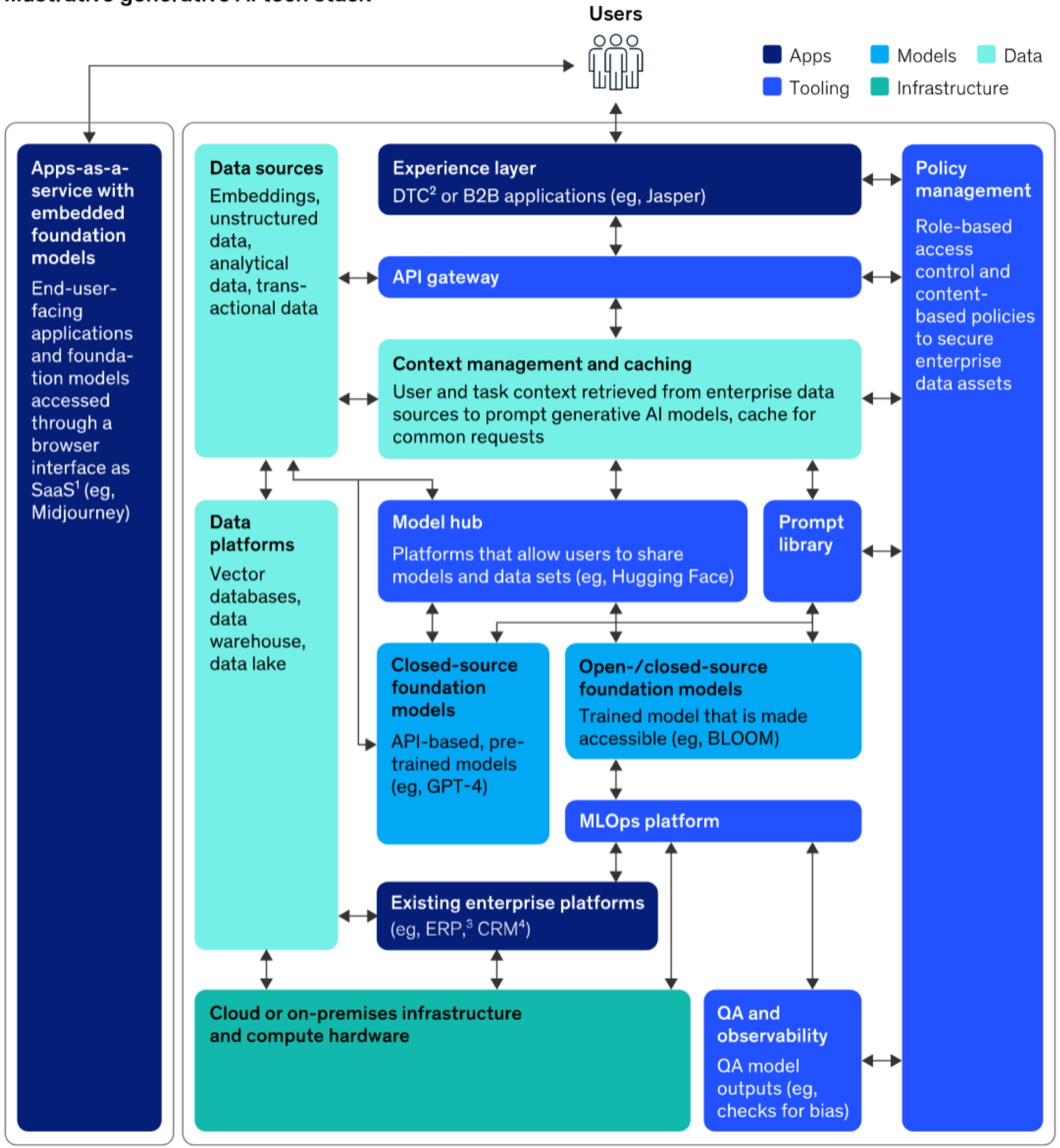
→ Queries submitted by users

→ Output returned to users

# Tech Stack

## The tech stack for generative AI is emerging.

### Illustrative generative AI tech stack



Users

Apps, Tooling, Models, Infrastructure, Data

**Apps-as-a-service with embedded foundation models**
End-user-facing applications and foundation models accessed through a browser interface as SaaS[1] (eg, Midjourney)

**Data sources**
Embeddings, unstructured data, analytical data, transactional data

**Experience layer**
DTC[2] or B2B applications (eg, Jasper)

**API gateway**

**Context management and caching**
User and task context retrieved from enterprise data sources to prompt generative AI models, cache for common requests

**Policy management**
Role-based access control and content-based policies to secure enterprise data assets

**Data platforms**
Vector databases, data warehouse, data lake

**Model hub**
Platforms that allow users to share models and data sets (eg, Hugging Face)

**Prompt library**

**Closed-source foundation models**
API-based, pre-trained models (eg, GPT-4)

**Open-/closed-source foundation models**
Trained model that is made accessible (eg, BLOOM)

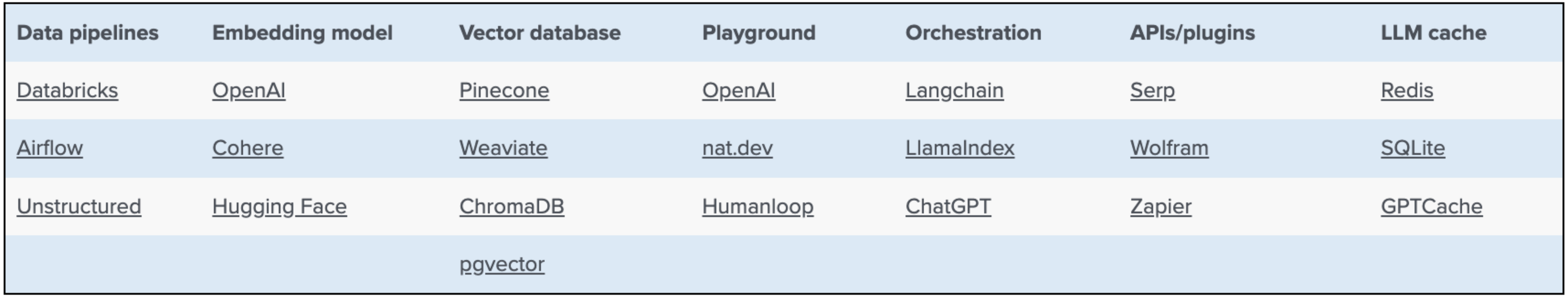
**MLOps platform**

**Existing enterprise platforms**
(eg, ERP,[3] CRM[4])

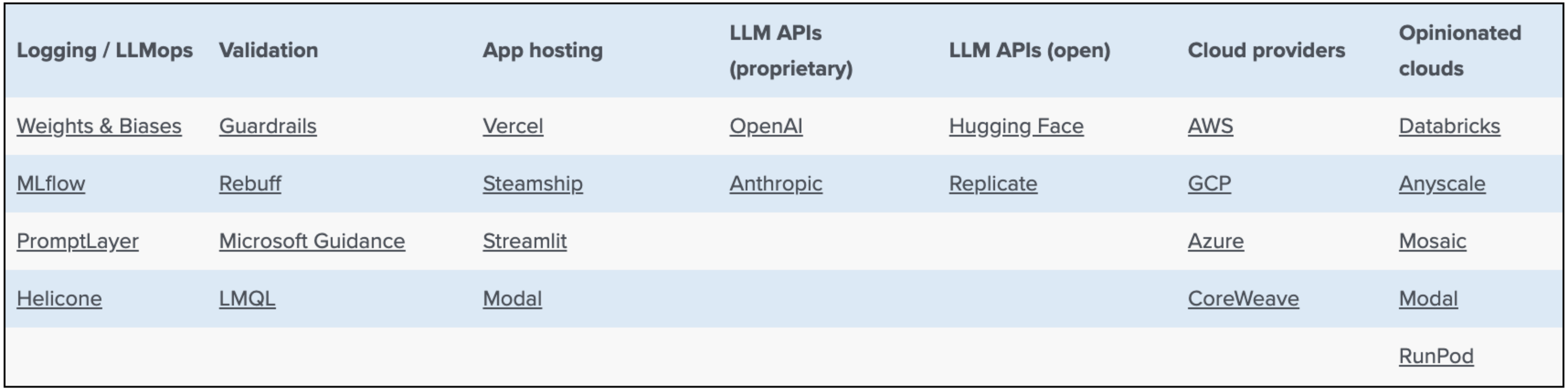
**Cloud or on-premises infrastructure and compute hardware**

**QA and observability**
QA model outputs (eg, checks for bias)

# **LLM** Technology Stack Choices

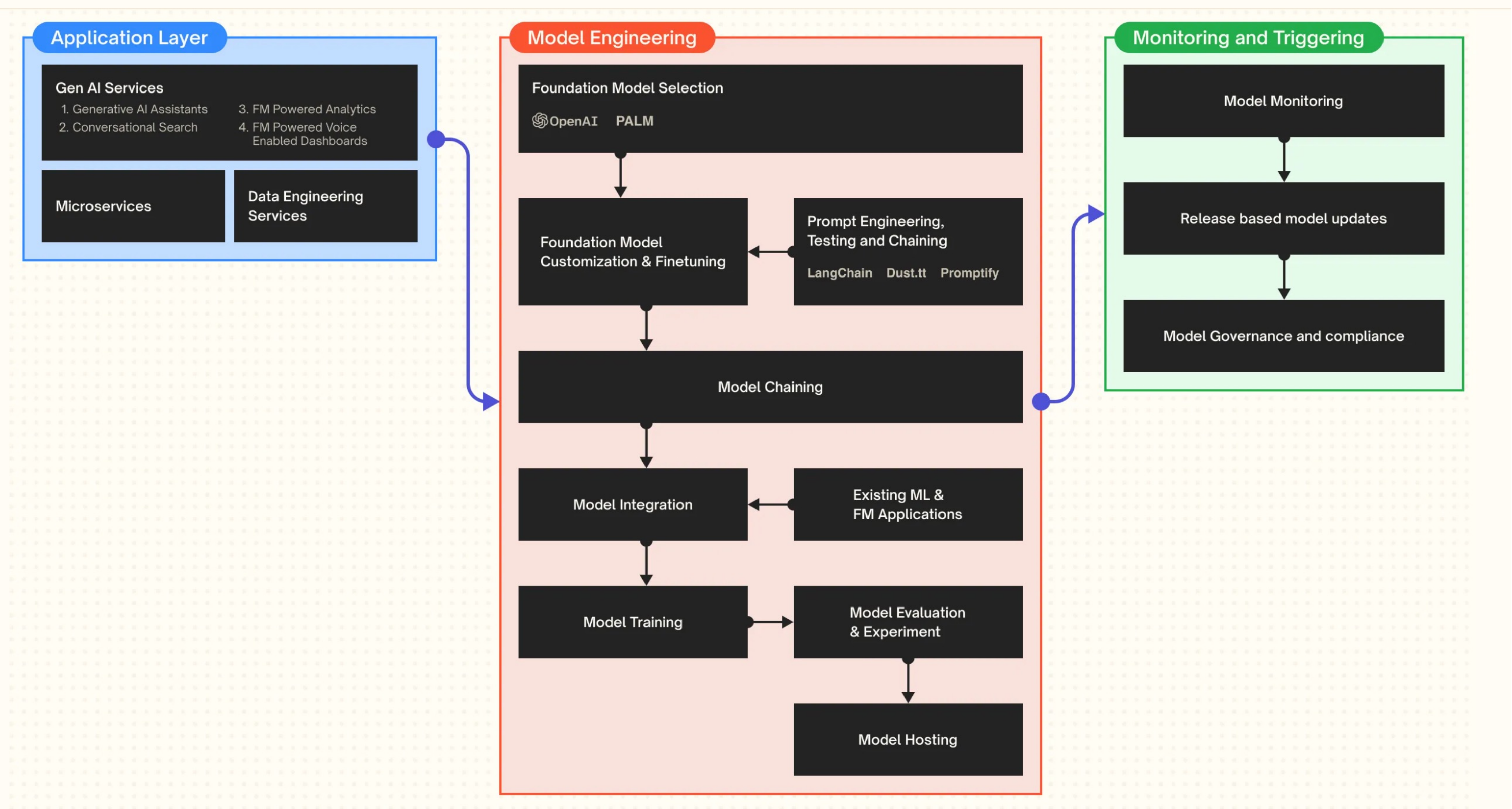


| Data pipelines | Embedding model | Vector database | Playground | Orchestration | APIs/plugins | LLM cache |
|---|---|---|---|---|---|---|
| Databricks | OpenAI | Pinecone | OpenAI | Langchain | Serp | Redis |
| Airflow | Cohere | Weaviate | nat.dev | LlamaIndex | Wolfram | SQLite |
| Unstructured | Hugging Face | ChromaDB | Humanloop | ChatGPT | Zapier | GPTCache |
|  |  | pgvector |  |  |  |  |



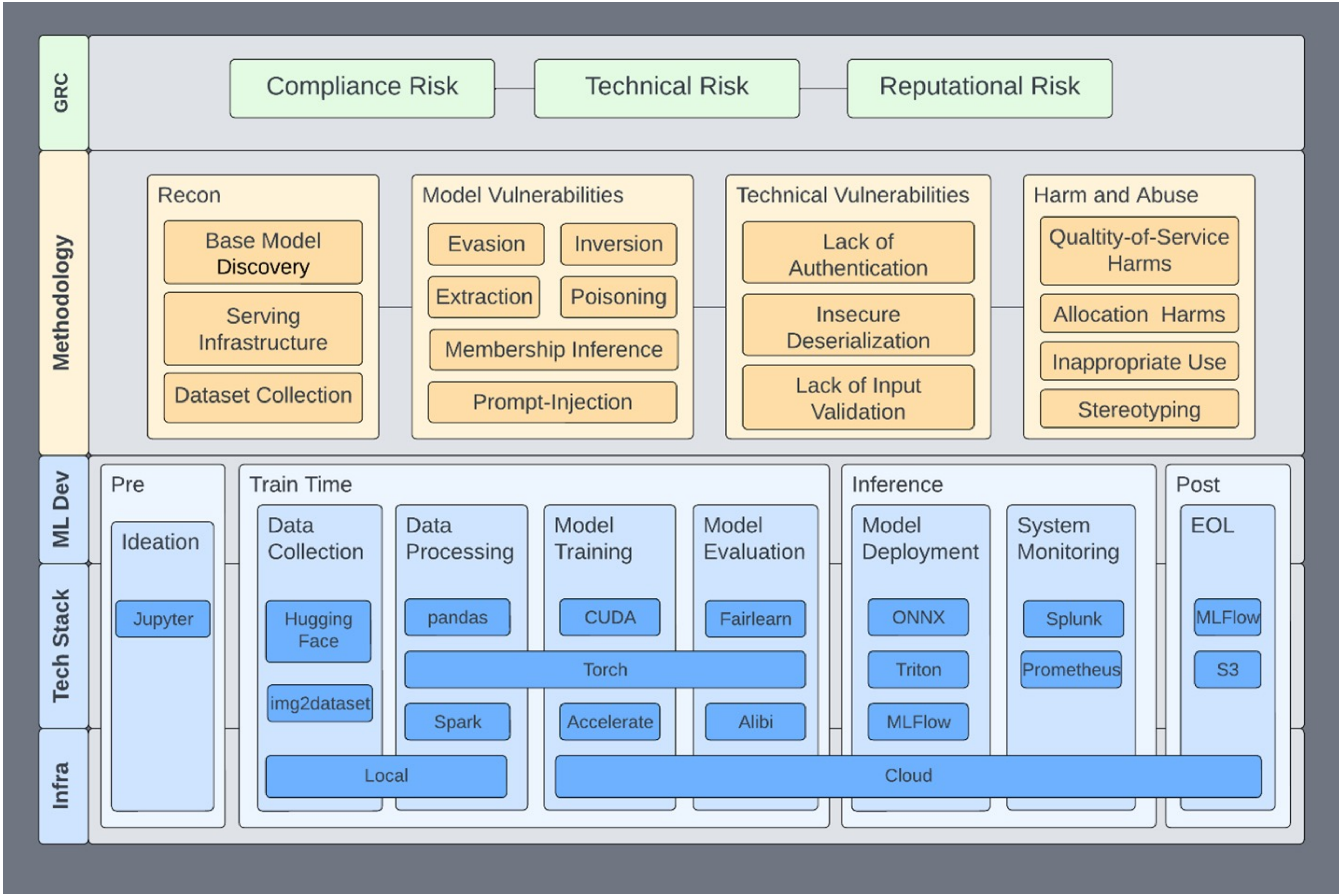
| Logging / LLMops | Validation | App hosting | LLM APIs (proprietary) | LLM APIs (open) | Cloud providers | Opinionated clouds |
|---|---|---|---|---|---|---|
| Weights & Biases | Guardrails | Vercel | OpenAI | Hugging Face | AWS | Databricks |
| MLflow | Rebuff | Steamship | Anthropic | Replicate | GCP | Anyscale |
| PromptLayer | Microsoft Guidance | Streamlit |  |  | Azure | Mosaic |
| Helicone | LMQL | Modal |  |  | CoreWeave | Modal |
|  |  |  |  |  |  | RunPod |

# **GEN-AI** Service Development



Application Layer
Gen AI Services
1. Generative AI Assistants
2. Conversational Search
3. FM Powered Analytics
4. FM Powered Voice Enabled Dashboards
Microservices
Data Engineering Services
Model Engineering
Foundation Model Selection
OpenAI
PALM
Foundation Model Customization & Finetuning
Prompt Engineering, Testing and Chaining
LangChain
Dust.tt
Promptify
Model Chaining
Model Integration
Existing ML & FM Applications
Model Training
Model Evaluation & Experiment
Model Hosting
Monitoring and Triggering
Model Monitoring
Release based model updates
Model Governance and compliance

# ML Tech Stack and Security Risk Management



GRC

Compliance Risk

Technical Risk

Reputational Risk

Methodology

Recon

Base Model Discovery

Serving Infrastructure

Dataset Collection

Model Vulnerabilities

Evasion

Inversion

Extraction

Poisoning

Membership Inference

Prompt-Injection

Technical Vulnerabilities

Lack of Authentication

Insecure Deserialization

Lack of Input Validation

Harm and Abuse

Qualtity-of-Service Harms

Allocation Harms

Inappropriate Use

Stereotyping

ML Dev

Tech Stack

Infra

Pre

Ideation

Jupyter

Train Time

Data Collection

Hugging Face

img2dataset

Data Processing

pandas

Spark

Model Training

CUDA

Accelerate

Model Evaluation

Fairlearn

Alibi

Torch

Local

Inference

Model Deployment

ONNX

Triton

MLFlow

System Monitoring

Splunk

Prometheus

Post

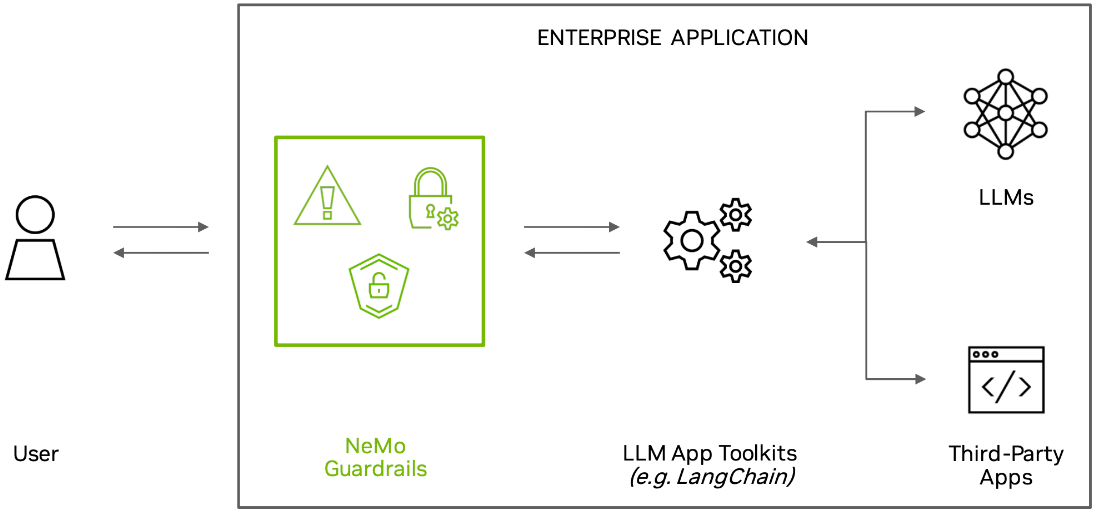
EOL

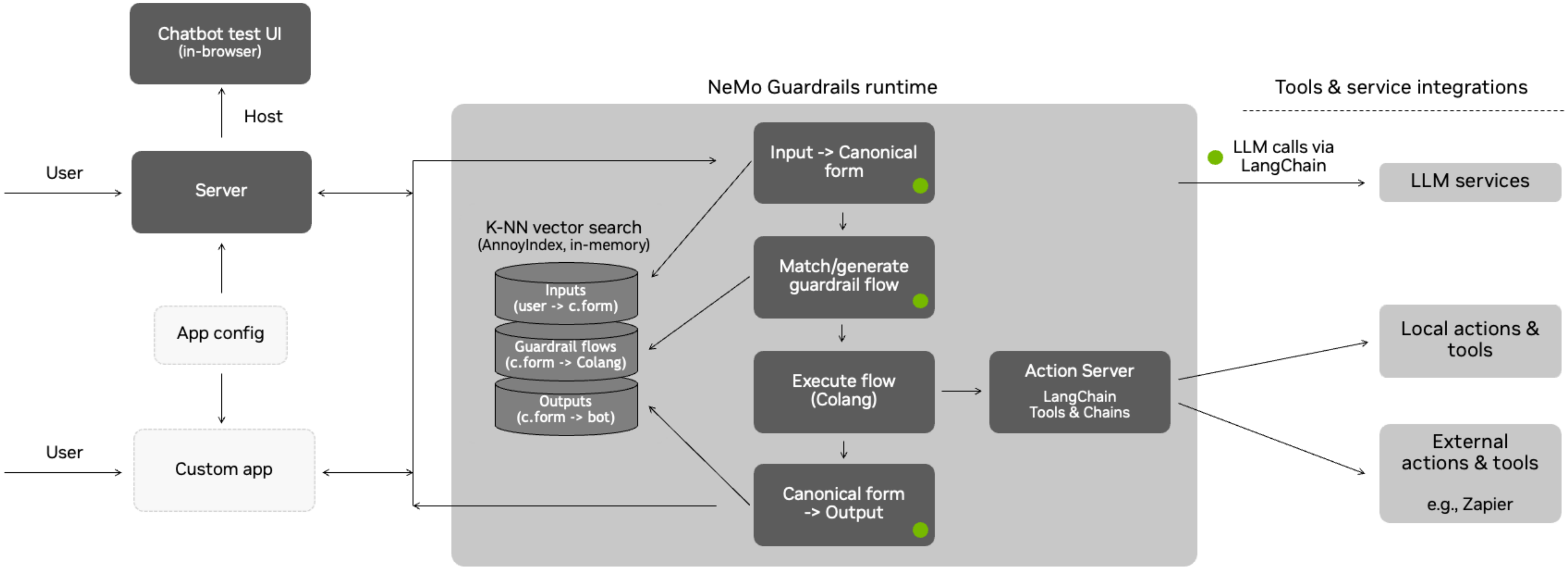
MLFlow

S3

Cloud

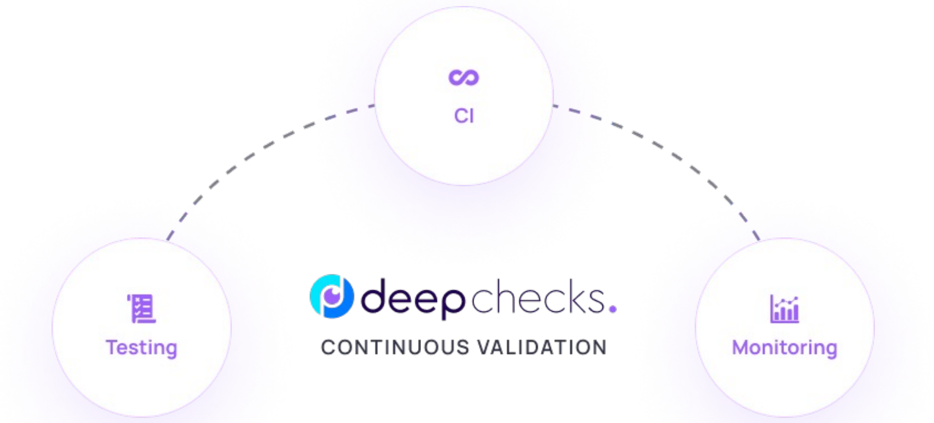
# NeMO Guardrails





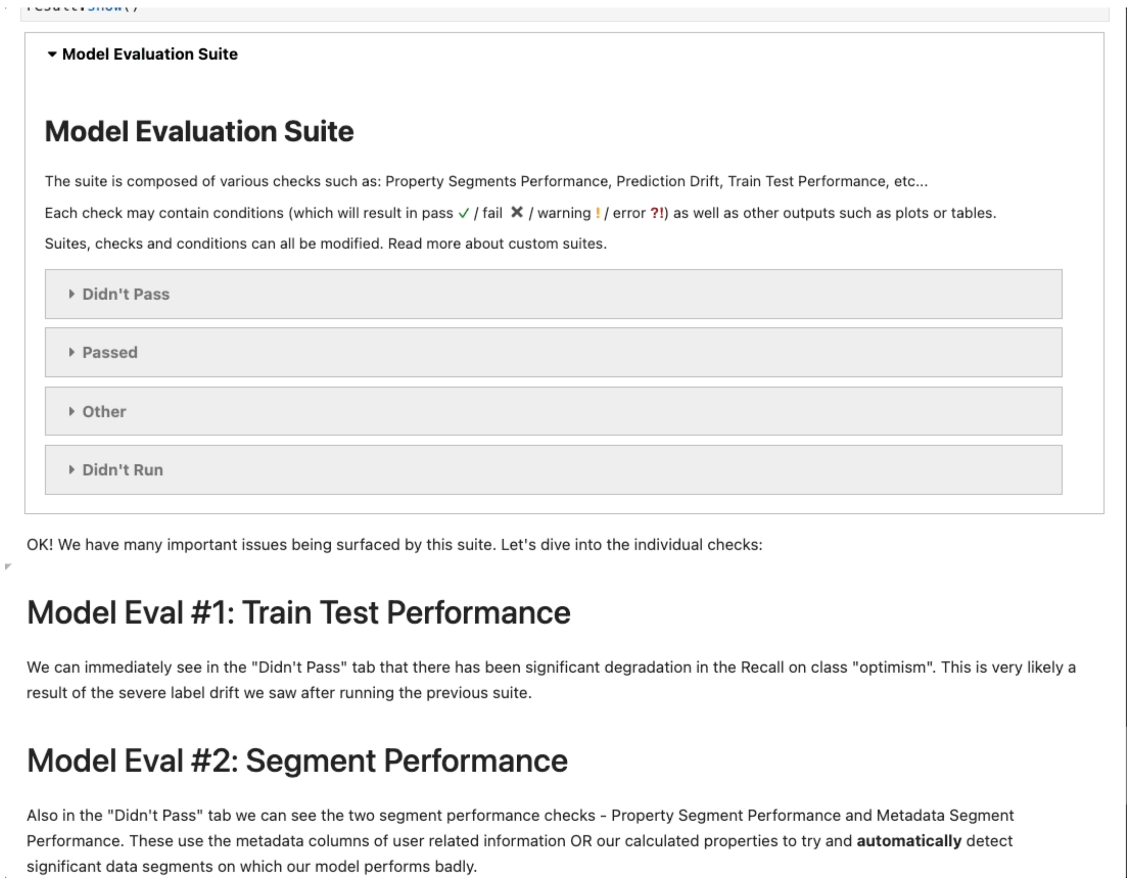
# Deepchecks

Tests for Continuous Validation of ML Models & Data. Deepchecks is a holistic open-source solution for all of your AI & ML validation needs, enabling to thoroughly test your data and models from research to production.



Deepchecks includes:

- **Deepchecks Testing** (Quickstart, docs):
  - Running built-in & your own custom Checks and Suites for Tabular, NLP & CV validation (open source).
- **CI & Testing Management** (Quickstart, docs):
  - Collaborating over test results and iterating efficiently until model is production-ready and can be deployed (open source & managed offering).
- **Deepchecks Monitoring** (Quickstart, docs):
  - Tracking and validating your deployed models behavior when in production (open source & managed offering).



▾ Model Evaluation Suite

## Model Evaluation Suite

The suite is composed of various checks such as: Property Segments Performance, Prediction Drift, Train Test Performance, etc...

Each check may contain conditions (which will result in pass ✓ / fail ✖ / warning ! / error ?!) as well as other outputs such as plots or tables.

Suites, checks and conditions can all be modified. Read more about custom suites.

OK! We have many important issues being surfaced by this suite. Let's dive into the individual checks:

## Model Eval #1: Train Test Performance

We can immediately see in the "Didn't Pass" tab that there has been significant degradation in the Recall on class "optimism". This is very likely a result of the severe label drift we saw after running the previous suite.

## Model Eval #2: Segment Performance

Also in the "Didn't Pass" tab we can see the two segment performance checks - Property Segment Performance and Metadata Segment Performance. These use the metadata columns of user related information OR our calculated properties to try and **automatically** detect significant data segments on which our model performs badly.

twelvefold.

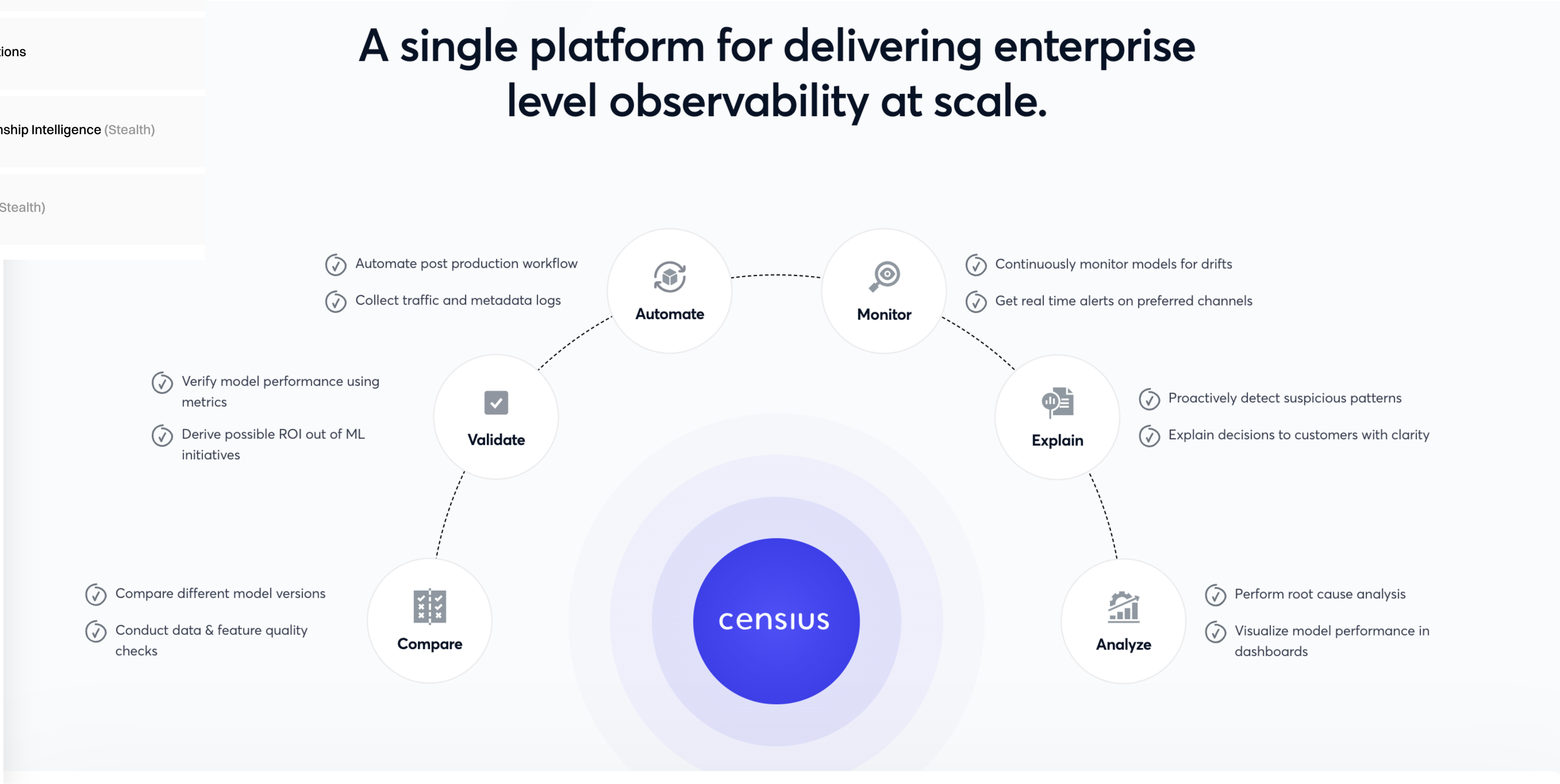
built in ATX

census — AI Observability Platform

attri — Generative AI Solutions

Contextual Relationship Intelligence (Stealth)

Simulated Agents (Stealth)



## A single platform for delivering enterprise level observability at scale.

Automate
- Automate post production workflow
- Collect traffic and metadata logs

Monitor
- Continuously monitor models for drifts
- Get real time alerts on preferred channels

Validate
- Verify model performance using metrics
- Derive possible ROI out of ML initiatives

Explain
- Proactively detect suspicious patterns
- Explain decisions to customers with clarity

Compare
- Compare different model versions
- Conduct data & feature quality checks

Analyze
- Perform root cause analysis
- Visualize model performance in dashboards

# Sentima

The Contextually Aware Converged Security Platform

http://www.sentima.io

### Contextual Awareness

Intent and Context aware platform that defines Why, Where, What, When of a request so proactive security decisions can be made instantly

### Identity Verification and Attestation

Attestation and Verification based User, Workload, Machine, Process, Network, Service Verification and Secure Communication

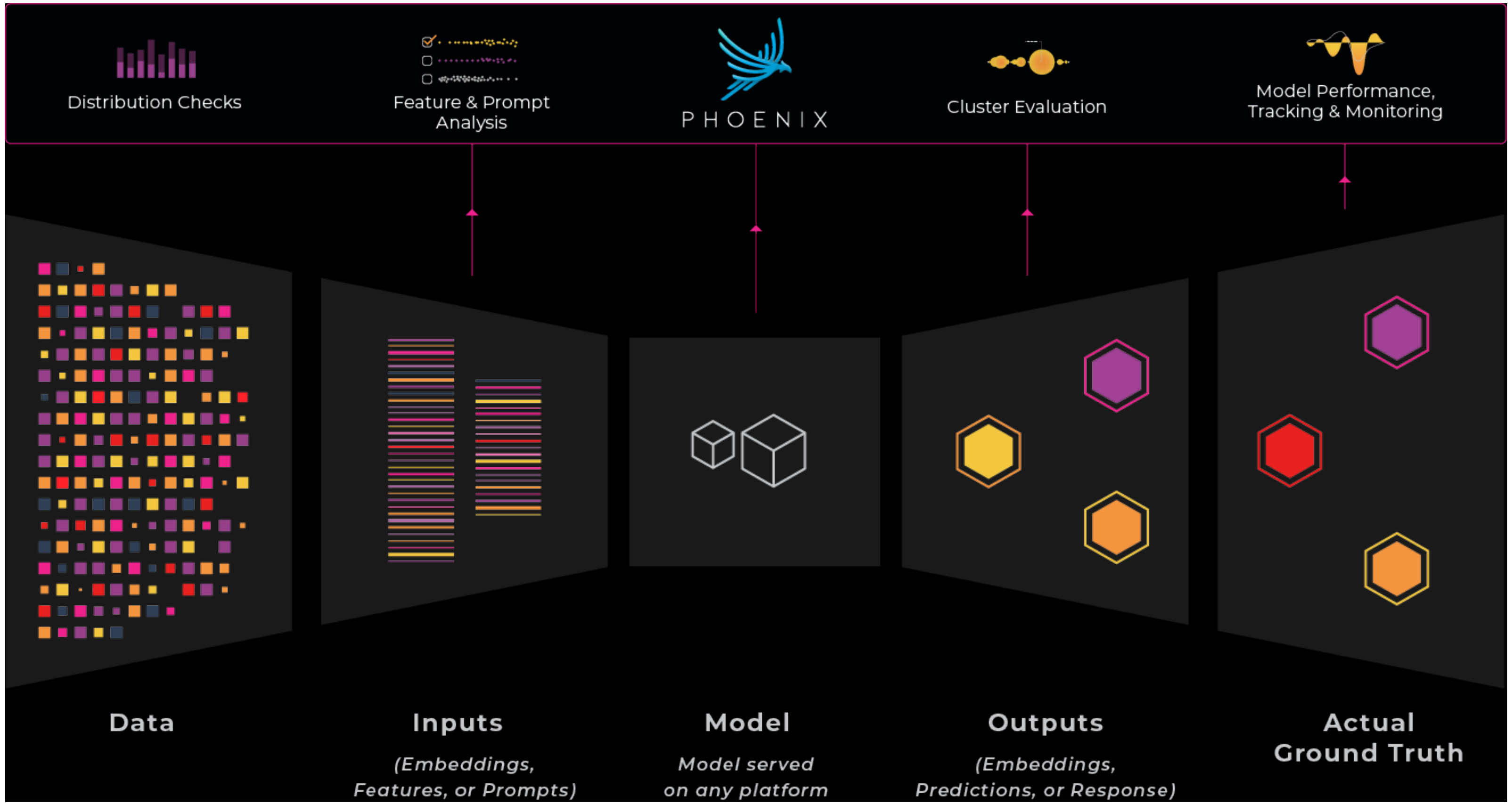
### Point to Point Zero Trust

Point to Point Zero Trust between Service to Data Stores, Service to Service, Users to Services, without Passwords

# Phoenix ML Observability in a Notebook

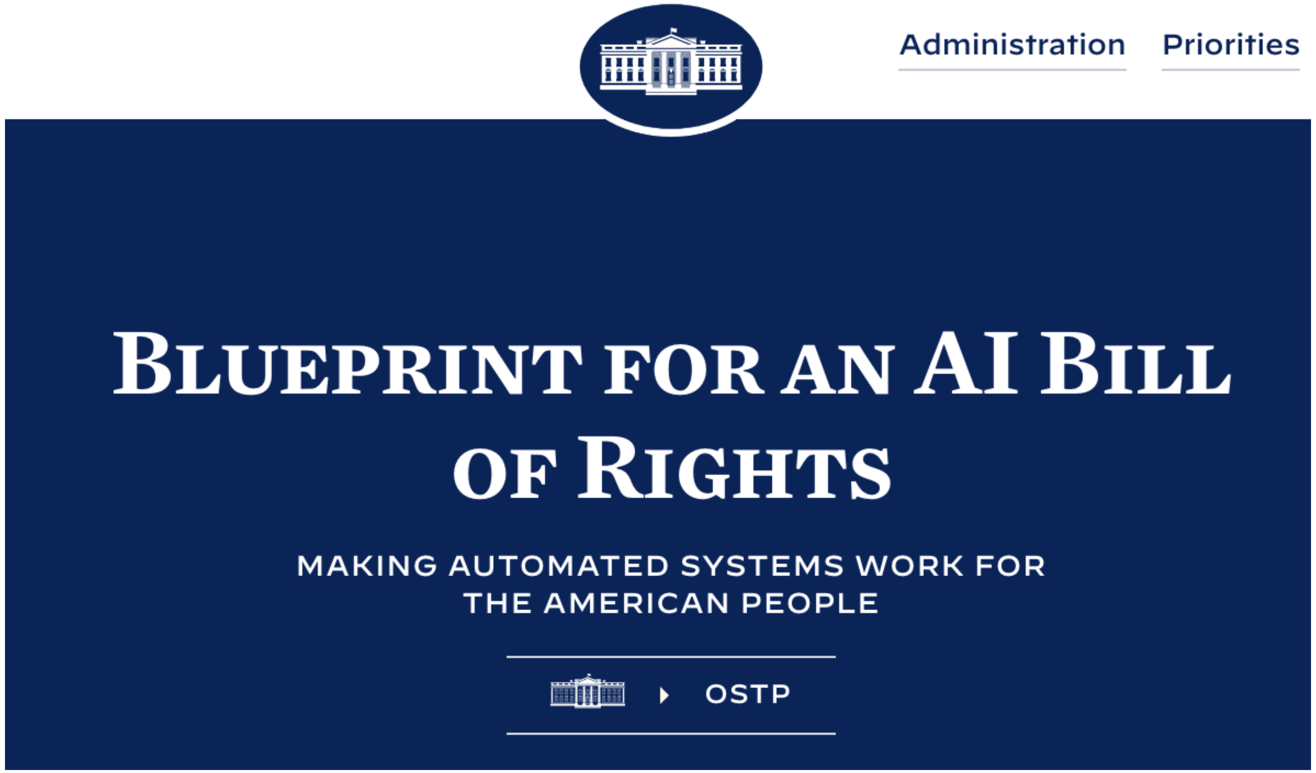
Phoenix provides ML insights at lightning speed with zero-config observability for model drift, performance, and data quality.

Phoenix is an Open Source ML Observability library carefully designed for the Notebook. The toolset is designed to ingest inference data for LLMs, CV, NLP and tabular datasets. It allows Data Scientists and AI Engineers to quickly visualize their inference data, monitor performance, track down issues & insights, and easily export to improve.


Distribution Checks
Feature & Prompt Analysis
PHOENIX
Cluster Evaluation
Model Performance, Tracking & Monitoring
Data
Inputs
(Embeddings, Features, or Prompts)
Model
Model served on any platform
Outputs
(Embeddings, Predictions, or Response)
Actual Ground Truth

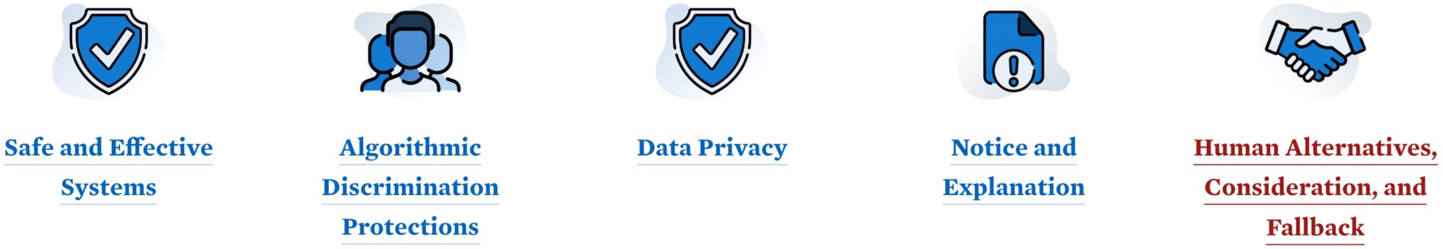
**Phoenix Functionality**

**Discover How Embeddings Represent Your Data:** Map structured features onto embeddings for deeper insights into how embeddings represent your data.
**Evaluate LLM Tasks:** Troubleshoot tasks such as summarization or question/answering to find problem clusters with misleading or false answers.
**Find Clusters of Issues to Export for Model Improvement:** Find clusters of problems using performance metrics or drift. Export clusters for fine-tuning workflows.
**Detect Anomalies:** Using LLM embeddings
**Surface Model Drift and Multivariate Drift:** Use embedding drift to surface data drift for generative AI, LLMs, computer vision (CV) and tabular models.
**Easily Compare A/B Datasets:** Uncover high-impact clusters of data points missing from model training data when comparing training and production datasets.



# Blueprint for an AI Bill of Rights

Making Automated Systems Work for the American People


Safe and Effective Systems
Algorithmic Discrimination Protections
Data Privacy
Notice and Explanation
Human Alternatives, Consideration, and Fallback

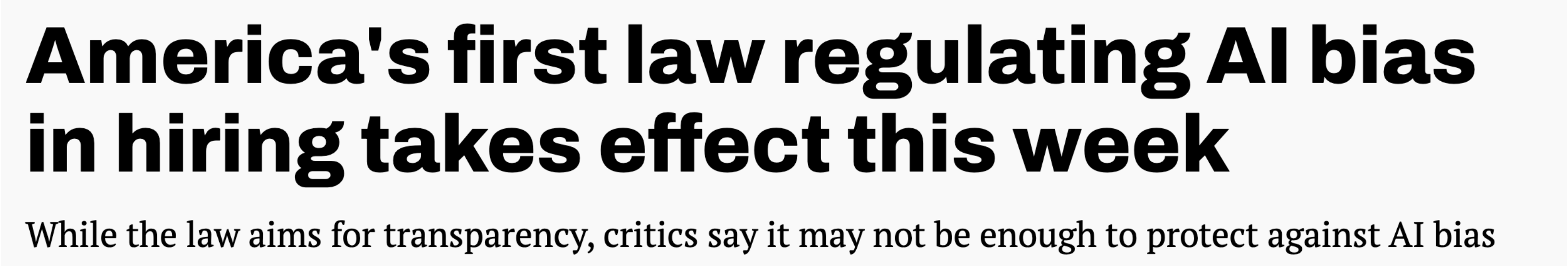
# EU lawmakers pass landmark artificial intelligence regulation

PUBLISHED WED, JUN 14 2023·9:45 AM EDT | UPDATED WED, JUN 14 2023·1:28 PM EDT

- The European Union’s AI Act is the first comprehensive set of regulations for the artificial intelligence industry.

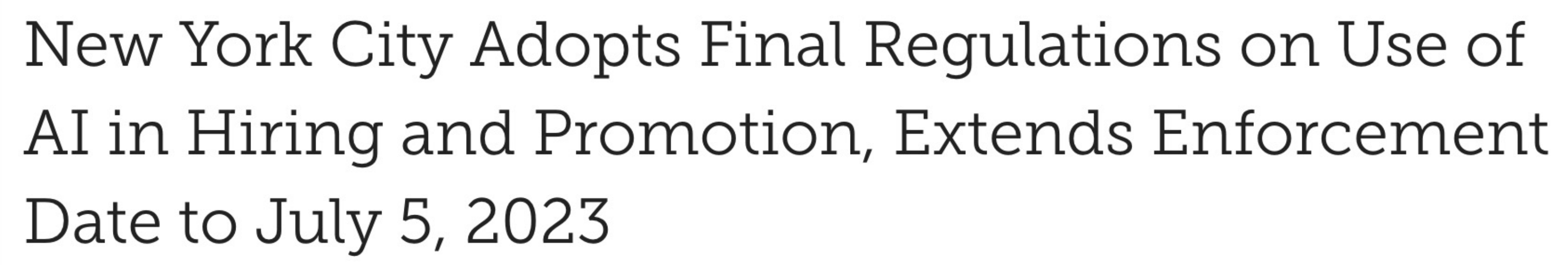
- The law proposes requiring generative AI systems, such as ChatGPT, to be reviewed before commercial release. It also seeks to ban real-time facial recognition.

- It comes as global regulators are racing to get a handle on the technology and limit some of the risks to society, including job security and political integrity.



# America's first law regulating AI bias in hiring takes effect this week

While the law aims for transparency, critics say it may not be enough to protect against AI bias



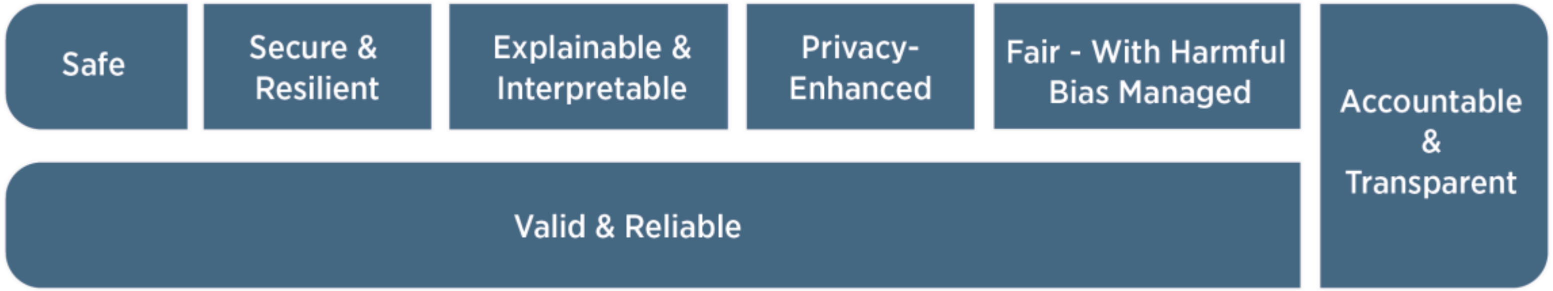
## New York City Adopts Final Regulations on Use of AI in Hiring and Promotion, Extends Enforcement Date to July 5, 2023

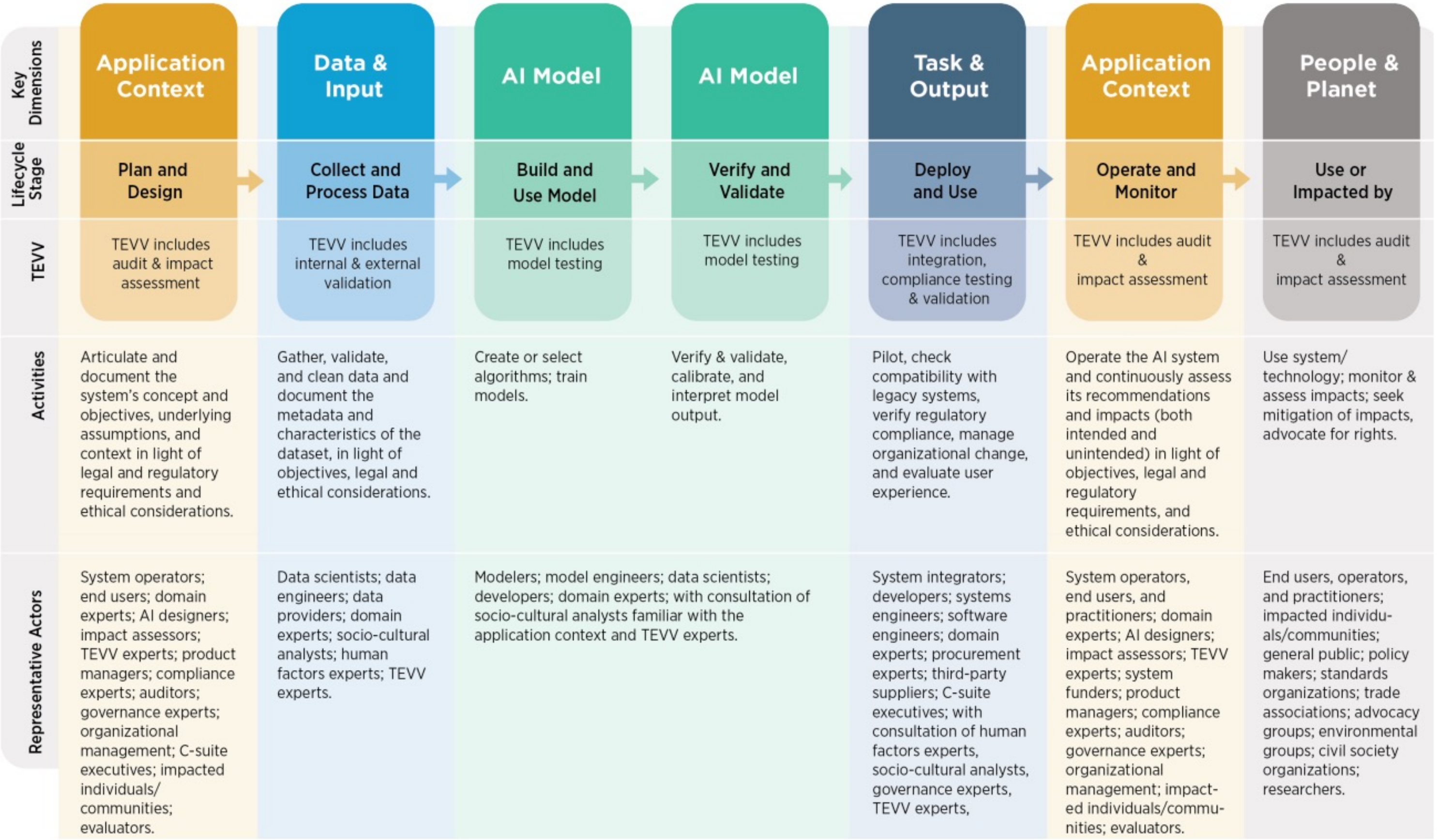
- **Automated resume screeners** that read job applications and recommend the best candidates for an open role
- **Matchmaking algorithms** that scour millions of job postings to recommend roles to candidates—and vice versa
- **Social media scrapers** that collect data on applicants to compile personality profiles based on what they've found online
- **AI-based chatbots** that ask candidates questions about their qualifications, then decide if they'll proceed in the interview process
- **Algorithmic video platforms** that have candidates answer interview questions on camera, record their replies, transcribe their responses, and analyze their vocal or facial patterns for subjective traits like "openness" or "conscientiousness"
- **Logic games** that purport to identify qualities like "risk-taking" or "generosity"

NIST

# AI RMF 1.0

On January 26, 2023, NIST released the AI Risk Management Framework (AI RMF 1.0) along with a companion NIST AI RMF Playbook, AI RMF Explainer Video, an AI RMF Roadmap, AI RMF Crosswalk, and various Perspectives.

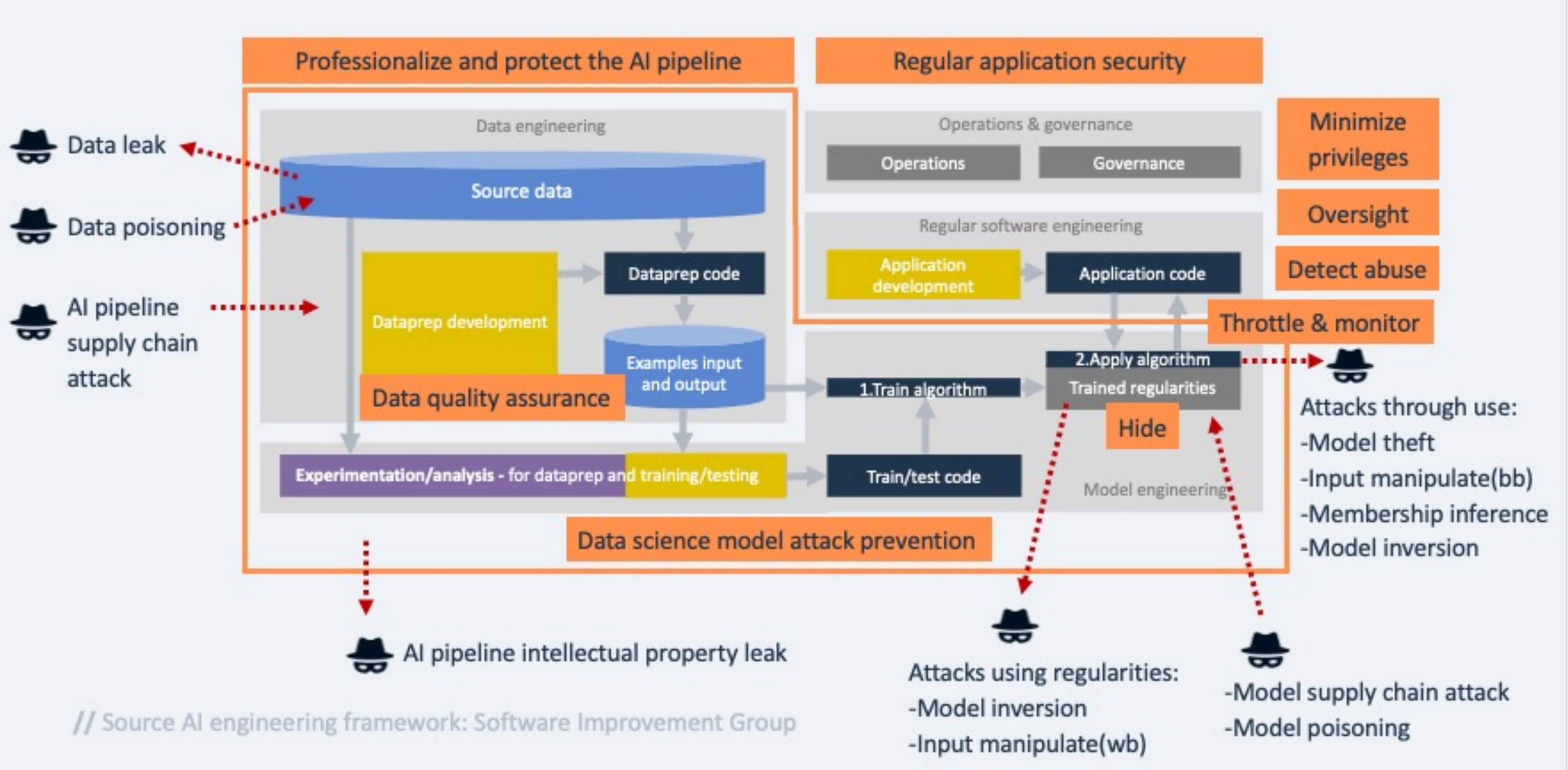

Safe
Secure & Resilient
Explainable & Interpretable
Privacy-Enhanced
Fair - With Harmful Bias Managed
Accountable & Transparent
Valid & Reliable



| Key Dimensions | Application Context | Data & Input | AI Model | AI Model | Task & Output | Application Context | People & Planet |
|---|---|---|---|---|---|---|---|
| Lifecycle Stage | Plan and Design | Collect and Process Data | Build and Use Model | Verify and Validate | Deploy and Use | Operate and Monitor | Use or Impacted by |
| TEVV | TEVV includes audit & impact assessment | TEVV includes internal & external validation | TEVV includes model testing | TEVV includes model testing | TEVV includes integration, compliance testing & validation | TEVV includes audit & impact assessment | TEVV includes audit & impact assessment |
| Activities | Articulate and document the system's concept and objectives, underlying assumptions, and context in light of legal and regulatory requirements and ethical considerations. | Gather, validate, and clean data and document the metadata and characteristics of the dataset, in light of objectives, legal and ethical considerations. | Create or select algorithms; train models. | Verify & validate, calibrate, and interpret model output. | Pilot, check compatibility with legacy systems, verify regulatory compliance, manage organizational change, and evaluate user experience. | Operate the AI system and continuously assess its recommendations and impacts (both intended and unintended) in light of objectives, legal and regulatory requirements, and ethical considerations. | Use system/technology; monitor & assess impacts; seek mitigation of impacts, advocate for rights. |
| Representative Actors | System operators; end users; domain experts; AI designers; impact assessors; TEVV experts; product managers; compliance experts; auditors; governance experts; organizational management; C-suite executives; impacted individuals/communities; evaluators. | Data scientists; data engineers; data providers; domain experts; socio-cultural analysts; human factors experts; TEVV experts. | Modelers; model engineers; data scientists; developers; domain experts; with consultation of socio-cultural analysts familiar with the application context and TEVV experts. | | System integrators; developers; systems engineers; software engineers; domain experts; procurement experts; third-party suppliers; C-suite executives; with consultation of human factors experts, socio-cultural analysts, governance experts, TEVV experts. | System operators, end users, and practitioners; domain experts; AI designers; impact assessors; TEVV experts; system funders; product managers; compliance experts; auditors; governance experts; organizational management; impacted individuals/communities; evaluators. | End users, operators, and practitioners; impacted individuals/communities; general public; policy makers; standards organizations; trade associations; advocacy groups; environmental groups; civil society organizations; researchers. |

OWASP
AI security & privacy guide

# OWASP Machine Learning Security Top Ten



Professionalize and protect the AI pipeline

Regular application security

Data engineering

Source data

Dataprep code

Dataprep development

Examples input and output

Data quality assurance

Experimentation/analysis - for dataprep and training/testing

Operations & governance

Operations

Governance

Regular software engineering

Application development

Application code

Minimize privileges

Oversight

Detect abuse

Throttle & monitor

1.Train algorithm

2.Apply algorithm

Trained regularities

Hide

Train/test code

Model engineering

Data science model attack prevention

Data leak

Data poisoning

AI pipeline supply chain attack

AI pipeline intellectual property leak

Attacks through use:
-Model theft
-Input manipulate(bb)
-Membership inference
-Model inversion

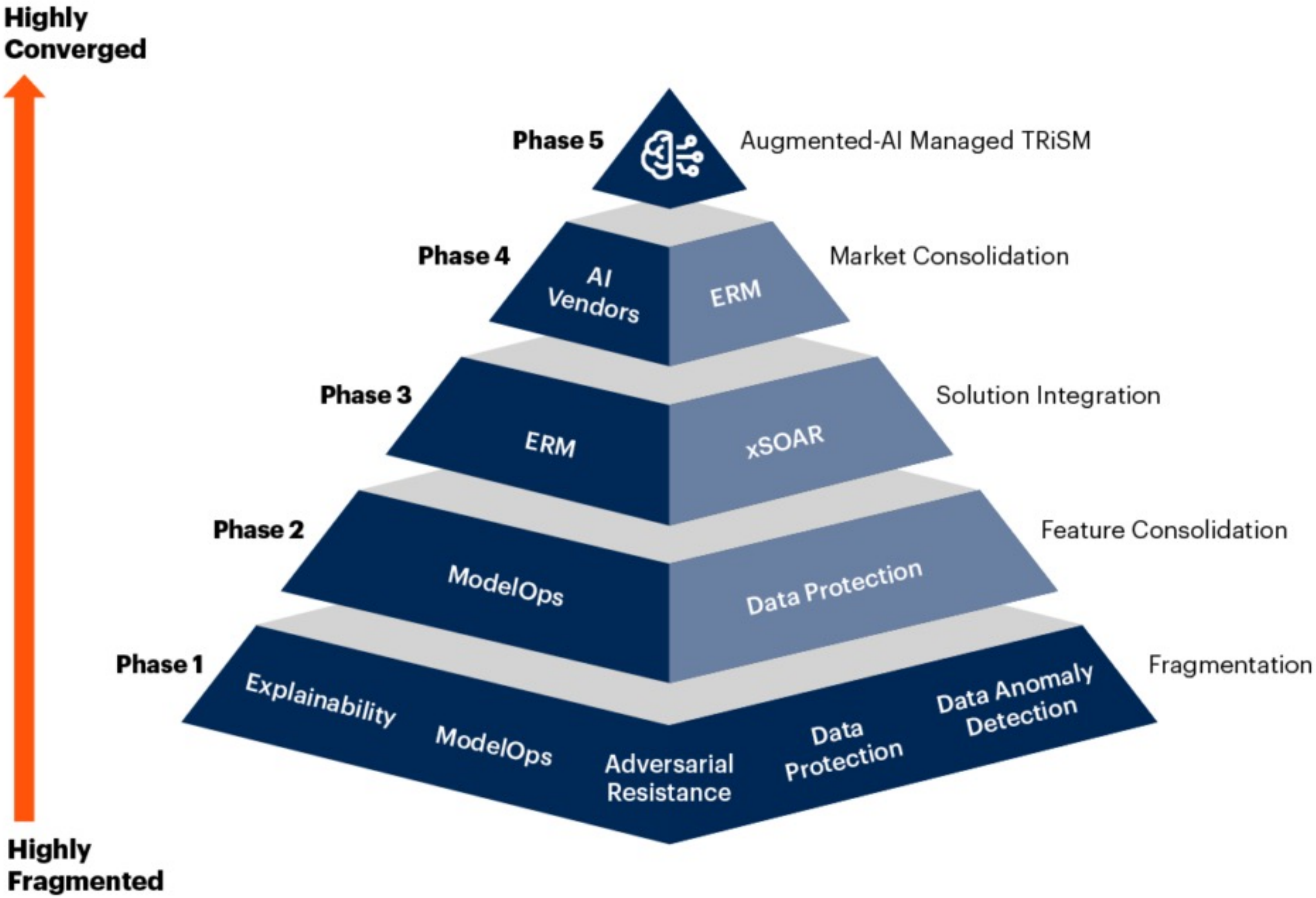
Attacks using regularities:
-Model inversion
-Input manipulate(wb)

-Model supply chain attack
-Model poisoning

// Source AI engineering framework: Software Improvement Group

- ML01:2023 Adversarial Attack
- ML02:2023 Data Poisoning Attack
- ML03:2023 Model Inversion Attack
- ML04:2023 Membership Inference Attack
- ML05:2023 Model Stealing
- ML06:2023 Corrupted Packages
- ML07:2023 Transfer Learning Attack
- ML08:2023 Model Skewing
- ML09:2023 Output Integrity Attack
- ML10:2023 Neural Net Reprogramming

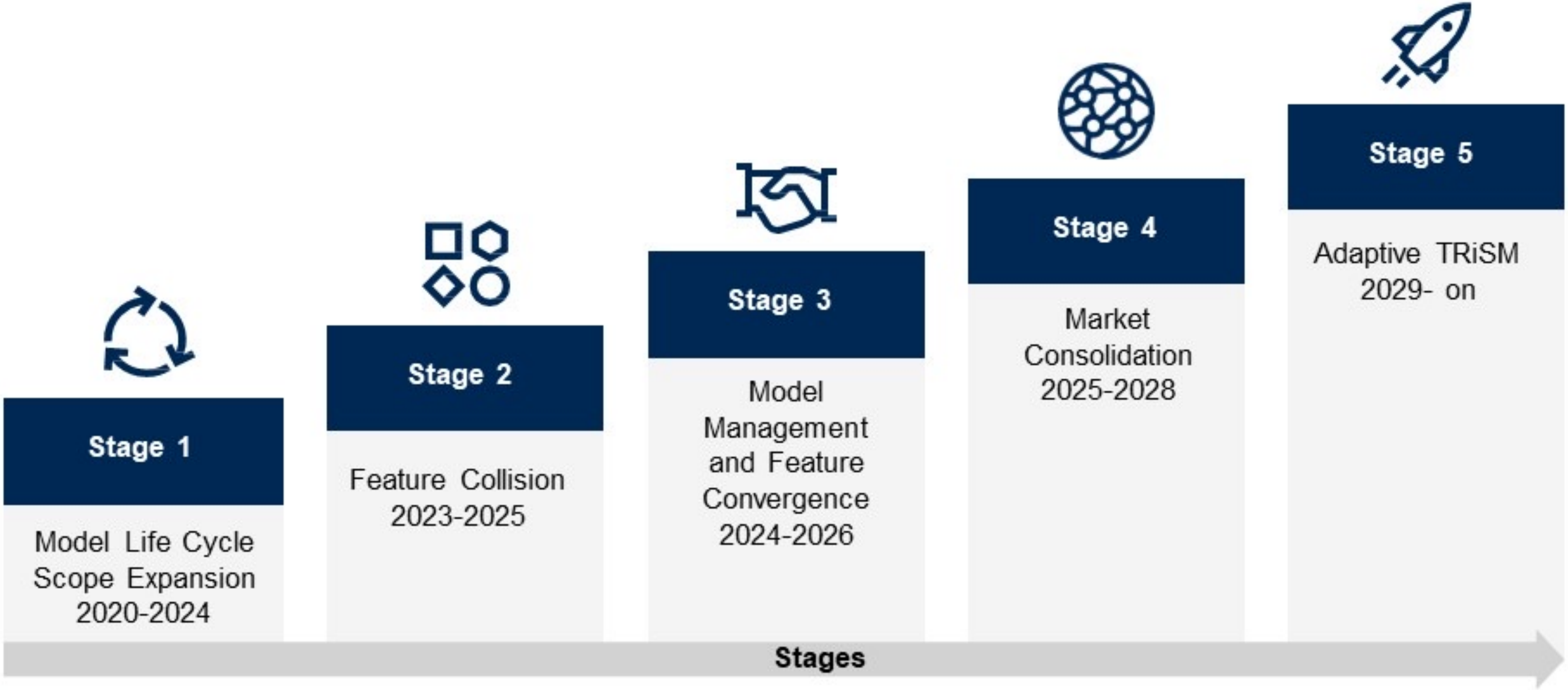
## Future Direction AI TRiSM Market



Source: Gartner

750738_C

# Future Direction of the AI TRiSM Market



**Stage 1**

Model Life Cycle Scope Expansion
2020-2024

**Stage 2**

Feature Collision
2023-2025

**Stage 3**

Model Management and Feature Convergence
2024-2026

**Stage 4**

Market Consolidation
2025-2028

**Stage 5**

Adaptive TRiSM
2029- on

**Stages**

Source: Gartner

758388_C

# North Austin Tech User Group : AI Focused

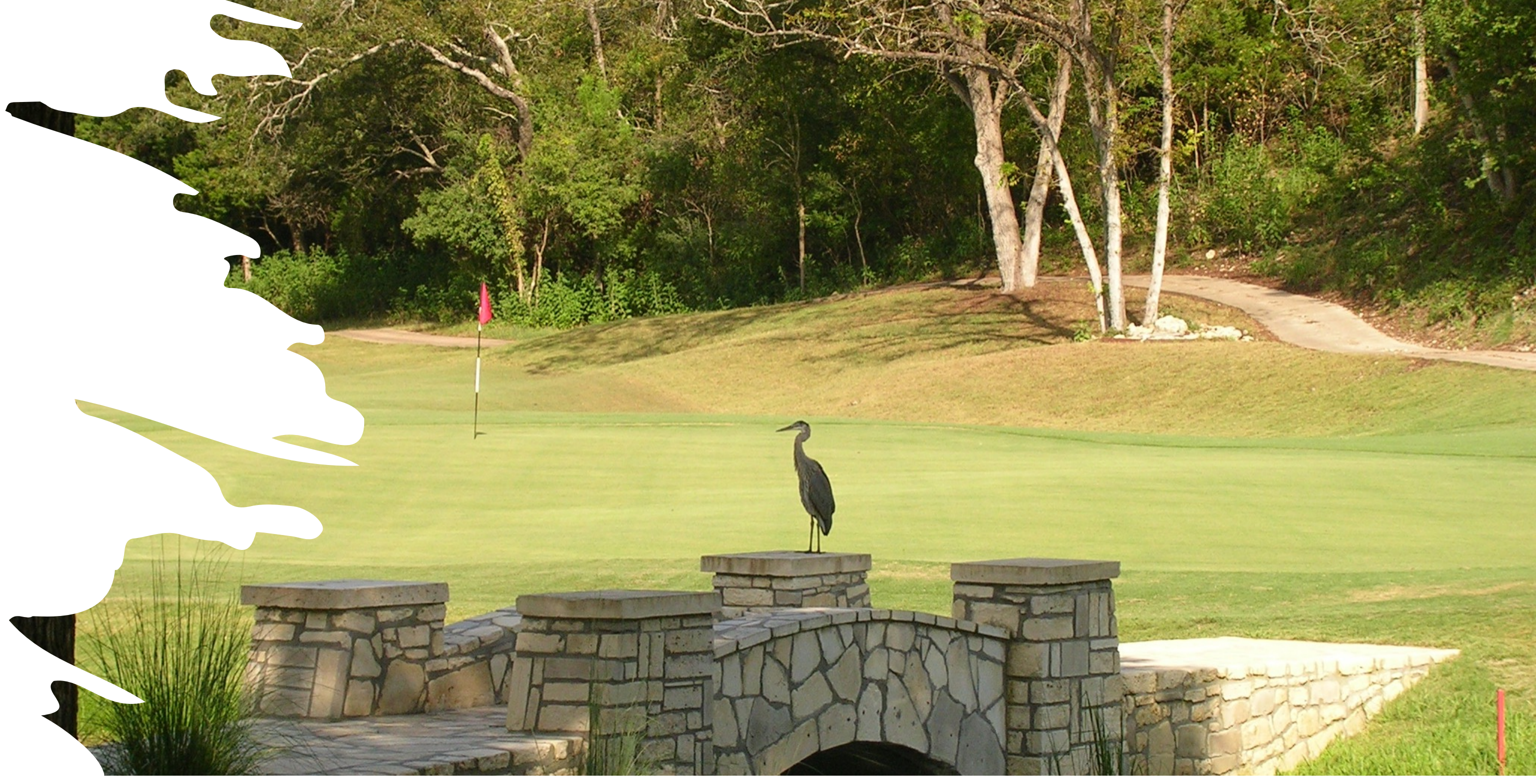
# NATU.AI

Connect Me at LinkedIn

**/in/mandavasuresh**

**Suresh Mandava**
SVP/Chief Architect
Cloud-Native AI/ML Platforms and Security
Leander, Austin, Texas : 636-634-0552

Linkedin : /in/MandavaSuresh
Twitter : @sureshmandava



**AITX Meetup**