

# Enabling Healthcare Operations with Agentic AI - DISHA Overview and Architecture

## Dr Sucheendra K Palaniappan, VP of Data Science & Engineering of SBX Corporation

Sucheendra K. Palaniappan is Vice President of Data Science and Engineering at SBX Corporation and Chief Technology Officer of SBX Technologies, a subsidiary focused on healthcare innovation. He is also a Senior Scientist at the Systems Biology Institute in Tokyo.

He holds a Ph.D. in Computer Science from the National University of Singapore and has primarily worked at the intersection of biomedicine and healthcare, with a focus on translational data science and applied machine learning.

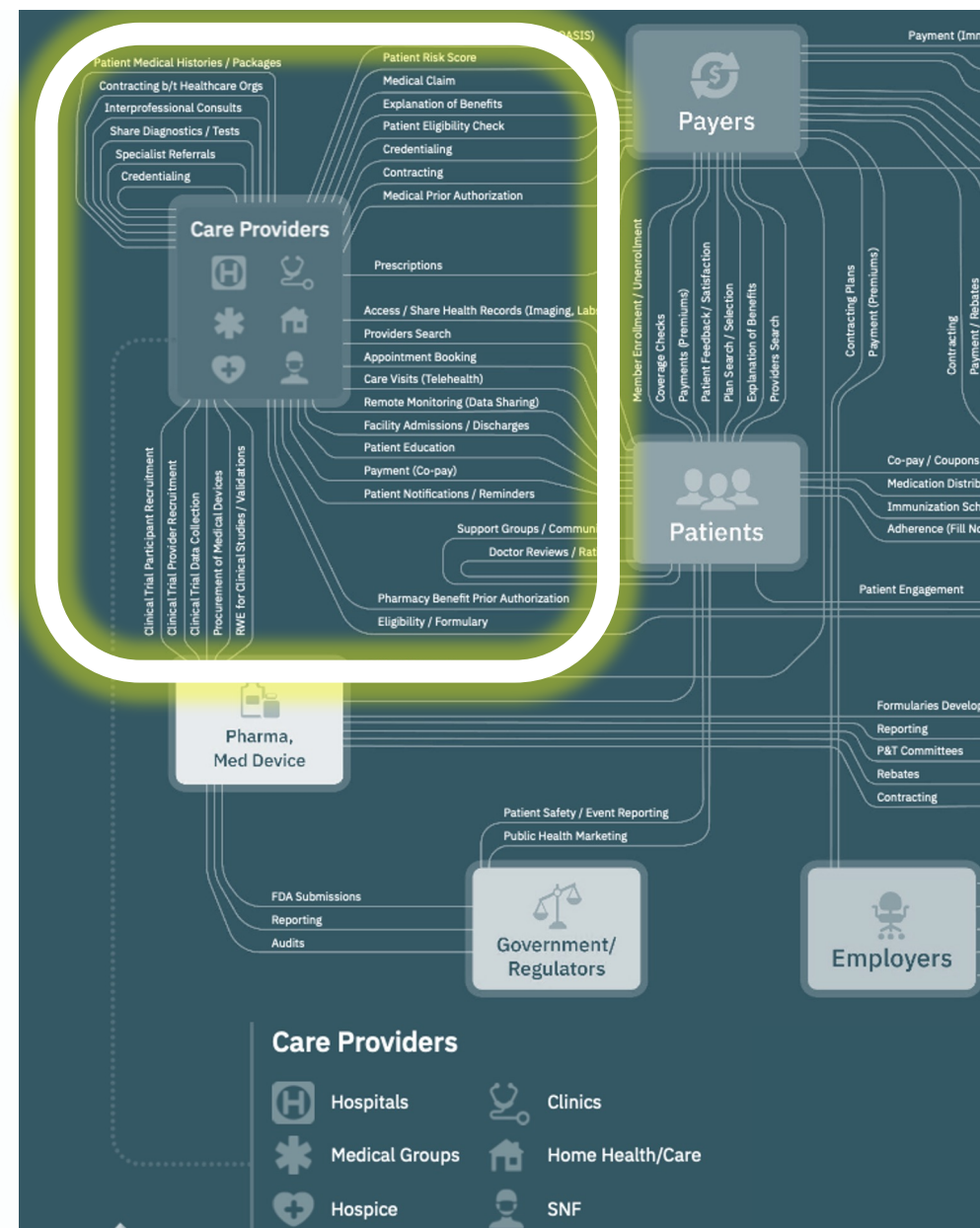


# Enabling Healthcare Operations with Agentic AI : DISHA Overview and Architecture

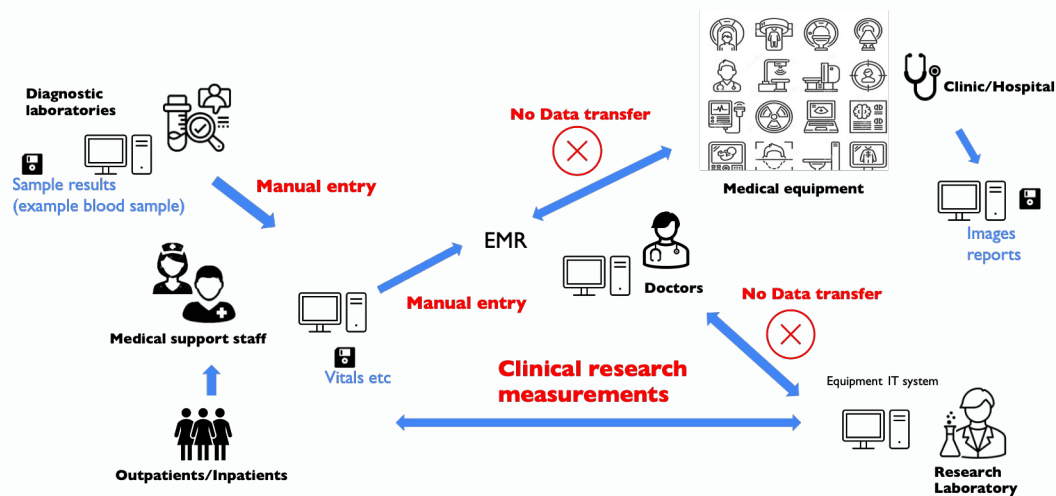
Setting the direction for the future of Clinical Decision Intelligence (CD<sub>x</sub>)



**SBX Group**  
Circa 2000 - Present



# The Supply Side Bottleneck: Lack of “Agency”



Typical workflow in a small hospital/clinic

- Fragmented workflows and siloed data disrupt coordination between teams and increase administrative burden for clinicians
- Manual processes and legacy IT slow processes, introduce errors, and add to provider workload
- Poor data readiness blocks analytics, AI, and real-time clinical decision support
- Inefficiency erodes margins, limits scalability, and contributes to staff burnout
- Inflexible, non-interoperable systems make modernization costly and frustrate care delivery



# Overview of Agentic AI

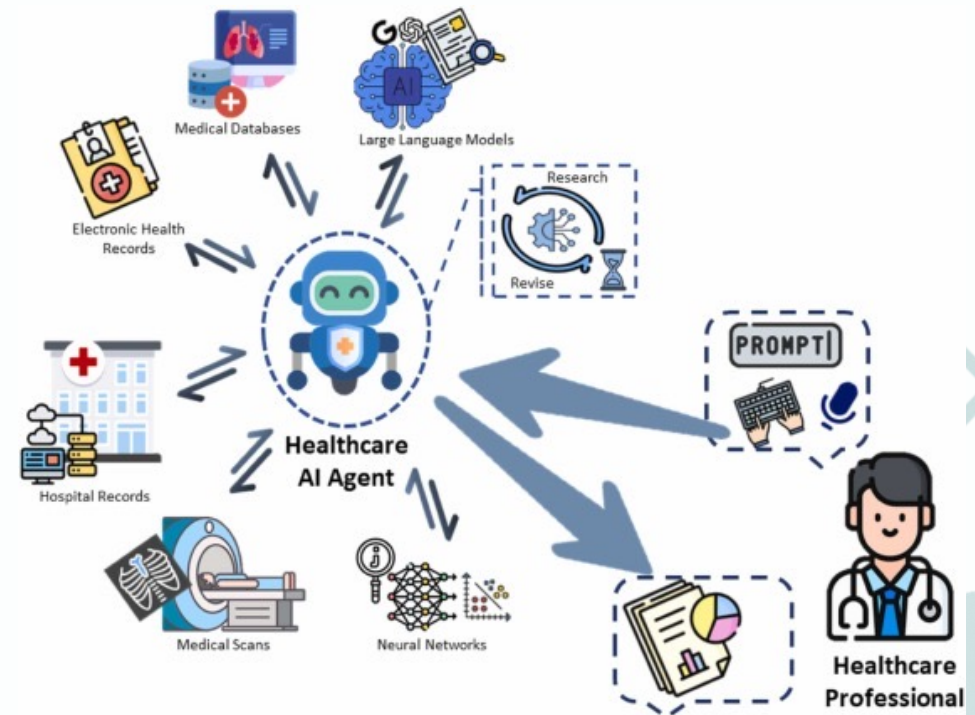
## Agentic $\neq$ LLMs

**LLMs** answer on prompt; no tools/state

**Reasoning systems** can plan/Chain of Thought but still passive/reactive

**Agentic systems** plan  $\rightarrow$  act (tools/APIs)  
 $\rightarrow$  observe  $\rightarrow$  adapt with aspects of  
Memory, policies, goal-directed outcomes

**Agents** add structure and orchestration for goal-directed behavior



Karunanayake, Nalan. "Next-generation agentic AI for transforming healthcare." Informatics and Health 2.2 (2025): 73-83.

# Core components of an Agentic Stack



- Planner & policy engine
- Tool executor (connectors to systems – Model context protocol)
- Memory/state & caches
- Guardrails (allow-lists, abuse filters)
- Observability (traces, audit), Human-in-loop gates

## Standards for Communication

- **MCP (Model Context Protocol)**: consistent agent ↔ tool communication
- **A2A (Agent-to-Agent, Google)**: discovery & messaging across agents/systems



# Some Agentic Frameworks

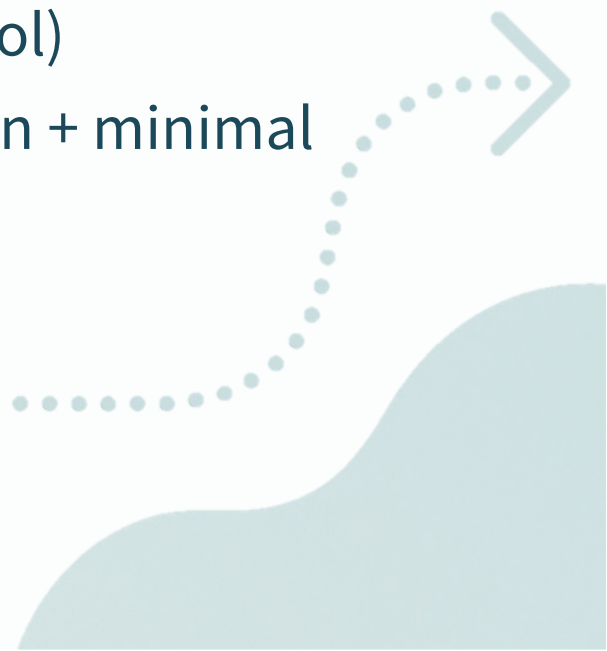


Category	Framework / Tool	Highlights / Example Use Case
Orchestration Runtimes	LangGraph	Stateful graph programming for long-running, agentic flows with human-in-loop.
	LangChain Agents (ReAct)	Prompt-based agent model; now evolving into LangGraph for production.
Workflow Engine	n8n (AI Agent Node)	Low-code orchestrator; integrates agents ↔ triggers ↔ APIs seamlessly.
Multi-agent Coordination (A2A)	AutoGen	Facilitates multi-agent dialogue, tool use, and supervisor agents.
	CrewAI	Manages role-based agent teams ("crews") for collaborative workflows.
Guardrail Toolkit	NVIDIA NeMo Guardrails	Policy layer to restrict/outbound calls and unsafe LLM generations.



# Healthcare : Agentic ≠ Agency



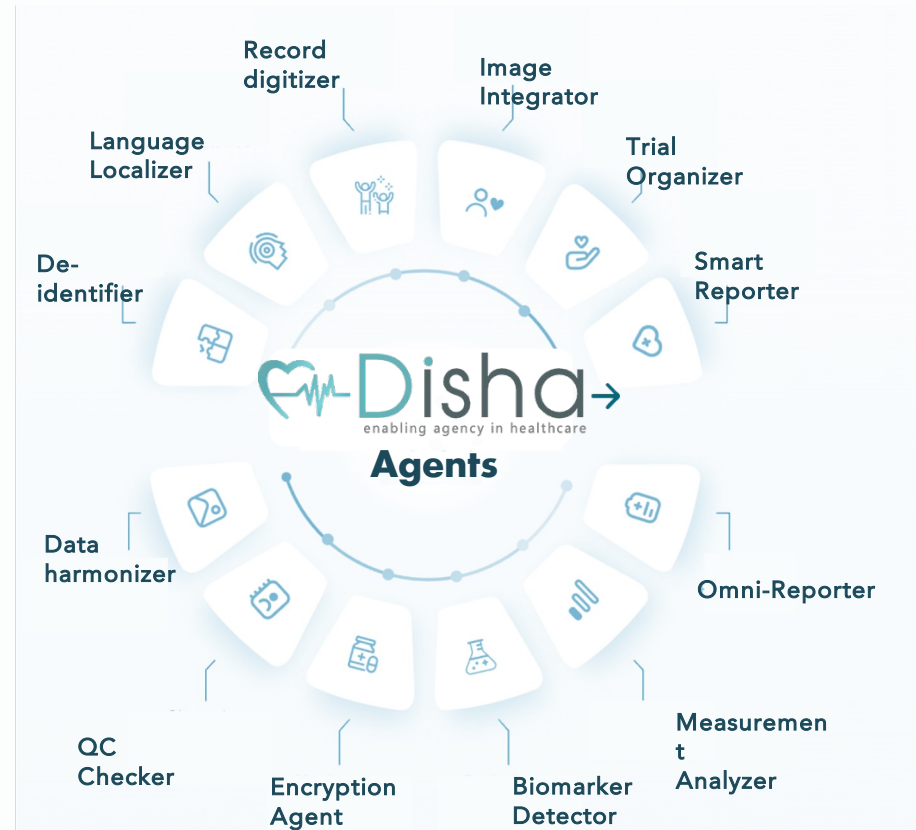
- Safety & Compliance Risks with Unbounded Autonomy
  - PHI compliance ( Edge vs Cloud), What is shared? where it is shared? how it is shared? With whom it is shared?
  - **Strong** Need for Human Oversight ( Human in control)
  - Need for plug and play solutions + easy IT integration + minimal setup
  - Role scoped solutions
- 



# Building a system of intelligent agents

**Disha Platform** powers hospital operations with a **system of interoperable, task-specific agents**

- Agents automate key tasks and workflows, reducing manual effort and errors
- Seamlessly operate within existing hospital systems-no disruption
- Each agent delivers the 3 Is: Information, Intelligence, and Interface for actionable results



# Disha Box

## Flexible, Secure, and Future-Ready Deployment

- Deploy Anywhere
  - On-premises, cloud, or as a portable “Disha Box” (edge device)
  - Local deployment ensures patient data privacy and regulatory compliance-no data leaves your premises
- Plug & Play Integration
  - Secure APIs, custom UIs, or hybrid modes
  - Designed to be interoperable with diverse devices, and legacy systems
- Minimal IT Burden
  - Pre-configured Device, Minimizes need for specialist IT staff and reduces operational complexity



### Disha Box\*

( comes with agents preinstalled, plug and play)

- 2048-core Ampere GPU, 64GB LPDDR5 RAM, 4TB NVMe SSD, with HDMI, USB, Gigabit Ethernet, and CSI camera ports.
- DISHA Box runs compact LLMs and multimodal models tuned for hospital use—up to 12B parameters—locally on-site.
- Plug-and-play setup with built-in privacy; designed for steady, reliable performance.

\*under development

# Disha Box

## Flexible, Secure, and Future-Ready Deployment

- Scalable & Modular
  - Add/remove agents as needs change; supports multi-site and multi-department rollouts
- Attach physical sensors seamlessly
  - Camera, Microphone, Speaker etc
- Role based Provisioning
  - Each box can be provisioned for a specific role ( i.e bed, Nurse, technician etc)
- Future-Ready
  - Built for new protocols, instantly compatible with LLMs and AI tools



### Disha Box\*

( comes with agents preinstalled, plug and play)

- 2048-core Ampere GPU, 64GB LPDDR5 RAM, 4TB NVMe SSD, with HDMI, USB, Gigabit Ethernet, and CSI camera ports.
- DISHA Box runs compact LLMs and multimodal models tuned for hospital use—up to 12B parameters—locally on-site.
- Plug-and-play setup with built-in privacy; designed for steady, reliable performance.

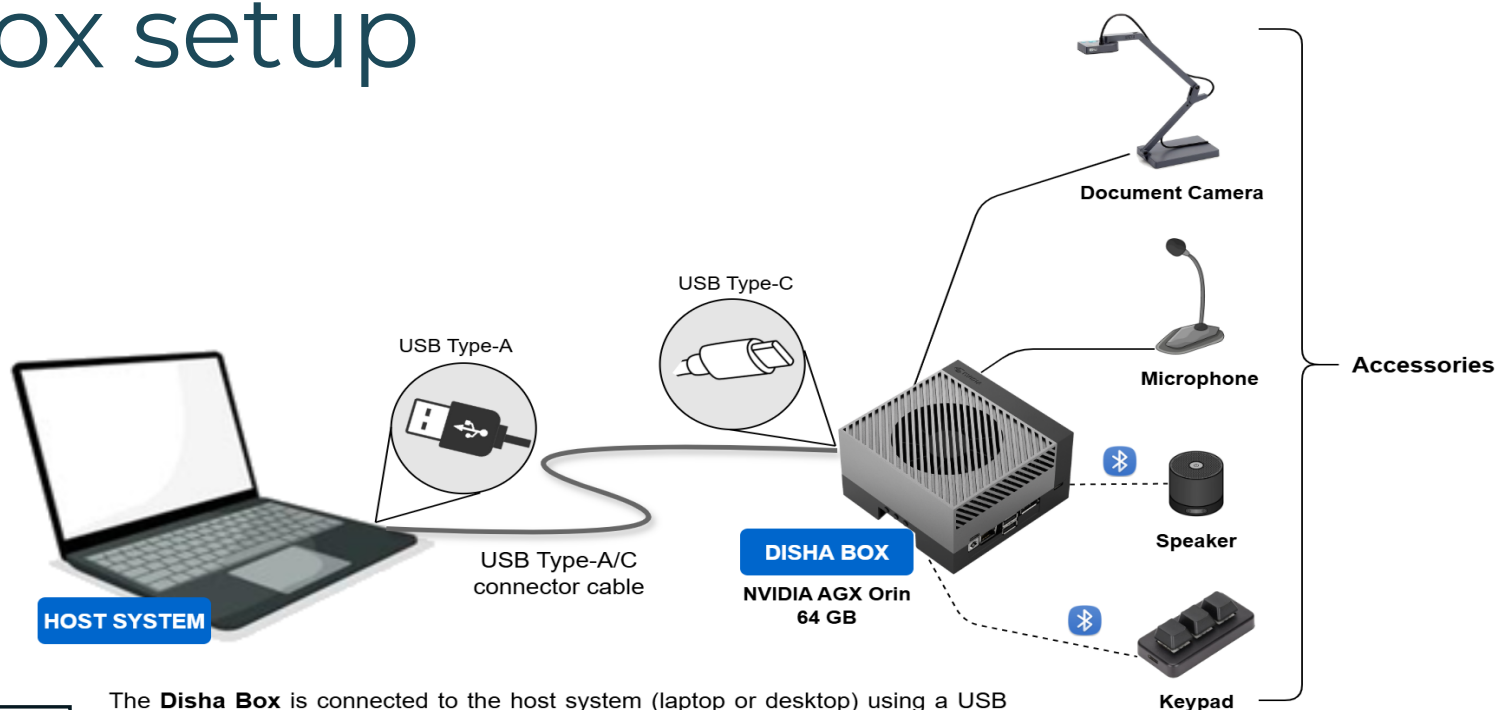
\*under development

# Disha box setup

## Host System

The host system can be any laptop or desktop where the Disha Box is connected using a USB Type-A and/or Type-C connector cable.

User can access the agents using browser only. The box will be available by the domain name <https://disha.local>



## NVIDIA AGX Orin 64GB Specification

**GPU:** NVIDIA Ampere architecture with 2048 NVIDIA CUDA® cores and 64 Tensor Cores

**CPU:** 12-core Arm Cortex-A78AE v8.2 64-bit CPU

**Memory:** 64GB LPDDR5 (256-bit interface) with 204.8 GB/s bandwidth

**Storage:** 64GB eMMC 5.1 + 2TB SSD

The **Disha Box** is connected to the host system (laptop or desktop) using a USB Type-A and/or Type-C connector cable. The accessories are connected via USB or bluetooth.



NO INTERNET  
CONNECTION REQUIRED



POWERED BY LARGE  
LANGUAGE MODELS (LLM)



PLUG AND PLAY SUPPORT



AGENT-DRIVEN SERVICES



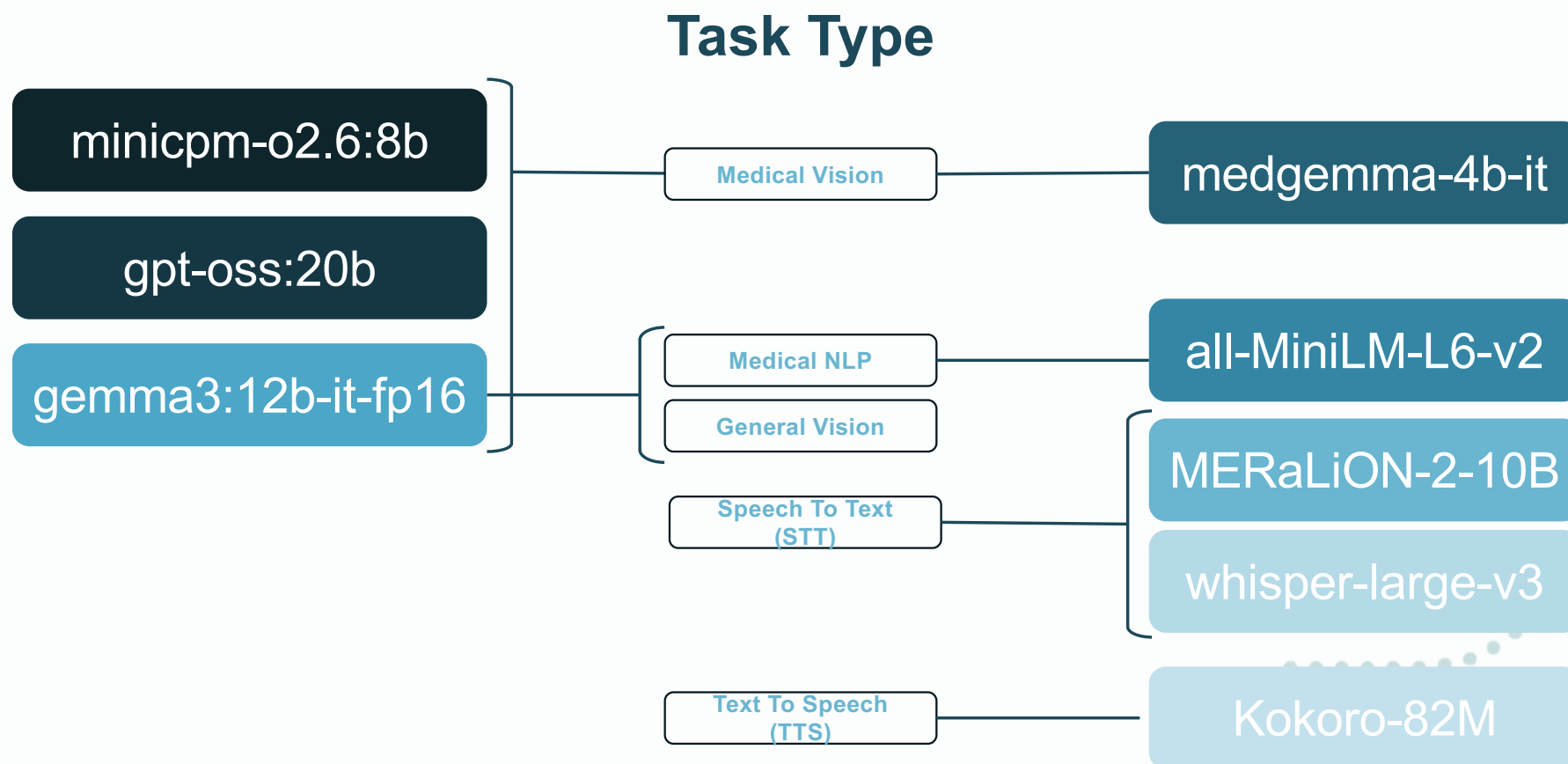
MEETS ALL HIPPA AND  
MEDICAL GUIDELINES



enabling agency in healthcare



# Model(s) Summary



# Agent Model Mapping

Agent	Primary Model	Secondary Model	Task Type
Prescription Reader	<a href="#">google/medgemma-4b-it</a>	<a href="#">gemma3:12b-it-fp16</a>	Medical Vision
Discharge Summary	<a href="#">gpt-oss:20b</a>	<a href="#">openbmb/minicpm-o2.6:8b</a>	Medical Vision
Speech To Text (MERaLiON-2-10B)	<a href="#">MERaLiON/MERaLiON-2-10B</a>		STT
ICDMapper	<a href="#">gemma3:12b-it-fp16</a>	<a href="#">all-MiniLM-L6-v2</a>	Medical NLP
Notes Reader	<a href="#">gemma3:12b-it-fp16</a>		General Vision
Radiology Explainer	<a href="#">google/medgemma-4b-it</a>	<a href="#">gemma3:12b-it-fp16</a>	Medical Vision
Measurement Analyzer	<a href="#">google/medgemma-4b-it</a>	<a href="#">gemma3:12b-it-fp16</a>	Medical Vision
Text To Speech (TTS)	<a href="#">hexgrad/Kokoro-82M</a>		TTS
Speech To Text (STT)	<a href="#">openai/whisper-large-v3</a>		STT
Report of Reports (RoR)	<a href="#">gpt-oss:20b</a>	<a href="#">google/medgemma-4b-it</a> <a href="#">openbmb/minicpm-o2.6:8b</a>	Medical Vision

# Tech Stack

## Frontend



**React.JS**  
v19.0 RC

## Middleware



**Next.JS**  
v15.0

## Backend



**Python**  
v3.10

## DB / Object Store / Secret Store



**MongoDB**  
v8.0.4



**MinIO**  
2024-12-18T13:15:44Z



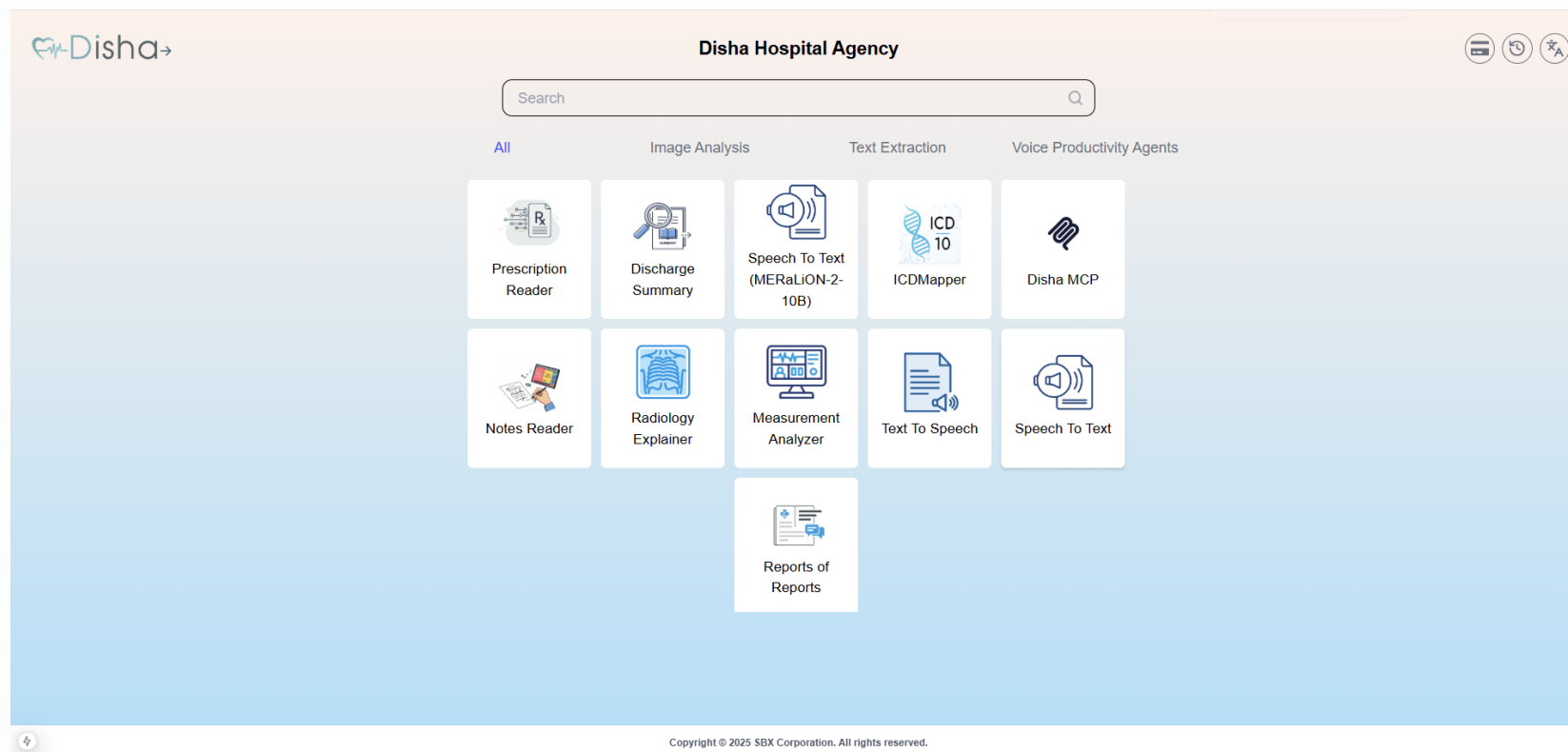
**Vault**  
v1.20.0

## Deployment



Docker  
Docker Compose

# The Disha Agency

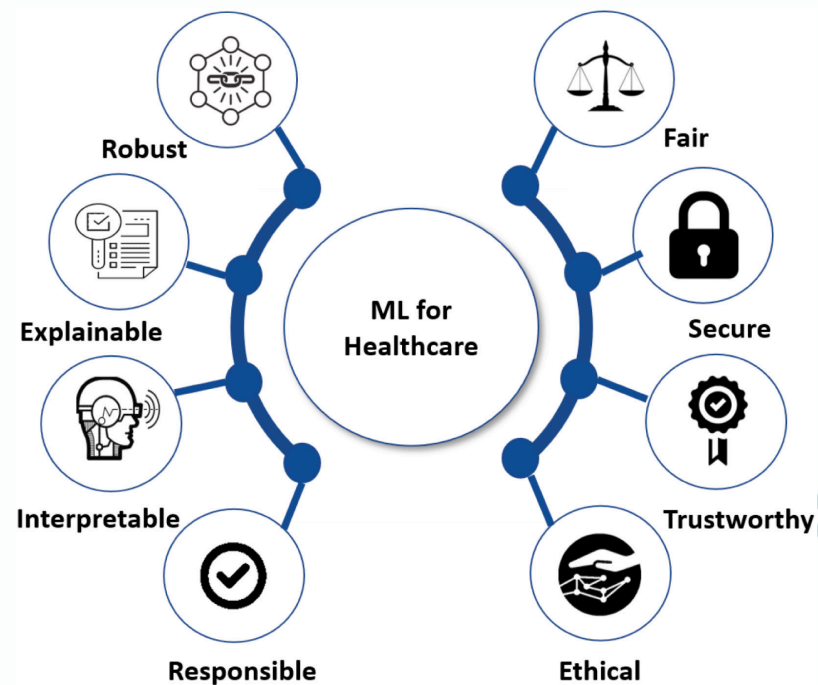




# Agency Showcase: ROI

## Reporting & actionable Insights

- Reports (both in text narrative, tabular quantities and images) are a critical component of hospital operations and clinical practice
- Deliberate Integration: Data and reports need to operate with existing HMS
- Ability to generate actionable insights in a trustable, explainable and interpretable manner is central to healthcare



Rasheed, Khansa, et al. "Explainable, trustworthy, and ethical machine learning for healthcare: A survey." *Computers in Biology and Medicine* (2022): 106043.

LIMS: Laboratory Information Management System  
HMS: Hospital Management Systems

# ROI Agents showcase



Prescription  
Reader



Notes Reader



Radiology  
Explainer



Measurement  
Analyzer



Reports of  
Reports



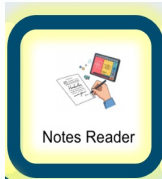
Trial Organizer



Disha Clinic



Disha Lab



Notes Reader

Disha→ Notes Reader

## Reading, structuring and reporting of clinical notes



Samples :

Sample 1

Sample 2

Drag & drop a file here

OR

Select File

Text (English)

Text (Japanese)

Structured (English)

Structured (Japanese)

Date: 8/27/08

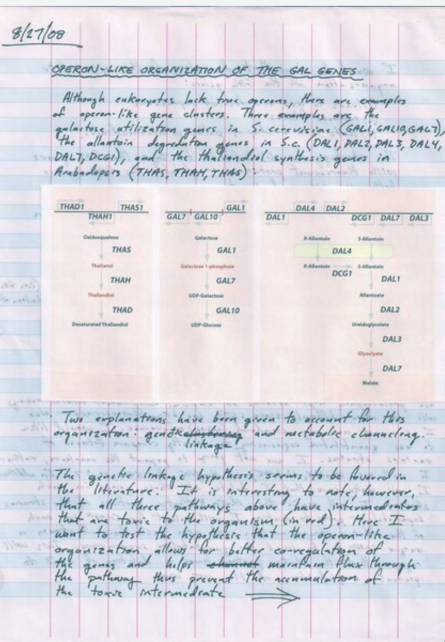
Title: OPERON-LIKE ORGANIZATION OF THE GAL GENES

Content:

- Although eukaryotes lack true operons, there are examples of operon-like gene clusters. Three examples are the galactose utilization genes in *S. cerevisiae* (GAL1, GAL10, GAL7), the allantoin degradation genes in *S.c.* (DAL1, DAL2, DAL3, DAL4, DAL7, DCG1), and the thalidomide synthesis genes in *Arabidopsis* (THAS, THAH, THAS).
- Two explanations have been given to account for this organization: genetic linkage and metabolic channeling.
- The genetic linkage hypothesis seems to be favored in the literature. It is interesting to note, however, that all three pathways above have intermediates that are toxic to the organism (in red). Here I want to test the hypothesis that the operon-like organization allows for better co-regulation of the genes and helps channel maintain flux through the pathway thus prevent the accumulation of the toxic intermediate

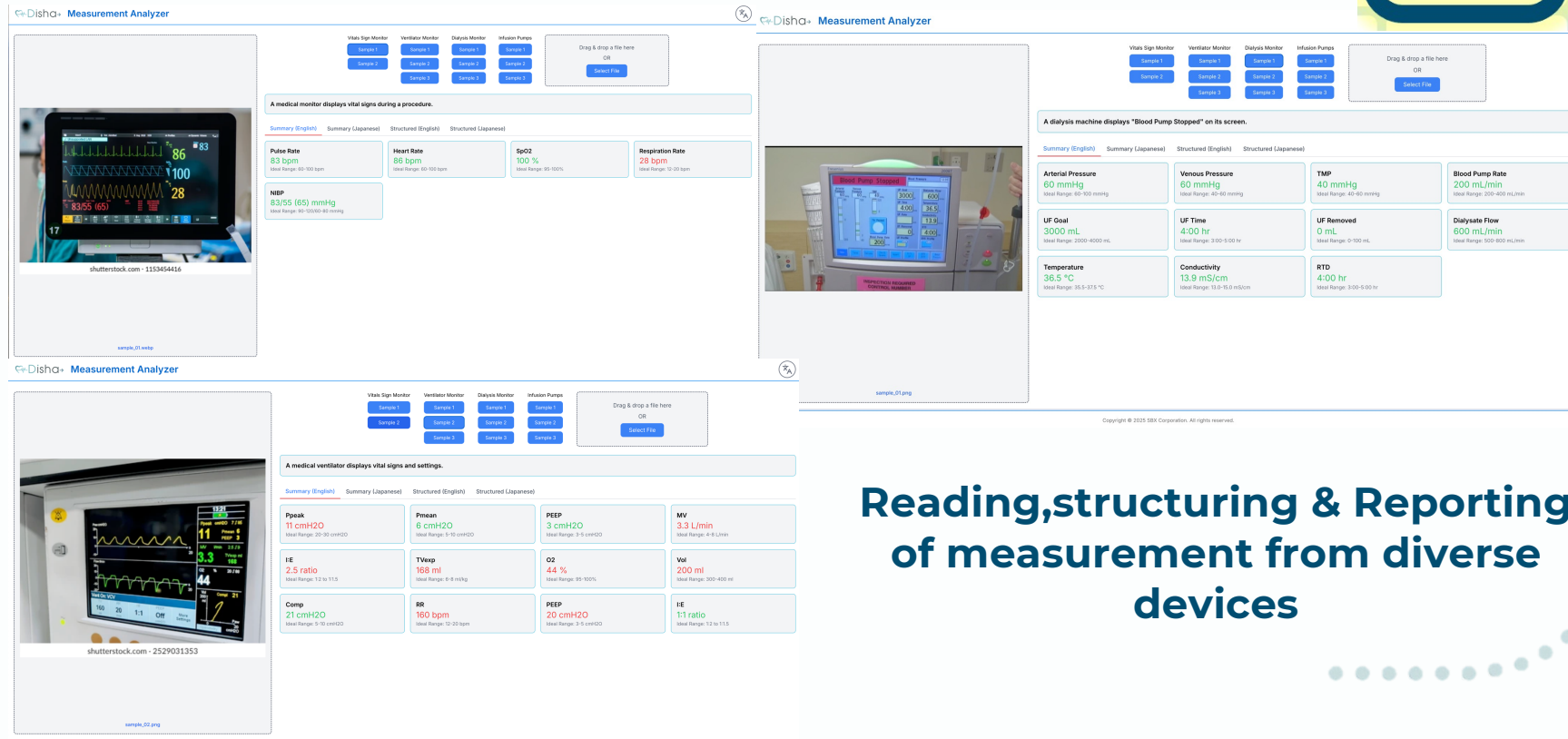
Diagrams:

- Title: THAD1 THAS1
  - Oxidosqualene
  - THAS
  - Thallanol
  - THAH
  - Thallandiol
  - THAD
  - Desaturated Thallandiol
- Title: GAL1 GAL7 GAL10
  - Galactose
  - GAL1
  - Galactose 1-phosphate
  - GAL7
  - UDP-Galactose
  - GAL10
  - UDP-Glucose
- Title: DAL1 DAL4 DAL2 DCG1 DAL7 DAL3
  - 5-Allantoin
  - DAL4
  - 5-Allantoin



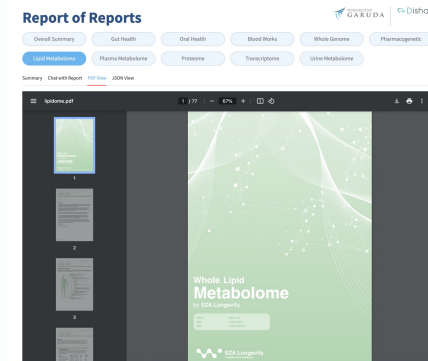
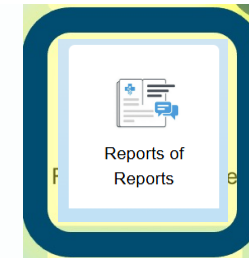
sample2.png

# ROI Agents showcase



Reading, structuring & Reporting  
of measurement from diverse  
devices

# ROI Agents showcase



**Report of Reports**

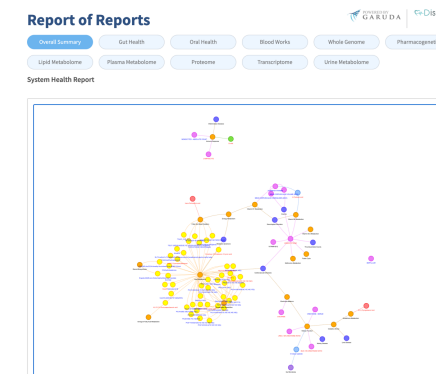
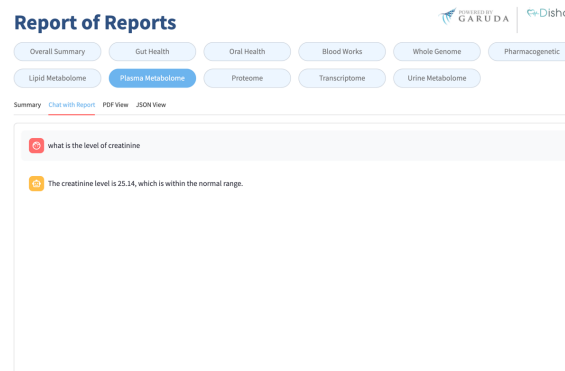
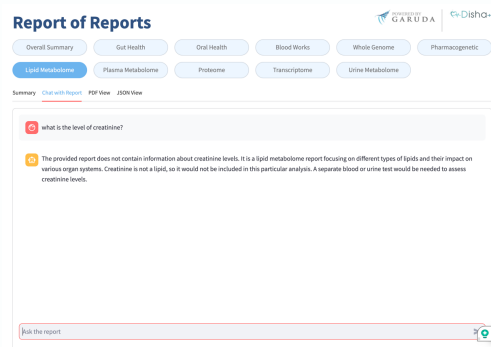
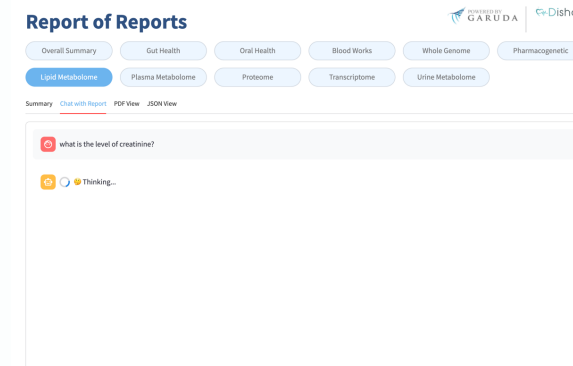
Overall Summary | Gut Health | Oral Health | Blood Works | Whole Genome | Pharmacogenetic

Lipid Metabolome | Plasma Metabolome | Proteome | Transcriptome | Urine Metabolome

Summary | Chat with Report | PDF View | JSON View

**Lipid Metabolome**

Type	Sub Type	Name	Formula	Range	Value	Analysis
Fatty Acids	Unsaturated Fatty Acids	6,8,12,15,18- eicosapentaenoic acid	C20:5(5Z,8Z,11Z,14Z,17Z)	17.75 - 23.83	24.28	High
Fatty Acids	Unsaturated Fatty Acids	8,10,12,14- docosahexaenoic acid	C22:6(4Z,7Z,10Z,13Z,16Z,19Z)	19.29 - 24.89	25.81	High
Polyketides	Polyketides	6,8,12,15- tetrahydroxypentadecanoic acid	C15:0(4Z,7Z,10Z,13Z)	17.28-21.38	22.81	High
Polyketides	Macrolides	14:0(4Z,7Z,10Z,13Z)	C14:0(4Z,7Z,10Z,13Z)	16.47-18.80	22.19	High
Glycerolipids	Triacylglycerols	1,3-bis(sn-3'-phosphatidyl)- sn-glycero-3-phosphate	C55H103O19	8.10-31.34	22.12	Low
Glycerolipids	Triacylglycerols	1,3-bis(sn-3'-phosphatidyl)- sn-glycero-3-phosphate	C45H85O13	10.74-28.75	16.71	Low
Glycerolipids	Triacylglycerols	1,3-bis(sn-3'-phosphatidyl)- sn-glycero-3-phosphate	C45H85O13	8.07-31.84	22.36	Low
Glycerolipids	Triacylglycerols	1,3-bis(sn-3'-phosphatidyl)- sn-glycero-3-phosphate	C45H85O13	22.22-29.34	21.67	Low
Glycerolipids	Monoglycerols	1,3-bis(sn-3'-phosphatidyl)- sn-glycero-3-phosphate	C29H53O9	8.47-26.81	21.68	Low







# Agency Stories

Discharge Summary Agent

Prescription Reader Agent

Speech To Text (MERaLiON) | ICDMapper

Report of Reports (RoR) | Text To Speech (TTS)

Measurement Analyzer | EHR (XLS)

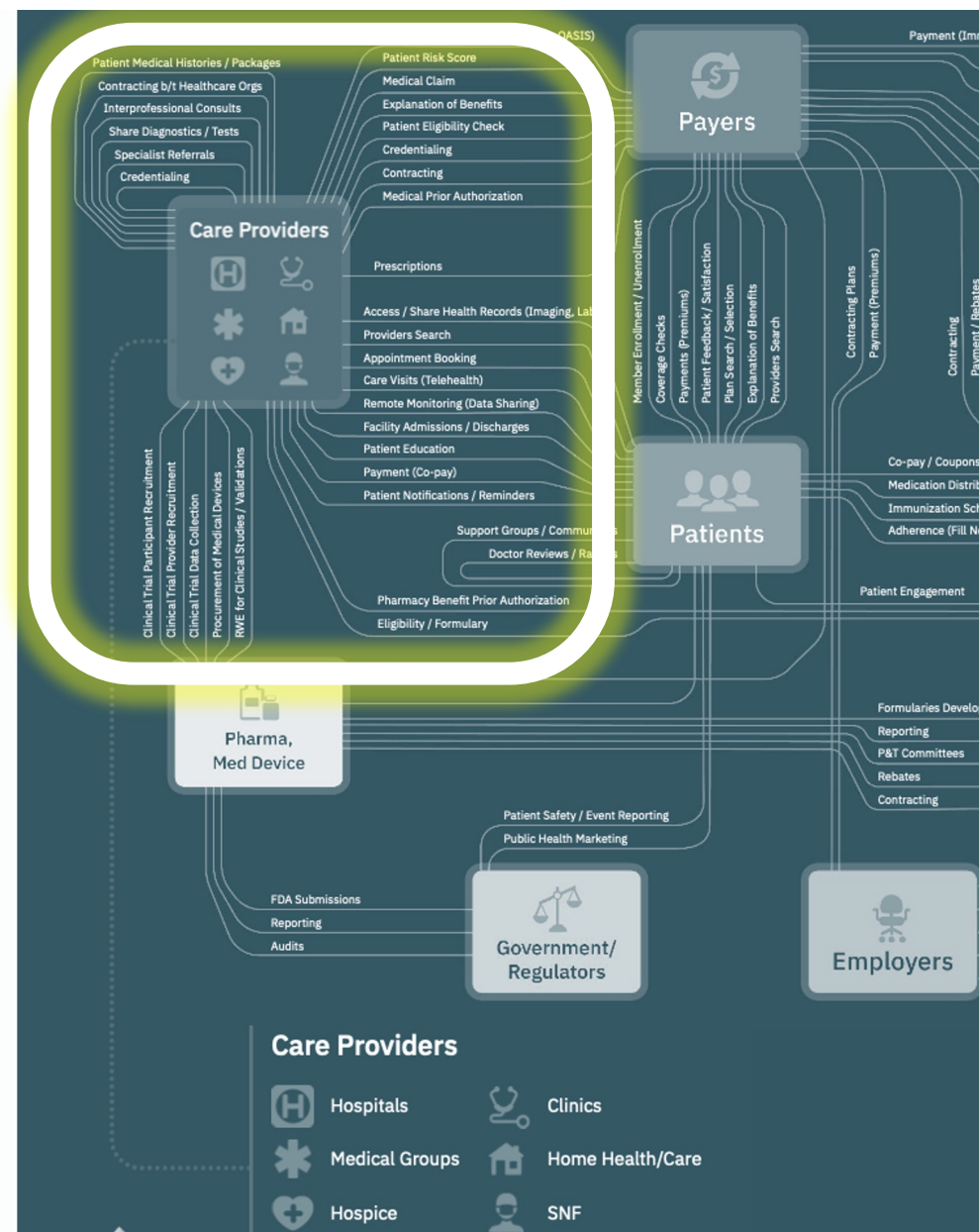
Measurement Analyzer | Text To Speech (TTS)

LangGraph workflows



# Agentic AI for Clinical Decision Intelligence: The TGWA Case Study

Setting the direction for the future of Clinical Decision Intelligence (CD<sub>x</sub>)

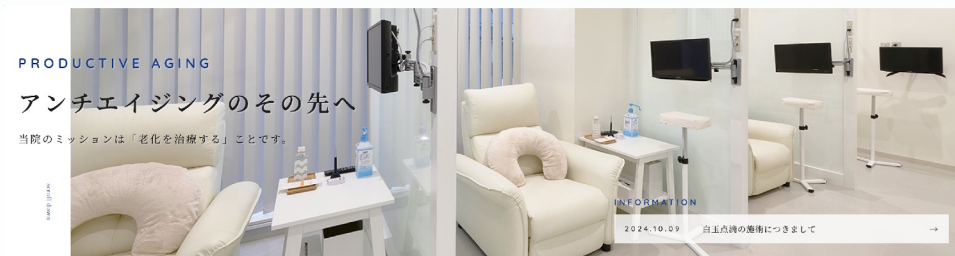


# Agents in Action – AI Clinic

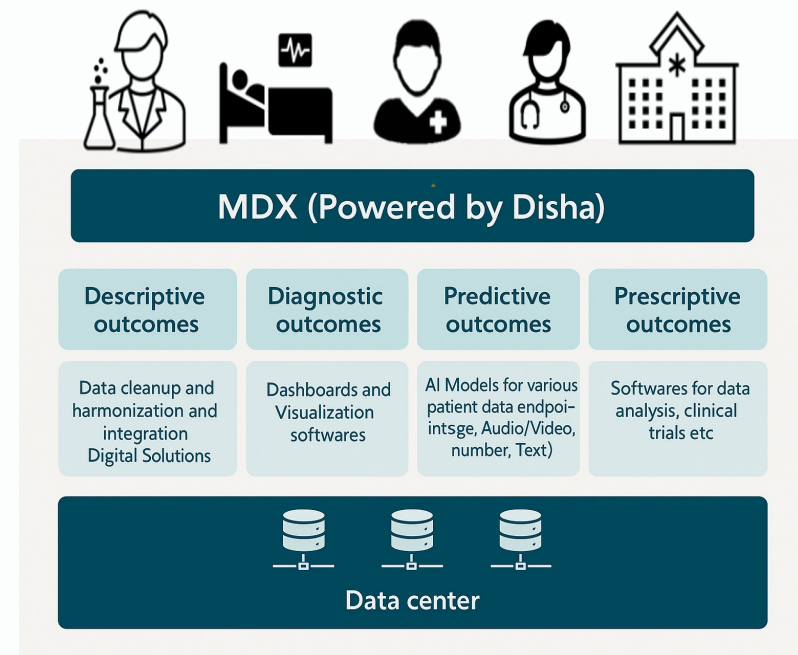
**Developed and deployed an AI Clinic with Disha agents to operationalize an Aging and Wellness Clinic in Tokyo, Japan**

**1,500** Patients

**4** Clinical Trials

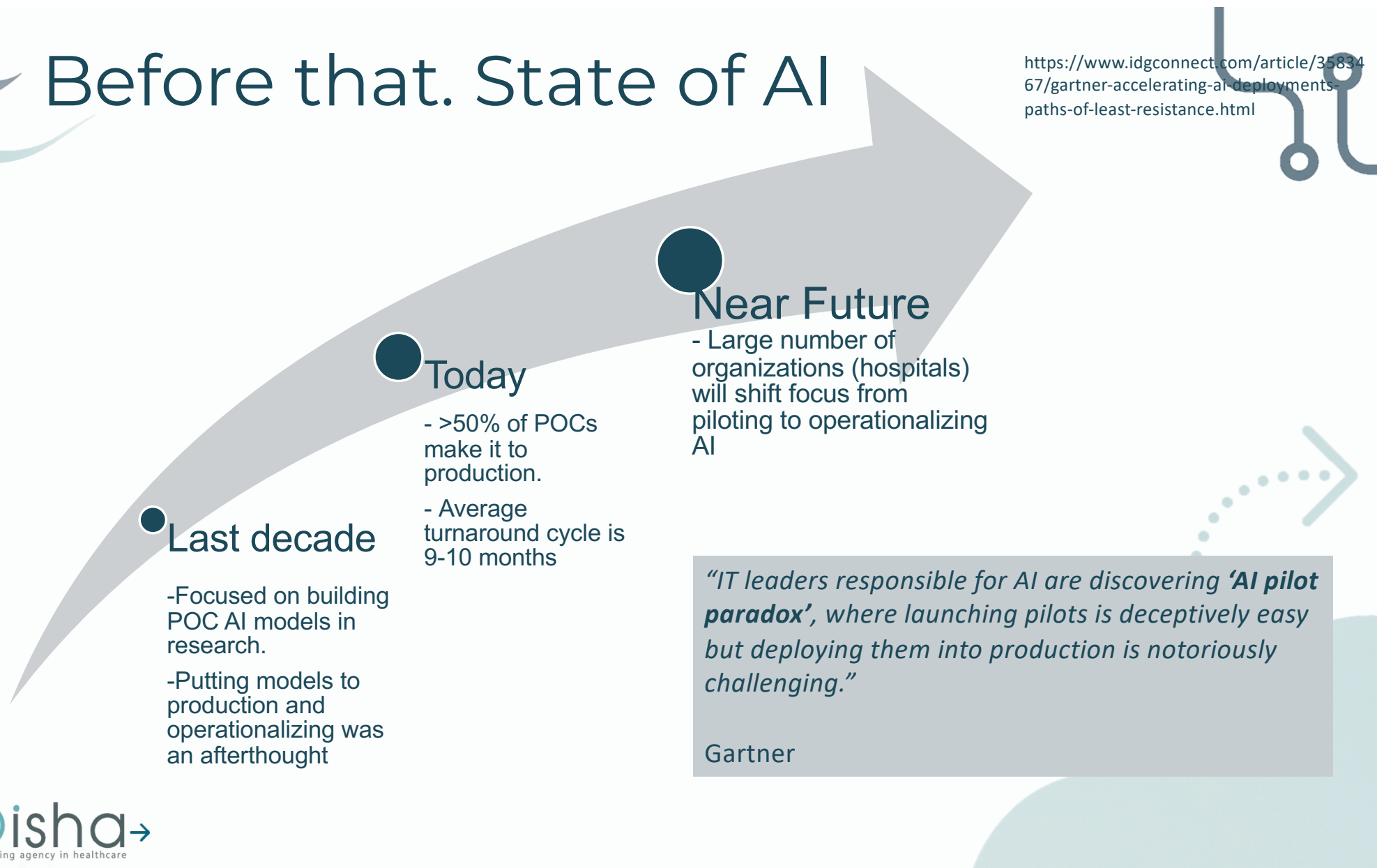


**Aging and Wellness focused Clinic in Tokyo**



# Before that. State of AI

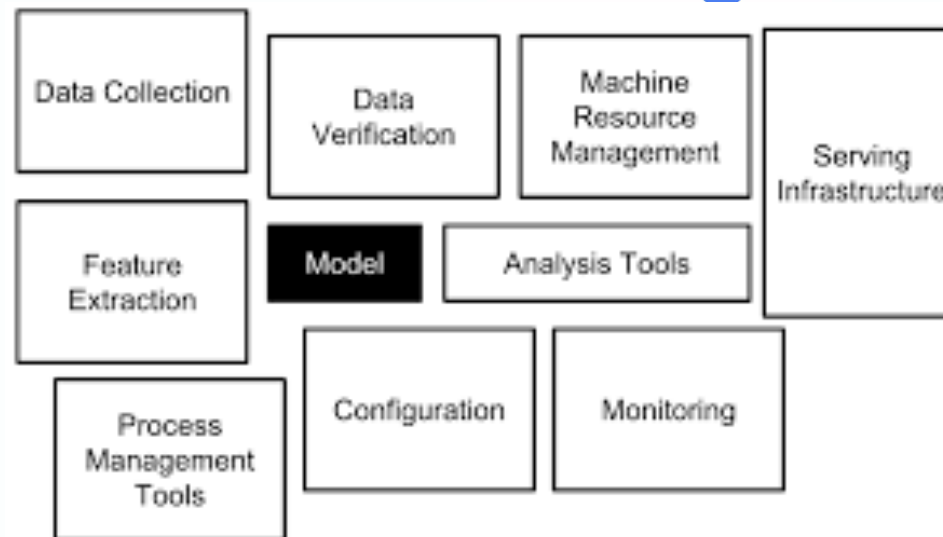
<https://www.idgconnect.com/article/3583467/gartner-accelerating-ai-deployments-paths-of-least-resistance.html>



*"IT leaders responsible for AI are discovering '**AI pilot paradox**', where launching pilots is deceptively easy but deploying them into production is notoriously challenging."*

Gartner

# How much of real world AI “systems” is the AI “model”?

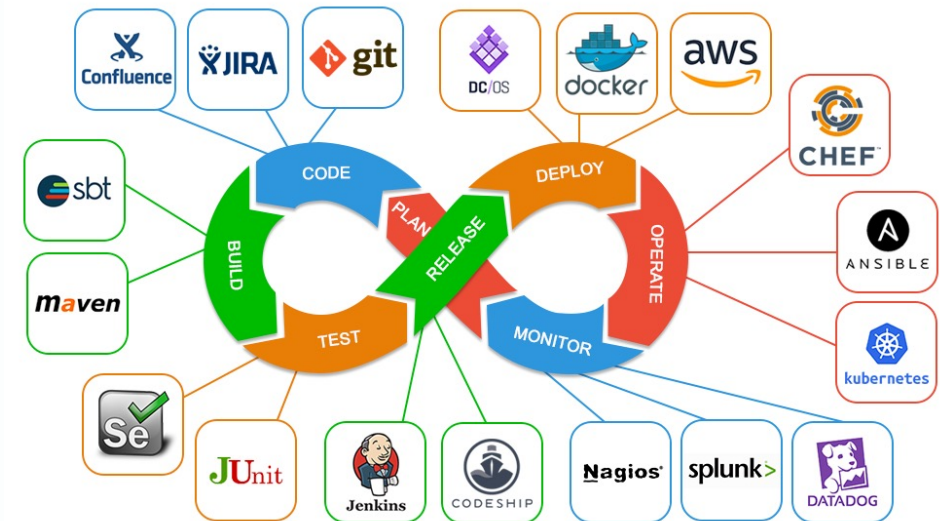


Phase 1  
(Research/experimental)  
Can we use AI to solve  
this problem?

Phase 2  
How do we implement these models into IT systems at scale?

# We are now at the point where software engineering was a (few) decades ago!

- Devops
  - Software Development + IT Operations is a subarea of software engineering
  - Scalable software products are built by embracing best Devops practices



<https://medium.com/swlh/how-to-become-an-devops-engineer-in-2020-80b8740d5a52>

## Embracing similar practices for AI ?

## Rise of MLOps

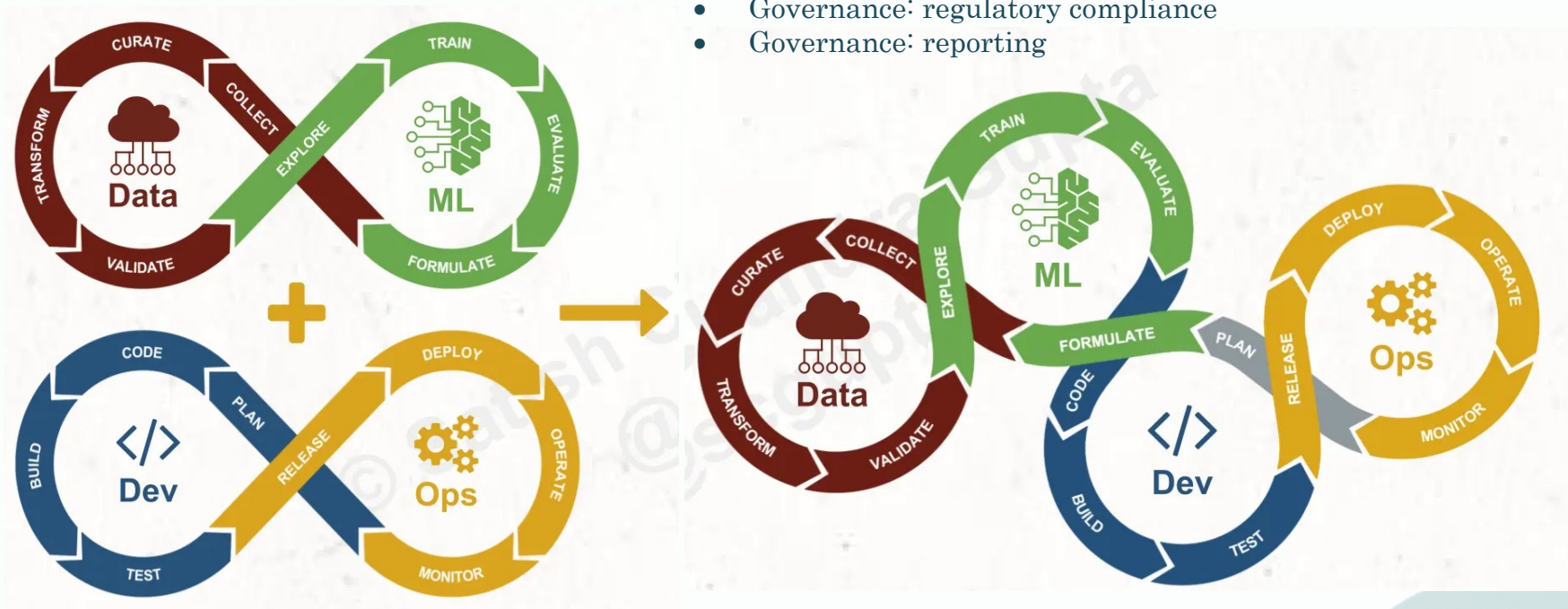


# MLOps

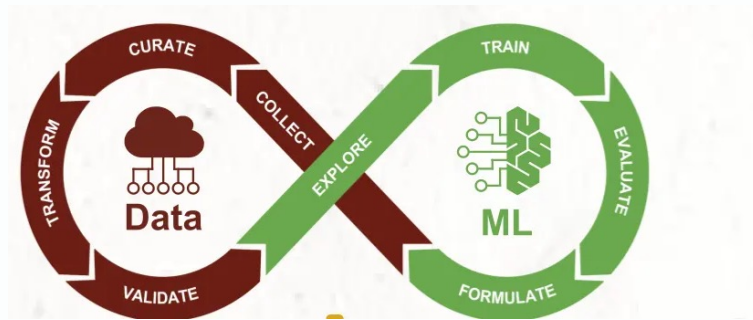
- Data + ML + Dev + Ops = MLOps
- Extending Devops for AI-ML

<https://www.ml4devs.com/articles/mlops-machine-learning-life-cycle/>

- Model versions tracking
- Monitor data versions and transformations
- Automation: data cleanup, transformation, re-training, optimization and scoring
- Monitor model performance (accuracy and robustness)
- Logging and notifications
- Governance: regulatory compliance
- Governance: reporting

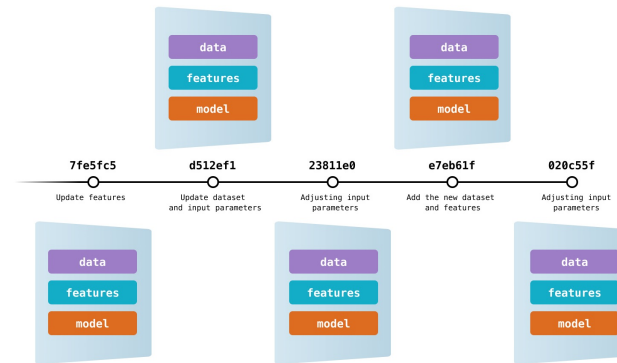


# Data and Model



- **Assumptions:** Data Scientist's decisions and assumptions must be explicit.
- **Randomness:** Considering that some machine learning experiments contain pseudo-randomness, this needs to be in some kind of control so it can be reproduced. For example, using "seed".
- **Data:** The same data of the experiment must be available.
- **Settings:** Repeat and reproduce experiments with the same settings from the original.
- **Implementation:** Especially with complex models, different implementations can have different results. This is important to keep in mind when debugging.
- **Environment:** It's crucial to have the same runtime configurations among all data scientists.

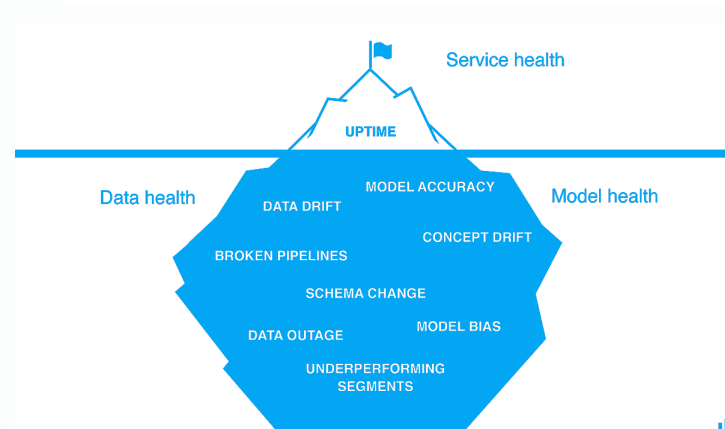
## Data and model Versioning



## Automation



ML Stages	Tasks
Data Engineering	Data acquiring, validation, and processing
Model Development	Model training, evaluation, and testing
Continuous Integration	Build and testing
Continuous Delivery	Deployment new implementation of a model as a service
Monitoring	Setting alerts based on pre-defined metrics



## Monitoring

<https://mlops-guide.github.io/>

# AI Clinic Case Study – Overview & Impact

- **Before DISHA, The clinic faced challenges common in healthcare operations**

- Fragmented data across EMRs and equipment
- Manual data entry and Limited insights

- **MDx Platform**

- Centralized data portal that consolidated diverse sources
- Automated data processing
- Extensible infrastructure to expand new AI models and applications

- **Key Features of the Platform**

- In-house compute for seamless processing
- Full Integration of all equipment & data sources
- DISHA agents to clean, harmonize, and analyze data
- MLOps & AIOps in action for continuous optimization
- AI-ready for future applications

**Intelligent Agents**

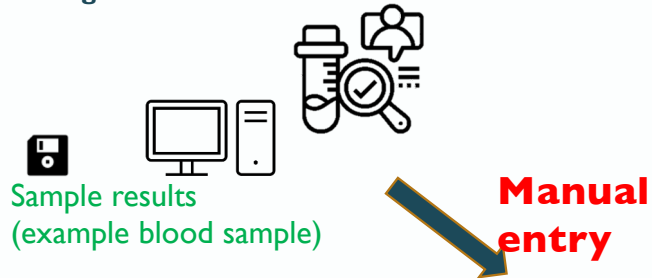
**Operationalization(O16n\*)**

**Data & Security**

**Compute & Infrastructure**

# Operations then

## - Diagnostic laboratories



Medical support staff

- Facilitate smooth operation
- Provide hassle free service to patients

Outpatients

Vitals etc

**Manual entry**

EMR  
(M3 digikar)

Doctors



- Medical equipment

Clinic



**No Data transfer**

Equipments IT system



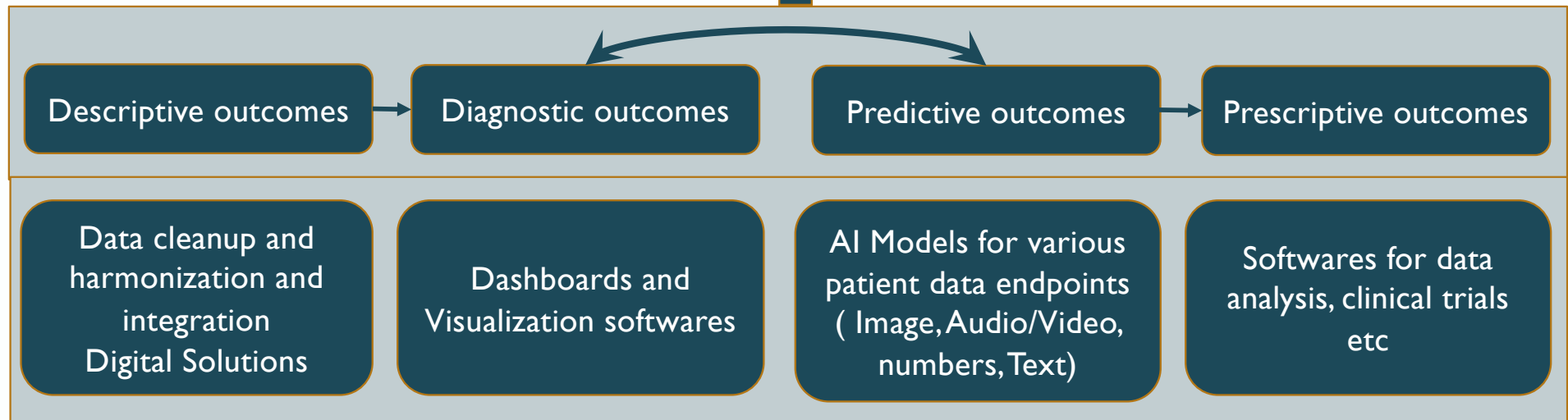
Research Laboratory

**Clinical research measurements**

# MDx Platform

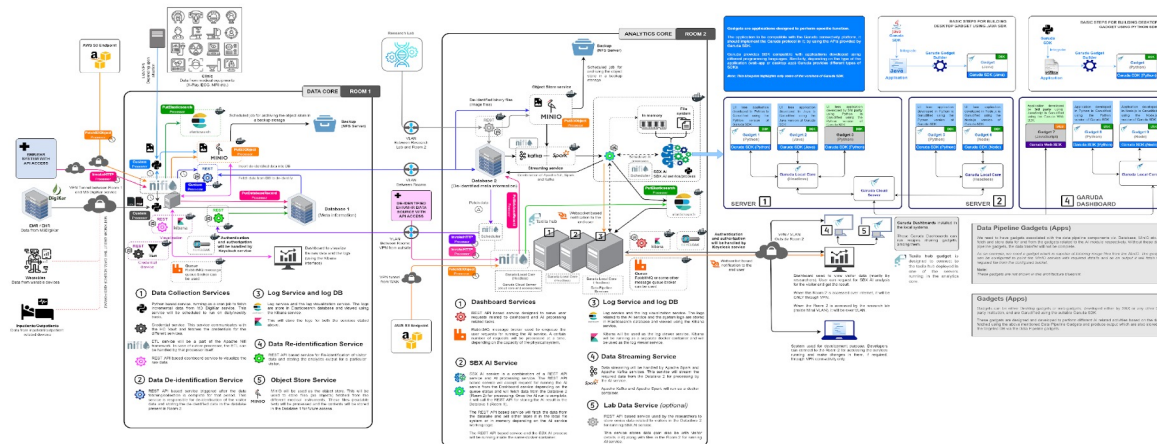


Improved decision making  
and outcomes

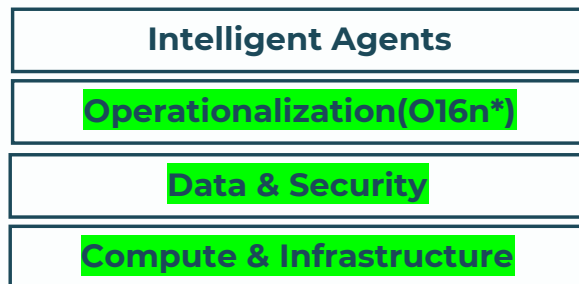




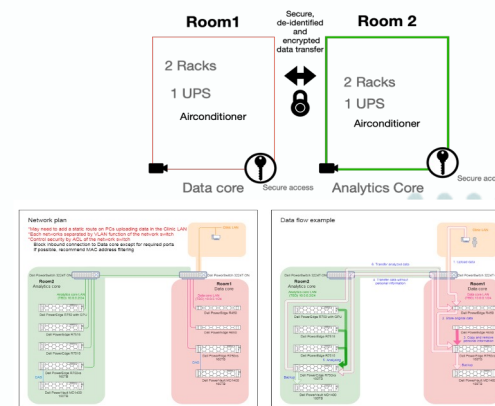
# Robust, Secure Infrastructure for Scalable AI



Operationalization (O16n): Technology stack for DevOps & MLOps



✓ Secure, redundant infrastructure ensures compliance, uptime, and data protection

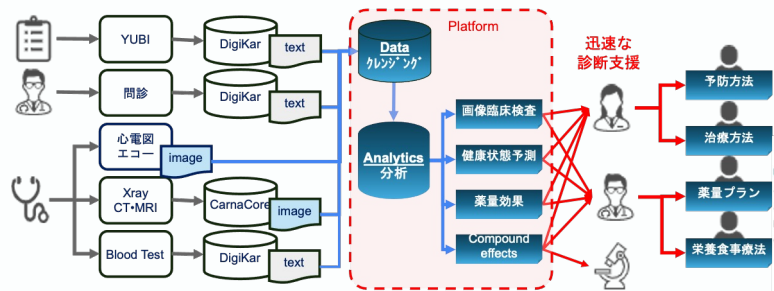


Compute and Infrastructure built ground up

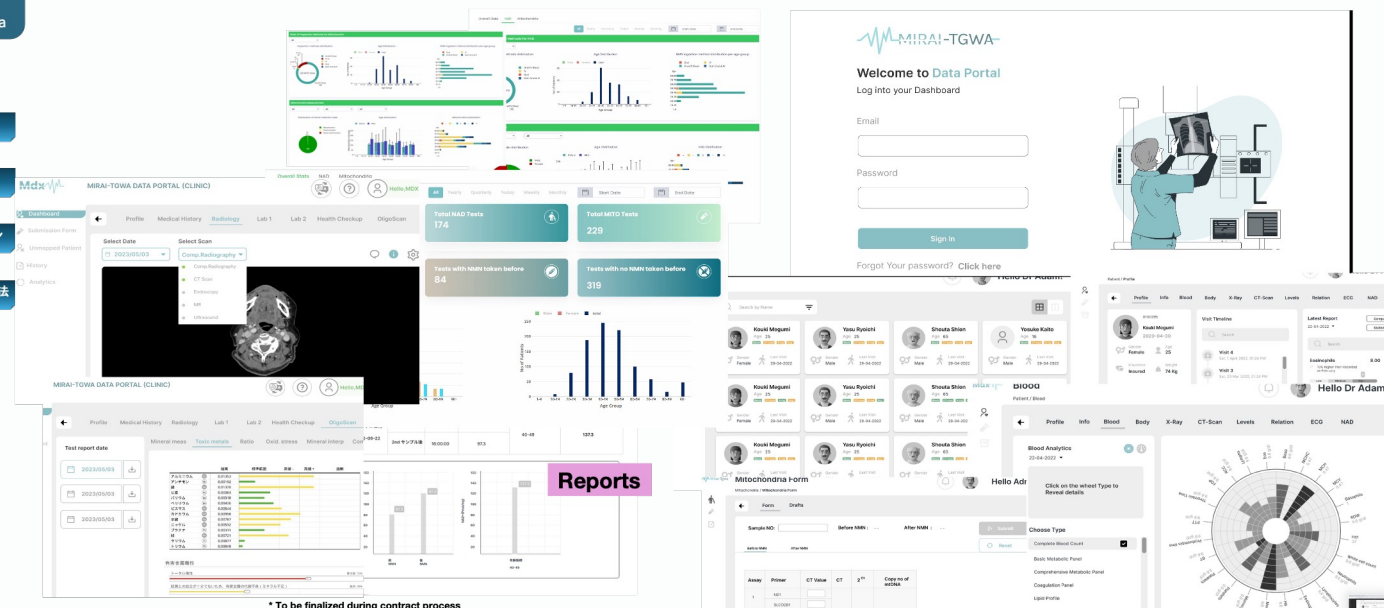


# AI Clinic – Data Flow & Analytics in Action

- Visualize predictions, verifications, and effects from clinic data, support doctors in making quick decisions, and provide data-based advice to patients
- Collect and organize data, and enable access and viewing of various data



## Data Portal All Clinic and operational data in one place ready for AI



Intelligent Agents

Operationalization(O16n\*)

Data & Security

Compute & Infrastructure

- ✓ Unified data enables real-time analytics, evidence-based advice, and automated reporting-all in one portal



# Disha Pilots



## Lighthouse Projects

Reaching out to collaborate with hospitals & clinics to  
develop, deploy and test  
Disha Agents

### Current Disha Partners

Hopeful Aging, USA  
<https://hopefulaging.com/>

SRM University and Corporate Hospitals, India

Redkar Hospital, India

The University of Osaka, Faculty of medicine, Japan

Graduate School of Medicine, The University of Tokyo

Keio University Hospital, Japan



**THANK YOU**



**INFOCOMM  
MEDIA  
DEVELOPMENT  
AUTHORITY**