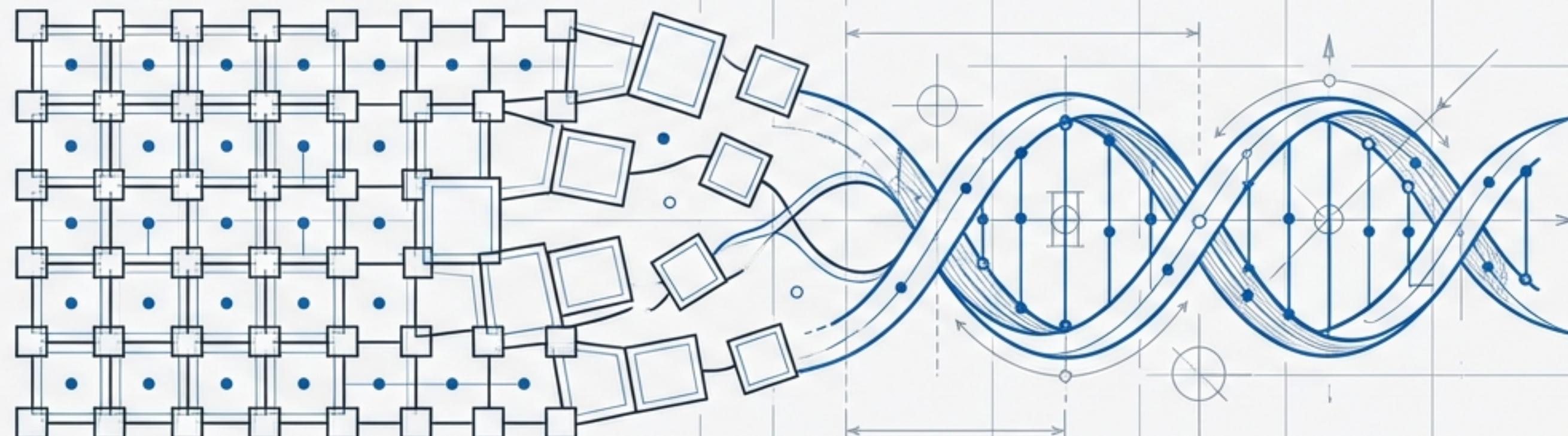


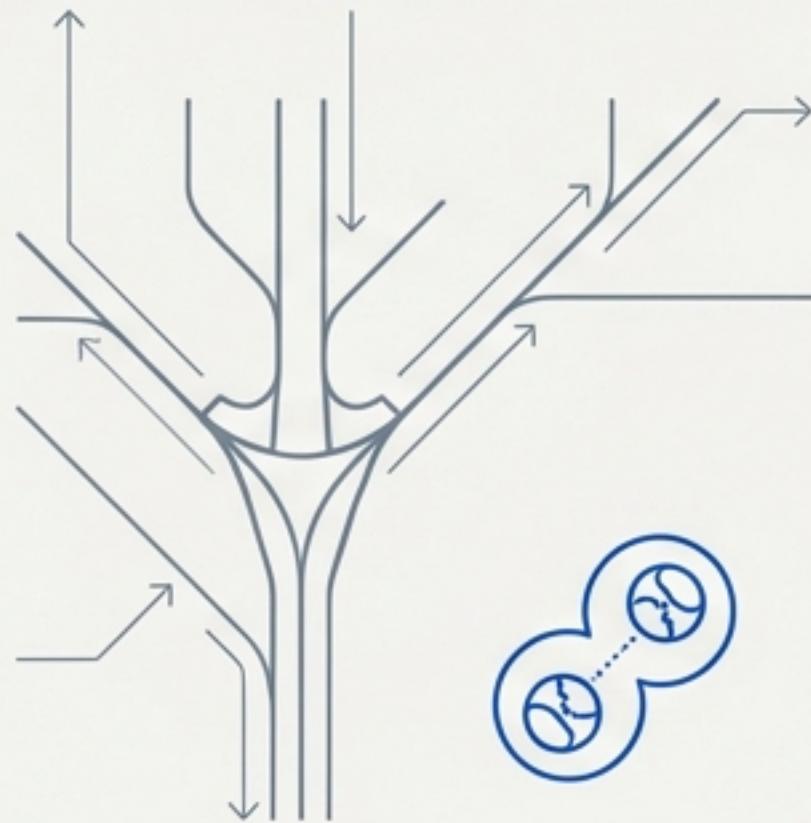
# Redefining the Data Scientist: Evolution to Version 6

## Adapting Skill Sets for the Era of Generative AI



Based on findings from the Data Scientist Society (Japan) 12th Symposium (2025.11.25).

# Executive Summary: The 2025 Standard



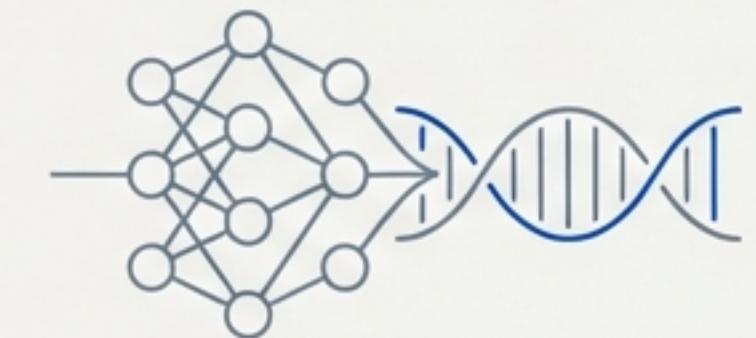
## Status: Released

The Data Scientist Society has officially released Skill Set Ver. 6.



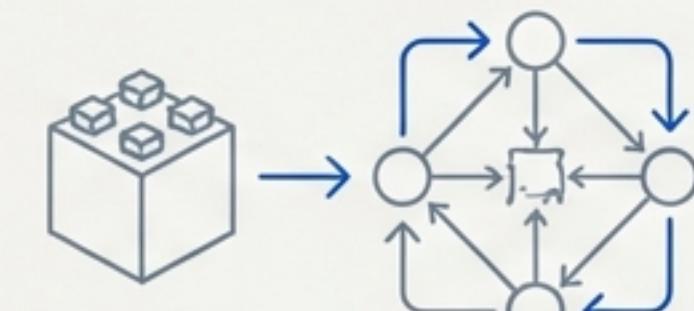
## Primary Driver: Generative AI

The explosive integration of Foundation Models has necessitated a structural overhaul of the profession.



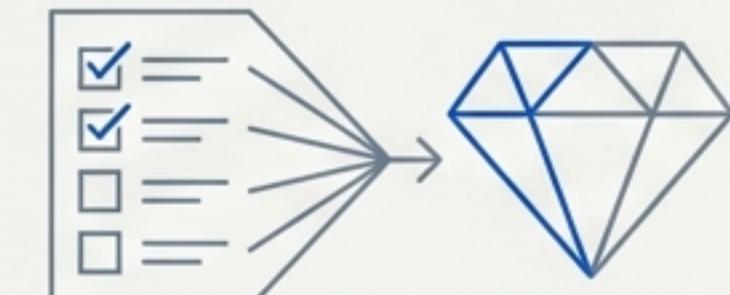
## The Shift: Architecture over Construction

The role has evolved from building models from scratch to architecting, fine-tuning, and evaluating AI systems.



## The Outcome: Value Creation

A streamlined checklist of approx. 110 items prioritizing unstructured data, multimodal processing, and AI ethics.

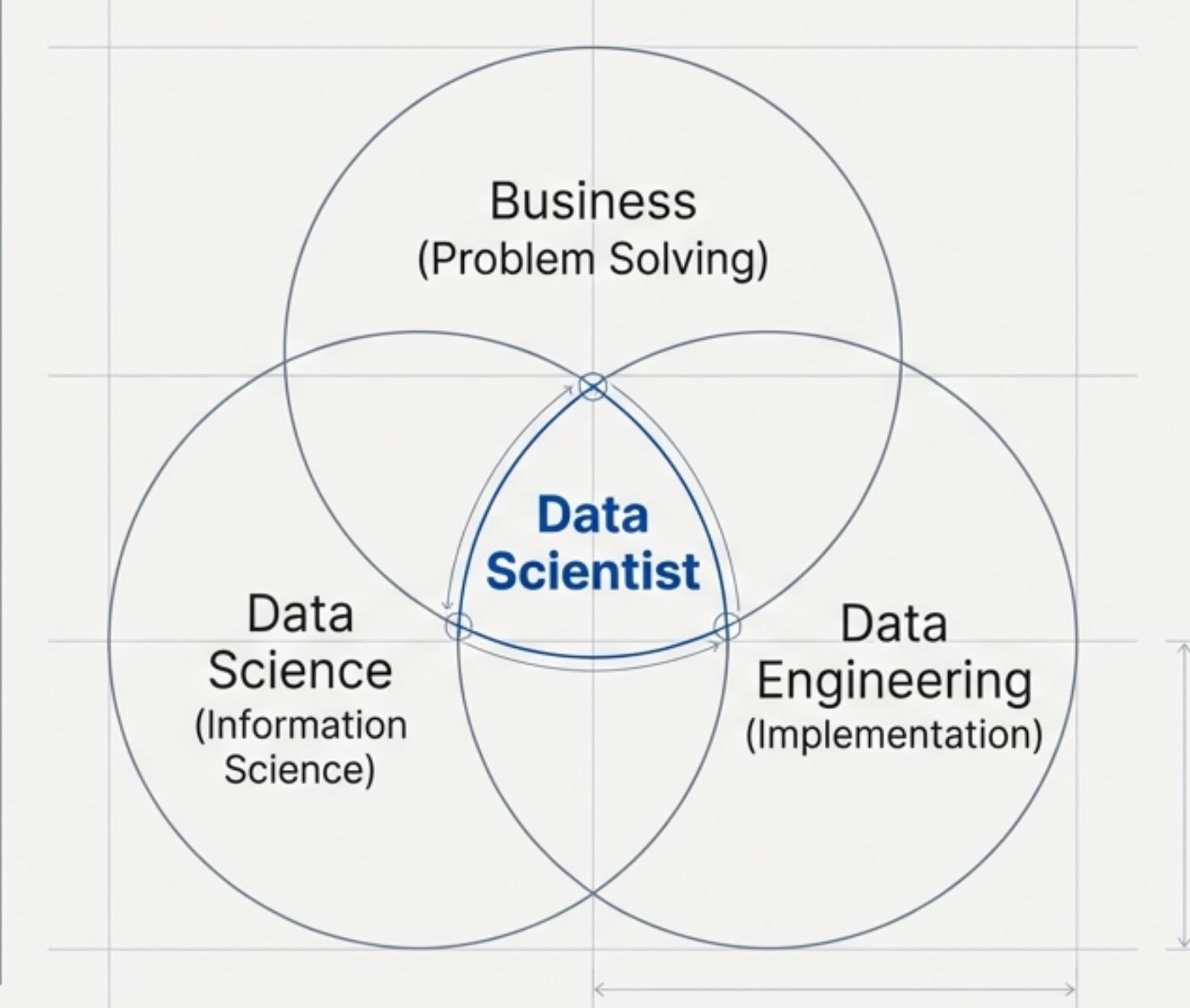


# 2014: Defining the Undefined

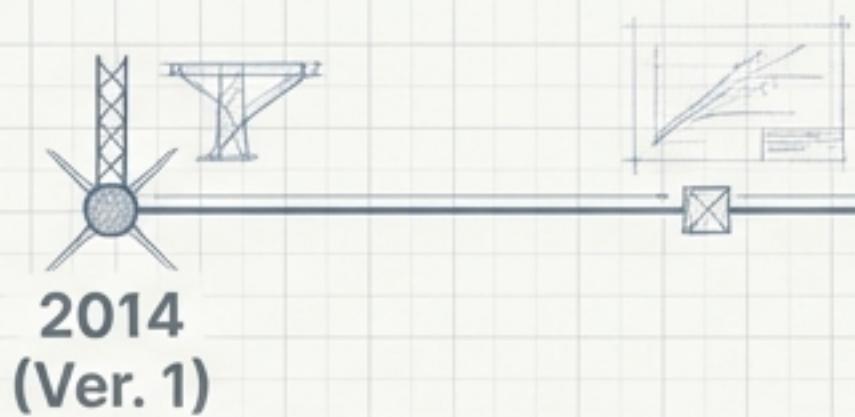
## The Original Challenge

In 2014, “Data Scientist” was a buzzword without a definition. A critical gap existed between employer expectations and talent capabilities.

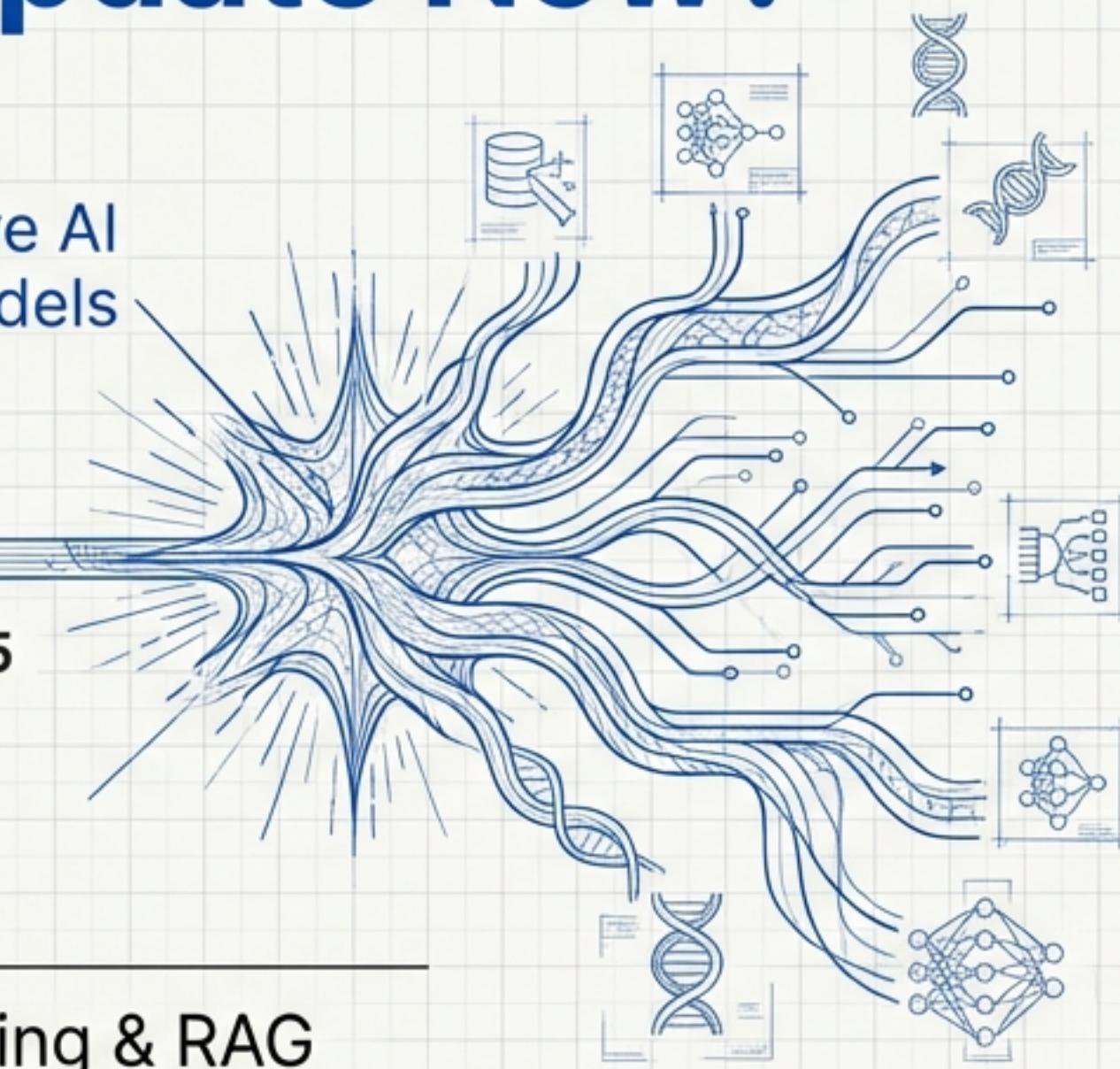
The mission was to provide a training map for young talent and facilitate dialogue with society.



# The 2025 Catalyst: Why Update Now?



**Key Driver:** Generative AI & Foundation Models



## FROM (Pre-GenAI)

- Mathematical modeling
- Structured data analysis (Tables/SQL)

## TO (Ver. 6)

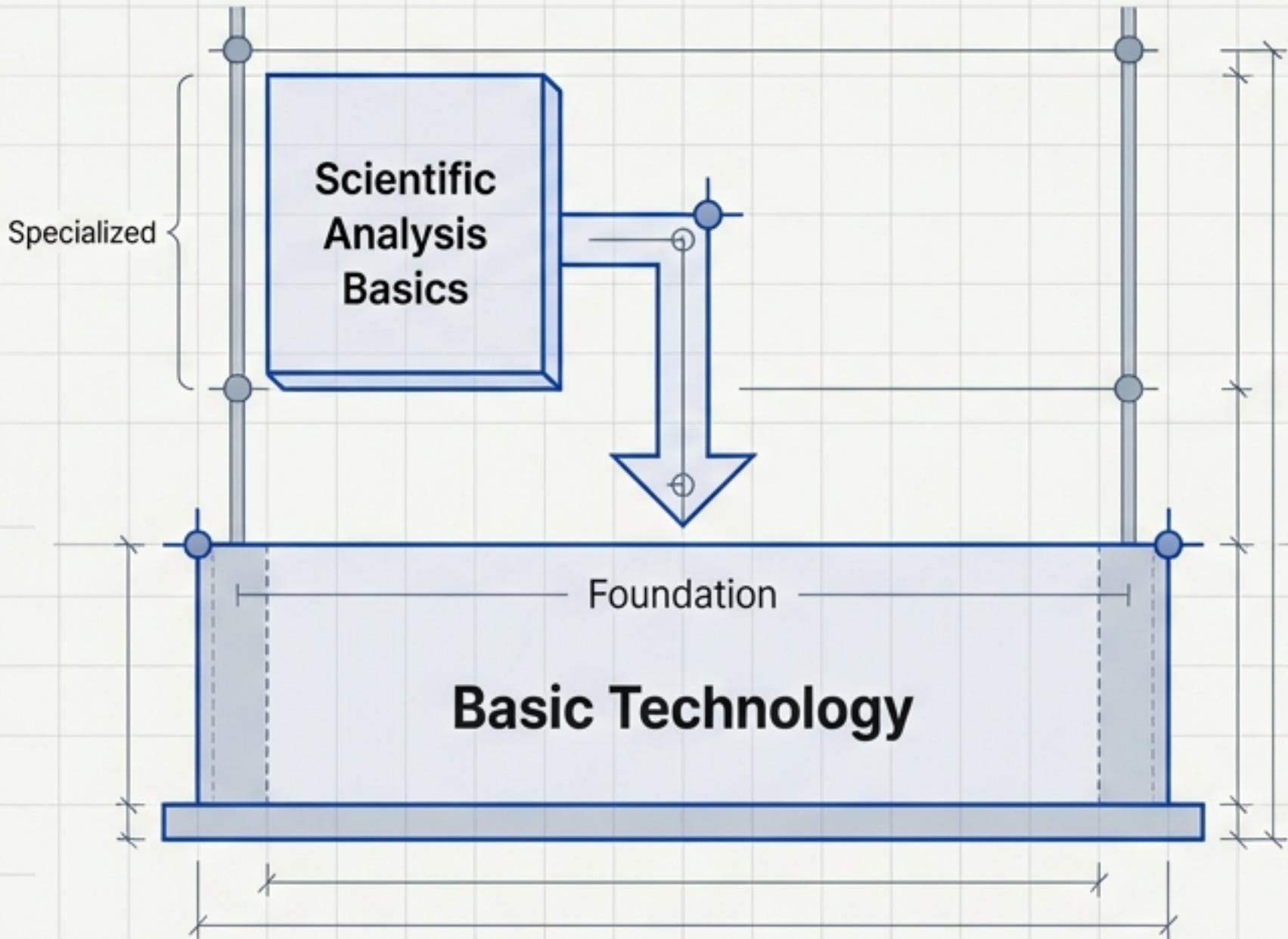
- Prompt Engineering & RAG
- Foundation Model Adaptation
- **Unstructured Data (Text, Image, Audio, Video)**

Handling Text, Image, Audio, and Video as standard inputs.

# Structural Shift: Raising the Floor

## Re-categorization

'Scientific Analysis Basics' is no longer a niche skill. It has been redefined as "Basic Technology".



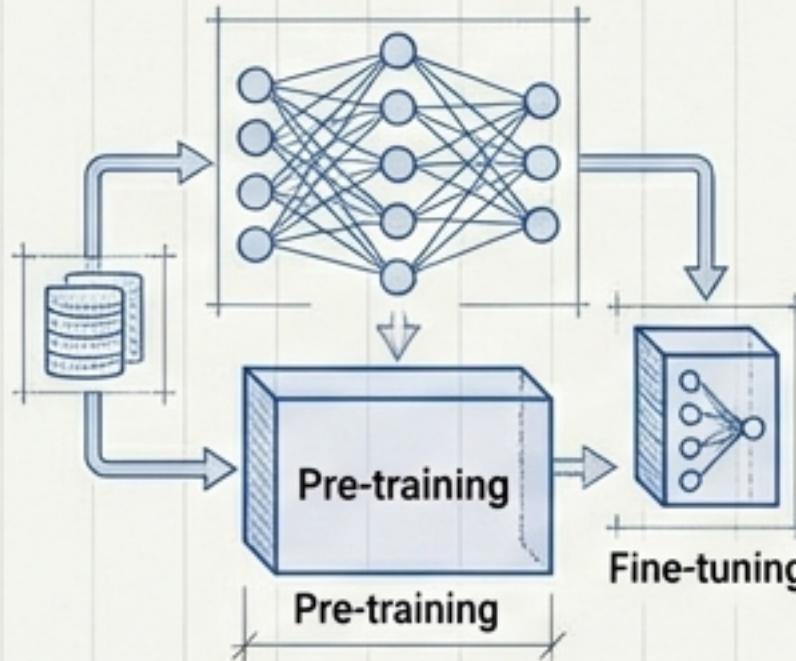
## The Implication

Statistical literacy is now the non-negotiable floor for all practitioners. The foundation has been raised.

## New Categories Added

- AI Utilization Technology
- Unstructured Data Technology

# Data Science Pillar: The GenAI Mechanics



## Structural Understanding

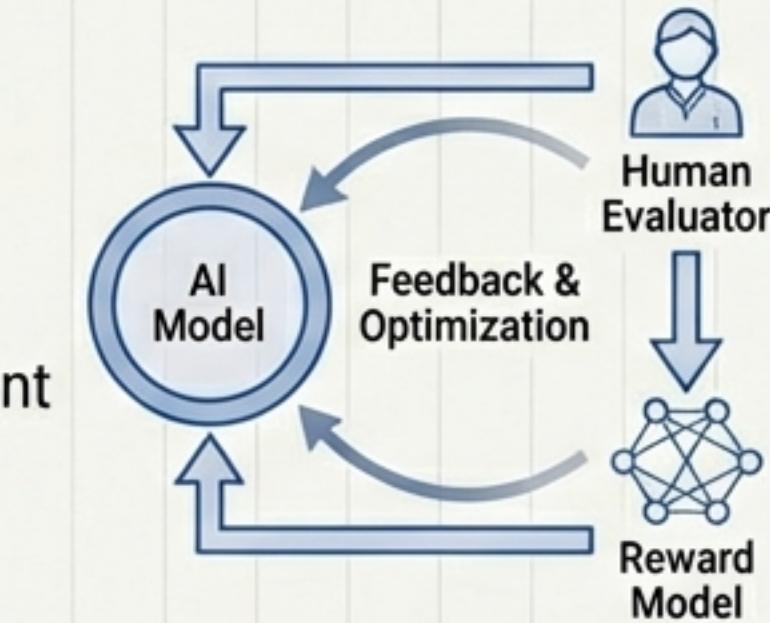
### LLM Structure & Training Procedures

Understanding Architecture, Pre-training, and Fine-tuning.

## Advanced Learning Methods

### Reinforcement Learning

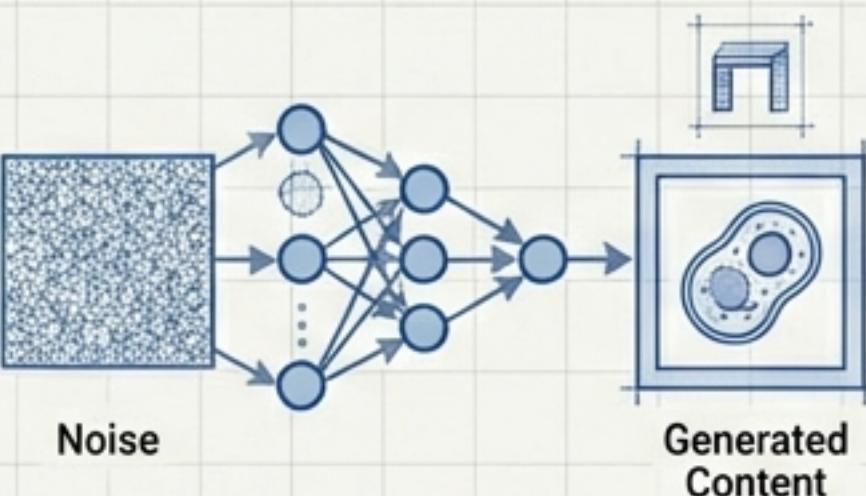
Focus on RLHF (Reinforcement Learning from Human Feedback) and RLAIF.



## Generation Mechanisms

### Diffusion Models

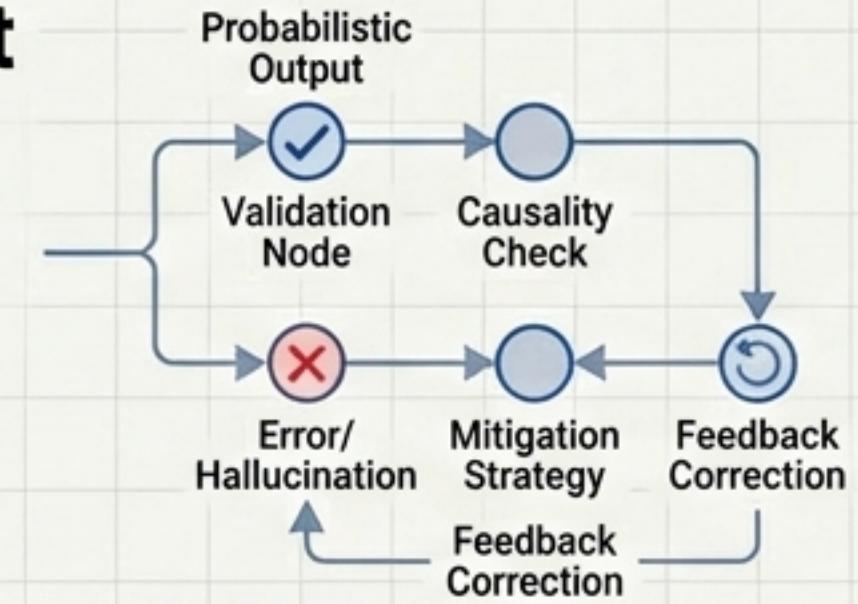
Mechanisms behind image and video generation.



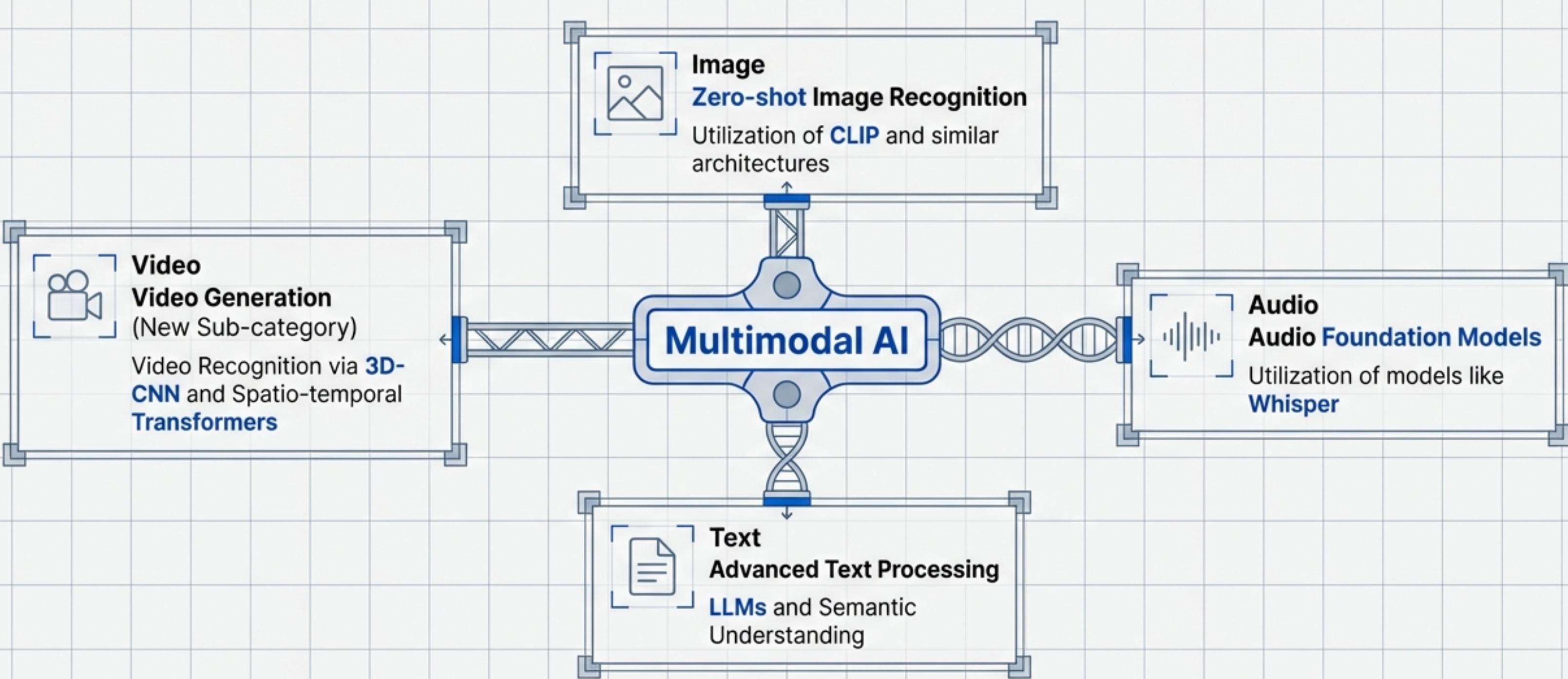
## Risk Management

### Handling Hallucinations & Causality

Managing the probabilistic nature of AI outputs.

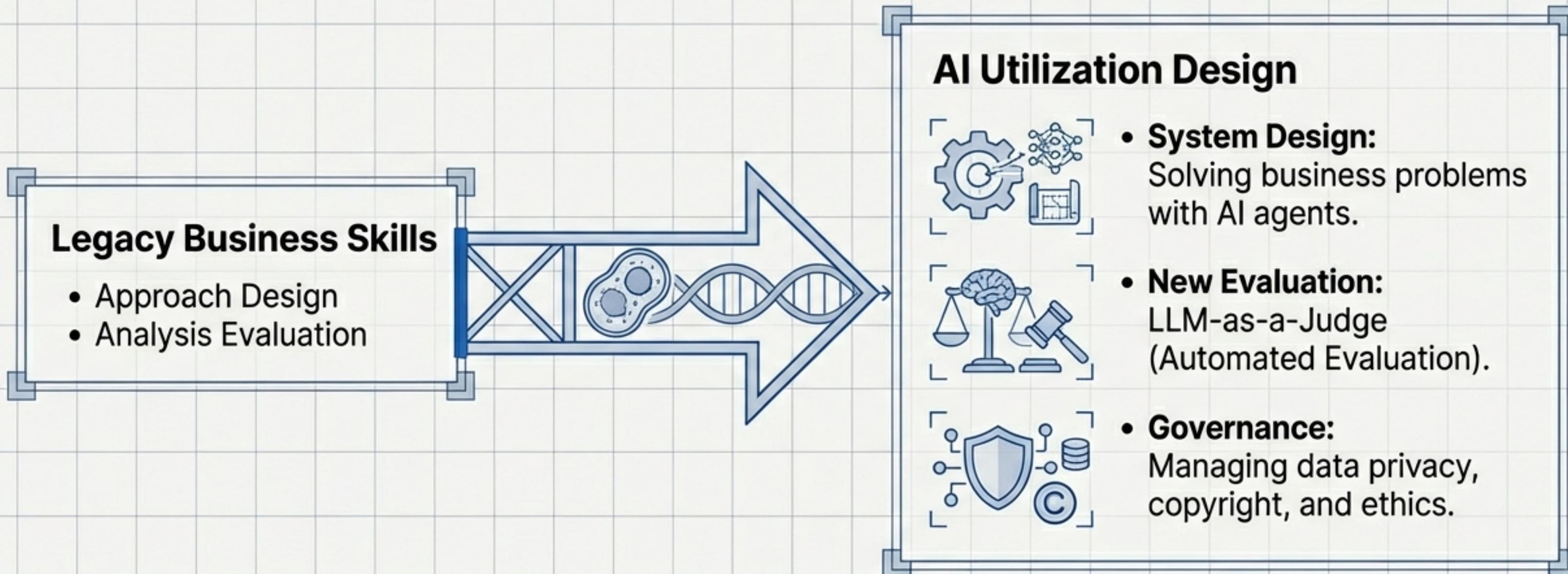


# Multimodal Expansion: Beyond Text and Numbers



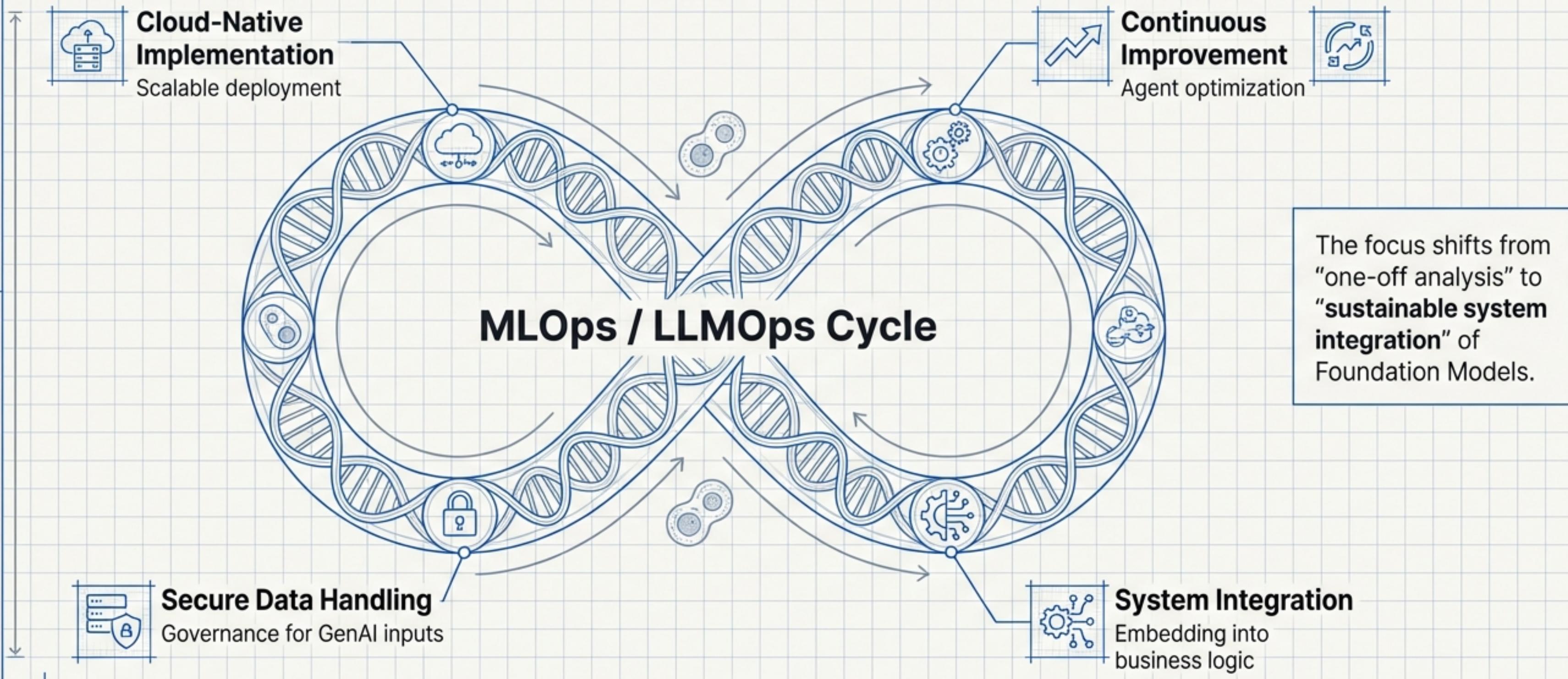
The Data Scientist must now function as a Multimedia Architect, processing sensory data streams.

# Business Pillar: From Analysis to System Design

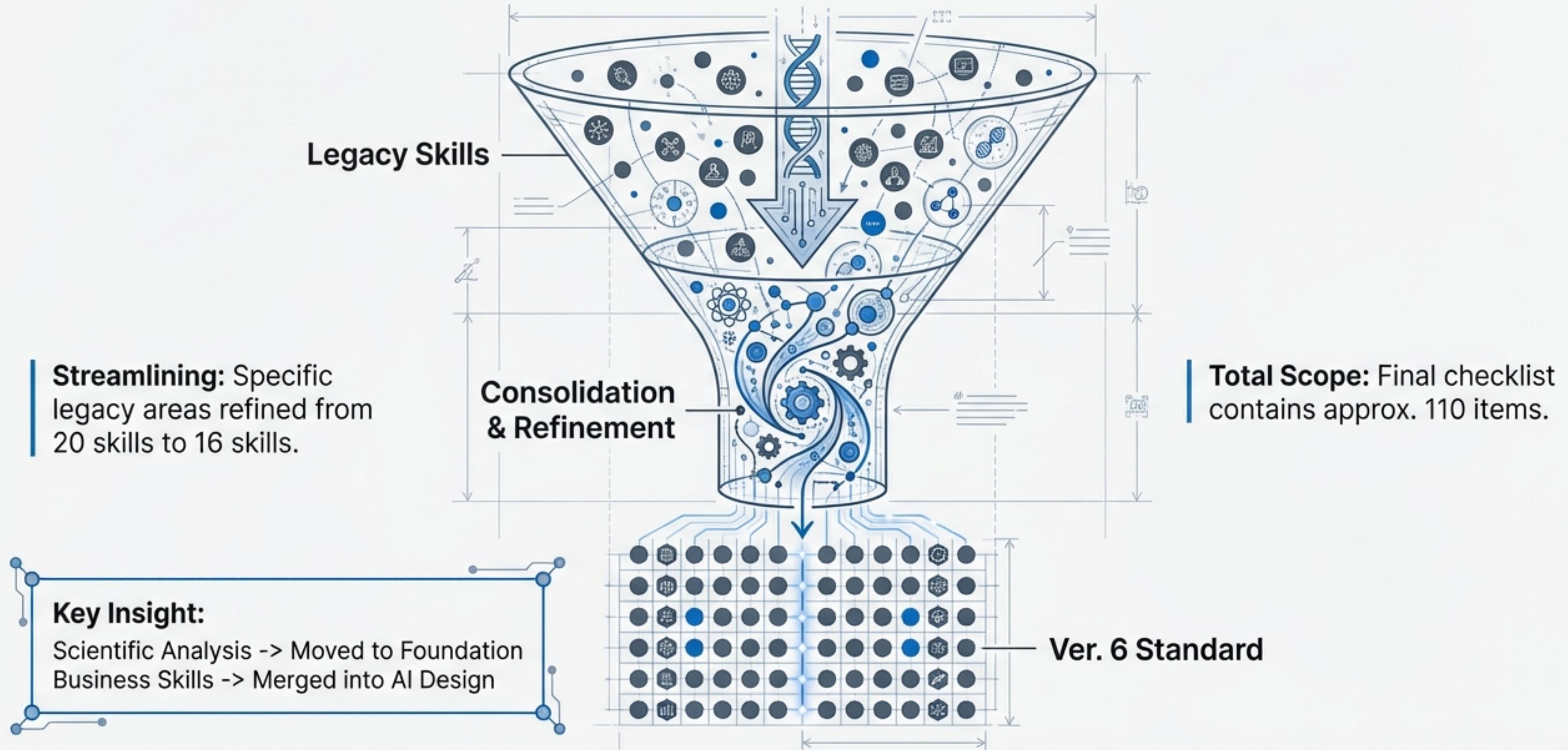


Old “Business Power” skills have been consolidated and repurposed for designing AI systems.

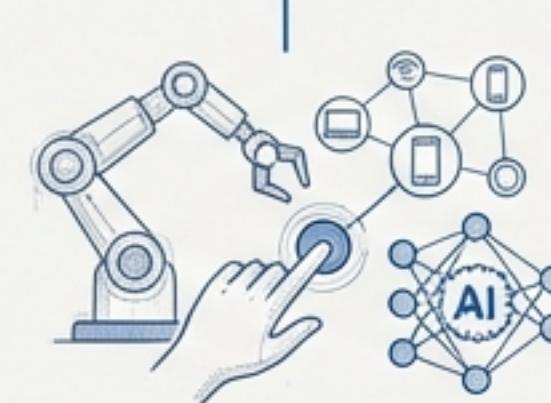
# Data Engineering: Systematizing the Agent



# By The Numbers: Refinement to Ver. 6



# The New Identity: From Analyzer to Value Creator



## Integration

Merging Robotics,  
IoT, and AI Agents.

**Role Expansion:**  
**“Analyzer” → “Value Creator”**



## Leadership

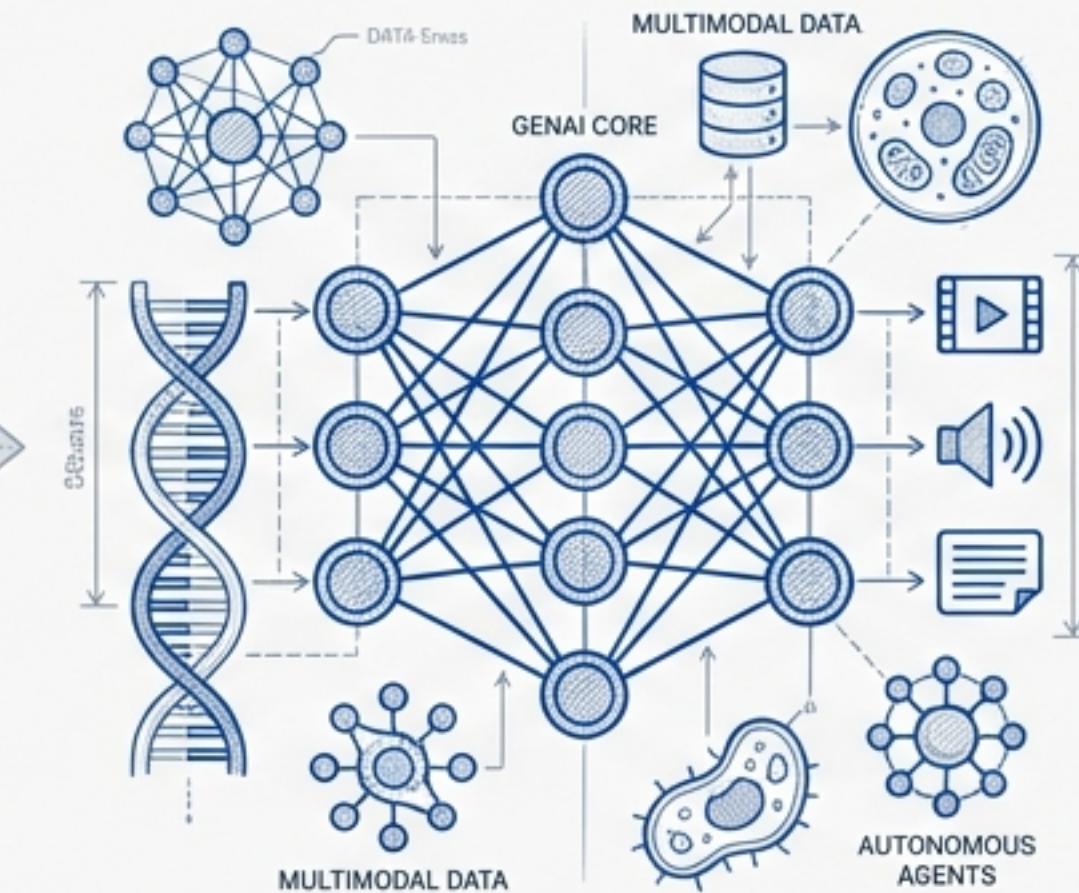
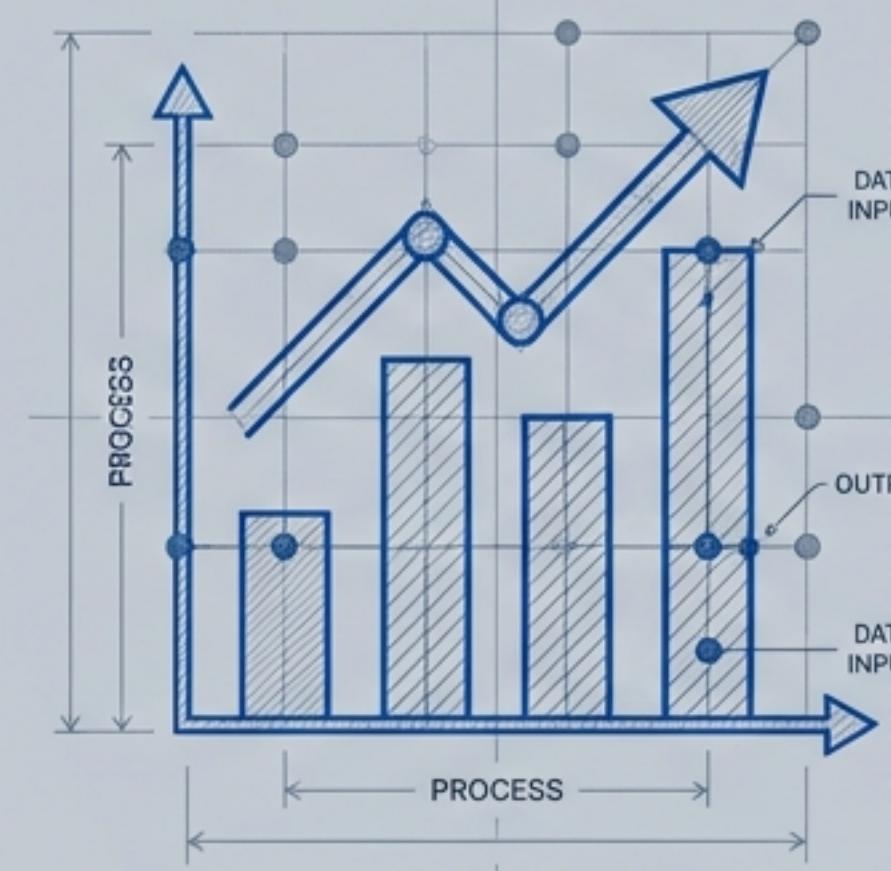
Change Leadership to  
drive AI adoption.



## Dialogue

Bridging the gap between AI  
capabilities and societal needs.

# Evolutionary Leap: 2014 vs. 2025



## 2014: The Analyzer

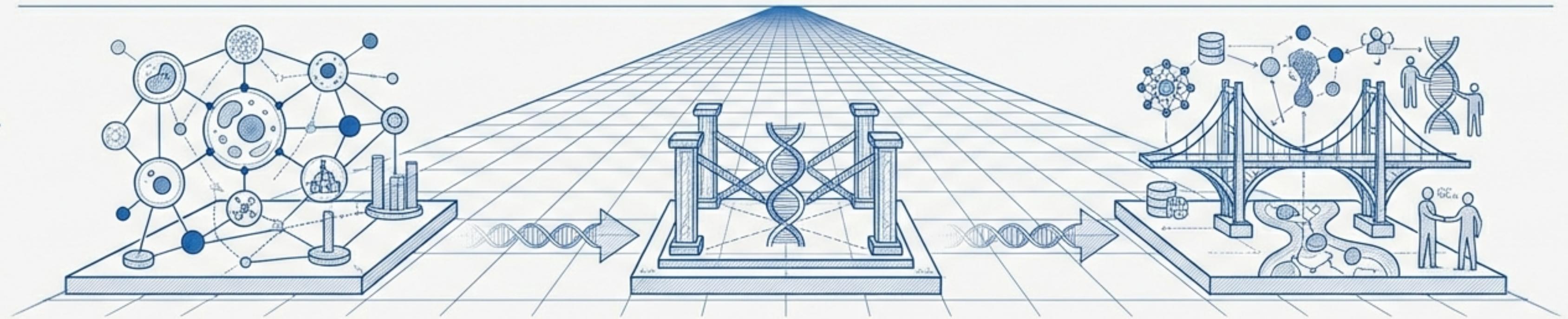
**Focus:** Statistical Analysis & Predictive Modeling  
**Data:** Structured (Tables/SQL)  
**Output:** Reports & Insights

## 2025: The Value Creator

**Focus:** GenAI Integration & System Design  
**Data:** Unstructured (Video/Audio/Text)  
**Output:** Content Generation & Autonomous Agents

# The Future Roadmap

## Dialogue with Society



### Open Ecosystem

Maintaining a collaborative environment for professionals.

### Ver. 6 Platform

The baseline for the next wave of AI-native talent.

### Human Value Creation

Bridging Generative AI capabilities with human needs.

**The era of merely ‘analyzing’ data is over.**

**The era of ‘generating value’ with AI has begun.**

### **Next Steps**

1. Review the **Ver. 6 Checklist**.
2. Assess team capabilities against the new Foundation baseline.
3. Update training curricula for Generative AI.

# Skill Definition Committee

**Kazuto Ataka**

(Committee Chair, Keio University/LINE Yahoo)

**Kazuhiro Moriya**

(Data Analytics Lab)

**Satoshi Saeki**

(Vice Chair, Data Scientist Society)

**Satoshi Sugiyama**

(Atrae / Alcia Solid Project)

**Nobumitsu Takahashi**

(Digital Growth Academia)

**Takahiro Tanaka**

(Hitachi Academy)

**Yukiko Kan**

(Rejoui)



DataScientist Society

Data Scientist Society - 12th Symposium (2025)