

# Technology-Driven, Robust Customs Administration

How does the Korean Customs Service leverage data and relevant technologies?



**Korea Customs Service**  
**Miryang Kim, 25.11.25**



## IT Organizations

ICT and Data Policy Bureau

ICT Planning Division

AI Innovation Division

ICT and Data Management Division

R&D and Equipment Division

UNI-PASS Operation Division

Customs Border Targeting Center

Designing ICT Direction, Conducting ISP, Controlling Data Quality

Operating Big Data Platform(+ML, DL), Fostering ICT Experts, Adopting Generative AI

Operating Customs Data Warehouse, Data Security

Researching AI Embedded in Devices

Operating UNI-PASS System(AI-driven Risk Management), Metadata System

Establishing AI Risk Management Models

## Business Organizations

Clearance Facilitation and Control Bureau

Audit and Revenue Bureau

Investigation Bureau

International Affairs Bureau

Requesting AI Capabilities for Business Improvement

Providing Feedback on AI and Data Models

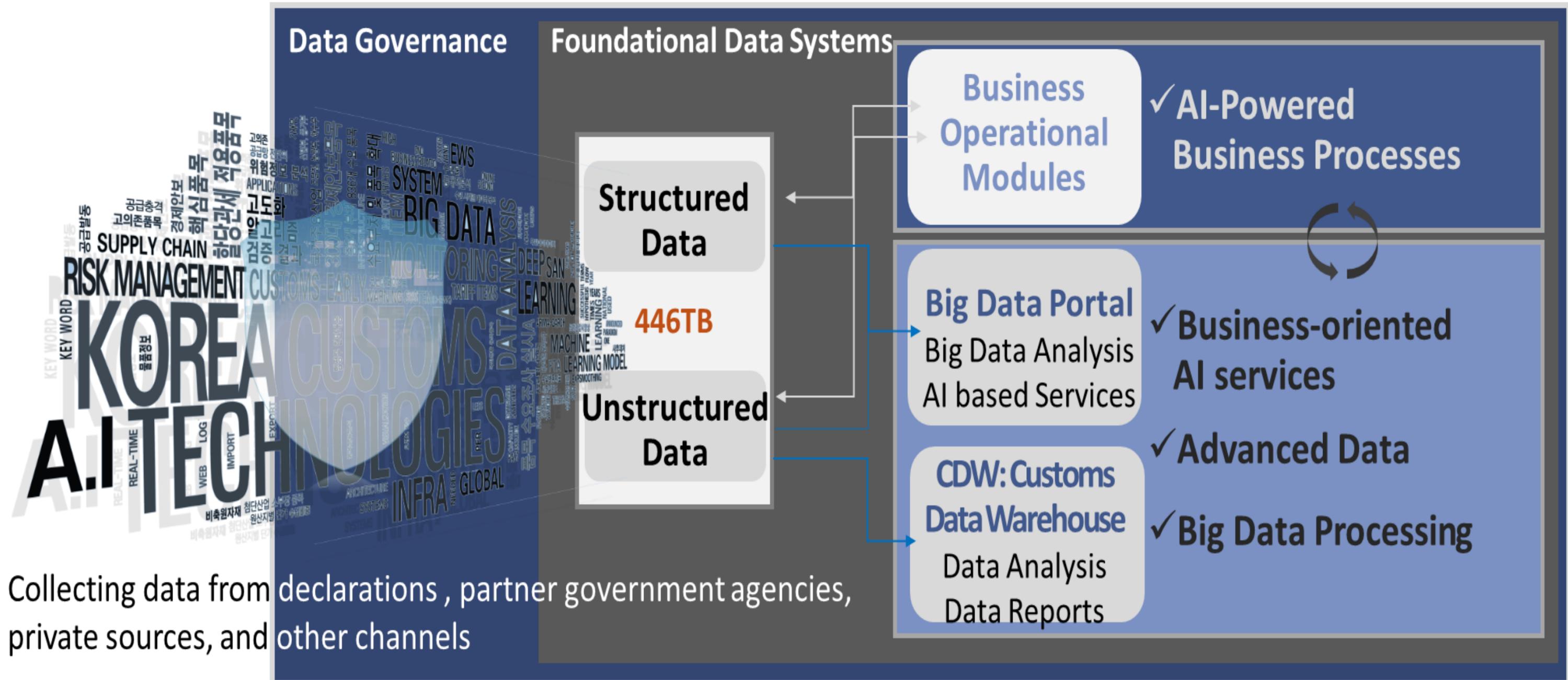
Providing Domain Knowledge

## Branch Customs Office

+ Trained ICT Experts

Utilizing Data Models

✓ Securing big data environment is a fundamental aspect of achieving desired results from data utilization



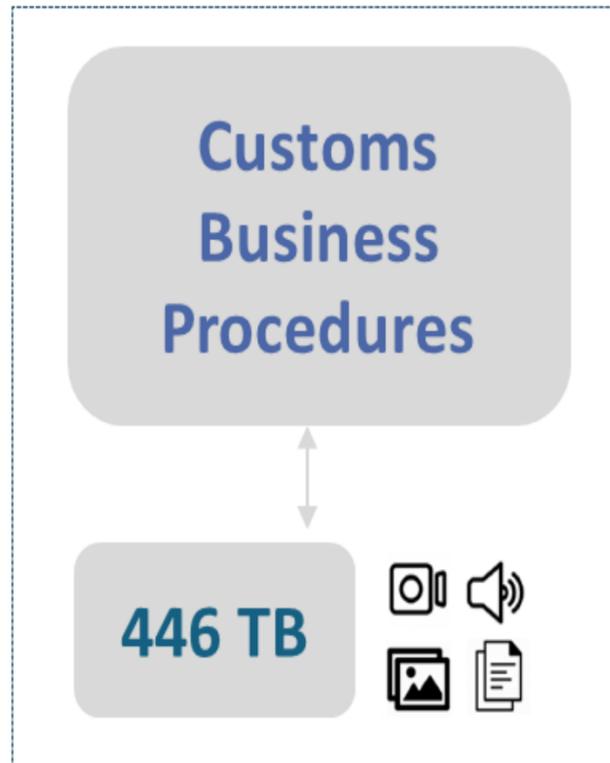
# Systems (Big Data and AI Technology)

✓ AI application requires a deep understanding of technology to grasp its mechanisms and capabilities.

## Big data · AI Platform

Data Analytics Environment + Big Data and AI Services

### Customs Procedures



### Big Data Processing Environment for Big Data Processing

#### Data Analytics Life Cycle

Extract → Store → Pre-process →  
Data Analytics → Visualize

#### Big Data Processing

ETL, ELT tools  
Big Data framework: Hadoop  
Cloud Computing: Kubernetes  
Analytics Tools: Python, SAS

#### High Performance computing Power

GPU, vRAM etc.

### Advanced Data analytics

#### Business Oriented Services

#### Traditional + Advanced Data Analytics

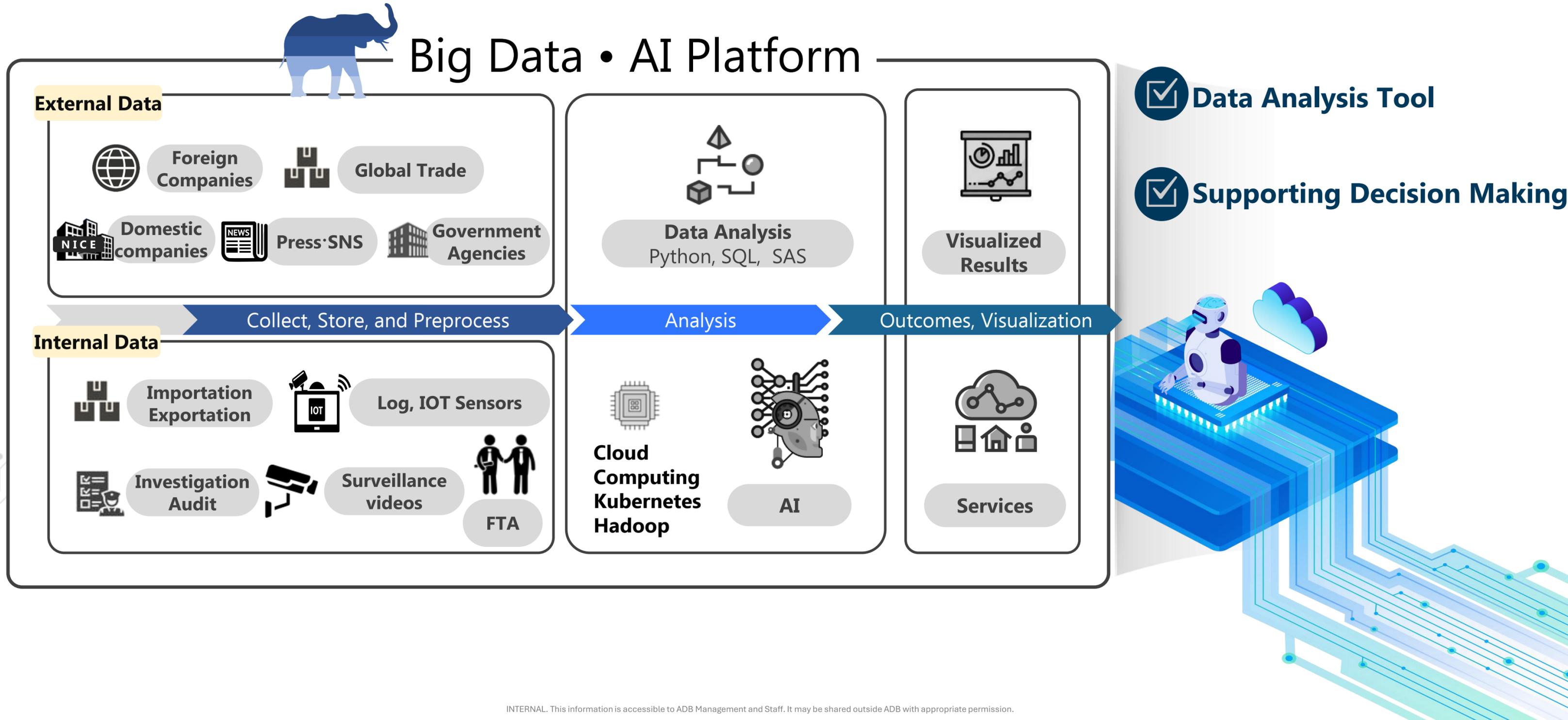
Structured Data + Unstructured Data  
Rule-based AI + Advanced AI

#### AI Technologies

Advanced statistical analytics  
Machine learning  
Deep learning  
Natural language processing  
Generative AI

# Systems (Big Data Portal)

✓ System environment allows handling big data and providing services based on AI technology



✓ Capability on technology and domain is the key aspect allowing us to achieve the desired results from AI application

## • Fostering Data Experts

### Training Course

#### CDW

Beginner Course, Intermediate Course, Advance Course

(2017~) Data Skills, Data Projects, Data-based solutions

#### Big Data Analytics

Understanding the Big Data Portal

Big Data Analytics (Beginner, Intermediate, Advanced)

Data Expert Fostering Program

☑ **160** Trainees

☑ **56** Data projects

#### Data Visualization

Understanding Application of Cutting-edge Technology in Customs Administration

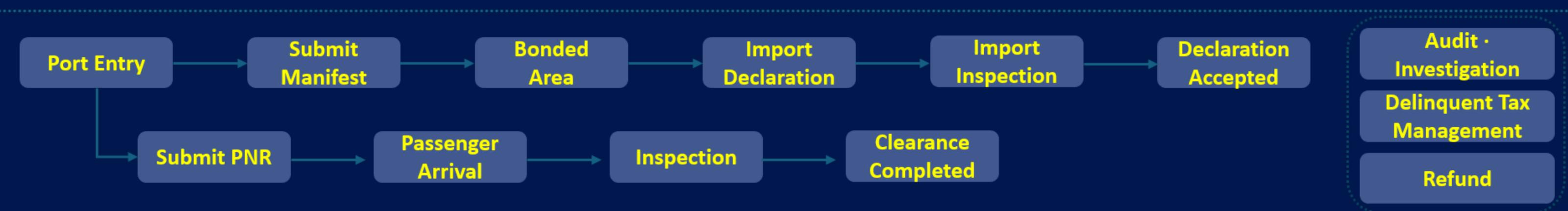
#### Practical Application of Generative AI

Training AI Courses for High Level Officers

(e-Learning) CDW, Practical Application of Generative AI

## • Establishing big data culture

# Artificial Intelligence in Customs Procedures



# Cases I: AI Models Providing Information

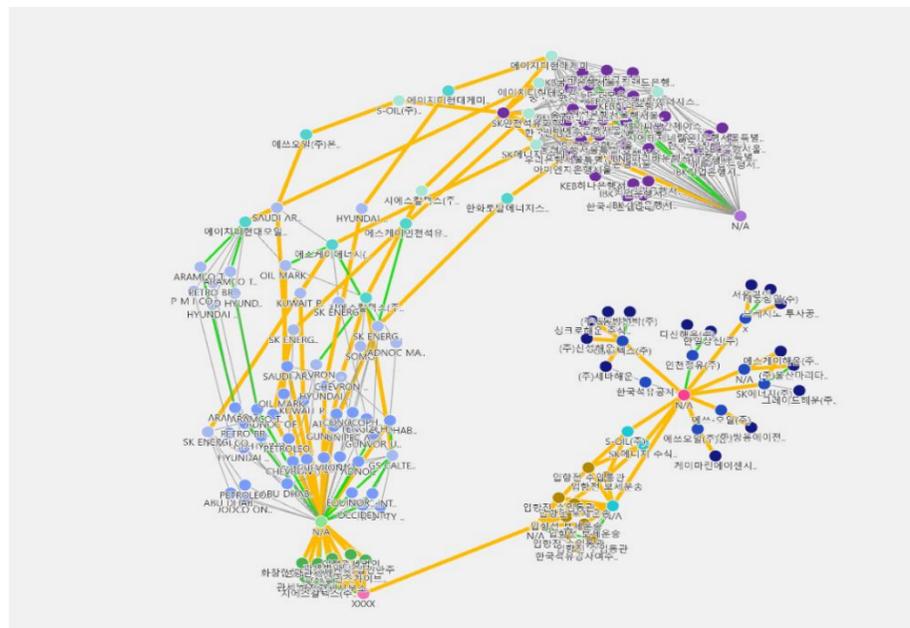
✓ AI models provides information that explains business status based on past and present data

\* **Clustering, FCNN, Random Forest, LightBGM, SVM, RNN etc.**

## Company Profiling

### Supporting Audit

- Natural Language Processing
- Collecting data from internal and external sources
- Information leveraged for auditing by providing details on companies
- \* company overview, financial status, import and export history, stakeholders, comparisons with other companies in the same industry and more



## Network of Stakeholders

### Understanding relationship between key stakeholders

- Visualizing relationship for better understanding
- Providing information including criminal history for each stakeholder to find risk considering the relationship between them

# Cases II : AI Models Controlling Diverse Risks

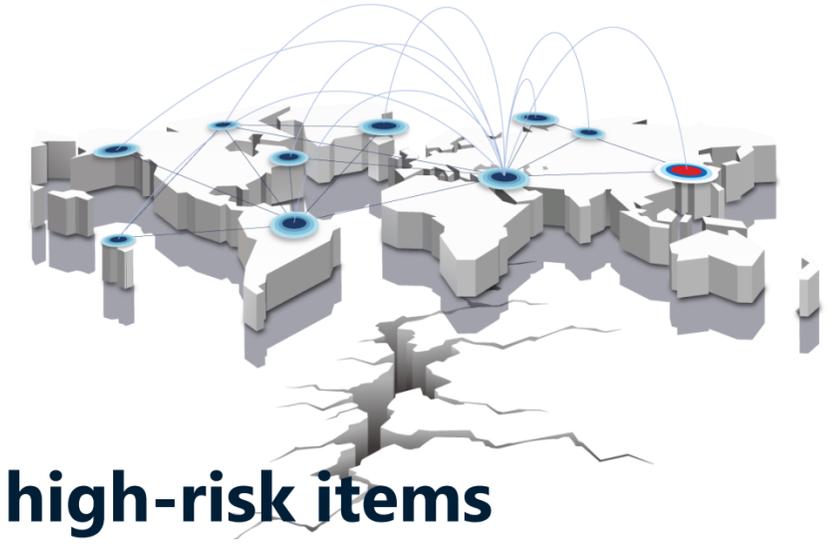
✓ AI models play the prominent roles in the risk management by controlling risk factors of importation

\* **Clustering, FCNN, Random Forest, LightBGM, SVM, RNN etc.**

## AI Customs Clearance Model

### Finding risk levels of importation

- Machine Learning and Deep learning
- Generating group of high-risk importation for a certain period of time
- Measuring level of risks in importation by learning pattern of data



**Focusing on high-risk items  
processing safe one through fast track**

## AI Prediction for possible delinquent taxpayers

### Finding taxpayers with high risk

- Machine learning + Business Factors
- Calculate risk scores of taxpayers from the results of a machine learning model and business factors

Company	Risk Score	Risk Grade
정원	998	01등급
(주)에이 & M	998	01등급
기...무	997	01등급
주식회사 & 산	997	01등급
주식회사 듀얼	997	01등급
농... 주식회사 산...	997	01등급
농업 호	997	01등급
세... Cor	996	01등급
(... 0	996	01등급
주식회사 스타 3C	996	01등급

# Cases III: AI Models with Expanded Range of Data

✓ Diverse data and technologies are used in AI modelling.

## AI Identifying Counterfeit Documents

### Comparing Similarity of Images

- CNN-based models
- Detecting objects(stamps and signatures from images) from the given images
- Comparing similarity of extracted images to the registered ones

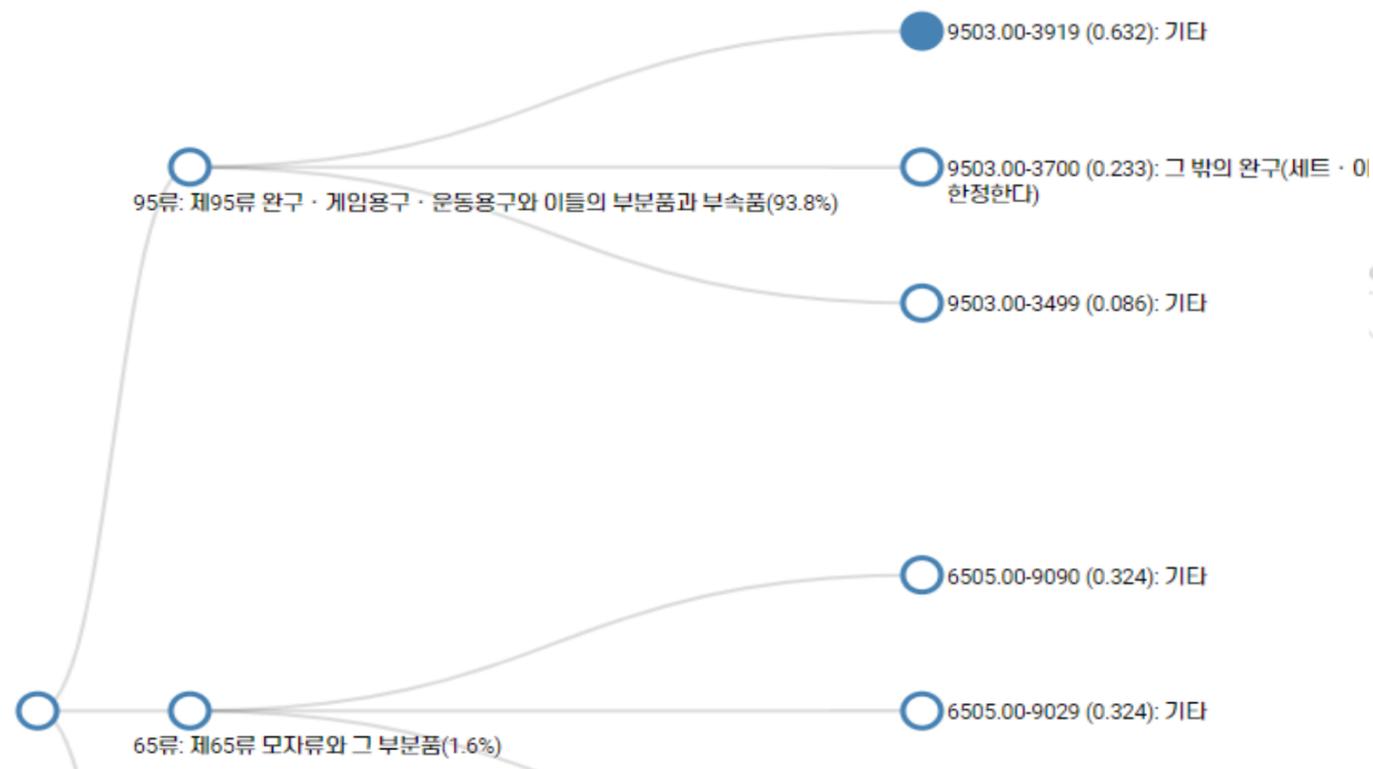
Identifying counterfeit Stamps/Signatures

		Issuing Authority Stamps		Issuer's Signature	
1	Original		NSW BU (S		Rol llum
	Results	60.05%	25.99%	49.17%	100%

## AI HS(Commodity Classification Code) Prediction Model

### Recommending possible Commodity Code

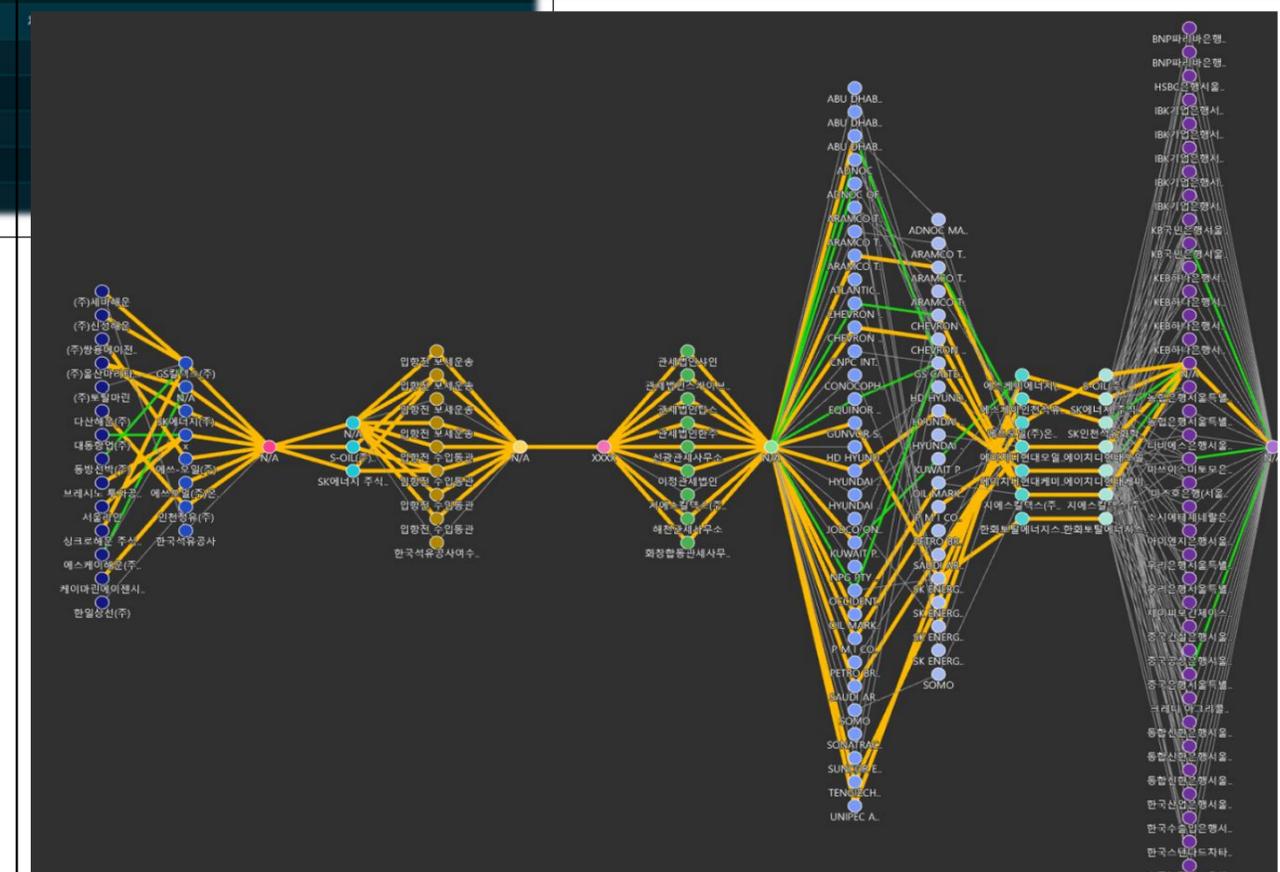
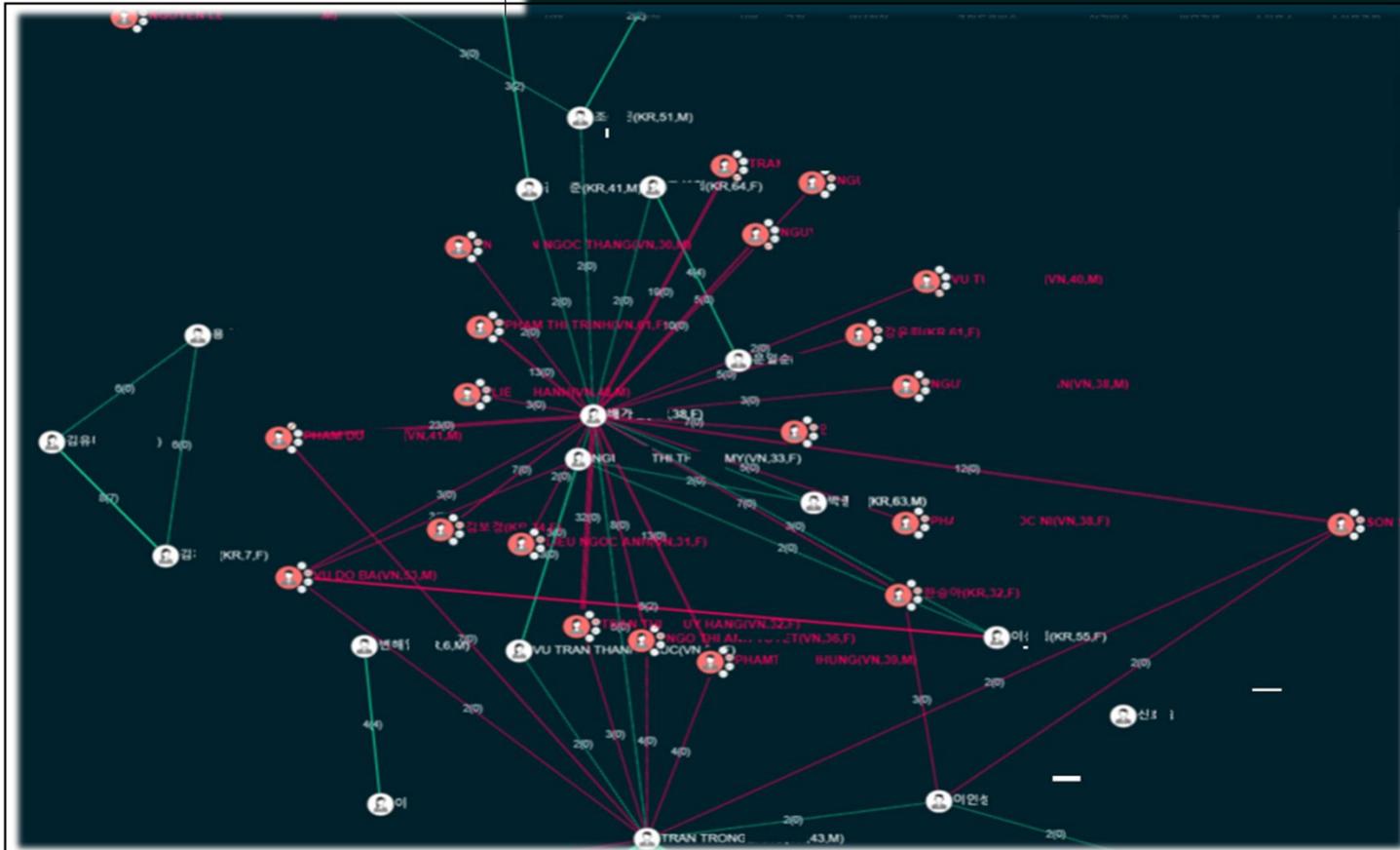
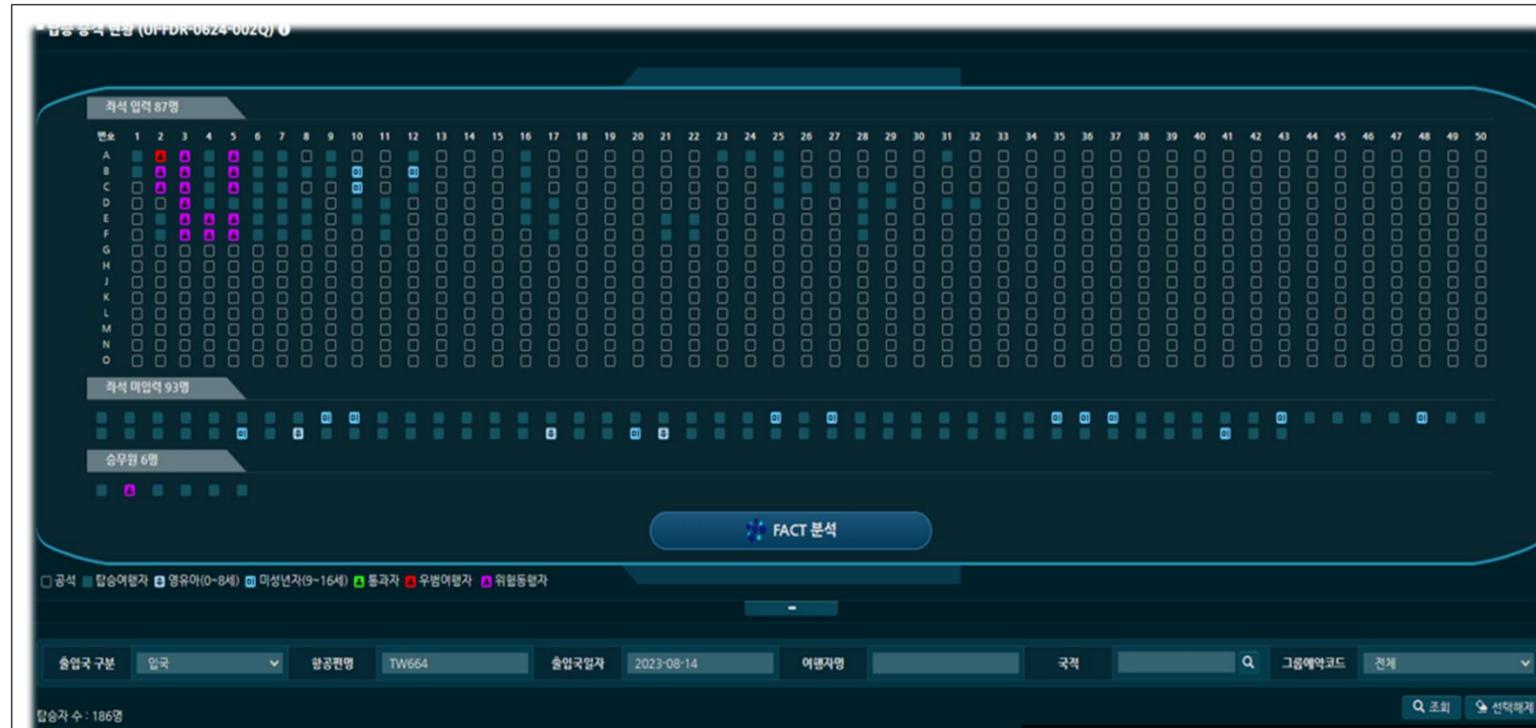
- Random forest, SVM, Logistic regression
- Suggesting possible HS code based on the trained model from the declared item names, descriptions and their HS codes



# Cases IV: Advance Analytics(Network Analysis)

## GraphDB, Network Analysis

- Discovering high-risk travelers, cargo, and delinquent taxpayers by analyzing their potential risk and identifying related factors based on network analysis.



# Applying Generative AI into Customs Procedures

✓ Korean Customs is developing in-house services based on Large Language Models(LLMs)

## ▪ **Task Force Team to build generative AI services tailored for customs administrative**

- **Period:** '25.4.1~7.5

- **Members:** 6 IT experts

- **Goals**

- ① Enhancing capability on generative AI development

- ② Securing generative AI human resources

- **Activities**

- ① Researching and verifying generative AI technology

- ② Building LLM-based pilot services

- ③ Fostering generative AI experts

## ■ Building LLM-based pilot services

- **Scope**

- Building services to verifying the core technology and generating data

- **Goals**

- ① Understanding principal and mechanism of core technology required to build generative AI services
- ② Securing high quality data utilized to optimize generative AI services



### Generative AI LAB.

General Q&A Chatbot

Management Datasets of Q&A Conversation

Automatic Generation Q&A Datasets

Management Datasets of automatically generated Q&A

Q&A on Taxable value calculation

✓ Korean Customs is developing in-house services based on Large Language Models(LLMs)

## Customs Valuation Q&A Pilot Service

Increase Accessibility in Knowledge

- Applied RAG to improve the performance of LLM-based services

Hi, I have a question about AI technology. What are the key technologies that enable generative AI, especially those related to LLM-based AI applications?

Okay, let's break down the key technologies that enable generative AI, with a focus on those powering LLM (Large Language Model) based applications. I'll organize this into categories: **Core Architecture, Training Techniques, and Optimization/Deployment.**

### 1. Core Architecture & Mechanisms:

- **Transformer Architecture:** This is *the* foundational architecture. It replaced Recurrent Neural Networks (RNNs) due to its ability to handle long-range dependencies and parallelize computations. Key features:
  - **Self-Attention:** Allows the model to weigh the importance of different words in a sequence when understanding context. This is what allows LLMs to "understand" relationships between words that are far apart.
  - **Encoder-Decoder Structure (often simplified):** While the original Transformer had both, many LLMs use only the decoder portion for generative tasks.
- **Attention Mechanisms (Beyond Self-Attention):** Variations like Multi-Head Attention allow the model to attend to different aspects of the input simultaneously.
- **Positional Encoding:** Since Transformers don't inherently understand word order, positional encoding adds information about the position of words in the sequence.

### 2. Training Techniques:

- **Self-Supervised Learning:** LLMs are primarily trained using self-supervised learning. This means they learn from unlabeled text data. Common techniques include:
  - **Next Token Prediction (Causal Language Modeling):** The model predicts the next word in a sequence. This is the most common training objective for generative LLMs.
  - **Masked Language Modeling (MLM):** (Used in models like BERT, but less common for pure generative models) The model predicts masked words in a sentence.
- **Reinforcement Learning from Human Feedback (RLHF):** This is *critical* for aligning LLMs with human preferences. It involves:
  - **Reward Modeling:** Training a model to predict human preferences for different model outputs.

Hi, I have a question about AI technology. What are the key technologies that enable generative AI, especially those related to LLM-based AI applications?

Okay, let's break down the key technologies that enable generative AI, with a focus on those powering LLM (Large Language Model) based applications. I'll organize this into categories: **Core Architecture, Training Techniques, and Optimization/Deployment.**

## 1. Core Architecture & Mechanisms:

- **Transformer Architecture:** This is *the* foundational architecture. It replaced Recurrent Neural Networks (RNNs) due to its ability to handle long-range dependencies and parallelize computations. Key features:
  - **Self-Attention:** Allows the model to weigh the importance of different words in a sequence when understanding context. This is what allows LLMs to "understand" relationships between words that are far apart.
  - **Encoder-Decoder Structure (often simplified):** While the original Transformer had both, many LLMs use only the decoder portion for generative tasks.
- **Attention Mechanisms (Beyond Self-Attention):** Variations like Multi-Head Attention allow the model to attend to different aspects of the input simultaneously.
- **Positional Encoding:** Since Transformers don't inherently understand word order, positional encoding adds information about the position of words in the sequence.

## 2. Training Techniques:

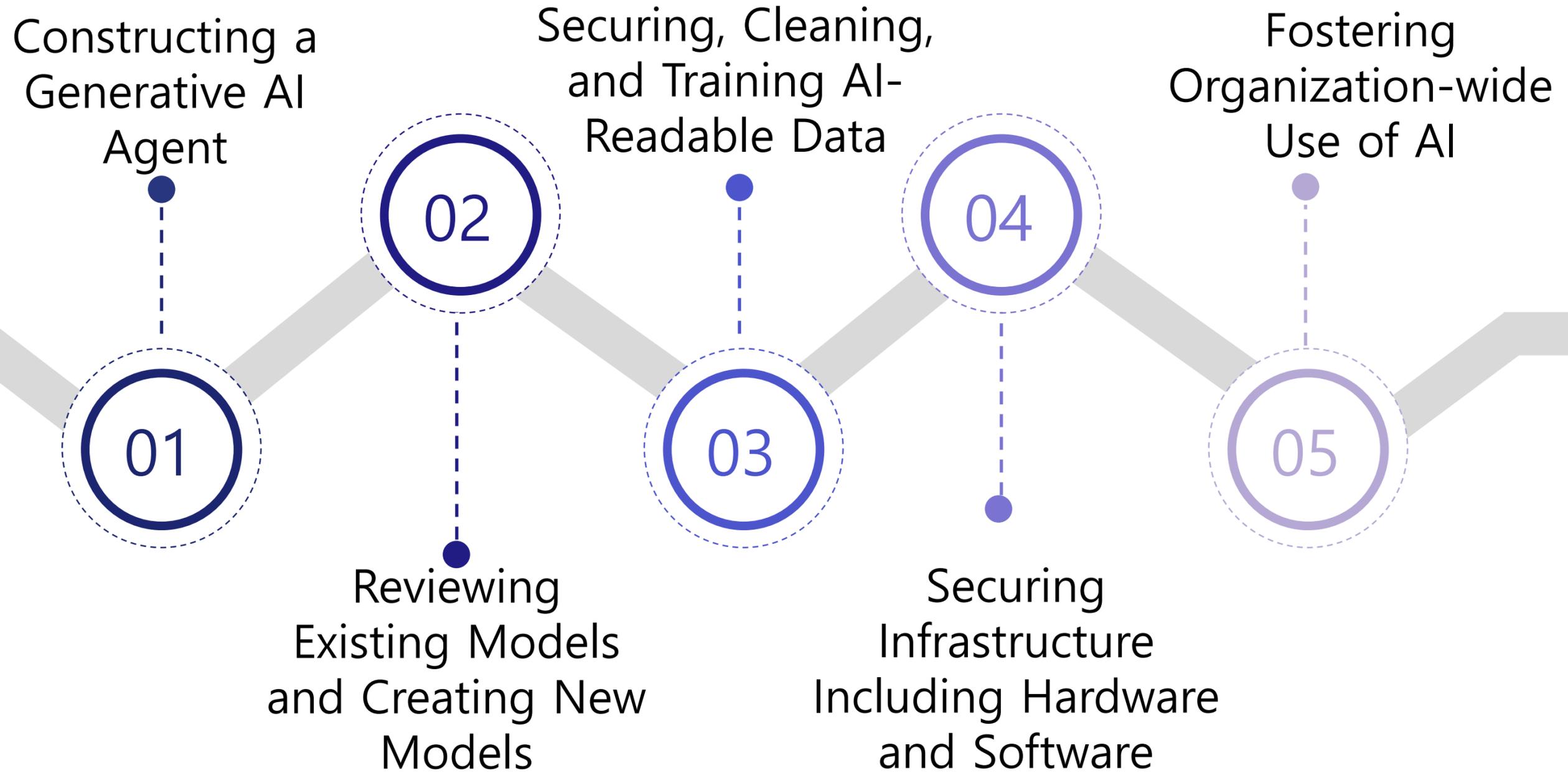
- **Self-Supervised Learning:** LLMs are primarily trained using self-supervised learning. This means they learn from unlabeled text data. Common techniques include:
  - **Next Token Prediction (Causal Language Modeling):** The model predicts the next word in a sequence. This is the most common training objective for generative LLMs.
  - **Masked Language Modeling (MLM):** (Used in models like BERT, but less common for pure generative models) The model predicts masked words in a sentence.
- **Reinforcement Learning from Human Feedback (RLHF):** This is *critical* for aligning LLMs with human preferences. It involves:
  - **Reward Modeling:** Training a model to predict human preferences for different model outputs.

Agent

## AI Models in Customs Procedures



# Roadmap for Generative AI



 **Business-Centered Applications**

 **In-Depth Understanding of Technology**

 **Large-Scale and High Quality Data**

 **Diagnosing Status, Defining Strategy**

Thank you