

Технологичная и надёжная таможенная администрация

Как Таможенная служба Кореи использует данные и современные технологии?



Таможенная служба Кореи
Мирян Ким, 25.11.25

IT-подразделения

Бюро политики в области ИКТ и данных

Департамент планирования ИКТ

Департамент инноваций в области ИКТ

Департамент управления ИКТ и данными

Департамент И&P и оборудования

Деп. эксплуатации системы UNI-PASS

Центр таможенного целевого контроля на границе

Бизнес-подразделения

Бюро содействия таможенному оформлению и контролю

Бюро аудита и доходов

Бюро расследований

Бюро международных связей

Региональные таможни

Подготовленные ИКТ-специалисты

Определение направления развития ИКТ, проведение ПИС, контроль качества данных

Эксплуатация платформы больших данных (включая ML и DL), подготовка специалистов ИКТ, внедрение генеративного ИИ

Эксплуатация хранилища таможенных данных, обеспечение информационной безопасности

Исследования ИИ, встроенного в устройства

Эксплуатация системы UNI-PASS (рискоориентированное управление на основе ИИ), управление метаданными

Разработка моделей управления рисками на основе ИИ

Запросы на применение ИИ для улучшения бизнес-процессов

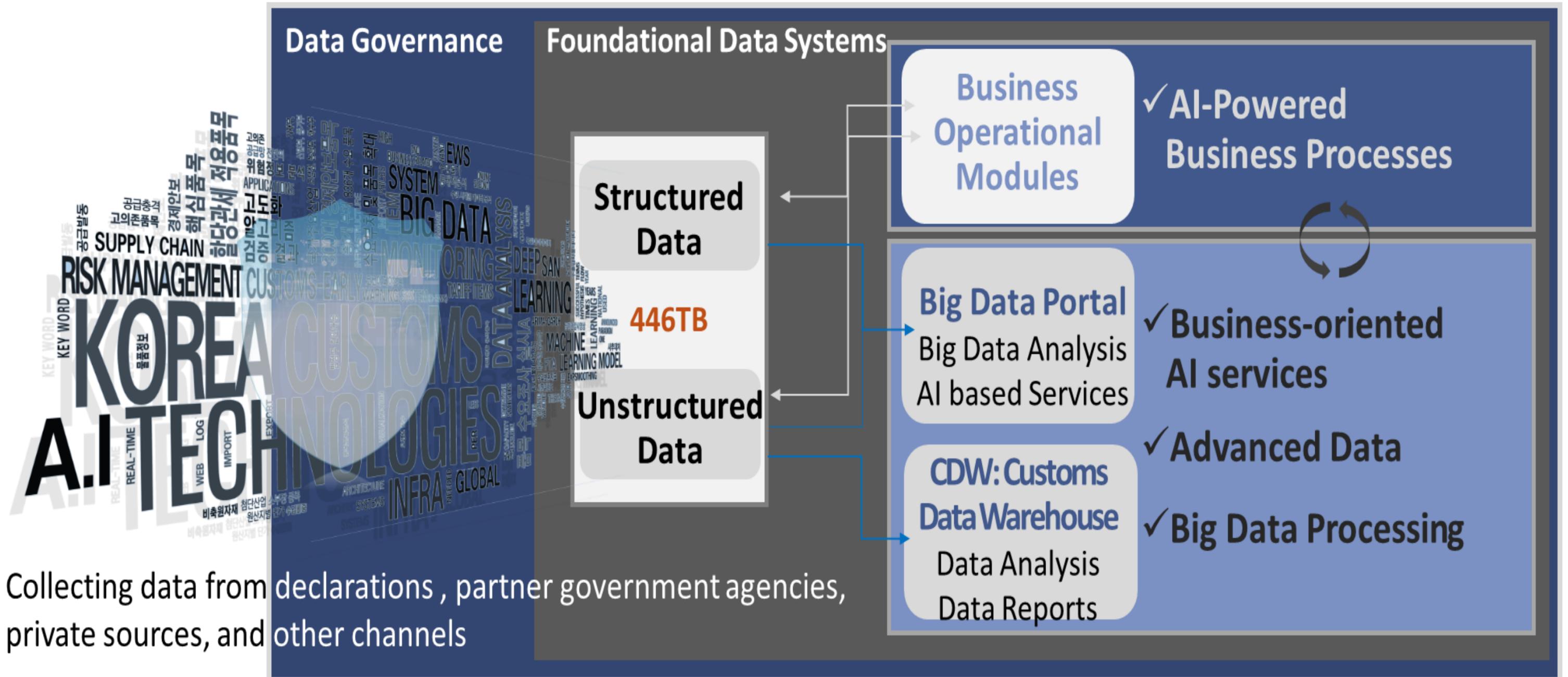
Предоставление обратной связи по моделям ИИ и данным

Предоставление отраслевых знаний

Использование разработанных моделей данных

Большие данные

- ✓ Обеспечение среды для работы с большими данными является фундаментальным условием для достижения желаемых результатов от использования данных.



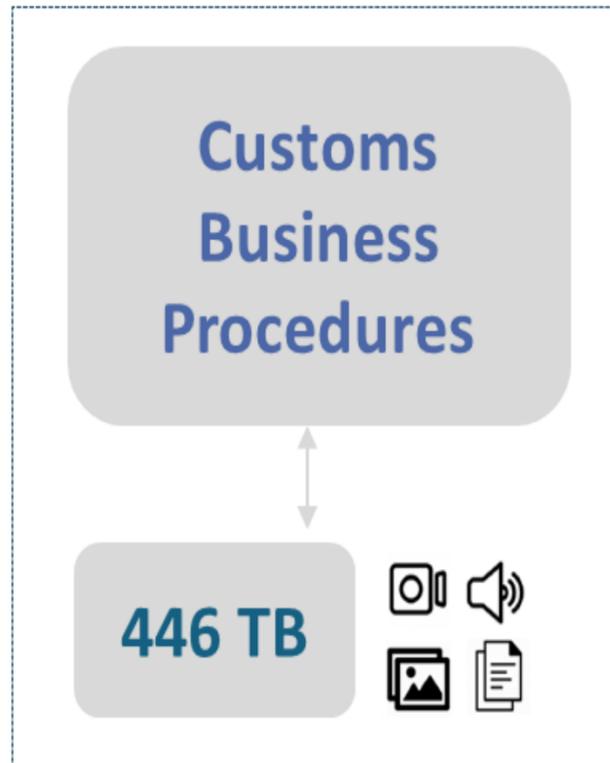
Системы (технологии больших данных и ИИ)

- ✓ Применение технологий искусственного интеллекта требует глубокого понимания самой технологии, её механизмов и возможностей.

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

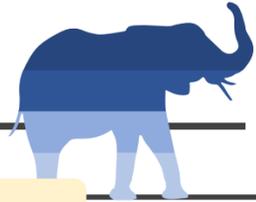
Structured Data + Unstructured Data
Rule-based AI + Advanced AI

AI Technologies

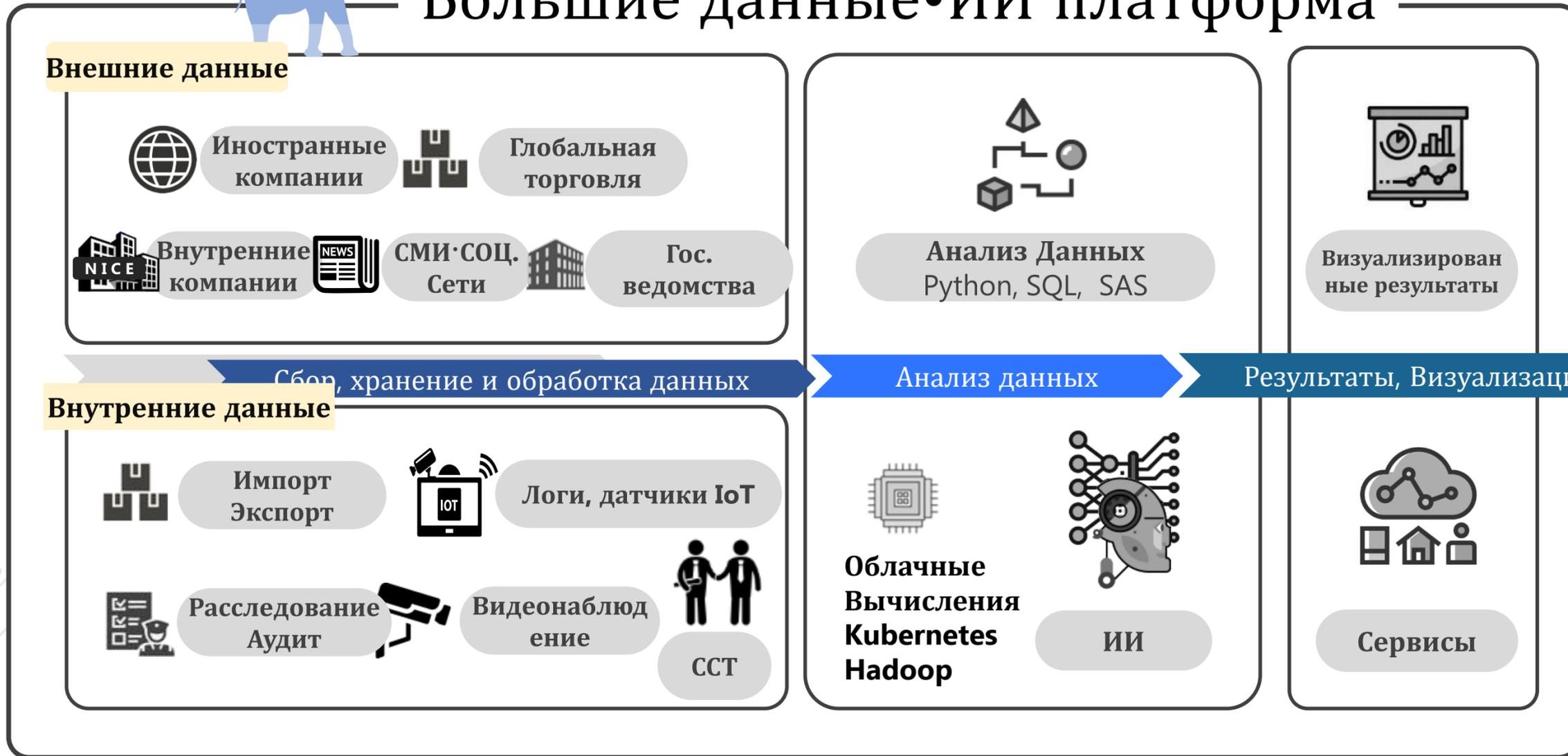
Advanced statistical analytics
Machine learning
Deep learning
Natural language processing
Generative AI

Системы (портал больших данных)

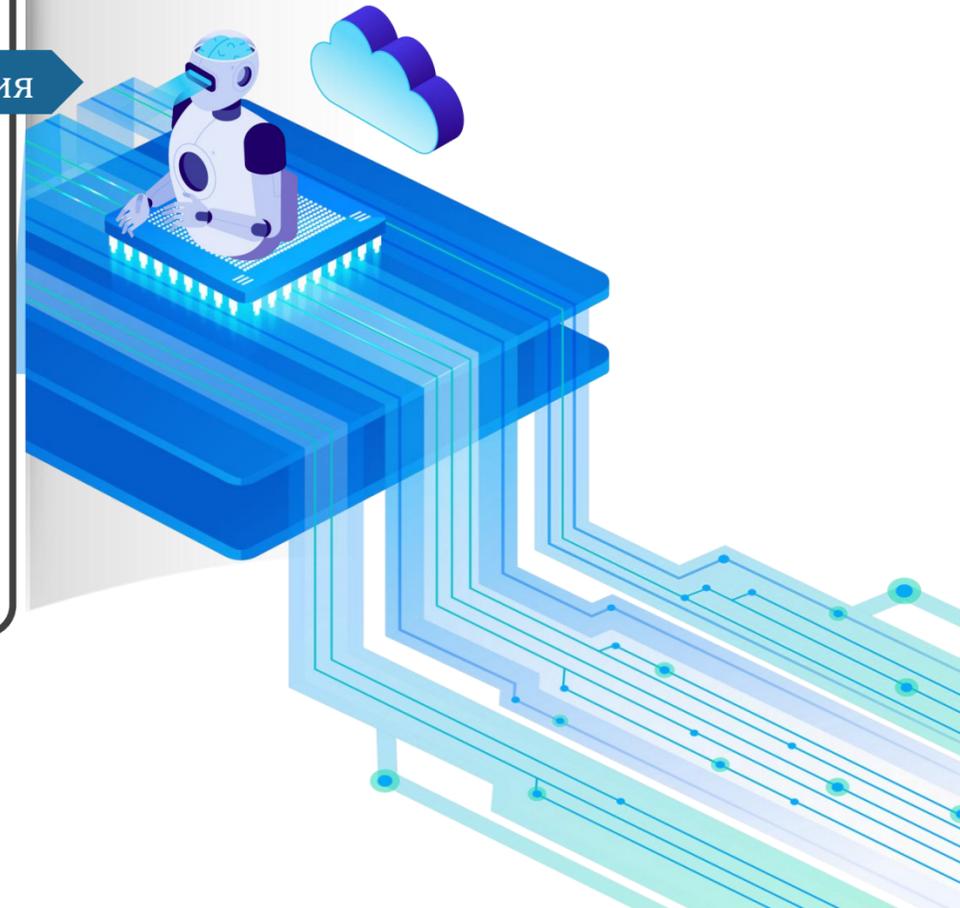
- ✓ Системная среда обеспечивает обработку больших данных и предоставление сервисов на основе технологий искусственного интеллекта.



Большие данные • ИИ платформа



- ✓ **Инструмент для анализа данных**
- ✓ **Поддержка принятия решений**



✓ Компетенции в области технологий и отраслевой специфики — ключевой фактор, позволяющий достигать желаемых результатов при применении ИИ.

• Подготовка экспертов по данным

Учебные курсы

ТХД

Базовый Курс, Средний Курс, Продвинутой Курс

(2017~) Навыки работы с данными, Проекты основанные на данных, Решения на основе данных

Аналитика больших данных

Понимание работы портала больших данных

Аналитика больших данных (базовый, средний, продвинутой уровни)

Программа подготовки экспертов по данным

☑ **160** Обучающихся

☑ **56** Проектов с данными

Визуализация данных

Понимание применения передовых технологий в таможенном администрировании

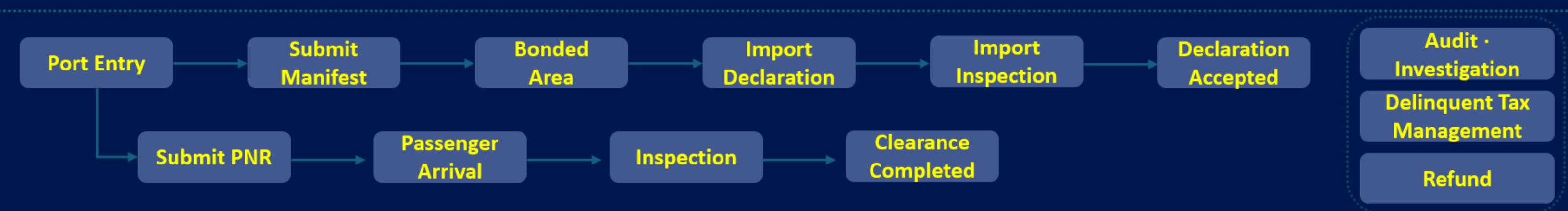
Практическое применение генеративного ИИ

Курсы по ИИ для руководящего состава

(Онлайн-Обучение) ТХД, Практическое применение генеративного ИИ

• Формирование культуры работы с большими данными

Искусственный интеллект в таможенных процедурах



Случай I: ИИ-модели, предоставляющие информацию

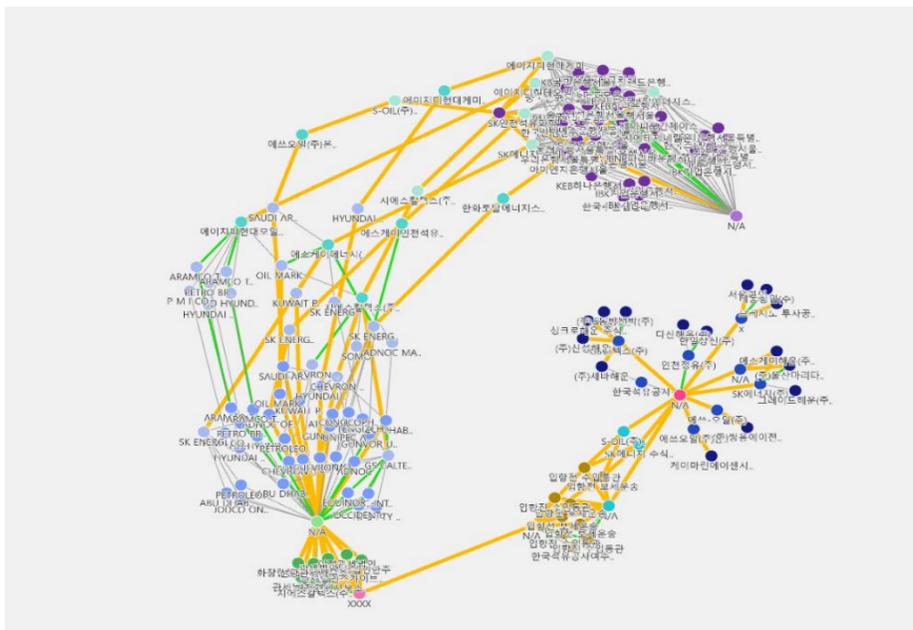
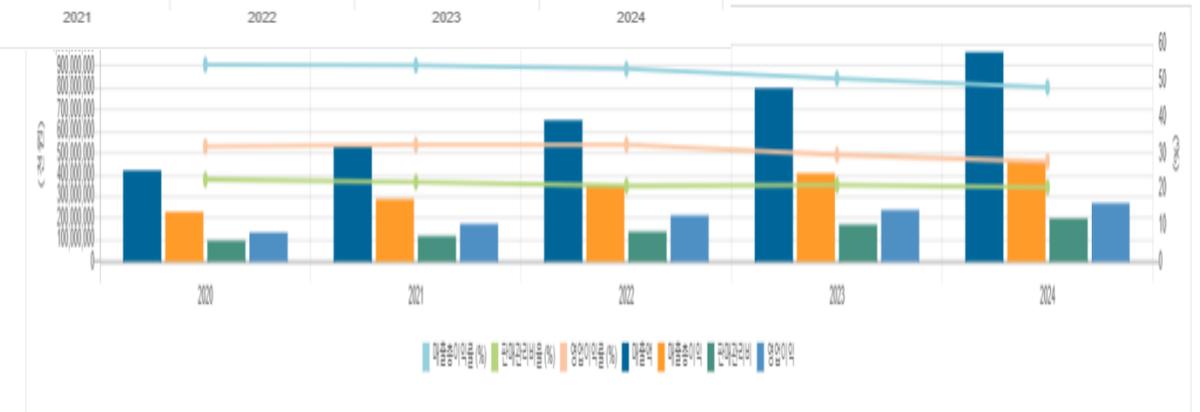
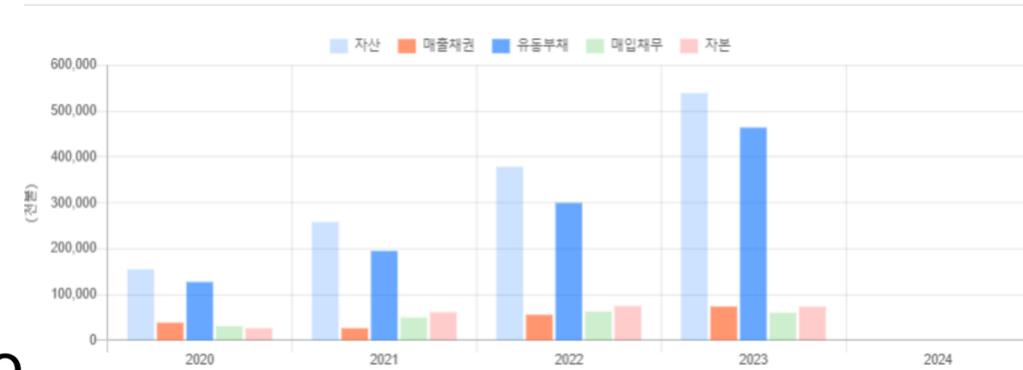
✓ Модели ИИ предоставляют информацию, которая объясняет состояние деятельности на основе прошлых и текущих данных.

* **Clustering (кластеризация), FCNN, Random Forest, LightGBM, SVM, RNN и др.** *

Профилирование компаний

Поддержка аудита

- Обработка естественного языка (NLP)
 - Сбор данных из внутренних и внешних источников
 - Использование информации для проведения аудита, включая сведения о компаниях
- * общий профиль компании, финансовое состояние, история импорта и экспорта, ключевые участники, сравнение с другими компаниями в той же отрасли и др. *



Сеть заинтересованных сторон

Понимание Связи Между Ключевыми Участниками

- Визуализация отношений для лучшего понимания
- Предоставление информации, включая криминальную историю каждого участника, чтобы выявить риски, основанные на связях между ними

Случай II: ИИ-модели для контроля различных рисков

✓ Модели ИИ играют ключевую роль в управлении рисками, контролируя факторы риска при импорте.

Clustering (кластеризация), FCNN, Random Forest, LightGBM, SVM, RNN и др. *

Модель ИИ для таможенного оформления

Определение уровней риска при импорте

- Машинное обучение и глубокое обучение
- Формирование группы «высокого риска» импорта за определённый период
- Оценка уровня рисков при импорте на основе изучения закономерностей в данных



Фокус на товарах с высоким риском
Обработка безопасных товаров через «быструю полосу» (**Fast Track**)



Прогнозирование потенциальных проблемных налогоплательщиков

Выявление налогоплательщиков с высоким риском

- Машинное обучение + бизнес-факторы
- Расчёт баллов риска налогоплательщиков на основе результатов модели машинного обучения и бизнес-факторов

Company	Risk Score	Risk Grade
정원	998	01등급
(주)에이오 & M	998	01등급
기흥 무역	997	01등급
주식회사 & 산	997	01등급
주식회사 듀얼	997	01등급
늘 주식회사 산...	997	01등급
농업 호	997	01등급
세원 Cor	996	01등급
(주) O	996	01등급
주식회사 스타 3C	996	01등급

Случай III: ИИ-модели с расширенным набором данных

✓ Для моделирования с использованием ИИ применяются разнообразные данные и технологии.

ИИ для выявления поддельных документов

Сравнение схожести изображений

- Модели на основе CNN (Convolutional Neural Network — сверточные нейронные сети)
- Обнаружение объектов (печати и подписи на изображениях)
- Сравнение схожести извлечённых объектов с зарегистрированными образцами

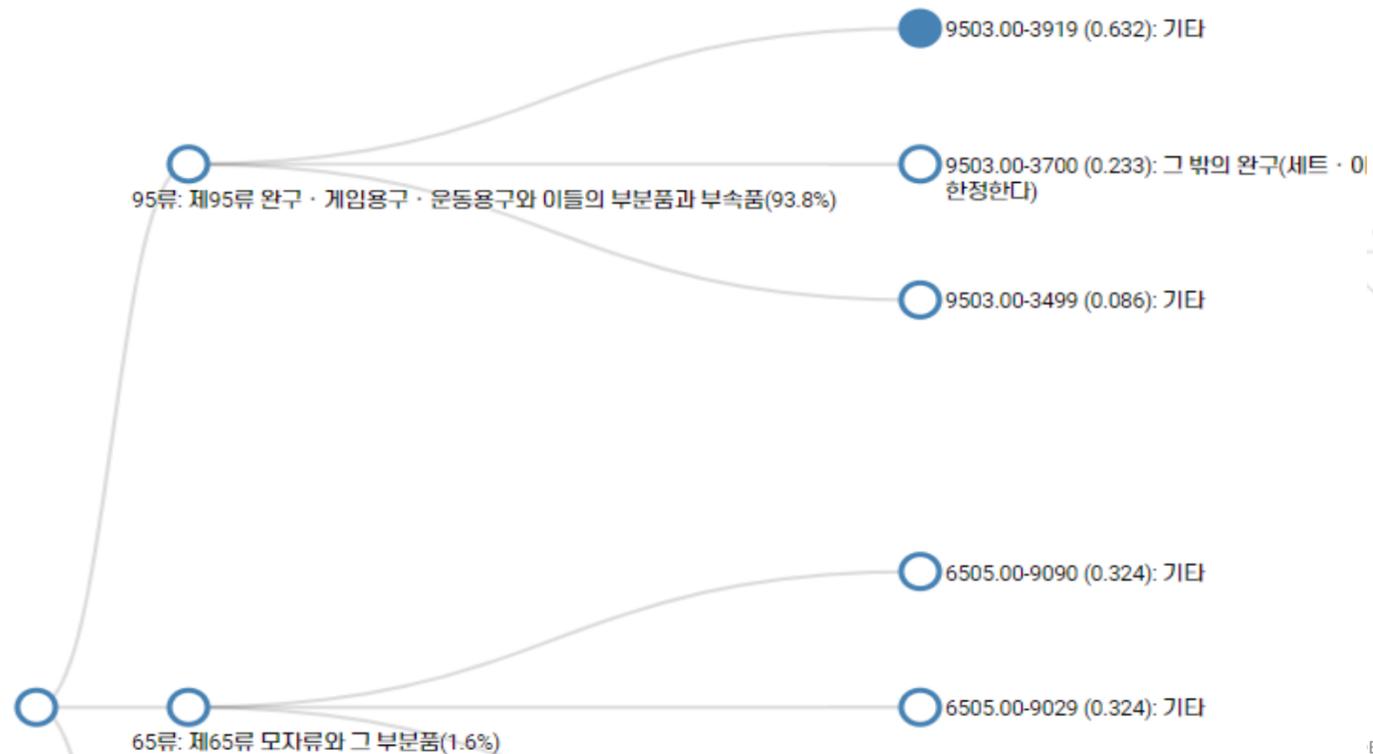
Identifying counterfeit Stamps/Signatures

		Issuing Authority Stamps		Issuer's Signature	
Targets	Original		SY CS		RC .UM
	Results	60.05%	25.99%	49.17%	100%
1	Original		NSW BU (S) IBER		Rol llum
	Results	60.05%	25.99%	49.17%	100%

ИИ-модель прогнозирования HS-кодов (классификационный код товаров)

Рекомендация возможного кода товара

- Random Forest, SVM, Logistic Regression
- Предложение возможного HS-кода на основе обученной модели, учитывающей заявленные названия товаров, описания и их HS-коды



✓ Таможенная служба Кореи разрабатывает внутренние сервисы на основе Large Language Models (LLMs) — больших языковых моделей.

- **Целевая группа для разработки сервисов генеративного ИИ, адаптированных под таможенное администрирование**

- **Период: '25.4.1~7.5**

- **Состав: 6 ИТ-специалистов**

- **Цели**

- ① Повышение компетенций в разработке генеративного ИИ

- ② Обеспечение кадровыми ресурсами в области генеративного ИИ

- **Деятельность**

- ① Исследование и проверка технологий генеративного ИИ

- ② Создание пилотных сервисов на базе LLM

- ③ Подготовка экспертов по генеративному ИИ

■ Создание пилотных сервисов на базе LLM

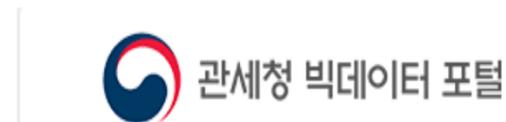
• Область применения

- Разработка сервисов для проверки ключевых технологий и генерации данных

• Цели

① Понимание принципов и механизмов ключевых технологий, необходимых для создания сервисов генеративного ИИ

② Обеспечение высококачественными данными для оптимизации сервисов генеративного ИИ



OCR 문자 추출

생성형 AI 실험실

질의 응답 챗봇

(챗봇) 질의 응답 품질 관리

질의 응답 자동 생성

(자동 생성) 질의 응답 품질 관리

과세가격 질의응답

Лаборатория генеративного ИИ

Общий чат-бот для вопросов и ответов (General Q&A Chatbot)

Управление наборами данных диалогов Q&A

Автоматическая генерация наборов данных Q&A

Управление наборами данных автоматически сгенерированных Q&A

Q&A по расчёту налогооблагаемой стоимости

✓ Таможенная служба Кореи разрабатывает внутренние сервисы на основе Large Language Models (LLMs).

Пилотный сервис Q&A по таможенной оценке

Повышение доступности знаний

- Применение RAG (Retrieval-Augmented Generation) для улучшения работы сервисов на базе LLM

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:

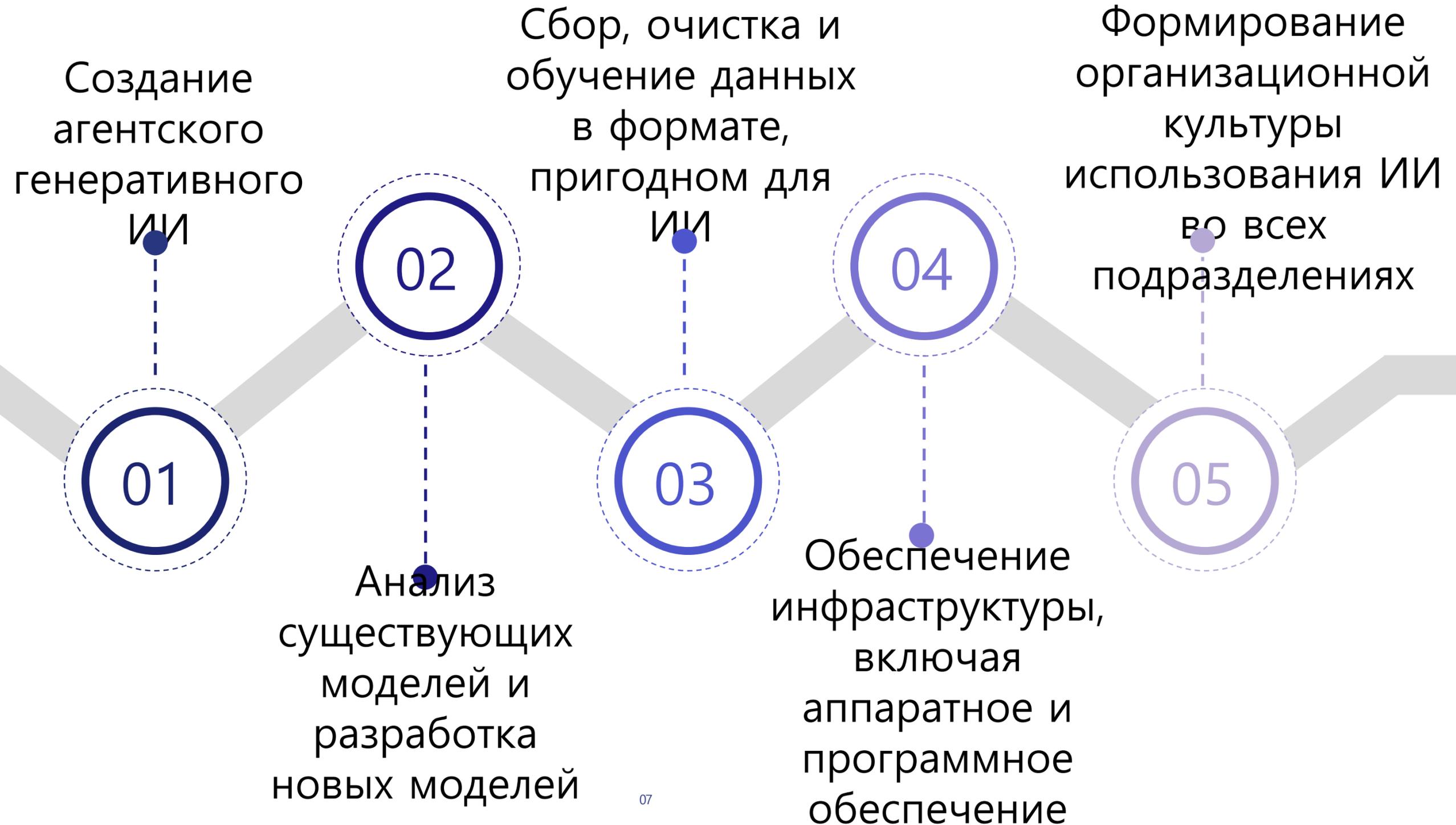
- **Self-Supervised Learning:** LLMs are primarily trained using self-supervised learning. This means they learn from unlabeled text data. Common techniques include:
 - **Next Token Prediction (Causal Language Modeling):** The model predicts the next word in a sequence. This is the most common training 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



Дорожная карта по генеративному ИИ



- Применение ИИ с ориентацией на бизнес**
- Глубокое понимание технологий**
- Масштабные и высококачественные данные**
- Диагностика текущего состояния и определение стратегии**

Спасибо!