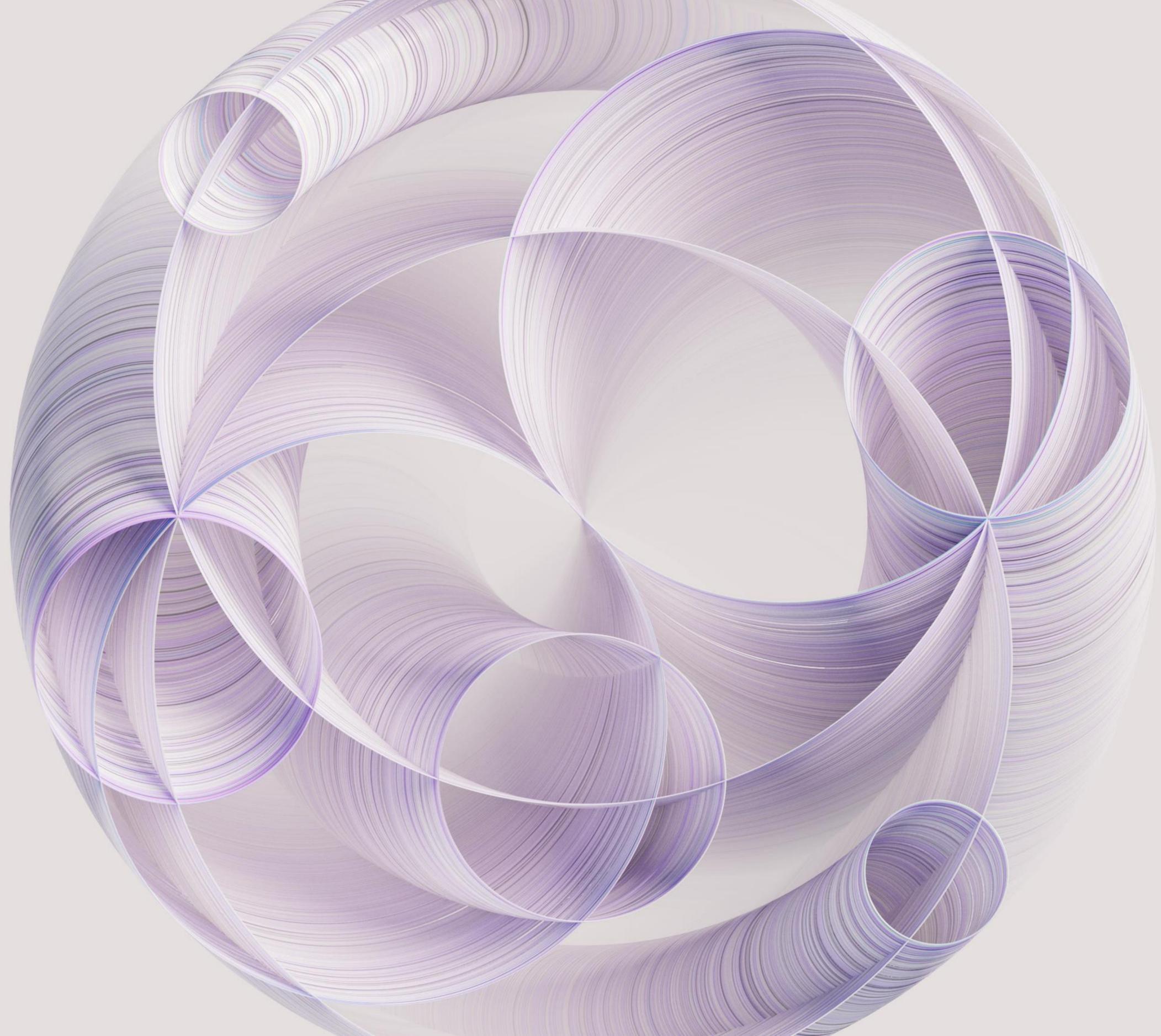


IBM

Контролируемая адаптация ИИ как конкурентное преимущество

Павел Мальцев
pmaltsev@cz.ibm.com
Subject Matter Expert, Data and AI,
Europe

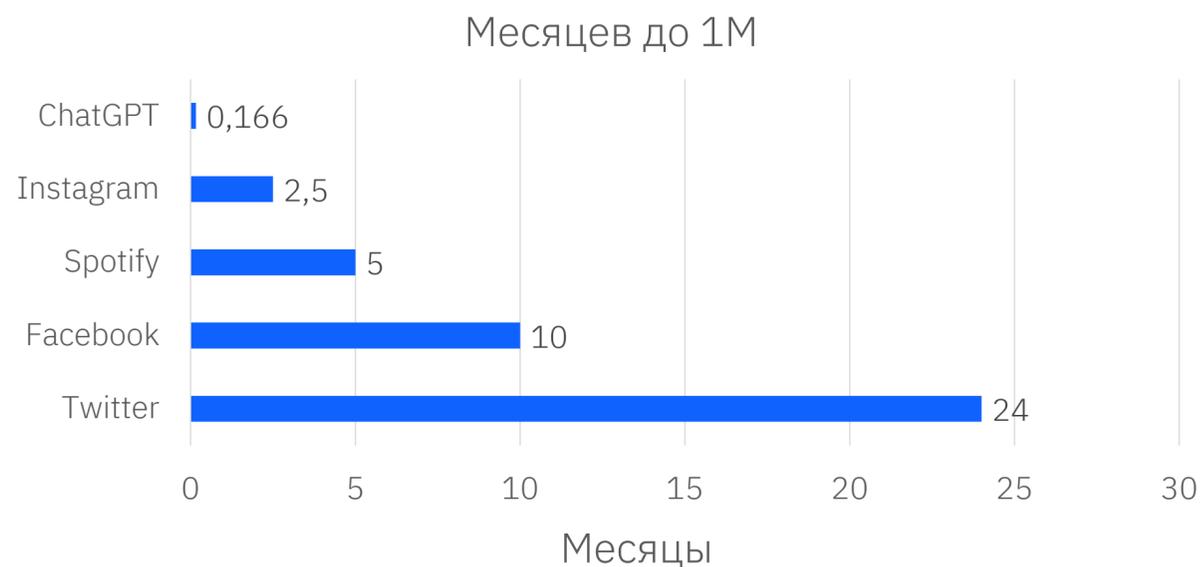


Скорость, масштаб и размах воздействия генеративного ИИ беспрецедентны

Массовое принятие рядовыми потребителями



1 М пользователей за 1 неделю **100 М** пользователей за 2 месяца



Широкомасштабное принятие корпоративными пользователями

29%

Организаций в США, Германии и Великобритании запустили проекты и используют ИИ¹

34%

В основном используют GenAI встроенный в существующие приложения как главный способ реализации GenAI кейсов¹

Generative AI – это не только чат-боты



Чат боты популярны и интересны для персонального использования

- Они не эквивалентны generative AI моделям, а являются кастомными приложениями
- Клиентам необходимо их адаптировать под конкретные задачи и существующие процессы



Prompt Engineering?
Как использовать различные модели?

Как настроить?
Мониторить? Управлять?
Оценивать?



Что такое fine-tuning?
Что такое prompt tuning?

Множество моделей...
Вендорские?
Open source?

Влияние размера?
Как оценить?
Где взять данные для test/dev?

Как контролировать чат?

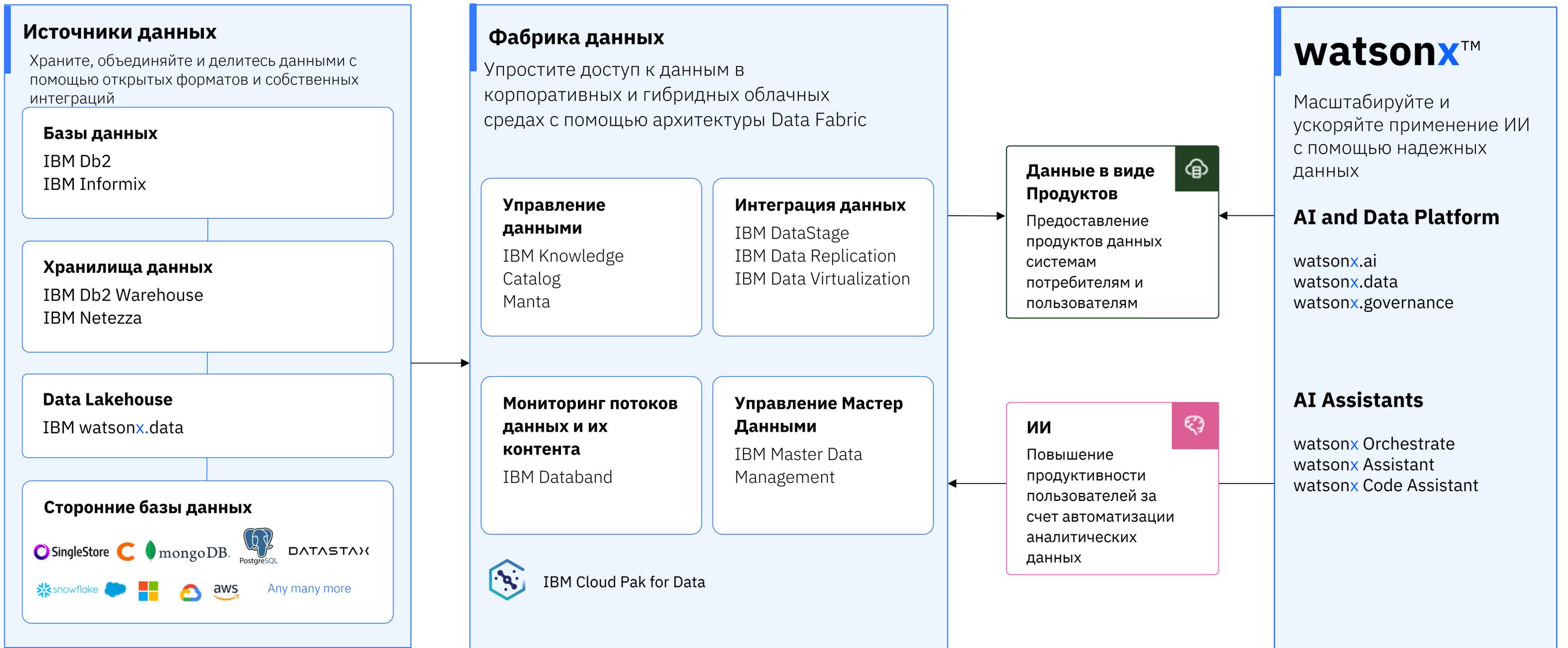
Что насчет API, SDK?

Как оптимизировать затраты?

Как использовать сообщества экспертов (SME) для улучшения модели?



Инвестиции в понятные и надежные исходные данные ускорят и отмасштабируют ваши инициативы в области ИИ



← Автоматизированный data lineage →

Получите полное представление о своих данных и их пути от источника до конечного использования с целью соблюдения нормативных требований и вариантов использования в ИИ с Manta (компанией IBM)

Встроенное, основанное на LLM обогащение метаданных

Что мы представляем?

- **Новое** – Автоматическая генерация смысловых наименований колонок и описаний на основании контекста
- **Улучшение** – привязка технических метаданных к терминам на основании анализа с LLM
- **Предварительные достижения** – внутреннее тестирование показало **удвоение автоматического корректного маппинга** (уменьшение ложно-положительных результатов), и **увеличение точности привязки терминов** с использованием LLMs

Почему?

Повышение производительности в задачах каталогизирования с повышенной точностью присвоения терминов через использование ИИ и LLM.

Унификация контекста бизнес-метаданных между Knowledge Catalog и watsonx.data

The screenshot shows the 'DB2 Bank Database Asset Enrichment' interface in watsonx.data. It displays a table of assets with columns for Name, Display name, Description, Source, and Business terms. The table lists various tables like ACCOUNT_HOLDERS, BANK_CLIENTS, and SAVINGS_ACCOUNTS. AI-generated descriptions and display names are shown next to each asset. A callout box points to the AI-generated content with a confidence score of 83%. Another callout points to the AI-generated description, and a third points to the AI-generated display name. The interface also shows a sidebar with details for the selected asset, including its source, type, and enrichment options.

Name	Display name	Description	Source	Business terms
ACCOUNT_HOLDERS Table	AI Account holders	Table containing data of debit acco...	DB2 Bank.. / BANK3	AI 2 Sugges
BANK_CLIENTS Table	Banking clients	AI This table contains client data.	DB2 Bank.. / BANK3	
BANK_CUSTOMERS Table	Banking customers	AI This database table contains custome...	DB2 Bank.. / BANK3	
CLIENTS Table	Clients		DB2 Bank.. / BANK3	
CREDIT_CHECK Table	Credit score check		DB2 Bank.. / BANK3	Credit score +2 more
MORTGAGE_APP Table	Mortgage applications	Mortgage application data and data from mo...	DB2 Bank.. / BANK3	
SAVINGS_ACCOUNTS Table	AI Saving accounts data	AI Savings accounts transactional data...	DB2 Bank.. / BANK3	
SYSTEMS Table		AI Table containing system records.	DB2 Bank.. / BANK3	AI 3 Sugges
SYSTEMS_2 Table		AI Table containing system records.	DB2 Bank.. / BANK3	AI 2 Sugges

Что предлагает IBM

Платформа для
ИИ и данных

watsonx

Масштабируйте и
ускоряйте
внедрение ИИ с
помощью
надежных данных.

watsonx.ai

Обучайте, проверяйте,
настраивайте и
развертывайте модели
ИИ

Корпоративная студия нового поколения для разработчиков ИИ для обучения, проверки, настройки и развертывания как традиционного машинного обучения, так и новых генеративных возможностей ИИ на основе базовых моделей. Она позволяет вам создавать приложения ИИ за ограниченное время на ограниченных данных.

watsonx.data

Масштабируйте рабочие
нагрузки ИИ на все ваши
данные, находящиеся где
угодно

Целевое хранилище данных,
созданное на основе
открытой архитектуры
Lakehouse, с поддержкой
запросов, управления и
открытых форматов данных
для доступа к данным и их
совместного использования.

watsonx.governanc

e

Контроль
ответственности,
прозрачности и
объяснимости рабочих
процессов ИИ
Комплексный инструментарий,
охватывающий как управление
данными, так и ИИ,
позволяющий реализовать
контроль ответственности,
прозрачности и объяснимости
рабочих процессов ИИ

Архитектура стека generative AI



AI ассистенты

watsonx Assistant
watsonx Code Assistant
watsonx Discovery
watsonx Orchestrate
+++

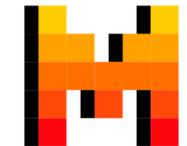
Gen AI платформа

watsonx.ai
watsonx.governance
watsonx.data



Базовые модели

granite-7b
granite-13b
granite-20b
more ...



Инфраструктура



HYBRID CLOUD



Watsonx.ai

инструментарий GenAI

Prompt Engineering

- Изучение и развитие Generative AI навыков
- Исследование разных моделей и параметров
- Обучение на экспериментах и примерах в виде prompt взаимодействия

Chat UI

- Интерактивное чат взаимодействие с большими языковыми моделями (LLM)
- Sample Chat инструкции для корректного результата
- Чат для адаптации RAG сценария

Code Samples

- Генерация и перевод примеров кода
- Примеры кода на Curl, Node.js и Python
- Сохранение результатов в виде шаблонов, сессий, или в виде Jupyter Notebook

Prompt/Fine-tuning

- Использование размеченных данных LLM для поддержки специализированных задач
- Развитие моделей с помощью индустриальных/бизнес данных и информации

InstructLab

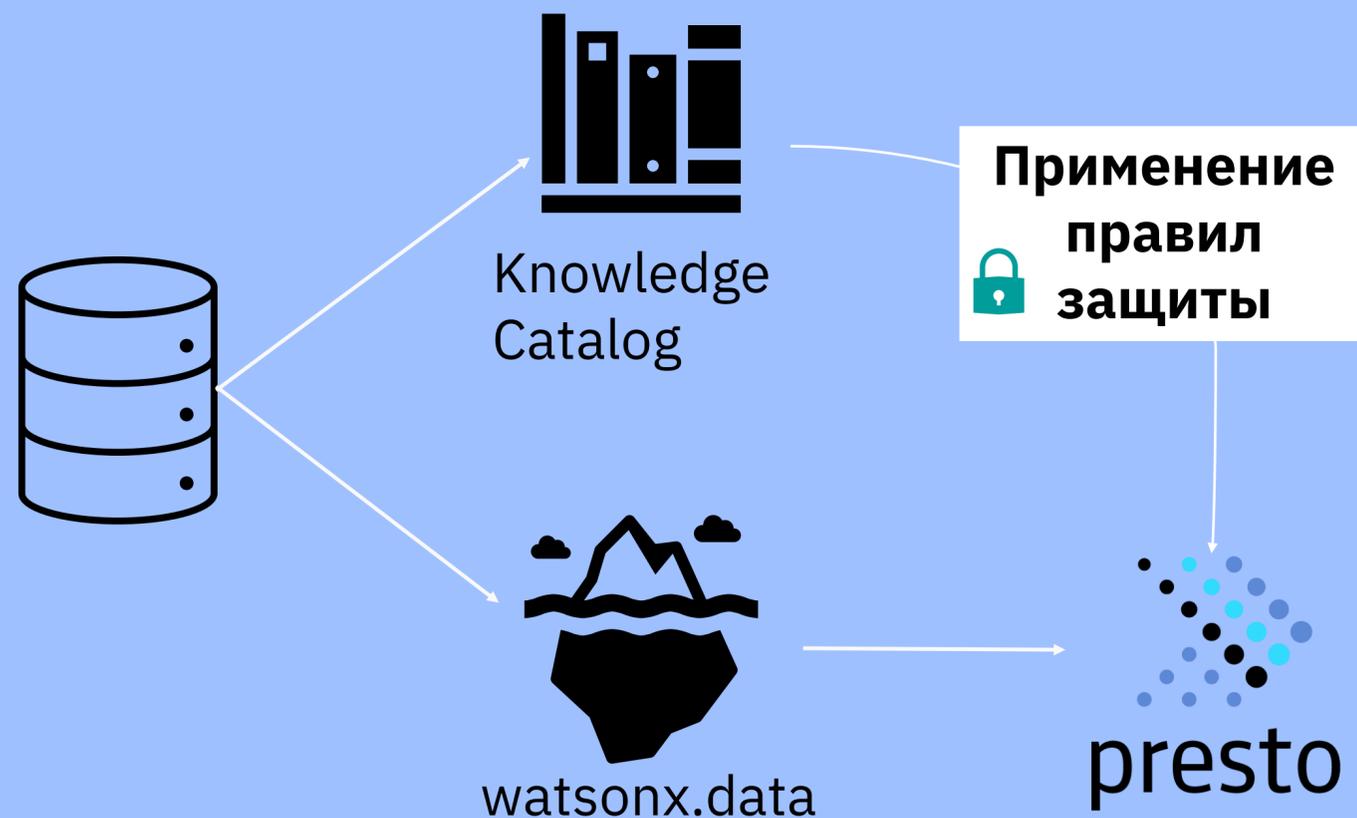
- Возможность сообществу экспертов сотрудничать и совместно развивать модели LLM
- Развитие модели LLM и кейсов ее применения
- Деплоймент адаптированной модели LLM для решения бизнес задач

Synthetic Data

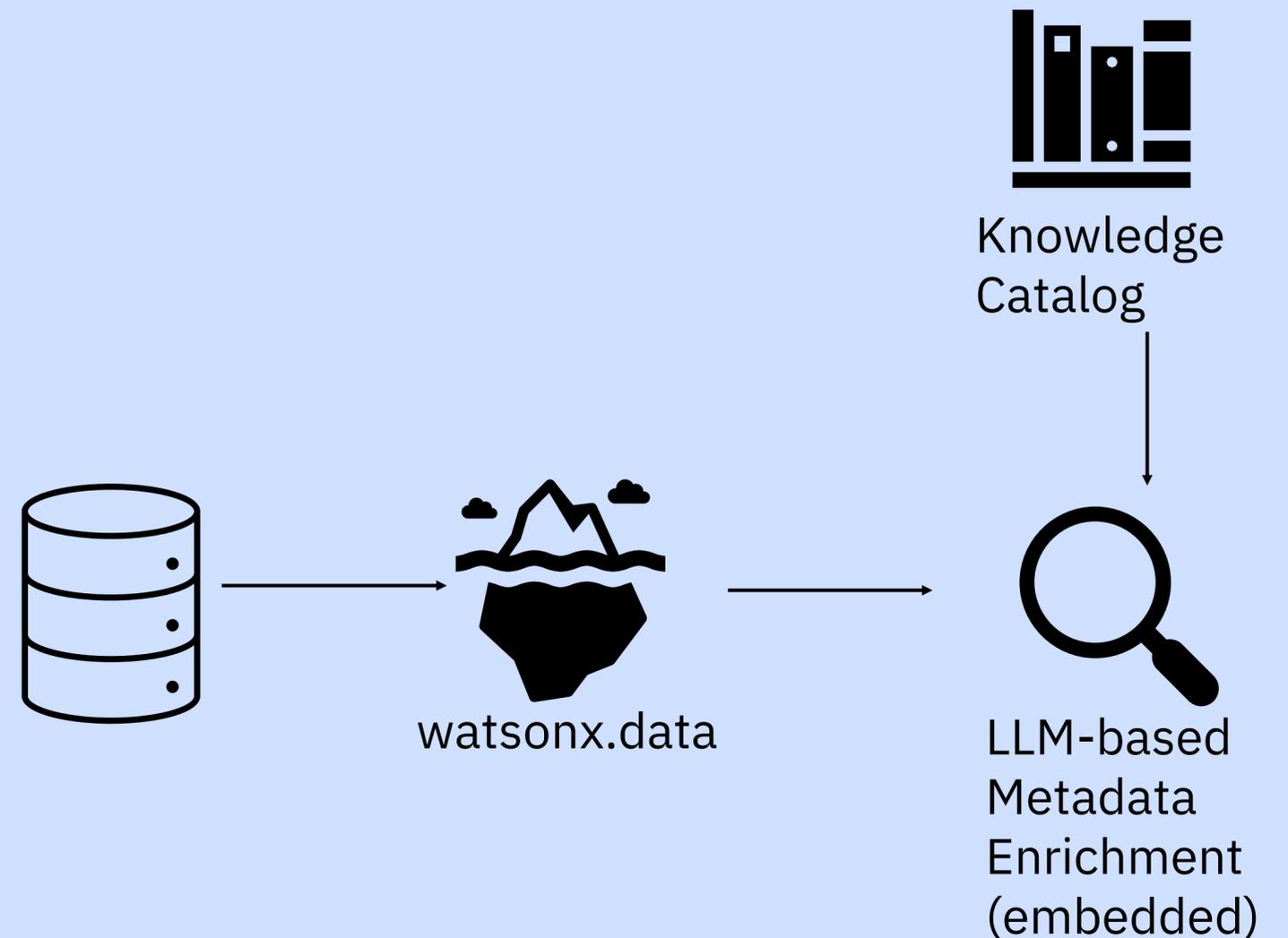
- Синтезирование данных для тестирования, развития приложений ИИ
- Синтезирование по схемам и правилам распределения
- Синтезирование на основе примера данных
- Приватность и безопасность

Интеграция Knowledge Catalog и watsonx.data

Вариант #1 – Защита конфиденциальных данных с помощью Knowledge Catalog Data Protection Rule в watsonx.data Presto Engine для унификации управления доступом

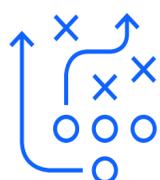


Вариант #2 – Обогащение метаданных через Knowledge Catalog в watsonx.data с целью автоматизированного каталогизирования и классификации



Три способа изучения компонентов платформы **watsonx** сегодня

IBM открыта к инвестированию в совместное партнерство



ЗАПРОСИТЕ ДЕМО

Опробуйте watsonx и ознакомьтесь с основными возможностями с помощью бесплатной демоверсии

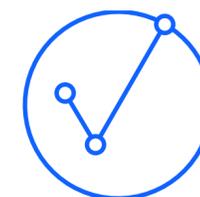
Доступно в любой момент



СТРАТЕГИЧЕСКАЯ СЕССИЯ

Обсуждение и индивидуальная демонстрация точки зрения и возможностей генеративного ИИ IBM watsonx. Оцените применимость watsonx в вашей стратегии проектов ИИ.

2-4 часа



ПИЛОТИРОВАНИЕ

Совместное с инженерами IBM глубокое тестирование возможностей **watsonx**. Оценка значения **watsonx** для выбранных кейсов, создание планов внедрения и развития ИИ в Вашей организации.

1-4 недели

IBM