

Интеграция Business Intelligence и Artificial Intelligence

Владимир Некрасов



info@contourcomponents.com



contourcomponents.com



Что такое Business Intelligence(BI)?

Многомерный анализ и богатая визуализация

Определение: Набор инструментов и процессов для сбора, анализа и визуализации данных

История:

- 1960-е: Ранние системы поддержки принятия решений
- 1990-е: Появляются хранилища данных и OLAP
- 2000-е-сейчас: BI, облачные BI

Использование: Отчеты (например, многомерный статистический показатель), информационные панели (например, демографические показатели), планирование политики на основе данных



Что такое Artificial Intelligence(AI)?

Компьютер как очень умный человек

Определение: Технология, которая имитирует человеческий интеллект с использованием алгоритмов

История:

- 1950-е годы: тест Тьюринга и ранние концепции ИИ
- 1980-е годы: экспертные системы для разных задач, основанных на правилах
- 2010-е годы — настоящее время: Глубокое обучение и большие данных

Применение: Предиктивное моделирование (например, экономическое прогнозирование), NLP (Natural Language Processing – обработка естественного языка) для запросов данных, автоматизация статистических задач



Применение BI в статистике

Высокоэффективная платформа для статистического производства

- ❑ Сбор и хранение данных
- ❑ Валидация и гармонизация данных
- ❑ Публикация и распространение
- ❑ Многомерный анализ и богатая визуализация



Применение BI в Статкомитете СНГ

Огромный международный статистический хаб

- ❑ Коннекторы с десятками международных источников данных
- ❑ Большое Хранилище данных
- ❑ Многоязычный статистический портал
- ❑ Сотни многомерных информационных панелей
- ❑ Каталог открытых данных



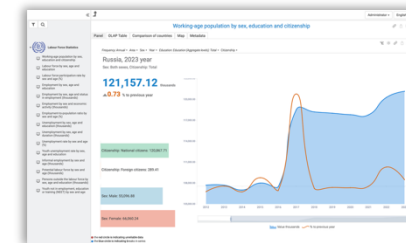
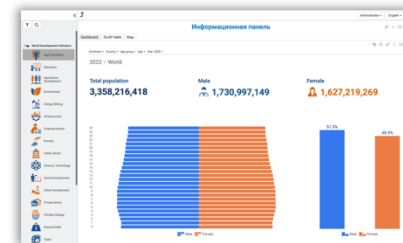
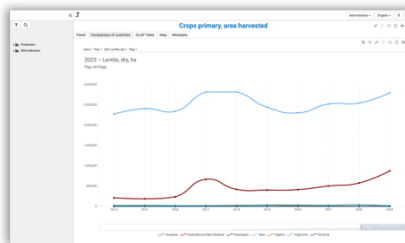
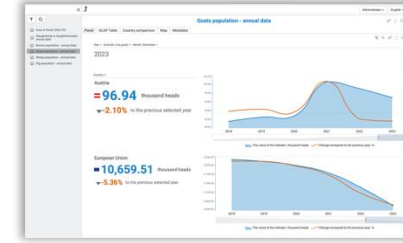
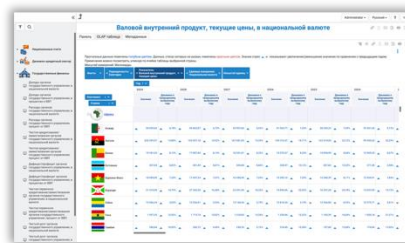
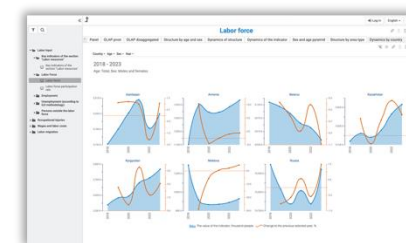
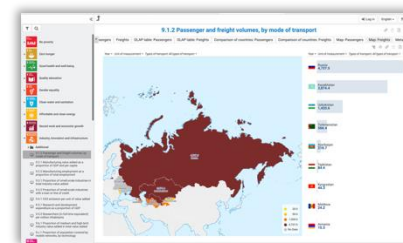
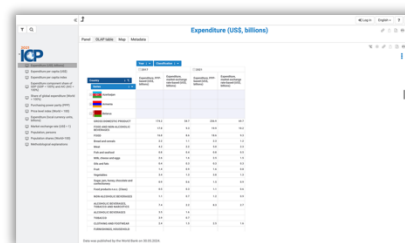
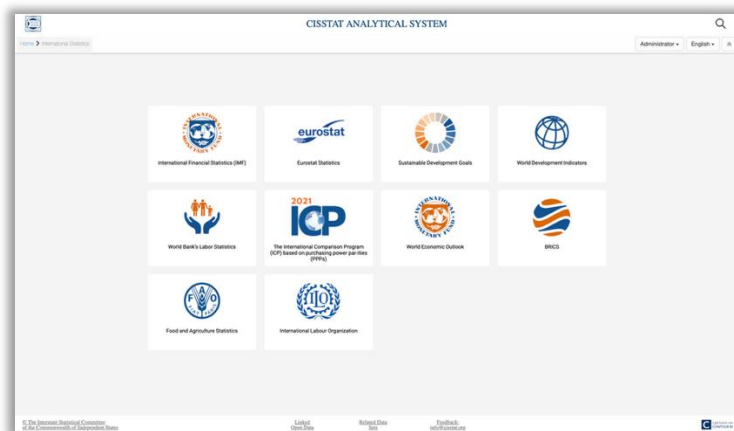
info@contourcomponents.com



contourcomponents.com



Аналитическая система Статкомитета СНГ на основе BI



info@contourcomponents.com



contourcomponents.com

BI + AI = BI-AI

BI

- Хранилище данных
- OLAP
- Визуализация



AI

- NLP
- Модель
- Текстовый и голосовой запрос



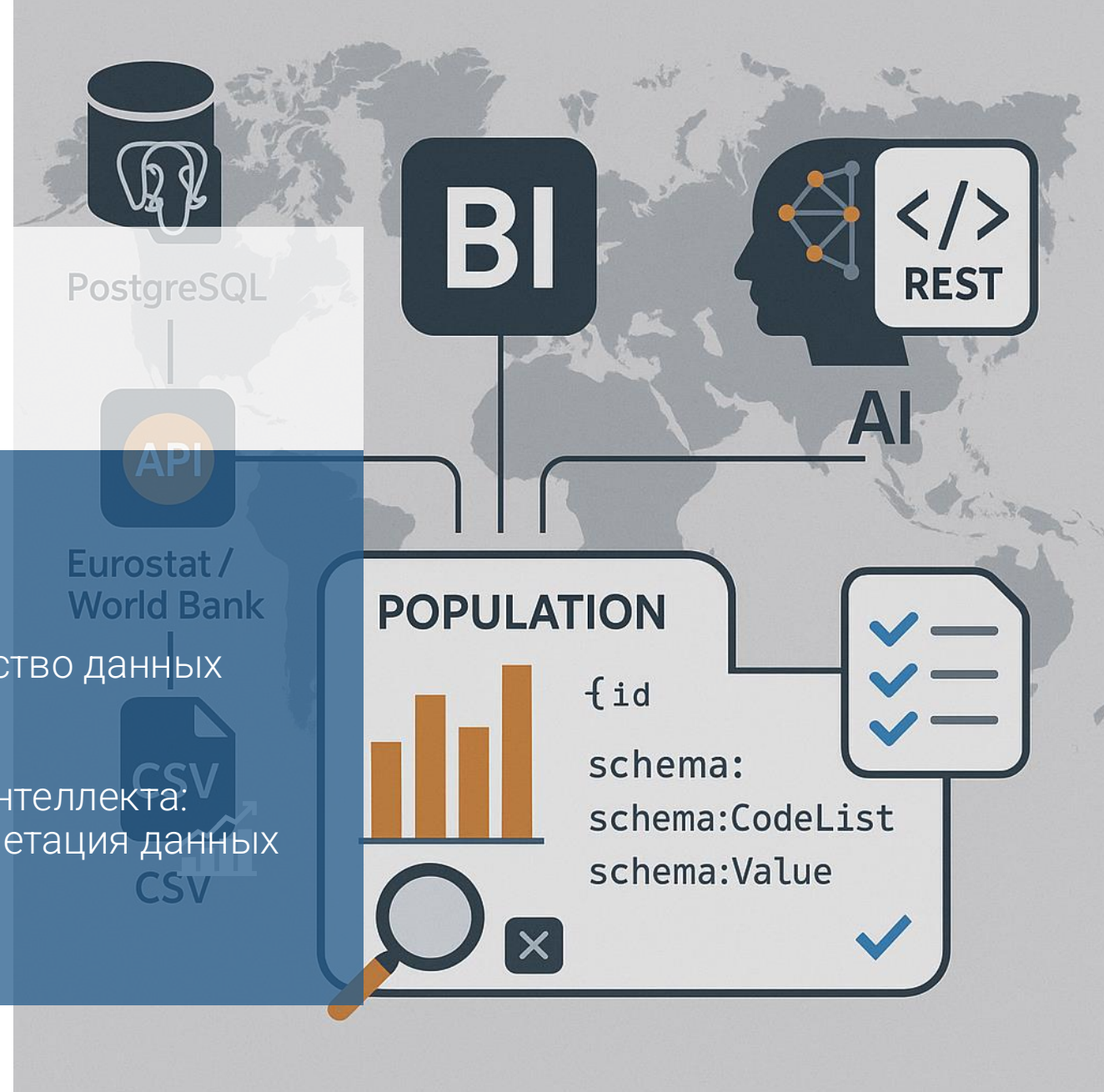
BI-AI

- Что-то совершенно новое



Применение BI-AI

- ❑ Коннекторы для источников данных
- ❑ Извлечение и гармонизация метаданных
- ❑ Информационное моделирование и качество данных
- ❑ Генерация визуализаций
- ❑ Расширенные функции искусственного интеллекта: запросы на естественном языке, интерпретация данных



Создание коннекторов к источникам данных

Разработка коннекторов к различным источникам статистики:

- ❑ Анализ API, документации и форматов источников данных
- ❑ Поддержка SDMX, API, SQL, CSV и веб-скрейпинга
- ❑ Получение и понимание описаний структуры данных (DSD)
- ❑ Расширение модели хранилища данных
- ❑ Импорт метаданных и данных в хранилище данных первоначальный и регулярный



Пример: Коннектор для базы данных ОЭСР

Коннектор
разработанный AI

- Язык: JavaScript
- Метод:
итеративная
серия заданий

</> Oecd Sdmx Exporter

// OECD SDMX Exporter - Fixed for null-safe codelist export

```
const fs = require('fs').promises;
const path = require('path');
const fetch = require('node-fetch');
const xml2js = require('xml2js');
```

```
const BASE_DATA_URL = 'https://stats.oecd.org/SDMX-JSON';
const BASE_STRUCTURE_URL = 'https://stats.oecd.org/restsdmx/sdmx.ashx';
const files_path = '/Users/vladimir/Documents/Doc/Contour BI 8/Connectors/OECD/Data';
const dataset = 'DP_LIVE';
const timeRange = { start: '2020', end: '2021' };
```

```
async function saveFile(filename, content) {
  const filePath = path.join(files_path, filename);
  await fs.mkdir(path.dirname(filePath), { recursive: true });
  await fs.writeFile(filePath, content, 'utf8');
}
```

```
async function saveCSV(filename, rows) {
  const csv = rows.map(row => row.map(v => `"${v}"`).join(',')).join('\n');
  await saveFile(filename, csv);
}
```

```
async function saveJSON(filename, jsonData) {
```



Гармонизация и очистка метаданных и данных

- ❑ Стандартизация терминологии и концепций
- ❑ Сопоставление локальных и глобальных классификаций на основе переходных ключей и ассоциаций
- ❑ Применение обнаружения и исправления ошибок



Сравнение AI и программиста

Программист

- 2 дня на версию
- Растущая сложность кода замедляет разработку

3 итерации в неделю

AI

- 2 секунды на версию
- Растущая сложность кода не влияет на скорость разработки

30 итераций в час



Функция очистки данных, созданная AI

```
DROP FUNCTION IF EXISTS tools.strip_html;
CREATE OR REPLACE FUNCTION tools.strip_html
(
    html_text TEXT,
    save_structure BOOLEAN DEFAULT true, -- save returns in Ansi text, ignored for HTML
    destination_format TEXT DEFAULT 'html'
)
RETURNS TEXT
LANGUAGE PLSQL
AS $$
DECLARE
    clean_text TEXT;
    destination_format TEXT;
BEGIN
    destination_format := TRIM(LOWER(destination_format));

    IF NOT destination_format = ANY ('text', 'html') THEN
        RAISE EXCEPTION 'Unknown destination format: %', destination_format;
    END IF;

    IF destination_format = 'html' THEN
        save_structure := true;
    END IF;

    -- Remove conditional comments and their content
    clean_text := regexp_replace(html_text, '<!--\[\[if\^\]>.*?<!\[\[endif\^\]>-->', '', 'gi');

    -- Remove XML-like tags and their content
    clean_text := regexp_replace(clean_text, '<xml\[\^\]>.*?</xml>', '', 'gi');

    -- Remove specific tags like <o:p> and their content
    clean_text := regexp_replace(clean_text, '<o:p\[\^\]>.*?</o:p>', '', 'gi');

    -- Remove <style> and <script> tags and their content
    clean_text := regexp_replace(clean_text, '<style\[\^\]>.*?</style>', '', 'gi');
    clean_text := regexp_replace(clean_text, '<script\[\^\]>.*?</script>', '', 'gi');
```

Сотни проверок и преобразований

Нечеловеческий стиль
сжатого
программирования

```
-- Remove <head> and <meta> tags and their content
clean_text := regexp_replace(clean_text, '<head\[\^\]>.*?</head>', '', 'gi');
clean_text := regexp_replace(clean_text, '<meta\[\^\]>.*?</meta>', '', 'gi');

-- If any { ... } block appears, it's most likely CSS, remove cautiously
clean_text := regexp_replace(clean_text, '{\[\^\]>.*?}', '', 'gi');

-- Replace HTML entities (like &nbsp;)
clean_text := regexp_replace(clean_text, '&nbsp;', ' ', 'gi');

-- Handle the rest based on the desired output format
IF destination_format = 'text' THEN
    IF save_structure THEN
        -- Add a newline after </o:p></font></p> and </div>
        clean_text := regexp_replace(clean_text, '</o:p></font></p>', '\n', 'gi');
        clean_text := regexp_replace(clean_text, '</div>', '\n', 'gi');

        -- Replace <p> and <br> tags with newline for structure preservation
        clean_text := regexp_replace(clean_text, '<\/p\[\^\]>.*?', '\n', 'gi');
        clean_text := regexp_replace(clean_text, '<br\[\^\]>.*?', '\n', 'gi');
        -- Replace <div> and </div> tags with newline for structure preservation
        clean_text := regexp_replace(clean_text, '<\/div\[\^\]>.*?', '\n', 'gi');

        -- Remove all other HTML tags while preserving the inner text content
        clean_text := regexp_replace(clean_text, '<[\^\]>+>', '', 'g');

        -- Adjust the regex to collapse multiple newlines into one newline, not two
        clean_text := regexp_replace(clean_text, '\\n\\s*{2,}', '\n', 'g');

        -- Trim leading and trailing whitespace and newlines
        clean_text := trim(both E'\n' FROM clean_text);
    ELSE
        -- For ANSI format without structure, replace </p> and </b> with a space
        clean_text := regexp_replace(clean_text, '<\/(p|b)>', ' ', 'gi'); -- Replace </p> and </b> with a space
        clean_text := regexp_replace(clean_text, '<[\^\]>+>', '', 'g'); -- Remove all other HTML tags
        clean_text := regexp_replace(clean_text, '\\s+', ' ', 'g'); -- Collapse multiple spaces
        clean_text := trim(both ' ' FROM clean_text); -- Trim leading and trailing spaces
    END IF;
ELSE
    -- For HTML format, replace <br> with <br /> and <p> with <p>
    clean_text := regexp_replace(clean_text, '<br>', '<br />', 'gi');
    clean_text := regexp_replace(clean_text, '<p>', '<p>', 'gi');
```

Комментарий к
каждому шагу



Очистка статистических метаданных

Ошибки в справочниках

- ☐ Пропущенные коды
- ☐ Пропущенные значения
- ☐ Неправильные значения
- ☐ Дубликаты
- ☐ Неправильные типы данных...



Проверка списка кодов

Справочник

Code	Value
1	
2	Fish
	Meat
4	Male
5	Vegetable
01/01/2022	
7	\$123

Процедура, созданная AI



Отчет об ошибках

Code	Value	Error Description
1	(empty)	Missing value
2	Fish	No error
(empty)	Meat	Missing code
4	Male	No error
5	Juice	No error
01/01/2022	Bread	Invalid code format (date instead of a number)
7	\$123	Invalid format (contains special character "\$")

Преимущества программирования на AI по сравнению с запросами к облачному AI

- ❑ Бесплатно: Мы не используем ChatGPT для обработки данных
- ❑ Скорость: Мы используем собственные серверы и базы данных
- ❑ Удобство: Созданная функция встроена в приложение



Создание визуализаций и информационных панелей

- ☐ Подключение к хранилищу данных
- ☐ Настройка словаря данных
- ☐ Генерация запросов
- ☐ Настройка OLAP-куба
- ☐ Выбор и настройка лучших диаграмм, карт, таблиц



Рутинные операции – передача опыта аналитиков
Творчество – сотрудничество с человеком

Рутинна:

- Именованние полей
- Настройка запросов
- Алгоритмы расчетов
- ...

Творчество:

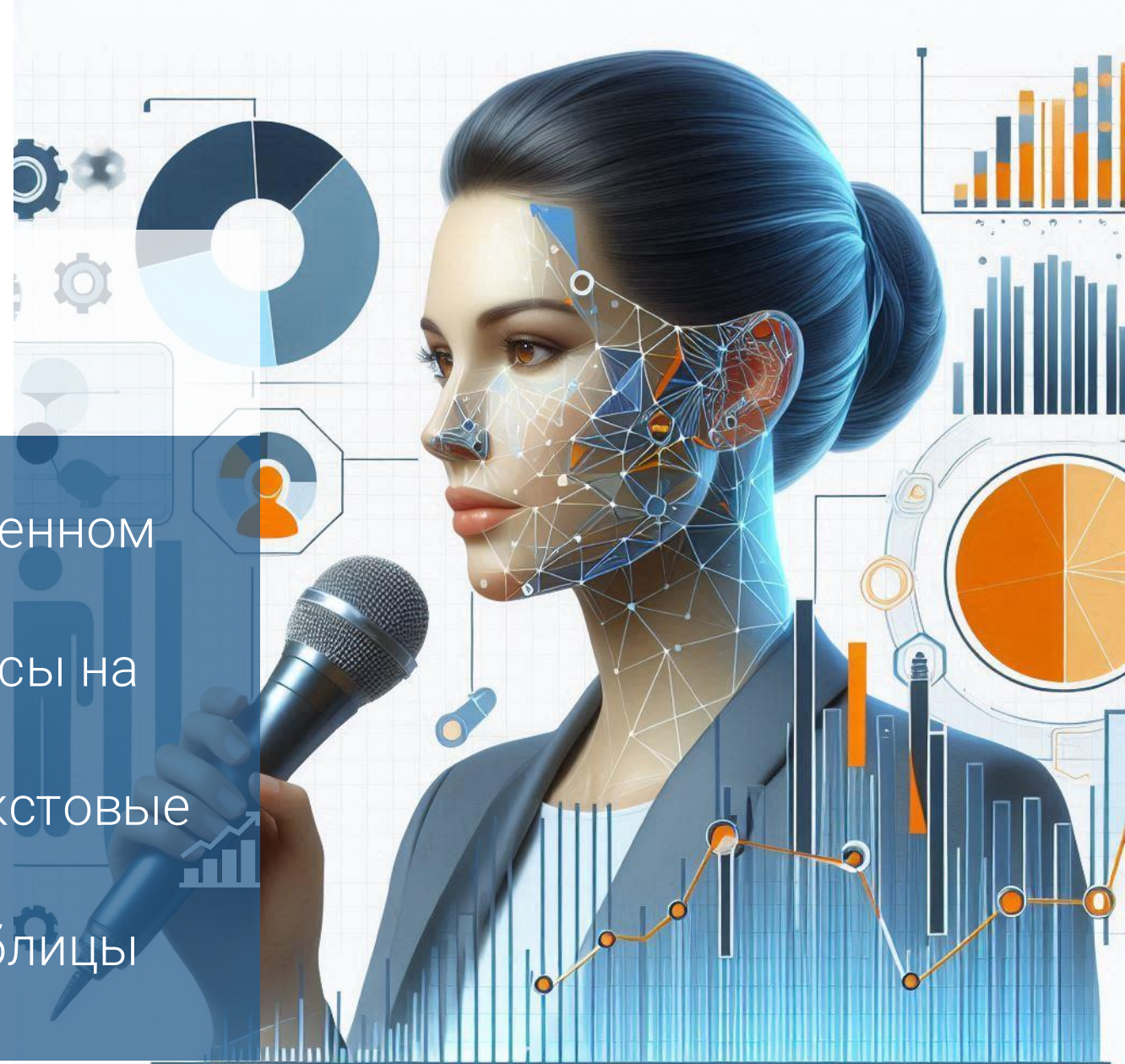
- Выбор цветов
- Добавление картинок
- Выбор типов графиков
- Структура дашборда
- ...

[illegible]

Интерфейс текстовых и голосовых запросов

Выполнение запросов на естественном языке:

- ❑ Текстовые и голосовые запросы на данные
- ❑ Мгновенные визуальные и текстовые ответы
- ❑ Одно точное число вместо таблицы



Предоставление интерпретаций и прогнозов

- ❑ Текстовые обзоры, созданные AI
- ❑ Оценка значений и структуры статистических показателей
- ❑ Объяснения тенденций и создание прогнозов



Заключение

- ❑ VI, интегрированный с AI, преобразует официальную статистику, сделав ее более точной и качественной
- ❑ AI, интегрированный в VI, даст пользователю новый, более естественный способ доступа к данным
- ❑ AI сможет помочь лучше понимать данные



Thank you!



info@contourcomponents.com



contourcomponents.com