

# Применение искусственного интеллекта в переписи населения

Объединение современных технологий и ресурсов в цикле GSBPM



Владимир Некрасов, ООО «Контур Компонентс»

# Модернизация переписи населения

Модернизация всех областей жизни касается и переписей населения

- Одновременное применение всех современных информационных технологий:
  - BI
  - Big Data
  - скоростной интернет с полным покрытием страны
  - гаджеты у населения
- Искусственный интеллект как одна из новых информационных технологий
- Взрывной рост данных – извлечение демографических данных из административных данных



# Покрытие всех этапов GSBPM с ИИ

Подготовка	Сбор	Интервью	Обработка	Публикация
Проверка адресов Прогноз нагрузки Логистика	Проведение переписи Сбор данных	Диалог с респондентом	Агрегация Анализ	Распространение Публикация
				
BI - Хранение - Визуализация данных ИИ - Геоанализ - Планирован - Оптимизация ресурсов	BI - Сбор данных ИИ Извлечение демографических данных из админданных – регистров, миграционного учета, соцсетей	BI - Сбор анкет ИИ - Диалог - Выявление ошибок online -- подсказки	BI - Агрегация - Расчеты (OLAP) ИИ - Аномалии - Дубликаты - Импутация	BI - Визуализация - Микроданные ИИ - Деперсонификация - Запросы на естественном языке

# Этап подготовки

## ИИ

- Проверка адресов, анализ спутниковых снимков, расчет координат по адресу
- Прогнозирование нагрузки, расчет потребности в персонале, оборудовании, офисах
- Оптимизация переписных участков и маршрутов переписчиков

## BI

- Дашборды с картами
- Интегрированные ИИ-советники



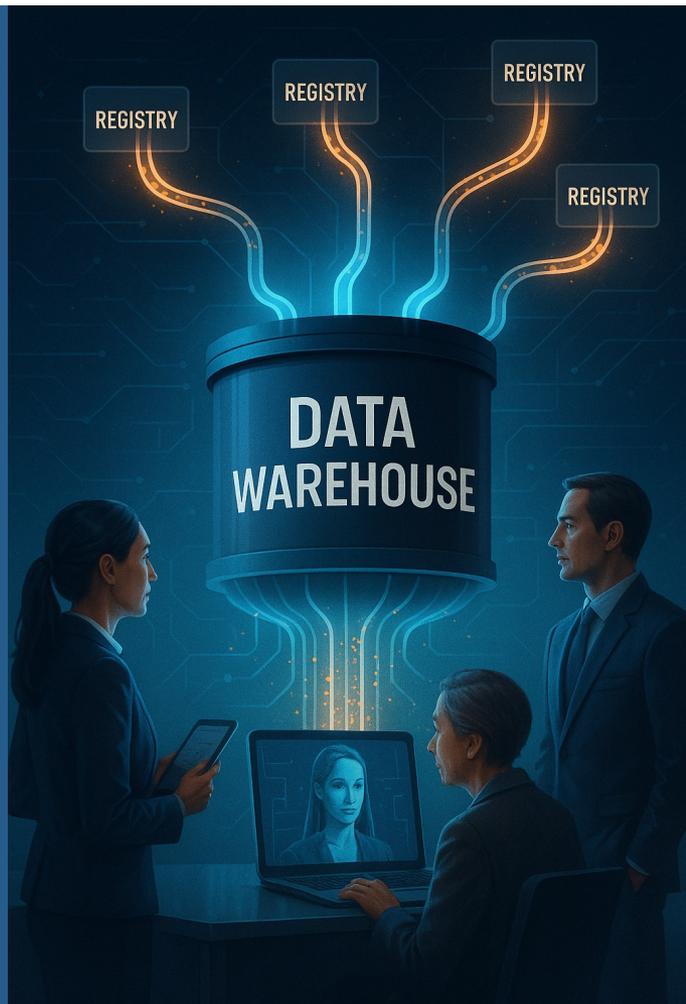
# Этап сбора данных – админданные

ВІ

- Сбор административных данных:
  - Налоговые реестры
  - Миграционный учет
  - Социальная помощь
- Хранилище данных

ИИ

- Извлечение демографических данных из административных данных
- Наполнение хранилища данных переписи



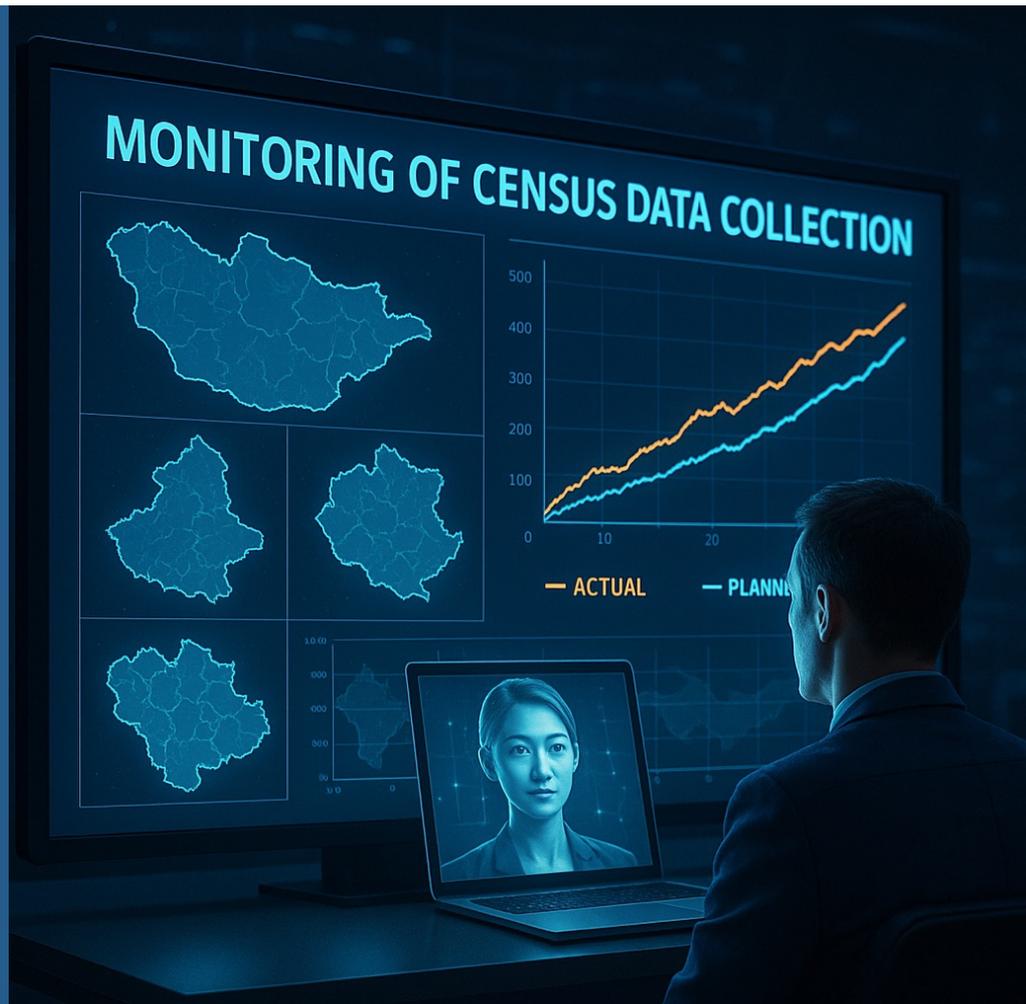
# Этап сбора данных – мониторинг

BI

- Сбор и хранение результатов
- Дашборды по стране и региону
- Прогресс в процентах

ИИ

- Раннее обнаружение регионов со слабым откликом
- Наполнение хранилища данных переписи



# Стадия интервью

- Чат-бот – искусственный интервьюер
- Голос, текст, жесты
- Многоязычный, с диалектами
- Кастомизированный образ для респондента, вызывающий у него доверие
- Обнаружение ошибок, предвзятости, коррекция
- Доступность на компьютерах, планшетах и смартфонах респондента



# Стадия обработки данных

## AI

- Валидация и очистка данных
- Выявление дублей и мусора
- Выявление аномалий – выбросов и несоответствий в разных ответах
- Перекрестное сравнение с реестрами и админданными
- Геокодирование адресов
- Импутация пропусков

## BI

- Агрегация и интерактивный анализ OLAP
- Хранение исправленных и расчетных данных в хранилище данных
- Дашборды для мониторинга качества данных



# Стадия распространения

## BI

- Публикации в фиксированной форме: брошюры, PDF
- Интерактивные дашборды (OLAP)
- Готовые наборы данных (каталог открытых данных)
- Запросы к микроданным

## AI

- Запросы на естественном языке
- Защита деперсонализации в BI инструментах
- Поиск наборов данных в каталоге открытых данных



# Дополнительные возможности

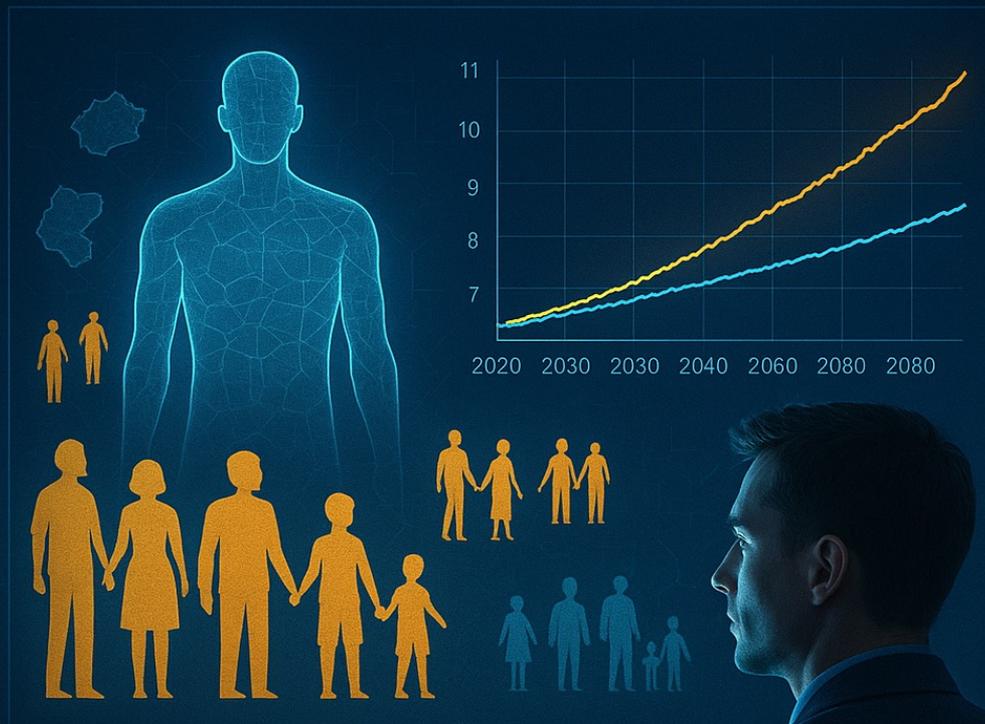
BI

Демографические расчеты на основе данных переписи и демографического и миграционного учета

AI

Демографические прогнозы: изменение численности населения, сценарии миграции и т.д.

## POPULATION FORECAST



# Потенциальные преимущества

- Эффективность: оптимизированная логистика
- Инклюзивность: многоязычные интервьюеры
- Качество: глубокая проверка, мониторинг в режиме реального времени
- Прозрачность: открытые данные, информационные панели, вопросы и ответы на естественном языке



Спасибо за внимание