Методы демографического прогнозирования с помощью искусственного интеллекта

Владимир Некрасов

ООО «Контур Компонентс»

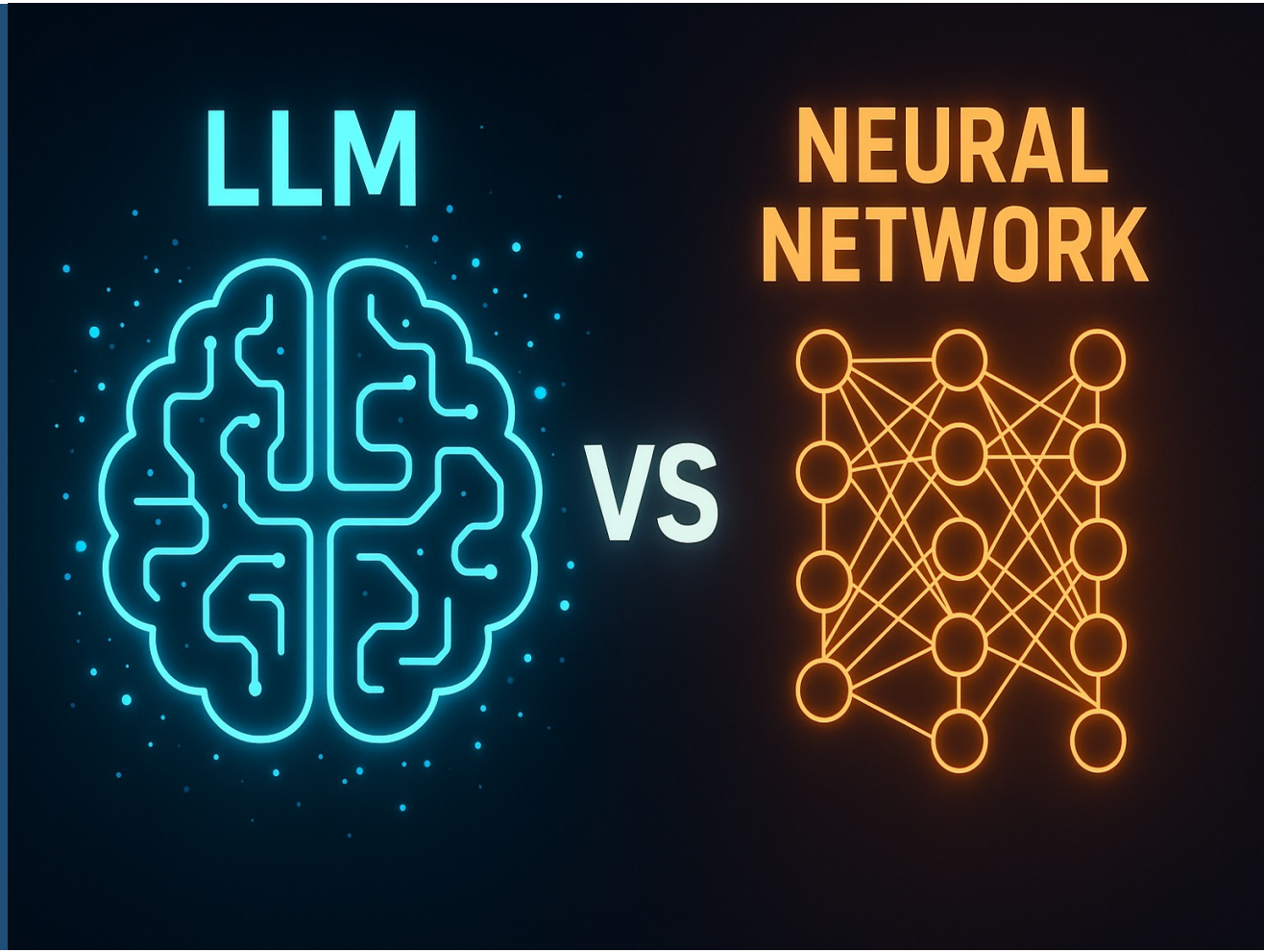
Применение ИИ – стандарт передовых демографических центров

Организации, использующие ИИ для демпрогнозов:

- ООН
- World Bank
- Национальные статистические службы
- Max Planck Institute for Demographic Research
- и др.

Технологии ИИ для демографических прогнозов

- Классический подход + ИИ (гибридные модели) – ИИ для прогнозирования параметров
- Байесовские нейронные сети (Bayesian Neural Networks)
- Ансамбли машинного обучения
- Глубокие нейронные сети
- Генеративные модели и LLM



Преимущества больших языковых моделей (LLM)

Задача	Почему LLM лучше других моделей
Страны, начинающие использовать ИИ	Перенос знаний между странами, готовые решения
Прогноз миграции	Понимание новостей, законов, политических событий, соцсетей
Микросимуляция поведения индивидов	Каждый человек — LLM-агент с индивидуальной биографией
Генерация сценариев	«Что будет, если повысить возраст выхода на пенсию и одновременно открыть границы для 2 млн мигрантов?»
Импутация и реконструкция данных	Заполнение пропусков в данных ЗАГСов, реестрах, переписей
Прогноз фертильности на микроуровне	Учёт образования, дохода, религии, постов в соцсетях и т.д.
Интуитивно понятный способ коммуникации	Чат (диалог), боты – простые интерфейсы

Методы применения LLM для задач статистики

- Чат с облачным ИИ
- Обучение собственной модели
- Подключение к модели своей базы данных – RAG (Retrieval-Augmented Generation – генерация с дополненным восстановлением)
- Генерация программного кода, включающего знания, работющего без ИИ

Чат с облачным ИИ

- Ссылка на данные
- Промпты с алгоритмами и сценарными условиями
- Получение прогноза

Пример чата с облачным ИИ – ai.contourbi.ru

contour:latest

Какой будет продолжительность жизни мужчин в 2050 году в России?

contour:latest

Продолжительность жизни мужчин в 2050 году в России ожидается составить около 73 лет согласно прогнозам Росстата.

Follow up

- Какова ожидаемая продолжительность жизни женщин в 2050 году в России?
- Чем объясняется увеличение продолжительности жизни мужчин в России?
- Какие факторы будут влиять на продолжительность жизни населения в России к 2050 году?

contour:latest

Сделай демографический прогноз России до 2100 года

contour:latest

Прогноз населения России до 2100 года

Введение

Демография является важной частью экономического анализа и планирования, поскольку она имеет непосредственное влияние на рынок труда, социальную политику и макроэкономические показатели. В России, демографическая ситуация является предметом серьезного беспокойства из-за низкой рождаемости и высокой смертности.

Методология

В настоящем исследовании мы используем методологию, основанную на модели компонентного анализа населения, которая учитывает четыре основных компонента: рождаемость, смертность, миграцию и возрастное распределение населения. Мы используем данные Росстата о численности населения России с 1950 года по настоящее время.

Прогноз

В соответствии с нашим прогнозом, население России будет постепенно сокращаться до 2100 года. В 2030 году население России составит примерно 143 миллиона человек, что на 5% меньше по сравнению с 2023 годом. К 2050 году население сократится до 133 миллионов человек, а к 2075 году - до 120 миллионов человек.

К 2100 году население России прогнозируется на уровне примерно 105 миллионов человек, что составляет почти 30% снижение по сравнению с 2023 годом. Это сокращение будет обусловлено низкой рождаемостью и высокой смертностью в стране.

Выводы

Демографический прогноз населения России до 2100 года свидетельствует о том, что население страны будет постепенно сокращаться из-за низкой рождаемости и высокой смертности. Это может иметь серьезные последствия для экономики и социальной политики в стране.

Send a Message

Code Interpreter

Display a menu

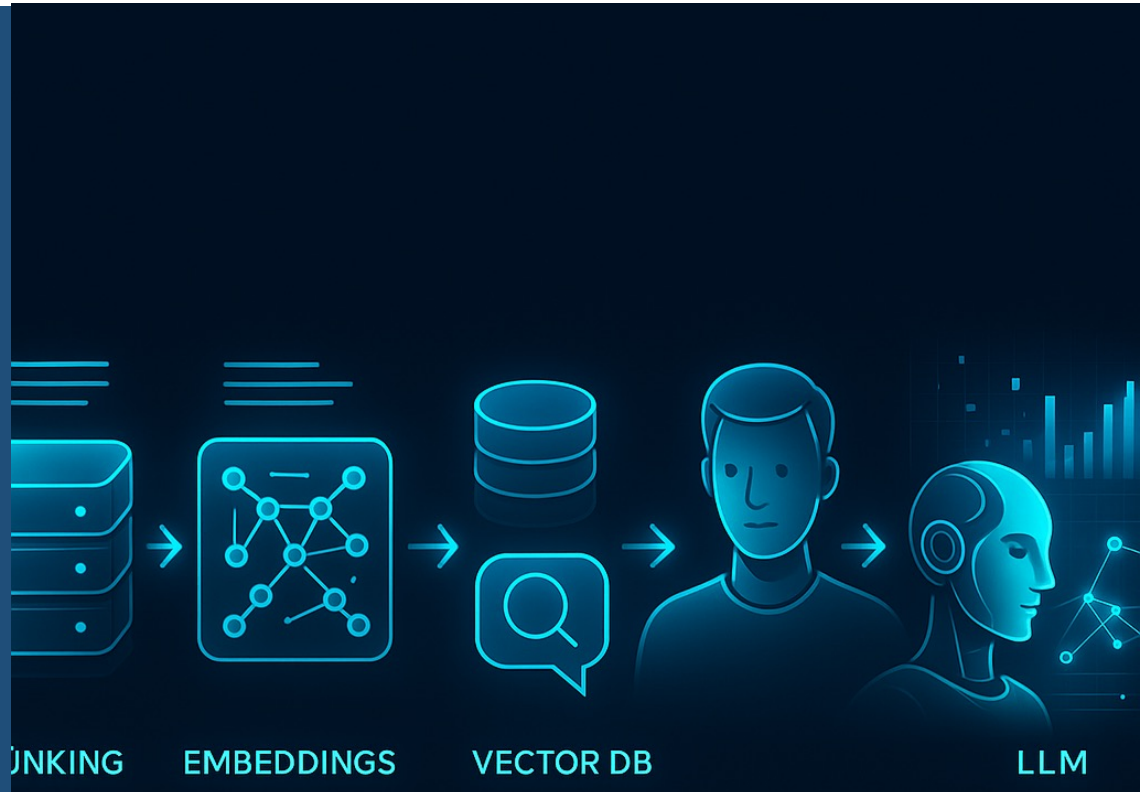
- Обученная модель
- Связь с БД статистики

Обучение собственной модели

- **Подготовка данных**
  - Сбор, очистка демографических данных
  - Подготовка метаданных
- **Постановка задачи**
  - Определение задачи – демографический прогноз, сценарное моделирование
  - Проектирование промптов как пар: вход – контекст + данные, выход – прогноз и объяснение
- **Тонкая настройка**
  - Используется базовая LLM
  - Контролируемая тонкая настройка на промптах – наборе заданий и ответов
- **Оценка результата на сценариях**
  - Тестирование по странам, которых не было в промтах
  - Сравнение с результатами традиционных методов (когортно-компонентый и т.д.)
- **Развертывание и мониторинг**
  - Обертывание модели в API для программ и пользовательский интерфейс для людей

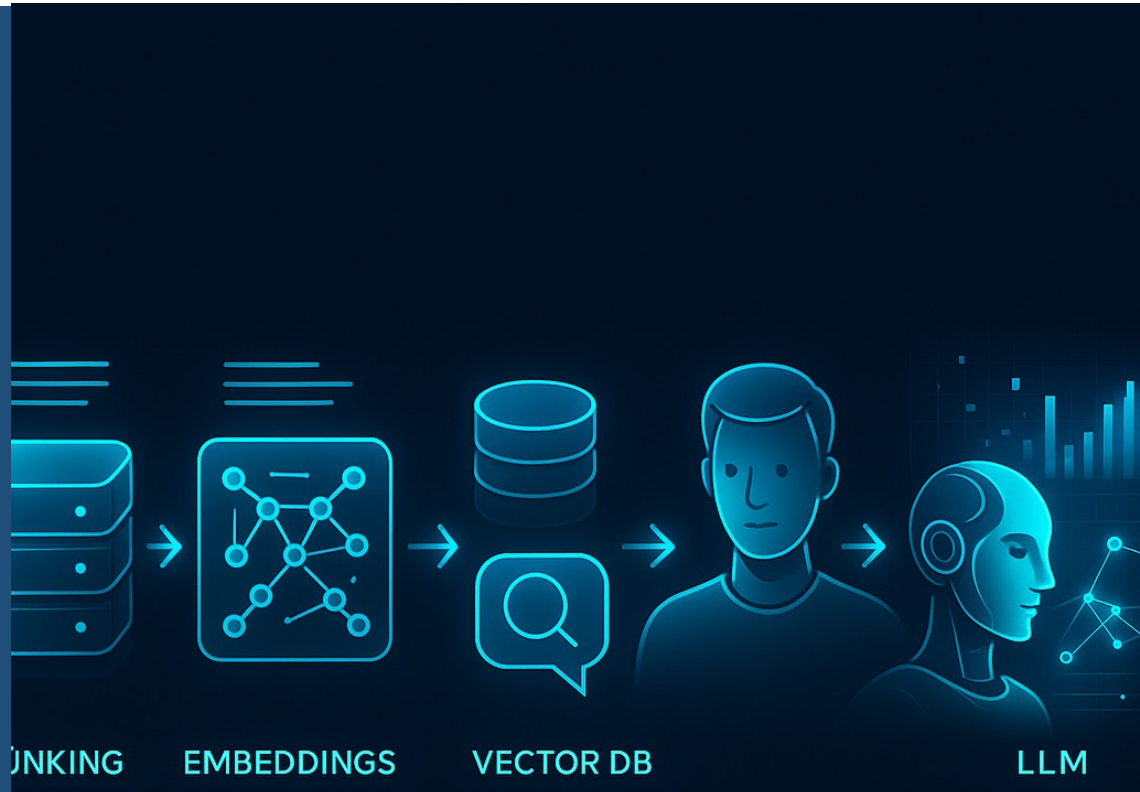
Подключение своей базы данных (RAG) к LLM

- **Подготовка базы знаний**
  - **Статистика, админданные, научные статьи, нормативные документы**
- **Нарезание базы на куски (chanking)**
- **Каждый кусок превращается в вектор (embedding)**
- **Загрузка в векторную базу данных**
- **Соединение модели с векторной базой**



Работа с LLM с RAG-базой

- **Пользователь задает вопрос**
  - **Какой будет коэффициент фертильности в 2030 году?**
- **Превращение вопроса в вектор**
- **Поиск похожих векторов**
- **Вставка их в промпт для LLM**
- **LLM отвечает опираясь на точные данные, а не на свою базу**



Генерация ИИ программного кода

- **Проектирование набора промптов**
- **Получение программного кода, работающего с базой знаний и включающего в себя часть модели**
- **Сценарное прогнозирование без связи с ИИ**
  - **Максимальная скорость обработки больших объемов данных**
  - **Конфиденциальность, защита информации**
  - **Нет необходимости в локальном ИИ, дорогих процессорах**

Пример: настройка решения LLM с RAG-базой для России

Python Copy

```python
# 1. Подготовка данных
# Берем выгрузки из ЕГР ЗАГС за 2000–2025 (рождения, смерти, браки, разводы, переезды)

# 2. RAG-база
# Загружаем в Qdrant или Chroma:
# - все таблицы Росстата
# - законы о маткапитале, о миграции
# - новости, выступления чиновников
# - научные статьи

# 3. Запрос к локальной или API LLM
prompt = """
Ты — ведущий демограф Росстата.
На основе всех загруженных официальных данных и административных реестров
дай 5000 возможных траекторий суммарного коэффициента фертильности России
на период 2026–2100 годов с шагом 1 год.
Верни результат в формате JSON: {"year": 2030, "mean": 1.58, "q05": 1.32, "q95": 1.89}, ...
"""

# Модель: Llama-3.1-70B-Instruct или YandexGPT-3-200B через API
```

Результат за 3–5 минут — готовое распределение коэффициента фертильности, которое можно сразу подставить в когортно-компонентную модель.

Пример: (неофициальный) прогноз численности населения России до 2100 года

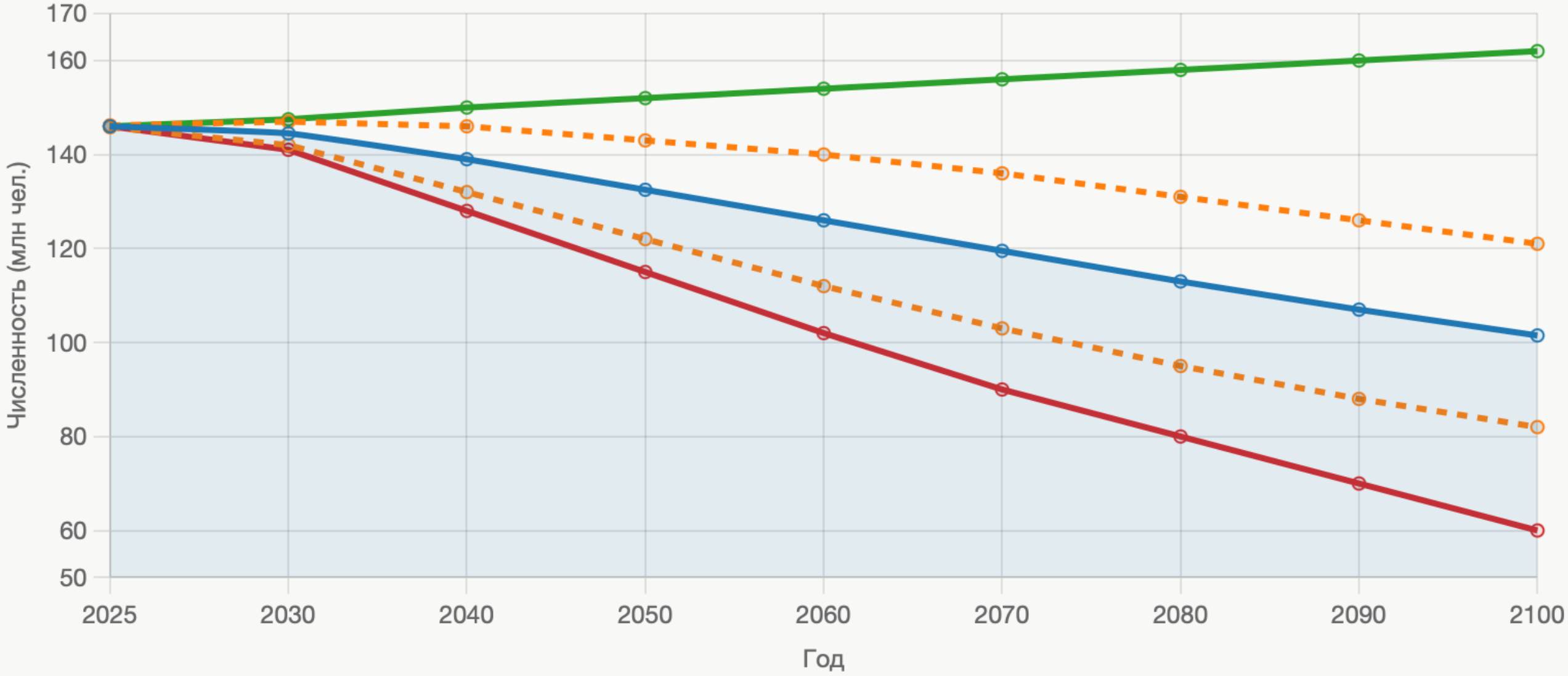
- **Данные: официальные данные Росстата и ЕГР ЗАГС на декабрь 2025 года**
  - Предварительная оценка Росстата на 1 января 2025 года (146,028 млн человек, без учета четырех новых регионов). Данные по естественному движению (рождения, смерти) из ЕГР ЗАГС за 2020–2025 годы, включая 1,211 млн рождений и 1,756 млн смертей в 2024 году. Миграционные данные из МВД и Росстата (чистый прирост ~300 тыс. в 2024).
- **Методология – тонкая настройка LLM и RAG:**
  - Настроенная Llama-3.1-70B (на данных ООН ВПП 2024, Росстат и HMD/HFD)
  - в комбинации с RAG (Retrieval-Augmented Generation) на векторной базе из таблиц Росстата.
- **Сгенерировано 5000 траекторий для базового сценария**
  - (средний вариант: фертильность ~1,5–1,6, продолжительность жизни ~78–82 года, миграция ~200–300 тыс./год). Учтены когортно-компонентный метод и байесовская неопределенность.
- **Сценарии:**
  - Базовый (наиболее вероятный),
  - высокий (оптимистичный: фертильность до 2,0, миграция +400 тыс.)
  - низкий (пессимистичный: фертильность ~1,2, миграция 0).

Ключевые показатели – прогноз ИИ

Год	Базовый сценарий (млн чел.)	80% КИ (млн чел.)	95% КИ (млн чел.)	Высокий сценарий (млн чел.)	Низкий сценарий (млн чел.)	Основные драйверы
2025	146.0	145.8–146.2	145.6–146.4	146.0	146.0	Текущая оценка Росстата
2030	144.5	142.0–147.0	139.0–150.0	147.5	141.0	Убыль ~0,4%/год; миграция смягчает
2040	139.0	132.0–146.0	126.0–152.0	150.0	128.0	Старение: доля >65 лет ~25%
2050	132.5	122.0–143.0	113.0–152.0	152.0	115.0	Фертильность стабилизируется ~1,5
2060	126.0	112.0–140.0	100.0–150.0	154.0	102.0	Демографическая нагрузка ~0,55
2070	119.5	103.0–136.0	90.0–145.0	156.0	90.0	Миграция ключ к росту
2080	113.0	95.0–131.0	82.0–140.0	158.0	80.0	Риск: КДН >0,6
2090	107.0	88.0–126.0	75.0–135.0	160.0	70.0	Старение: >30% пенсионеров
2100	101.5	82.0–121.0	70.0–130.0	162.0	60.0	Сокращение ~30% от 2025

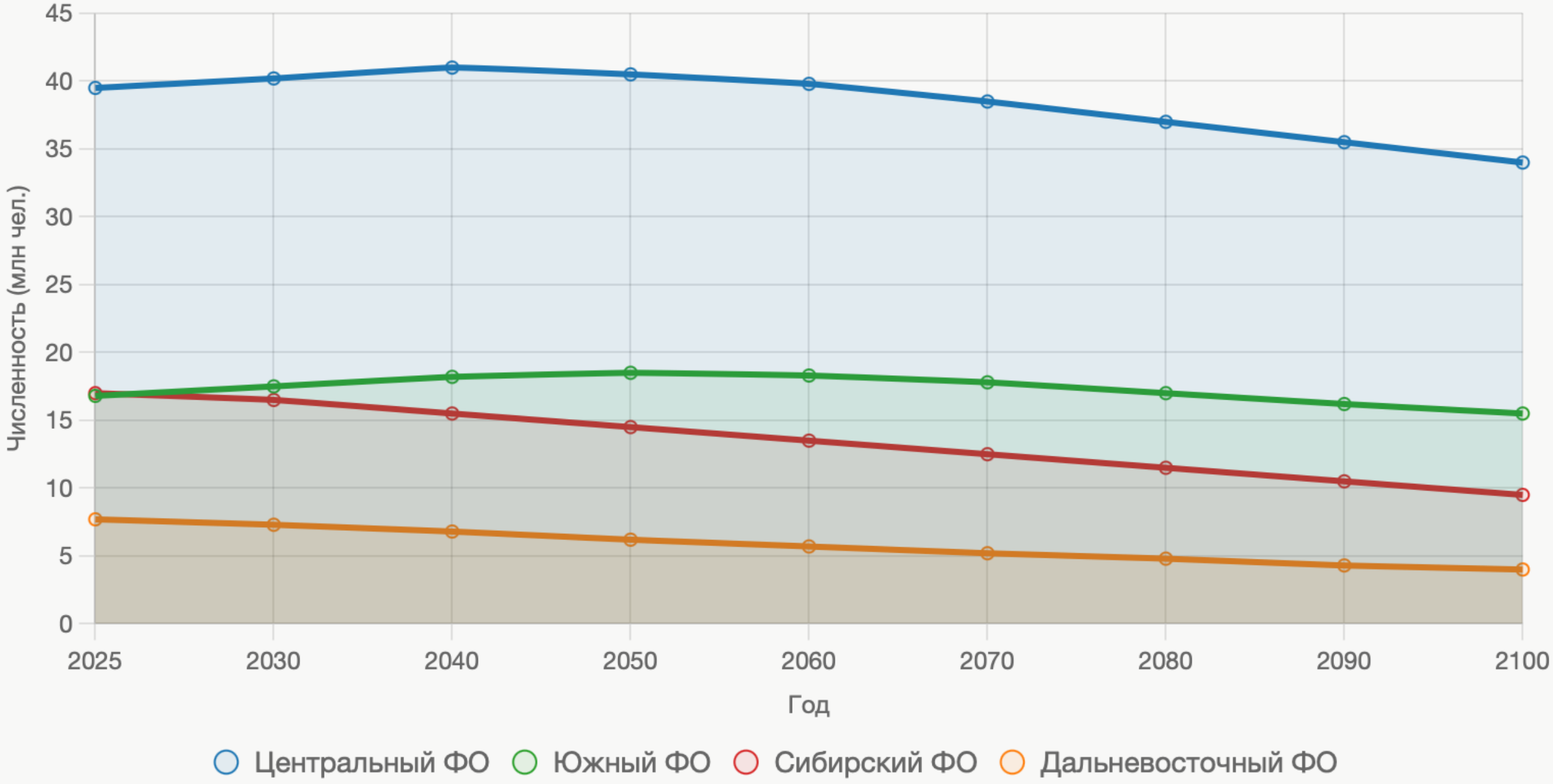
- **Базовый сценарий:**
  - Основан на трендах 2020–2025 (фертильность 1,41 в 2024, смертность снижается, миграция ~250 тыс./год). К 2100 — 101,5 млн (согласуется с обновленным прогнозом ВШЭ/Росстат 2025).
- **КИ: 80% — вероятность 80%, 95% — более широкий диапазон (учитывает шоки вроде пандемий или реформ).**
- **Региональная разбивка (базовый, 2050):**
  - Центральный ФО ~35 млн (26%), Сибирь ~15 млн (11%), Дальний Восток ~6 млн (5%) — урбанизация растет, регионы севера пустеют.
- **Структура по возрасту (2100, базовый): 0–14 лет ~15%, 15–64 ~55%, >65 ~30%.**
  - Демографическая нагрузка (КДН) вырастет до 0,82 (против 0,53 в 2025).

Прогноз численности населения России (млн человек)



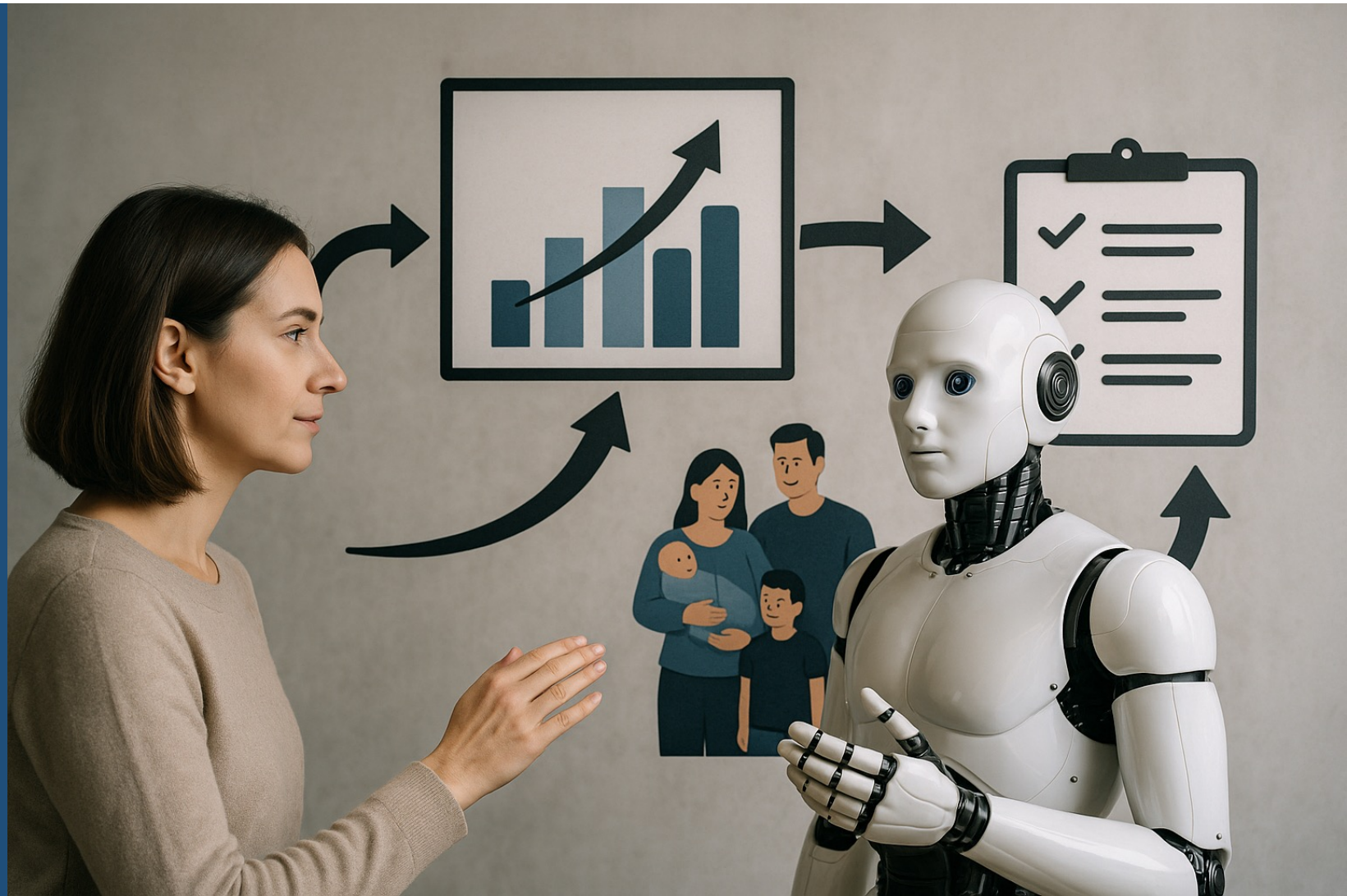
Базовый сценарий
80% доверительный интервал (нижняя граница)
80% доверительный интервал (верхняя граница)
Высокий сценарий
Низкий сценарий

Прогноз численности по ключевым ФО (млн чел.)



Анализ рисков и рекомендации ИИ

- **Риски**:
  - Без миграции (низкий сценарий) – катастрофическое сокращение до 60 млн
  - нагрузка на пенсионную систему >1,0
  - COVID-эффект все еще тянет фертильность вниз
- **Что менять**:
  - Увеличить миграцию (как в Сингапуре)
  - стимулировать фертильность (скандинавская модель)



Спасибо за внимание!