

# Предварительный план работ по проекту CityAir

---

## □ Этап 1. MVP — Minimum Viable Product (до 1 октября 2025)

### 1. Подготовка и обработка данных

- Формирование датасетов для lab и field (g1, g2).
- Интеграция эталонных значений (field reference, NO2, O3) в field-данные.
- Первичная очистка и визуализация (EDA — exploratory data analysis).
- Создание масок (baseline, stat).

### 2. Базовая модель восстановления концентраций по лабораторным данным

- 1 этап регрессии: построение зависимости Baseline от температуры и влажности ( $T, H \rightarrow \text{Baseline}(T, H)$ ).
- 2 этап регрессии: восстановление концентрации газа на стационарных участках (сигнал,  $T, H \rightarrow \text{концентрация газа}$ ).
- Анализ кросс-влияния газов и проверка мультиколлинеарности признаков.
- (При наличии кросс-влияния: расширенная регрессия — сигнал1, сигнал2,  $T, H \rightarrow \text{концентрация газа1, газа2}$ .)
- 2 этап регрессии на нестационарных участках (добавление производных в признаки).
- Составление запроса на дополнительные лабораторные эксперименты (диапазоны температур, влажности и т.д.).
- Тестирование 1 этапа ( $\text{Baseline}(T, H)$ ) на field-данных (участки без дрейфа).
- Тестирование 2 этапа на field-данных (по возможности).

### 3. Первая модель учёта деградации

- Анализ дрейфа (drift) сигнала по времени.
- Простая модель коррекции деградации (например, временной тренд).
- Проверка корректировки по field reference.

### 4. Анализ результатов MVP

- Анализ ошибок и ограничений текущего этапа.
- Составление списка необходимых улучшений.

---

## □ Этап 2. Итеративное улучшение моделей (октябрь 2025 – январь 2026)

### 1. Улучшение данных и разметки

- Запрос и интеграция новых данных (g3, g1, g2).
- Улучшение масок (baseline, stat).

### 2. Расширение регрессионной модели

- Повышение точности на всех этапах.

- Учет кросс-влияния газов на 2 этапе регрессии (многофакторная регрессия):
  - сигнал1, сигнал2, сигнал3, ..., T, H → концентрация газа1, газа2, газа3 и т.д.
- Тестирование сложных ML-моделей (CatBoost, XGB, нейросети) на обоих этапах регрессии.
- Улучшение регрессии для нестационарных данных по мере появления новых лабораторных экспериментов (разные диапазоны температур, влажности и т.д.).
- Итоговый отчёт по НИР (ГОСТ Р 7.32-2017).

### 3. Полноценный учёт деградации сенсоров

- Расширение признаков деградации (время, климатические параметры).
- Подбор и тестирование различных моделей деградации.
- Валидация моделей на field reference.

### 4. Валидирование и кросс-валидация моделей

- Кросс-валидация на разных постах.
- Оценка качества (RMSE, Pearson, t-Студент, относительная неопределённость/погрешность).

### 5. Интеграция с API CityAir

- Работа по интеграции идёт параллельно с вышеуказанными задачами.

---

## □ Этап 3. Программная реализация (январь 2026)

- Финальная реализация пайплайна коррекции концентраций.
- Финальная реализация коррекции деградации.
- Интеграция с API CityAir.
- Подготовка документов для регистрации РИД (ФИПС).
- Подготовка комплекта программной документации.

---

## ✓ Уже сделано (на текущий момент)

- Разработана начальная структура данных и пайплайна.
  - Проведен базовый EDA и обработка лабораторных данных.
  - Выполнена первая разметка (baseline, stat).
  - Построены первичные модели регрессии сигналов (пока только для сигналов, не концентраций).
  - Интегрированы эталонные значения (field reference, NO2, O3) в field-данные.
  - Получены минимальные метрики качества ( $R^2$ , RMSE).
  - Сформирован общий мастер-документ проекта.
  - Составлен список текущих ограничений и задач для доработки.
-

## Важные замечания:

- Переход к предсказанию концентраций или коэффициентов коррекции — уже на этапе MVP.
- Field\_reference используется с самого начала для анализа деградации.
- Итеративный подход: после MVP все этапы могут быть повторены и улучшены по мере поступления новых данных и идей.