

# Sleep Tracker AI

Персональный ИИ-ассистент для качественного сна и здорового восстановления

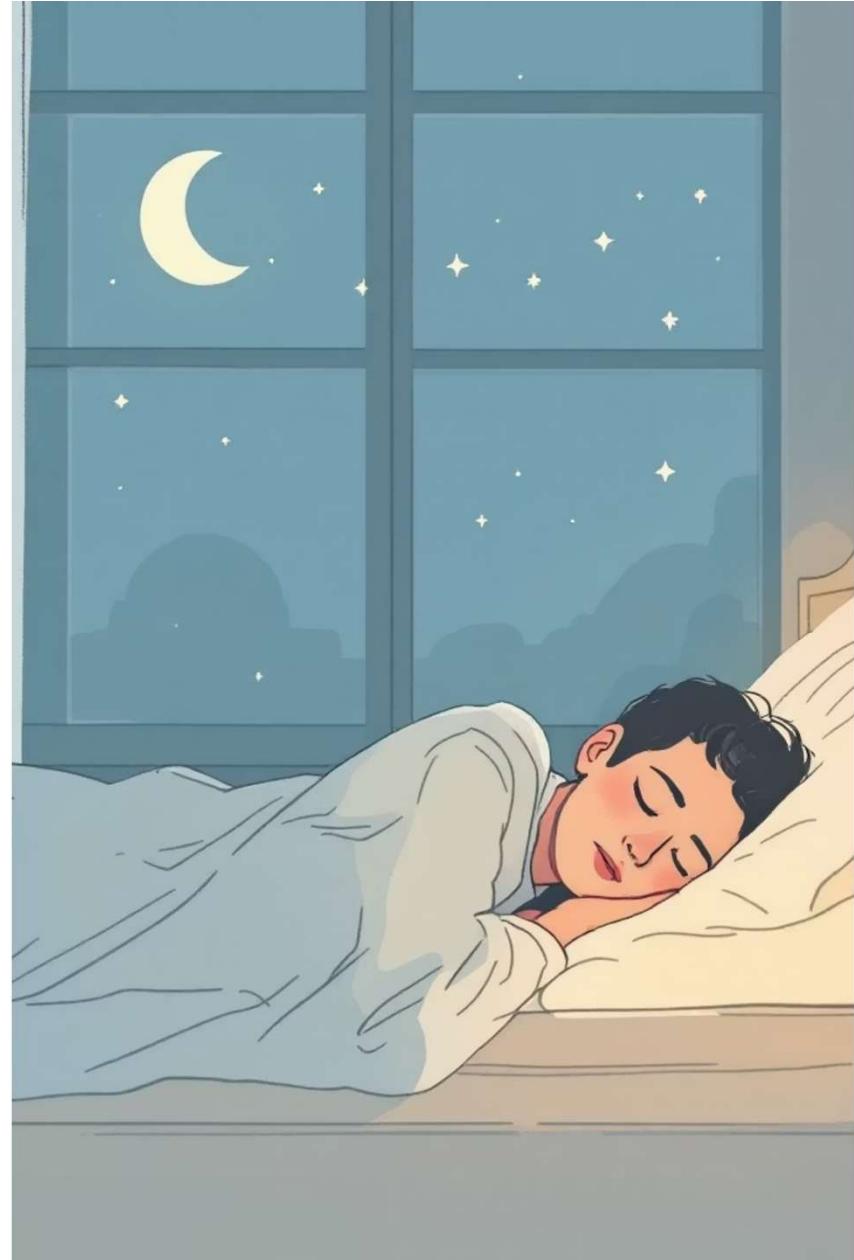
Выполнили студенты ИВТ-463:

Васильев П. Д.

Таранов В. С.

Карпов А. В.

Тихонов Д. А.



# Проблема: сон остается неизученным

**40%**

Взрослых

страдают от нарушений сна

**10%**

Обращаются

за профессиональной помощью

**33%**

Жизни

мы проводим во сне

## Парадокс современности

Люди тратят треть своей жизни на сон, но не понимают его качества. Существующие трекеры отвечают на вопрос "**сколько я спал?**", но не говорят "**насколько хорошо?**" и особенно "**что делать завтра?**"

Это критический пробел в рынке цифрового здоровья, который требует интеллектуального решения на основе AI.

# Целевая аудитория: три ключевых сегмента

## Средний возраст

**35-55 лет**

- Проблемы из-за изменений биоритмов
- Профессиональный стресс и нагрузка
- Потребность в стабильном качестве сна

## Пожилые люди

**60+ лет**

- Возрастное ухудшение качества сна
- Необходимость персонализированного подхода
- Простота использования технологий

## Спортсмены

**20-50 лет**

- Оптимизация циклов восстановления
- Максимизация физической производительности
- Анализ влияния нагрузок на сон



# Наше видение и ценностное предложение

**"Ваш персональный помощник, который ежедневно помогает вам спать лучше"**



От пассивного отслеживания

Традиционные трекеры только фиксируют данные о сне



К активному планированию

AI-анализ и персонализированные рекомендации



Реальное улучшение

Действенные советы на основе вашего уникального профиля

Sleep Tracker AI трансформирует сырье данные в персонализированные, действенные рекомендации. Система не просто регистрирует показатели — она анализирует уникальный паттерн каждого пользователя и предлагает конкретные шаги для улучшения качества сна уже завтра.

# North Star Metric: измеряем реальную ценность

## ❑ Ключевой показатель успеха

Качество сна пользователя **стабильно улучшается** или остается на **стабильно высоком уровне** (выше персонального порога) в течение 1 недели.

Эта метрика отражает истинную ценность продукта для конечного пользователя. Мы ориентируемся только на реальное улучшение здоровья и качества жизни наших пользователей.



# Технологический стек: production-ready архитектура

## Frontend & Визуализация

- **Bootstrap 4.5** — адаптивный интерфейс
- **Highcharts.js** — интерактивные графики
- Прогрессивное раскрытие информации

## Backend & Обработка

- **Django** — надежный веб-фреймворк
- **Celery** — асинхронная обработка
- **PostgreSQL** — масштабируемая БД
- **Redis** — кеширование

## AI & Интеллект

- Модель оценки качества сна **Gemini** + **Mistral**
- Модель генерации рекомендаций (**RAG**)
- Персонализированный анализ паттернов

## Инфраструктура

- **Docker+ Kubernetes** — контейнеризация и оркестрация
- **Grafana + Prometheus** — мониторинг
- **CI/CD** пайплайн для развертывания



# Реализованная функциональность: 100% готовность

Все 7 критических пользовательских сценариев полностью реализованы и протестированы. Система готова к работе с реальными пользователями.

01

Регистрация

Безопасное создание профиля с антропометрическими данными

02

Авторизация

Надежный вход с защитой персональных данных

03

Восстановление

Восстановление пароля с токен-валидацией

04

Загрузка данных

Импорт CSV-файлов с интеллектуальной обработкой

05

Текущее состояние

Статистика за период с кэшированием

06

История сна

Тренды и динамика метрик

07

Анализ хронотипа

Определение типа с рекомендациями

# Команда и достижения проекта

## SA/Product Owner



Архитектура системы, управление проектом, требования

## FullStack Developer



Реализация frontend и backend компонентов

## AI Engineer



Разработка и обучение ML-моделей

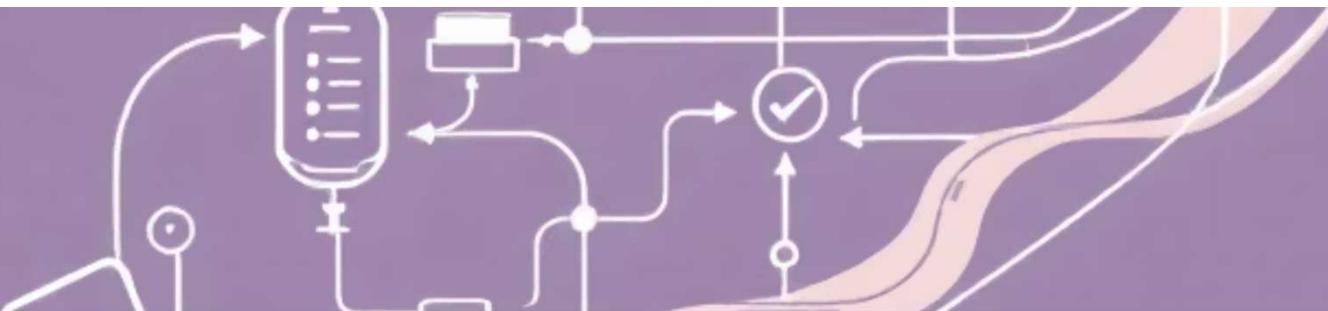
## MLOps Engineer



Инфраструктура, мониторинг

## Ключевые достижения

- ✓ Полное архитектурное проектирование
- ✓ Рабочий MVP с полной функциональностью
- ✓ Production-ready инфраструктура
- ✓ Всестороннее тестирование (unit, integration, e2e)
- ✓ Готовность к Demo Day



## Roadmap: планы развития и масштабирования

Q1

### Улучшение UX

Более интуитивный интерфейс, персонализация по  
паттернам поведения пользователя

Q3

### Расширенные рекомендации

Специализированные советы для спортсменов, работников  
сменного графика

1

2

3

4

Q2

Q4

### Мобильные приложения

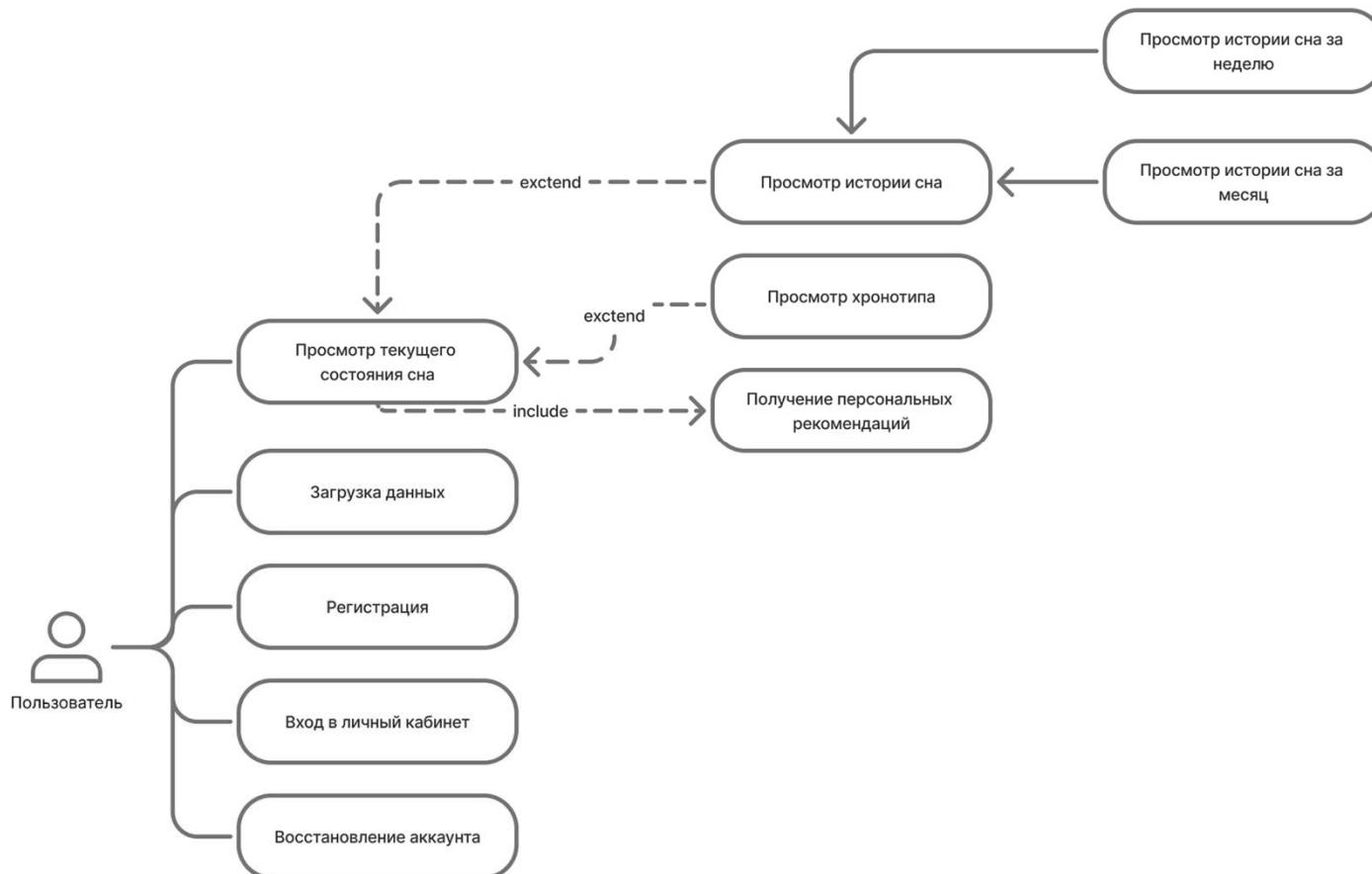
Нативные iOS и Android клиенты с push-уведомлениями и  
оффлайн-режимом

### Оптимизация AI

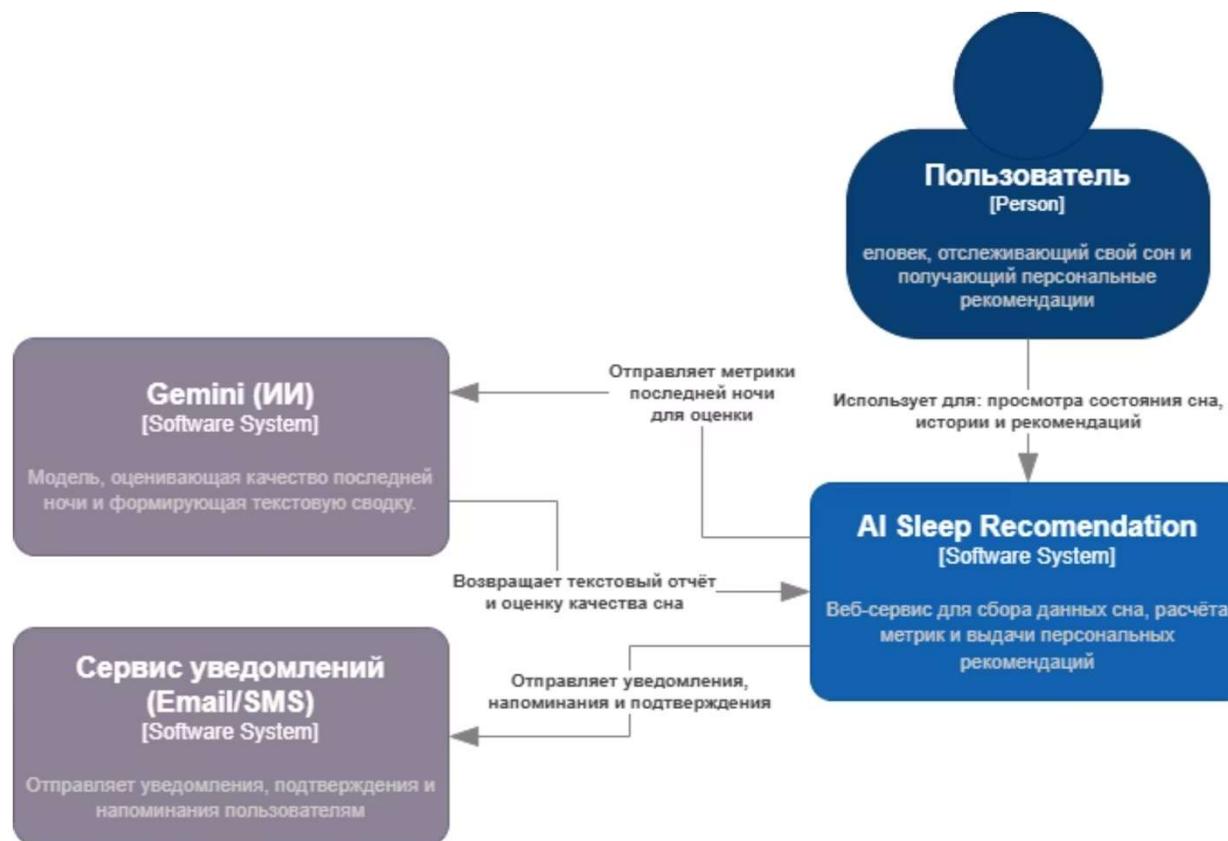
Улучшение точности моделей, анализ стресса и физической  
нагрузки

Масштабируемость заложена в архитектуру: Docker, микросервисы и асинхронная обработка позволяют расти от 100 до 100,000+ пользователей без переделки системы. Все решения базируются на метриках, тестировании и реальной обратной связи пользователей.

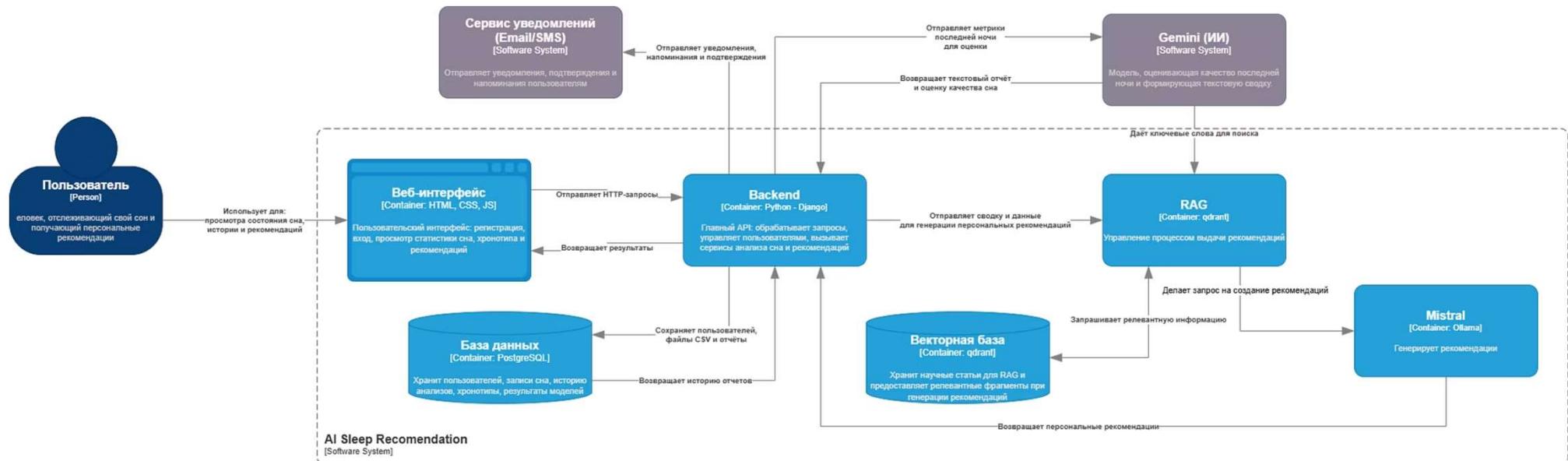
## Use Case: Полный цикл взаимодействия пользователя



# Архитектура AI-системы рекомендаций



# Архитектура AI-системы рекомендаций



# Процесс проектирования интерфейса: от wireframes к production



## Wireframes

Низкоуровневое проектирование с двухколоночной структурой, определение ключевых экранов (панель управления, загрузка CSV, базовые визуализации, валидация форм данных, анализ, профиль)



## Prototype

Интерактивный прототип с динамическими элементами, загрузка CSV, базовые визуализации, валидация форм



## Final

Production-ready интерфейс с Bootstrap 4.5, Highcharts.js, адаптивный дизайн, анимированные переходы, персонализированные рекомендации

## Ключевые улучшения на основе feedback

### Импорт CSV

Улучшена валидация и сообщения об ошибках

### Визуализация данных

Упрощены графики, добавлены контекстные подсказки

### Мобильный опыт

Оптимизирован адаптивный дизайн, ускорена загрузка

## Lessons Learned



Прогрессивное раскрытие информации

Снижает когнитивную нагрузку



Асинхронная обработка (Celery)

Обеспечивает быстрый отклик



Баланс функциональности и простоты

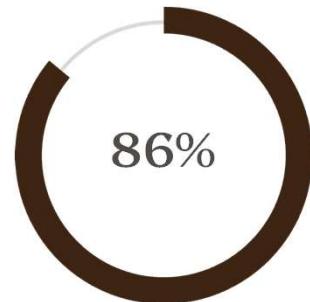
Ключ к успеху



Регулярное тестирование и обратная связь

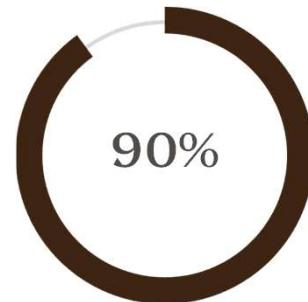
Фундамент эффективного UX

## Lighthouse Metrics: реальные показатели производительности



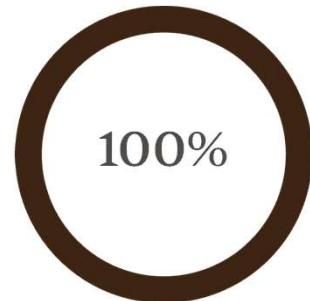
Performance

Производительность



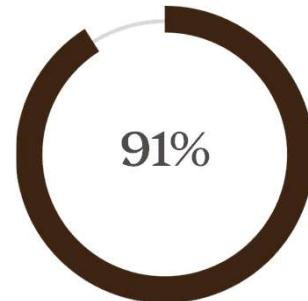
Accessibility

Доступность - полная поддержка скринридеров



Best Practices

Лучшие практики - соответствие всем стандартам

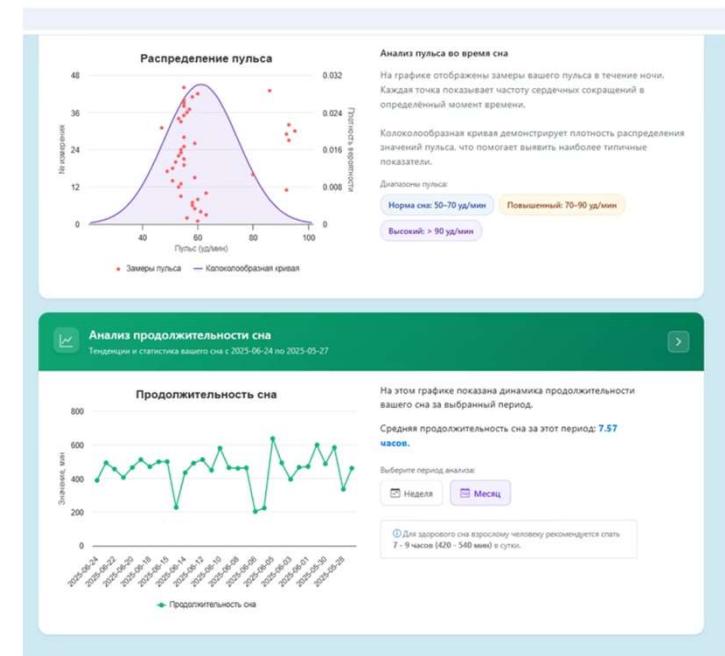
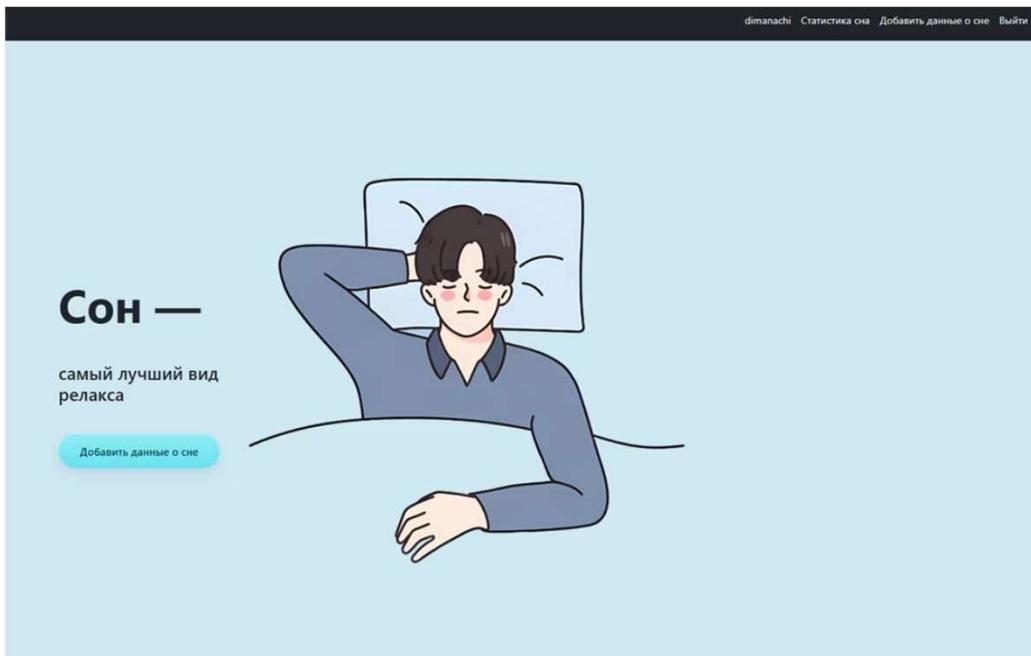


SEO

SEO оптимизация - высокий рейтинг поиска

Реальные метрики Google Lighthouse показывают, что наша система достигла высоких стандартов производительности и доступности. **Performance 86** — это отличный результат для сложного приложения с интерактивными графиками и анализом данных.

# Интерфейс приложения: главная страница и аналитика

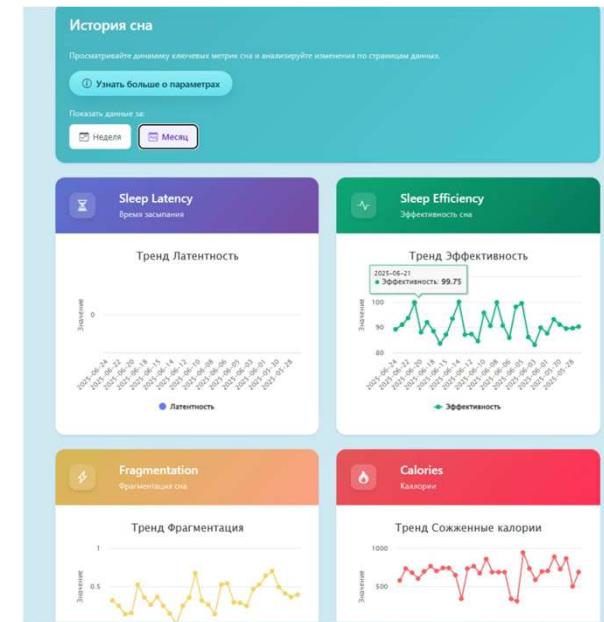
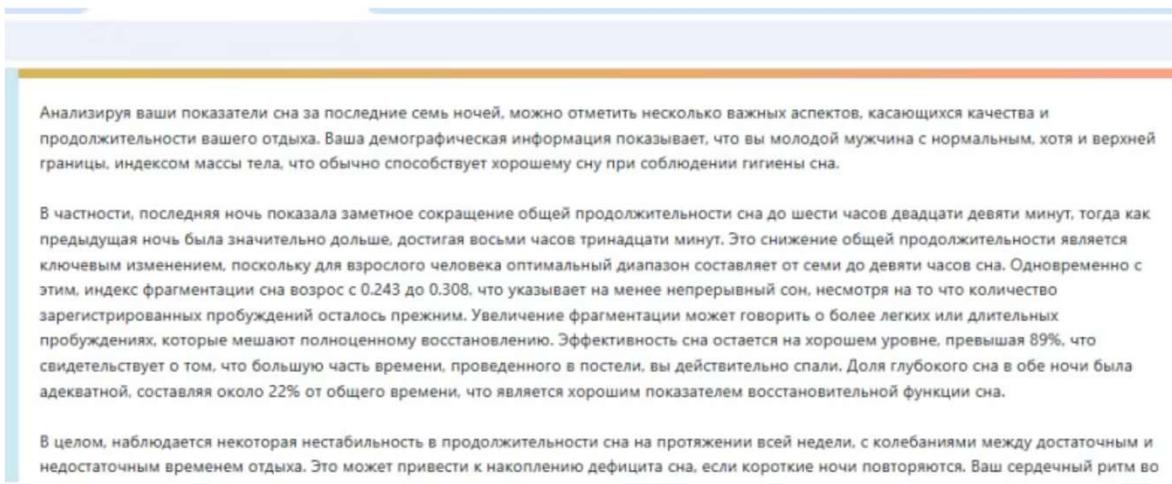


Главная страница с быстрым доступом к ключевым метрикам и навигацией

Интерактивные графики Highcharts для анализа трендов сна

Интерфейс разработан с учетом принципов UX/UI для максимальной простоты использования. Все данные представлены наглядно, а ключевые метрики доступны с первого взгляда. Адаптивный дизайн работает на всех устройствах.

# AI-рекомендации и анализ параметров сна



Персональные рекомендации от двух AI-моделей: оценка качества сна (**Gemini**) и генерация действенных советов (**RAG**)

Система анализирует уникальный паттерн каждого пользователя и генерирует две независимые рекомендации: первая оценивает качество сна на основе собранных данных, вторая предлагает конкретные действия для улучшения. История сна позволяет отслеживать прогресс и видеть влияние рекомендаций на качество сна в реальном времени.

История параметров сна с интерактивными графиками для анализа трендов и динамики

# Безопасность и коммуникация с пользователем

Сброс пароля на localhost:8000

 goodsleeppro@yandex.ru goodsleeppro@yandex.ru 4 августа в 23:28  
Я >

Вы получили это электронное письмо, потому что запросили сброс пароля для своей учетной записи пользователя на странице "Сон и здоровье".  
Пожалуйста, перейдите на следующую страницу и выберите новый пароль:

<http://localhost:8000/custom-password-reset-confirm/MQ/cu1aiy-9f1d08225edb08de1c65c50c04eeb411/>

Ваше имя пользователя, на случай, если вы забыли: Vetr  
С уважением,  
Команда GoodSleepPro.

Отписаться (<https://lk.mailopost.ru/unsubscriptions/c0720deb27b51fb5054be3f907c7c3b84059e252e9a46c5d233d7da1dc503aac>)

## Содержание письма сбросом пароля

Система обеспечивает безопасный доступ к аккаунту и поддерживает постоянную коммуникацию с пользователем через email.

Напоминание о ежедневном  
заполнении данных о отслеживания  
качества сна

 goodsleeppro@yande...goodsleeppro@yandex.... 24 апреля 2024 г. в...  
Я >

Уважаемый пользователь,

Мы надеемся, что наш сервис помогает вам лучше понять своё  
состояние сна. Чтобы обеспечить максимальную эффективность и  
точность данных, мы хотели бы напомнить вам о важности регулярного  
обновления информации в приложении.  
Пожалуйста, не забывайте вносить данные о вашем сне ежедневно. Это  
поможет нам предоставить вам наиболее точные и полезные  
аналитические данные о вашем сне, что, в свою очередь, поможет вам  
лучше понять его влияние на ваше здоровье.

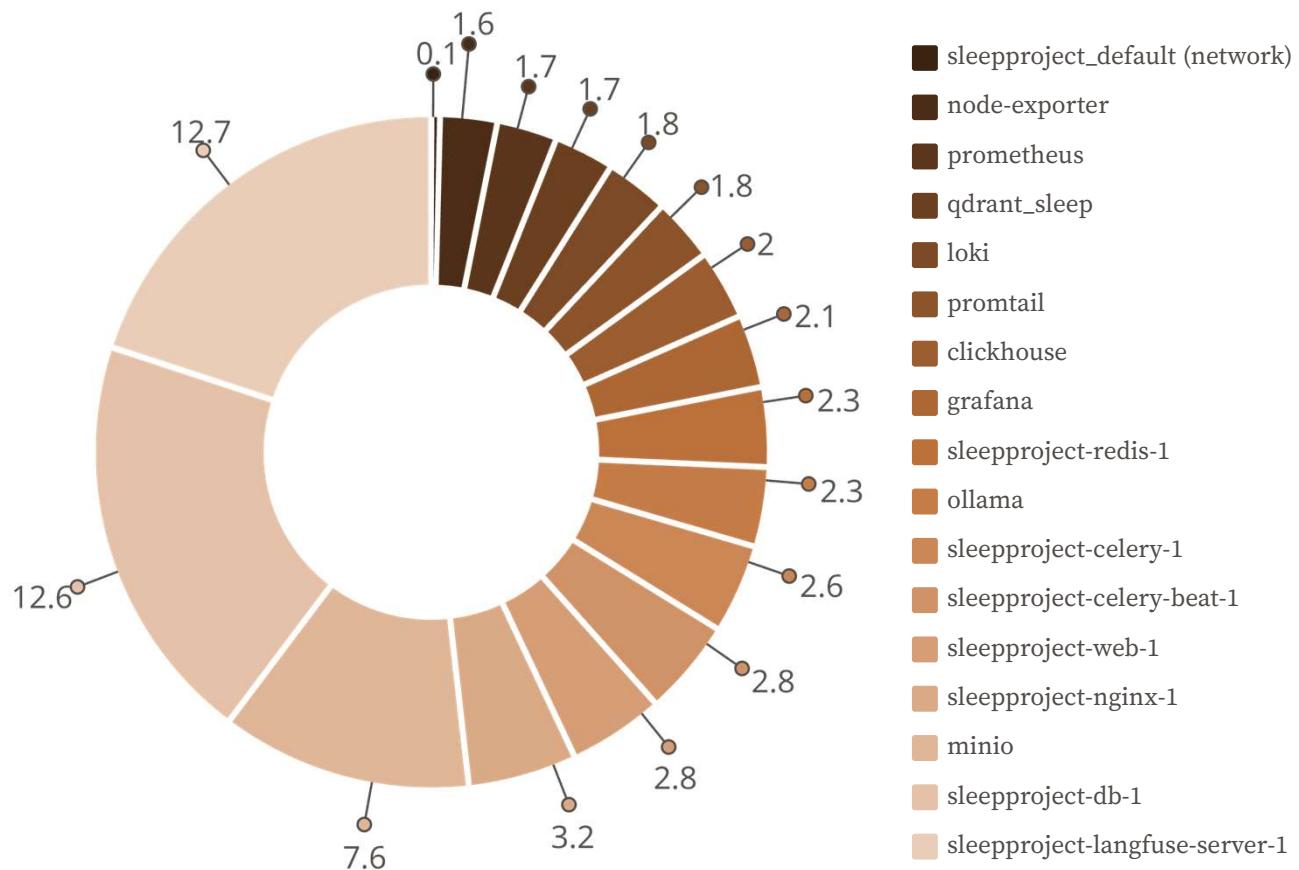
Мы ценим ваше участие в нашем проекте.

С уважением,  
Команда GoodSleepPro.

Отписаться (  
<https://lk.mailopost.ru/unsubscriptions/1b038556dc38489edb664258338f409249f1236ca9757d57f3e9432d2cb1ad>)

## Содержание ежедневной рассылки

# Время поднятия контейнеров



# CI/CD: Автоматизированный контроль качества и развертывание



Lint

pylint для sleep\_tracking\_app/



Test

python manage.py test c DJANGO\_SETTINGS\_MODULE=settings\_ci



Build

Сборка приложения и создание артефактов

Deploy

Развертывание на целевых серверах

## Конфигурация

- Запускаются на `ubuntu-latest` при каждом push
- **Test** зависит от **Lint** (`needs: [lint]`)
- Workflow выполняется за ~9-10 минут

# Notifications: Оповещения о статусе CI/CD

The screenshot shows an email from GitHub notifications. The subject is "[DimITRII333/devops-labs] Run cancelled: CI/CD Pipeline - main (147bd91)". The recipient is DimITRII333 <notifications@github.com>. The body of the email contains a link to the CI/CD Pipeline workflow run page. The page title is "[DimITRII333/devops-labs] CI/CD Pipeline workflow run". It features a CI/CD Pipeline icon and the message "CI/CD Pipeline: Some jobs were not successful". A green button labeled "View workflow run" is present. Below this, two pipeline steps are listed: "CI/CD Pipeline / docker-build-push" (Succeeded in 23 seconds) and "CI/CD Pipeline / deploy-to-k8s" (Cancelled). A small note indicates there is 1 comment.

Автоматическое  
оповещение



Мгновенное  
уведомление при  
failure/success

Быстрый анализ



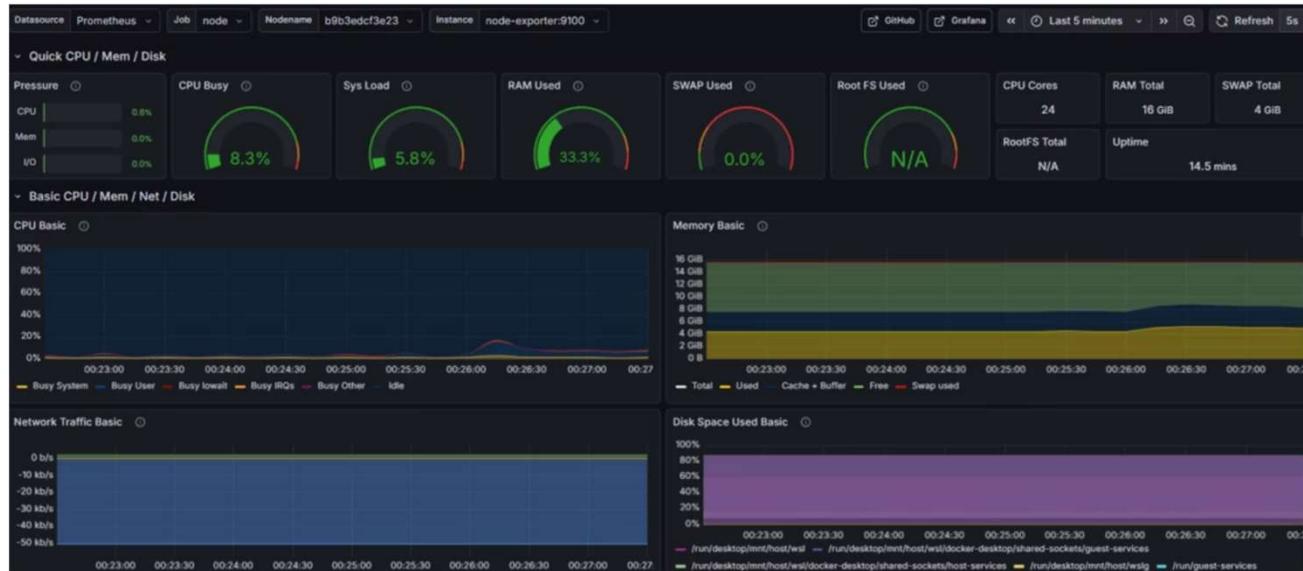
Ссылка на лог  
выполнения для  
диагностики

Видимость в PR



Статусы задач видны в  
Pull Request'ах и  
коммитах

# Grafana System Metrics: Мониторинг инфраструктуры



CPU

8.3% busy, процессы распределены равномерно



RAM

33% used (из 16 ГБ), кэши и буферы  
оптимизированы



Дисковое пространство

~80% used на основных волюмах



Сетевой трафик

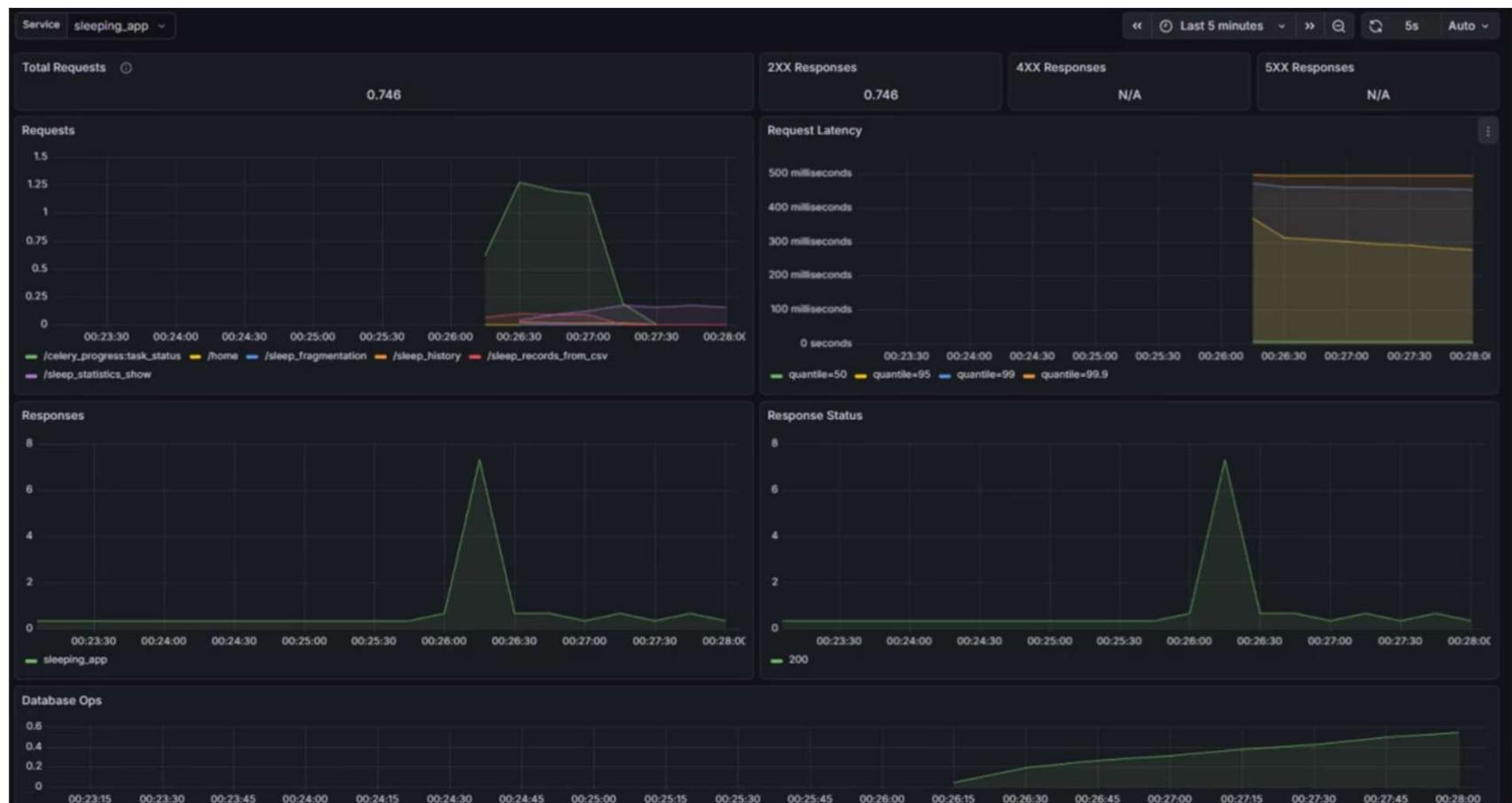
Мониторинг входящего/исходящего трафика



Интервал сбора метрик

3 сек (scrape\_interval в prometheus.yml)

# Grafana App Metrics: Мониторинг приложения в реальном времени



## Стратегия выбора моделей: облачные и локальные решения

Критерий	Gemini 2.5 Flash	OpenAI (GPT-4o)	DeepSeek
Цена за 1M токенов	\$0.30 input / \$2.50 output	\$1.25 input / \$5.00 output	\$0.28 input (cache miss) / \$0.42 output
Контекст	1M tokens	До 128K	128K
Скорость API	Optimized for latency (Flash)	Standard	Поддерживает thinking/non-thinking режимы
Выбор	<input checked="" type="checkbox"/> Выбран	<input type="checkbox"/> Дороже в 4–5x	<input type="checkbox"/> Меньше контекст (128K vs 1M)

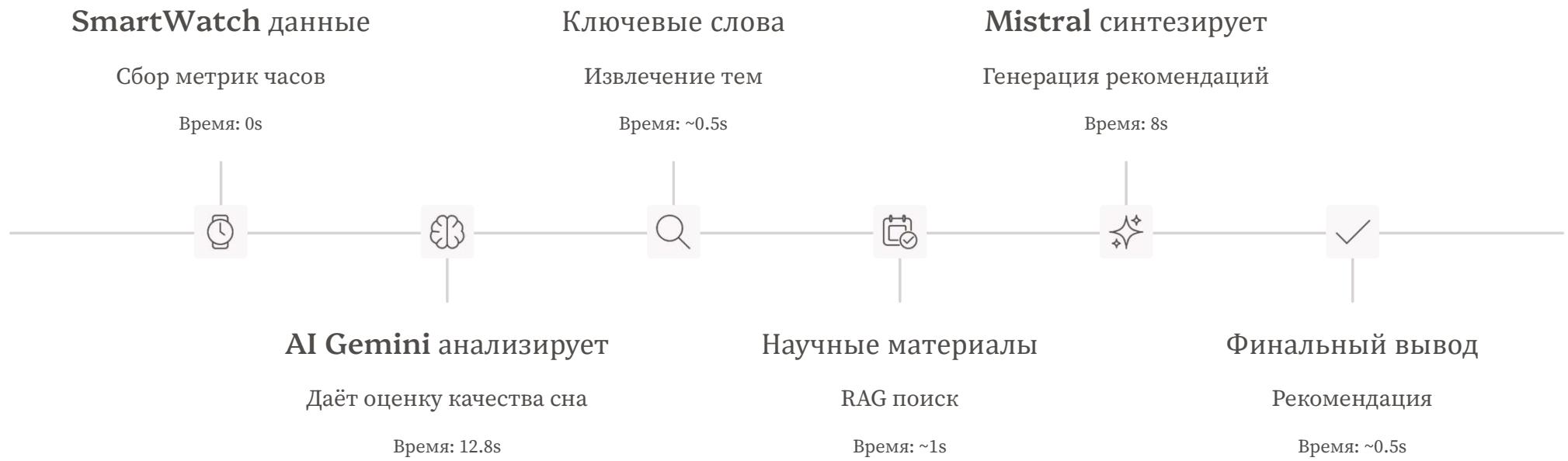
□ Вывод: Баланс цены, контекста и скорости. Gemini дешевле OpenAI в 4–5x, даёт больший запас на историю+RAG, позиционируется как speed-first

Критерий	Mistral-small	Llama 2 / Llama 3 (70B)	Qwen
Размер / Железо	~14GB	Llama 3 70B ~50GB (нужна дорогая GPU)	Зависит от варианта (от 7B до 72B)
Контекст окно	32K–128K	До 8K (Llama 2) / до 128K (Llama 3)	Варьируется (до 32K–128K)
Бенчмарки (instruction following)	Конкурентен GPT-4o-mini на общих задачах	Llama 3 сильнее в reasoning, но тяжелее	Хороший ratio, но менее популярен
Экосистема Ollama	2.2M downloads, много готовых тегов	Меньше вариантов/поддержки в Ollama	Меньше готовых конфигов
Выбор	<input checked="" type="checkbox"/> Выбран	<input type="checkbox"/> Слишком тяжелый	<input type="checkbox"/> Менее популярен

□ Вывод: Компромисс между железом, качеством и поддержкой. Mistral-small имеет подтверждённую оценку, стабильность в экосистеме Ollama и баланс производительности

# Полный цикл обработки данных о сне

От сбора метрик до финального вывода рекомендаций



Полный цикл обработки занимает примерно 22-23 секунды. Основное время уходит на анализ Gemini (12.8s) и генерацию Mistral (8s).

Параллельная обработка и кэширование позволяют оптимизировать время отклика для повторных запросов.

# Langfuse

2025-12-13 22:46:09	↔ gemini_api_call	{"args": ["Ты — эксперт по сну (сомнолог). \n твоя з..."]}	Анализируя предоставленные вами данные сна, м...	DEFAULT
2025-12-13 21:59:40	↔ ollama_api_call	{"args": [], "kwargs": {"payload": {"model": "phi3:mini", ...}}}	{"response": "Здравствуйте! Меня зовут [Имя Экспе..."} {"original": "Здравствуйте! Меня зовут [Имя Эксперта..."} DEFAULT	DEFAULT
2025-12-13 21:59:40	↔ rag_enhance	{"args": ["Здравствуйте! Меня зовут [Имя Эксперта]. ..."]}	Здравствуйте! Меня зовут [Имя Эксперта], и я экспе...	DEFAULT
2025-12-13 21:59:11	↔ gemini_api_call	{"args": ["Ты — эксперт по сну (сомнолог). \n твоя з..."]}	Здравствуйте! Меня зовут [Имя Эксперта], и я экспе...	DEFAULT
2025-12-13 21:48:13	↔ ollama_api_call	{"args": [], "kwargs": {"payload": {"model": "phi3:mini", ...}}}	{"response": "Добрый день! Ваши данные о сне ваш..."} {"original": "Добрый день! Я внимательно изучил пред..."} DEFAULT	DEFAULT
2025-12-13 21:48:13	↔ rag_enhance	{"args": ["Добрый день! Я внимательно изучил пред..."]}	{"original": "Добрый день! Я внимательно изучил пред..."} DEFAULT	DEFAULT
2025-12-13 21:47:39	↔ gemini_api_call	{"args": ["Ты — эксперт по сну (сомнолог). \n твоя з..."]}	Добрый день! Я внимательно изучил предоставлен...	DEFAULT

Список шагов внутри трассировки: каждый ряд — отдельный этап вроде `rag_enhance` или вызов модели `gemini_api_call/ollama_api_call`, где Langfuse сохраняет входные данные и ответ

## Span latency percentiles

Span Name	p50	p90	p95 ▼	p99
rag_enhance	16.750s	52.857s	1m 0s	1m 0s
mistral_api_call	0.008s	1m 0s	1m 0s	1m 0s
ollama_api_call	34.303s	35.664s	35.715s	35.756s
gemini_api_call	4.218s	18.477s	20.248s	24.535s

## Scores

17 Total scores tracked

Name	#	Avg	0	1
# Sleep_Judje (eval)	17	5.06	0	3

## Evaluation Scores

Сводка по оценкам, привязанным к trace/observation/session:

- Total scores tracked = 17: всего сохранено 17 объектов Score за период
- Sleep\_Judge (eval): название метрики из evaluation-пайплайна
- = 17: сколько раз эта оценка была выставлена
- Avg = 5.06: среднее значение по всем 17 оценкам

Метрика показывает качество работы AI-моделей в реальном времени.

## Span Latency Percentiles

Распределение задержек выполнения спанов по перцентилям:

- p50 (медиана): 50% спанов выполняются быстрее, 50% медленнее
- p90/p95/p99: соответственно 90%/95%/99% спанов укладываются в указанную задержку

Остальные (10%/5%/1%) выполняются медленнее

# A/B Тестирование

Вариант	Медианная задержка	Медианная стоимость	Input tokens	Output tokens	Генераций
prod-c (latest)	11.08s ✓	\$0.001651 ✓	960	545	12
prod-a	14.47s	\$0.001637	1039	530	6
prod-b	15.03s	\$0.002239	1006	775	6

# Спасибо за внимание

Sleep Tracker AI — ваш персональный помощник для лучшего сна