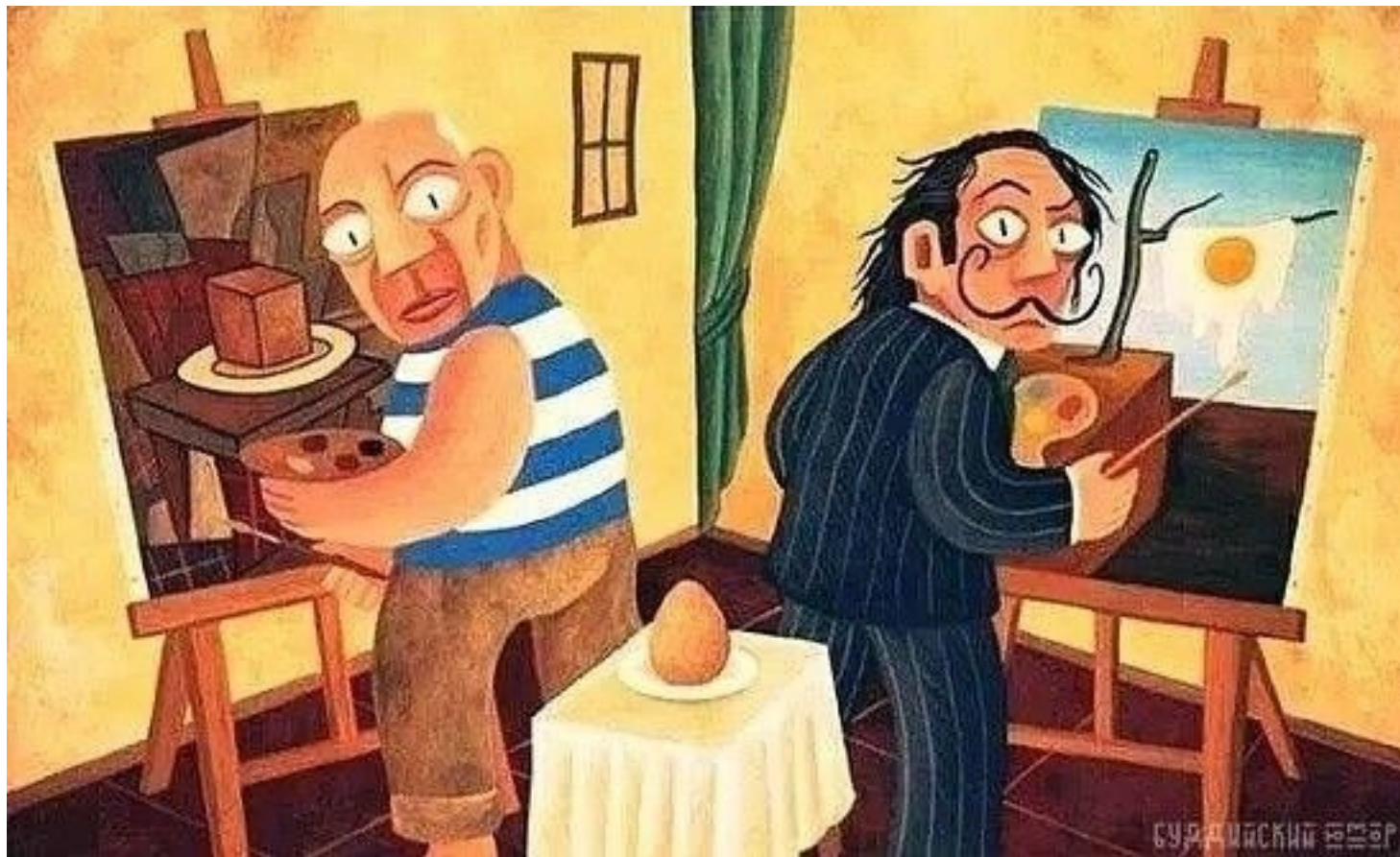


# ML System Design

Что такое ML System Design

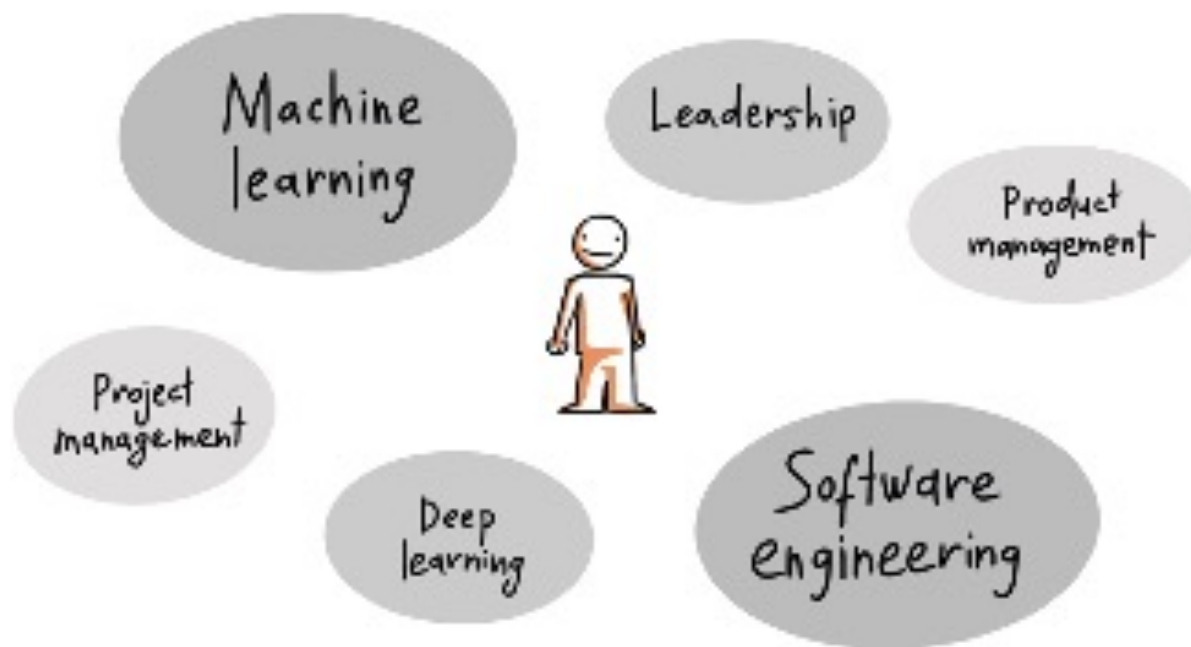
# У каждого свое понимание



# Что пишут в книгах

- MACHINE LEARNING SYSTEM DESIGN is a complex, multistep process of designing, implementing, and maintaining machine learning-based systems that involves a combination of techniques and skills from various fields and roles, including machine learning, software engineering, project management, product management, and leadership

[ENG] Machine Learning System Design With end-to-end examples. Valerii Babushkin and Arseny Kravchenko

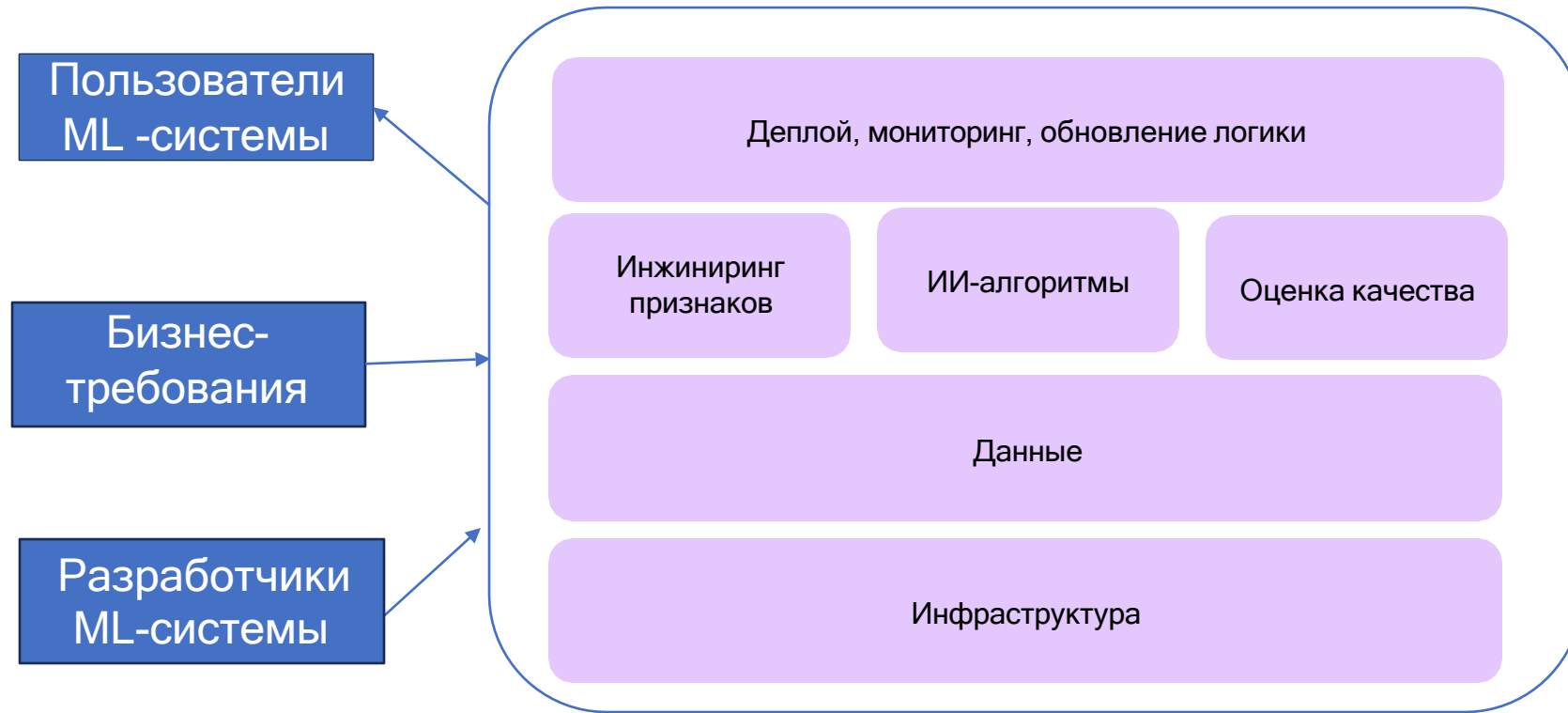


# ML System Design != ML-алгоритмы

- ML-алгоритмы - часть системы, но не единственная
- Система - шире и включает больше вопросов

# Компоненты ML-системы

# Компоненты ML-системы



# ML-разработка != Разработка

- Код и данные связаны
- Тестируется и версионизируется не только код, но и данные
- Инженерные челленджи



# Ожидание от работы в ML в индустрии

- Собрать данные
- Обучить модель
- Внедрить



# Реальность работы в индустрии

- Договориться, какую метрику оптимизируем
- Придумать/адаптировать/выбрать лосс
- Собрать данные
- Натренировать модель
- Ужаснуться качеству, набрать новых данных, переразметить
- Обучить
- Еще раз ужаснуться качеству, добрать данных
- Обучить
- Задеплоить
- Крепко спать, пока идет эксперимент
- Проснуться в 4 утра от звонка инцидент-менеджментов, что поехали метрики, отменить раскатку
- Провести анализ ошибок, дообучить модель
- Задеплоить
- Поставить свечку
- Продакшн не упал, а бизнесовые метрики ухудшаются
- Записаться к психотерапевту
- Пересмотреть оптимизируемую метрику
- Поздравляю, все сначала



# ML для исследований и в продакшне

	Исследования	Продакшн
Цель	Качество модели	У каждого стейкхолдера своя цель

# ML для исследований и в продакшне

	Исследования	Продакшн
Цель	Качество модели	У каждого стейкхолдера своя цель
Данные	Статичный слепок	Постоянно изменяются

# ML для исследований и в продакшне

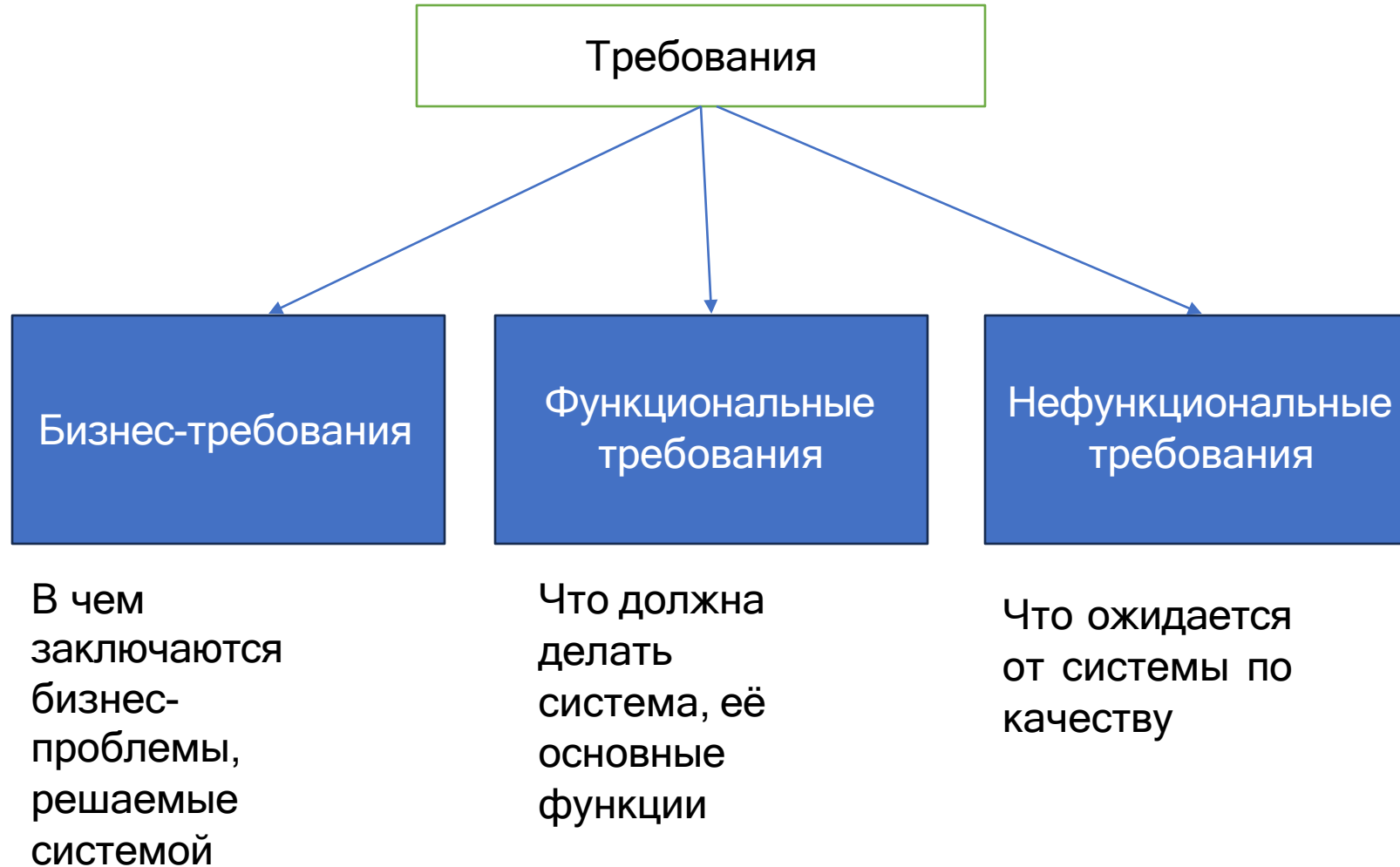
	Исследования	Продакшн
Цель	Качество модели	У каждого стейкхолдера своя цель
Данные	Статичный слепок	Постоянно изменяются
Интерпретируемость	Приятный бонус	Важен

# ML для исследований и в продакшне

	Исследования	Продакшн
Цель	Качество модели	У каждого стейкхолдера своя цель
Данные	Статичный слепок	Постоянно изменяются
Интерпретируемость	Приятный бонус	Важен
Вычислительные особенности	Быстрое обучение, высокая пропуская способность	Быстрый инференс, малая задержка

# Требования к ML-системам

# Требования можно разбить на три группы





# Нефункциональные требования

- Reliability (надёжность)

# Нефункциональные требования

- Reliability (надёжность)
- Scalability (масштабируемость)

# Нефункциональные требования

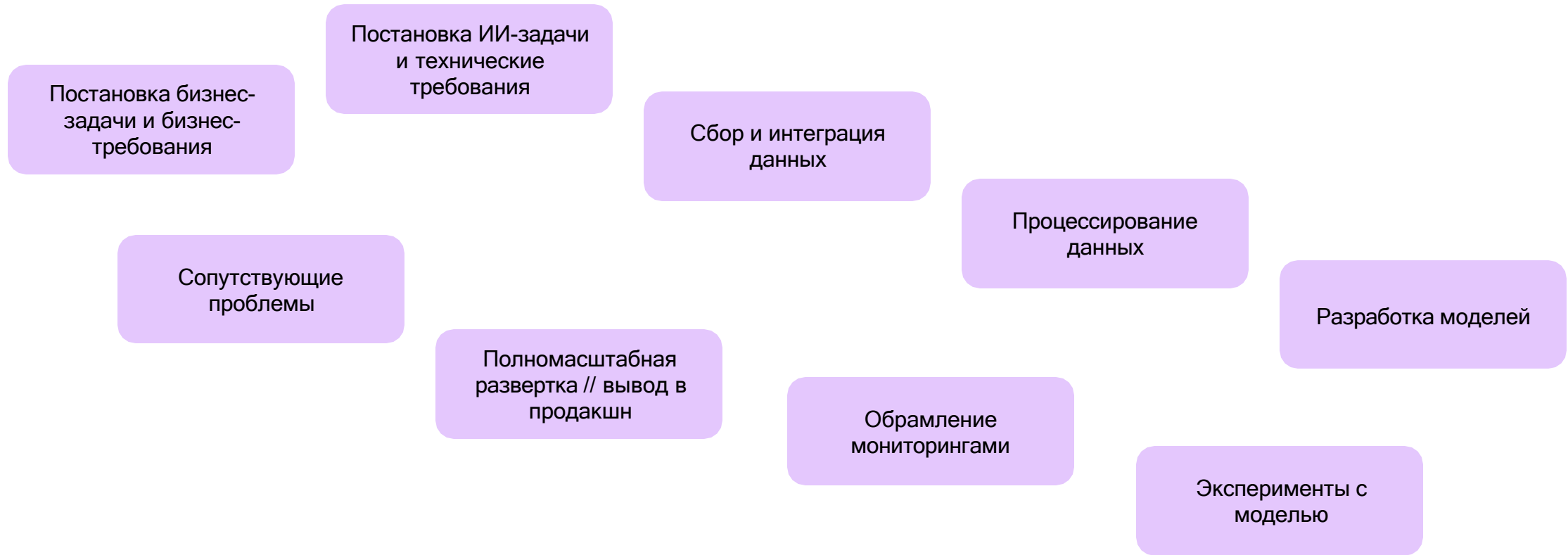
- Reliability (надёжность)
- Scalability (масштабируемость)
- Maintainability (обслуживаемость)

# Нефункциональные требования

- Reliability (надёжность)
- Scalability (масштабируемость)
- Maintainability (обслуживаемость)
- Adaptability (адаптируемость)

# Общий цикл жизни ML/AI-инициатив

# Основные этапы жизни ИИ-проекта



# Основные этапы жизни ИИ-проекта

Постановка бизнес-задачи и бизнес-требования

В чем заключается бизнес-проблема?  
Насколько масштабна проблема?  
Решается ли проблема сейчас?  
На какую бизнес-метрику хочется повлиять новым решением?

**Артефакт:** документ с бизнес-требованиями

# Основные этапы жизни ИИ-проекта

Постановка бизнес-задачи и бизнес-требования

Постановка ИИ-задачи и технические требования

Регрессия / классификация / кластеризация / генерация?  
Как часто модель должна работать?  
На устройстве / или в облаке?  
Какие требования к нагрузке?  
Какие требования к скорости выдачи ответа?

**Артефакт:** документ с поставленной задаче в терминах ИИ и техническими требованиями



# Основные этапы жизни ИИ-проекта

Постановка бизнес-задачи и бизнес-требования

Постановка ИИ-задачи и технические требования

Сбор и интеграция данных

**Артефакт:**

- ТЗ на сбор данных
- Инструкции по разметке данных
- описание источников данных и схема взаимосвязи данных

# Основные этапы жизни ИИ-проекта

Постановка бизнес-задачи и бизнес-требования

Постановка ИИ-задачи и технические требования

Сбор и интеграция данных

Процессирование данных

**Артефакт:**

- сформированный датасет
- описание признаков и всех манипуляций

# Основные этапы жизни ИИ-проекта

Постановка бизнес-задачи и бизнес-требования

Постановка ИИ-задачи и технические требования

Сбор и интеграция данных

Процессирование данных

Разработка моделей

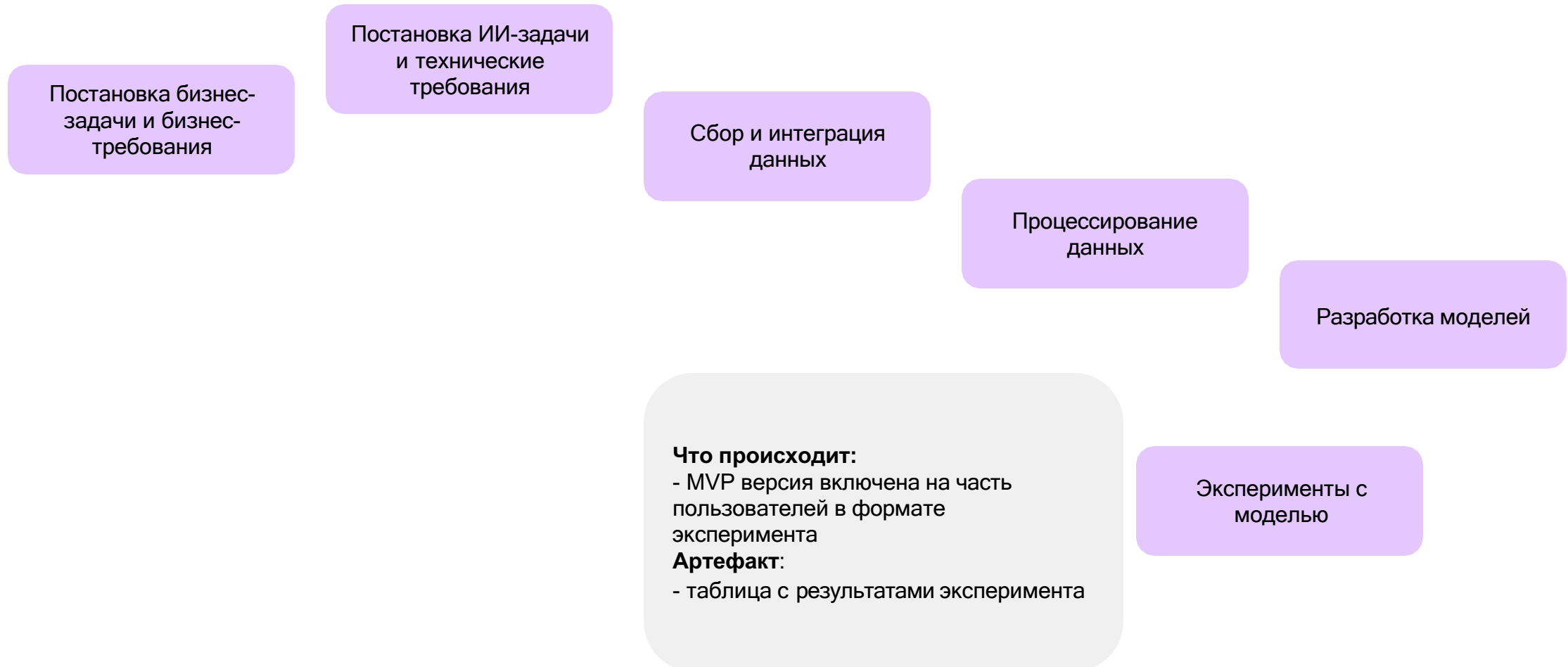
## Что происходит:

- выбор моделей / архитектур
- обучение моделей
- валидация
- оценка качества моделей
- может включать и написание сервисов

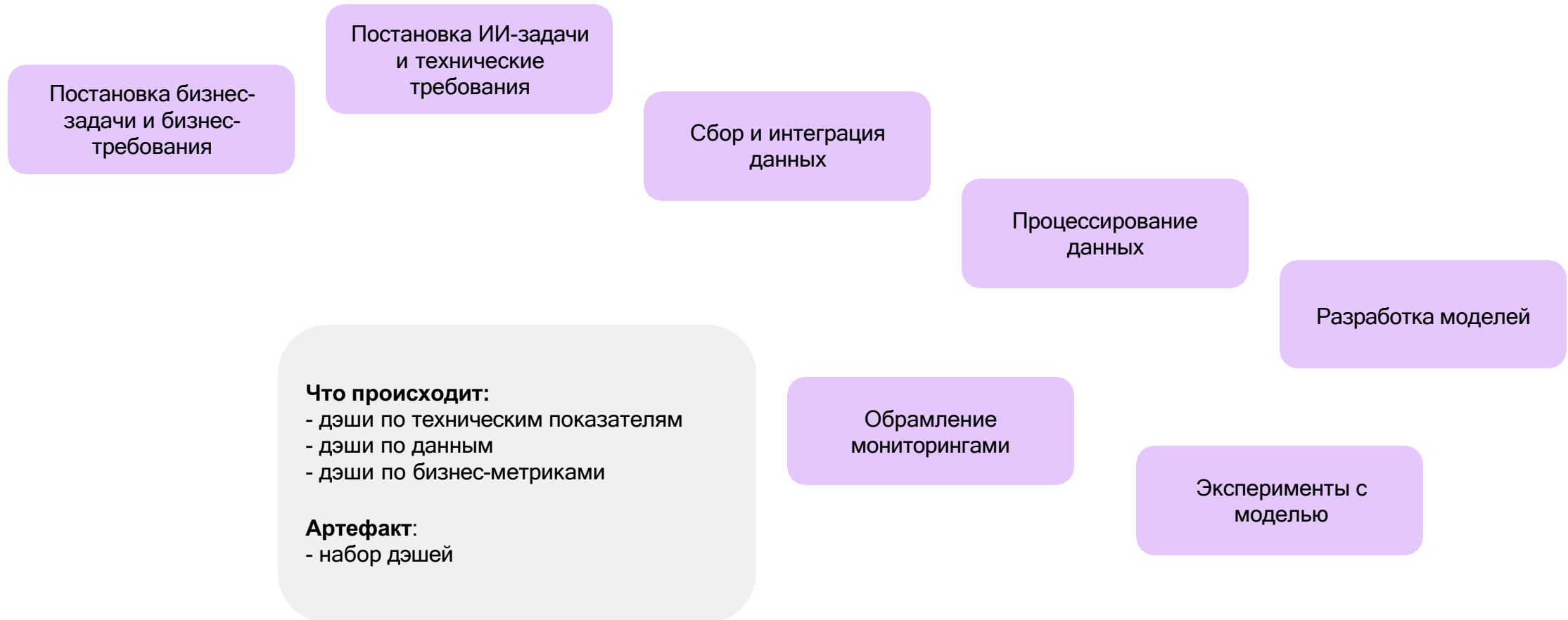
## Артефакт:

- код на Git
- файл с весами модели
- описание процедур тренировки и валидации модели

# Основные этапы жизни ИИ-проекта



# Основные этапы жизни ИИ-проекта



# Основные этапы жизни ИИ-проекта

Постановка бизнес-задачи и бизнес-требования

Постановка ИИ-задачи и технические требования

Сбор и интеграция данных

Процессирование данных

Разработка моделей

Полномасштабная развертка // вывод в продакшн

Обрамление мониторингами

Эксперименты с моделью

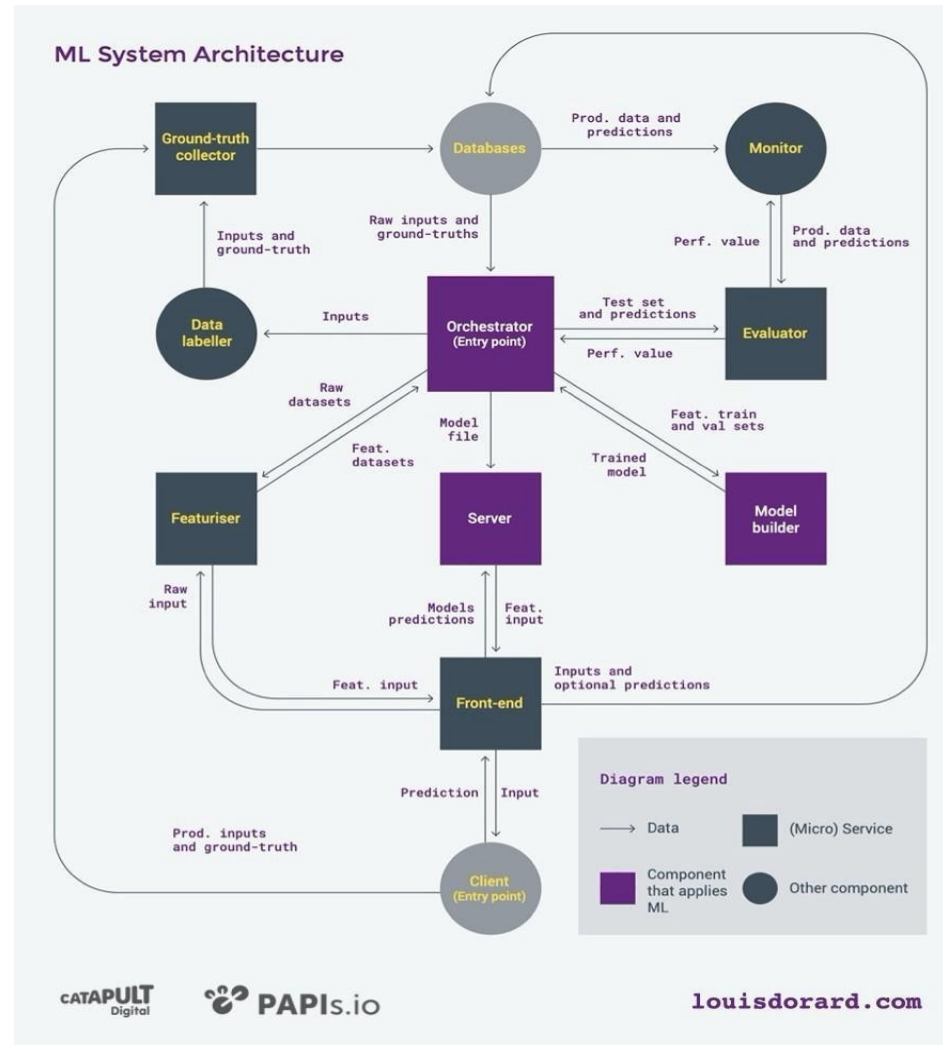
## Что происходит:

- увеличиваем влияние AI-системы на 100% пользователей или других объектов

## Артефакт:

- код на Git

# Типовая архитектура ИИ-проектов



Картинка в лучшем формате:

[https://miro.medium.com/v2/resize:fit:786/format:webp/0\\*YAdydPMicy2NqyTy](https://miro.medium.com/v2/resize:fit:786/format:webp/0*YAdydPMicy2NqyTy)

# ML System Design Doc AWS

## 1. Обзор документа:

3-5 предложений о сути документа

## 2. Мотивация:

зачем делаем проект

## 3. Метрики успеха

## 4. Требования и ограничения

## 5. Методология:

- Постановка проблемы
- Данные
- Технологии
- Эксперименты и валидация

## 6. Имплементация:

- архитектура решения - общая схема
- Инфраструктура
- Нагрузка на систему
- Безопасность
- Приватность данных
- Мониторинги и системы предупреждения
- Стоимость системы
- Точки интеграции
- Риски и неопределенность

## 7. Приложения

- Риски, альтернативы, дедлайны и тп



# ML System Design Doc Reliable ML

## 1. Цели и предпосылки

1. Зачем идти в разработку продукта
2. Бизнес-требования и ограничения
3. Что входит в скоуп проекта/итерации
4. Предпосылки решения

## 2. Методология

1. Постановка задачи
2. Блок-схема решения
3. Этапы решения задачи

## 3. Подготовка пилота

1. Способы оценки пилота
2. Что считаем успешным пилотом
3. Подготовка пилота

## 4. Внедрение

1. архитектура решения - общая схема
2. Инфраструктура
3. Нагрузка на систему
4. Безопасность
5. Приватность данных
6. Мониторинги и системы предупреждения
7. Стоимость системы
8. Точки интеграции
9. Риски и неопределенность

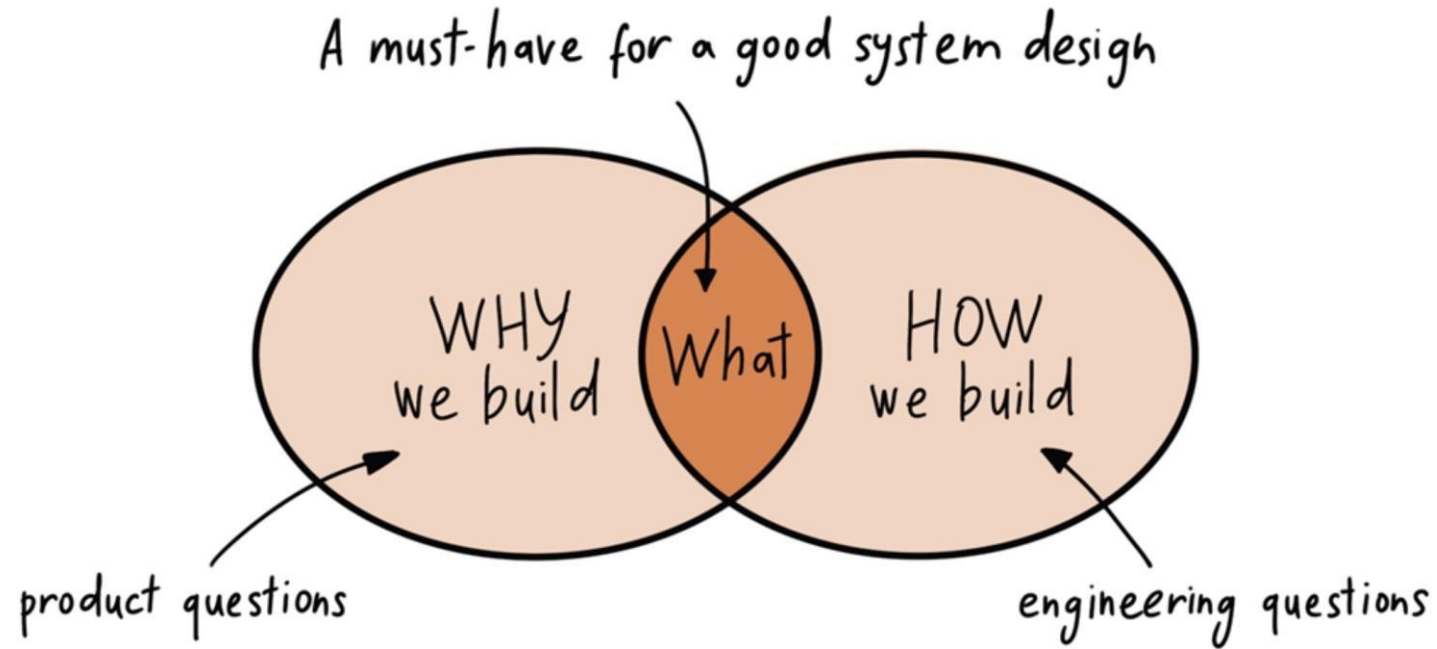
А нужен ли ML вообще?

# Цель бизнеса

- Максимизировать прибыль
- Может принять альтернативную форму - захватить рынок, стать масштабными, чтобы выйти в прибыль и максимизировать прибыль

# Фокус ML-команд

- Отличные метрики качества моделей
- Разумно устроенный код



[ENG] Machine Learning System Design With end-to-end examples.  
Valerii Babushkin and Arseny Kravchenko

# Лучшая стратегия – стать "почемучкой"

- Почему нам нужно сделать это решение?
- Какую проблему решаем?
- Как часто появляется проблема?
- Какие альтернативы?
- Почему мы строим решения в одних ограничениях, а не других?

# Иерархический подход

- 1 уровень - разобраться в постановке проблемы
- 2 уровень - детализировать постановку проблему почемучка-вопросами
- 3 уровень - специализированные детализированные вопросы

# 4 стадии решения проблем

Эвристика

Простая модель

Усложнение  
модели

Переход к сложным  
моделям



# Краткое резюме

- Что мы строим?
- Зачем мы строим?
- Как мы строим?