

От DataOps к MLOps: как раскрыть потенциал монетизации данных



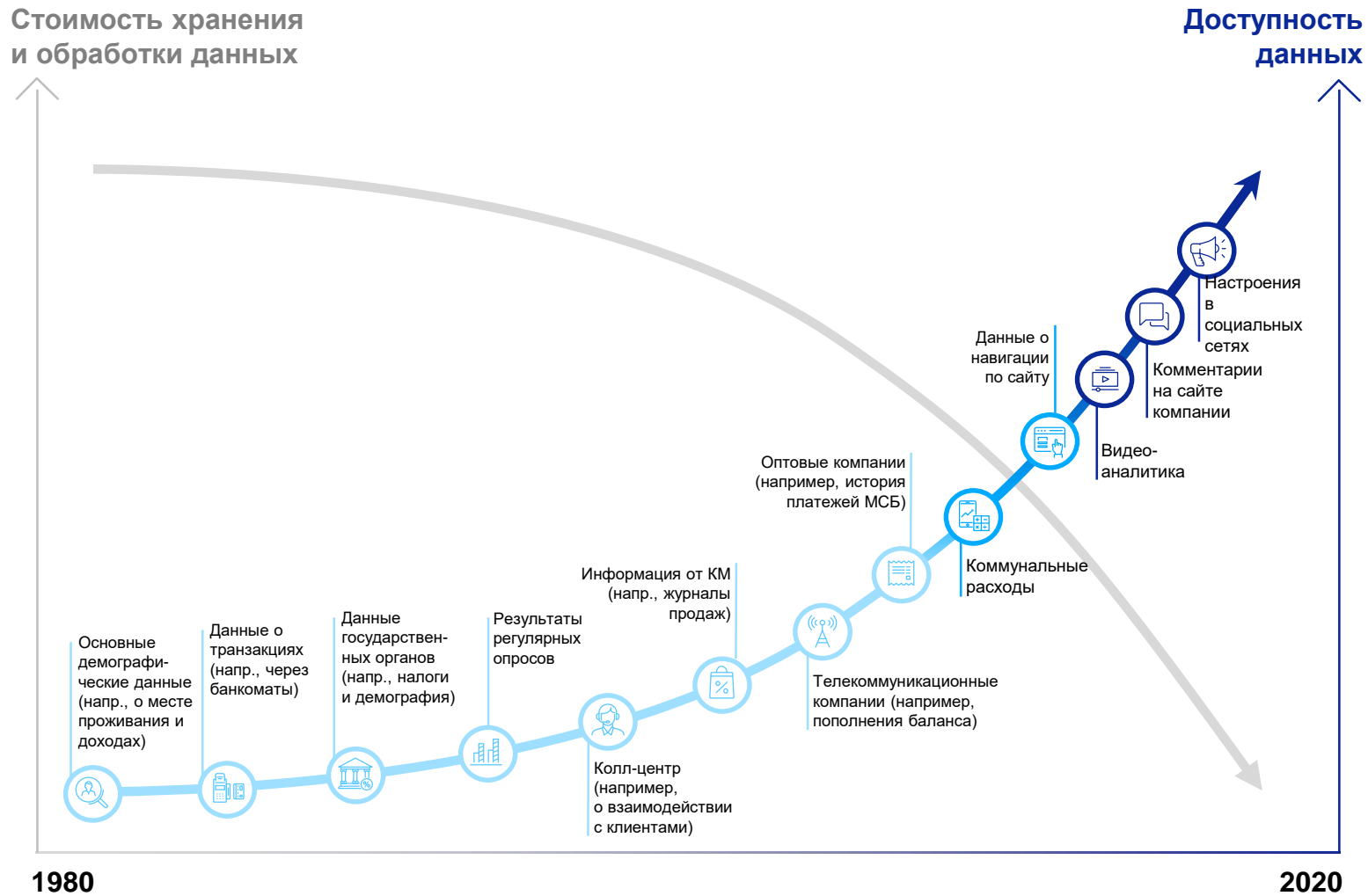
Что такое большие данные?

Для соответствия определению больших данных необходимо наличие как минимум 3 из 5 «V»

Для реализации ценности больших данных необходимы технологии, одной из которых является искусственный интеллект.



Что изменилось за последнее время?



~90% все данных
доступных сегодня были
сформированы в последние 2 года



Более, чем в 20 раз
Снизилась стоимость хранения
данных за последние 10 лет

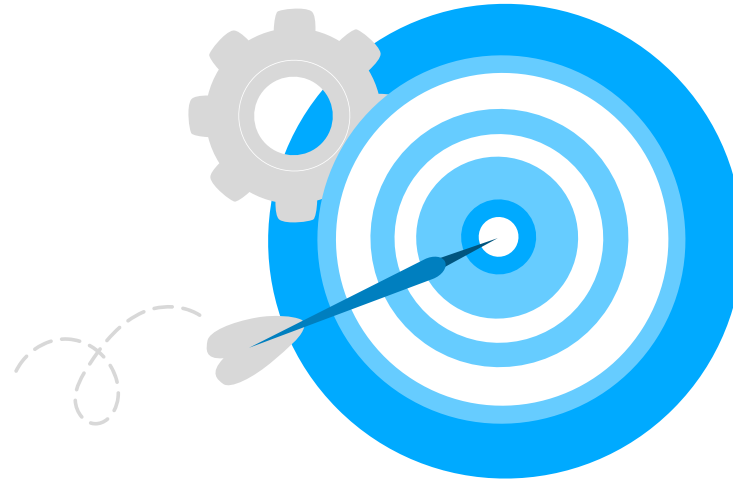
Каковы новые возможности в результате изменений?



Быстрее

Принятие решений за минуты вместо дней и часов

Ежедневное создание новых промо; изменение цены в цифровом канале каждые 15 минут



Точнее

Повышение точности оценок и прогнозов, и как следствие – принятие оптимальных решений

Определение оптимальных условий и предложений для максимизации выручки

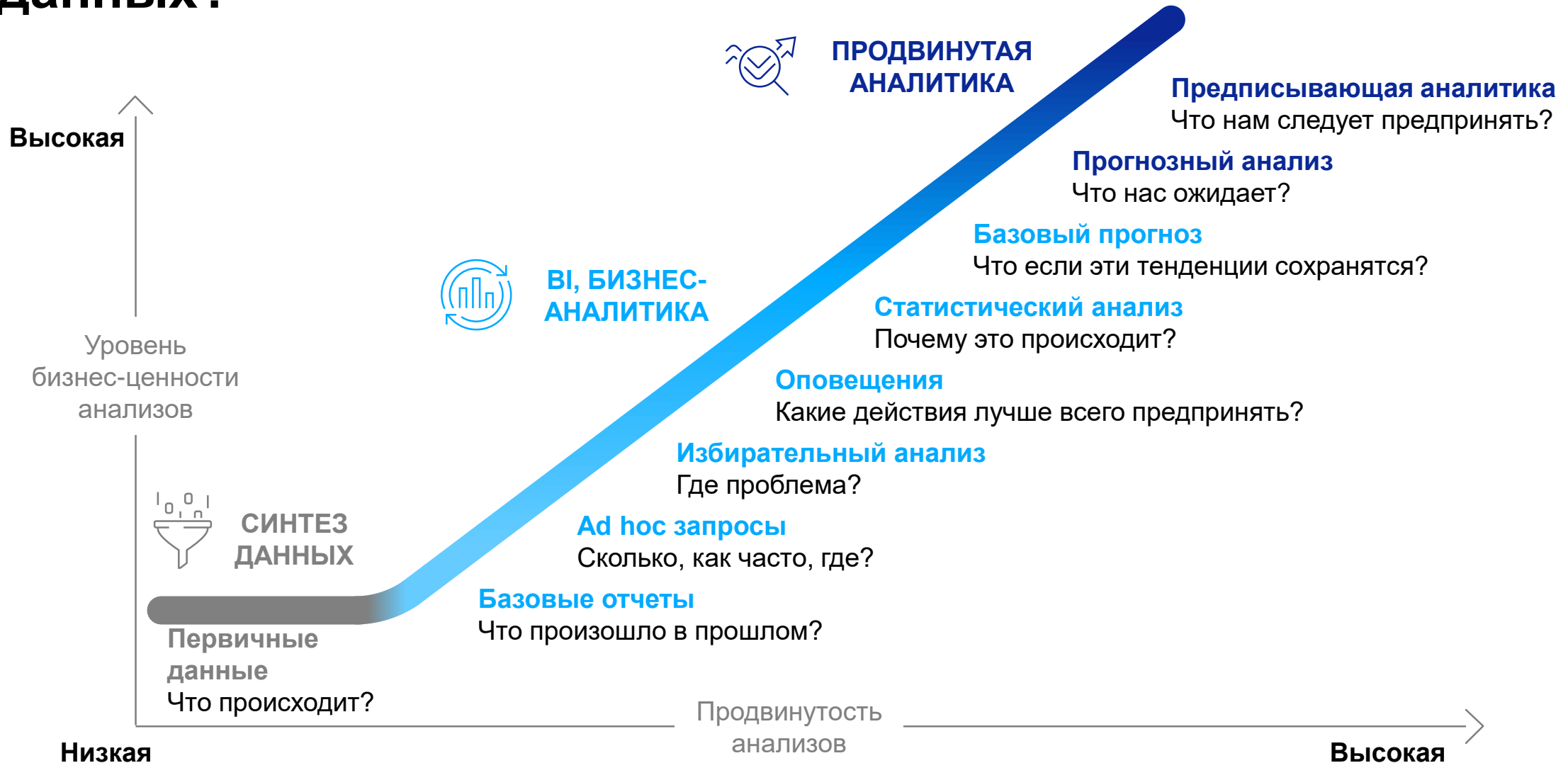


Дешевле

Автоматизация ручного труда и сокращение потерь бизнеса от неоптимальных решений

Автоматическая оценка рисков; разработка чат-ботов

Чем продвинутая аналитика отличается от BI и синтеза данных?



Что ведущие игроки стали делать иначе?

От традиционных методов анализа ...

Математические модели



Методы анализа на основе математической статистики

Заданный график анализов



Аналитика выполняется с заданной регулярностью с задержкой в реализации

Ограниченный набор данных



Аналитика, основанная только на доступных внутри организации данных

Локальное хранение данных



Доступ к данным ограничен локальными устройствами и их ограничениями в объемах хранения

Внутреннее использование данных



Данные используются только внутри компании



... к углубленной аналитике

Искусственный интеллект

Машинное обучение

Глубокое обучение

Нелинейные методы анализа для поиска сложных связей

Аналитика в режиме реального времени

Анализ данных в режиме реального времени для моментального принятия решений

Новые источники информации

Аналитика включает любые внутренние и внешние данные

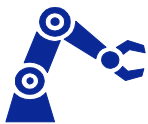
Повсеместный доступ к данным

Доступ к данным в любое время, в любом месте и с любого устройства

Монетизация данных

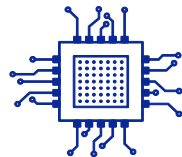
Данные и аналитика – монетизируемый актив, как внутри организации, так и вовне

Что такое искусственный интеллект, машинное обучение и глубокое обучение?



Искусственный интеллект

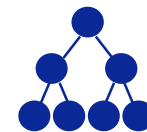
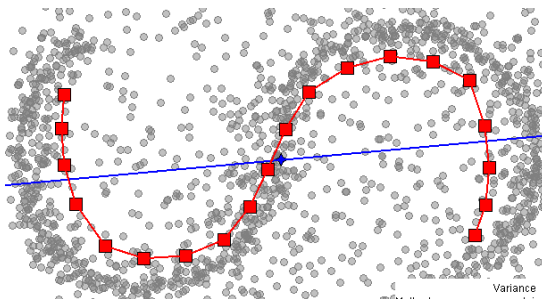
Интеллект, проявляемый машинами и **имитирующими когнитивные функции**, которые люди ассоциируют с другими людьми. К когнитивным функциям относятся все аспекты **восприятия, рассуждения, обучения и решения проблем**



Машинное обучение

Основной подход к реализации искусственного интеллекта путем обучения и прогнозирования на основе данных и опыта. Категории машинного обучения включают:

- “с учителем”,
- “без учителя”
- “с подкреплением”



Глубокое обучение

Ветви машинного обучения, соединяющие искусственные калькуляторы, которые **имитируют функцию нейронов в мозге**. Они образуют “нейронную сеть”, которая получает исходные данные, анализирует их, принимает решение, а затем получает информацию о том, правильно ли оно определено



Запрос бизнеса к ИИ в России

Янв. 2019

“” 3 лидера на рынке – Сбербанк, Яндекс, Mail.ru, остальные – игроки нишевые (ЦРТ, Касперский, АБВУУ)

Центр НТИ ИИ МФТ

Фев. 2019

Принята национальная программа по ИИ

Мар. 2019

“” 30% топ менеджеров РФ сказали, что активно внедряют ИИ

Исследование Microsoft

Дек. 2019

- “”
- 90% знают о технологиях ИИ
 - У 40% опрошенных получилось извлечь хотя бы минимальную прибыль
 - 60% не планируют по разным причинам
 - 70% заявляют о нехватке квалифицированных кадров, но наращивать компетенции не планируют

Обзор РБК

Авг. 2020

- “”
- Порочный круг бесплатных пилотов
 - Недостаточная вовлеченность бизнес руководства
 - Фокус на технологии, а не на задачах бизнеса
 - Недостаток инвестиций

Accenture

MLOps – инструмент эффективного внедрения искусственного интеллекта

1 Цифровой двойник

Классические BI инструменты для более точного управления бизнес-процессами, напр., создание виртуального прототипа реального объекта

2 Big Data технологии

Повышение эффекта от классических BI инструментов - становится все более значимым потенциал монетизации данных с использованием технологий BigData, продвинутой аналитики, машинного обучения

3 Сервисная модель - MLOps

Лавинообразный рост количества задач бизнеса к использованию решений с ИИ

Ключевые цели MLOps:

- Сокращение TTM (time-to-market)
- Масштабируемость

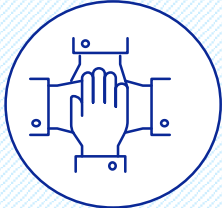
ИИ продукт – модель + ?

Бизнес заказчик



Продукт

**Бизнес процесс
бизнес эффект**



**Заказчик, Бизнес
аналитик, Data
Scientist**

Модель



Data Scientist

**Регламент
МаaS+Интеграция**



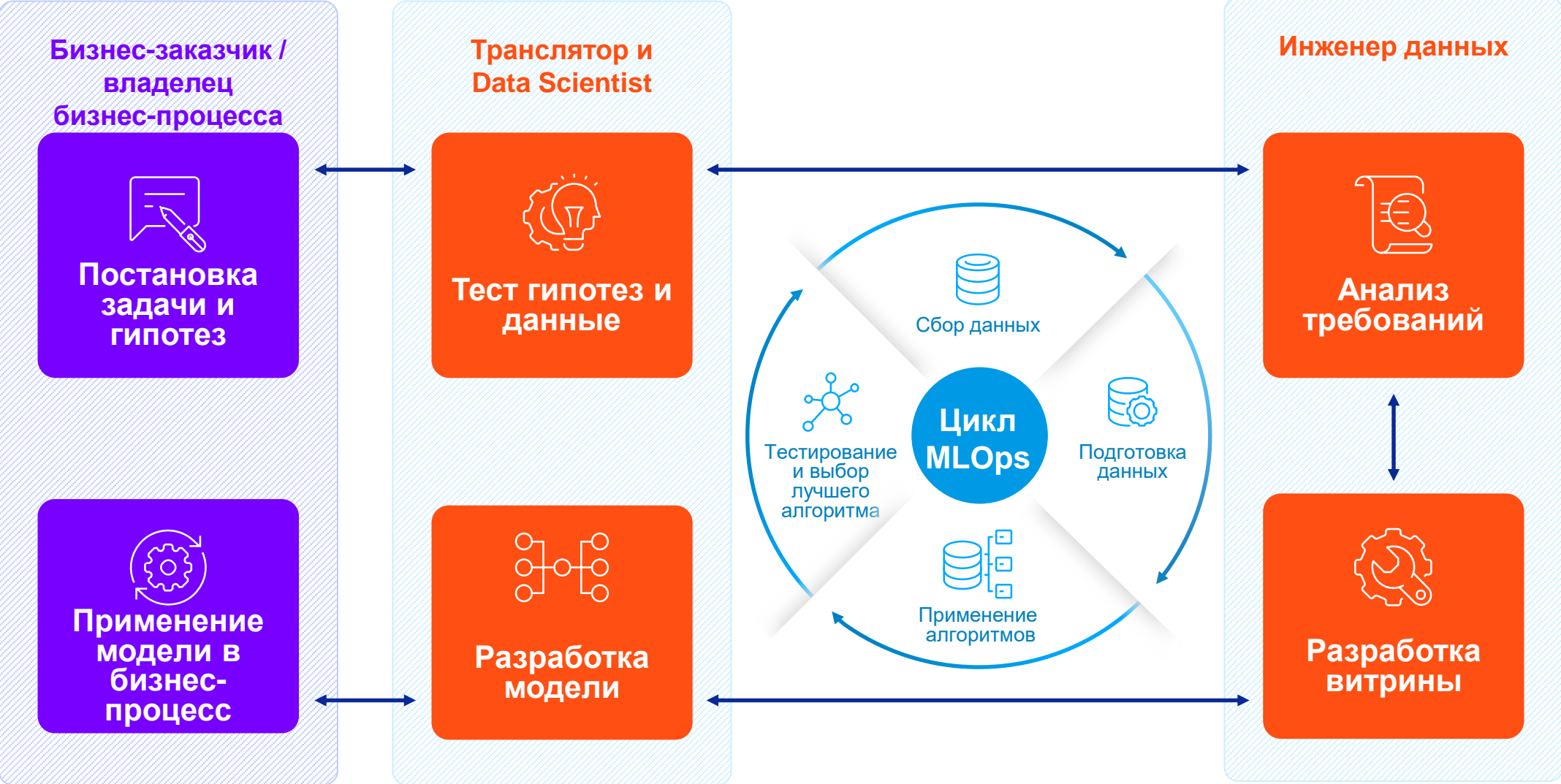
**ML Engineer, Data
Engineer**

Разработка решения – лишь часть процесса, необходимого для достижения бизнес-эффекта

Приоритетные этапы

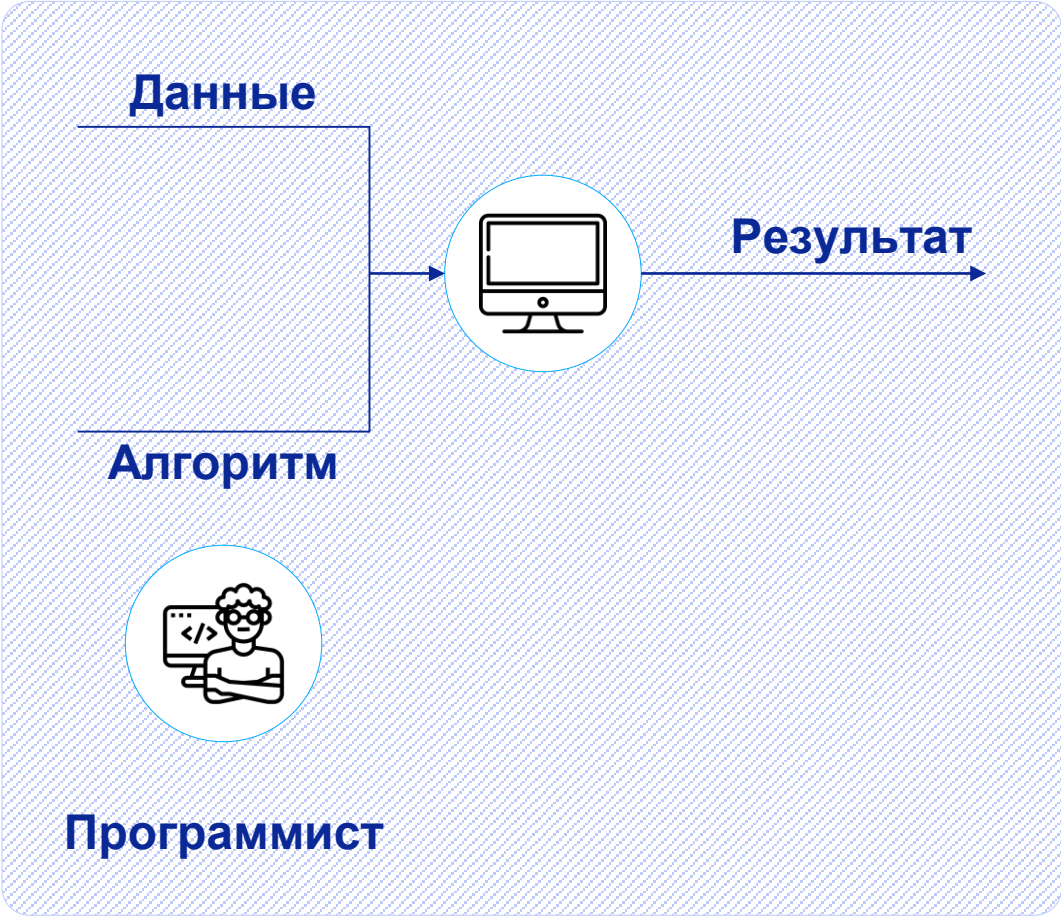


Интеграция моделей в бизнес-процессы



Модель = Программа?

Программирование



Машинное обучение



MLOps — модель ключевых компетенций



- Бизнес / Функция**
- A** — **Бизнес-заказчик** формирует требования для разрабатываемого решения, ответственен за интеграцию и реализацию бизнес-эффекта
 - B** — **Предметные эксперты (частичное вовлечение)** формируют бизнес-гипотезы и проверяют результаты аналитики со стороны бизнес-опыта
 - C** — **Конечные пользователи (частичное вовлечение)** предоставляют вводные по бизнес-процессу
- Подразделение аналитики**
- D** — **Транслятор** — проводит синтез выводов, ответственен за постановку задач для data scientist и инженеров данных на основании требований бизнес-заказчика
 - E** — **Data scientist** разрабатывает продвинутые статистические модели и алгоритмы
 - F** — **Инженер данных** собирает, структурирует и фильтрует данные для обеспечения их качества и доступности
 - G** — **Разработчик / инженер ML (опционально)** интегрирует решения, создает отчеты и информационные панели
 - H** — **UX дизайнер (опционально)** проектирует целевой пользовательский опыт

Способы сокращения среднего времени разработки и внедрения решений

Рычаги влияния

A

Технологии



Конвейер MLOps с автоматизацией

- Управление версиями кода модели
- Автоматизированное компонентное тестирование
- Автоматическое ведение документации по коду модели

B

Процессы и инструменты



Надлежащий процесс разработки моделей машинного обучения, соответствующий agile-принципам и существующим ролям

- Операционная модель в области машинного обучения, соответствующая agile-принципам
- Проверка кода коллегами

C

Развитие навыков



Обучение Data scientist-ов эффективным методам разработки ПО

- Создание кода, который можно повторно использовать и адаптировать
- Контроль версий ПО

Потенциальный эффект

Производительность рабочей группы по машинному обучению

 **50%**



Сроки вывода новой модели на рынок

 **25%**



Срок вывода нового релиза на рынок

 **50%**



Компоненты DevOps + модели



Выводы



Продуктом, создающим ценность для бизнеса является не только модель, но и технологическое решение по ее применению в бизнес процессе



Сервисная модель MLOps появляется в ответ на рост запроса бизнеса к ИИ



Ключевые цели создания MLOps для решения возрастающего потока задач - сокращение TTM (time-to-market) и масштабируемость

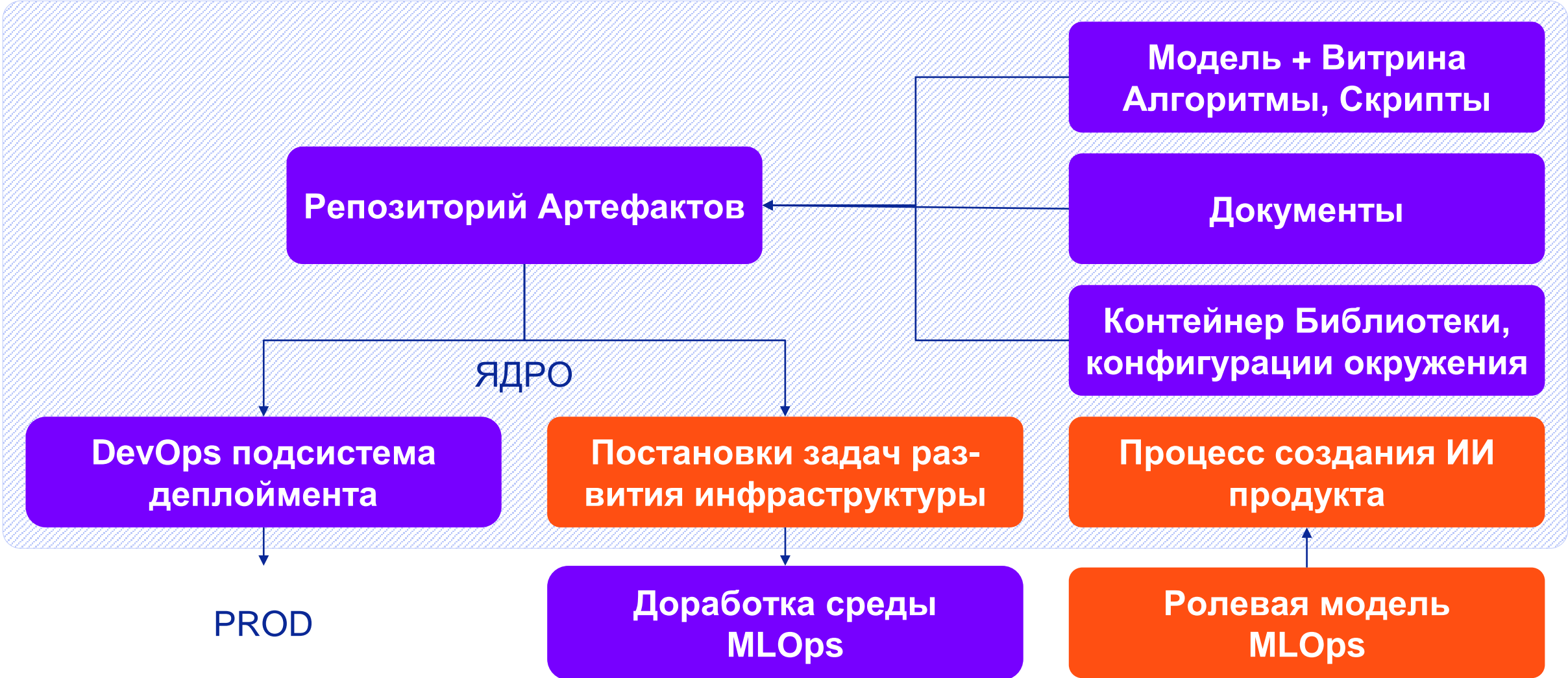


Сервисная модель (MLOps) определяет ключевые компетенции, участвующие в разработке моделей и типовой процесс взаимодействия между ними



Инструментарий для моделирования должен быть дополнен инструментарием для деплоя, основанным на DevOps стеке, адаптированном под специфику разработки моделей

Система управления MLOps

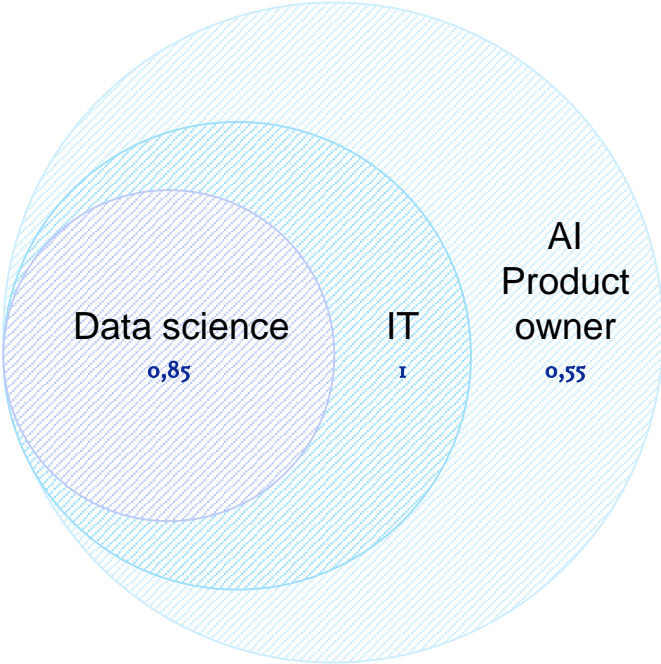


ИИ продукт — дилемма компетенций

xx Индекс Coursera 2019 г., max=1

Опция 1.

AI product owner с углубленными знаниями data science, IT и бизнеса



Опция 2.

Product owner обладает компетенциями транслятора между Data Scientist/IT и бизнесом

