

Data Science и инструменты анализа данных

2024/2025 учебный год

Доцент кафедры ИВЭ, Махно В.В.

©Создано при помощи <https://sberuniversity.ru/>



Структура курса

Модуль состоит из двух частей:

- Искусственный интеллект: понятие, структура научной области, история, интеграция в бизнес-процессы.
- Машинное обучение: задачи машинного обучения, постановка задач, работа с данными, алгоритмы и инструменты, примеры применения.

Понятие искусственного интеллекта

В самом широком смысле искусственным интеллектом (ИИ) называют способность компьютера решать те же интеллектуальные задачи, которые способен решать человек.

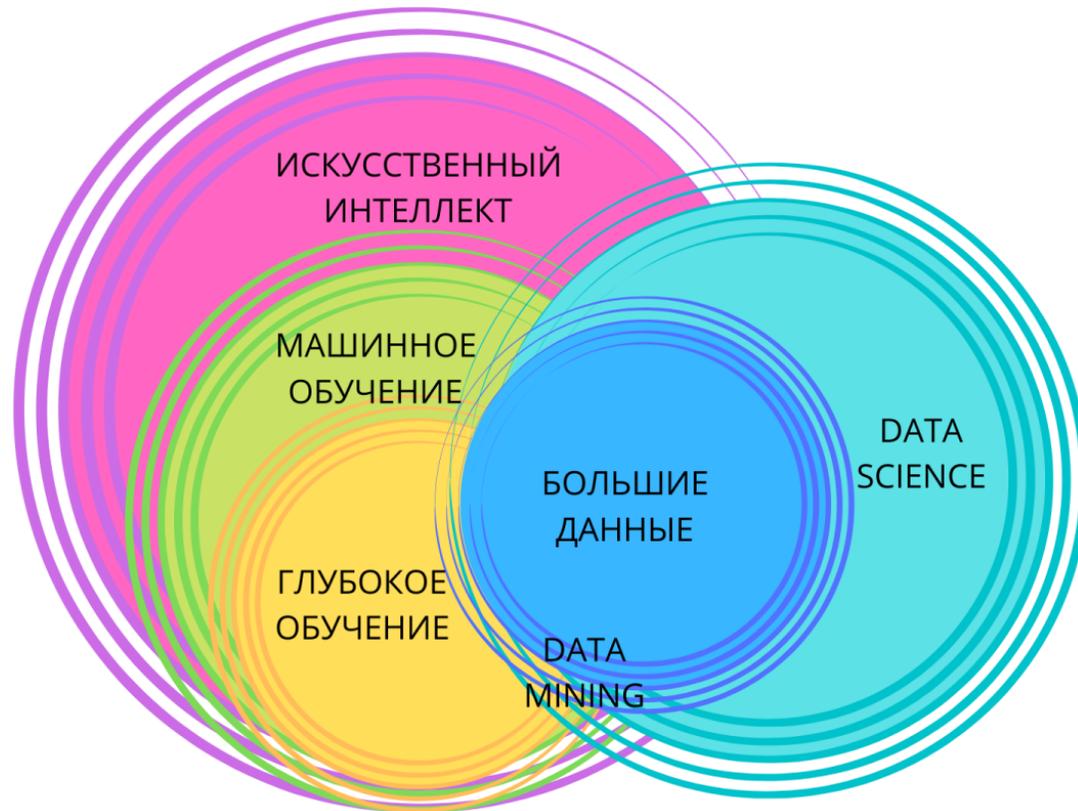


История искусственного интеллекта

- **1940-е: Тест Тьюринга** (Элиза, Джозеф Вейценбаум, 1966 г; «Женя Густман», Владимир Веселов, Санкт-Петербург, 2014 Г)
- **1950-е: Персептрон Розентблатта** («зима искусственного интеллекта»)
- **1980-е: Экспертные системы**
- **1990-е: Deep Blue** ([обыгравший в 1997 году Гарри Каспарова](#))
- **2000-е: Данные**
- **2010-е: Watson и DeepMind**
- **2020: GPT-3, AlphaFold**



Области искусственного интеллекта

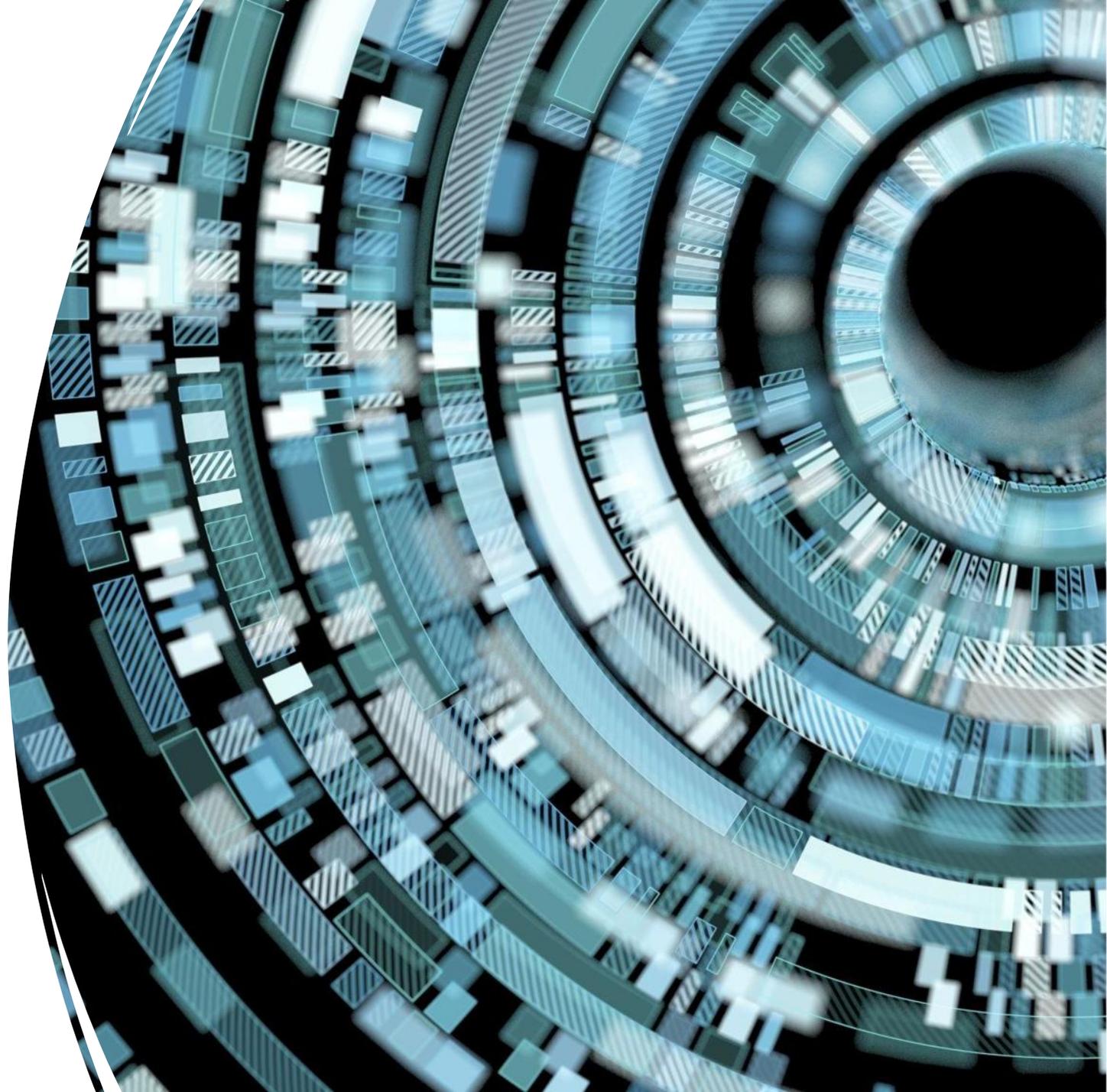


Области искусственного интеллекта

Машинное обучение

Глубинное обучение
(«глубокое обучение», от англ.
Deep learning)

- Data Science
- Data Mining
- Большие данные



Технологии работы с большими данными

Вместо последовательной обработки данных о миллиардах клиентов используют распределенное хранение данных и параллельную их обработку, например в [Hadoop](#) или [Spark](#).

Для построения моделей машинного обучения наиболее популярен язык программирования Python и его библиотеки Scikit-learn, LightGBM, CatBoost и другие. Для обучения нейронных сетей используются отдельные библиотеки: PyTorch и TensorFlow. Также существуют графические интерфейсы, например RapidMiner, но они предоставляют более ограниченный функционал по сравнению с Python и поэтому не так популярны.



Онлайн-курс СберУниверситета

Генеративное искусство

Подробнее о курсе



Бесплатный курс от Сбера
по генеративному
искусству

https://courses.sberuniversity.ru/generative-art?utm_source=tg&utm_medium=organic&utm_campaign=courses&utm_content=gen_i&utm_term=01_09_2023

Интеграция в бизнес-процессы

Принцип работы алгоритма машинного обучения следующий:

по большому количеству примеров вида вход — выход настраивают алгоритм, который сможет по входу предсказывать выход.

Процесс настройки алгоритма называется **обучением (learning)**.

Огромные массивы информации (наборов данных, или по-английски dataset — датасеты)



Направления применения ИИ

Автоматизация

ИИ успешно применяется в задачах, в которых вид входных и выходных данных всегда один и тот же, однако четкого алгоритма получения результата не существует



Прогнозирование

Речь идет о решении задач, для которых ответ неизвестен, но его можно спрогнозировать на основе исторических данных.





Кейсы применения ИИ

Кредитный скоринг

Уже сейчас при принятии решения для подавляющего количества кредитов используется ИИ, для этого алгоритм будет анализировать кредитную историю клиента и информацию о его доходах и тратах.

[AI Journey: Рекомендательные системы и системы поддержки принятия решений в кредитном скоринге](#)



Кейсы применения ИИ

Agro AI

Перед агробизнесом стоит ряд задач:

- Как повысить урожайность хозяйств;
- Как купить или арендовать хорошее поле;
- Как увеличить посевную площадь;
- Как оценить спрос и предложение.



Кейсы применения ИИ

Борьба с мошенничеством

Технологии ИИ позволяют более точно и быстро детектировать подозрительные операции и блокировать их. Для создания таких алгоритмов можно использовать как данные о реальных попытках мошенничества, так и строить модели поведения пользователя, основываясь только на «чистых» данных.



Кейсы применения ИИ

Робот-юрист

Юристы занимаются вычиткой документов и тратят на рутинный процесс большое количество времени.

Алгоритмы машинного обучения находят в тексте имена, адреса, число голосов, наименование юр. лица и связывают эти фрагменты информации друг с другом.

Кейсы применения



Чат-боты



Для упрощения взаимодействия клиента с банком используется автоматический сервис, задавать вопросы которому можно текстом или голосом.



[AI Journey: DialoGPT для генерации диалогов.](#)

Кейсы применения

Снижение аварийности на транспорте с помощью компьютерного зрения

Пассажирский транспорт мегаполиса входит в группу повышенного риска в связи с внушительным пассажиропотоком, плотностью движения и высокой нагрузкой на водителя.

70% дорожно-транспортных инцидентов происходит из-за потери внимания человеком.



Дополнительные кейсы для изучения

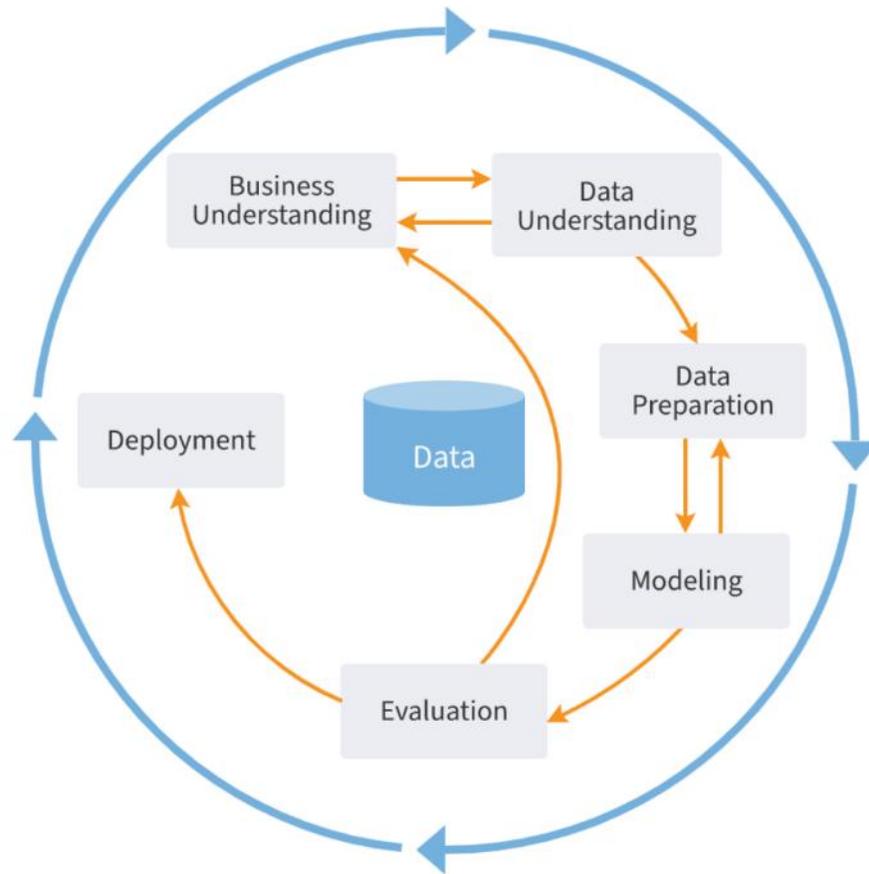
- промышленность: настройка оборудования под производство конкретных объектов, автоматическая диагностика оборудования и прогнозирование сбоев, контроль производственных процессов;
- торговля: предсказание спроса, разработка персонализированных программ лояльности;
- медицина: автоматическая диагностика и расшифровка результатов исследований, автоматизация составления медицинских отчетов и рекомендаций пациенту;
- транспорт: прогнозирование спроса, диагностика транспортных средств, маркетинговые компании;
- общепит, телеком



Условия применения ИИ

- Данные
- Ресурсы
- Специалисты
- Метрики

Методология управления проектами по анализу данных



Бизнес анализ
(Business Understanding)

Анализ данных
(Data Understanding)

Подготовка данных
(Data Preparation)

Моделирование
(Modeling)

Оценка решения
(Evaluation)

Внедрение
(Deployment)