

Сквозные технологии цифровой экономики. Искусственный интеллект и большие данные

Информационные технологии в цифровой экономике. Лекция 8

К.Т.Н., доцент
Буряченко Владимир Викторович
buryachenkoVV@gmail.com

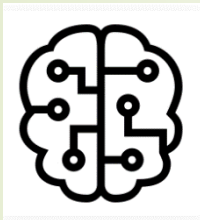
СибГУ им. М.Ф. Решетнева, 2022



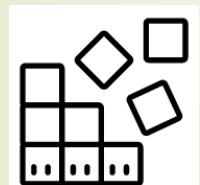
2

0101
1001

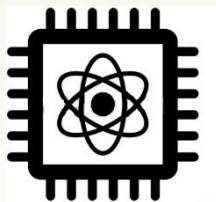
большие данные



**нейротехнологии и
искусственный интеллект**



**системы распределенного
реестра**



квантовые технологии



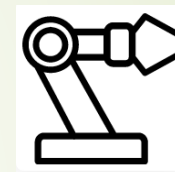
**новые производственные
технологии**



**промышленный
интернет**



**технологии беспроводной
связи**



**компоненты робототехники
и сенсорики**



**технологии виртуальной и
дополненной реальностей**

*Представлены в программе "Цифровая экономика Российской Федерации" (утверждена в 2017 г.)
В рамках федерального проекта «Цифровые технологии» (национальная программа "Цифровая экономика Российской Федерации",
утверждена в 2018 г.) разработаны дорожные карты* сквозных технологий.*

**Кроме дорожных карт «Большие данные» и «Промышленный интернет»*

Основные сквозные технологии цифровой экономики



3

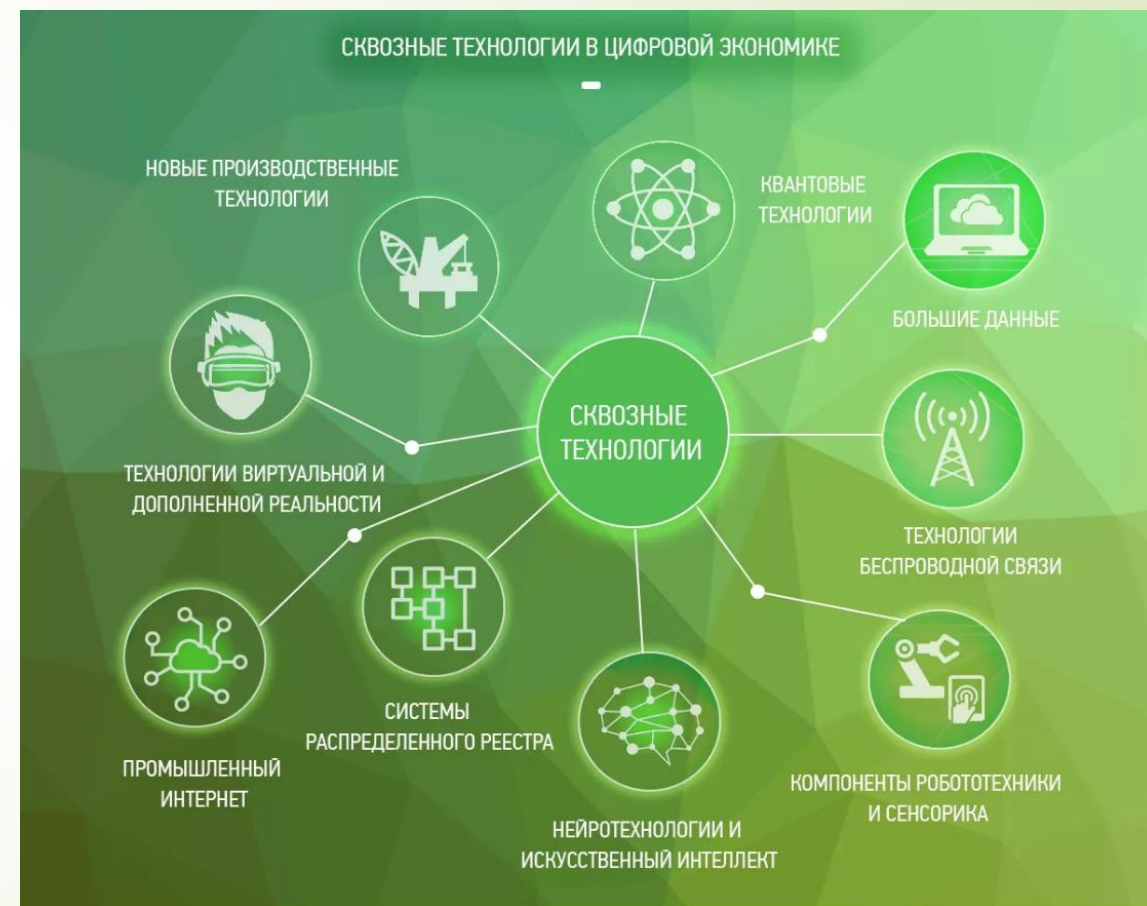
- **Сквозные технологии** универсальны, используются не только в частном (коммерческом), но и в государственном секторе экономики.
- Применение сквозных технологий является одной из профессиональных компетенций участника команды цифровой трансформации в государственном управлении.
- **Государственная поддержка** по стимулированию развития сквозных технологий осуществляется в рамках федерального проекта «**Цифровые технологии**» национальной программы «цифровая экономика РФ».
- Также поддержка исследований и бизнеса по направлениям сквозных технологий осуществляется в рамках проекта «**Наука. Технологии. Инновации**». **Инициатива НТИ.**



4

Основные сквозные технологии цифровой экономики

- Нейротехнологии и искусственный интеллект,
- Большие данные
- Системы распределённого реестра (блокчейн),
- Виртуальная и дополненная реальности
- Квантовые технологии,
- Новые производственные технологии,
- Промышленный интернет,
- Робототехника и сенсорика,
- Беспроводная связь.

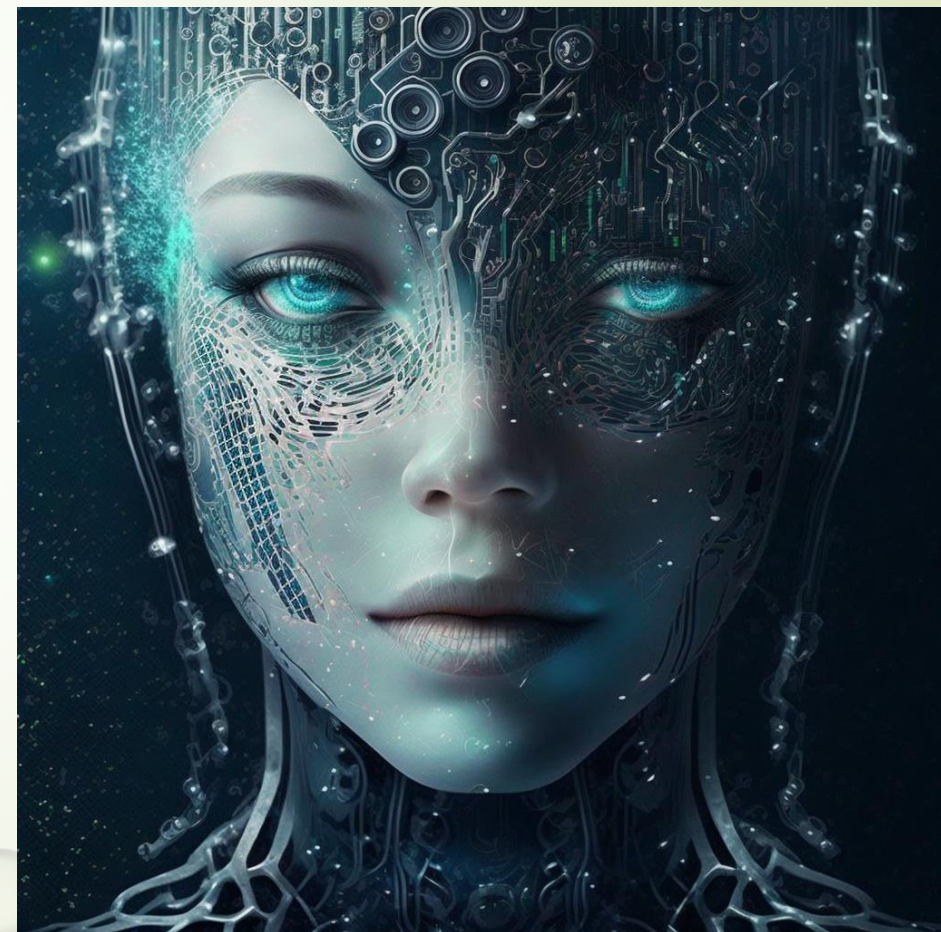




5

Нейротехнологии и Искусственный интеллект (ИИ)

- **Искусственный интеллект (ИИ)** — комплекс технологических решений, **имитирующий когнитивные функции человека** (включая самообучение и поиск решений без заранее заданного алгоритма) и позволяющий при выполнении задач достигать результаты, как минимум сопоставимые с результатами интеллектуальной деятельности человека.
- Комплекс технологических решений включает информационно-коммуникационную инфраструктуру, программное обеспечение, в котором в том числе используются методы **машинного обучения**, процессы и **сервисы по обработке данных и принятию решений**.
- **Нейротехнологии** – технологии, которые **используют или помогают понять работу мозга, мыслительные процессы, высшую нервную деятельность**, в том числе технологии по **усилению, улучшению работы мозга** и психической деятельности.





6

Нейротехнологии и Искусственный интеллект (ИИ)

Основными тенденциями развития рынка искусственного интеллекта являются:

- Развитие роботизированных сервисов с помощью ИИ и избавление от человеческого фактора и освобождения человека от монотонной работы путем автоматического создания программного обеспечения.
- Расширение вычислительных и функциональных возможностей программных продуктов.
- Новые методы машинного обучения, которые ускоряют разработку и реализацию решений в области ИИ в условиях ограниченного количества данных.
- Повсеместное применение ИИ.

Основными драйверами рынка искусственного интеллекта являются:

- Увеличение объема данных для анализа и повышение доступности данных надлежащего качества.
- Развитие вычислительной архитектуры следующего поколения.
- Развитие перспективных методов анализа данных.



Нейротехнологии и Искусственный интеллект (ИИ)

В российской федерации программными нормативными документами по ИИ являются:

- **Национальная стратегия развития искусственного интеллекта** на период до 2030 года
- **Дорожная карта развития СЦТ нейротехнологии и искусственный интеллект (ИИ)**

Оператор разработки дорожной карты - сбербанк. Искусственный интеллект может применяться для реализации новых возможностей человека во всех сферах деятельности, в том числе для:

- Освобождения человека от монотонной работы путем автоматического создания программного обеспечения;
- Поддержки в принятии решений;
- Автоматизации опасных видов работ;
- Поддержки коммуникаций между людьми.

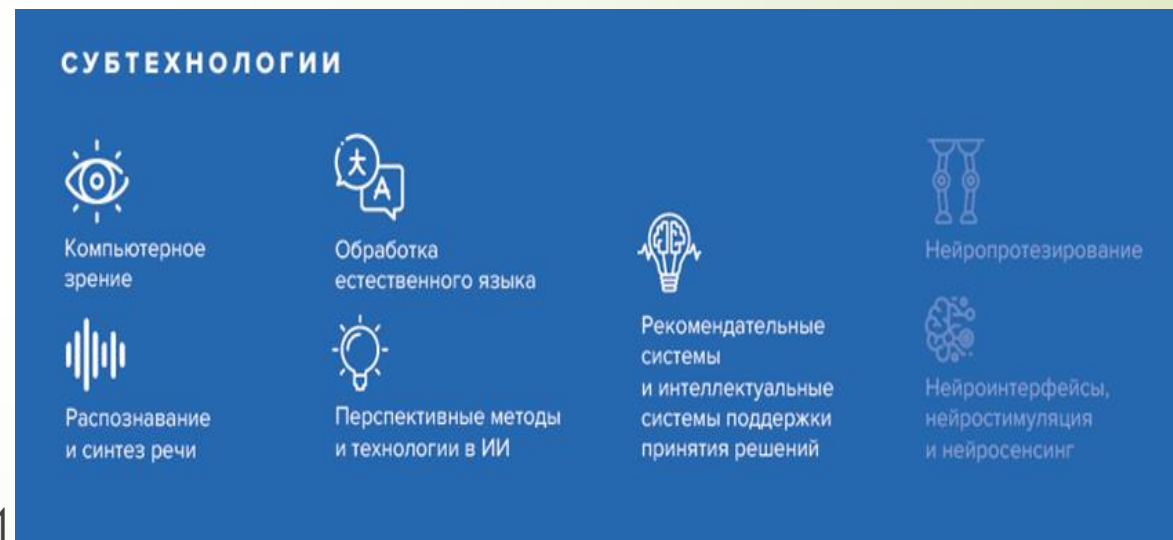
Применение искусственного интеллекта необходимо во всех экономических и социальных отношениях для повышения качества жизни и улучшения благосостояния общества.



Нейротехнологии и Искусственный интеллект (ИИ)

Субтехнологии

- Компьютерное зрение
- Обработка естественного языка
- Распознавание и синтез речи
- Системы поддержки принятия решений
- Перспективные методы и технологии в ИИ
- *Нейропротезирование и нейроинтерфейсы (пока исключены из приоритета)*





Компьютерное зрение

- Область искусственного интеллекта (ИИ), которая позволяет компьютерам и системам получать значимую информацию из цифровых изображений, видео и других визуальных входов - и предпринимать действия или давать рекомендации на основе этой информации.

Некоторые сферы применения:

- Системы технического зрения.
- Распознавание образов.
- Обработка визуальной информации.
- Анализ поведения людей.
- Генерация контента.

Обработка естественного языка



- Отрасть искусственного интеллекта, которая помогает компьютерам понимать, интерпретировать и манипулировать человеческим языком.
- **Обработка естественного языка (natural language processing, NLP) - общее направление искусственного интеллекта и математической лингвистики.** Оно изучает проблемы компьютерного анализа и синтеза текстов на естественных языках.
- *Примеры разработчиков и решений в рамках субтехнологии «обработка естественного языка»:*
 - *ABBYY — российская компания-разработчик решений в области распознавания текстов (OCR) и лингвистики. Наиболее известные продукты в данной области — система потокового ввода данных ABBYY flexicapture и анализа/понимания текста ABBYY compero;*
 - *Алиса — виртуальный голосовой помощник, созданный компанией «Яндекс». Алиса распознает естественную речь, имитирует живой диалог, дает ответы на вопросы пользователя и, благодаря запрограммированным навыкам, решает прикладные задачи;*



Распознавание и синтез речи

- Использование компьютерного оборудования и программных технологий для **идентификации и обработки человеческого голоса**, а также для **искусственного моделирования** речи человека.
- **Распознавание речи** - использование компьютерного оборудования и программных технологий для **идентификации и обработки человеческого голоса**. В основном используется для преобразования произнесенных слов в компьютерный текст. Кроме того, автоматическое распознавание речи используется для **аутентификации пользователей** с помощью их голоса и выполнения действий на основе инструкций, определенных человеком (Techopedia).
- **Синтез речи** - это искусственное моделирование речи человека с помощью компьютера или другого устройства. В отличие от распознавания голоса, синтез речи в основном используется для перевода текстовой информации в аудиоинформацию и в приложениях, таких как службы с поддержкой голоса и мобильные приложения. Помимо этого, оно также используется в вспомогательных технологиях для помощи людям с нарушениями зрения в чтении текстового контента.
 - Алиса/yandex.Speechkit — голосовой помощник от компании «яндекс», умеет распознавать речь человека, вести простые разговоры, управлять большим количеством ПО. Yandex.Speechkit – набор инструментов для распознавания и синтеза речи, позволяющий сторонним разработчикам создавать свои приложения.



Рекомендательные системы и интеллектуальные системы поддержки принятия решений

- Класс решений, обеспечивающий выполнение процесса без участия человека, поддержку в выборе решений, а также предсказание объектов, которые будут интересны пользователю.
- **Системы поддержки принятия решений (СППР)** – компьютерные автоматизированные системы, целью которых является помощь людям, **принимающим решение в сложных условиях**, для полного и объективного анализа предметной деятельности. СППР возникли в результате слияния управленческих информационных систем и систем управления базами данных.

Основные виды технологий, используемых для построения систем поддержки принятия решения (рекомендательных систем):

1. Контентно-ориентированные рекомендательные системы.
2. Рекомендательные системы, основанные на коллаборативной фильтрации.
3. Экспертные системы.



Нейропротезирование и нейроинтерфейсы

- Класс решений, позволяющих **отслеживать и влиять на нервную систему человека, посредством инвазивных и неинвазивных методов.**
- В рамках данной суб-сцт было выделено три перспективных задачи/технологии (например, «технологии распознавания потребительской информации», «технологии обработки и интерпретации данных о состоянии пользователя при помощи нейрофизиологии» и другие), а также три перспективных target use-cases (например, «системы нейромаркетинга», «системы контроля состояния пользователя» и другие).

Примеры разработчиков и решений в рамках суб-сцт «нейропротезирование и нейроинтерфейсы»:

- Экзоатлет разрабатывает продукты в области **протезирования и экзоскелетов.** Их продукты помогают людям с ограниченными возможностями, а также улучшают физические способности пользователя.

Примеры разработчиков и решений в рамках суб-сцт «нейросенсинг и нейростимуляция»:

- Викиум разрабатывает программное обеспечение для **отслеживания, анализа, визуализации мозговой активности** и усиления когнитивных способностей пользователя.

Согласно протоколу заседания наблюдательного совета АНО «цифровая экономика» от 24 мая 2019г. №13 было рекомендовано исключить нейротехнологии (нейропротезирование, нейроинтерфейсы, нейростимуляция и нейросенсинг) из перечня приоритетных субтехнологий и продолжить их развитие в рамках национальной технологической инициативы.

Большие данные (Big data)

Основные свойства больших данных

- **Объем.** Из названия «большие данные» становится понятно, что они содержат в себе много информации. И это действительно так: компании могут ежедневно получать десятки терабайт различных данных, некоторые — сотни петабайт. То есть «большие данные» не были бы таковыми без объема.
- **Скорость.** Большие данные поступают и обрабатываются из разных источников с высокой скоростью. При отсутствии этого свойства информацию уже нельзя будет назвать «big data». А еще они генерируются без остановки.
- **Разнообразие.** Большие данные содержат в себе информацию, относящуюся к разным типам. Это одно из главных отличий от простых данных — они всегда структурированы и могут быть сразу сохранены в базе данных.





15 Этапы работы с большими данными

► 1 этап. Интеграция

На этом этапе компания интегрирует в свою работу технологии и системы, позволяющие собирать большие объемы информации из разных источников. Внедряются **механизмы обработки и форматирования данных** для упрощения работы аналитиков с «big data».

► 2 этап. Управление

Полученные данные нужно где-то хранить, этот вопрос решается до начала работы с ними. Решение принимается на основе множества критериев, главными из которых считаются предпочтения по формату и технологии обработки. Как правило, для хранения компании используют **локальные хранилища, публичные или частные облачные сервисы**.

► 3 этап. Анализ

Большие данные начинают приносить пользу после анализа. Это заключительный этап взаимодействия с ними. Для этого применяют **машинное обучение, ассоциацию правил обучения, генетические алгоритмы и другие технологии**. После анализа данных остается только самое ценное для бизнеса.



Что такое BIG DATA?

Различают структурированные и неструктурированные данные.

Структурированные данные четко упорядочены и поступают из установленных источников в единообразных форматах.

Для их обработки легко составить корреляционные зависимости и посредством принятых алгоритмов заставить машину проводить анализ данных под поставленную задачу.

Неструктурированные данные — неупорядоченные данные, поступающие из различных источников, не всегда заведомо установленных, в различных форматах, что усложняет процесс их сбора, обработки, сопоставления и анализа.

Как следствие — возникновение проблем потери времени и валидации результатов.



Определения BIG DATA

- **Big data** — это многоплановое и многоуровневое понятие, основными характеристиками которого являются объем данных, скорость их генерации и доступности, разнообразие источников и типов данных, форм их хранения.

Таким образом под **big data** нужно понимать не какой-то конкретный объём данных и даже не сами данные, а **методы их обработки, которые позволяют распределенно обрабатывать информацию.**

Эти методы можно применить как к огромным массивам данных (таким как содержание всех страниц в интернете), так и к небольшим объемам данных.

Источники BIG DATA



Несколько примеров того, что может быть источником данных, для которых необходимы методы работы с большими данными:

- Gps-сигналы от автомобилей для транспортной компании
- Данные, снимаемые с датчиков в большом адронном коллайдере
- Оцифрованные книги в российской государственной библиотеке
- Информация о транзакциях всех клиентов банка
- Информация о всех покупках в крупной ритейл сети и т.Д.

Количество источников данных стремительно растёт, а значит технологии их обработки становятся всё более востребованными.

Принципы работы BIG DATA



1. **Горизонтальная масштабируемость.** Поскольку данных может быть сколь угодно много – любая система, которая подразумевает обработку больших данных, должна быть расширяемой.

2. **Отказоустойчивость.** Принцип горизонтальной масштабируемости подразумевает, что машин в кластере может быть много. Это означает, что часть этих машин будет гарантированно выходить из строя. Методы работы с большими данными должны учитывать возможность таких сбоев и переживать их без каких-либо значимых последствий.

Например, кластеры google имеют более 100 000 машин.

3. **Локальность данных.** По возможности обрабатывать данные следует на той же машине, на которой они хранятся.

В больших системах данные распределены по большому количеству машин. Если данные физически находятся на одном сервере, а обрабатываются на другом – расходы на передачу данных могут превысить расходы на саму обработку.



20

Методы работы с большими данными

- **Association rule learning.** Набор методик для выявления взаимосвязей, т.е. Ассоциативных правил, между переменными величинами в больших массивах данных. Используется в **data mining**.
- **Classification.** Набор методик, которые позволяет предсказать поведение потребителей в определенном сегменте рынка (принятие решений о покупке, отток, объем потребления и проч.). Используется в **data mining**.
- **Cluster analysis.** Статистический метод классификации объектов по группам за счет выявления наперед не известных общих признаков. Используется в **data mining**.
- **Data fusion and data integration.** Набор методик, который позволяет анализировать комментарии пользователей социальных сетей и сопоставлять с результатами продаж в режиме реального времени.
- **Data mining.** Добыча данных, интеллектуальный анализ данных, глубинный анализ данных) — собирательное название, используемое для обозначения совокупности методов обнаружения в данных ранее неизвестных, нетривиальных, практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности.



21 Методы работы с большими данными

- **Genetic algorithms.** В этой методике возможные решения представляют в виде `хромосом`, которые могут комбинироваться и мутировать. Как и в процессе естественной эволюции, выживает наиболее приспособленная особь.
- **Machine learning.** Направление в информатике (исторически за ним закрепилось название `искусственный интеллект`), которое преследует цель создания алгоритмов самообучения на основе анализа эмпирических данных.
- **Natural language processing (nlp).** Набор заимствованных из информатики и лингвистики методик распознавания естественного языка человека.
- **Network analysis.** Набор методик анализа связей между узлами в сетях. Применительно к социальным сетям позволяет анализировать взаимосвязи между отдельными пользователями, компаниями, сообществами и т.П.
- **Optimization.** Набор численных методов для редизайна сложных систем и процессов для улучшения одного или нескольких показателей. Помогает в принятии стратегических решений, например, состава выводимой на рынок продуктовой линейки, проведении инвестиционного анализа и проч.



22 Методы работы с большими данными

- **Predictive modeling.** Набор методик, которые позволяют создать математическую модель наперед заданного вероятного сценария развития событий. Например, анализ базы данных CRM-системы на предмет возможных условий, которые подтолкнут абонента сменить провайдера.
- **Regression.** Набор статистических методов для выявления закономерности между изменением зависимой переменной и одной или несколькими независимыми. Часто применяется для прогнозирования и предсказаний. Используется в data mining.
- **Sentiment analysis.** В основе методик оценки настроений потребителей лежат технологии распознавания естественного языка человека. Они позволяют вычлнить из общего информационного потока сообщения, связанные с интересующим предметом (например, потребительским продуктом). Далее оценить полярность суждения (позитивное или негативное), степень эмоциональности и проч.
- **Supervised learning.** Обучение с учителем. Набор основанных на технологиях машинного обучения методик, которые позволяют выявить функциональные взаимосвязи в анализируемых массивах данных за счёт обучения на заранее подготовленных и размеченных наборах данных.

23 Методы работы с большими данными



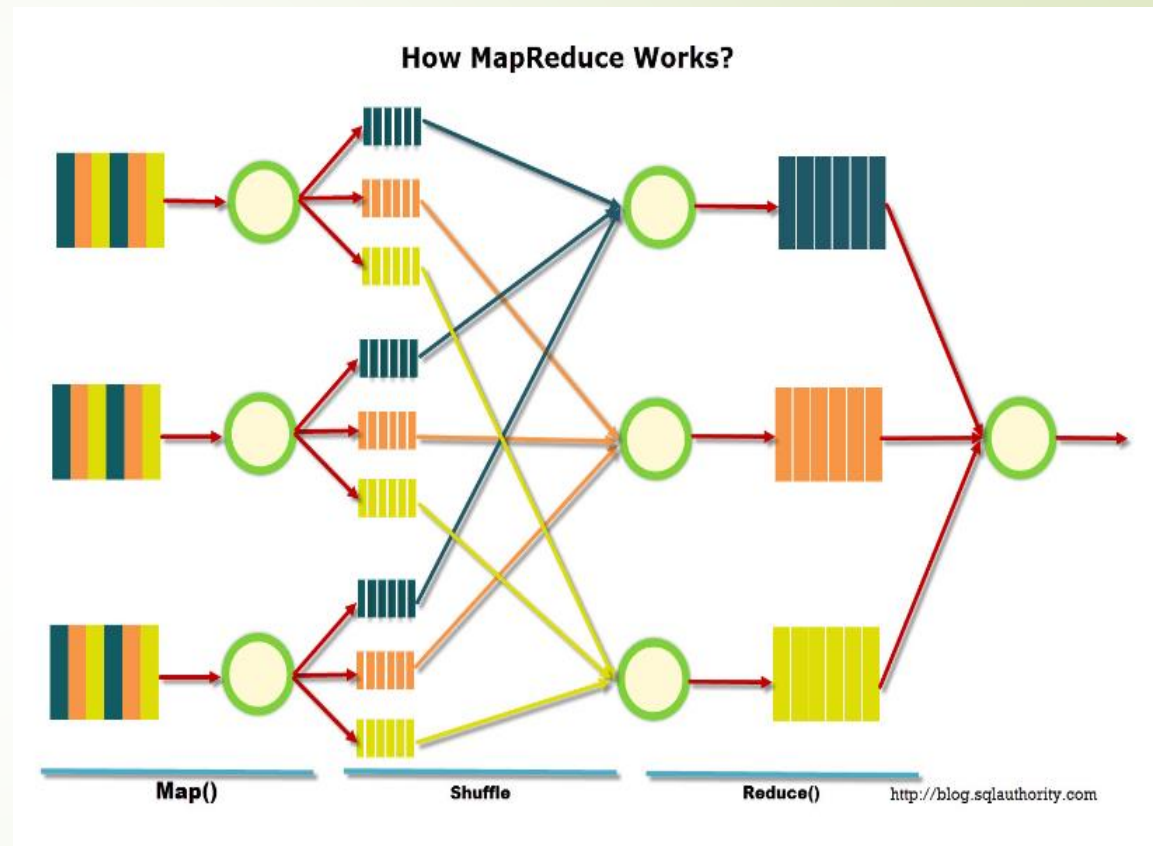
Один из методов обработки данных:

MapReduce — модель распределённых вычислений, представленная компанией Google, используемая для параллельных вычислений над очень большими, вплоть до нескольких петабайт, наборами данных в компьютерных кластерах.

Map-reduce предполагает, что данные организованы в виде некоторых записей.

Обработка данных происходит в 3 стадии:

Map, shuffle, reduce.



методы работы BIG DATA



1. **Стадия map.** На этой стадии данные преобразовываются при помощи функции map, которую определяет пользователь.

Работа этой стадии заключается в преобработке и фильтрации данных с целью получения только важных для дальнейшего анализа и хранения данных. Например из всего массива больших данных формируются пары «Атрибут-значение», а остальные данные отбрасываются.

2. **Стадия shuffle (перетасовка).** Проходит незаметно для пользователя.

В этой стадии вывод функции map «разбирается по корзинам» – каждая корзина соответствует одному ключу вывода стадии map. В дальнейшем эти корзины послужат входом для reduce.

3. **Стадия reduce (сводить).** Каждая «корзина» со значениями, сформированная на стадии shuffle, попадает на вход функции reduce.

Функция reduce задаётся пользователем и вычисляет финальный результат для отдельной «корзины».

Множество всех значений является финальным результатом mapreduce-задачи.



Большие данные и аналитика

Big data меняет многие принципы управления и работы с потребителями, способствуют созданию новых цепочек формирования стоимости.

Сейчас проблема перемещается от big data в сторону **fast data** — обработки потока данных в реальном масштабе времени, а также **smart data** — интеллектуальных данных, которые помогают понять суть или смысл проблемы, явления, сообщения на основе семантического (смыслового) анализа.



Большие данные и аналитика

В сфере аналитики данных выделяют несколько последовательных стадий:

- 1) **descriptive analytics** — дескриптивная, описательная аналитика;
- 2) **diagnostic analytics** — аналитика, связанная с распознаванием образов, определением того, что представляет собой объект анализа или воздействия;
- 3) **predictive analytics** — предиктивная, предсказательная, прогнозная аналитика, которая ориентирована на формирование прогноза данных или событий будущего на основе использования методов математической статистики, анализа данных, теории игр, моделирования (выдвигается гипотеза и проверяется, что произойдет, если...?) и других методов обработки накопленных данных.



Большие данные и аналитика

- 4) **prescriptive analytics** — прескриптивная, предписывающая аналитика. Она позволяет предупреждать о каких-то явлениях, выдавать рекомендации к действию, предписания.
- **Prescriptive analytics** основана на статистике прошлого и информации, получаемой в режиме реального времени. В рамках этого подхода строятся модели будущего и в зависимости от полученных результатов вырабатываются решения, которые принимаются без участия человека.



Большие данные и аналитика

- Несколько лет назад появился тренд, названный **perceptual computing (перцептивные вычисления)**.
- В рамках этого тренда разрабатываются технологии и устройства, способные по голосу, жестам и даже поведению владельца определять, что именно ему требуется.
- Другое направление в рамках этого тренда — создание систем информационной безопасности на основе отслеживания поведения потребителя, например владельца кредитной карты, для выявления отклонений, которые могут быть связаны с несанкционированным доступом к карте.



29

ГЛАВНЫЕ ПРОБЛЕМЫ ЭТИКИ (Big Data)

- 16 января 2020 года центр подготовки руководителей цифровой трансформации, сформированный на базе **высшей школы государственного управления РАНХИГС**, сообщил, что совместно с **Высшей школой экономики** составил доклад о главных этических проблемах и рисках, вызываемых в России цифровыми технологиями, а также возможных путях их решения.
- В докладе эксперты представили главные проблемы этики работы с данными, **этики Искусственного Интеллекта и Интернета Вещей**.



30

ГЛАВНЫЕ ПРОБЛЕМЫ ЭТИКИ (Big Data)

Пользователи почти никогда не знают:

- **Какие данные** о них собираются;
- **У кого** и насколько надежно они хранятся;
- Сколько компаний, людей и государственных органов каких именно стран **имеют доступ** или возможность получить доступ **к ЭТИМ ДАННЫМ**;
- **Какие их социальные связи с** какими другими людьми прослеживаются;
- **Как удалить свои данные** в базах компаний, отозвать свое согласие.

Государство является владельцем большого количества данных граждан и организаций и в перспективе может собирать их еще больше. При этом, государство отстраняется от законодательного ограничения бизнеса при обороте данных, предпочитая использовать собранные данные о гражданах в своих целях.



31

ГЛАВНЫЕ ПРОБЛЕМЫ ЭТИКИ (Работа с данными)

Некорректная работа с данными государства, как утверждают эксперты, может быть чревата следующими рисками:

- **Непреднамеренное создание алгоритмов** и систем принятия решений, дискриминирующих отдельные группы населения из-за недостаточно тщательного отбора данных для обучения алгоритмов;
- **Конфликт интересов общества и отдельных групп граждан** и общественное неприятие инициатив, в том числе из-за отсутствия информационной поддержки и предварительной работы с общественным мнением;
- **Принятие на государственном уровне неверных решений** с далеко идущими последствиями в виде финансовых, репутационных и иных потерь для государства;
- **Непреднамеренное нарушение законов** госслужащими по причине неэтичного использования данных, повлекшего за собой значительный ущерб.
- **Сложности и замедление развития технологий искусственного интеллекта** как результат регулирования и ограничения доступа к данным;
- **Утечки данных.**



32 ГЛАВНЫЕ ПРОБЛЕМЫ ЭТИКИ (ИИ)

По мнению экспертов, привнести этичность в саму технологию искусственного интеллекта сложно, так как моральный выбор не имеет четко обозначенных нормами закона границ.

К проблемам в сфере этики ИИ относятся следующие:

- **Ответственность** за этическое/**неэтичное** поведение искусственного интеллекта, за принятие ошибочных решений, ущерб из-за сбоев и т.п.;
- Решением данной проблемы считаются различные варианты: **от личной ответственности авторов до полного освобождения кого-либо от ответственности за действия искусственного интеллекта.**
- **Предвзятость** систем искусственного интеллекта в целом и программ-советчиков в частности;
- **Непрозрачность решений искусственного интеллекта и приватности** пользовательских данных при его применении;
- Одним из подходов к решению проблем приватности, прозрачности и предвзятости искусственного интеллекта эксперты считают **создание стандартов на разработку этически ориентированных интеллектуальных систем.**
- **Надежность** искусственного интеллекта.



33

ГЛАВНЫЕ ПРОБЛЕМЫ ЭТИКИ (IoT)

- Ключевыми проблемами интернета вещей авторы обозначают нарушение конфиденциальности и падение доверия к интернету в целом.

«Во многих случаях данные, используемые для аналитики, генерируются автоматически - например, путем отслеживания онлайн-активности, а не предоставляются людьми сознательно. Возникает проблема скрытого сбора данных, которые «могут быть использованы против вас». Пользователю приходится доверять устройствам интернета вещей, подключенным к сети, не зная при этом, собирают ли и передают ли устройства его данные или нет. Повышается риск утраты максимально чувствительных данных», - отмечают авторы доклада

- Принцип информированного согласия на сбор и обработку личных данных, полученных при использовании устройств и услуг интернета вещей, по мнению авторов доклада, неэффективен для защиты прав пользователей, так же как и согласие на обработку прочих данных.

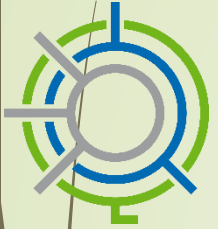


34

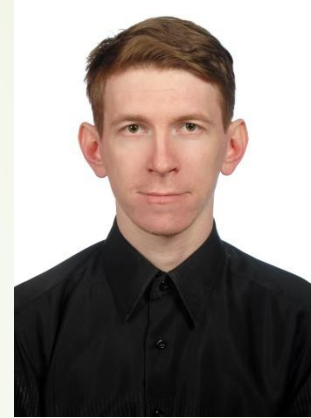
Направления решения этических проблем цифровых технологий

Это такие направления, как:

- **Обеспечение необходимого регулирования** (регулятор должен находить баланс между поддержкой технологических инноваций и необходимостью защищать общество от возможных рисков, связанных с их внедрением);
- **Создание этических кодексов и руководств** применительно как к определенным технологиям, так и к различным сферам деятельности;
- **Обеспечение доступности цифровых услуг** для граждан и создание «аналоговых интерфейсов»;
- **Включение этики в образовании** в области технологий.



Искусственный интеллект и большие данные



к.т.н., доцент, доцент каф. ИВТ
Буряченко Владимир Викторович
Кафедра ИВТ, Л304
BuryachenkoVV@gmail.com

СибГУ им. М.Ф. Решетнева, 2024