

Интеллектуальные технологии и машинное обучение в повышении информативности официальной статистики

Зарова Елена Викторовна, доктор экономических
наук. профессор



План курса

- 1. Введение в методологию и практику интеллектуальных технологий и машинного обучения для целей официальной статистики.**
- 2. Методы автоматизации сбора и обработки статистических данных.**
- 3. Применение машинного обучения для классификации и аналитической группировки в официальной статистике.**
- 4. Кластеризация и выявление скрытых закономерностей в официальной статистике.**
- 5. Использование генеративных моделей и технологий text mining в семантическом анализе текстовых данных.**
- 6. Геопространственная официальная статистика: методы машинного обучения и технологии ИИ.**
- 7. Этические, правовые и организационные аспекты внедрения интеллектуальных технологий в статистике.**

План лекции № 1

- 1. Актуальность курса**
- 2. Основные понятия: искусственный интеллект, машинное обучение, статистический ИИ**
- 3. Краткий обзор мировых и российских кейсов внедрения**



1. Актуальность курса

В эпоху цифровой трансформации государственного управления **интеграция искусственного интеллекта в официальную статистику** становится не просто технологическим новшеством, а **стратегической необходимостью** для обеспечения:

- **качественного анализа социально-экономических процессов** и принятия обоснованных решений **на всех уровнях власти,**
- **повышения качества удовлетворения информационных потребностей** бизнеса, СМИ, населения



КонсультантПлюс

Указ Президента РФ от 15.02.2024 N 124
"О внесении изменений в Указ Президента
Российской Федерации от 10 октября 2019 г.
N 490 "О развитии искусственного
интеллекта в Российской Федерации" и в
Национальную стратегию, утвержденную
этим Указом"

Связь с национальными стратегиями развития ИИ и статистики



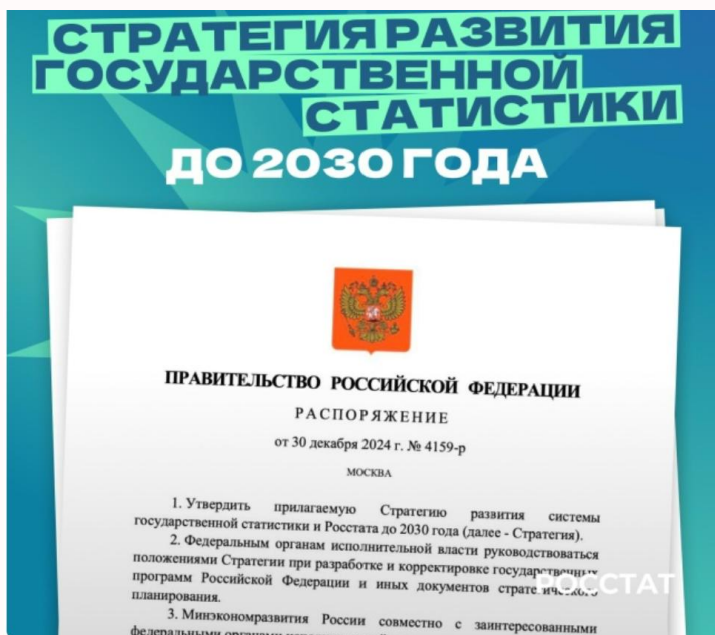
Указ Президента РФ от 15.02.2024 № 124

Обновлённая Национальная стратегия развития искусственного интеллекта до 2030 года от 15 февраля 2024 года устанавливает амбициозные цели по **масштабированию ИИ-технологий** во всех секторах экономики и государственного управления.



СТРАТЕГИЯ РАЗВИТИЯ СТАТИСТИКИ

Активное внедрение искусственного интеллекта (ИИ) для улучшения **качества и оперативности сбора, обработки и анализа статистических данных.**



СИНЕРГИЯ СТРАТЕГИЙ

Курс отвечает задачам обоих документов, предоставляя методологические и прикладные знания для внедрения методов МО и ИИ-технологий в систему государственной статистики для повышения качества, оперативности и глубины анализа данных.

Регулирование ИИ в Российской Федерации

Указ Президента Российской Федерации от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации» утверждена **Национальная стратегия развития искусственного интеллекта на период до 2030 года**

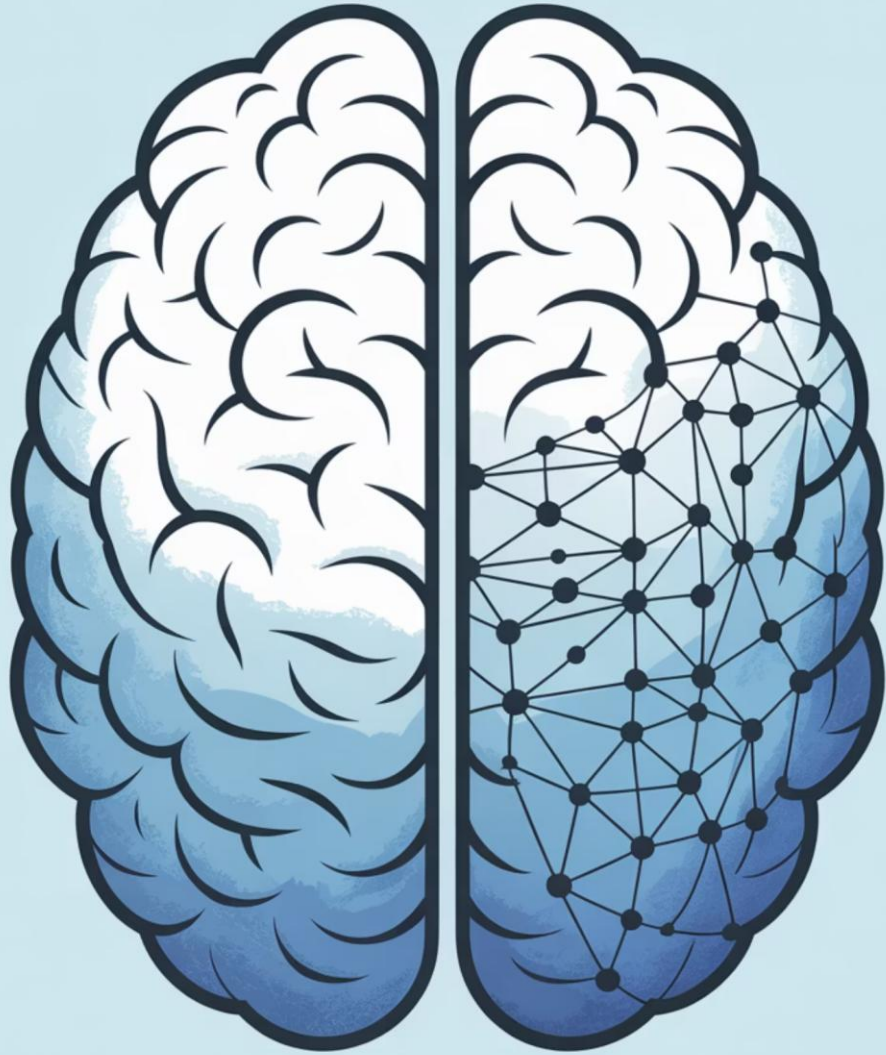
Указ Президента Российской Федерации от 15 февраля 2024 г. № 124 были **внесены изменения** как в указ 2019 года, так и в саму «Стратегию»



■ ОПРЕДЕЛЕНИЕ ИИ

«Искусственный интеллект — комплекс технологических решений, позволяющий **имитировать когнитивные функции человека (включая поиск решений без заранее заданного алгоритма)** и получать при выполнении конкретных задач **результаты, сопоставимые с результатами интеллектуальной деятельности человека или превосходящие их**. Комплекс технологических решений включает в себя информационно-коммуникационную инфраструктуру, программное обеспечение (в том числе **использующее методы машинного обучения**), процессы и сервисы по обработке данных и поиску решений»

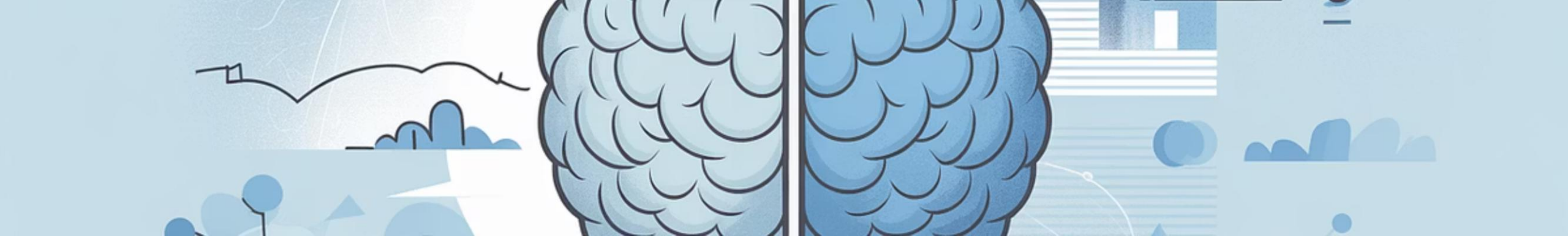
■ **ДОВЕРЕННЫЕ
ТЕХНОЛОГИИ
ИИ**



КОГНИТИВНЫЕ ФУНКЦИИ ЧЕЛОВЕКА И ИХ РОЛЬ В РАБОТЕ ИИ С БОЛЬШИМИ ДАННЫМИ

Когнитивные функции — это высшие психические процессы, позволяющие человеку воспринимать, обрабатывать и использовать информацию.

При работе с большим объемом данных понимание этих механизмов становится ключом к созданию **эффективных систем искусственного интеллекта**



Когнитивные функции человека: ключевые способности



Память

Хранение и воспроизведение информации, критична для анализа больших данных и **построения связей между разрозненными фактами**



Внимание

Выбор и **концентрация на релевантных данных** среди огромного объёма информационного шума



Критическое мышление

Оценка, анализ и синтез информации для принятия взвешенных и обоснованных решений

Логика и аналитика

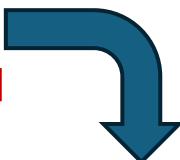
Выявление скрытых закономерностей, построение предсказательных моделей и систематизация больших объёмов данных

Креативность и интуиция

Генерация новых гипотез и нестандартных решений, которые ИИ пока не способен полноценно заменить без человеческого участия

Какие когнитивные функции являются **специфическими** для работников официальной статистики?



- 
- **Логическое и образно-теоретическое мышление**, которое позволяет анализировать, синтезировать и обобщать статистические данные для формирования выводов и рекомендаций.
 - **Чувствительность к информационным изменениям**, то есть способность фиксировать и адекватно реагировать на новые данные и изменения в статистической информации.
 - **Таксономичность** — умение систематизировать и классифицировать информацию, что важно при работе с большими массивами данных и построении статистических моделей.
 - **Внимание и память**, обеспечивающие точность и надежность в сборе и анализе данных, а также удержание в сознании ключевых параметров и методик.
 - **Исполнительные функции**, включающие планирование, организацию работы и контроль результатов анализа.
 - **Языковые и коммуникационные навыки**, необходимые для грамотного представления статистических данных и взаимодействия с другими специалистами.

Ключевые принципы

Доверенные технологии ИИ — это системы, обеспечивающие безопасность, прозрачность, этичность и надежность ИИ-решений для общества и бизнеса.

В России действуют ГОСТ Р 59276-2020 и ГОСТ Р 71539-2024, которые задают общие принципы и процессы жизненного цикла ИИ для повышения доверия к технологиям.

Продолжается проработка единого федерального закона "О системах искусственного интеллекта", который должен закрепить общие принципы, риски, маркировку, ответственность и процесс внедрения ИИ.



Безопасность данных

Защита персональной информации и предотвращение утечек



Прозрачность



Этичность










Соблюдение моральных норм и предотвращение манипуляций

Международный опыт: три подхода к доверию

Европейский союз и Евростат



«Европейский акт об ИИ» — первое в мире комплексное регулирование с жесткими требованиями к безопасности, прозрачности и контролю рисков.

- а          <

Вызовы внедрения ИИ в системы сбора и обработки данных



Указ Президента Российской Федерации от 15.02.2024 № 124 о внесении изменений в Национальную Стратегию развития ИИ на период до 2030 г.

Новые вызовы:

- ▶ нехватка вычислительных мощностей, отечественных решений в области ИИ (программно-аппаратные комплексы и электронная компонентная база);
- ▶ дефицит высококвалифицированных специалистов и инновационных разработок в области ИИ;
- ▶ низкий уровень внедрения технологий ИИ в госуправлении;
- ▶ нехватка кадров для обеспечения массового внедрения технологий ИИ;
- ▶ недостаточное субсидирование организаций в сфере ИИ, дефицит частных инвестиций;
- ▶ нормативные барьеры для внедрения технологий ИИ в отдельных отраслях экономики (отсутствие методологической базы для обеспечения систем ИИ достоверными исходными данными);
- ▶ необходимость обеспечения безопасности при разработке и использовании технологий ИИ;
- ▶ необходимость обеспечения защиты персональных данных и иной информации ограниченного доступа, объектов интеллектуальных прав при создании и обучении моделей ИИ;
- ▶ ограничение доступа к технологиям ИИ в связи с недобросовестной конкуренцией со стороны недружественных иностранных государств и введением ими односторонних ограничительных мер;
- ▶ дополнительные барьеры для развития международного сотрудничества в сфере ИИ

ФЕДЕРАЛ
ПРЕСС

!!! Нехватка кадров
для обеспечения
развития и
внедрения ИИ

Февраль 2025 - В Екатеринбурге прошло заседание круглого стола на тему «Искусственный интеллект во власти и бизнесе», который организовал медиахолдинг «ФедералПресс». <https://fedpress.ru/article/3363337>



Задачи по внедрению ИИ в соответствии со Стратегией развития государственной статистики до 2030 года

“

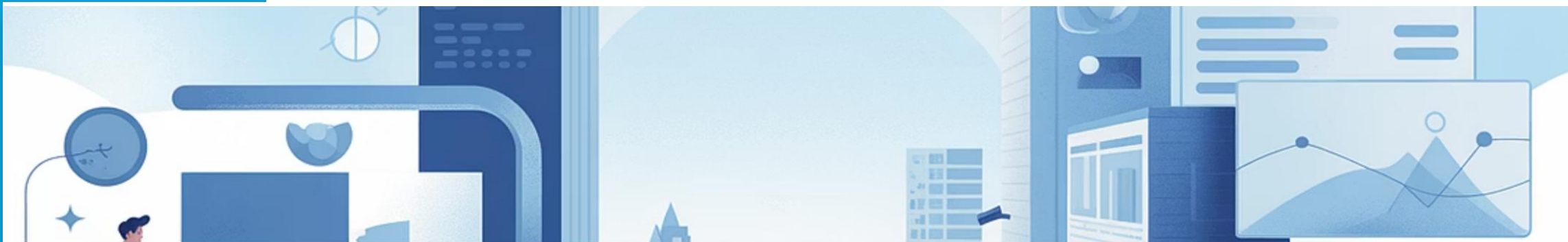
«Внедрение технологий искусственного интеллекта позволит обеспечить **качественно новый уровень обработки и анализа статистической информации**»

“

«Создание интеллектуальных **систем поддержки принятия решений** на основе больших данных — приоритетная задача цифровой трансформации государственной статистики»

“

«Развитие **отечественных решений в области ИИ** обеспечит **технологический суверенитет и безопасность** статистических данных»



Вызовы при реализации задач ИИ в государственной статистике

Технологические санкции

Ограниченный доступ к зарубежным компонентам и программному обеспечению, усложняющие развитие и внедрение ИИ-решений в статистическую систему

Кадровый дефицит

Острый недостаток квалифицированных специалистов и необходимость масштабного повышения квалификации действующих сотрудников и преподавателей в области ИИ












Безопасность данных

Обеспечение конфиденциальности и защиты персональных данных при использовании ИИ, предотвращение несанкционированного доступа и злоупотреблений технологиями

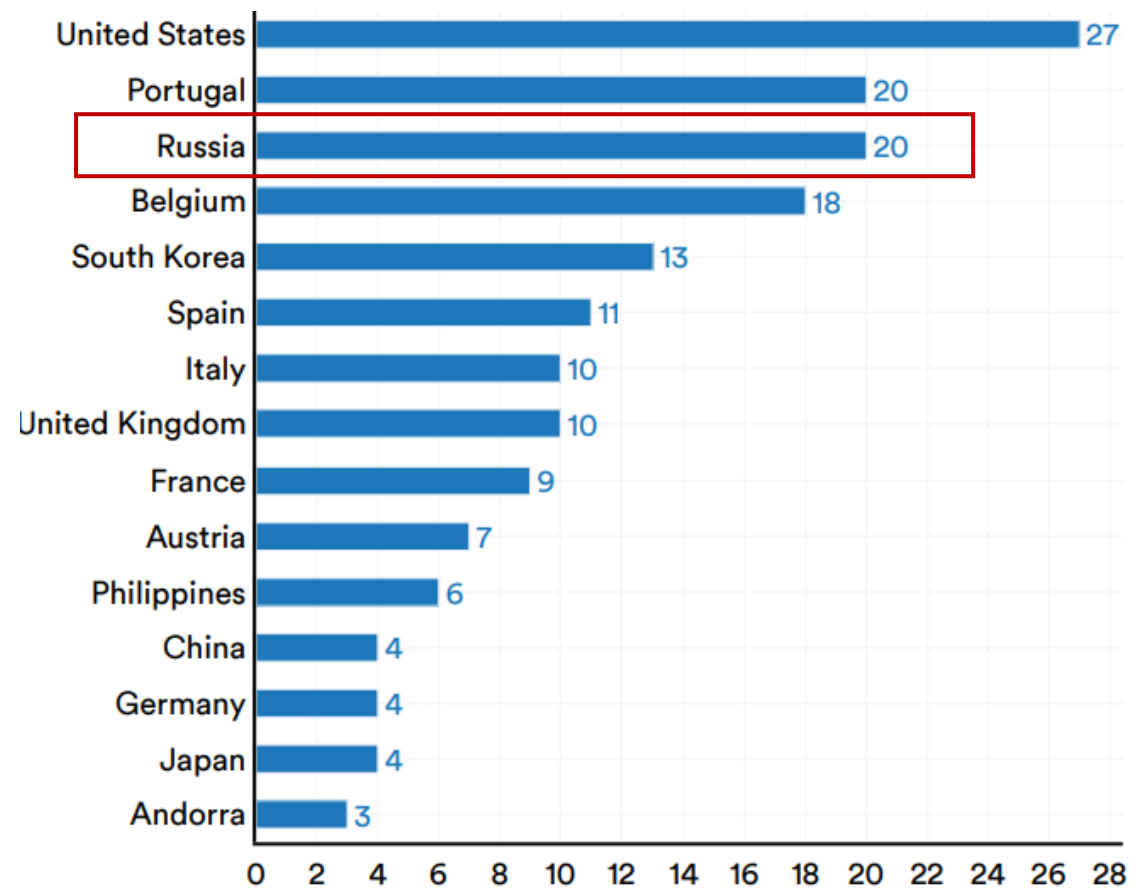
Интеграция процессов

Необходимость бесшовной интеграции ИИ в существующие статистические процессы без потери качества данных и с учётом этических норм применения технологий

The Global AI Index 2024

	Overall	Talent	Infrastructure	Operating Environment	Research	Development	Government Strategy	Commercial	Scale	Intensity
 United States	1	1	1	2	1	1	2	1	1	3
 China	2	9	2	21	2	2	5	2	2	21
 Singapore	3	6	3	48	3	5	10	4	11	1
 United Kingdom	4	4	17	4	4	16	7	5	3	9
 France	5	10	14	19	6	4	9	8	6	10
 South Korea	6	13	6	35	13	3	4	12	7	11
 Germany	7	3	13	8	8	11	8	9	5	15
 Canada	8	8	18	16	9	10	3	6	8	8
 Israel	9	7	26	65	7	6	32	3	14	2
 India	10	2	68	3	14	13	11	13	4	36
 Japan	11	23	5	53	20	14	12	14	9	31
 Switzerland	12	5	11	58	5	19	64	20	29	4
 Portugal	29	29	37	6	32	23	53	30	30	29
 Brazil	30	26	36	28	44	29	27	33	19	44
 Russia	31	58	44	30	37	20	21	40	22	46

Число законопроектов,
связанных с ИИ, принятых
в отдельных
географических регионах,
2016–2024 гг. (суммарно)



https://d-russia.ru/wp-content/uploads/2024/12/hai_ai_index_report_2025.pdf

**2. Основные
понятия:
искусственный
интеллект,
машинное обучение,
статистический ИИ**

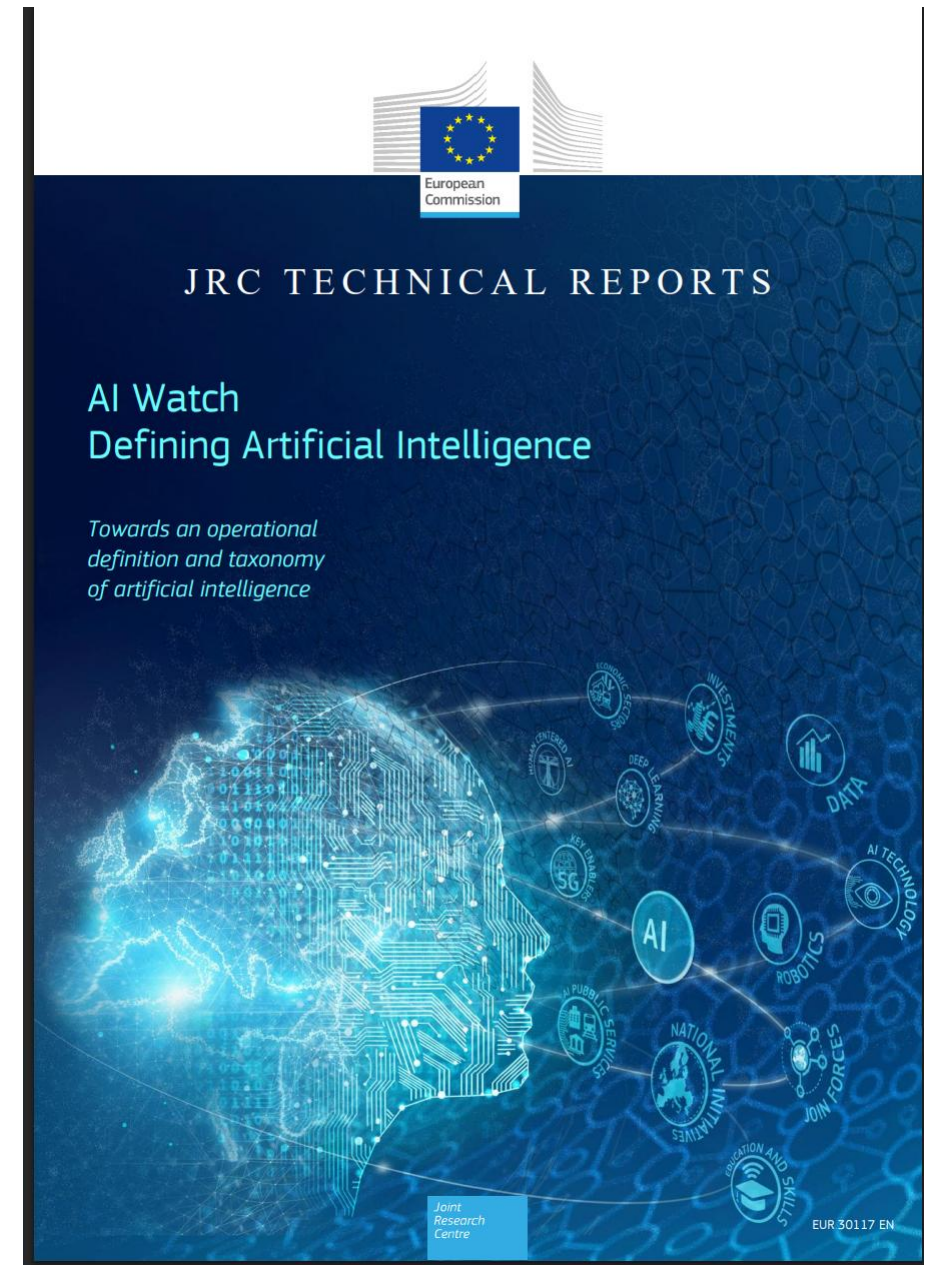


Коллекция определений ИИ

Определения ИИ, представленные:

- Международными организациями
 - Национальными службами
 - Институтами

<https://publications.jrc.ec.europa.eu/repository/handle/JRC118163>



Определения ИИ международными статистическими организациями

eurostat 

Statistics Explained

Искусственный интеллект относится к системам, использующим **такие технологии**, как **интеллектуальный анализ текста**, компьютерное зрение, распознавание речи, генерация естественного языка, **машинное обучение**, глубокое обучение, для **сбора и/или использования данных для прогнозирования, рекомендации или принятия решений с различной степенью автономности** относительно наилучших действий для достижения конкретных целей.

Искусственный интеллект (ИИ) — это обширная область компьютерной науки, ориентированная **на создание систем, способных выполнять задачи, которые обычно требуют человеческого интеллекта**. Это включает в себя широкий спектр возможностей, таких как обучение, рассуждение, решение задач, восприятие, понимание языка и другие. Системы ИИ часто включают в себя **машинное обучение**, что позволяет им учиться и совершенствоваться.- ILOSTAT



International
Labour
Organization



«Искусственный интеллект (ИИ) или **самообучающиеся системы** — это собирательное название для машин, которые **воспроизводят когнитивные способности человека.**»

В более широком технологическом ландшафте **предиктивное обслуживание в когнитивную эпоху может преобразовать глобальные производственные системы**», 2019.



United Nations

Искусственный интеллект (ИИ) включает в себя широкий спектр **технологий**, которые можно определить как «**самообучающиеся, адаптивные системы**». Его можно классифицировать по **технологиям, целям** (например, распознавание лиц или изображений), **функциям** (например, понимание языка и решение задач) или типам агентов (включая роботов и беспилотные автомобили).

ИИ и МО в «Стратегии развития государственной статистики до 2030 года»



- ИИ рассматривается как инновационная технология для расширения перечня анализируемых показателей и оперативной обработки собираемых статистических сведений.
- Интеграция ИИ в государственную статистику направлена на эффективное использование больших данных, повышение качества информации и снижение нагрузки на респондентов за счет автоматизации и интеллектуального анализа.
- Использование ИИ обозначено как ключевой вектор цифровой трансформации — от сбора данных к аналитике, построению индивидуальных наборов показателей, применению административных и иных источников информации для формирования официальной статистики.
- В стратегии акцентируется, что внедрение ИИ и МО способствует развитию новых подходов к анализу и интерпретации статистики. В частности, это новые методы группировки, замены пропущенных данных, поиска выбросов, статистического моделирования распределений, основанные на данных и алгоритмах.

Машинное обучение

МО позволяет выявлять скрытые закономерности, структуру и взаимосвязи в данных (**скрытые паттерны**)

Машинное обучение — это область искусственного интеллекта, целью которой является автоматизация решения сложных профессиональных задач **за счет способности алгоритмов учиться на данных без явного программирования.**

Задачи машинного обучения :

- ☐ классификация — распределение объектов по заранее заданным категориям,
- ☐ регрессия — предсказание числовых значений,
- ☐ кластеризация — выделение скрытых групп в данных,
- ☐ уменьшение размерности данных — упрощение структуры данных для анализа

Методы МО

- Обучение с учителем: обучение **на размеченных** данных.
- Обучение без учителя: анализ и выделение паттернов **без заранее известных меток.**
- Обучение с подкреплением: обучение стратегии через **взаимодействие с окружением и получение обратной связи.**
- Глубокое обучение: **использование нейронных сетей** с большим числом слоев для моделирования сложных зависимостей

Основные алгоритмы МО

Бэггинг (bootstrap aggregating) — это метод, при котором несколько моделей обучаются параллельно на разных случайных подвыборках исходных данных, а их результаты усредняются или голосуются для повышения устойчивости и точности.

Бустинг — это последовательный метод, где каждая следующая модель обучается на ошибках предыдущей, направленный на снижение систематической ошибки (смещения).

Стекинг (stacking) — это метод ансамблирования в машинном обучении, при котором несколько различных моделей ("базовых") обучаются на одной задаче, а их прогнозы передаются на вход другой модели — "мета-модели". Мета-модель учится оптимально комбинировать результаты базовых моделей для улучшения итогового прогноза. В отличие от бэггинга и бустинга, стекинг может объединять алгоритмы разной природы.

Вставка 1. «Без явного программирования» означает, что компьютерная система не получает заранее прописанные пошаговые инструкции для решения задачи; вместо этого она самостоятельно извлекает закономерности из данных и учится выполнять требуемые действия на основе примеров и опыта.

Программист не строит алгоритм вручную, а предоставляет обучающую выборку, по которой модель формирует свои правила и прогнозы. Алгоритмы машинного обучения становятся способны работать с новыми ситуациями, опираясь на обобщённые знания, а не на фиксированный набор инструкций.

Пример самой простой статистической задачи с "явным программированием":

Пусть требуется для набора чисел (например, измерений роста учеников: 158, 160, 162, 159, 161) вычислить среднее арифметическое вручную или с помощью простого кода. Для этого мы строго по шагам:

Суммируем все значения: $158 + 160 + 162 + 159 + 161 = 800$.

Делим сумму на количество наблюдений: $800 / 5 = 160$.

Алгоритм очевиден и всегда одинаков для любых данных. Программист явно указывает вычисляемую формулу и все шаги расчёта, не предоставляя компьютеру искать закономерности или самостоятельно строить модель, — это и есть "явное программирование".

Дерево регрессии — это пример задачи машинного обучения без явного программирования, потому что модель автоматически строится на основе данных, а не по заранее прописанным правилам. Алгоритм анализирует обучающие примеры с числовыми целевыми значениями, рекурсивно разбивая данные на подмножества с помощью простых правил "если..., то...", которые он сам выбирает для минимизации ошибки прогноза. В конечных листах дерева регрессии получается прогноз — усреднённое значение целевой переменной для соответствующего подмножества. Таким образом, человек задаёт только данные и цель, а алгоритм сам создаёт структуру дерева и правила, что и означает отсутствие явного программирования решения задачи.

Пример задачи дерева регрессии и правила "если-то":

Допустим, нужно предсказать цену дома в зависимости от площади и возраста дома. Дерево регрессии автоматически строит последовательность правил, например:

Если площадь дома больше 100 кв.м, то перейти к следующему правилу,

Если возраст дома меньше 10 лет, тогда предсказать цену как 5 млн рублей,

Иначе, если возраст больше 10 лет, предсказать цену как 3 млн рублей,

Если площадь меньше или равна 100 кв.м, предсказать цену как 2 млн рублей.

Каждое такое правило "если-то" создаётся автоматически на основе анализа данных с целью минимизации ошибки прогноза, без необходимости вручную прописывать формулы или уравнения. В итоге, дерево регрессии — это набор таких логических условий, оформленных в виде дерева, где в листьях находятся предсказанные значения.

Машинное обучение (МО)

Машинное обучение — это междисциплинарная область знаний, основанная на

статистических методах, численных алгоритмах и теории оптимизации,

которая занимается разработкой моделей и алгоритмов, способных обучаться на данных для выявления закономерностей и принятия решений без явного программирования.

В основе машинного обучения лежит статистическая теория обучения, направленная на построение итоговых моделей с использованием тренировочных данных и оценку их качества на тестовых данных с помощью статистических критериев.

Машинное обучение сочетает в себе задачи классификации, регрессии и кластеризации, широко применяемые для анализа больших объемов статистических данных и решения практических проблем в различных областях.

Вставка 2. Статистическая теория обучения — это направление, изучающее обобщающие свойства алгоритмов машинного обучения и строящее математические модели, позволяющие оценить, насколько хорошо обученная на данных модель будет работать на новых, ранее не встречавшихся примерах. Теория оперирует такими понятиями, как ошибка обобщения, переобучение и объем обучающей выборки.

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ:

- (а) методы обучения без учителя (*unsupervised learning*)
- (б) методы обучения с учителем (*supervised learning*)
- (в) обучение с частичным привлечением учителя (*semi-supervised learning*)
- (г) обучение с подкреплением (*reinforcement learning*)
- Современные методы ансамблирования и улучшения моделей



Машинное обучение «с учителем»

(контролируемое) - то один из основных методов машинного обучения, при котором модель обучается на размеченных данных. В процессе обучения модель получает входные данные и соответствующие им правильные ответы (метки), что позволяет ей научиться делать предсказания на новых, ранее не виденных данных.

- **Линейная регрессия**
- **Логистическая регрессия**
- **Деревья регрессии и деревья классификации**
- **Метод опорных векторов**
- **Дискриминантный анализ**
- **Нейронные сети**

Машинное обучение "без учителя"

(неконтролируемое). Эти алгоритмы обнаруживают скрытые закономерности или группировки данных без необходимости вмешательства человека.

Semi-supervised learning -Полуконтролируемое обучение

Во время обучения используется меньший набор помеченных данных для классификации или регрессии из большего набора немаркированных данных.

- **Кластерный, бикластерный анализ**
- **Метод главных компонент**

Машинное обучение с подкреплением (Reinforcement machine learning) — это модель машинного обучения, которая похожа на контролируемое обучение, но **алгоритм не обучается с использованием выборочных данных. Эта модель обучается по ходу дела, используя метод проб и ошибок.** Последовательность успешных результатов будет подкреплена для разработки лучшей рекомендации или политики для данной проблемы.

Немашинное обучение в ИИ

(1) Приложение для прогнозирования погоды, которое предсказывает сегодняшнюю погоду на основе вчерашней погоды. Оно построено на основе одной жестко запрограммированной системы правил и демонстрирует возможности, подобные «естественному интеллекту».

(2) Методы поиска по дереву или графу

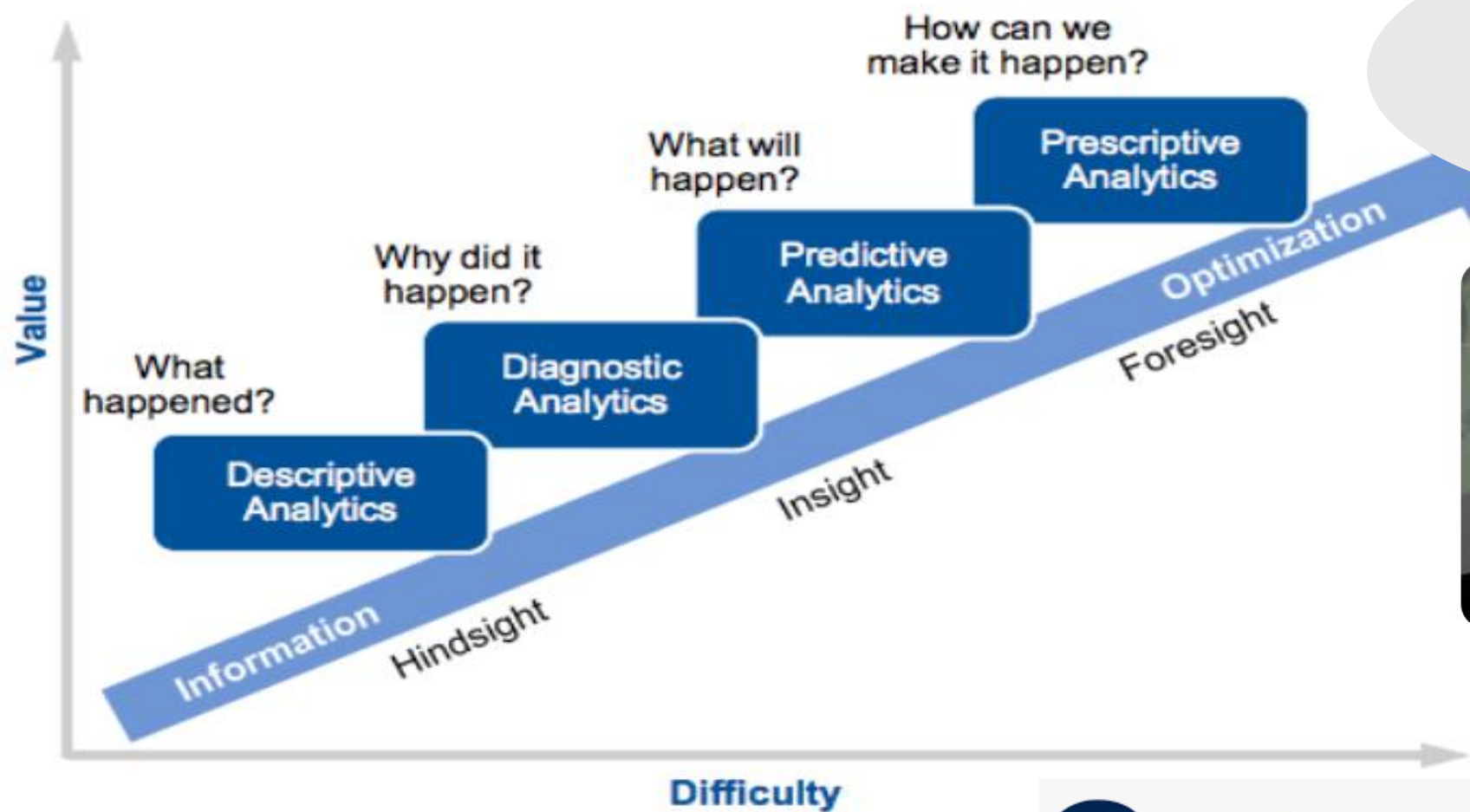
ПЯТЬ КАТЕГОРИЙ АНАЛИТИКИ ДАННЫХ:

- (1) описательная аналитика
- (2) диагностическая аналитика
- (3) предиктивная аналитика
- (4) предписывающая аналитика
- (5) когнитивная аналитика

В аналитической практике эти категории сосуществуют и дополняют друг друга



Модель восходящей аналитики Г. Гартнера



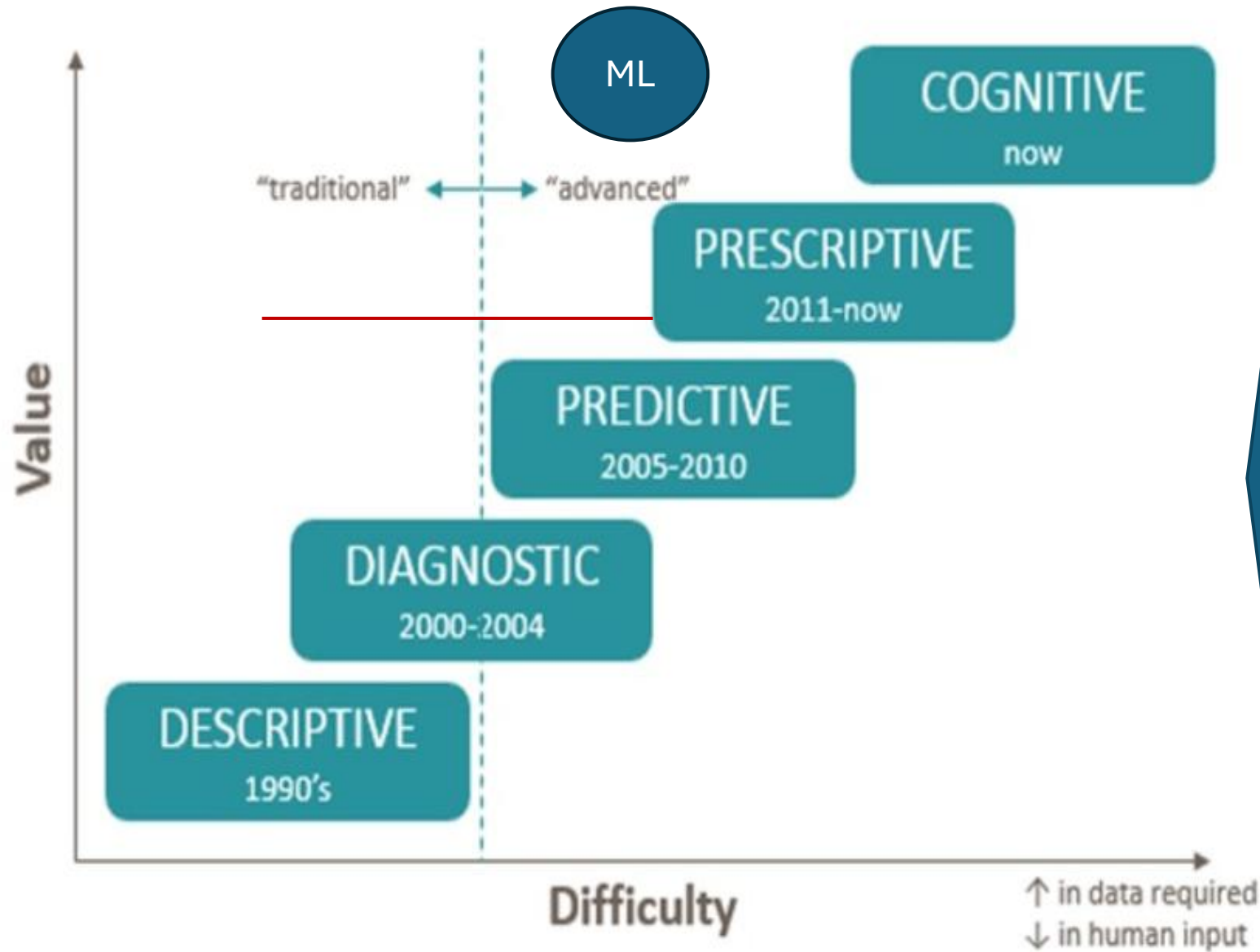
...для оценки зрелости
данных и повышения
ценности аналитических
результатов



Gideon I. Gartner (March 13, 1935 – December 12, 2020)

Source: Gartner (March 2012)

Gartner®



- *Descriptive analytics* - “What happened?”
- *Diagnostic analytics* - “Why did it happen?”
- *Predictive analytics* - “What will happen? When and Why?”
- *Prescriptive analytics* – “How can we make it happen? Informed decision making”
- *Cognitive analytics* – “Automation of decision making”



Описательная аналитика

Основная задача описательной аналитики как стадии — предоставить систематизированную и наглядную картину того, что произошло, путем обработки, агрегирования и интерпретации «исторических» статистических данных.

Она отвечает на вопрос "Что случилось?" и выявляет ключевые закономерности и тенденции в прошлом, тем самым создавая базу для принятия обоснованных решений и последующего более глубокого анализа данных.

Диагностическая аналитика

Основная задача диагностической аналитики как стадии анализа данных — **выявить причины и факторы** произошедших событий или изменений в показателях, то есть ответить на вопрос «Почему это произошло?». Диагностическая аналитика используется **для поиска взаимосвязей, аномалий и факторов, влияющих на результаты, с помощью многофакторного анализа, проверки статистических гипотез и детализации данных**. Это позволяет глубже понять природу возникших проблем и найти пути для их решения.

Предиктивная (прогнозная) аналитика

Основная задача предиктивной (прогнозной) аналитики как этапа — **построение моделей**, позволяющих на основе исторических и текущих данных **прогнозировать вероятные сценарии развития событий или будущие значения ключевых показателей**. Эта стадия отвечает на вопрос «**Что произойдет?**» и используется для предсказания трендов, спроса, поведения клиентов, рисков или других изменений, важных для принятия обоснованных решений. В результате предиктивная аналитика обеспечивает возможность компании готовиться к будущим ситуациям и минимизировать потенциальные риски.

Прескриптивная (предписывающая) аналитика

Задача прескриптивной аналитики заключается в том, чтобы на основании анализа данных и прогнозных моделей вырабатывать конкретные рекомендации и оптимальные решения для достижения бизнес-целей в текущей ситуации.

Прескриптивная аналитика отвечает на вопрос "Что делать?" — то есть не только описывает или прогнозирует ситуацию, но и формирует сценарии действий, учитывающих последствия и ограничения доступных вариантов.

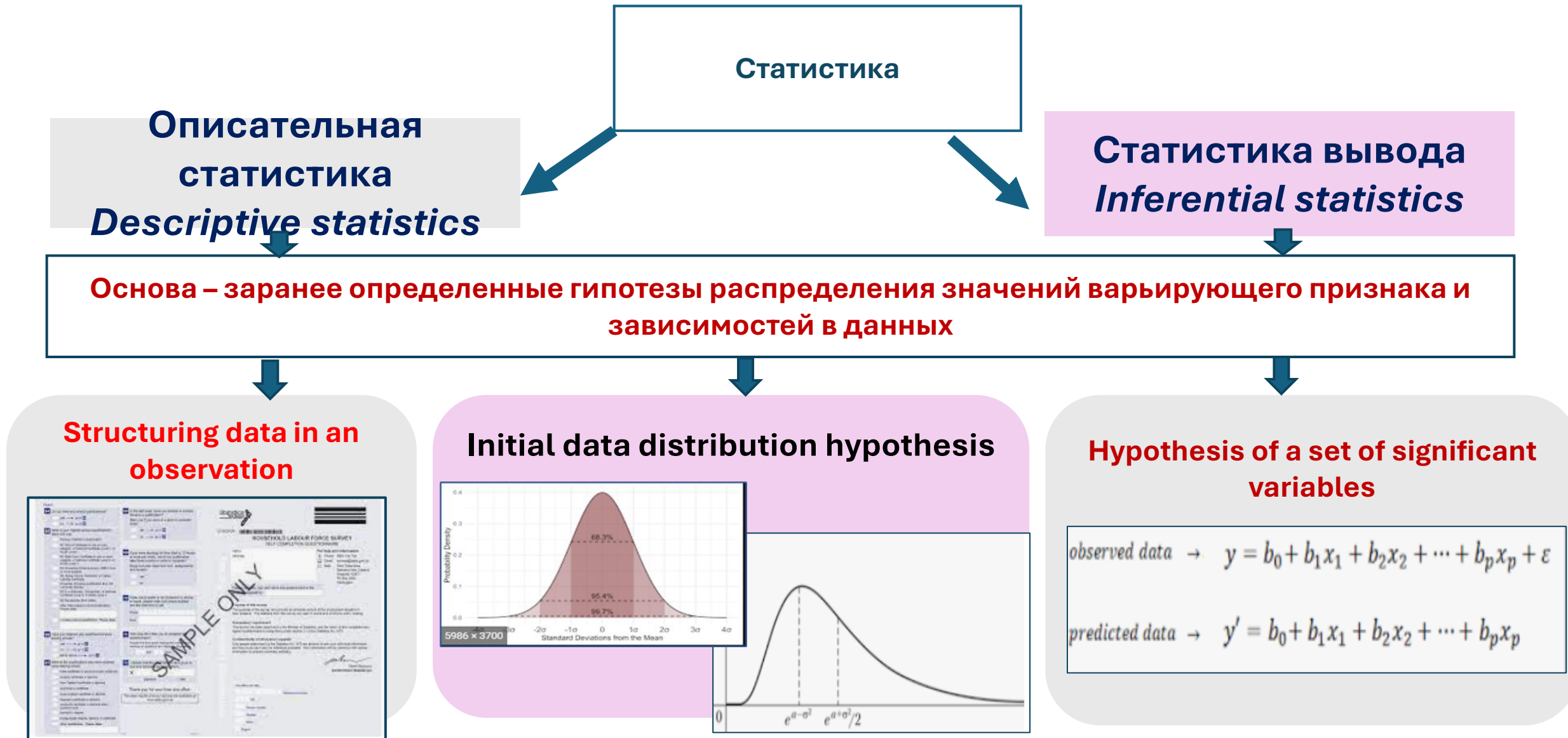
Она применяется для выбора наилучших стратегий и оперативных решений на основе анализа большого числа факторов и моделирования возможных сценариев развития событий.

Когнитивная аналитика

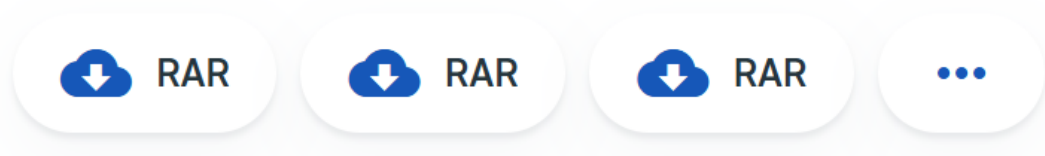
Основная задача когнитивной аналитики как стадии анализа данных — это извлечение, интерпретация и осмысление знаний из неструктурированной или сложной информации путём имитации когнитивных (мысленных) процессов человека.

Когнитивная аналитика стремится устанавливать смысловые связи, выявлять закономерности, причинно-следственные отношения и использовать эти знания для поддержки принятия решений в условиях высокой неопределённости, сложности и большого объёма данных. Она помогает преобразовать разрозненную информацию в осмысленные модели, интегрируя машинное обучение, семантический анализ и экспертные подходы.

Традиционные статистические методы основаны на гипотезах, выдвигаемых заранее исследователем (предварительно определённые гипотезы)



Итоги выборочного обследования рабочей силы



Таб.2.2. Структура занятого населения в возрасте 15 лет и старше по возрастным группам

Таб.2.10. Занятое население в возрасте 15 лет и старше по уровню образования

Таб.2.11. Структура занятого населения в возрасте 15 лет и старше по уровню образования

Таб.2.20. Занятое население в возрасте 15 лет и старше по месту основной работы

Данные ОРС

МО:

**Скрытый паттерн : возраст + тип
занятости + миграционный статус =**

**мигранты молодого и среднего возраста более
гибки в выборе типа занятости — они
быстрее адаптируются к гибридным и неполным
формам занятости, включая дистанционную
работу и самозанятость**

Скрытые паттерны ОРС – на основе МО по рекомендации МОТ

- ❑ **Скрытый сегмент временной и частичной занятости**

Машинное обучение может выделить группы работающих, сочетающих несколько видов временной или неполной занятости.

- ❑ **Взаимосвязь уровня образования, профессиональной мобильности и миграционного статуса**

- ❑ **Паттерны ухода из рабочей силы и возвращения к ней**

Скрывается динамика, связанная с временным выходом из рабочего процесса (например, декрет, уход за родственниками) и повторной интеграцией, что позволяет понять циклы участия в труде и риски долгосрочной безработицы.

- ❑ **Влияние автоматизации на смену профиля занятости разных возрастных и миграционных групп**

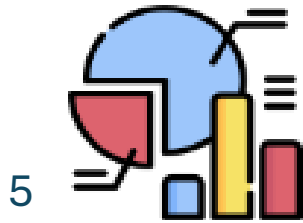
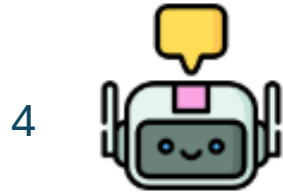
- ❑ **Психосоциальные факторы и профессиональное выгорание: возраст, стаж работы, частота смены занятости**

Алгоритмы выделяют паттерны, связанные с эмоциональным выгоранием, стрессом и влиянием условий труда на занятость в разных социально-демографических группах, что важно для разработки мер поддержки сотрудников.

Трансформация гипотетической основы статистических методов на этапах восходящей аналитики



Match the icons



1. Descriptive analytics

2. Diagnostic analytics

3. Predictive analytics

4. Prescriptive analytics

5. Cognitive analytics

Стадии ИИ

Указ Президента РФ от 10.10.2019 N 490
"О развитии искусственного интеллекта в
Российской Федерации"
(вместе с "Национальной стратегией развития
искусственного интеллекта на период до 2030
года")

(в) **перспективные** методы
искусственного интеллекта в целях
разработки универсального
(**сильного**) искусственного
интеллекта (.....**автоматическое**
машинное обучение....)

МО-класс
методов ИИ

Искусственный интеллект - комплекс технологических решений, позволяющий **имитировать когнитивные функции человека** (включая **самообучение** и поиск решений **без заранее заданного алгоритма**) и получать при выполнении конкретных задач результаты, сопоставимые, **как минимум**, с результатами интеллектуальной деятельности человека



Narrow
AI

NAI-Узкий ИИ - выполняет одну или **узкую задачу**, часто гораздо быстрее и лучше, чем человеческий разум

General
AI

AGI -Общий (сильный) ИИ- может использовать предыдущие знания и навыки для выполнения **новых задач** в другом контексте без необходимости обучения базовых моделей людьми

Super AI

SAI- Супер ИИ -будет обладать **когнитивными способностями, превосходящими человеческие**. Генерирует идеи, концепции и решения, выходящие **за пределы понимания человека** (но во благо человека)! **Самосовершенствуется!!!**

Стадия ИИ в Российской Федерации - оценка

- Россия на сегодняшний день находится на стадии развития **Narrow AI** (узкий ИИ), где акцент делается на практическом внедрении и масштабировании конкретных приложений искусственного интеллекта для решения прикладных задач в финансах, промышленности, госсекторе, медицине и других сферах. В стране активно создается технологическая инфраструктура — национальные центры развития ИИ, дата-центры, собственные экосистемы и модели (например, YandexGPT и GigaChat).

3. Краткий обзор мировых и российских кейсов внедрения

Количественный анализ тематической структуры (**940 докладов**)

Распределение по основным тематическим блокам

| Тематический блок | Количество докладов | Доля (%) |

|-----|-----|-----|

| Официальная статистика и национальные системы | 284 | 30.2% |

| **Машинное обучение, ИИ, большие данные** | **212** | **22.6%** |

| Теоретическая статистика и методология | 183 | 19.5% |

| Здоровоохранение и биостатистика | 112 | 11.9% |

| Социальные науки, рынок труда, неравенство | 87 | 9.3% |

| Экология, климат, устойчивое развитие | 62 | 6.6% |



Географическое распределение (по аффилиации первых авторов)



Европа: 42%

Азия и Тихоокеания: 24%

Северная и Южная Америка: 21%

Африка: 13%



Более 210 докладов затронули темы **машинного обучения, глубокого обучения, больших языковых моделей (LLM)** и их интеграции в статистическую практику.

Однако конгресс продемонстрировал зрелый подход:

ИИ рассматривается не как замена статистике, а как инструмент, требующий **статистической валидации, интерпретируемости и контроля смещений.**

Тесты с ответами на понимание ИИ и машинного обучения

Проверьте свои знания в области искусственного интеллекта и машинного обучения с помощью тщательно подобранных вопросов. Каждый тест содержит четыре варианта ответа, один из которых правильный.



Вопрос 1: Что такое машинное обучение?

1

Вариант А

Процесс ручного программирования всех действий компьютера

2

Вариант Б

Раздел информатики, изучающий автоматическое обучение и совершенствование систем на основе данных

3

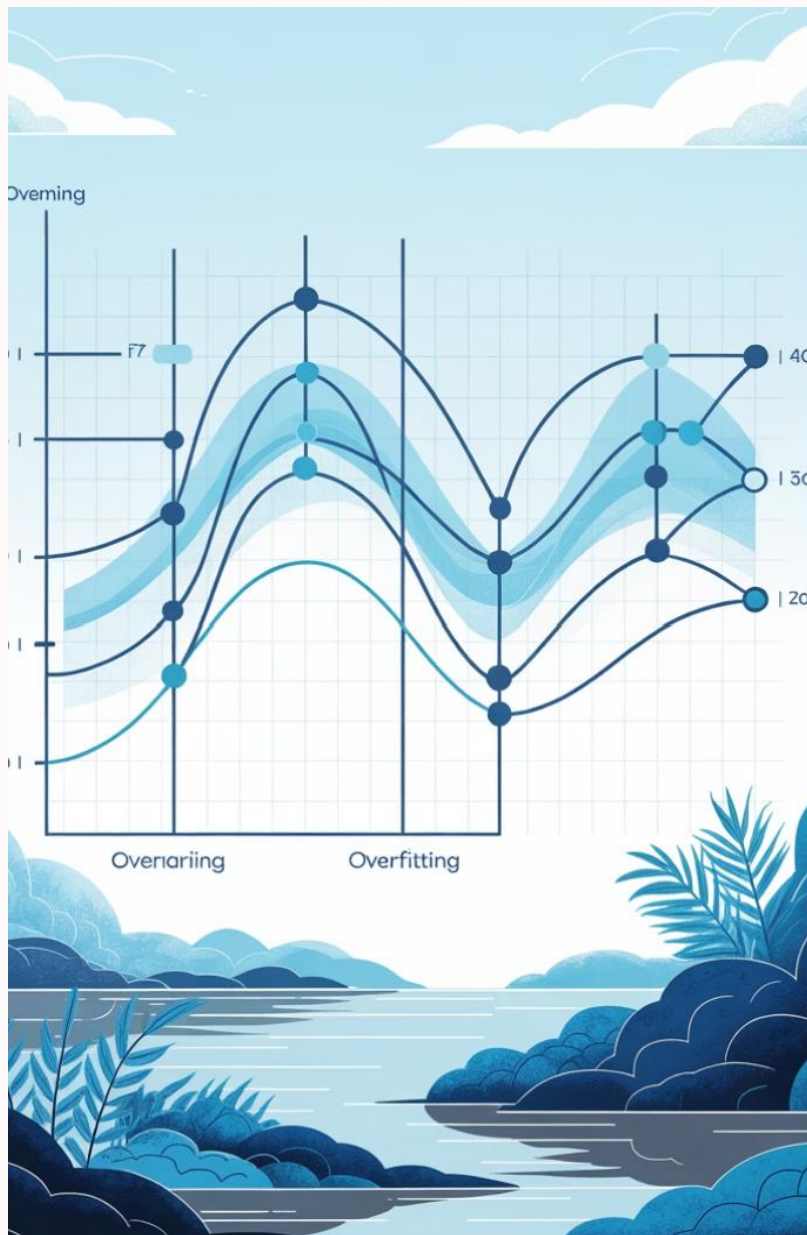
Вариант В

Искусство создания роботов для выполнения физических задач

4

Вариант Г

Метод хранения больших объёмов данных в облаке



Вопрос 2: Что такое переобучение в машинном обучении?

А

Когда модель слишком просто описывает данные и плохо учится

Б

Когда модель запоминает шум и случайные ошибки вместо закономерностей

В

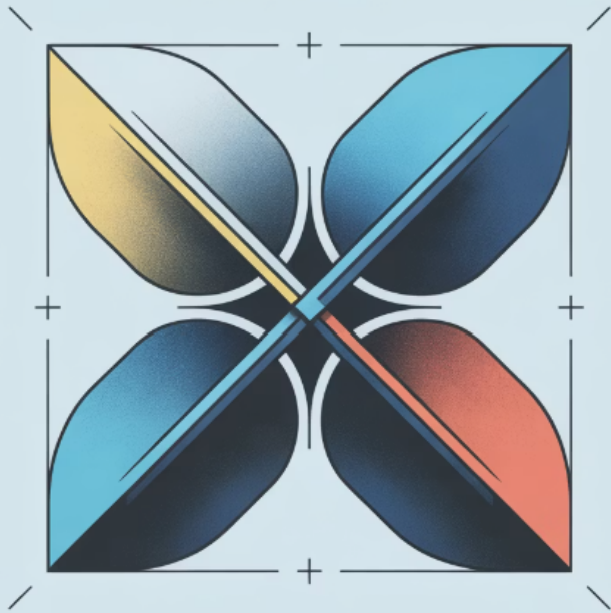
Процесс увеличения объёма обучающих данных

Г

Использование слишком маленькой модели для больших данных

Переобучение — это одна из главных проблем в машинном обучении, когда модель слишком хорошо подстраивается под обучающие данные, теряя способность к обобщению.

Вопрос 3: Как можно избежать переобучения?



Вариант А

Использовать меньше данных для обучения

Вариант Б

Применять перекрёстную проверку (cross-validation)

Вариант В

Увеличивать количество параметров модели

Вариант Г

Игнорировать тестовые данные

Перекрёстная проверка помогает оценить способность модели работать на новых данных, предотвращая переобучение.

Вопрос 4: Какие из перечисленных алгоритмов относятся к машинному обучению?



Вариант А

Деревья решений, нейронные сети, поддерживающие векторные машины



Вариант В

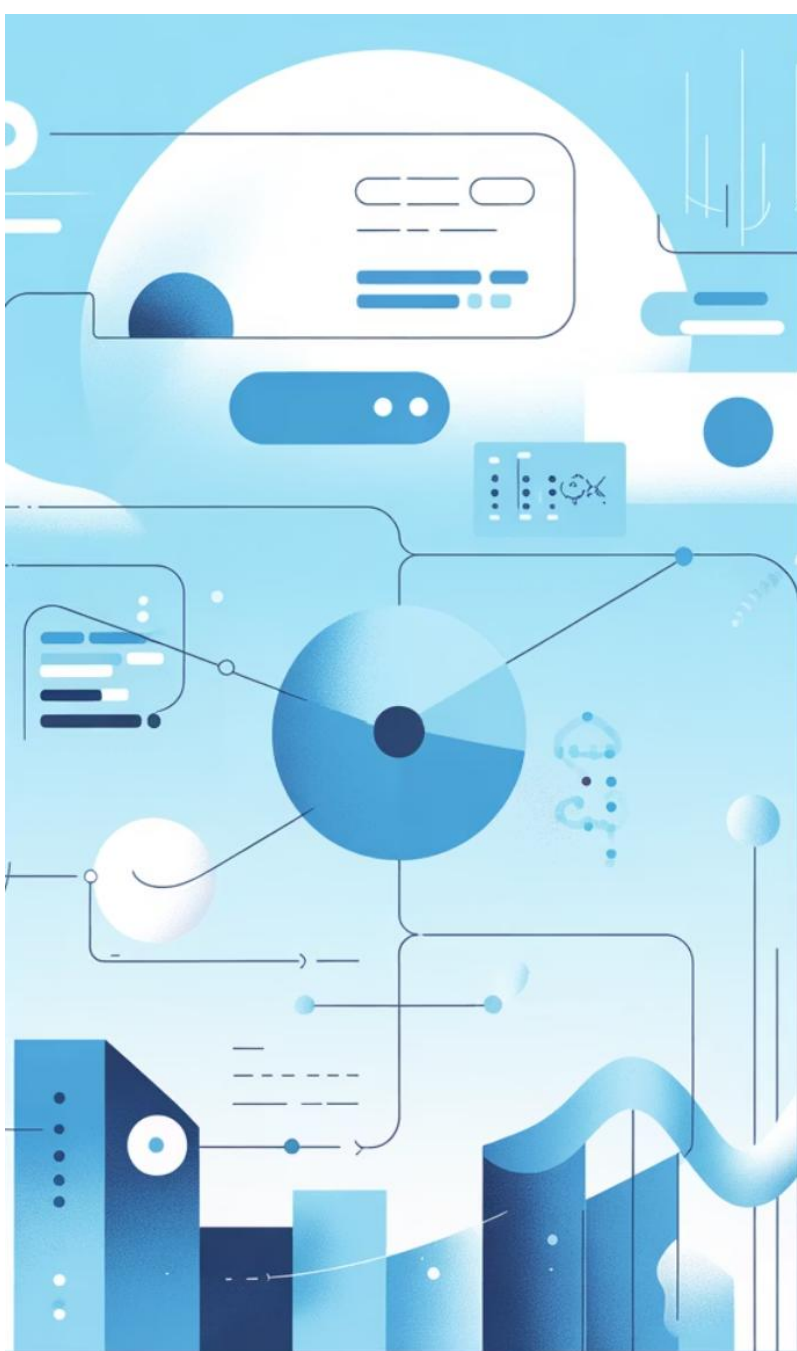
Фотошоп, Illustrator, Premiere Pro

Деревья решений, нейронные сети и SVM — это классические алгоритмы машинного обучения, широко используемые в различных задачах классификации и регрессии.



Вариант Г

Электрические цепи и схемы



Вопрос 5: Что такое обучение с учителем?

1

Обучение модели на размеченных данных с известными ответами

При обучении с учителем модель получает примеры с правильными ответами и учится предсказывать результаты для новых данных.

2

Обучение без использования данных

Это невозможно — любому алгоритму машинного обучения требуются данные для обучения.

3

Обучение модели на неразмеченных данных

Это описание обучения без учителя (unsupervised learning), а не с учителем.

4

Обучение модели с помощью случайных чисел

Случайные числа используются для инициализации, но не являются основой обучения.

Вопрос 6: Искусственная нейронная сеть была вдохновлена:



Вариант А

Работой человеческого мозга



Вариант Б

Законами физики



Вариант В

Принципами бухгалтерского учёта



Вариант Г

Механикой автомобилей



Поздравляем!



Тест завершён

Теперь вы лучше понимаете ключевые понятия искусственного интеллекта и машинного обучения



Применяйте знания

Используйте полученные знания для дальнейшего изучения и практического применения в реальных проектах



Новые вызовы

Готовы двигаться дальше и покорять новые высоты в захватывающем мире искусственного интеллекта?

Продолжайте учиться, экспериментировать и развивать свои навыки в области ИИ и машинного обучения!

* Кросс-валидация (cross-validation) — это метод оценки качества модели машинного обучения, заключающийся в многократном разбиении набора данных на обучающие и тестовые подмножества. Модель обучается на одной части данных, а затем проверяется на другой, что позволяет получить более надёжную и стабильную оценку её производительности по сравнению с одним простым разбиением на обучение и тест.

** SQL-запросы, HTML, CSS — это три разные технологии, используемые для работы с базами данных и веб-разработкой.

➤ SQL-запросы

SQL (Structured Query Language) — язык структурированных запросов, предназначенный для взаимодействия с реляционными базами данных. С помощью SQL-запросов можно получать, добавлять, изменять и удалять данные, а также управлять структурой базы данных (таблицами, индексами и пр.).

➤ HTML

HTML (HyperText Markup Language) — стандартный язык разметки для создания веб-страниц и их структуры. С помощью HTML задается содержание, структура и элементы страницы, такие как заголовки, абзацы, списки, изображения и ссылки.

➤ CSS

CSS (Cascading Style Sheets) — каскадные таблицы стилей, технология, используемая для оформления и визуального представления веб-страниц. CSS отвечает за цвета, шрифты, отступы, расположение элементов на странице и адаптивность дизайна.

Эти технологии часто используются совместно для создания современных веб-приложений, где:

SQL отвечает за хранение и обработку данных,

HTML — за структуру пользовательского интерфейса,

CSS — за его оформление и внешний вид

Задания



Игра «да и нет не говорите».

Ответить одним развернутым предложением

Вопрос: Можно ли считать машинное обучение видом искусственного интеллекта?

Вопрос: Обязательно ли для всех моделей машинного обучения нужны размеченные данные?

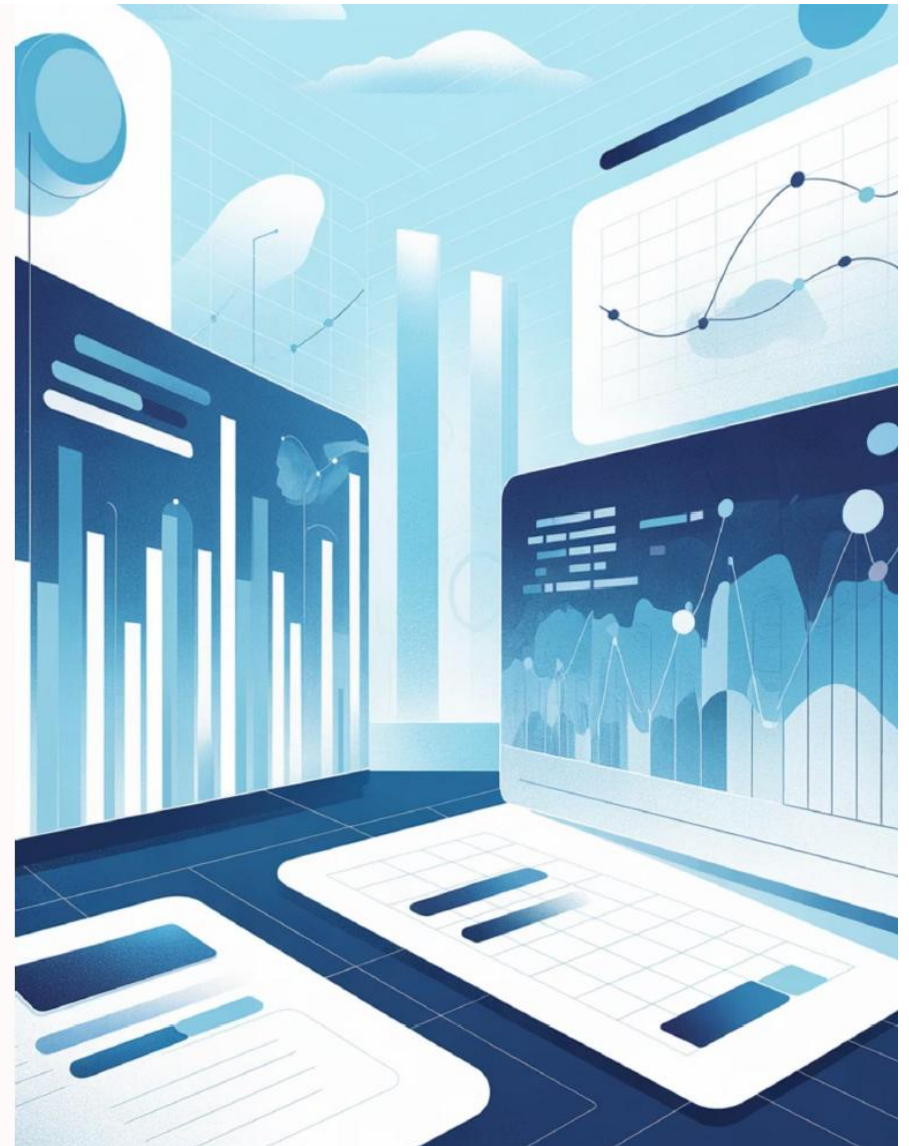
Вопрос: Может ли модель с переобучением давать хорошие результаты на новых данных?

Вопрос: Все ли задачи с данными лучше решать с помощью глубокого обучения?

Вопрос: Признак «высокая корреляция с целевой переменной» всегда гарантирует его полезность в модели?

Методы описательной статистики и статистики вывода в машинном обучении: ответить на вопросы

Глубокое погружение в статистические основы машинного обучения через
систематический анализ ключевых методов и их практического применения





1. Методы описательной статистики в ML

Описательная статистика формирует фундамент для понимания данных, предоставляя инструменты для систематизации, визуализации и количественного описания наборов данных перед применением сложных алгоритмов машинного обучения.



Табличное представление

Сводные таблицы и частотные распределения для структурирования больших объёмов данных



Графическая визуализация

Гистограммы, диаграммы рассеяния, χ^2 -тесты для наглядного анализа



Статистические показатели

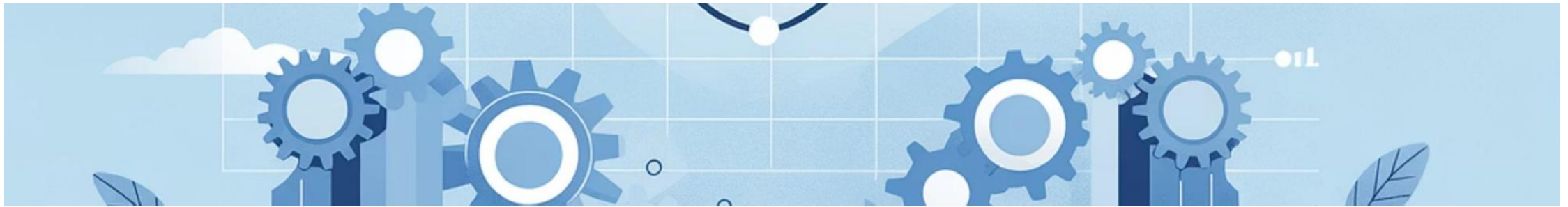
Меры центральной тенденции, разброса, формы распределения данных

Меры центральной тенденции

- Среднее арифметическое
- Медиана, \bar{x} и процентиль
- Мода (наиболее частое значение)

Меры разброса

- Дисперсия, σ^2 и стандартное отклонение
- Интерквартильный размах (IQR)
- Коэффициент вариации, симметрии, эксцесса



2: Методы статистики вывода в ML

Статистика вывода позволяет делать обоснованные заключения о генеральной совокупности на основе выборочных данных, что критически важно для построения надёжных и интерпретируемых моделей машинного обучения.

01

Проверка гипотез

T-тесты, F-тесты и χ^2 -тесты для статистической значимости различий и зависимостей

02

Оценка параметров

Построение доверительных интервалов для количественной оценки неопределённости

03

Регрессионный анализ

Линейная и логистическая регрессия как основа для предсказательного моделирования

04

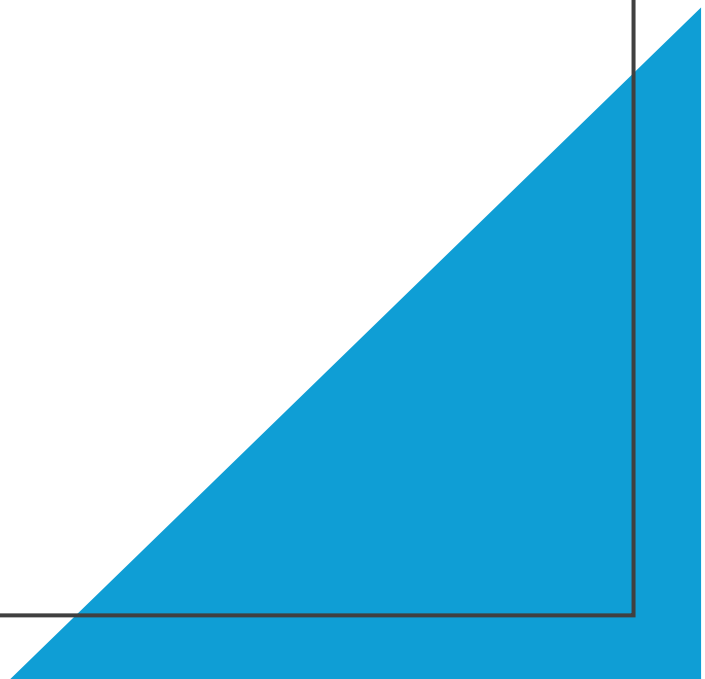
Валидация моделей

Кросс-валидация и бутстрэп для надёжной оценки обобщающей способности

Ключевая роль: Обоснование выводов о генеральной совокупности на основе выборки и проверка статистической значимости параметров в моделях

Спасибо за ВНИМАНИЕ!!!

Zarova.ru@gmail.com



ПРИЛОЖЕНИЯ

Приложение 1



An
Phríomh-Oifig
Staidrimh

Central
Statistics
Office

eurostat 

Artificial Intelligence & Machine Learning for Official Statistics

AIML4OS ESSnet Project

Data Science Leaders Network – January 2025

<https://unstats.un.org/bigdata/events/2025/ai-data-science/symposium/presentations/Day%203%20-%20Peer-learning%20talk%20-%20Brendan%20CSO%20Ireland.pdf>



Project Overview

- 14 participating countries
- €4 million in total funding
- 4 years (Apr 2024 – Mar 2028)
- 13 workpackages – CSO in charge of WP1, Project Management and Coordination
- 55 separate tasks, 39 deliverables



Project Overview

Provide a comprehensive suite of resources, guidelines, methodologies and use cases for the implementation of AI/ML in Official Statistics.

Build communities around AI/ML and innovation

Support transition from prototype to production

Provide working use cases

Provide sandbox for development









Promote shared standards and guidelines

Deliver economies of scale



Project Structure

#	Supporting Workpackages	Lead
1	Coordination	
2	Communication	
3	Lab	
4	State-of-play	
5	Standards, methodology, implementation	 
6	Knowledge repository and training	

#	Use Case Workpackages	Lead
7	Earth Observation Data	
8	Editing	
9	Imputation	
10	Text to code	
11	Supply Chain Networks	
12	Large Language Models	
13	Synthetic Data	 

As many as 12 countries on each workpackage

Приложение 2 – шкала риска ИИ

Организация AI Watch, созданная при Европейской комиссии, провела исследование по соответствию стандартов различных стандартизирующих организаций положениям [законопроекта AIA](#) от 21.04.2021, согласно которым ужесточаются требования к стандартизации систем искусственного интеллекта¹.

Новый подход ЕС к определению ИИ заключается в оценке ИИ с точки зрения рисков (см. Рис.)

Categories of AI systems



The European Commission's risk-based approach is structured around four categories of AI systems. Three of the four are regulated under the AI Act.

<https://rdc.grfc.ru/2021/10/ai-standards/>

Категории риска ИИ – новый подход ЕС

- К системам с *минимальным риском* новые правила не будут применяться, так как представляют минимальный или нулевой риск для прав или безопасности граждан. Компании и пользователи смогут использовать их бесплатно. Примеры включают спам-фильтры и видеоигры. По мнению членов Европейской комиссии большинство приложений искусственного интеллекта попадут в эту категорию.
- Системы ИИ с *ограниченным риском* будут подчиняться конкретным обязательствам по обеспечению прозрачности, чтобы пользователи могли принимать информированные решения, знать, что они взаимодействуют с машиной, и отключаться по желанию. Это относится, например, к чат-ботам.
- К *высокорисковым системам* ИИ относятся: биометрическая идентификация и категоризация физических лиц, образование и профессиональная подготовка, трудоустройство, управление персоналом и доступ к samozанятости, доступ к основным частным и государственным услугам и льготам, данные правоохранительных органов, данные миграционных и пограничных служб, данные институтов отправления правосудия и демократических процессов.
- Системы с *неприемлемо высоким риском* представляют «явную угрозу безопасности, средствам к существованию и правам людей». Примеры включают социальную оценку со стороны правительства (например, кредитную систему Китая), эксплуатацию уязвимых мест детей и использование подсознательных методов. Биометрия – для особых случаев (поиск ребенка)

Приложение 3

Самоадаптивные обучающиеся системы в контексте искусственного интеллекта (ИИ) и машинного обучения (МО) — это интеллектуальные системы, которые используют методы ИИ и МО для автоматической настройки своих моделей и алгоритмов в процессе работы. Эти системы способны самостоятельно анализировать данные, учиться на примерах и изменять свои параметры, чтобы улучшать точность, адаптироваться к изменяющимся условиям, а также предлагать персонализированные решения или образовательные траектории.

- В контексте ИИ и МО самоадаптивные обучающиеся системы реализуют способность к автономному обучению и самокоррекции без необходимости постоянного вмешательства человека. Они используют машинное обучение для выявления закономерностей и оптимизации своих функций в реальном времени, что значительно повышает эффективность и гибкость их применения в различных областях, в том числе в адаптивном обучении, автоматизации, робототехнике и принятии решений.