

Тема лекции 1:
**Введение в искусственный
интеллект**

Лектор: Найханова Лариса Владимировна



Определения терминов

Интеллект

Определение:

от лат. intellectus – ощущение, восприятие, разумение, понимание, понятие, рассудок, или ум – качество психики, состоящее из способности приспосабливаться к новым ситуациям, способности к обучению и запоминанию на основе опыта, пониманию и применению абстрактных концепций и использованию своих знаний для управления окружающей средой

Искусственный интеллект

Определение 1980 г:

это область информатики, которая занимается разработкой интеллектуальных компьютерных систем, то есть систем, обладающих возможностями, которые мы традиционно связываем с человеческим разумом, – понимание языка, обучение, способность рассуждать, решать проблемы и т.д.

Искусственный интеллект

Определение по ГОСТ Р 59277-2020:

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

Система искусственного интеллекта

Определение:

Техническая система, в которой используются технологии искусственного интеллекта

1 BA – BASIC ALGORITHMS

Это **простые интеллектуальные системы с обратной связью**

Это системы, которые принимают некоторые сигналы от внешнего мира и подстраивают свою работу под изменяющиеся

Пример – светофор

2 NAI – NARROW AI – узкий ИИ

Все **существующие** системы ИИ относятся к группе Narrow AI (**слабый ИИ**)

Такие системы строятся на одном из методов ИИ: чаще всего это нейронные сети глубокого обучения, но могут быть и экспертные системы, генетические алгоритмы, нечеткие системы и др.

Они решают какую-то одну определенную задачу в строго заданных граничных условиях

5 SAI – Strong AI - сверхразум

Это интеллект многократно превосходящий по уровню развития человеческий. Исследователи и ученые в области ИИ еще не создали сильного ИИ

Чтобы добиться успеха, им нужно найти способ сделать машины сознательными, запрограммировав полный набор когнитивных способностей

Классификация ИИ



3 WAI – Wide AI

Главная задача таких систем — преодолеть узость применимости

Решаемые проблемы: объяснимость, передача знаний, быстрое обучение, решение проблемы катастрофического забывания, инкрементальное обучение

4 AGI — общий искусственный интеллект

Сильный ИИ нацелен на создание интеллектуальных машин, неотличимых от человеческого разума. Но, как и ребенок, ИИ-машине придется учиться через вводимые данные и опыт, постоянно развивая свои способности с течением времени

Выводы по классификации

- Данная классификация позволяет точнее определить значения термина ИИ в каждом контексте его употребления, и, соответственно, ограничить круг задач и требований, предъявляемых к нему на каждом этапе, что, в свою очередь, улучшит коммуникацию между специалистами, использующими это понятие
- Мы находимся **в начале пути к сильному искусственному интеллекту**, и еще предстоит решить множество теоретических и практических проблем его понимания и создания

Почему мы можем говорить об ИИ как о глобальном тренде?

Искусственный интеллект как глобальный тренд

Наличие четырех факторов

Данных стало
гораздо
больше

Рост
вычислительн
ых
мощностей
компьютеров

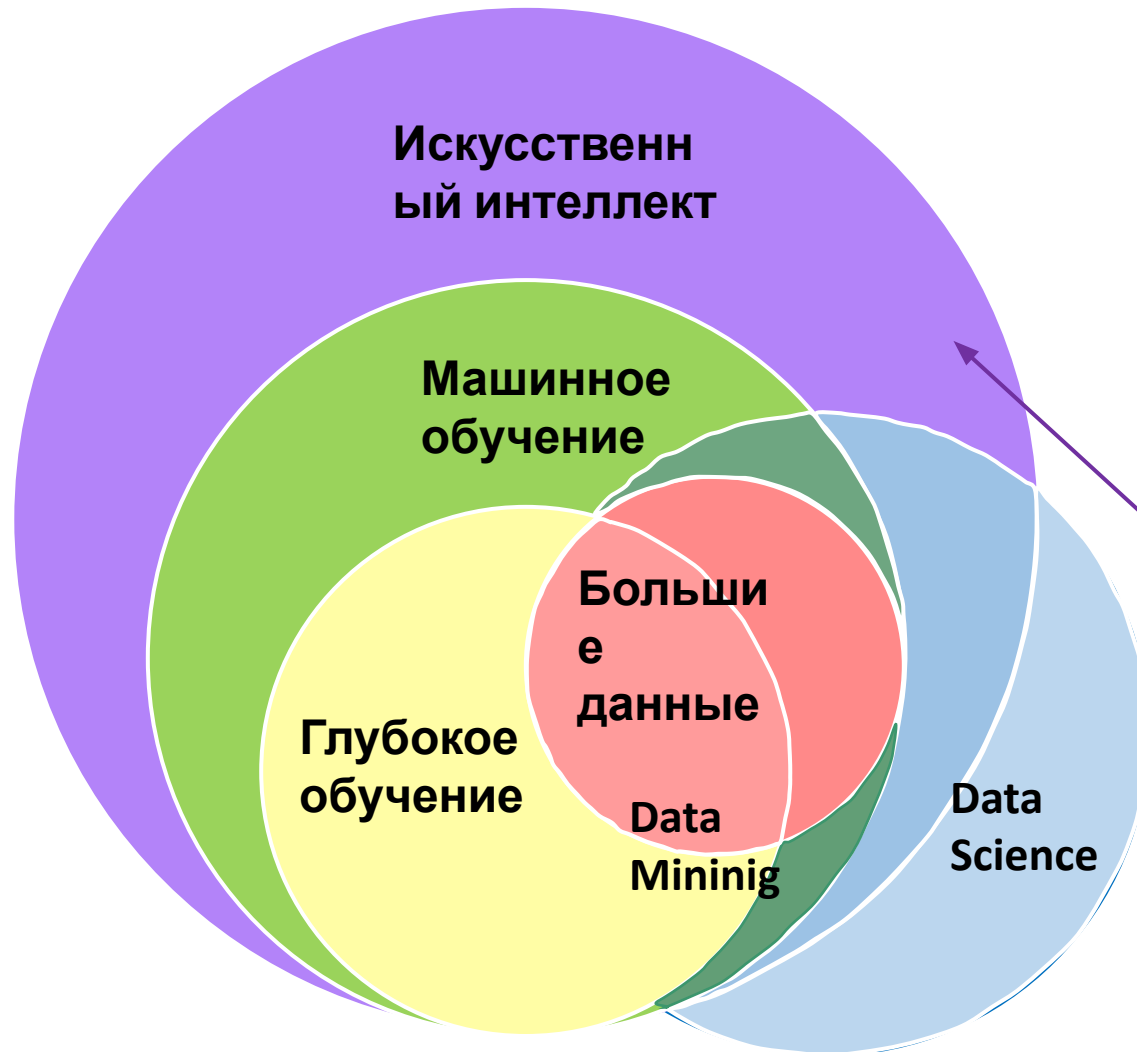
Развитие
соответствующи
х
областей
математики

Снижение
себестоимости
решений,
использующих
технологии
искусственного
интеллекта

Технологические направления ИИ. Данные Deloitte



Смежные области исследований



Технологии, приведенных на предыдущем слайде, разрабатывается в нескольких смежных областях

Сюда относятся и инженерия знаний, и мягкие вычисления и другие исследования, проводимые в области ИИ

Машинное обучение

Машинное обучение (machine learning, ML) - это одно из направлений разработки ИИ, основанное на выполнении компьютером множества сходных задач без использования прямых инструкций

Машинное обучение базируется на трех основных понятиях:

Признаки (свойства, метрики, фичи, features) - индивидуальные измеримые параметры наблюдаемых явлений, от правильности подбора которых зависит успешность и скорость machine learning

Алгоритмы - специальные программы, "подсказывающие" компьютеру, каким источником данных необходимо воспользоваться
Для каждой задачи подбираются отдельные алгоритмы, составленные с расчетом на ускорение обработки данных и получение точного результата

Наборы данных - информация (выборки данных) в виде текстовых, графических, видеофайлов, которую машина использует для накопления опыта при обучении
При этом для решения каждого конкретного типа задач в систему должны загружаться уникальные данные

Оптимальными для правильного проведения машинного обучения считаются выборки данных, составленные вручную и содержащие максимум информации разного качества - это позволяет компьютеру выявлять неочевидные взаимосвязи между данными и делать по ним полезные выводы

К а т е г о р и
и з а д а ч
м а ш и н н о г
о
о б у ч е н и я

1

Регрессия - составление прогнозов на основе выборки данных с отличающимися признаками

2

Классификация - получение конкретного ответа на основании набора признаков

3

Кластеризация - разбивка данных на несколько групп

4

Уменьшение размерности - сокращение большого количества признаков для удобства их дальнейшей визуализации

5

Выявление аномалий - поиска отличий в наборах данных от стандартной информации

Типы машинного обучения

1

Обучение с учителем – оно предполагает использование полного набора снабженных признаками данных (размеченного дата сета) для тренировки системы ИИ на всех стадиях ее создания















2

Обучение без учителя – при котором машина самостоятельно выявляет закономерности, определяет признаки и классифицирует полученные данные

3

Обучение с подкреплением - когда машине требуется правильно решить поставленные перед ней задачи во внешней среде, располагая несколькими возможными вариантами действия

Августовский рейтинг языков программирования на основе индекса TIOBE

Aug 2022	Aug 2021	Change	Programming Language	Ratings	Change
1	2	▲	 Python	15.42%	+3.56%
2	1	▼	 C	14.59%	+2.03%
3	3		 Java	12.40%	+1.96%
4	4		 C++	10.17%	+2.81%
5	5		 C#	5.59%	+0.45%
6	6		 Visual Basic	4.99%	+0.33%
7	7		 JavaScript	2.33%	-0.61%
8	9	▲	 Assembly language	2.17%	+0.14%
9	10	▲	 SQL	1.70%	+0.23%
10	8	▼	 PHP	1.39%	-0.80%
11	16	▲▲	 Swift	1.27%	+0.30%
12	12		 Classic Visual Basic	1.27%	+0.04%
13	22	▲▲	 Delphi/Object Pascal	1.22%	+0.60%
14	23	▲▲	 Objective-C	1.22%	+0.61%

Среды разработки

- В качестве среды разработки используются платформы и среды Visual Studio 22, R-Studio, R-Brain, Eclipse, PyCharm, Spyder, IntelliJ IDEA, Jupyter Notebooks, Juno
- После создания SQL Server 2019, ориентированного на использование технологии больших данных и машинного обучения в среде платформы ASP.NET Core MVC 5 следует ориентироваться на Visual Studio 22

Глубокое обучение (Deep learning)

Глубокое обучение не только может дать результат там, где другие методы не сработают, но и позволяет построить более точную модель или же сократить время на ее создание
Расплачиваться же за это приходится еще большими вычислительными мощностями
Еще одним недостатком глубокого обучения является сложность интерпретации получаемых моделей

Глубокое обучение не только может дать результат там, где другие методы не сработают, но и позволяет построить более точную модель или же сократить время на ее создание; расплачиваться же за это приходится еще большими вычислительными мощностями. Еще одним недостатком глубинного обучения является сложность интерпретации получаемых моделей

1

2

3

Определяющая характеристика глубокого обучения — наличие более одного слоя между входом и выходом. Обычно, говоря о глубинном обучении, подразумевают использование глубоких нейронных сетей. Есть, однако, несколько алгоритмов, которые реализуют глубинное обучение, используя иные типы слоев

Data Science

Это концепция объединения статистики, анализа данных, машинного обучения и связанных с ними методов для понимания и анализа реальных явлений

Data Science — это работа с большими данными (англ. Big Data)

Большие данные — это огромные объёмы неструктурированной информации: например, метеоданные за какой-то период, статистика запросов в поисковых системах, результаты спортивных состязаний, базы данных геномов микроорганизмов и многое другое

Ключевые слова здесь — "огромный объём" и "неструктурированность"

Чтобы работать с такими данными, используют математическую статистику и методы МО

1

2

3

Специалист, который делает такую работу, называется **датасайентист** (или Data Scientist)

Он анализирует большие данные (Big Data), чтобы делать прогнозы

Какие именно прогнозы — зависит от того, какую задачу нужно решить

Итог работы дата-сайентиста — прогнозная модель

Если упростить, то это программный алгоритм, который находит оптимальное решение поставленной задачи

Data Mining

Широкое понятие, означающее извлечение знаний из данных

Data mining переводится как "добыча" или "раскопка данных"

Нередко рядом с Data Mining встречаются слова "обнаружение знаний в базах данных" (knowledge discovery in databases) и "интеллектуальный анализ данных"

Их можно считать синонимами Data Mining

1

Технологию Data Mining достаточно точно определяет Григорий Пиатецкий-Шапиро – один из основателей этого направления:

Data Mining – это процесс обнаружения в сырых данных

- ранее неизвестных
- нетривиальных
- практически полезных и доступных интерпретации знаний, необходимых для принятия решений в различных сферах человеческой деятельности

2

3

Data Mining является мультидисциплинарной областью, возникшей и развивающейся на базе достижений прикладной статистики, распознавания образов, методов искусственного интеллекта, теории баз данных и др. Отсюда обилие методов и алгоритмов, реализованных в различных действующих системах Data Mining

Многие из таких систем интегрируют в себе сразу несколько подходов. Тем не менее, как правило, в каждой системе имеется какая-то ключевая компонента, на которую делается главная ставка

Большие данные

1

Большие данные – Это набор подходов и методов, разработанных для анализа данных огромных размеров

2

Надо отметить, что **не всякая работа с данными относится к искусственному интеллекту** (например, аналитик, делающий вывод по графикам, не относится к искусственному интеллекту), и не все алгоритмы искусственного интеллекта разрабатываются с использованием больших данных (например, экспертные системы)

3

Тем не менее, множество современных интеллектуальных систем основано именно на обучении по данным: машинный перевод, распознавание изображений и речи, прогнозирование поведения клиентов и др. Во время обучения по данным алгоритм "изучает" большое количество реальных случаев (например, поведения клиентов или переводов текстов) и благодаря этому делает качественные предсказания для новых случаев

Источники больших данных

Источники сбора больших данных делятся на три типа:

- социальные
- машинные
- транзакционные

Социальные

Все, что человек делает в компьютерной сети, — источник социальных больших данных. Каждую секунду пользователи загружают в Instagram 1 тыс. фото и отправляют более 3 млн электронных писем. Ежесекундный личный вклад каждого человека — в среднем 1,7 мегабайта.

Другие примеры социальных источников Big Data — статистики стран и городов, данные о перемещениях людей, регистрации смертей и рождений и медицинские записи.

Машинные

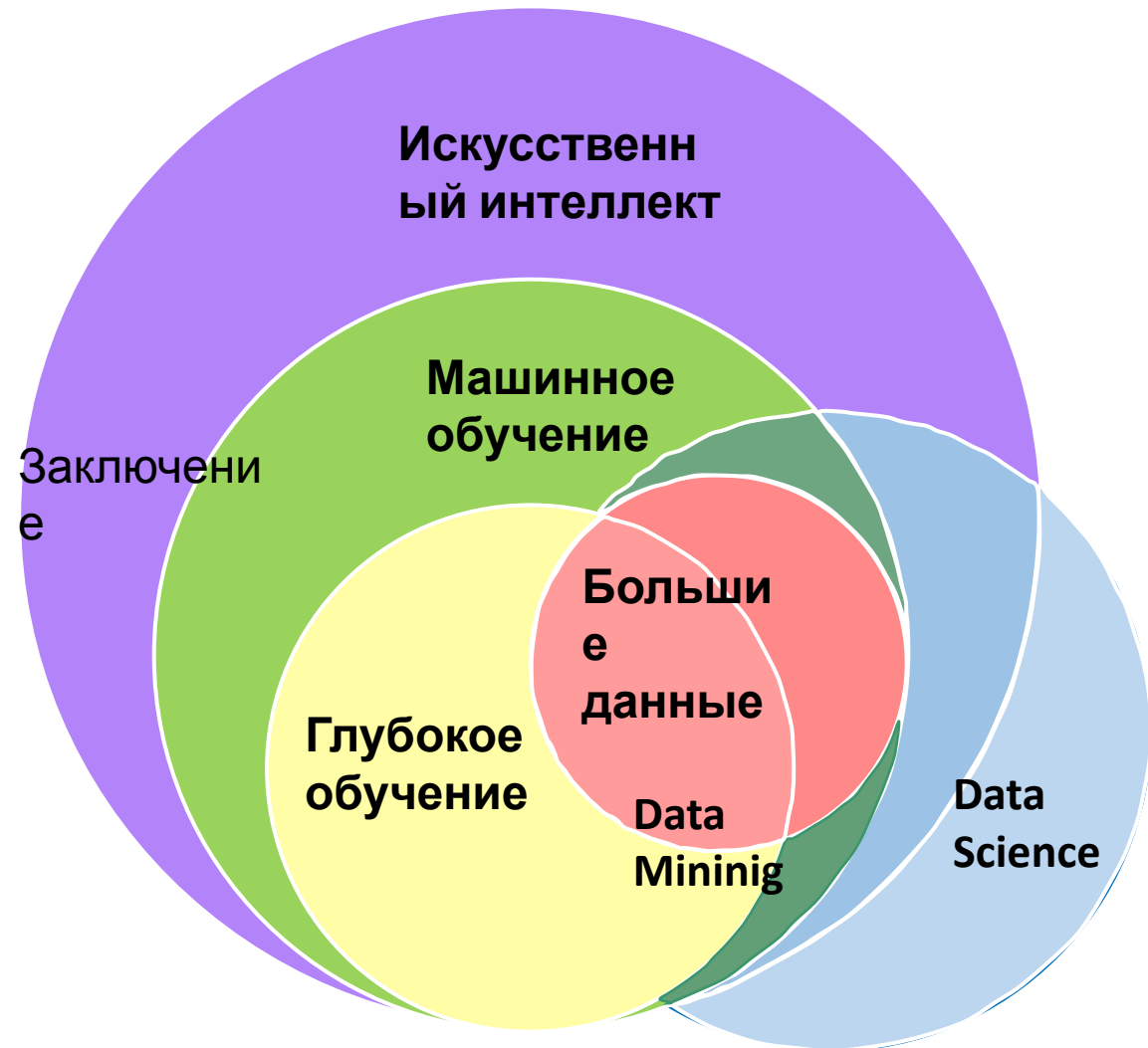
Большие данные также генерируются машинами, датчиками и "интернетом вещей". Информацию получают от смартфонов, умных колонок, лампочек и систем умного дома, видеокамер на улицах, метеоспутников и т.п.

Транзакционные

Транзакционные данные возникают при покупках, переводах денег, поставках товаров и операциях с банкоматами.

Заключение

Концептуально мы рассмотрели составляющие или технологии искусственного интеллекта:



Контрольные вопросы

1. Какое определение термина "Искусственный интеллект" наиболее близко к определению "Интеллект"? (Определение 1980 г)
2. О какой передаче знаний идет речь на этапе Wide AI?
3. На каком этапе развития ИИ находится современный ИИ? (NARROW AI)
4. На каком этапе развития ИИ произошел технологический взрыв? (NARROW AI)
5. Наличие каких 4-х факторов позволили ИИ стать глобальным трендом?
6. Согласно данным Deloitte относится ли создание чат-ботов к технологическим направлениям ИИ?
7. Перечислите категории задач МО (регрессия, классификация, кластеризация, уменьшение размерности, выявление аномалий).
8. Как называется тип МО, когда машине требуется правильно решить поставленные перед ней задачи во внешней среде, располагая несколькими возможными вариантами действия?
9. Какими недостатками характерны глубокому обучению? (большие вычислительные мощности и сложность интерпретации получаемых моделей)
10. Приведите пять примеров больших данных.
11. Кто один из основателей Data Mining (Григорий Пиатецкий-Шапиро)?
12. При разработке экспертных систем используются большие данные?
13. Перечислите источники больших данных.