

# Лекция. Искусственный интеллект и машинное обучение

Павлова

29 октября 2025 г.

ВОПРОСЫ:

- История развития искусственного интеллекта (ИИ)
- Проблема представления знаний
- Машинное обучение, подходы обучения
- Машинное обучение, подходы обучения

## 1 История развития искусственного интеллекта (ИИ)

Термин «искусственный интеллект» ввел в 1956 г. американский профессор математики Джон Маккарти. Он предположил, что точные науки способны описать человеческие мысли и рассуждения, и поэтому воспоминания и идеи могут быть выражены алгоритмами. Ученый вместе с коллегами организовал семинар в Дармутском колледже, по итогам которого была заложена общая методика исследования ИИ и создано профессиональное сообщество.

**Искусственный интеллект** — это широкий термин, который включает в себя компьютерные системы, имитирующие человеческий интеллект.

Исследования в области искусственного интеллекта начались в 1956 г., их философские корни уходят глубоко в прошлое. Вопрос, сможет ли машина думать, имеет долгую историю. Концепция начала формироваться с 1950-х, начала 20-го века.

Одним из способов определения ИИ является знаменитый тест Тьюринга: если вы можете принять компьютер за человека, он (искусственно) «умный».

**Тест Тьюринга эмпирический** - тест, идея которого была предложена Аланом Тьюрингом в статье «Вычислительные машины и разум», опубликованной в 1950 году в философском журнале Mind. Тьюринг задался целью определить, может ли машина мыслить.

Одно из самых первых упоминаний о «компьютерном интеллекте» было сделано в 1947 году. В докладе «Интеллектуальные машины» Тьюринг исследовал вопрос, может ли машина обнаруживать разумное поведение, и в рамках этого исследования предложил то, что может считаться предтечей его дальнейших исследований: «Нетрудно разработать машину, которая будет неплохо играть в шахматы.

А. Тьюринг назвал тест «игрой в имитацию». Стандартная интерпретация этого теста звучит следующим образом: «Человек взаимодействует с одним компьютером и одним человеком. На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Задача компьютерной программы — ввести человека в заблуждение, заставив сделать неверный выбор».

Все участники теста не видят друг друга. Если судья не может сказать определённо, кто из собеседников является человеком, то машина прошла тест. Для того чтобы протестировать интеллект машины, а не её возможность распознавать устную речь, беседа ведётся в режиме «только

текст», например, с помощью клавиатуры и экрана (компьютера-посредника). Переписка должна производиться через контролируемые промежутки времени, чтобы судья не мог делать заключения, исходя из скорости ответов. Во времена Тьюринга компьютеры реагировали медленнее человека. Оригинальный тест Тьюринга - не единственная версия - в 1990 году изобретатель Хью Лебнер изменил правила, увеличив время, которое человек проводит за разговором с компьютером (с пяти до 25 по Тьюрингу), а также количество людей, разговаривающих с внешней стороной (от двух до четырех).

**Робот** можно определить как универсальный автомат для осуществления механических действий, подобных тем, которые производит человек, выполняющий физическую работу. При создании первых роботов и вплоть до сегодняшнего дня образцом для них служат физические возможности человека. Именно стремление заменить человека на тяжелых работах и породило сначала идею робота, затем первые попытки ее реализации (в средние века) и, наконец, обусловило возникновение и развитие современной робототехники и роботостроения

**Термин «робот»**, как известно, славянского происхождения. Его ввел известный писатель К. Чапек в 1920 г. в своей фантастической пьесе «R.U.R.» («Россумовские универсальные роботы»), где так названы механические рабочие, предназначенные для замены людей на тяжелых физических работах. Название «робот» образовано от чешского слова *robot*, что означает тяжелый подневольный труд

Историк из Стэнфордского университета (США) доктор Адриен Мэр опубликовал исследование, доказывающее, что древние греки в своих мифах предсказали появление искусственного интеллекта, роботов, беспилотных машин и другие современные технологии.

**Корни робототехники** уходят в глубокую древность. Впервые возникли идеи и были предприняты первые попытки создания человекоподобных технических устройств, подвижных культовых статуй, механических слуг и т. п. Статуи богов с подвижными частями тела (руки, голова) появились еще в Древнем Египте, Вавилоне, Китае. В «Илиаде» Гомера божественный кузнец Гефест выковывает механических служанок. Аристотель упоминает о приводимых в движение с помощью ниток куклах-марионетках, из которых создавались целые механические театры. До нас дошли книги Герона Александрийского (I век н.э.), где описаны подобные и многие другие автоматы древности. В качестве источника энергии в них использовались вода, пар, гравитация (гири).

В средние века большой популярностью пользовались различного рода автоматы, основанные на использовании часовых механизмов. Были созданы всевозможные часы с движущимися фигурами людей, ангелов и т. п. К этому периоду относятся сведения о создании первых подвижных человекоподобных механических фигур — андроидов. Так, андроид алхимика Альберта Великого (1193—1280) представлял собой куклу в рост человека, которая, когда стучали в дверь, открывала и закрывала ее, кланяясь при этом входящему.

Целый ряд человекоподобных автоматов был создан швейцарскими часовщиками Пьером-Жаком Дро (1721—1790) и его сыном Анри Дро (1752—1791). От имени последнего было образовано позднее и само название «андроид». Эти человекоподобные игрушки представляли собой многопрограммные автоматы с оперативно сменяемыми программами. Программы задавались с помощью сменных кулачков, устанавливаемых на врачающемуся барабане, и других подобных механических устройств. Привод осуществлялся от часового механизма

## 2 Проблема представления знаний и экспертные системы

Экспертные системы-компьютерная система, способная частично заменить специалиста-эксперта в разрешении проблемной ситуации. Современные экспертные системы начали разрабатываться исследователями искусственного интеллекта в 1970-х годах, а в 1980-х годах получили коммерческое

подкрепление. Предшественники экспертных систем были предложены в 1832 году С. Н. Корсаковым, создавшим механические устройства, так называемые «интеллектуальные машины», позволявшие находить решения по заданным условиям, например, определять наиболее подходящие лекарства по наблюдаемым у пациента симптомам заболевания

**Экспертная система** – это вычислительная система, в которую включены знания специалистов о некоторой узкой предметной области в форме базы знаний. Экспертные системы должны уметь принимать решения вместо специалиста в заданной предметной области

Проблема представления знаний

1. Представление знаний. В рамках этого направления решаются задания, связанные с формализацией и представлением знаний в памяти системы ИИ. Для этого разрабатываются специальные модели представления знаний и языка описания знаний, внедряются разные типы знаний. Проблема представления знаний есть одной из основных проблем для системы ИИ, потому что функционирование такой системы опирается на знания о проблемной области, которые хранятся в ее памяти.

2. Манипулирование знаниями. Для того чтобы знаниями можно было пользоваться при решении задачий, стоит научить систему ИИ оперировать ими. В рамках данного направления разрабатываются способы пополнения знаний на основе их неполных описаний, создаются методы достоверного и правдоподобного вывода на основе имеющихся знаний, предлагаются модели рассуждений, которые опираются на знание и особенности, которые имитируют человеческие рассуждения. Манипулирование знаниями очень тесно связано с представлением знаний, и разделить эти два направления можно лишь условно.

Проблема представления знаний способствовала развивающееся параллельно направление компьютерной лингвистики. Системы, основанные на знаниях, нашли широкое применение в виде экспертных систем, с которыми одно время отождествлялся ИИ. Экспертные системы оказались способны строить формальные описания задач, сформулированных на ограниченном естественном языке для одной узкой предметной области, знания о которой были в них заложены

**В настоящее время наиболее известны четыре подхода к представлению знаний:** продукционная модель; логическая модель; семантические сети; фреймы.

**Продукционные правила** – наиболее простой способ представления знаний. Он основан на представлении знаний в форме правил, структурированных в соответствии с образцом "Если - То". Часть правила "Если" называется посылкой, а "То" – выводом или действием. Правило в общем виде записывается так: Если А<sub>1</sub>, А<sub>2</sub>, ..., А "То" В. Такая запись означает, что "если все условия от А<sub>1</sub> до А" являются истинными, то В также истинно" или "когда все условия от А<sub>1</sub> до А" выполняются, то следует выполнить действие В". Например, рассмотрим правило: Если у является отцом х, з является братом у, то з является дядей х. В данном случае описано универсальное правило для трех переменных: х, у, з. Подставляя конкретные значения (Вася, Федя, Петя) вместо переменных можно получить конкретные различные знания. Конечно, при решении задач в искусственном интеллекте, используются куда более сложные постановки вопроса.

**Знания, представленные в интеллектуальной системе**, образуют базу знаний. В интеллектуальную систему входит также механизм выводов, который позволяет на основе знаний, имеющихся в базе знаний, получить новые знания. Простота и наглядность представления такой модели знаний обусловила применения данной системы во многих системах. Кроме того, данная модель может служить основой языка программирования, ориентированного на знания.

Основные элементы экспертной системы:

**Интерфейс пользователя** – комплекс программ, реализующих диалог конечного пользователя с ЭС как на стадии ввода информации, так и при получении результатов.

**База знаний (БЗ)** – ядро ЭС, совокупность знаний предметной области, записанная на машинный носитель в форме, понятной эксперту и пользователю (обычно на некотором языке, при-

ближенном к естественному). Параллельно такому «человеческому» представлению существует БЗ во внутреннем «машинном» представлении.

**Решатель** (дедуктивная машина, машина вывода, блок логического вывода) — программа, моделирующая ход рассуждений эксперта на основании информации, имеющейся в БЗ.

**Подсистема объяснений** — программа, позволяющая пользователю получить ответы на вопросы: «Как была получена та или иная рекомендация?», т.е. описание всего процесса получения решения с указанием использованных фрагментов БЗ (всех шагов цепи умозаключений), «Почему система приняла такое решение?», т.е. ссылка на умозаключение, непосредственно предшествовавшее полученному решению — отход на один шаг назад. Развитые подсистемы объяснений поддерживают и другие типы вопросов.

**Интеллектуальный редактор БЗ** — программа, предоставляющая инженеру по знаниям возможность создавать БЗ в диалоговом режиме. Включает в себя систему вложенных меню, шаблонов языка представления знаний, подсказок («help»-режим) и других сервисных средств, облегчающих работу с базой. Данный элемент ЭС предназначен для работы инженера по знаниям

### 3    Машинное обучение

Машинное обучение основное направление исследований в области искусственного интеллекта, сформировалось в самостоятельное направление в 1980 гг.

Теория обучения машин (machine learning, машинное обучение) развилась благодаря: прикладной статистике, численным методам оптимизации, дискретному анализу. Методы машинного обучения составляют основу относительно новой дисциплины — интеллектуальный анализ данных (data mining).

**Машинное обучение (Machine Learning)** — обширный подраздел искусственного интеллекта, изучающий методы построения алгоритмов, способных обучаться. Машинное обучение находится на стыке математической статистики, методов оптимизации и классических математических дисциплин, но имеет также и собственную специфику, связанную с проблемами вычислительной эффективности и переобучения. Многие методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам. Многие методы тесно связаны с технологиями получения информации и интеллектуальным анализом данных (Data Mining).

Основная цель машинного обучения - предсказание результата на основе предудыщих накопленных данных.

**Обучение с учителем** — это направление машинного обучения, объединяющее алгоритмы и методы построения моделей на основе множества примеров, содержащих пары «известный вход — известный выход».

Обучение под контролем используется чаще всего. В этом случае исследователь данных выступает в качестве наставника, демонстрируя, какие результаты должен получить алгоритм. Как ребенок учится различать фрукты, запоминая рисунки в книге, так алгоритм практикуется на специальным образом маркированных наборах данных с преопределеными результатами.

Для этого типа обучения используются такие алгоритмы, как линейная и логистическая регрессия, мультиклассовая классификация и метод опорных векторов. Алгоритм с учителем, он должен работать с примерами, которые содержат не только вектор независимых переменных (атрибутов, признаков), но и значение, которое должна выдавать модель после обучения (такое значение называется целевым). Разность между целевым и фактическим выходами модели называется

**ошибкой обучения (невязкой, остатками)**, которая минимизируется в процессе обучения и выступает в качестве «учителя». Значение выходной ошибки затем используется для вычисления коррекций параметров модели на каждой итерации обучения.

К числу алгоритмов обучения с учителем для решения задач классификации относятся:

- деревья решений;
- машины опорных векторов;
- байесовский классификатор;
- метод k-ближайших соседей;
- линейный дискриминантный анализ;

Алгоритмами обучения с учителем для решения задачи регрессии являются:

- линейная регрессия;
- логистическая регрессия;
- нейронные сети.

**Data Mining** — исследование и обнаружение "машиной" (алгоритмами, средствами искусственного интеллекта) в сырых данных скрытых знаний, которые ранее не были известны, нетривиальны, практически полезны, доступны для интерпретации человеком. Термин предложил 1996 г. Григорий Пятецкий-Шапиро. Цель Data Mining состоит в извлечении знаний из больших объемов данных.

- Кластеризация: разделение на кластеры, группы (нет меток классов, без учителя)
- Классификации: разделение на кластеры, группы (с учителем)
- Регрессия: подобно задаче классификации позволяет определить по известным характеристикам объекта значение некоторого его параметра В отличие от задачи классификации значением параметра является не конечное множество классов, а множество действительных чисел
- Ассоциативные правила для прогнозирования: нахождение частых зависимостей (или ассоциаций) между объектами или событиями

**Большие данные** — технологии сбора, обработки и хранения структурированных и неструктурированных массивов информации, характеризующихся значительным объемом и быстрой скоростью изменений (в том числе в режиме реального времени), что требует специальных инструментов и методов работы с ними. Теоретические разделы машинного обучения объединены в отдельное направление теорию вычислительного обучения (Computational Learning Theory, COLT). Машинное обучение тесно связано с цифровыми технологиями.

**Цифровые технологии** — технологии сбора, хранения, обработки, поиска, передачи и представления данных в электронном виде.

«**Сквозные**» **цифровые** технологии используются для сбора, хранения, обработки, поиска, передачи и представления данных в электронном виде, в основе функционирования которых лежат программные и аппаратные средства и системы, создающие новые рынки и изменяющие бизнес-процессы.

**Искусственный интеллект AI** — это система программных и (или) аппаратных средств, способная с определенной степенью автономности воспринимать информацию, учиться и принимать решения на основе анализа больших массивов данных, в том числе имитируя человеческое поведение. Развитие AI связывают с киберфизическими системами (Cyber-Physical System, CPS). Такие системы частично или полностью замещают, дополняют функционирование нервной системы биологического объекта, в том числе на основе искусственного интеллекта. Термин «киберфизические системы» предложила в 2006 г Хелен Джайлл.

**Киберфизические системы**, состоят из различных природных объектов, искусственных подсистем и управляющих контроллеров, позволяющих представить такое образование как единое целое. Обеспечивает связь и координацию между вычислительными и физическими ресурсами. Компьютеры осуществляют мониторинг и управление физическими процессами с использованием такой петли обратной связи, где происходящее в физических системах оказывает влияние на вычисления и наоборот Киберфизические системы объединяют модели двух типов:

- традиционные инженерные модели (механические, строительные, электрические, биологические, химические, экономические и другие)
- модели компьютерные.

**Появилась новая дисциплина проектирование моделей (model engineering).**

**Умный транспорт (Smart Transportation)** оборудован различными компьютеризированными встроенными системами управления на разных уровнях.

Кибернетические решения в области разработки «умного транспорта» связаны с созданием полноценной связанной системы:

- между машинами (Vehicle-to-Vehicle, V2V)
- между машиной и внешней окружающей ее инфраструктурой (Vehicle-to-Infrastructure, V2I)

**Системы машинного обучения (подходы к машинному обучению).**

- С учителем и без учителя
- Частичное - semisupervised learning
- С подкреплением - Reinforcement Learning
- Динамическое обучение
- Глубокое обучение

**Обучение без учителя**

Самостоятельное обучение подразумевает большую независимость: компьютер учится распознавать сложные процессы и алгоритмы без постоянного контроля со стороны человека. Такой тип обучения подразумевает отсутствие маркировки данных и конкретных предопределенных результатов.

**Обучение без учителя** — это класс задач, при решении которых ответы неизвестны или не существуют, и требуется найти некоторые закономерности в данных на основе признаковых описаний. Методы обучения без учителя применяются при решении следующих задач:

- Кластеризация — задача разделения объектов на группы, обладающие некоторыми свойствами. Примером может служить кластеризация документов из электронной библиотеки или кластеризация абонентов мобильного оператора.
- Оценивание плотности — задача приближения распределения объектов. Примером может служить задача обнаружения аномалий, в которой на этапе обучения известны лишь примеры «правильного» поведения оборудования (или, скажем, игроков на бирже), а в дальнейшем требуется обнаруживать случаи некорректной работы (соответственно, незаконного поведения игроков). В таких задачах сначала оценивается распределение «правильных» объектов, а затем аномальными оказываются все объекты, которых в рамках этого распределения получают слишком низкую вероятность.
- Визуализация — задача изображения многомерных объектов в двумерном или трехмерном пространстве таким образом, чтобы сохранялось как можно больше зависимостей и отношений между ними.

- Понижение размерности — задача генерации таких новых признаков, что их меньше, чем исходных, но при этом с их помощью задача решается не хуже (или с небольшими потерями качества, или лучше — зависит от постановки). К этой же категории относится задача построения латентных моделей, где требуется описать процесс генерации данных с помощью некоторого (как правило, небольшого) набора скрытых переменных.
- Ассоциативный анализ. Целью ассоциативного анализа является выявление скрытых связей или правил между объектами в наборе данных. Алгоритмы ассоциативного анализа находят часто встречающиеся комбинации товаров или признаков и позволяют строить рекомендательные системы, анализировать покупательское поведение или проводить маркетинговые исследования.

При обучении без учителя использующиеся алгоритмы: кластеризация методом k-средних, анализ основных и независимых компонентов и ассоциативные правила.

Примеры: обнаружение аномалий в сетевой безопасности. Алгоритмы обучения без учителя могут помочь выявить подозрительное поведение в компьютерных сетях и обнаружить аномальную активность; группировка новостных статей. При анализе большого объема текстовых данных, таких как новостные статьи или блоги, алгоритмы кластеризации могут помочь автоматически группировать статьи по схожим темам; рекомендательные системы. Методы ассоциативного анализа могут применяться для поиска скрытых связей между продуктами или интересами пользователей, что позволяет строить персонализированные рекомендации.

**Частичное обучение** Обучение с частичным привлечением учителя занимает промежуточное положение между обучением с учителем и без учителя. Когда получение достаточного количества размеченных данных затруднено (например, когда при разметке данных привлекаются дорогостоящие устройства или квалифицированные лица), помимо размеченных данных можно также за- действовать и неразмеченные данные для построения более эффективных моделей, по сравнению с моделями, построенными с полным участием учителя или без него вовсе.

Существует несколько подходов к обучению с частичным привлечением учителя:

**методы на основе кластеризации.** В этом подходе неразмеченные данные сначала кластеризуются, а затем каждому кластеру присваивается метка класса на основе имеющихся помеченных данных;

**методы на основе графов.** В этом подходе данные представляются в виде графа, где узлы представляют собой примеры данных, а ребра — связи между ними. Затем используются методы распространения разметки, чтобы расширить метки классов на основе имеющихся.

**Обучение с подкреплением** В обучении с подкреплением присутствует понятие агента и среды. Среда может быть реальной или виртуальной. Агент взаимодействует со средой и обучается принимать последовательность действий в окружении, после чего получает обратную связь в виде награды или штрафа.

Одним из основных компонентов обучения с подкреплением является функция оценки состояния, которая предсказывает ожидаемую награду. Целью агента является настройка стратегии действий таким образом, чтобы максимизировать накопленную награду на протяжении взаимодействия с средой. Агент использует эту функцию для выбора оптимальных действий и оценки своего текущего состояния.

Одним из наиболее популярных алгоритмов в обучении с подкреплением является метод Q-обучения. В этом методе агент обучается оценивать и выбирать действия на основе значения Q-функции, которая представляет ожидаемую суммарную награду за выполнение действия в определенном состоянии. Алгоритм Q-обучения основан на принципе итерационного обновления значения Q-функции на основе накопленной награды и последующего выбора оптимальных действий.

Обучение с подкреплением имеет широкий спектр применений в различных областях. Например, в робототехнике агент может управлять роботом для преодоления препятствий или выполнения задач. В игровой индустрии методы обучения с подкреплением используются для обучения виртуальных персонажей или улучшения их стратегий в играх.

**Динамическое (потоковое) обучение** – алгоритмы способны к обучению на лету (или пакетные). Подход к задачам обучения — это концепция, парадигма, точка зрения на процесс обучения, приводящая к набору базовых предположений, гипотез, эвристик, на основе которых строится модель, функционал качества и методы его оптимизации. Разделение методов «по подходам» условно.

**Потоковое обучение** представляет собой метод машинного обучения, в котором модель обновляется непрерывно по мере поступления новых данных. В отличие от традиционных методов, где данные разбиваются на независимые партии, потоковое обучение позволяет обрабатывать данные непрерывно и реагировать на изменения в режиме реального времени.

Для реализации потокового обучения используются следующие подходы:

- стохастический градиентный спуск (SGD). Этот метод обновляет модель после каждого образца данных, что позволяет адаптироваться к изменению данных;
- алгоритмы с высокой скоростью обучения. В потоковом обучении важно эффективно использовать ресурсы и минимизировать время обучения модели. Поэтому разработаны алгоритмы с высокой скоростью обучения, такие как алгоритмы на основе решающих деревьев, каскадные классификаторы и алгоритмы с динамическим обновлением модели;
- алгоритмы обнаружения изменений. В потоковом обучении информация может изменяться со временем. Алгоритмы обнаружения изменений позволяют моделям отслеживать и реагировать на изменения в данных, поддерживая актуальность модели;
- параллельное и распределенное обучение. Потоковое обучение часто сочетается с параллельным или распределенным обучением. Путем распределения обработки данных между несколькими узлами можно сократить время обучения и обеспечить масштабируемость. Потоковое обучение находит свое применение во многих областях, таких как онлайн-рекомендации, финансовые прогнозы, обнаружение аномалий, анализ социальных медиа и многих других. Например, в задачах рекомендаций онлайн-платформ потоковое обучение позволяет моделям быстро адаптироваться к предпочтениям пользователей и изменениям в их поведении.

**Глубокое обучение** специализируется на создании и обучении нейронных сетей с большим количеством слоев. Глубокое обучение стало важным и мощным инструментом в решении сложных задач в различных областях, таких как компьютерное зрение, распознавание речи, естественный язык и многих других.

Преимущества глубокого обучения заключаются в его способности извлекать высокоуровневые признаки из сложных исходных данных. Это достигается благодаря глубокой архитектуре нейронных сетей, которая позволяет модели самостоятельно выявлять и иерархически представлять сложные зависимости в данных.

Одним из наиболее популярных типов глубоких нейронных сетей являются сверточные сети. Они весьма эффективны при обработке изображений, так как специализированы на выявление локальных шаблонов или признаков в изображениях.

Еще одним важным классом глубоких нейронных сетей являются рекуррентные сети, которые обладают способностью моделировать последовательные и временные зависимости в данных. Рекуррентные нейронные сети широко используются для задач обработки естественного языка, машинного перевода и анализа временных рядов.

В последние годы появились новые архитектуры, такие как генеративно-состязательные сети, которые используются для генерации новых данных с высоким уровнем реалистичности. Также стоит отметить трансформеры, которые получили широкое распространение в области обработки естественного языка и машинного перевода.