



Технологические возможности реализации искусственного интеллекта

Курс «Искусственный
интеллект»



Оглавление

Введение	2
Термины, используемые в лекции	2
Базовые технологии искусственного интеллекта	3
Инфраструктурный уровень	4
Уровень алгоритмов	7
Специализации в области ИИ	10
Ограничения в использовании ИИ	14
Что можно почитать ещё?	16

Введение

Всем привет!

Мы продолжаем путешествие в мир искусственного интеллекта (ИИ). На прошлой лекции мы рассмотрели некоторые области применения ИИ, позволяющие машинам выполнять ряд сложных задач, для решения которых требуется человек разумный. Очевидно, что для решения той или иной задачи требуются определённые технические возможности. Причём для реализации разных задач могут потребоваться совершенно разные инструменты.

Сегодня мы и займёмся рассмотрением ключевых компонентов, из которых строится платформа искусственного интеллекта.

Термины, используемые в лекции

Алгоритм — конечная совокупность точно заданных правил решения некоторого класса задач или набор инструкций, описывающих порядок действий исполнителя для решения определённой задачи.

Программное обеспечение (ПО) — программа или множество программ, используемых для управления компьютером.

Аппаратное обеспечение, аппаратные средства, компьютерные комплектующие, «железо» (англ. hardware) — электронные и механические части вычислительного устройства, входящие в состав системы или сети, исключая программное обеспечение и данные (информацию, которую вычислительная система хранит и обрабатывает).

Центральный процессор (ЦП, central processing unit, CPU) — электронный блок либо интегральная схема, исполняющая машинные инструкции (код программ), главная часть аппаратного обеспечения компьютера или программируемого логического контроллера.

Графический процессор (graphics processing unit, GPU) — это специализированная электронная схема, предназначенная для манипулирования и преобразования памяти с целью ускорения создания изображений, предназначенных для вывода на устройство отображения.

Тензорный процессор Google (Google Tensor Processing Unit, Google TPU) — тензорный процессор, относящийся к классу нейронных процессоров, являющийся специализированной интегральной схемой, разработанной корпорацией Google и предназначенной для использования с библиотекой машинного обучения TensorFlow.

Машинное обучение (machine learning, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение благодаря применению решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, математического анализа, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме.

Базовые технологии искусственного интеллекта

Ключевые технологии искусственного интеллекта можно разделить на два уровня: *инфраструктурный* и *уровень алгоритмов*.

Инфраструктурный уровень — это базовое оборудование, то есть машины, механизмы, приборы, устройства и т. п. Иными словами, это приспособления, которые служат для реализации той или иной задачи в области ИИ.

Поверх этого базового оборудования находится **уровень алгоритмов**. Алгоритм — это совокупность точно заданных правил или набор инструкций, описывающих порядок действий для решения определённой задачи. Также этот уровень включает данные. Этот уровень абстрактный.

Приведу простую аналогию. Если перед нами стоит задача добраться из одного города в другой, то инфраструктурный уровень — это автомобили, дороги, светофоры и так далее, а уровень алгоритмов — это порядок действий необходимых для управления автомобилем и правила дорожного движения, а также маршрут из точки А в точку Б.

Другая терминология, которую вы можете услышать — **это аппаратное и программное обеспечение**.

Аппаратное обеспечение или «железо» — это электронные и механические части вычислительного устройства, входящие в состав системы или сети. Программное обеспечение — это компьютерные программы.

Может возникнуть вопрос: аппаратное и программное обеспечение — это составляющие ИТ, информационных технологий, в чём же разница между ИИ и ИТ?

Информационные технологии — это то, с чем мы ежедневно сталкиваемся, сегодня ИТ проникло практически во все сферы нашей жизни: начиная от вашего смартфона, заканчивая изучением глубокого космоса. ИИ является частью ИТ, которая превратила обычные информационные технологии в интеллектуальные системы.

Но всё же искусственный интеллект и информационные технологии — это не одно и то же. ИИ направлен на выполнение задач, включающих обучение, адаптацию, выполнение, обработку информации, такой как распознавание речи, в то время как системы информационных технологий основаны на оценке, хранении, захвате и анализе данных.

Информационные технологии ограничиваются передачей и манипулированием данными. В то время как приложения, управляемые искусственным интеллектом, работают для решения проблем и повышения эффективности ИТ-операций. Это делает программное обеспечение с искусственным интеллектом более «интеллектуальным».

Рассмотрим более детально каждый уровень.

Инфраструктурный уровень

Базовое аппаратное обеспечение

Что мы понимаем под базовым аппаратным обеспечением инфраструктуры искусственного интеллекта? Этот уровень можно разделить на четыре направления: CPU, GPU, специализированный чип и высокоскоростная сеть. Поговорим о CPU и GPU.

CPU (central processing unit) — это центральный процессор — электронный блок либо интегральная схема, исполняющая машинные инструкции (код программ), главная часть аппаратного обеспечения компьютера.

Центральный процессор обрабатывает основные инструкции компьютера, такие как арифметические, логические функции и операции ввода-вывода. Обычно это небольшой, но мощный чип, встроенный в материнскую плату компьютера.

Центральный процессор считается «мозгом» компьютера, поскольку он интерпретирует и выполняет большинство аппаратных и программных инструкций компьютера.

Центральным архитектурным элементом CPU является ядро, в котором происходят все вычисления и логические операции. На заре своего появления процессоры были одноядерными, но современные процессоры являются многоядерными и имеют два или более ядер для повышения производительности. Центральный процессор обрабатывает задачи последовательно, распределяя их между несколькими ядрами для достижения многозадачности.

GPU (graphics processing unit) — это графический процессор, специализированная электронная схема, которая использует ускоренные вычисления для рендеринга (вычисления) изображений и графики высокого разрешения. Хотя изначально GPU предназначались для рендеринга 2D и 3D изображений, видео и анимации на компьютере, сегодня они используются в приложениях, выходящих далеко за рамки обработки графики, включая искусственный интеллект.

GPU функционируют аналогично центральным процессорам и содержат схожие компоненты (например, ядра, память и т. д.). Они могут быть интегрированы в CPU или быть дискретными (т. е. отдельными от CPU с собственной оперативной памятью).

Главным отличием GPU является то, что в нём используется параллельная обработка, когда задачи разделяются на более мелкие подзадачи, которые распределяются между огромным количеством процессорных ядер в GPU. Это

позволяет ускорить обработку специализированных вычислительных задач. За счёт пакетного выполнения инструкций и обработки огромных объёмов данных они могут ускорять рабочие нагрузки, превосходящие возможности центрального процессора.

Кроме того, существует специальный чип, специально разработанный для алгоритмов искусственного интеллекта, — чип TPU от Google.

Тензорный процессор Google (Google Tensor Processing Unit, Google TPU) — тензорный процессор, относящийся к классу нейронных процессоров, являющийся специализированной интегральной схемой, разработанной корпорацией Google и предназначенной для использования с библиотекой машинного обучения TensorFlow.

Тензоры — это просто математические объекты, которые можно использовать для описания физических свойств, так же как скаляры и векторы. TPU как и GPU выполняют тензорные операции, однако TPU ориентированы на выполнение больших тензорных операций, которые часто присутствуют в обучении нейронных сетей по сравнению с рендерингом 3D-графики.

Наконец, чтобы полностью использовать возможности искусственного интеллекта, нам нужна высокоскоростная сеть. Во время обучения и расчёта некоторых сложных моделей данных нам необходимы огромные гарантии пропускной способности сети. В настоящее время сеть стала важной частью общей производительности машинного обучения.

Вопрос — можно ли использовать среднестатистический домашний компьютер для разработки и функционирования ИИ? В целом да, это возможно. Однако, есть проблема, которая в первую очередь заключается во времени, которое вам нужно для проверки гипотез. Я сейчас немного забегаю вперёд, потому что о разработке приложений ИИ мы будем говорить в следующей лекции, но идея состоит в том, что для решения той или иной задачи вы не можете мгновенно предоставить готовое решение. Обычно у вас есть несколько предположительных вариантов (гипотез), которые вам нужно проверить. На обычном домашнем компьютере проверка одной гипотезы, то есть проведение всех вычислений может занимать огромное количество времени. Простая аналогия здесь — рендеринг одного изображения на слабой видеокарте занимает много времени, аналогично и с вычислениями для ИИ.

В качестве примера познакомимся с рабочей станцией NVIDIA DGX.

Это рабочая станция, специально разработанная для ИИ, основанная на технологии NVIDIA NVLink, с восемью GPU H100. Она может достигать производительности в 32

петафлопс, что в сотни раз превышает возможности традиционного сервера. Рабочая станция компактна (она может поместиться под столом).

В целом существует четыре варианта развёртывания рабочих станций для ИИ:

1. **Традиционные облачные провайдеры** — к ним относятся такие крупные поставщики, как Amazon Web Services (AWS), Microsoft Azure и Google Cloud.
2. **Облачные провайдеры, специализирующиеся на глубоком обучении** — это облачные предложения, специально разработанные для поддержки рабочих процессов глубокого обучения, например, с акцентом на программные возможности GPU. Примером может служить компания Paperspace.
3. **Готовые локальные серверы глубокого обучения** — рабочие станции глубокого обучения, предлагаемые такими компаниями, как NVIDIA (например, системы DGX).
4. **Проприетарные (собственные) рабочие станции DL** — они создаются организацией с нуля.

Уровень алгоритмов

Алгоритмы искусственного интеллекта представляют собой сложный набор правил, на основе которых работают программы, определяя их шаги и способность к обучению. В рамках ИИ используется большое количество всевозможных алгоритмов, но можно выделить два больших направления — это машинное обучение и глубокое обучение.

Машинное обучение (machine learning) — это одна из основных концепций искусственного интеллекта. Все мы должны учиться, и передача знаний человеком также осуществляется с помощью метода обучения. Мы усваиваем знания наших предков, а затем создаём новые знания путём умозаключений. Мы надеемся, что машина обладает такой же способностью: усвоив предыдущую информацию, машина становится похожей на интеллектуальную и может соответствующим образом реагировать на новые данные в будущем. Это называется машинным обучением.

В первом приближении посмотрим каков наиболее типичный процесс машинного обучения. Сначала мы собираем исторические данные, затем обучаемся на исторических данных. В результате обучения мы получаем модель. Затем, когда в эту модель вводятся новые данные, мы можем предсказать результаты. Это наиболее типичный процесс машинного обучения. Мы можем просто сказать, что в

результате обучения создаётся модель, а затем эта модель используется для прогнозирования новых данных.

В качестве примера мы возьмём прогнозирование цен на жильё. Мы предполагаем, что цены на жильё связаны только с площадью дома. Тогда мы можем ввести одномерные данные, такие как исторические цены на дома, и одномерные данные о площади домов. В модели цены дома мы можем предсказать цену дома после ввода данных о площади нового дома после обучения. Это пример простейшего машинного обучения.

В машинном обучении мы можем выделить алгоритмы обучения с учителем (supervised) и обучения без учителя (unsupervised), а также обучение с подкреплением (reinforcement). В чём между ними разница?

Допустим, нам нужно обучить модель, которая будет предсказывать есть ли у человека грипп или нет по двум признакам: температура тела и головная боль. Для этого мы представляем модели машинного обучения две группы случаев: если температура тела >37 и есть головная боль — это грипп, если хотя бы один из признаков не соответствует этим критериям — гриппа нет. То есть мы заранее определяем критерии, с помощью которых модель будет классифицировать случаи.

В случае обучения без учителя — мы не делаем меток: этот человек болен, а этот нет. Алгоритм должен сам отнести каждый случай к той или иной категории.

Обучение с подкреплением также является важной ветвью машинного обучения. Его суть заключается в том, чтобы решить проблему принятия решений, то есть принимать решения автоматически, а также принимать решения непрерывно. В основном оно включает две части, первая — это агент, а вторая — вознаграждение за действие в окружающей среде. Целью обучения с подкреплением является получение наибольшего суммарного вознаграждения. Рассмотрим пример: ребёнок, который учится ходить, — это процесс обучения с подкреплением. Ребёнок — это агент, и он надеется изменить состояние окружающей среды посредством такого действия, как ходьба. Каждый его шаг — это изменение состояния. Если он делает правильный шаг, мы даём ему вознаграждение. В противном случае, если он не делает правильный шаг, мы не даём ему вознаграждения. Благодаря этому процессу малыш постепенно научится ходить.

Работу этого алгоритма можно увидеть на видео: [Deep Learning Cars](#).

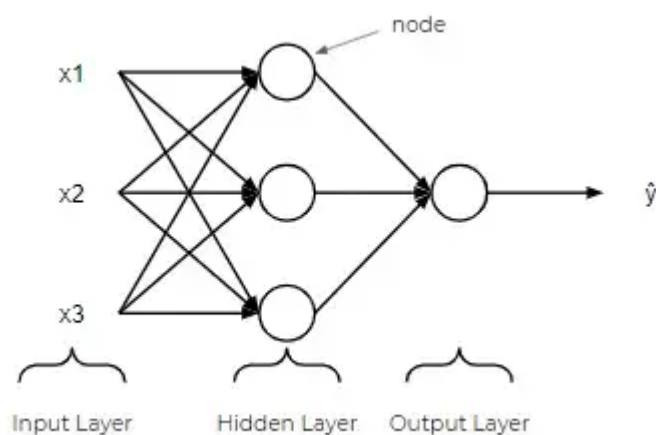
Алгоритмы глубокого обучения

Когда мы говорим об искусственном интеллекте, мы часто слышим такое понятие, как глубокое обучение. На самом деле глубокое обучение является особой формой машинного обучения. Глубокое обучение основано на алгоритмах нейронных сетей.

Большинство из того, что мы обсуждали на прошлой лекции основано на глубоком обучении, а именно распознавание изображений, распознавание речи, обработка естественного языка, распознавание аудио и т. д.

Когда речь заходит об алгоритмах глубокого обучения, мы должны понимать, что в основе лежат нейронные сети. Нейронная сеть похожа на нейронную передачу человеческого мозга, от одного входного блока к следующему входному блоку для получения результата. Принцип работы простой нейронной сети заключается в имитации передачи информации по нервам в человеческом мозге, где информация передаётся от одного нейрона к другому.

Я не хочу слишком углубляться в механику, но стоит показать вам, как выглядит структура базовой нейронной сети.



В нейронной сети есть входной слой, один или несколько скрытых слоёв и выходной слой. Входной слой состоит из одной или нескольких переменных признаков (или входных переменных, или независимых переменных), обозначаемых как x_1 , x_2 , ..., x_n . Скрытый слой состоит из одного или нескольких скрытых узлов или скрытых блоков. Узел — это просто один из кругов на диаграмме выше. Аналогично, выходная переменная состоит из одного или нескольких выходных узлов.

Поясню на примере. В вышеприведённой нейронной сети может быть три входные переменные: площадь квартиры, количество комнат и адрес. Подав в нейронную сеть эти три части информации, она выдаст на выходе цену квартиры. Как именно она это делает?

Не вдаваясь в детали, можно сказать, что нейронная сеть — это не что иное, как сеть уравнений. Каждый узел в нейронной сети — это математическое уравнение, которое «отбирает» определённые, значимые признаки и «разбирает» входящую информацию на детали, чтобы на основании этого выдать ответ.

Как правило, большее количество узлов и слоёв позволяет нейронной сети выполнять гораздо более сложные вычисления.

Мы разобрали технологическую базу, лежащую в основе ИИ, но пока машины не научились управлять сами собой, поэтому работу по настройке оборудования, программированию всё ещё приходится делать людям. Хотя это новая и нишевая область, карьера в сфере искусственного интеллекта не является однородной. В рамках искусственного интеллекта существуют различные виды профессий, требующие определённых навыков и опыта. Рассмотрим сначала специализацию, которая называется инженер по машинному обучению, а также навыки, которые требуются для такого специалиста.

Специализации в области ИИ

Инженер по машинному обучению

Стать инженером машинного обучения всё ещё не так просто, как стать веб-разработчиком или разработчиком мобильных приложений. Если спросить людей, носящих звание «инженер машинного обучения», чем они занимаются, можно получить совершенно разные ответы. Чтобы понять, чем же занимается инженер машинного обучения, я хочу сделать небольшой обзор навыков, которые необходимы в этой профессии, а именно soft-skills, основных технических навыков и более специализированных навыков.

Soft-skills

Прежде чем мы перейдём к техническим навыкам, поговорим о некоторых нетехнических навыках, которые следует иметь в виду. Прежде всего требуются знания математики и информатики. Однако, скорее всего, вам придётся искать способы применения этих знаний для решения реальных проблем.

Получение новых навыков. Эта область быстро меняется. Каждый месяц появляются новые модели нейронных сетей, которые превосходят предыдущую архитектуру. Производители графических процессоров ведут гонку вооружений. В 2017 году практически все крупные технологические гиганты выпустили собственные фреймворки машинного обучения. Нужно многое успеть, но, к счастью, способность быстро учиться — это то, что можно улучшить. Например, есть замечательная книга «Думай как математик» Барбары Оакли, в которой описаны техники, помогающие в усвоении новой информации.

Тайм-менеджмент. Умение управлять временем является гораздо более важным фактором успеха, чем природный талант или врождённый интеллект. И это универсальное правило. Время, которое вы тратите на обучение, является основным узким местом, особенно если учесть индивидуальные способности концентрироваться на одной задаче в течение длительного времени.

Знание бизнеса/домена. Наиболее успешными проектами в области машинного обучения будут те, что направлены на решение реальных проблем. В академической среде акцент делается на улучшении алгоритмов. В промышленности акцент делается на том, чтобы эти улучшения были направлены на решение реальной проблемы. Работая в сфере предпринимательства, важно понимать, какие потребности бизнеса нужно удовлетворять. А в области здравоохранения, какие задачи стоят перед врачами в диагностике или лечении, для этого придётся погрузиться в изучение, например, компьютерной томографии или микроскопического исследования тканей.

Коммуникация. Вам нужно будет проходить собеседования. Придётся объяснять концепции ИИ людям, практически не разбирающимся в этой области. Скорее всего, вам придётся работать с командой инженеров, а также со многими другими командами. Общение значительно облегчит всё это. Поэтому есть смысл пройти хотя бы один курс по риторике, актёрскому мастерству или речи.

Быстрое прототипирование — быстрая «черновая» реализация базовой функциональности будущего продукта/изделия, для анализа работы системы в целом. Быстрая итерация идей является обязательным условием для того, чтобы найти ту, которая будет работать. В машинном обучении это относится ко всему: от выбора правильной модели до реализации проекта в целом.

Хорошо, теперь, когда мы разобрались с soft-skills, перейдём к техническим навыкам или hard-skills.

Python является универсальным языком машинного обучения. Возможно, вы уже слышали о Python, даже если ранее не занимались программированием. Python — это язык программирования, который часто используется для создания веб-сайтов и программного обеспечения, автоматизации задач и анализа данных. Python — это язык общего назначения, то есть он может использоваться для создания множества различных программ и не специализирован для решения каких-либо конкретных задач. Эта универсальность, а также удобство для начинающих сделали его одним из самых используемых языков программирования на сегодняшний день. Исследование, проведённое аналитической компанией RedMonk, показало, что в 2022 году этот язык программирования был вторым по популярности среди разработчиков.

C++. Иногда Python бывает недостаточно. Часто вы будете сталкиваться с проектами, в которых необходимо использовать аппаратные средства для повышения скорости. Даже если вы в основном используете Python, знание C++ значительно облегчит использование библиотек Python, повышающих производительность.

Линейная алгебра. Вам потребуется глубокое знакомство с матрицами, векторами и операциями с матрицами.

Математический анализ. Нужно понимать хотя бы основные концепции, такие как производные и интегралы. В противном случае даже такие простые понятия, как градиентный спуск, будут непонятными. Дифференциальные уравнения также полезны для машинного обучения.

Статистика. Статистика в машинном обучении встречается очень часто. Статистика является базовым компонентом аналитики данных и машинного обучения. Она помогает анализировать и визуализировать данные для поиска невидимых закономерностей. Изучение статистики наряду с программированием и математикой является неотъемлемым первым шагом на пути к карьере в области ИИ.

Вся эта математика может поначалу показаться пугающей, если вы долгое время не занимались ею. Да, машинное обучение гораздо более математически сложное занятие, чем, например, фронтенд-разработка. Как и в любом другом навыке, совершенствование математики — это вопрос целенаправленной практики. Однако всё это возможно освоить, даже если вы ранее не изучали математику за рамками школьной программы.

Мы рассмотрели только некоторые, основные навыки, требующиеся для специалиста в области ИИ, в целом — это набор, который подходит и для других специализаций, но есть некоторые отличающиеся требования.

Дата сайентист — собирает данные, анализируют их и получают глубокие знания для самых разных целей. Они используют различные технологические инструменты, процессы и алгоритмы для извлечения знаний из данных и выявления значимых закономерностей. Это может быть как простая задача — выявление аномалий в данных временных рядов, так и сложная — прогнозирование будущих событий и выработка рекомендаций. К основным квалификационным требованиям, предъявляемым к специалисту по анализу данных, относятся:

- Образование в области статистики, математики, информатики и т. д.
- Понимание неструктурированных данных и статистического анализа.

- Опыт работы с облачными инструментами, такими как Amazon S3 и платформа Hadoop.
- Навыки программирования на Python, Perl, Scala, SQL и т. д.
- Знания Hive, Hadoop, MapReduce, Pig, Spark и т. д.

Разработчик бизнес-аналитики

Разработчики бизнес-аналитики (BI) обрабатывают сложные внутренние и внешние данные для выявления тенденций. Например, в компании, предоставляющей финансовые услуги, это может быть человек, отслеживающий данные фондового рынка для принятия инвестиционных решений. В компании, производящей продукцию, это может быть человек, отслеживающий тенденции продаж, чтобы определить стратегию распространения.

Однако, в отличие от аналитика данных, разработчики бизнес-аналитики не создают отчёты самостоятельно. Они обычно отвечают за проектирование, моделирование и поддержание сложных данных в высокодоступных облачных платформах данных, чтобы бизнес-пользователи могли использовать приборные панели. От разработчика BI ожидается следующая квалификация:

Инженер/архитектор больших данных

Инженеры и архитекторы больших данных разрабатывают экосистемы, обеспечивающие эффективное взаимодействие различных бизнес-вертикалей и технологий. По сравнению с учёными, изучающими данные, эта роль может показаться более вовлечённой, поскольку инженеры и архитекторы больших данных обычно занимаются планированием, проектированием и разработкой сред больших данных на базе систем Hadoop и Spark. Ожидается, что инженеры по большим данным будут обладать навыками программирования на C++, Java, Python или Scala. Они также должны иметь опыт в области добычи данных, визуализации данных и миграции данных.

Аналитик данных

Долгое время аналитиком данных называли человека, который собирал, очищал, обрабатывал и анализировал данные, чтобы получить представление о них. По большей части это были рутинные, повторяющиеся задачи. С развитием искусственного интеллекта большая часть рутинной работы была автоматизирована. Поэтому роль аналитика данных была усовершенствована, чтобы присоединиться к новому набору профессий, связанных с ИИ. Сегодня аналитики данных готовят данные для моделей машинного обучения и строят содержательные отчёты на основе полученных результатов.

В результате аналитик данных ИИ должен знать больше, чем просто электронные таблицы. Они должны владеть:

- SQL и другие языки баз данных для извлечения/обработки данных.
- Python для очистки и анализа данных.
- Аналитические панели и инструменты визуализации, такие как Tableau, PowerBI и т. д.
- Бизнес-аналитика для понимания рыночного и организационного контекста.

Ограничения в использовании ИИ

Последнее о чём мы поговорим на сегодняшней лекции — это какие ограничения в использовании ИИ имеются на сегодняшний день. Иными словами: когда имеет смысл использовать ИИ, а когда нет.

Машинное обучение — это не волшебство, а главное отличие ИИ от традиционного программирования заключается в том, что вместо написания программы или алгоритма, мы используем модели. Модель — это тоже программа, но эта программа обучена распознавать определённые закономерности. А как модель учится распознавать закономерности? На основе данных. Данные в машинном обучении — это всё. Без большого количества данных ваш алгоритм не сможет быть обучен или протестирован. Исключением является, например, обучение с подкреплением, о котором мы говорили, однако в реальных бизнес-задачах они имеют ограниченное применение. Если вы владелец компании, типа Яндекс, то обладаете колоссальным количеством данных и имеете возможность использовать машинное обучение для каких-либо целей. Однако, если у вас нет данных, и нет ресурсов, чтобы их собрать или купить, то это может поставить крест на возможности использования ИИ.

На самом деле 80% времени при разработке приложений ИИ уходит на задачи подготовки данных, которые включают сбор, очистку, нормализацию, стандартизацию, формирование. Только после этого вы сможете перейти к крутому моделированию и визуализации на основе искусственного интеллекта.

В сценарии использования нулевая толерантность к ошибкам и нет супервизора.

Почти все системы машинного обучения работают с вероятностями. С какой вероятностью животное на картинке — кошка? С какой вероятностью выбранный маршрут является самым коротким? Какова вероятность того, что завтра пойдёт

дождь, и какова вероятность того, что вам понравится предложенный Яндексом фильм?

Поэтому даже самые лучшие системы ML с точностью более 90 процентов время от времени ошибаются. Такова природа машинного обучения.

Поэтому в прикладных целях вы должны спросить себя, какой ущерб наносит «неточная рекомендация»? Если Яндекс предложит 9 фильмов, которые вы считаете отличными, и один фильм, который вам не нравится — никакого вреда. Точно так же, как и рекомендации продуктов Озон: практически любой рекомендованный продукт, на который вы НЕ нажимаете, является неправильным. Ущерб оттого, что продукты не принимаются? Не очень большой. А вот польза — ещё больше, потому что 25% продаж на Озон приходится именно на эти рекомендации.

Перейдём к случаям использования, где нет права на ошибку. В медицине системы машинного обучения сегодня применяются для помощи в диагностике. И «помощь» — самое важное слово здесь. Медицина — это область, в которой ошибки могут быть фатальными. И всё же системы ИИ используются. Это возможно только потому, что врач ВСЕГДА сохраняет контроль, проверяя и уравнивая рекомендации системы ИИ с другими факторами.

Поэтому если у вас есть сценарий использования, который допускает нулевую терпимость к ошибкам, и у вас нет специалиста, который может проверить (супервизора), то вам лучше не использовать машинное обучение.

Решение, основанное на правилах, работает отлично.

При использовании машинного обучения основное внимание всегда должно быть сосредоточено на добавленной стоимости для достижения конкретной цели. Другими словами, в какой степени новая технология может выполнить поставленные задачи лучше, чем существующие системы. Нет смысла использовать машинное обучение только потому, что оно «в тренде».

Приложения, основанные на правилах, — это системы с правилами, установленными человеком. Классический и простой пример — функция «Если, то...» в Excel: «Если клиент не совершал покупок более 3 месяцев, то пометьте его красным цветом».

Многие компании анализируют свои данные с помощью электронных таблиц Excel, и пока это работает хорошо — то есть служит достижению цели — и не слишком дорого, то нет необходимости в машинном обучении.

Это справедливо и для небольших и молодых компаний, у которых пока не так много данных. В какой-то момент системы, основанные на правилах, достигают своего предела, и тогда машинное обучение становится идеальным решением.

Подведём итоги того, что мы узнали в этом уроке. Мы сделали обзор ключевых технологий искусственного интеллекта. Он включает в себя инфраструктурный уровень и уровень алгоритмов. Затем мы разобрали некоторые из основных компонентов инфраструктурного слоя, а также представили некоторые классические алгоритмы искусственного интеллекта. Наконец, познакомились с некоторыми специализациями в области ИИ, а также ограничениями, связанными с использованием ИИ.

Спасибо за внимание!

Домашнее задание

1. Выберите один реальный пример внедрения ИИ в конкретной отрасли (например, здравоохранение, финансы, розничная торговля и т. д.). Исследуйте и напишите подробное объяснение технологических возможностей, которые сделали это внедрение возможным, и влияние, оказанное на отрасль.
2. Посмотрите лекцию TED [«Искусственный интеллект и технологическая сингулярность»](#) на ютубе, затем поразмышляйте над собственным пониманием технологических возможностей и напишите краткое изложение выступления, выделив основные выводы и то, как они соотносятся с последними технологическими возможностями ИИ.

Что можно почитать ещё?

1. [Квантовый компьютер: что это, как работает, возможности | РБК Тренды.](#)
2. [Самым быстрым суперкомпьютером в мире остаётся Frontier | Издательство «Открытые системы».](#)
3. [80 ГБ памяти, 700 Вт и 60 TFLOPS. Представлен ускоритель Nvidia H100 на основе GPU поколения Hopper.](#)